The 7th International Meeting on Image Formation in X-Ray Computed Tomography

CT MEETING 2022

June 12-16, BALTIMORE, MD, USA

Proceedings

CT MEETING 2022

Sponsors:

JOHNS HOPKINS
BIOMEDICAL ENGINEERING

SIEMENS

UNITED IMAGING

Canon

PHILIPS

The 7th International Conference on Image Formation in X-Ray Computed Tomography is supported by:

Johns Hopkins Biomedical Engineering

Siemens

General Electric

United Imaging

Canon

Philips

We thank our sponsors for their important and valuable contributions!

We would like to thank the members of the scientific committee for their contribution to this conference and their assistance in planning and scientific review.

Scientific Committee

Name	Institution
Adam Wang	Stanford University, USA
Bruno De Man	GE Research, USA
Cristian T. Badea	Duke University, USA
Cyril Riddell	GE Healthcare, France
Emil Y. Sidky	University of Chicago, USA
Frédéric Noo	University of Utah School of Medicine, USA
Ge Wang	Rensselaer Polytechnic Institute, USA
Guang-Hong Chen	University of Wisconsin-Madison, USA
Hengyong Yu	University of Massachusetts Lowell, USA
Hiroyuki Kudo	University of Tsukuba, Japan
Ioannis Sechopoulos	Radboud University Medical Center, The Netherlands
Jeffrey A. Fessler	University of Michigan, USA
Jeffrey H. Siewerdsen	Johns Hopkins University School of Medicine, USA
Jerome Liang	Stony Brook University, USA
Jing Wang	University of Texas Southwestern Medical Center, USA
Jingyan Xu	Johns Hopkins University, USA
Johan Nuyts	KU Leuven, Belgium
John Boone	University of California, Davis, USA
Jongduk Baek	Yonsei University, South Korea
Karl Stierstorfer	Siemens Healthineers, Germany
Katsuyuki Taguchi	Johns Hopkins University School of Medicine, USA
Ke Li	University of Wisconsin-Madison, USA
Ken Sauer	University of Notre Dame, USA
Kirsten Boedeker	Canon Medical Systems Corporation
Lifeng Yu	Mayo Clinic, USA

Marc Kachelrieß	German Cancer Research Center, Germany
Mats Danielsson	KTH Royal Institute of Technology, Sweden
Michael McNitt-Gray	David Geffen School of Medicine at UCLA, USA
Peter B. Noël	University of Pennsylvania, USA
Rongping Zeng	U.S. Food and Drug Administration, USA
Scott Hsieh	Mayo Clinic, USA
Srinivasan Vedantham	University of Arizona, USA
Stanislav Žabic	Philips Healthcare, USA
Stefan Sawall	German Cancer Research Center, Germany
Thomas Koehler	Philips Research Hamburg, Germany
Thomas Koenig	Ziehm Imaging, Germany
Web Stayman	Johns Hopkins School of Medicine, USA
Wenli Wang	Avant Tomography Consulting LLC, USA
Xiaochuan Pan	University of Chicago, USA
Xun Jia	University of Texas Southwestern Medical Center, USA
Yuxiang Xing	Tsinghua University, China
Zhicong Yu	Accuray, USA
Zhou Yu	Canon Medical Research Institute, USA
Zhye Yin	GE Research, USA

Conference Chair

Name	Institute
J. Webster Stayman	Johns Hopkins University

Organization Committee

Name	Institute
Grace J. Gang	Johns Hopkins University
Junyuan Li	Johns Hopkins University
Quinn Ma	Johns Hopkins University
Alex Sisniega	Johns Hopkins University
Ken Taguchi	Johns Hopkins University
Wenying Wang	Johns Hopkins University
Wojtek Zbijewski	Johns Hopkins University
Yijie Yuan	Johns Hopkins University

Schedule CT Meeting 2022

	Sunday *June 12*	**Monday** *June 13*	**Tuesday** *June 14*	**Wednesday** *June 15*	**Thursday** *June 16*
7:00		Breakfast	Breakfast	Breakfast	Breakfast
7:20					
7:40					
8:00					
8:20		Welcome + Novel CT Technologies (3 Talks)	Invited Talks on Photon Counting CT	Modeling and Assessment (4 Talks)	Invited Talk on Interventional CT
8:40					
9:00					Interventional Imaging (2 Talks)
9:20					
9:40		Coffee	Coffee	Coffee	Coffee
10:00		Reconstruction and Deep Learning (5 Talks)	Spectral CT (5 Talks)	Invited Talk on Deep Learning	Cardiac CT and Motion Compensation (4 Talks) + Conclusion
10:20					
10:40				Deep Learning Assessment (3 Talks)	
11:00					
11:20					
11:40		Lunch	Lunch	Lunch	Lunch
12:00					
12:20					
12:40					
1:00					
1:20		Monday Poster Session (~25 Posters)	Tuesday Poster Session (~25 Posters)	Wednesday Poster Session (~25 Posters)	Optional Tours JHU Laboratories and Hospital
1:40					
2:00					
2:20					
2:40					
3:00		Coffee	Coffee	Coffee	
3:20	Registration	CT Acquisition (5 Talks)	Artifacts and Sparse CT (5 Talks)	Spectral and Polyenergetic CT Reconstruction (5 Talks)	
3:40					
4:00					
4:20					
4:40					
7:00	Welcome Reception (Great Hall)	Dinner (Charles Commons Banquet Room)	Dinner Out (R. House)	Dinner Out (Ministry of Brewing)	
7:20					
7:40					
8:00					
8:20					
8:40					
9:00					

Monday, June 13

Welcome Address 08:20 – 08:40 Web Stayman

Oral Session : Novel CT Technologies
Time : 08:40 – 09:40
Moderators : Bruno De Man, Ke Li

Coffee Break 09:40 – 10:00

Oral Session : Reconstruction and Deep Learning
Time : 10:00 – 11:40
Moderators : Marc Kachelrieß, Koen Michielsen

This session is made possible by a generous gift from United Imaging Healthcare

Lunch 11:40 – 13:20

Poster Session 13:20 – 15:00

The poster session will begin with 30-second teaser presentation given by presented in **the Glass Pavilion**. General presentation and discussion of posters will follow in **the Great Hall.**

M19	Deep learning enabled wide-coverage high-resolution cardiac CT	Tzu-Cheng Lee, Jian Zhou, John Schuzer, Masakazu Matsuura, Takuya Nemoto, Hiroki Taguchi, Zhou Yu, Liang Cai	142
M20	Preliminary study on image reconstruction for limited-angular-range dual-energy CT using two-orthogonal, overlapping arcs	Buxin Chen, Zheng Zhang, Dan Xia, Emil Y. Sidky, and Xiaochuan Pan	145
M21	Correcting spurious signal using an automated Deep Learning based reconstruction workflow	Matthew Andrew, Andriy Andreyev, Faguo Yang and Lars Omlor	149
M22	Dual-Energy Head Cone-Beam CT Using a Dual-Layer Flat-Panel Detector: Physics-Based Material Decomposition	Zhilei Wang, Hao Zhou, Shan Gu, Hewei Gao	157
M23	Combining Deep Learning and Adaptive Sparse Modeling for Low-dose CT Reconstruction	Ling Chen, Zhishen Huang, Yong Long, Saiprasad Ravishankar	153
M24	X-ray Dissectography Enables Stereotography	Chuang Niu and Ge Wang	161
M25	Mixed coronary plaque characterization with the first clinical dual-source photon-counting CT scanner a phantom study	Thomas Wesley Holmes, Leening P. Liu, Nadav Shapira, Elliot McVeigh, Amir Pourmorteza, Peter B. Noël	165

Coffee Break 15:00 – 15:20

Oral Session : CT Acquisition
Time : 15:20 – 17:00
Moderators : Adam Wang, Rolf Clackdoyle

Time	Title	Authors	Page
15:20 – 15:40	LaBr3:Ce and Silicon Photomultipliers: Towards the Optimal Scintillating Photon-Counting Detector	Stefan J. van der Sar, David Leibold, Stefan E. Brunner, and Dennis R. Schaart	49
15:40 – 16:00	Preliminary Investigations of a Novel Dynamic CT Collimator	Web Stayman, Nir Eden, Yiqun Q. Ma, Grace J. Gang, Allon Guez	53
16:00 – 16:20	X-ray CT Data Completeness Condition for Sets of Arbitrary Projections	Gabriel Herl, Andreas Maier, and Simon Zabler	66
16:20 – 16:40	CT imaging with truncation data over limited-angular ranges	Dan Xia, Zheng Zhang, Buxin Chen, Emil Y. Sidky, and Xiaochuan Pan	57
16:40 – 17:00	Cone Beam Field of View Extension through Complementary Short Scan Trajectories with Displaced Center of Rotation	Gabriele Belotti, Simon Rit, Guido Baroni	62

Dinner 19:00 – 21:20
(Charles Commons Banquet Room)

Tuesday, June 14

Oral Session : Invited Talks on Photon Counting CT
Time : 08:20 – 09:40
Moderator : Web Stayman

Time	Title	Presenter
08:20 – 09:00	Photon counting detector computed tomography: technical background	Peter Noel
09:00 – 09:40	Photon counting detector computed tomography: clinical applications	Shuai Leng

Coffee Break 09:40 – 10:00

Oral Session : Spectral CT
Time : 10:00 – 11:40
Moderators : Kevin Brown, Cristian Badea

This session is made possible by a generous gift from Philips Healthcare

Time	Title	Authors	Page
10:00 – 10:20	Consistency-based auto-calibration of the spectral model in dual-energy CT	Jérome Lesaint, Simon Rit	170
10:20 – 10:40	Direct binning for photon counting detectors	Katsuyuki Taguchi and Scott S. Hsieh	174
10:40 – 11:00	Co-clinical photon counting CT research for multi-contrast imaging	Cristian T. Badea, Darin P. Clark, Alex Allphin, Juan Carlos Ramirez-Giraldo, Prajwal Bhandari, Yvonne M. Mowery, Ketan B. Ghaghada	178
11:00 – 11:20	Reproducibility in dual energy CT: the impact of a projection domain material decomposition method	Viktor Haase, Frederic Noo, Karl Stierstorfer, Andreas Maier, and Michael McNitt-Gray	182
11:20 – 11:40	Dual-source photon-counting CT: consistency in spectral results at different acquisition modes and heart rates	Leening P. Liu, Nadav Shapira, Pooyan Sahbaee, Harold I. Litt, Marcus Y. Chen, Peter B. Noël	186

Lunch 11:40 – 13:20

Poster Session 13:20 – 15:00

The poster session will begin with 30-second teaser presentation given by presented in **the Glass Pavilion**. General presentation and discussion of posters will follow in **the Great Hall.**

T18	Self-trained Deep Convolutional Neural Network for Noise Reduction in CT	Zhongxing Zhou, Akitoshi Inoue, Cynthia McCollough, and Lifeng Yu	279
T19	2D-3D motion registration of rigid objects within a soft tissue structure	Nargiza Djurabekova, Andrew Goldberg, David Hawkes, Guy Long, Felix Lucka and Marta M. Betcke	283
T20	Gas Bubble Motion Artifact Reduction through Simultaneous Motion Estimation and Image Reconstruction	Kai Wang, Hua-Chieh Shao, You Zhang, Chunjoo Park, Steve Jiang, Jing Wang	288
T21	Comparing One-step and Two-step Scatter Correction and Density Reconstruction in X-ray CT	Alexander N. Sietsema, Michael T. McCann, Marc L. Klasky, and Saiprasad Ravishankar	292
T22	Material decomposition from unregistered dual kV data using the cOSSCIR algorithm	Benjamin M. Rizzo, Emil Y. Sidky, and Taly Gilat Schmidt	296
T23	PixelPrint: Three-dimensional printing of patient-specific soft tissue and bone phantoms for CT	Kai Mei, Michael Geagan, Nadav Shapira, Leening P. Liu, Pouyan Pasyar, Grace J. Gang, Web Stayman, and Peter B. Noël	300
T24	Practical Workflow for Arbitrary Non-circular Orbits for CT with Clinical Robotic C-arms	Yiqun Ma, Grace J. Gang, Tess Reynolds, Tina Ehtiati, Junyuan Li, Owen Dillon, Tom Russ, Wenying Wang, Clifford Weiss, Nicholas Theodore, Kelvin Hong, Ricky O'Brien, Jeffrey Siewerdsen, Web Stayman	304
T25	Rigid motion correction based on locally linear embedding for helical CT scans with photon-counting detectors	Mengzhou Li, Chiara Lowe, Anthony Butler, Phil Butler, and Ge Wang	308

Coffee Break 15:00 – 15:20

Oral Session : Artifacts and Sparse CT
Time : 15:20 – 17:00
Moderators : Xiaochuan Pan, Jerome Z. Liang

Time	Title	Authors	Page
15:20 – 15:40	Deep Scatter Estimation for Coarse Anti-Scatter Grids as used in Photon-Counting CT	Julien Erath, Jan Magonov, Joscha Maier, Eric Fournié, Martin Petersilka, Karl Stierstorfer, and Marc Kachelrieß	190
15:40 – 16:00	Cross-Domain Metal Segmentation for CBCT Metal Artifact Reduction	Maximilian Rohleder, Tristan M. Gottschalk, Andreas Maier and, Bjoern W. Kreher	194
16:00 – 16:20	Sparsier2Sparse: Weakly-supervised learning for streak artifacts reduction with unpaired sparse view CT data	Seongjun Kim, Byeongjoon Kim, and Jongduk Baek	198
16:20 – 16:40	Dual Domain Closed-loop Learning for Sparse-view CT Reconstruction	Yi Guo, Yongbo Wang, Manman Zhu, Dong Zeng, Zhaoying Bian, Xi Tao and Jianhua Ma	202
16:40 – 17:00	Hybrid Reconstruction Using Shearlets and Deep Learning for Sparse X-Ray Computed Tomography	Andi Braimllari, Theodor Cheslerean-Boghiu, Tobias Lasser	206

Dinner 19:00 – 21:20
(R. House)

Wednesday, June 15

Oral Session : Modeling and Assessment
Time : 08:20 – 09:40
Moderators : Grace J. Gang, Kirsten L. Boedeker

Time	Title	Authors	Page
08:20 – 08:40	Trade-offs between redundancy and increased rank for tomographic system matrices	Feriel Khellaf and Rolf Clackdoyle	414
08:40 – 09:00	Stationary X-ray Tomography for Hemorrhagic Stroke Imaging - Sampling and Resolution Properties	A. Lopez-Montes, T. McSkimming, W. Zbijewski, J. H. Siewerdsen, A. Skeats, B. Gonzales, A. Sisniega	418
09:00 – 09:20	Angular normalized glandular dose coefficient in breast CT: clinical data study	Hsin Wu Tseng, Andrew Karellas, and Srinivasan Vedantham	422
09:20 – 09:40	Estimating the accuracy and precision of quantitative imaging biomarkers as endpoints for clinical trials using standard-of-care CT	Paul Kinahan, Darrin Byrd, Hao Yang, Hugo Aerts, Binzhang Zhao, Andrey Fedorov, Lawrence Schwartz, Tavis Allison, Chaya Moskowitz	426

Coffee Break 09:40 – 10:00

Oral Session : Invited Talk on Deep Learning
Time : 10:00– 10:40
Moderator : Web Stayman

Time	Title	Presenter
10:00 – 10:40	Hallucinations and objective assessments of deep learning technologies for image formation	Mark Anastasio

Oral Session : Deep Learning Assessment
Time : 10:40 – 11:40
Moderators : Saiprasad Ravishankar, Rongping Zeng

Time	Title	Authors	Page
10:40 – 11:00	Reconstructing Invariances of CT Image Denoising Networks using Invertible Neural Networks	Elias Eulig, Björn Ommer, and Marc Kachelrieß	430
11:00 – 11:20	Local Linearity Analysis of Deep Learning CT Denoising Algorithms	Junyuan Li, Wenying Wang, Matt Tivnan, Jeremias Sulam, Jerry L Prince, Michael McNitt-Gray, Web Stayman and Grace J. Gang	434
11:20 – 11:40	Evaluation of deep learning-based CT reconstruction with a signal-Laplacian model observer	Gregory Ongie, Emil Y. Sidky, Ingrid S. Reiser, & Xiaochuan Pan	438

Lunch 11:40 – 13:20

Poster Session 13:20 – 15:00

The poster session will begin with 30-second teaser presentation given by presented in **the Glass Pavilion**. General presentation and discussion of posters will follow in **the Great Hall.**

W21	Iodine-enhanced Liver Vessel Segmentation in Photon Counting Detector-based Computed Tomography using Deep Learning	Sumin Baek, Okkyun Lee, and Dong Hye Ye	394
W22	Optimization of Empirical Beamhardening Correction Algorithm	Andriy Andreyev, Faguo Yang, Lars Omlor, and Matthew Andrew	402
W23	Deep Learning-based Prior toward Normalized Metal Artifact Reduction in Computed Tomography	Jeonghyeon Nam, Dong Hye Ye, and Okkyun Lee	398
W24	On use of augmentation for the DNN-based CT	Prabhat Kc, Kyle J. Myers, M. Mehdi Farhangi, Rongping Zeng	406
W25	Joint Multi-channel Total Generalized Variation Minimization and Tensor Decomposition for Spectral CT Reconstruction	Huihua Kong, Xiangyuan Lian, Jinxiao Pan, and Hengyong Yu	410

Coffee Break 15:00 – 15:20

Oral Session : Spectral and Polyenergetic CT Reconstruction
Time : 15:20 – 17:00
Moderators : Emil Sidky, Johan Sunnegaardh

Time	Title	Authors	Page
15:20 – 15:40	Tunable Neural Networks for Multi-Material Image Formation from Spectral CT Measurements	Matthew Tivnan, Grace Gang, Peter Noël, Jeremias Sulam, and J. Webster Stayman	442
15:40 – 16:00	Self-supervised nonlocal spectral similarity induced material decomposition network	Lei Wang, Yongbo Wang, Zhaoying Bian, Dong Zeng, and Jianhua Ma	446
16:00 – 16:20	Likelihood-based bilateral filtration in material decomposition for photon counting CT	Okkyun Lee	450
16:20 – 16:40	Experimental Evaluation of Polychromatic Reconstruction for Quantitative CBCT	Michał Walczak, Pascal Paysan, Mathieu Plamondon, Stefan Scheib	454
16:40 – 17:00	Dual-energy cone-beam CT with three-material decomposition for bone marrow edema imaging	Stephen Z. Liu, Magdalena Herbst, Thomas Weber, Sebastian Vogt, Ludwig Ritchl, Steffen Kappler, Jeffrey H. Siewerdsen, and Wojciech Zbijewski	458

Dinner 19:00 – 21:20
(Ministry of Brewing)

Thursday, June 16

Oral Session : Invited Talk on Interventional CT
Time : 08:20 – 09:00
Moderator : Web Stayman

Time	Title	Presenter
08:20 – 09:00	Engineering the Future of Spine Surgery	Nick Theodore

Oral Session : Interventional Imaging
Time : 09:00 – 09:40
Moderator : Cyrill Riddell

Time	Title	Authors	Page
09:00 – 09:20	Real-time Liver Tumor Localization via a Single X-ray Projection Using Deep Graph Network-assisted Biomechanical Modeling	Hua-Chieh Shao, Jing Wang, and You Zhang	464
09:20 – 09:40	3D Reconstruction of Stents and Guidewires in an Anthropomorphic Phantom From Three X-Ray Projections	Tim Vöth, Thomas König, Elias Eulig, Michael Knaup, Veit Wiesmann, Klaus Hörndler, and Marc Kachelrieß	468

Coffee Break 09:40 – 10:00

Oral Session : Cardiac CT and Motion Compensation
Time : 10:00 – 11:20
Moderators : Ken Taguchi, Simon Rit

Time	Title	Authors	Page
10:00 – 10:20	Context-Aware, Reference-Free Local Motion Metric for CBCT Deformable Motion Compensation	H. Huang, J.H. Siewerdsen, W. Zbijewski, C.R. Weiss, M. Unberath, and A. Sisniega	472
10:20 – 10:40	Simulation of Random Deformable Motion in Soft-Tissue Cone-Beam CT with Learned Models	Y. Hu, H. Huang, J. H. Siewerdsen, W. Zbijewski, M. Unberath, C. R. Weiss, and A. Sisniega	476
10:40 – 11:00	A five-dimensional cardiac CT model for generating virtual CT projections for user-defined bolus dynamics and ECG profiles	Eri Haneda, Bernhard Claus, Jed Pack, Darin Okerlund, Albert Hsiao, Elliot McVeigh, and Bruno De Man	480
11:00 – 11:20	A Virtual Imaging Trial Framework to Study Cardiac CT Blooming Artifacts	Ying Fan, Jed Pack, and Bruno De Man	484

Conclusion 11:20 – 11:40 Web Stayman

Lunch 11:40 – 13:20

Optional Tours 13:20 – 15:00 (JHU Laboratories and Hospital)

Monday, June 13

Novel CT Technologies

Reconstruction and Deep Learning

Monday Poster Session

CT Acquisition

Dark-Field Imaging on a Clinical CT System: Realization of Talbot-Lau Interferometry in a Gantry

Manuel Viermetz, Nikolai Gustschin, Clemens Schmid, Jakob Haeusele,
Roland Proksa, Thomas Koehler, and Franz Pfeiffer

***Abstract*—Computed tomography (CT) is a foundation of modern clinical diagnostics but it presently only retrieves information from X-rays attenuation. However, it is known that micro structural texture or porosity – which is well below the spatial resolution of CT – can be revealed by grating-based dark-field imaging. Diagnostic value of this sub-resolution tissue information has been demonstrated in pre-clinical studies on small-animal disease models and recently also in a first clinical radiography system [1]. These studies show that dark-field imaging is particularly useful for early detection and staging of lung diseases.**

While dark-field CT is regularly realized in laboratory environment, the transfer to human scale and bringing it to clinical application poses several technical challenges. Switching from a step-and-shoot acquisition to a mode where the gantry and acquisition operate continuously as well as reducing scan times to below seconds and ensuring stability against vibrations are key concerns when it comes to the translation of the established laboratory dark-field technology to full-body medical CT. In [2], we recently demonstrated the first dark-field CT implementation, which collectively solves these roadblocks and therefore is a milestone in the development of clinical CT imaging. The prototype we present allows to reconstruct the attenuation and dark-field channels of a human thorax phantom from a one second long acquisition and covers a 45 cm diameter field of view.

In this work, we present how the first dark-field CT prototype works and focus particularly on the technical design, optimized design of the gratings for CT application and the first characterization of the interferometer in the rotating gantry. We discuss which steps where particularly important for the realization and where we see potential for further improvements. These results provide key insights for future dark-field CT implementations.

***Index Terms*—X-ray Imaging, Dark-field contrast, Computed Tomography, Talbot-Lau interferometer**

I. Introduction

X-RAY Computed Tomography (CT) is a well-established technique, which is – in its conventional form – limited to attenuation contrast. There it allows high resolution imaging, fast acquisition times, and decent contrast for many purposes such as material research, inspection, and medical diagnostics. Throughout the last decade Talbot-Lau interferometry, which started as a synchrotron-only technology, gained attention as it unlocks additional contrast channels [3], [4]. A Talbot-Lau interferometer, as illustrated in Fig. 1, allows simultaneous imaging of attenuation, refraction, and small-angle scattering of the transmitted object [5], [6].

Particularly the access to sub-resolution tissue properties by analyzing the small-angle scattering power – also referred to as dark-field signal – of the material is a promising application in biomedical imaging. Previous work at laboratory setups demonstrated that the dark-field modality is useful for lung diagnostics as it enables the detection of micro-structural changes in lung parenchyma [7]–[10].

Recently the first clinical study on dark-field chest radiography of COPD and COVID-19 patients validated these pre-clinical results and clearly showed the potential of this additional image contrast [1], [11]. This first in-vivo human

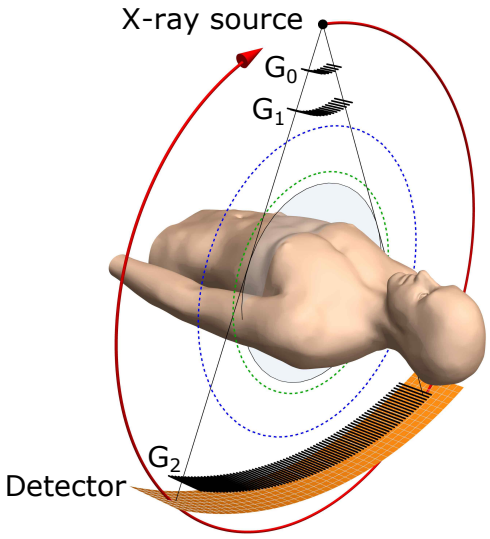


Fig. 1. Talbot-Lau interferometer in a CT gantry. It consists of three cylindrically bent gratings illustrated as line patterns, referred to as G_0, G_1, and G_2, which structure the X-ray beam as they absorb or phase-shift the radiation. By the Talbot effect a self-image of G_1 appears at the location of G_2. The object induces attenuation, refraction, and small-angle scattering, which subsequently change this self-image characteristically. Using the Moiré effect and the G_2 grating which has a period matching to the G_1 self-image these distortions can be retrieved with a common X-ray detector. Since this approach relies on coherent radiation, a G_0 source-grating is required to convert the incoherent source spot into a set of sufficiently coherent slit sources.

M. Viermetz, N. Gustschin, C. Schmid, J. Haeusele, and F. Pfeiffer are with the Chair of Biomedical Physics, Department of Physics, Technical University of Munich, 85748 Garching, Germany. (e-mail: contact.viermetz@tum.de)

M. Viermetz, N. Gustschin, C. Schmid, J. Haeusele, and F. Pfeiffer are with the Munich Institute of Biomedical Engineering, Technical University of Munich, 85748 Garching, Germany.

R. Proksa, and T. Koehler are with Philips Research, 22335 Hamburg, Germany.

T. Koehler and F. Pfeiffer are with the Institute for Advanced Study, Technical University of Munich, 85748 Garching, Germany.

F. Pfeiffer is with the Department of Diagnostic and Interventional Radiology, School of Medicine and Klinikum rechts der Isar, Technical University of Munich, 81675 Munich, Germany.

Manuscript was compiled on January 24, 2022.

dark-field imaging system can now retrieve otherwise undetectable changes in the porous lung tissue which are correlated with diseases progressions. However, this chest radiography provides no 3D information and anatomical structures overlap in image domain. To overcome these restrictions our approach is to install the interferometer into a clinical CT gantry. However, this step involves a variety of problems since the space on a clinical CT gantry is limited, an extremely large field of view must be covered by the gratings (compared to current state of the art dark-field implementations), and vibrations of the system can quickly degrade the measured signal. In this work we discuss our design of a Talbot-Lau interferometer which has been integrated into a Brilliance iCT (Royal Philips, The Netherlands) and produces the first dark-field CT reconstructions of human sized phantoms [2].

II. Talbot-Lau interferometer layout

In our presented design the most important design rules have been (1) to maintain the bore diameter of the conventional CT, which is 70 cm (blue dashed line in Fig. 1), (2) to not modify the conventional gantry and its primary imaging components, i.e., source and detector, and (3) to have at least 45 cm coverage which is 90% of a conventional CT which has an diameter of 50 cm (highlighted area and green dashed line in Fig. 1, respectively).

The presented Talbot-Lau interferometer consists of three gratings and must be optimized for the 80 kVp spectrum, which is the lowest available energy on the iCT platform. As this method only works for coherent radiation a source grating G_0 is required [12]. It splits the radiation from the incoherent X-ray source into several slit sources which fulfill the coherence requirement for the length of the interferometer setup. Because of the weak interaction between hard X-rays and grating the aspect ratio of the structures must be high – which is difficult to fabricate – and gold as a good absorber is used as the grating material. The grating G_1 introduces a fine intensity modulation on the incident radiation, which is basically a fine line pattern with a periodicity of a few micrometers. Attenuation of the radiation by the object causes a decrease of intensity, whereas refraction and small-angle scattering induce small distortions of the pattern. To resolve these small changes an analyzer grating G_2 is positioned in front of the detector and makes use of the Moiré effect to translate the changes to an intensity pattern which can be measured by the detector.

The grating periods and distances define the sensitivity of the system to small-angle scattering. Basically, smaller periods and longer distances lead to a more sensitive setup. Note that a too small sensitivity results in a poor signal to noise ratio and a too high sensitivity can lead to artefacts due visibility starvation. From previous work in [13], [14] a sensitivity range that is meaningful for lung imaging can be derived. To achieve a sufficient sensitivity the position and the period of the grating structures must be selected carefully. Certainly higher sensitivity is preferred, however, there are limitations in the grating fabrication and the available space on the CT gantry which limit the maximum sensitivity of our prototype.

In Fig. 2 two sketches of different geometries illustrate possible implementations of the three gratings into a CT gantry. Increasing inter-grating distances as well as decreasing the grating periods improve sensitivity. Evidently, it is therefore a good design choice to maximize the distances within the constraints given by the CT platform [15]. Consequently G_0 and G_2 are positioned as close to the X-ray source and the detector, respectively, and G_1 as close to the bore as possible. The inverse geometry is advantageous for our CT implementation as it requires only one large coverage G_2 with relatively coarse period and the gratings G_0 and G_1 are small enough to be manufactured as single parts.

III. Integration into the CT gantry

In our implementation we position a combined mount which holds G_0 and G_1 into the collimator box which is positioned downstream the X-ray source before the patient. This assembly

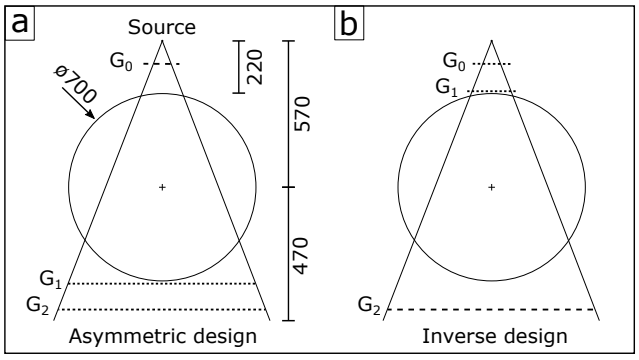


Fig. 2. Asymmetric and inverse interferometer geometry. **a,** in an asymmetric design the gratings G_1 and G_2 are both behind the patient and must cover a large area due to the large field of view. Furthermore, the periods of these two gratings are rather fine, compared to the G_0 grating. **b,** in the inverse geometry, G_0 and G_1 are in front of the patient, their size is relatively small due to the strong divergence of the setup, and the period of the largest grating G_2 is coarse, which makes fabrication easier. Another advantage of this design is its dose efficiency because G_1 is positioned in front of the patient.

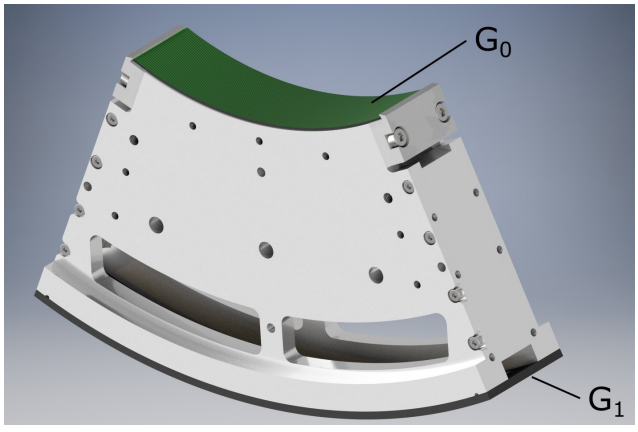


Fig. 3. G_0 and G_1 bending frame for CT implementation. The two gratings are rigidly mounted on a machined structure which ensures a precise inter-grating distance and bends the gratings cylindrically. This focuses the trenches between the grating lamellae towards the X-ray source spot to compensate for the strongly divergent radiation and reduce shadowing artefacts.

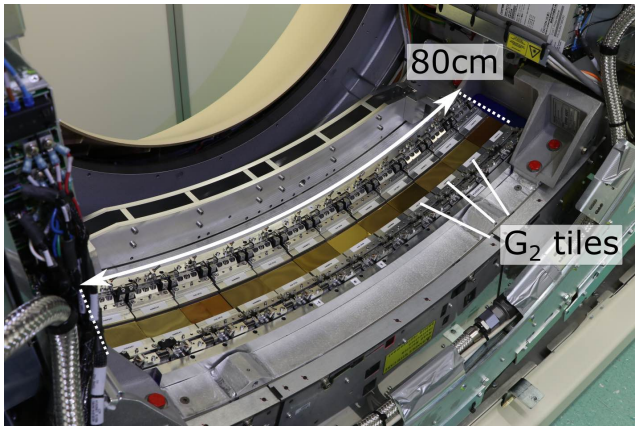


Fig. 4. Modular G_2 implementation covering 80 cm width. The specialized mount allows adjustment of each G_2 tiles individually to minimize the gap between the tiles and to ensure parallel alignment of the G_2 lamellae to the line pattern produced by the G_1. Again, each grating tile is bent cylindrically and also the combined assembly follows this curvature to focus into the source spot, which reduces shadowing artefacts.

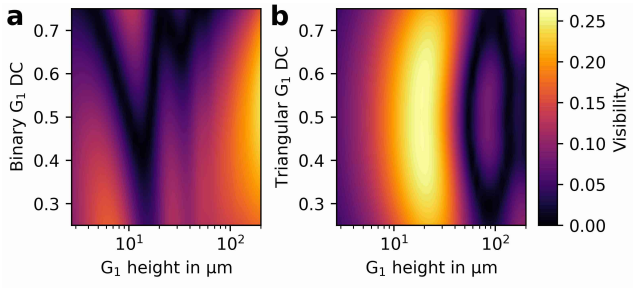


Fig. 5. Wave optical simulation results of the interferometer for different G_1 parameters (using gold as a grating material). **a,** shows the performance for a conventional Talbot-Lau interferometer with rectangular or binary grating profile. Here we find no significant performance in the height range typical for phase-shifting G_1 (below 25 µm) merely a design with an absorption G_1 (height exceeding 200 µm) would give good performance. **b,** only for a triangular G_1 grating profile with a height of 18 µm a good performance exceeding 26% visibility is found.

is shown in Fig. 3 where the two gratings are cylindrically bent to focus towards the X-ray source spot of the CT. This is an important design characteristic since grating G_0 exceeds an aspect ratio of 100 (relation between height and width of the lamellae). In an un-bent state, a significant fraction of the radiation would traverse the grating not with a perpendicular incident angle because of the strong divergence. This would lead to a degradation of the slit-sources by shadowing, which should be generated by the G_0, and consequently the performance of the interferometer vanishes.

The large G_2 grating must cover about 80 cm of arc length which is realized by combining several smaller tiles to one large grating. We developed a specialized mount which allows us to adjust the tiles inside the gantry individually and which additionally bends each tile to focus towards the source spot. It is shown in Fig. 4 covering 90% of the CT detector columns and 32 pixel rows which leads to a total coverage of 20 mm in the iso-center and a reconstructible volume diameter of 450 mm.

IV. Optimization of the Grating Parameters

Based on the state of the art fabrication technology for X-ray absorption gratings, which are compatible with the 80 kVp spectrum used in our dark-field CT prototype and a ballpark sensitivity estimation that shows that we will not reach the upper sensitivity limit for human lung imaging where signal saturation occurs, the period of G_0 is fixed to 4.8 µm. From this parameter and the introduced inverse interferometer geometry we derive the optimal grating parameters using a wave propagation simulation. It is similar to previous work by [16] and includes the geometrical effects of the diverging beam, the selected X-ray source spot, the clinical 80 kVp spectrum, its filtration, and the imperfections in the extended gratings, i.e. bridges in the layout.

Apart from the sensitivity (which is defined by distances and periods), another key parameter of the interferometer is the so-called fringe visibility, which directly translates into signal to noise ratio of the dark-field image [17]. This visibility is strongly influenced by the height, shape, and duty cycle of the grating G_1. In Fig. 5 simulation results of these free parameters reveal significant performance differences, particularly for two different G_1 grating profiles. We find that the best system performance can be achieved with a triangular shaped G_1 grating profile [18]. The elegance of a triangular G_1 profile is that it requires only about 18 µm high structures and thus, X-ray flux through this grating remains high whereas the alternative – an absorption G_1 – would reduce the flux at least by 50%. Furthermore, an absorption G_1 would require a high aspect ratio which makes its fabrication difficult.

The dark-field CT prototype is equipped with a triangular G_1 on a flexible polyimide substrate. The flexibility of the substrate enables us to bend the structure as discussed earlier and ensures that the performance is not degraded, e.g. by scattering of radiation in the substrate material. G_0 and G_2 use graphite substrates which are also flexible for bending to the required radii.

V. Performance of the Talbot-Lau Interferometer

The design utilizes the standard iCT detector and can be rotated with rotation times between 0.27 and 1.5 s which are the standard settings for this clinical CT model. From the centrifugal acceleration and the vibrations induced by the X-ray tube as well as other sources of instabilities the interferometer is periodically leading to a corresponding distortion of the fringe pattern.. Fortunately, the frame rate of the detector is high enough to resolve the distortions with high accuracy, thus, with a sophisticated processing framework the attenuation and dark-field signal can be extracted from the measurements.

The visibility of the presented interferometer, which is the most important measure for the performance of a Talbot-Lau interferometer, is around 22% for the central 50% of the field of view. Fig. 6 shows the flat-field performance of the interferometer where a characteristic reduction towards the left and right is induced by remaining partial shadowing in G_0 due to the large source spot. For larger fan angles, the width of

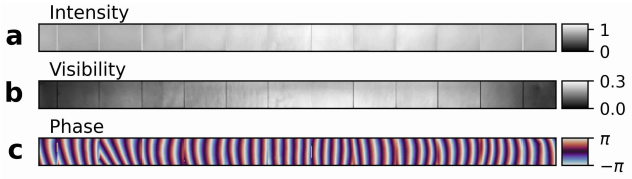


Fig. 6. Performance evaluation of the Talbot-Lau interferometer during rotation. **a,** the intensity has a characteristic maximum in the center and decreases to the left and right due to partial shadowing in the G_0. **b,** the system visibility is highest in the center exceeding 26%, but due to partial shadowing there is a drop to the left and right. **c,** the interferometer phase is optimized for about 10 pixels per fringe for ideal phase sampling during processing and reconstruction.

the X-ray source spot appears to be bigger, because the actual source area is an inclined plane on the anode target surface. Only under the projection through the iso-centre the source spot is smallest.

VI. Conclusion

We conclude that the presented design is most suitable for clinical dark-field CT implementation within the limitations of current grating fabrication and the geometrical boundary conditions imposed by the desire to use a conventional clinical CT as a platform for the prototype. We analyzed the impact of the most important design parameters to identify potential problems and could verify that a triangular grating profile for G_1 has the best performance, besides an absorption G_1.

In our implementation shadowing is avoided by cylindrically bending all gratings to focus into the source spot. Nevertheless, partial shadowing in the G_0 due to the extended source causes a visibility decrease towards larger fan-angles.

With this proposed design, we demonstrate that implementation of a Talbot-Lau interferometer into a clinical CT is feasible with state of the art gratings. This development now brings dark-field CT an important step closer to the clinics as it demonstrates that up-scaling to human sized field of views works and sub-second scan times are now easily possible.

Acknowledgment

The authors wish to thank Julia Herzen, Daniela Pfeiffer, Alexander Fingerle, Ernst Rummeny, Pascal Meyer, Jürgen Mohr, Maximilian von Teuffenbach, Amanda Pleier, Michael Heider, Sven Prevrhal, Thomas Reichel, Stefan Löscher, Frank Bergner, Ami Altman, and Shlomo Gotman for their help, support, and dedication to launch this complex project. This work was carried out with the support of the Karlsruhe Nano Micro Facility (KNMF, www.kit.edu/knmf), a Helmholtz Research Infrastructure at Karlsruhe Institute of Technology (KIT). We acknowledge the support of the TUM Institute for Advanced Study, funded by the German Excellence Initiative, the European Research Council (ERC, H2020, AdG 695045) and Philips GmbH Market DACH.

References

[1] K. Willer *et al.*, "X-ray dark-field chest imaging for detection and quantification of emphysema in patients with chronic obstructive pulmonary disease: A diagnostic accuracy study," *The Lancet Digital Health*, vol. 3, e733–e744, 2021.

[2] M. Viermetz *et al.*, "Darkfield computed tomography reaches the human scale," accepted.

[3] F. Pfeiffer *et al.*, "Phase retrieval and differential phase-contrast imaging with low-brilliance x-ray sources," *Nature Physics*, vol. 2, pp. 258–261, 4 2006.

[4] T. Weitkamp *et al.*, "Tomography with grating interferometers at low-brilliance sources," International Society for Optics and Photonics, SPIE, 2006, pp. 249 –258.

[5] F. Pfeiffer *et al.*, "Hard-x-ray dark-field imaging using a grating interferometer," *Nature Materials*, vol. 7, pp. 134–137, 2 2008.

[6] M. Bech *et al.*, "Quantitative x-ray dark-field computed tomography," *Physics in Medicine and Biology*, vol. 55, pp. 5529–5539, 18 2010.

[7] F. G. Meinel *et al.*, "Improved diagnosis of pulmonary emphysema using in vivo dark-field radiography," *Investigative Radiology*, vol. 49, 2014.

[8] K. Hellbach *et al.*, "In vivo dark-field radiography for early diagnosis and staging of pulmonary emphysema," *Invest Radiol*, vol. 50, 7 2015.

[9] A. Velroyen *et al.*, "Grating-based x-ray dark-field computed tomography of living mice," *EBioMedicine*, vol. 2, pp. 1500–1506, 10 2015.

[10] P. Modregger *et al.*, "Small angle x-ray scattering with edge-illumination," *Scientific Reports*, vol. 6, 2016.

[11] M. Frank *et al.*, "Dark-field Chest X-ray Imaging for the Assessment of COVID-19-Pneumonia," *submitted*,

[12] F. Pfeiffer *et al.*, "Hard x-ray phase tomography with low-brilliance sources," *Physical Review Letters*, vol. 98, p. 108 105, 10 2007.

[13] W Yashiro *et al.*, "On the origin of visibility contrast in x-ray Talbot interferometry," *Optics Express*, vol. 18, p. 16 890, 2010.

[14] V. Ludwig *et al.*, "Exploration of different x-ray Talbot-Lau setups for dark-field lung imaging examined in a porcine lung," *Physics in Medicine and Biology*, vol. 64, 6 2019.

[15] T. Donath *et al.*, "Inverse geometry for grating-based x-ray phase-contrast imaging," *Journal of Applied Physics*, vol. 106, 5 2009.

[16] A. Ritter *et al.*, "Simulation framework for coherent and incoherent X-ray imaging and its application in Talbot-Lau dark-field imaging," *Optics Express*, vol. 22, p. 23 276, 2014.

[17] T. Weber *et al.*, "Noise in x-ray grating-based phase-contrast imaging," *Medical Physics*, vol. 38, pp. 4133–4140, 7 2011.

[18] A. Yaroshenko *et al.*, "Non-binary phase gratings for x-ray imaging with a compact Talbot interferometer," *Optics Express*, vol. 22, pp. 547–556, 1 2014.

Dark-Field Imaging on a Clinical CT System: Performance and Potential based on first Results

Nikolai Gustschin, Manuel Viermetz, Clemens Schmid, Jakob Haeusele, Frank Bergner, Tobias Lasser, Thomas Koehler, and Franz Pfeiffer

***Abstract*—X-ray computed tomography (CT) has been established as a daily tool in clinical diagnostics and has been continuously refined by more recent innovations in the last years. These systems are, however, limited by fundamental constraints since they are only capable of mapping X-ray attenuation differences in the tissue. Phase-contrast and dark-field imaging provide complementary contrast, which originates from physically different interaction processes of X-rays with matter. Particularly the dark-field signal is considered to have significant diagnostic potential since it is capable to retrieve micro-structural information below the actual resolution limit of the imaging system. This was demonstrated in various laboratory setups and recently also in the fist study with human patients in a clinical radiography system based on a grating interferometer.**

In a recent work, we presented the first implementation of such an X-ray interferometer into a clinical CT gantry. Upscaling and adapting this technology for a rotating CT gantry involves several challenges and tradeoffs ranging from limitations in interferometer design over fast, continuous signal acquisition requirements to tolerances in applied patient dose. In this work we discuss the performance of the first clinical dark-field CT prototype. For this purpose, we present results of our phantom studies which were designed to evaluate whether and how the dark-field contrast generated by the system is capable to provide additional structural sample information. The key aspects include the possibility of quantitative imaging and a gradual approach to simulate results that come as close as possible to a real application in a human patient.

***Index Terms*—X-ray imaging, dark-field contrast, computed tomography**

I. Introduction

X-RAY computed tomography has evolved to one of the most commonly used and indispensable diagnostic 3D imaging modalities. Moreover, the technique is continuously optimized by recent innovations like dual energy or photon-counting technologies as well as advanced data acquisition, reconstruction, and evaluation procedures. Those are, however, still based on a contrast which is measured by retrieving the X-ray attenuation properties of different tissues and hence fundamentally limited. Considering the wave nature of X-rays, additional and complementary contrast can be achieved by measuring the small-angle scattering (dark-field) properties of an object [1]. Contrary to attenuation contrast, the dark-field signal hence retrieves information on micro-structures without the need to actually resolve them on the detector [2]. The potential of the dark-field channel in clinical diagnostics has been discussed in various studies, which were mainly focused on lung diseases like chronic obstructive pulmonary disease (COPD) [3], [4], fibrosis [5] or lung cancer [6]. This is because, particularly in lung parenchyma, the dark field signal is significantly more sensitive to small changes in the alveolar structure compared to the attenuation.

One of the most promising approaches for dark-field imaging with respect to clinical applications is Talbot-Lau interferometry [7]. While first clinical prototypes of this kind are limited to 2D radiographic imaging, all bench top or small-animal CT systems feature a step and shoot data acquisition which leads to total scan times of at least several minutes, which is unfeasible for realistic clinical applications [8], [9]. Recently, we reported on the first successful integration of a Talbot-Lau interferometer into a rotating clinical CT gantry [10]. Restrictions in total interferometer length and grating positioning on a compact CT gantry along with limitations in state of the art grating fabrication technology constrain the theoretically achievable system sensitivity.

Here, we present first results from our prototype system starting from a basic proof-of-concept towards more realistic phantoms for an actual clinical application. This allows to practically evaluate the system performance and discuss its potential and limitations for clinical application. In this context, we also consider a possibility of quantitative dark-field imaging and discuss the patient dose resulting from the applied scanning protocols.

II. Dark-field signal calibration

A physical interpretation of the dark-field signal is a scattering distribution, which is continuously broadened along the beam path [11]. Similar to the exponential decay of the attenuation signal, the measured dark-field signal therefore depends on the penetrated sample thickness. Compared to 2D radiographic imaging, computed tomography provides a more suitable basis for quantitative imaging. However, the dark-field signal strength further depends on several additional parameters

N. Gustschin, M. Viermetz, C. Schmid, J. Haeusele, and F. Pfeiffer are with the Chair of Biomedical Physics, Department of Physics, Technical University of Munich, 85748 Garching, Germany. (email: niko.gustschin@tum.de)

N. Gustschin, M. Viermetz, C. Schmid, J. Haeusele, T. Lasser, and F. Pfeiffer are with the Munich Institute of Biomedical Engineering, Technical University of Munich, 85748 Garching, Germany.

T. Lasser is with Computational Imaging and Inverse Problems, Department of Informatics, Technical University of Munich, 85748 Garching, Germany.

T. Koehler and F. Bergner are with Philips Research, 22335 Hamburg, Germany.

T. Koehler and F. Pfeiffer are with the Institute for Advanced Study, Technical University of Munich, 85748 Garching, Germany.

F. Pfeiffer is with the Department of Diagnostic and Interventional Radiology, School of Medicine and Klinikum rechts der Isar, Technical University of Munich, 81675 München, Germany

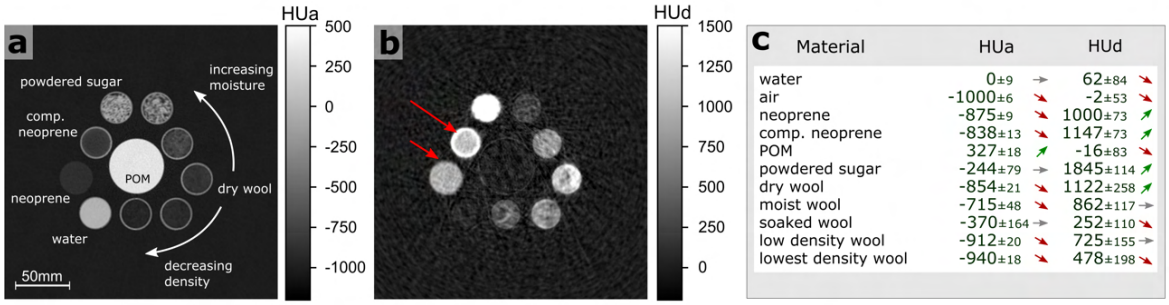


Fig. 1. Reconstruction results from a multi-material cylinder phantom. **a,** Conventional attenuation and **b,** dark-field tomograms calibrated to the attenuation Hounsfield (HUa) and dark-field Hounsfield scales (HUd). The dark-field channel maps the scattering power of different materials and hence provides complementary sub-resolution information. The arrows in b indicate a locally stronger dark-field signal from the same neoprene material, where it is slightly compressed by the tube housing. **c,** Summary of the measured HUa and HUd units from the reconstruction shown in a and b including standard deviation. The coloured arrows indicate a rough classification of the signal within the overall measured signal range of the respective contrast modality. They demonstrate that different HUa and HUd signal combinations contain uncorrelated information which can facilitate material differentiation.

like the feature size of the small-angle scattering structures and also system specific characteristics like the X-ray spectrum or the interferometer sensitivity [2]. Similar to the attenuation Hounsfield scale, a quantitative consistency between different system parameters can be achieved by a calibration using predefined reference materials [8]. For this purpose we use a neoprene foam and air as reference materials in order to transform the measured linear diffusion coefficient $\varepsilon(x)$ to a dark-field Hounsfield unit (HUd) according to:

$$\mathrm{HUd}\,(x) = 1000 \;\cdot\; \frac{\varepsilon(x) - \varepsilon_{\mathrm{air}}}{\varepsilon_{\mathrm{neoprene}} - \varepsilon_{\mathrm{air}}}. \qquad (1)$$

In this HUd scale, the signal from the strongly scattering and weakly attenuating neoprene is defined by a value of 1000 while the non-scattering air yields a value of zero. Darker regions thereby indicate a lower density of scattering interfaces. For a better differentiation to the attenuation Hounsfield scale (HUa), the HUd scale reference values are chosen such that the signal from lung tissue is expected to give positive values.

III. Multi-material phantom

For a proof-of-principle that the dark-field CT prototype is capable to differentiate a varying density of scattering structures below its resolution limit, we use a multi-material phantom. It is composed of different materials in plastic tubes which are arranged around a larger polyoxymethylene (POM) cylinder. The reconstructed attenuation and dark-field images are shown in Fig. 1a and b and demonstrate that our system is able to retrieve the complementary nature of the dark-field signal. Dry wool for example, has a relatively weak attenuation but strong dark-field signal. A decreasing density is hardly noticeable in the attenuation while it is clearly visible in the dark-field. On the other hand, the dark-field signal also fades with an increasing water content, which gradually neutralizes the amount of scattering interfaces and is more apparent in the attenuation modality.

Similarly, pathological changes in lung parenchyma can either be a loss of pulmonary structures (e.g. emphysema), or a replacement with conjunctive tissue (e.g. pulmonary fibrosis). The example demonstrates that the system can successfully provide perfectly registered multi-modal images which enable a better differentiation of material compositions relevant for diagnostic purposes. Moreover, the results demonstrate that also material inhomogeneities which are hardly noticeable in the attenuation contrast can be more distinctly revealed in the dark-field modality. This is for example apparent in the cylinders stuffed with wool or also in the periphery of the top neoprene cylinder, where the material is locally compressed by a tube housing.

IV. Anthropomorphic thorax phantom

To evaluate the dark-field CT prototype regarding clinical application, we employed an anthropomorphic human chest phantom. First, it allows us to assess the performance of the system on a large field of view (FOV). Secondly, it also simulates a more realistic attenuation of the incoming photon flux by artificial bones and soft tissue, which is important when it comes to dose and noise considerations. To simulate lung tissue, we modified the commercially available phantom (Lungman, Kyoto Kagaku, Tokyo, Japan) [12] with a neoprene insert. It features a relatively weak attenuation as well as a porous micro-structure in the size range of lung alveoli [13]. A POM cylinder in the centre simulates additional attenuation by the heart and several plastic tube inserts allow to evaluate additional embedded materials.

The respective attenuation and dark-field images are depicted in Fig. 2a and b. The FOV of the prototype system (45 cm) is capable to map the entire human thorax along with the patient table. As expected, the conventional reconstruction yields a good contrast for highly attenuating materials like bones, POM and the artificial soft tissue. In contrast, the scattering properties of the neoprene insert and apparently also of the synthetic bones stand out in the dark-field domain. It is, however, evident that the latter modality features a distinctly lower resolution compared to the conventional attenuation image. This is caused by low-frequency artefacts originating from the sliding window phase retrieval approach [10] and is acceptable, since the dark-field signal inherently already contains sub-resolution information.

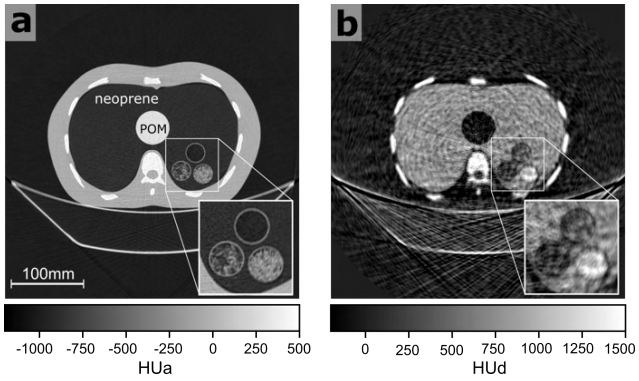


Fig. 2. Reconstruction results of an anthropomorphic human chest phantom where a neoprene foam is used to simulate the attenuating and scattering characteristics of lung parenchyma. The field of view (FOV) comprises the whole thorax cross-section along with the patient table. **a,** The conventional reconstruction shows the expected good contrast for bones and a very low signal from the lung area. **b,** In the dark-field reconstruction, homogeneous materials like POM or soft tissue generate no signal, while the scattering properties of the foam material and powdered sugar are clearly retrieved.

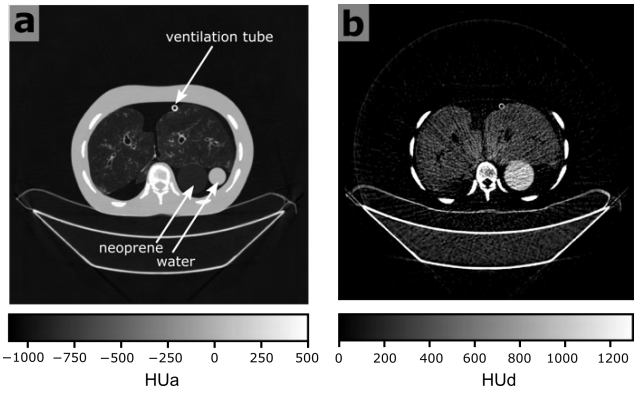


Fig. 3. Reconstruction results of a dissected porcine lung placed inside the human chest phantom. **a,** Conventional reconstruction shows the additional bronchial system while the alveolar structure has comparable attenuation characteristics to the neoprene material. **b,** The scattering properties of the porcine lung tissue are clearly retrieved in the dark-field, however, the signal is significantly lower compared to the neoprene reference.

V. Porcine lung sample

Although the used neoprene material features micro-structures in size comparable to pulmonary alveoli, it cannot fully model the anatomical structure of actual lung parenchyma. The phantom material is highly uniform while lung tissue additionally comprises a capillary system and bronchioles of different sizes. We therefore extended the measurements examining a porcine lung. The additional attenuation of human tissue was again modeled by inserting the lung into the thorax phantom along with a water and neoprene reference sample. A dissected porcine lung was ordered from a butcher and the measurements were conducted more than 24 h post-mortem. The lung was inflated externally with 30 mbar in order to partially revert an incipient collapse of lung alveoli. The resulting attenuation and dark-field images are shown in Fig. 3a and b. From the conventional reconstruction it is evident that the attenuation of lung tissue is similar to the neoprene foam, but also reveals additional bronchi. The scattering properties of the lung tissue are clearly captured in the dark-field domain, however, the signal is significantly lower compared to the neoprene reference material. Previous work with dark-field radiography already showed a significant dependence of the dark-field signal on the applied ventilation pressure [14]. It hence can be justifiably assumed that a higher density of scattering interfaces in the case of a decreased ventilation pressure will favour dark-field signal intensity. Moreover, a further increased signal can also be expected in living subjects since a decomposition processes of lung parenchyma starts immediately after death when the tissue is no longer perfused.

VI. Non-clinical applications

Beyond clinical application, dark-field imaging finds general use when examining microscopic defects and fibrous or porous materials in the field of non-destructive testing or for quality control purposes. The key features of the presented dark-field CT prototype system enable new possibilities for applications where the FOV or acquisition time has been a significant limiting factor. Here, we demonstrate a potential for security screening applications with large samples. For this purpose a cloth bag was imaged after filling it with several items along with a sample of fine baby powder. In the attenuation image in Fig. 4a the powder appears to be a microscopically homogeneous object. On the dark-field Housfield scale in Fig. 4b it reaches values of up to 13×10^3 HUd which is close to a total extinction of the maximum visibility of the system (see [10]). Soft homogeneous materials could be thus efficiently separated from similarly absorbing explosives or drugs that incorporate micro-granular scatterers. Depending on the respective scattering power and the thickness or volume of the material, however, a higher maximum visibility might be required. Moreover, metallic structures which are not uncommon in such applications will be a major challenge for a reasonable performance of a grating interferometer.

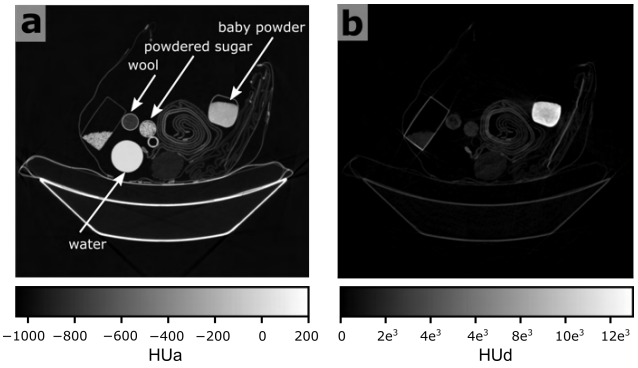


Fig. 4. Reconstruction results of a bag filled with different objects (water bottle, wool, powdered sugar, packaging material, rolled towels and cardboard). **a,** In the conventional reconstruction, micro-granular powder appears to be homogeneous and cannot be distinguished from bulk material or fluids with similar attenuation. **b,** In the dark-field image, the powder can be clearly differentiated even from other strongly scattering materials like powdered sugar.

VII. Dose considerations

The presented phantom scans were acquired in axial data acquisition mode using an 80 kVp spectrum and a tube current of 550 mA. The integrated grating interferometer consists out of three optical X-ray gratings which absorb a significant part of the generated X-ray flux. With an unmodified CT system, our settings would result in a volume CT dose index ($CTDI_{vol}$) of about 13 mGy. The measured $CTDI_{vol}$ of 7.39 mGy with these settings according to the constancy test protocol [15] with a standard body phantom (32 cm diameter) and a calibrated dosimeter (NOMEX, PTW, Germany) indicates that the combination of the first two gratings can be considered to absorb around 50% of the generated X-ray photons in front of the patient. This value lies well within the clinically applicable range for chest CT of adults at state of the art CT systems [16]. Since one of the absorbing gratings is positioned after the patient directly in front of the detector, a substantial part of the applied patient dose does not actually reach the detector. This is, however, compensated by retrieving sub-resolution information in the additional dark-field modality, which can not be accessed by attenuation contrast only.

VIII. Conclusion

In this work we presented results of recent phantom measurements with the first clinical dark-field CT prototype system. Despite the restrictions posed by the compact interferometer geometry and limitations in current grating fabrication technology, the system performs reasonably well on a sufficient FOV and within a clinically conceivable dose range. We propose a HUd scale for quantitative dark-field imaging and discuss the expected signal strength of lung tissue by means of different phantom materials. Although the presented phantom studies are mainly dedicated to lung imaging, further clinical as well as non-clinical applications are now accessible due to a large FOV in combination with a fast data acquisition procedure.

Acknowledgments

The authors wish to thank Julia Herzen, Daniela Pfeiffer, Alexander Fingerle, Ernst Rummeny, Pascal Meyer, Jürgen Mohr, Maximilian von Teuffenbach, Amanda Pleier, Tom Kumschier, Michael Heider, Sven Prevrhal, Thomas Reichel, Ami Altman, and Shlomo Gotman for their help, support, and dedication to launch this complex project. This work was carried out with the support of the Karlsruhe Nano Micro Facility (KNMF, www.kit.edu/knmf), a Helmholtz Research Infrastructure at Karlsruhe Institute of Technology (KIT). We acknowledge the support of the TUM Institute for Advanced Study, funded by the German Excellence Initiative, the European Research Council (ERC, H2020, AdG 695045) and Philips GmbH Market DACH .

References

[1] F. Pfeiffer *et al.*, "Hard-x-ray dark-field imaging using a grating interferometer," *Nature Materials*, pp. 134–137, 2008.

[2] W Yashiro, Y Terui, K Kawabata, and A Momose, "On the origin of visibility contrast in x-ray Talbot interferometry," *Optics Express*, p. 16 890, 2010.

[3] P. Modregger, T. P. Cremona, C. Benarafa, J. C. Schittny, A. Olivo, and M. Endrizzi, "Small angle x-ray scattering with edge-illumination," *Scientific Reports*, 2016.

[4] K. Willer *et al.*, "X-ray dark-field chest imaging for detection and quantification of emphysema in patients with chronic obstructive pulmonary disease: A diagnostic accuracy study," *The Lancet Digital Health*, e733–e744, 2021.

[5] A. Yaroshenko *et al.*, "Improved In vivo Assessment of Pulmonary Fibrosis in Mice using X-Ray Dark-Field Radiography," *Scientific Reports*, p. 17 492, 2015.

[6] K. Scherer *et al.*, "X-ray Dark-field Radiography - In-Vivo Diagnosis of Lung Cancer in Mice," *Scientific Reports*, p. 402, 2017.

[7] A. Momose, "X-ray phase imaging reaching clinical uses," *Physica Medica*, pp. 93–102, 2020.

[8] A. Velroyen *et al.*, "Grating-based X-ray Dark-field Computed Tomography of Living Mice," *EBioMedicine*, pp. 1500–1506, 2015.

[9] Z. Wu *et al.*, "Prototype system of non-interferometric phase-contrast computed tomography utilizing medical imaging components," *Journal of Applied Physics*, p. 074 901, 2021.

[10] M. Viermetz *et al.*, "Dark-field computed tomography reaches the human scale," *PNAS*, accepted.

[11] M. Bech, O. Bunk, T. Donath, R. Feidenhans'l, C. David, and F. Pfeiffer, "Quantitative x-ray dark-field computed tomography," *Physics in Medicine and Biology*, pp. 5529–5539, 2010.

[12] *Multipurpose Chest Phantom N1 "LUNGMAN"*, https://www.kyotokagaku.com/en/products_data/ph-1_01/, (accessed 03 January 2021), 2020.

[13] K. Taphorn, F. De Marco, J. Andrejewski, T. Sellerer, F. Pfeiffer, and J. Herzen, "Grating-based spectral X-ray dark-field imaging for correlation with structural size measures," *Scientific Reports*, p. 13 195, 2020.

[14] F. D. Marco *et al.*, "Contrast-to-noise ratios and thickness-normalized, ventilation-dependent signal levels in dark-field and conventional in vivo thorax radiographs of two pigs," *PLOS ONE*, e0217858, 2019.

[15] IEC 61223-2-6:2006, *Evaluation and routine testing in medical imaging departments - Constancy tests*, 2006.

[16] American Association of Physicists in Medicine, *Adult Routine Chest CT Protocols*, https://www.aapm.org/pubs/CTProtocols/, (accessed 10 December 2021), 2016.

Non-invasive real-time thermometry via spectral CT physical density quantifications

Nadav Shapira, Leening P. Liu, Johoon Kim, David P. Cormode, Gregory J. Nadolski, Matthew Hung, Michael C. Soulen, Peter B. Noël

***Abstract*— Efficient removal of solid focal tumors is a major challenge in modern medicine. Percutaneous thermal ablation is a first-line treatment for patients not fit for surgical resection or when the disease burden is low, mainly due to expedited patient recovery times, lower rates of post-operative morbidity, and reduced healthcare costs. While continuously gaining popularity, ~100,000 yearly thermal hepatic ablation procedures are currently performed without actively monitoring temperature distributions, leading to high rates of incomplete ablations, local recurrences, and damage to surrounding structures. Recent advancements in computed tomography (CT), especially spectral CT, provide promising opportunities for lowering these rates. The additional information available with spectral CT can provide the necessary capabilities to achieve accurate, reliable, on-demand, and non-invasive thermometry during ablation procedures. By taking advantage of our newly developed spectral physical density maps and their direct relation with temperature changes, we performed experiments on phantoms and *ex vivo* tissue to develop, evaluate, optimize, and refine a method for generating thermometry maps from spectral CT scans. Our results validate the accuracy of the spectral physical density model, allowing "whole-organ" mass quantifications that are accurate within one percent, as well as demonstrate an ability to extract temperature changes (linear correlation coefficient of 0.9781) non-invasively and in real-time.**

Index Terms—Dual-Energy CT, Spectral-CT, Quantitative imaging, Image-guided therapy, Interventional oncology, Tumor ablation, Thermometry.

I. Introduction

With more than 900,000 yearly new cases worldwide, liver cancer is the fifth most common cancer in men and the ninth most common cancer in women[1]. Percutaneous thermal ablation techniques provide minimally invasive and inexpensive focal treatment strategies for hepatic tumors[2,3]. They are considered a first-line treatment for patients with small hepatocellular carcinomas (HCC)[4], the most common primary malignancy in the liver and the second leading cause of cancer-related mortality in the world[5], and are used to bridge patients to liver transplantation[6]. Safe and effective ablation treatments rely on complete coverage of the target lesion with lethal temperatures[7] (≥60 °C), while sparing as much surrounding tissue as possible and keeping safety margins to adjacent critical structures[6]. However, despite technological advancements over the past years, local recurrence rates are high (**Figure 1**), burdening patients and healthcare systems.

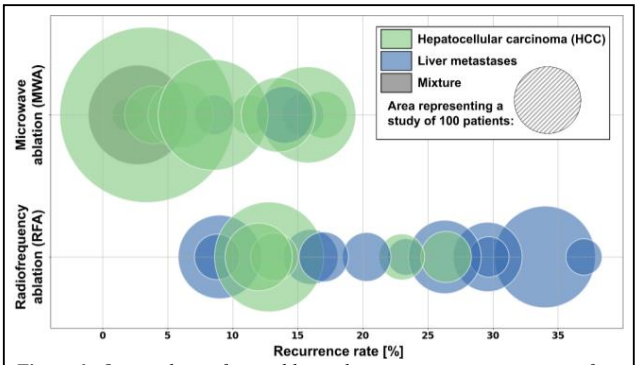


Figure 1: *Our analysis of post-ablation liver cancer recurrence rates from twenty-five past studies shows high recurrence rates of up to 17% for microwave ablation and even higher rates, up to 37%, for radiofrequency ablation. Disc sizes represent the number of patients included in each study.*

Published clinical requirements include temperature accuracy of ≤2 °C, spatial resolution of ≤2 mm, short acquisition, and image generation times for volumetric coverage (≤1 minute), metal artifact suppression, and radiation dose levels that meet safety standards[8]. CT-thermometry provides the most promising solution for monitoring thermal ablation treatments. This is mainly because most ablation treatments are already performed under CT guidance, the compatibility of CT with all commercially available ablation systems[9], and the ability to detect immediate complications, e.g., bleeding. However, despite decades of CT-based thermometry investigations[10], there is a growing demand for solutions to monitor temperatures during ablation procedures.

Temperature dependence of CT Hounsfield units (HU) has been observed since the late 1970s[10]. Since then, several research groups partially addressed some of the clinical requirements for image-based thermometry listed above with *ex vivo* and *in vivo* experiments. However, several crucial obstacles remain for the anticipated clinical translation of CT-based thermometry.

Temperature assessments from conventional CT rely on attributing shifts in HU to changes in tissue density, which in

This work was supported in part by the National Center for Advancing Translational Sciences of the NIH (award #UL1TR001878), by the Institute for Translational Medicine and Therapeutics' (ITMAT) Transdisciplinary Program in Translational Medicine and Therapeutics, and by Philips Healthcare.

N. Shapira, D.P. Cormode, G.J. Nadolski, M. Hung, M.C. Soulen, and P.B. Noël are with the Department of Radiology, University of Pennsylvania, Philadelphia, PA, USA. L.P. Liu and J. Kim are with the Department of Radiology and the Department of Bioengineering, University of Pennsylvania, Philadelphia, PA, USA. e-mail: nadav.shapira@pennmedicine.upenn.edu.

turn are affected by temperature changes according to thermal volumetric expansion $\rho(T)/\rho_o(T_0) = (1 + \alpha\Delta T)^{-1}$, where $\rho(T)$ is the tissue density at temperature T, ρ_o is the tissue density at a baseline temperature T_0, ΔT is the change from baseline temperature (in °C), and α is the thermal expansion coefficient associated with the tissue[10]. All previous CT-based thermometry studies employed linear or quadratic approximations of the relation above (assuming small ΔT) to correlate temperature and HU changes, which are assumed to be proportional to ρ. While conventional HU depend linearly on tissue density, they are also affected (non-linearly) by changes in tissue composition, which previous studies report as a limiting factor that corrupts temperature accuracy[10]. In addition, conventional HU are greatly affected by the scanner model, e.g., tube filtration, and acquisition parameters, e.g., kVp, as well as the patient habitus. This leads to large inconsistencies in thermal sensitivity values seen in previous studies (**Figure 2**). We propose to employ the exact mathematical relationship between tissue density and temperature changes[10], $\Delta T \propto \rho_o(T_0)/\rho(T)$, by utilizing our recently developed spectral physical density quantifications[11].

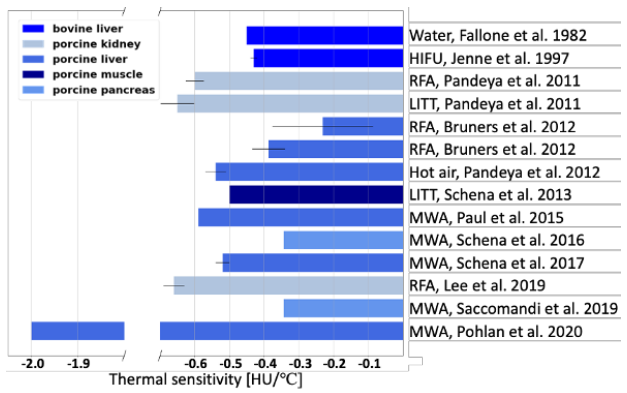


Figure 2: *Dispersity of reported thermal sensitivity values. A large range of HU changes, from 0.5 HU to 4 HU, are associated with an increase of 2 °C.*

To systematically evaluate our method, we require dedicated phantoms designed to support the development of CT-based solutions for thermometry applications. Tissue-mimicking phantoms that emulate various physical properties of biological tissue are central for the development and evaluation of novel clinical technologies and applications. Compared to *ex vivo* and *in vivo* experiments, phantoms simplify safety, logistical, and cost considerations. For CT-based thermometry applications, phantoms are required to exhibit similar thermal and x-ray to those of human tissue, i.e., thermal conductivity and diffusivity, linear attenuation coefficients within the entire photon energy range relevant to clinical CT imaging. Such phantoms will enable repeatable and controlled experiments that allow rigorous comparisons of temperature sensitivity and reliability on different CT platforms at different imaging protocols.

In this work we report on recent developments aimed for making non-invasive real-time thermometry a reality. We have previously demonstrated accurate physical density quantifications from clinical virtual mono-energetic images (VMI) and effective atomic number (Z_{eff}) spectral results in phantom experiments[11]. Here we further validate the accuracy of these new spectral maps by utilizing them for non-invasive whole-organ mass estimations on *ex vivo* tissue. While our excellent results attest to the high accuracy of our spectral physical density quantifications, they also present opportunities for novel stand-alone clinical applications. In addition, we demonstrate high correlations between temperature changes and physical density quantifications on *ex vivo* tissue and on thermo-spectral tissue-mimicking phantoms that we developed specifically for this purpose.

II. Methods

A. *Thermo-spectral tissue-mimicking phantom development*

Our phantoms were developed by iteratively modifying the synthesis method detailed by Negussie and Mikhail *et al.*[12,13] in order to match the x-ray attenuation curve to that of human liver tissue calculated from well-accepted elemental composition and physical density values[14]. These phantoms are particularly useful for thermal therapy experiments since they present comparable thermal properties to those of human tissue. Briefly, 287.5 ml of deionized water in a 1000 ml flask were degassed by purging N_2 for 15 minutes. After degassing, 202.5 ml of 40% (w/v) acrylamide/bis-acrylamide solution was added to the degassed water to achieve 490 ml solution of 16.5% (w/v) acrylamide/bis-acrylamide. While stirring, 6 grams of calcium chloride dissolved in 10 ml of deionized water was added to the acrylamide solution. A single gram of ammonium persulfate in 2 ml of deionized water and a single milliliter of N,N,N',N'-tetramethylethylenediamine were subsequently added to the solution. After stirring for additional 15 seconds, the final solution was immediately transferred to a 475 ml plastic jar.

B. *Physical density spectral map generation*

A complete description of the development of our spectral physical density model, its optimization, and its verification on a tissue characterization phantom (Gammex Model 467, Sun Nuclear) was provided before[11]. Briefly, 70 keV VMI and Z_{eff} voxel values, which are clinically available on any spectral CT platform, are converted into physical density values through a parametrized Alvarez-Macovski model[15].

C. *Non-invasive mass measurements*

Ex vivo bovine muscle physical density quantifications were evaluated to determine the effect of acquisition parameters on the resulting accuracy. The specimen was weighed with a precision balance (Fisher Scientific Education Precision Balance, Fisher Scientific) to provide ground-truth mass values. Next, the sample was placed on a rectangular block of polyfoam within the 20 cm bore a multi-energy CT phantom (MECT, Sun Nuclear) and scanned with a spectral detector dual-energy CT (IQon spectral scanner, Philips Healthcare) (**Figure 3A**).

Scans were repeated three times at a tube voltage of 120 kVp for each set of collimations {16x0.625, 64x0.625 mm}, dose levels {15.2, 30.3, 45.5 mGy}, and acquisition mode {axial, helical} combinations. No helical scan was acquired with 16x0.625 mm at 45.5 mGy due to tube output limitations. Images were reconstructed with a clinical standard body kernel,

a field of view of 350 mm, and at a slice thickness of 2.5 mm with 2.5 mm slice intervals. 70 keV VMIs and Z_{eff} maps were reconstructed from every acquisition and corresponding physical density maps were generated using our spectral physical density model. Regions of interest (ROI) with a diameter of 13.6 mm were positioned in the center of the sample on four consecutive image slices to assess physical density quantification and their dependence on scan parameters. Finally, the sample was weighed after scanning to account for any losses in blood or minor changes in temperature.

To calculate the total mass of the specimen, physical density values were summed from non-air voxels, i.e., 70 keV VMI values larger than -950 HU, and multiplied by the voxel volume (0.68 x 0.68 x 2.5 mm^3). Calculated mass was compared to the average of the two weights, pre- and post-scanning, to evaluate its accuracy with different scanning parameters.

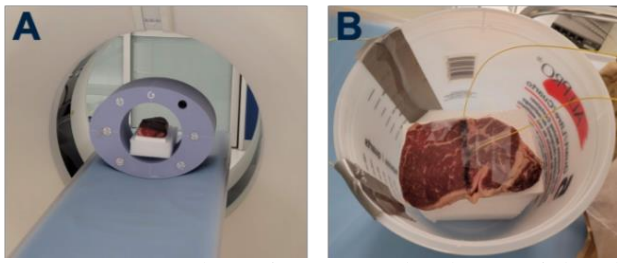


Figure 3: *Experimental setup for two potential applications of physical density quantifications with spectral CT. An ex vivo muscle was (A) scanned within a multi-energy CT phantom on a spectral detector dual-energy CT to investigate non-invasive mass measurements, and (B) subjected to a range of*

D. Non-invasive temperature monitoring

To assess the correlation between changes in temperature and changes in physical density, relative to the physical density values at baseline temperature, optical fiber temperature probes were inserted into the same sample used in the non-invasive mass measurement experiment, as well as three of our dedicated thermo-spectral tissue-mimicking phantoms, using 13-gauge medical trocars to continuously record local internal temperatures (**Figure 3B**). The sample or the phantom was placed in a plastic container, and pre-heated water was poured in to completely submerge it, consequently subjecting it to a wide range of temperatures. After the sample, or phantom, reached an equilibrium temperature, ice was added to cool the water. During heating and cooling, scans were performed approximately every minute with a spectral detector dual-energy CT at a tube voltage 120 kVp, a 16x0.625 mm collimation, a revolution time of 0.75 seconds, and three different radiation dose levels {15.2, 30.3, 45.5 mGy}. For each scan, physical density maps were generated from 70 keV VMI and Z_{eff} spectral results using the same reconstruction parameters detailed in the section above.

The locations of optical fiber temperature probes were determined by thresholding 70 keV VMIs at 90 HU. 4.1 mm diameter ROIs were placed adjacent to the tip of the optical fiber to measure physical density. To elucidate thermal volumetric expansion, physical density values were normalized by dividing the last temporal physical density value with the physical density at a given timepoint. Similarly, the change in temperature was determined relative to the last temporal scan. Linear regressions were fit to the data, where the slopes were associated with the thermal volumetric expansion coefficient. R-values were determined to characterize the correlation between normalized physical density and temperature change.

III. Results

A. Thermo-spectral tissue-mimicking phantom development

Attenuation curves that were measured on multiple VMIs, at energies between 40 and 200 keV, with increments of 10 keV, are presented in **Figure 4**. The curves present the iterative developmental process. It enabled us to achieve a maximum error of 4 HU compared to human liver tissue across the entire energy range (Iteration #0 is the scale-down formulation from Negussie and Mikhail *et al.*). Since human livers differ in their elemental compositions, e.g., different fat contents, we deduce that these attenuation errors were sufficiently small for the intended purposes of these phantoms, i.e., the development and testing of spectral CT thermometry approaches.

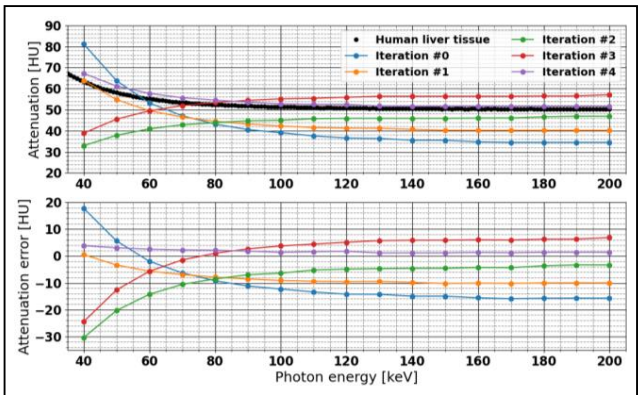


Figure 4: *Iterative development process of a thermo-spectral tissue-mimicking phantom, achieving errors below 4 HU across all x-ray energies.*

B. Non-invasive mass measurements

Spectral physical density quantifications from scans at different doses, collimations, and axial/helical scans revealed that with dose matched scans and different collimations, physical density decreased approximately by 0.003 g/ml with increased collimation. In addition, we observed no effect of dose and axial/helical scans on physical density quantifications.

Non-invasive "whole-organ" mass estimations of the *ex vivo* bovine muscle from scans of varying collimation and dose levels illustrated extremely high accuracies (**Figure 5**). Similar to the physical density quantification, dose levels did not impact the mass value. Between the five different combinations of collimation and dose, estimated mass values were within ±1.1 grams of ground-truth mass measurements with a scale. For 16x0.625 and 64x0.625 mm collimations, the errors in mass were -0.34% and -0.04%, respectively. The accuracy of estimated mass further validates our physical density results and demonstrates a clinical application of non-invasive mass measurements for determining the presence of pathology, which is currently utilized only in post-mortem autopsies.

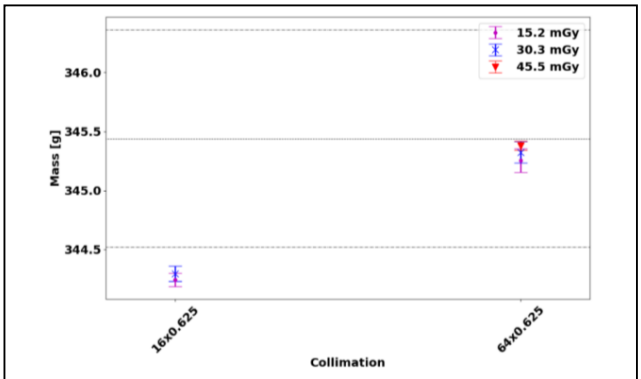


Figure 5: *Mass estimation from spectral physical density maps, in excellent agreement with measured mass. The dashed lines represent two independent weight measurements, and the dotted line represents their average.*

C. Non-invasive temperature monitoring

In addition to non-invasive mass measurements, physical density quantifications also enabled non-invasive temperature monitoring as temperature changes are reflected in physical density changes (**Figure 6**). Specifically, a linear fit between normalized physical density and change in temperature demonstrated a slope of 0.00042 ± 0.00001 °C^{-1} and an intercept of 1.000 ± 0.0003 for temperatures between 22.0 and 45.5 °C. These fit parameters correspond to a 0.42% decrease in physical density with an increase of 10 °C. High linear correlation ($R = 0.9781$) between normalized physical density and change in temperature recapitulated the theoretical relation.

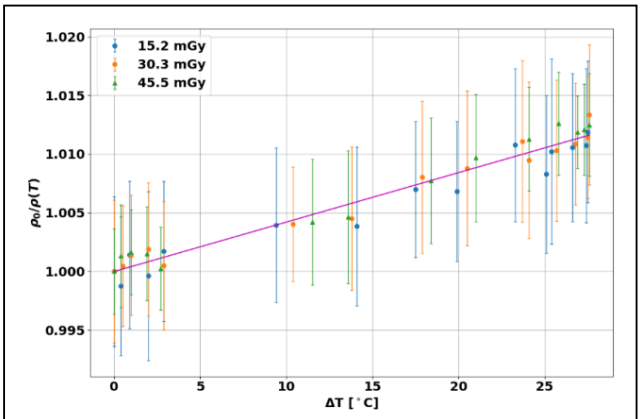


Figure 6: *Normalized physical density changes with heating and cooling of ex vivo bovine muscle. The linear relationship between normalized physical density and changes in temperature reflected thermal volumetric expansion.*

IV. Conclusion

We have demonstrated the quantitative accuracy of our physical density model on *ex vivo* tissue and on dedicated thermo-spectral phantoms that we have developed specifically for CT-based thermometry applications. In addition, our results demonstrate our ability to employ a well-established and direct relation, i.e., approximation-free, between changes in physical density estimations obtained from spectral CT and changes in temperature. This ability can serve as the backbone of future non-invasive real-time thermometry that is based on non-retrieved spectral information. With the increase in spectral CT utilization and the foreseen replacement of conventional CT scanners by this newer generation systems, we recognize a great opportunity to improve the monitoring and guidance of thermal therapy procedures, which will help reduce the currently high rates of local recurrence.

References

1. Liver cancer statistics | World Cancer Research Fund. Available at: https://www.wcrf.org/dietandcancer/cancer-trends/liver-cancer-statistics. (Accessed: 24th March 2021)
2. Dodd, G. D., Soulen, M. C., Kane, R. A., Livraghi, T., Lees, W. R., Yamashita, Y., Gillams, A. R., Karahan, O. I. & Rhim, H. Minimally invasive treatment of malignant hepatic tumors: At the threshold of a major breakthrough. *Radiographics* **20**, 9–27 (2000).
3. Froelich, M. F., Schnitzer, M. L., Rathmann, N., Tollens, F., Unterrainer, M., Rennebaum, S., Seidensticker, M., Ricke, J., Rübenthaler, J. & Kunz, W. G. Cost-effectiveness analysis of local ablation and surgery for liver metastases of oligometastatic colorectal cancer. *Cancers* **13**, (2021).
4. Izzo, F., Granata, V., Grassi, R., Fusco, R., Palaia, R., Delrio, P., Carrafiello, G., Azoulay, D., Petrillo, A. & Curley, S. A. Radiofrequency Ablation and Microwave Ablation in Liver Tumors: An Update. *The Oncologist* **24**, e990 (2019).
5. McGlynn, K. A., Petrick, J. L. & London, W. T. Global Epidemiology of Hepatocellular Carcinoma: An Emphasis on Demographic and Regional Variability. *Clinics in Liver Disease* **19**, 223–238 (2015).
6. Ringe, K. I., Wacker, F. & Raatschen, H. J. Is there a need for MRI within 24 hours after CT-guided percutaneous thermoablation of the liver? *Acta Radiologica* **56**, 10–17 (2015).
7. Vogl, T. J., Nour-Eldin, N. E. A., Hammerstingl, R. M., Panahi, B. & Naguib, N. N. N. Microwave Ablation (MWA): Basics, Technique and Results in Primary and Metastatic Liver Neoplasms - Review Article. *RoFo Fortschritte auf dem Gebiet der Rontgenstrahlen und der Bildgebenden Verfahren* **189**, 1055–1066 (2017).
8. Frich, L. Non-invasive thermometry for monitoring hepatic radiofrequency ablation. *Minimally Invasive Therapy and Allied Technologies* **15**, 18–25 (2006).
9. Bruners, P., Pandeya, G. D., Levit, E., Roesch, E., Penzkofer, T., Isfort, P., Schmidt, B., Greuter, M. J. W., Oudkerk, M., Schmitz-Rode, T., Kuhl, C. K. & Mahnken, A. H. CT-based temperature monitoring during hepatic RF ablation: Feasibility in an animal model. *International Journal of Hyperthermia* **28**, 55–61 (2012).
10. Fani, F., Schena, E., Saccomandi, P. & Silvestri, & S. CT-based thermometry: An overview. *International Journal of Hyperthermia* **30**, 219–227 (2014).
11. Hwang, M., Litt, H. I., Noël, P. B. & Shapira, N. Accurate physical density assessments from clinical spectral results. in *Medical Imaging 2021: Physics of Medical Imaging* (eds. Bosmans, H., Zhao, W. & Yu, L.) **11595**, 45 (SPIE, 2021).
12. Mikhail, A. S., Negussie, A. H., Graham, C., Mathew, M., Wood, B. J. & Partanen, A. Evaluation of a tissue-mimicking thermochromic phantom for radiofrequency ablation. *Medical Physics* **43**, 4304–4311 (2016).
13. Negussie, A. H., Partanen, A., Mikhail, A. S., Xu, S., Abi-Jaoudeh, N., Maruvada, S. & Wood, B. J. Thermochromic tissue-mimicking phantom for optimisation of thermal tumour ablation Thermochromic tissue-mimicking phantom for optimisation of thermal tumour ablation. *International Journal of Hyperthermia* (2016). doi:10.3109/02656736.2016.1145745
14. *Tissue Substitutes in Radiation Dosimetry and Measurement, Report 44 of the International Commission on Radiation Units and Measurements (ICRU)*. (1989).
15. Alvarez, R. E. & Macovski, A. Energy-selective reconstructions in X-ray computerised tomography. *Physics in Medicine and Biology* **21**, 002 (1976).

Cone-beam reconstruction for a circular trajectory with transversely-truncated projections based on the virtual fan-beam method

Mathurin Charles, Rolf Clackdoyle, and Simon Rit

***Abstract*—We describe a new procedure for three-dimensional (3D) region-of-interest (ROI) reconstruction from transversely-truncated cone-beam projections acquired with a circular source trajectory. This method is an extension to 3D of the virtual fan-beam (VFB) method. It is based on a VFB formula that performs the backprojection in the acquisition geometry. Our simulation results show that the ROI reconstruction of the 3D Shepp-Logan phantom is very similar to the one obtained by the Feldkamp, Davis, Kress (FDK) algorithm without truncation. However the reconstruction of the Forbild head phantom shows artefacts which are absent from the FDK truncation-free reconstruction.**

I. Introduction

In three-dimensional (3D) cone-beam computed tomography (CBCT), a common source trajectory is a circular scanning around the object. The plane containing the circular source trajectory is usually called the central plane, midplane or source plane. From Tuy's data sufficiency condition [1], we know that mathematically exact reconstruction of the object density is possible only in the midplane. In case of non-truncated cone-beam projections, the well-known and widely used Feldkamp-Davis-Kress (FDK) algorithm [2] provides exact reconstruction in the central plane and approximate reconstruction elsewhere. This algorithm, which can be seen as a heuristic extension of the fan-beam filtered backprojection (FBP) formula for two-dimensional (2D) reconstruction, applies a ramp filter to each projection row. Consequently, the FDK formula is not suitable for treating transversely-truncated cone-beam projections.

We distinguish two kinds of situations with transverse truncation. In the first one, the detector is placed off-center so that, even if the detector does not cover the object laterally, each ray-line in the midplane is measured at least once during a 360° scan. It is thus possible, in the midplane, to obtain the missing information of a truncated projection from other projections. Elsewhere, the same procedure is applied to all the other rows of the projections even though the missing rays and measured rays have a different angle with the central plane. Using this idea, several methods have been proposed such as a pre-convolution weighting of the projections before applying the FDK algorithm [3] and a 3D version of a Katsevich-type FBP [4].

M. Charles (email: mathurin.charles@univ-grenoble-alpes.fr) and R. Clackdoyle are with Université Grenoble Alpes, CNRS, TIMC UMR 5525, Grenoble, France.

S. Rit is with Université de Lyon, INSA-Lyon, Université Claude Bernard Lyon 1, UJM-Saint Etienne, CNRS, INSERM, CREATIS UMR 5220, U1294, F-69373, Lyon, France.

The second kind of situation with truncated cone-beam projections is when the detector, which still does not cover the whole object, is centered (when a ray-line passing through the center of rotation of the source hits the center of the detector). In that case, we define the field-of-view (FOV) as the region imaged by every source position, and it corresponds to the volume inside a cylinder which does not contain the whole object. In this situation, it is not possible to obtain missing line-integrals in the midplane from other source positions so the previous methods cannot be applied. However, there are two analytical methods which can perform region-of-interest (ROI) reconstruction from truncated projections in the midplane: the differentiated back-projection (DBP) [5], [6] method, also called back-projection filtration (BPF) [7]; and the virtual fan-beam (VFB) method [8]. The BPF has been extended to three dimensions [9] but, to our knowledge, not the VFB method. In this work, we propose to extend the VFB method to 3D ROI reconstruction in the case of transversely-truncated cone-beam projections acquired with a circular source trajectory.

In the usual 2D context, the principle of the VFB method is to identify a virtual source trajectory for which we have non-truncated projections and to rebin the truncated projections into this geometry. Then, super-short-scan formulas [10] can be used to perform the reconstruction. To choose the virtual trajectory, we use the fact that acquired data can be rebinned into non-truncated projections for any point inside the FOV and outside the convex hull, as we have access to the integral of any half-line extending from this point. In a previous contribution, we proposed a VFB formula [11] for a circular fan-beam acquisition geometry, for which the backprojection was performed in the acquisition geometry. In this work, we extend this approach to 3D.

II. Theory

A. Notation

Let $\vec{\theta}_\lambda = (\cos\lambda, \sin\lambda, 0)$, $\vec{\eta}_\lambda = (-\sin\lambda, \cos\lambda, 0)$ and $\vec{e}_z = (0, 0, 1)$ (in 2D , the last component of $\vec{\theta}_\lambda$ and $\vec{\eta}_\lambda$ is discarded). Let f denote the 3D object density to be reconstructed. The cone-beam projections of f for a circular source trajectory of radius R_A acquired on a flat detector placed at the origin O are defined by

$$\bar{g}^{R_A}(\lambda, u, v) = \int_0^{+\infty} f\left(R_A\vec{\theta}_\lambda + l\frac{-R_A\vec{\theta}_\lambda - u\vec{\eta}_\lambda + v\vec{e}_z}{\sqrt{R_A^2 + u^2 + v^2}}\right) \mathrm{d}l \tag{1}$$

where $\lambda \in \Lambda = [0, 2\pi)$ and $S(\Lambda) = R_A\vec{\theta}_\Lambda$ is the set of vertices (cone-beam source locations) of the trajectory (see figure 1).

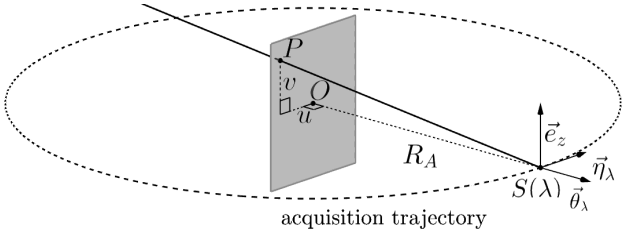


Fig. 1. The circular acquisition geometry of center O and radius R_A. The point $S(\lambda)$ is a vertex of the trajectory. A ray passing through the point P on the detector placed at the origin O is identified by the parameters (λ, u, v).

In 2D, the fan-beam projections of f for a circular source trajectory of radius R_A with angular parametrization are defined by

$$g^{R_A}(\lambda, \gamma) = \int_0^\infty f(R_A\vec{\theta}_\lambda - t\vec{\theta}_{\lambda+\gamma})\,\mathrm{d}t \tag{2}$$

where $\gamma \in (-\pi/2, \pi/2)$ is the usual ray-angle measured counterclockwise with respect to the central ray (which is defined by the source and the center of rotation). The parameters u and γ (respectively for equispaced rays and equiangular rays) are linked by $u = R_A \tan\gamma$ so, in the midplane, we have

$$g^{R_A}(\lambda, \gamma) = \bar{g}^{R_A}(\lambda, R_A\tan\gamma, 0), \tag{3}$$

$$\bar{g}^{R_A}(\lambda, u, 0) = g^{R_A}(\lambda, \arctan(u/R_A)). \tag{4}$$

B. Configuration studied

We consider the following configuration. The FOV is the volume inside a cylinder of center O, extended axially without limit since we consider no axial truncation. We assume that the support of the object function is contained within a known ellipsoid which extends outside the FOV (see figure 2).

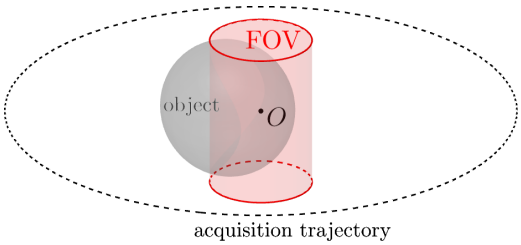


Fig. 2. The ellipsoid object is partially covered by the cylindrical FOV.

C. The VFB formula used in the midplane

In the midplane, the 2D slice of the object has an elliptic support and the FOV has a circular support. The chosen virtual trajectory is the arc of circle at the border of the FOV and outside the object (see figure 3). In that case, the area for which the VFB method is mathematically exact is the convex hull of the virtual trajectory.

We now recall our VFB formula from [11]. The rebinning relations between two trajectories with different radius can

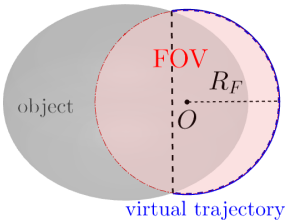


Fig. 3. Situation in the midplane: the circular FOV of center O and radius R_F covers only a part of the elliptic object. The virtual trajectory is the arc of circle of center O and radius $R_V = R_F$ in blue and the vertical black dashed line is the boundary of its convex hull.

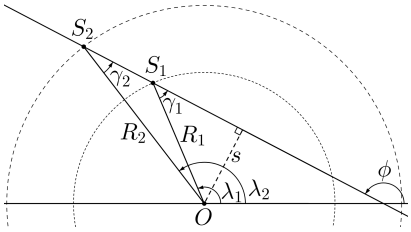


Fig. 4. The parameters (λ_i, γ_i) of a ray for source trajectories of radius R_i with $i \in \{1, 2\}$ are linked through $s = R_i \sin\gamma_i$ and $\phi = \lambda_i + \gamma_i$.

be seen on figure 4. The data are first rebinned from the acquisition geometry of radius R_A to the virtual geometry of radius R_V using

$$g^{R_V}(\lambda, \gamma) = g^{R_A}(\lambda + \gamma - \gamma^{R_V}_{R_A}, \gamma^{R_V}_{R_A}) \tag{5}$$

where

$$\gamma^{R_V}_{R_A} = \arcsin\left(\frac{R_V}{R_A}\sin\gamma\right). \tag{6}$$

Then, differentiation and Hilbert filtering is performed on the non-truncated projections in the virtual geometry with

$$g_F^{R_V}(\lambda, \gamma) = \frac{1}{2\pi}\int_{-\pi}^{\pi} h_H(\sin(\gamma - \gamma'))(\partial_1 - \partial_2)g^{R_V}(\lambda, \gamma')\,\mathrm{d}\gamma'. \tag{7}$$

where $h_H(s) = \int_{\mathbb{R}} -\mathrm{i}\ \mathrm{sign}(\sigma)\mathrm{e}^{2\mathrm{i}\pi\sigma s}\,\mathrm{d}\sigma$ denotes the Hilbert filter and ∂_i corresponds to the partial derivative with respect to the i-th variable. As the virtual trajectory is not a full scan, the redundancy in the filtered projections is handled by applying a weight w^{R_V} (that we do not detail) to $g_F^{R_V}$: $\tilde{g}_F^{R_V}(\lambda, \gamma) = w^{R_V}(\lambda, \gamma)\ g_F^{R_V}(\lambda, \gamma)$. Next, the filtered projections in the acquisition geometry are obtained from the filtered projections in the virtual geometry by

$$g_F^{R_A}(\lambda, \gamma) = \frac{R_A\cos\gamma}{\sqrt{R_V^2 - R_A^2\sin^2\gamma}}\tilde{g}_F^{R_V}(\lambda + \gamma - \gamma^{R_A}_{R_V}, \gamma^{R_A}_{R_V}) \tag{8}$$

where

$$\gamma^{R_A}_{R_V} = \arcsin\left(\frac{R_A}{R_V}\sin\gamma\right). \tag{9}$$

Finally, the backprojection is performed in the acquisition geometry. For every $\vec{x}$ in the convex hull of the virtual trajectory, we have:

$$f(\vec{x}) = -\int_0^{2\pi} \frac{1}{||R_A\vec{\theta}_\lambda - \vec{x}||} g_F^{R_A}(\lambda, \gamma_{\vec{x},\lambda})\,\mathrm{d}\lambda \tag{10}$$

where

$$\gamma_{\vec{x},\lambda} = \arctan\left(\frac{-\vec{x}\cdot\vec{\eta}_\lambda}{R_A - \vec{x}\cdot\vec{\theta}_\lambda}\right). \quad (11)$$

We can see that this formula is designed for equiangular data $g^{R_A}(\lambda,\gamma)$. As we consider equispaced data $\bar{g}^{R_A}(\lambda,u,v)$ in this paper, an additional rebinning using equation (3) is required before using the VFB formula above.

D. Modifying the VFB method for cone-beam projections

We now detail how the VFB formula above is modified to be used on cone-beam projections. First, we perform a weighting of the cone-beam projections:

$$\bar{g}_W^{R_A}(\lambda,u,v) = \bar{g}^{R_A}(\lambda,u,v)\frac{\sqrt{R_A^2+u^2}}{\sqrt{R_A^2+u^2+v^2}} \quad (12)$$

For an object that is constant in z, (12) ensures that for all v: $\bar{g}_W^{R_A}(\lambda,u,v) = \bar{g}_W^{R_A}(\lambda,u,0)$, so the exact reconstruction area will be extended axially if each row of the weighted projections is treated as the row in the midplane.

Then, for all the weighted data rows $\bar{g}_W^{R_A}(\lambda,u,v)$ of parameter v fixed, we perform the following steps as if the transaxial plane of height $z=v$ was the source plane, using the same virtual source trajectory as in the midplane:

1) Rebinning of the data rows to the virtual geometry:

$$\bar{g}^{R_V}(\lambda,\gamma,v) = \bar{g}_W^{R_A}(\lambda+\gamma-\gamma_{R_A}^{R_V}, R_A\tan\gamma_{R_A}^{R_V}, v) \quad (13)$$

2) Differentiation and Hilbert-filtering of the virtual data:

$$\bar{g}_F^{R_V}(\lambda,\gamma,v) = \frac{1}{2\pi}\int_{-\pi}^{\pi} h_H(\sin(\gamma-\gamma'))\,(\partial_1-\partial_2)\bar{g}^{R_V}(\lambda,\gamma',v)\,\mathrm{d}\gamma' \quad (14)$$

3) Rebinning to acquisition geometry with weighting w^{R_V}:

$$\bar{g}_F^{R_A}(\lambda,u,v) = \frac{R_A\cos\gamma}{\sqrt{R_V^2 - R_A^2\sin^2\gamma}}\,(\bar{w}^{R_V}\bar{g}_F^{R_V})(\lambda+\gamma-\gamma_{R_V}^{R_A},\gamma_{R_V}^{R_A},v) \quad (15)$$

where we take $\bar{w}^{R_V}(\lambda,\gamma,v) = w^{R_V}(\lambda,\gamma)$ for all v, and $\gamma = \arctan(u/R_A)$.

Finally, the backprojection is performed in the acquisition geometry to give $\hat{f}$, the 3D VFB reconstruction :

$$\hat{f}(\vec{x},z) = -\int_0^{2\pi}\frac{1}{||R_A\vec{\theta}_\lambda - \vec{x}||}\bar{g}_F^{R_A}(\lambda,u_{\vec{x},\lambda},v_{z,\lambda})\,\mathrm{d}\lambda \quad (16)$$

for $\vec{x}$ in the convex hull of the virtual source trajectory,

$$u_{\vec{x},\lambda} = \frac{-R_A\vec{x}\cdot\vec{\eta}_\lambda}{R_A - \vec{x}\cdot\vec{\theta}_\lambda} \text{ and } v_{z,\lambda} = \frac{R_A z}{R_A - \vec{x}\cdot\vec{\theta}_\lambda}. \quad (17)$$

III. Experiments and results

A. Simulations

The simulations were performed on a 3D version of the Shepp-Logan phantom and on the 3D head Forbild phantom[1]. The reconstructed image was computed on a cubic grid of size [401, 401, 401] voxels. The data were acquired on a circular trajectory of center $O=(0,0,0)$ and radius R_A, using the software RTK [12]. The projections were transversely truncated such that the FOV was a cylinder of center O and radius R_F. The virtual source trajectory radius was $R_V = R_F$. The acquisition trajectory along $[0,2\pi)$ was sampled with N_λ vertices and each projection was composed of $N_u \times N_v$ ray-lines. The virtual trajectory was composed of $N_{\lambda_{\text{virt}}}$ virtual segments and each virtual projection was composed of $N_{\gamma_{\text{virt}}}$ ray-lines. For the Shepp-Logan phantom, we took $R_A = 4$, $R_F = 0.8$, $N_\lambda = 1256$, $N_u = 409$, $N_v = 517$, $N_{\lambda_{\text{virt}}} = 879$ and $N_{\gamma_{\text{virt}}} = 1257$. For the head Forbild phantom, we took $R_A = 45$, $R_F = 9$, $N_\lambda = 1256$, $N_u = 409$, $N_v = 603$, $N_{\lambda_{\text{virt}}} = 693$ and $N_{\gamma_{\text{virt}}} = 1257$.

[1] See http://www.imp.uni-erlangen.de/phantoms/head/head.html.

B. Results

Figures 5 and 6 show, for three slices of the 3D Shepp-Logan phantom and the head Forbild phantom respectively, the reference image, the reconstructed image using the FDK algorithm with non-truncated data, the reconstructed image using our modified 3D VFB method for transversely-truncated data, and the profiles of the lines drawn in white on the reference and the 3D VFB reconstructions. The mathematically exact reconstruction area (convex hull of the virtual source trajectory), which we also call the recoverable area, is delimited by a black dashed line on the 3D VFB reconstructions.

Looking at figure 5, we can see that the 3D VFB reconstruction is excellent in the recoverable area in the midplane (left column). In the planes at $x=0$ (middle column) and at $y=0.4$ (right column), the reconstruction is still very good when we are close to the midplane. Further away from the midplane, we observe a slow decrease of the intensity when $|z|$ increases, similar to that on the FDK reconstruction, although not exactly the same. There are also slight horizontal streak artefacts, tangent to the white ellipse, which are less marked on the FDK reconstruction.

The 3D VFB reconstruction of the Forbild head phantom (figure 6) is good in the recoverable area in the midplane (left column), but far less accurate that what we obtained for the Shepp-Logan phantom in figure 5. The difference is that the Forbild phantom consists of many more and finer anatomical structures than the Shepp-Logan phantom, making it a far more challenging phantom to reconstruct. Consequently, we observe that the FDK and 3D VFB reconstructions suffer from many artefacts for planes at $x=0$ (middle column) and at $y=-1$ (right column). The artefacts are stronger for the 3D VFB reconstruction, as we observe for instance with the white area at the right of the black ellipse at plane $x=0$ (middle column), and also with the large black horizontal streak covering the top of the two circular structures at plane $y=-1$ (right column).

IV. Conclusion

In this work, we proposed a 3D version of the VFB method, based on a VFB formula performing the backprojection in the circular acquisition geometry and detailed in a previous contribution [11]. This method was used for ROI reconstruction from transversely-truncated cone-beam projections acquired

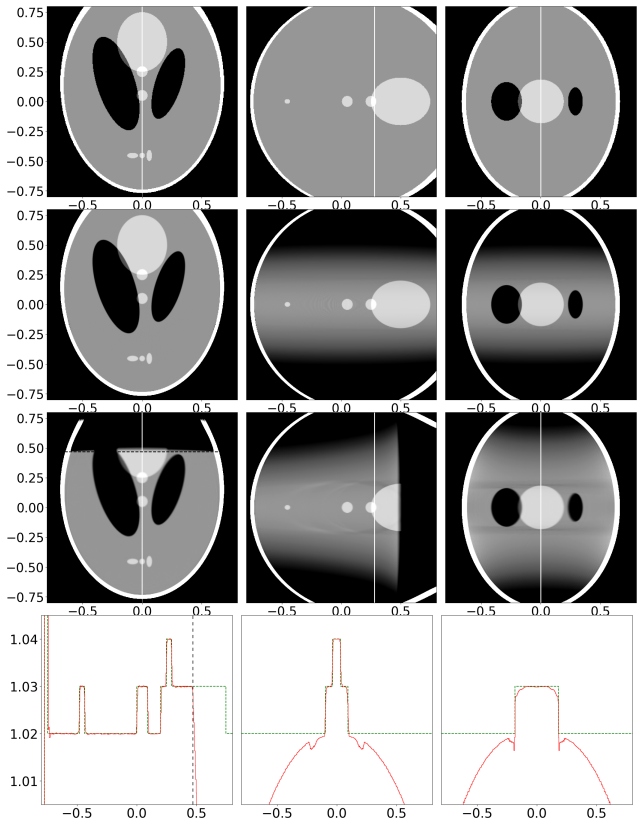


Fig. 5. Left column: (x, y) plane at $z = 0$. Middle column: (y, z) plane at $x = 0$. Right column: (x, z) plane at $y = 0.4$. Top row: 2D slices of the reference Shepp-Logan phantom. Middle row 1: reconstruction using the FDK algorithm without truncation. Middle row 2: reconstruction using our 3D VFB algorithm with truncation. The black dashed line defines the boundary of the possible reconstruction area. The plotting scale is [1.0 (black), 1.04 (white)]. Bottom row: profile corresponding to the white line, plotted with scale [1.005, 1.045]. The reference profile is plotted in green dashed line and the real one in red.

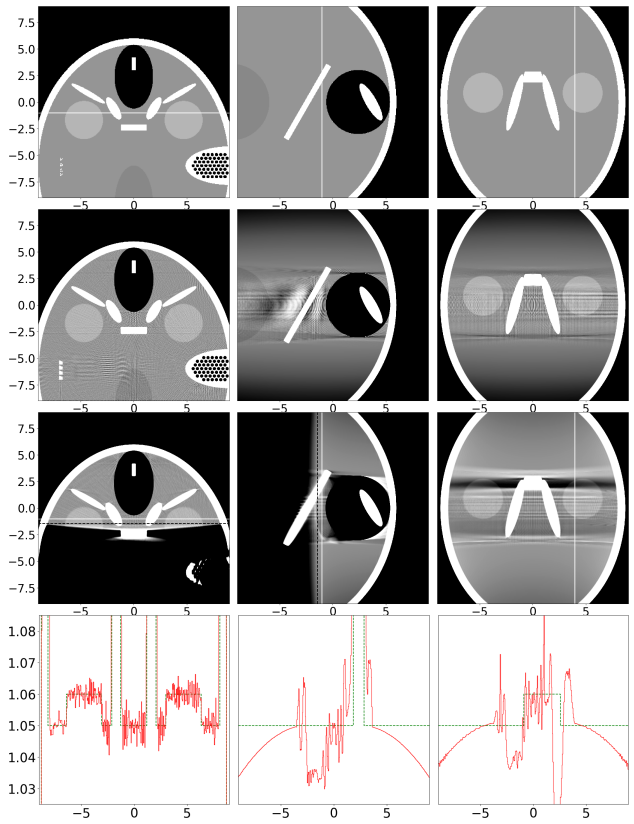


Fig. 6. Left column: (x, y) plane at $z = 0$. Middle column: (y, z) plane at $x = 0$. Right column: (x, z) plane at $y = -1$. Top row: 2D slices of the reference head Forbild phantom. Middle row 1: reconstruction using the FDK algorithm without truncation. Middle row 2: reconstruction using our 3D VFB algorithm with truncation. The black dashed line defines the boundary of the possible reconstruction area. The plotting scale is [1.0 (black), 1.1 (white)]. Bottom row: profile corresponding to the white line, plotted with scale [1.025, 1.085]. The reference profile is plotted in green dashed line and the real one in red.

with a circular source trajectory. The numerical results were satisfactory for the 3D Shepp-Logan phantom but mixed for the more challenging Forbild head phantom, for which strong artefacts appeared that were absent from the FDK truncation-free reconstruction. Both the FDK algorithm and the 3D VFB had to address the incompleteness of a circular cone-beam trajectory, but the 3D VFB was also handling truncated data, so it was not surprising that different artefacts appeared in the off-plane reconstructed images.

Acknowledgment

This work was supported by grants ANR-17-CE19-0006 (ROIdoré) and ANR-21-CE45-0026 (SPECT-Motion-eDCC) from the Agence Nationale de la Recherche, France.

References

[1] H. K. Tuy. An Inversion Formula for Cone-Beam Reconstruction. *SIAM J. Appl. Math.*, 43(3):546–552, 1983.

[2] L. A. Feldkamp, L. C. Davis, and J. W. Kress. Practical cone-beam algorithm. *J. Opt. Soc. Amer. A*, 1(6):612, 1984.

[3] P. S. Cho, A. D. Rudd, and R. H. Johnson. Cone-beam CT from width-truncated projections. *Comput. Med. Imag. Graph.*, 20(1):49–57, 1996.

[4] D. Schäfer, M. Grass, and P. van de Haar. FBP and BPF reconstruction methods for circular X-ray tomography with off-center detector. *Med. Phys.*, 38(S1):S85–S94, 2011.

[5] F. Noo, R. Clackdoyle, and J. D. Pack. A two-step Hilbert transform method for 2D image reconstruction. *Phys. Med. Biol.*, 49(17):3903–3923, 2004.

[6] T. Zhuang, S. Leng, B. E. Nett, and G. Chen. Fan-beam and cone-beam image reconstruction via filtering the backprojection image of differentiated projection data. *Phys. Med. Biol.*, 49(24):5489–5503, 2004.

[7] Y. Zou, X. Pan, and E. Y. Sidky. Image reconstruction in regions-of-interest from truncated projections in a reduced fan-beam scan. *Phys. Med. Biol.*, 50(1):13–27, 2004.

[8] R. Clackdoyle, F. Noo, J. Guo, and J. A. Roberts. Quantitative reconstruction from truncated projections in classical tomography. *IEEE Trans. Nucl. Sci.*, 51(5):2570–2578, 2004.

[9] L. Yu, Y. Zou, E. Y. Sidky, C. A. Pelizzari, P. Munro, and X. Pan. Region of interest reconstruction from truncated data in circular cone-beam CT. *IEEE Trans. Med. Imag.*, 25(7):869–881, 2006.

[10] F. Noo, M. Defrise, R. Clackdoyle, and H. Kudo. Image reconstruction from fan-beam projections on less than a short scan. *Phys. Med. Biol.*, 47(14):2525–2546, 2002.

[11] M. Charles, R. Clackdoyle, and S. Rit. Implementation of the virtual fan-beam method for 2D region-of-interest reconstruction from truncated data. In *Fully 3D Image Recon. Radiol. Nucl. Med.*, pages 44–48, 2021.

[12] S. Rit, M. Vila Oliva, S. Brousmiche, R. Labarbe, D. Sarrut, and G. C. Sharp. The Reconstruction Toolkit (RTK), an open-source cone-beam CT reconstruction toolkit based on the Insight Toolkit (ITK). *Journal of Physics: Conference Series*, 489:012079, 2014.

Iterative image reconstruction for CT with unmatched projection matrices using the generalized minimal residual algorithm

Emil Y. Sidky, Per Christian Hansen, Jakob S. Jørgensen, and Xiaochuan Pan

***Abstract*—The generalized minimal residual (GMRES) algorithm is applied to image reconstruction using linear computed tomography (CT) models. The GMRES algorithm iteratively solves square, non-symmetric linear systems and it has practical application to CT when using unmatched back-projector/projector pairs and when applying preconditioning. The GMRES algorithm is demonstrated on a 3D CT image reconstruction problem where it is seen that use of unmatched projection matrices does not prevent convergence, while using an unmatched pair in the related conjugate gradients for least-squares (CGLS) algorithm leads to divergent iteration. Implementation of preconditioning using GMRES is also demonstrated.**

***Index Terms*—Linear iterative image reconstruction, GMRES, unmatched projector/back-projector, preconditioning**

I. Introduction

LINEAR models for computed tomography (CT) play an important role for iterative image reconstruction. The most common approach to CT processing involves taking the negative logarithm of the projection data, so that the line integration model leads to a linear relation between the image and processed data. Accordingly, the CT image reconstruction problem can be written as a large linear system

$$Ax = b, \tag{1}$$

where b, a vector of length m, represents the processed projection data; x, a vector of length n, contains the image pixel values; and the $m \times n$ system matrix A contain the weights that model line-integration. Linear tomographic models can include quadratic regularization, cf. [1] and [2, Chapter 12], or more sophisticated modeling such as noise correlation and blur due to accurate detector physics [3]. Even when non-linear models for CT [4] are considered for iterative image reconstruction, there is usually a large linear system that is involved in the algorithm. Novel techniques for solving large linear systems may thus be of practical use for iterative image reconstruction in CT.

The most common iterative algorithm for solving linear CT models, excluding row-action, sequential, or SIRT-type data processing methods, is the conjugate gradients (CG) algorithm [2, Chapter 11]. For least-squares problems with non-symmetric system matrices, in particular, there is the conjugate-gradients least-squares (CGLS) algorithm that solves the optimization problem

$$\min_x \frac{1}{2}\|Ax - b\|_2^2. \tag{2}$$

The minimizer of this optimization problem can also be found from solving the normal equations directly

$$A^\top Ax = A^\top b, \tag{3}$$

which are derived from (2) by taking the gradient of the objective function and setting it to zero. In applying CGLS, the implementation for back-projection B must be the matrix transpose $A^\top$. If $B \neq A^\top$ then the resulting method is not well defined, there is no convergence theory, and if it does converge it does not solve Eqs. (2) and (3). Nevertheless there are practical motivations for considering back-projection implementations B different than $A^\top$. These motivations are outlined in Ref. [5] in connection with SIRT-type iterative solvers, where the authors explain that B can be a preconditioner, B may be an efficient but approximate implementation of $A^\top$, or A may involve complex physics modeling that may make computer implementation of $A^\top$ prohibitively expensive. As shown in [6] we can guarantee convergence of SIRT-type methods with $B \neq A^\top$ (but not CGLS) by shifting the complex eigenvalue spectrum of BA so that eigenvalues with negative real part are eliminated; but it forces a modification of the problem that is being solved.

Use of the GMRES algorithm allows for use of back-projectors B that are not equal to $A^\top$ without modification of the desired reconstruction model. Furthermore, the algorithm does not involve any parameters other than the iteration number. In Sec. II we present the ABBA framework [7] which involves two forms of GMRES called AB-GMRES and BA-GMRES. In Sec. III we demonstrate use of BA-GMRES for unmatched projector/back-projector pairs and for preconditioning. We conclude this abstract in Sec. IV.

II. The ABBA framework

The GMRES algorithm solves a linear system

$$Sx = v,$$

where the coefficient matrix S is a square matrix that is not necessarily symmetric. The relevance for CT image reconstruction is that a square non-symmetric matrix arises when multiplying unmatched back-projection B and projection A

E. Y. Sidky and X. Pan are with the Department of Radiology, University of Chicago, Chicago, IL, 60637 USA. emails: sidky@uchicago.edu, xpan@uchicago.edu

P. C. Hansen and J. S. Jørgensen are with the Department of Applied Mathematics and Computer Science, Technical University of Denmark, Kgs. Lyngby, DK-2800, Denmark. emails: pcha@dtu.dk, jakj@dtu.dk

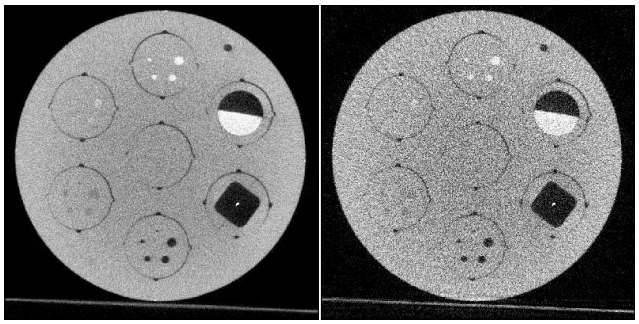


Fig. 1. Mid-slice images of QA phantom. (Left) FBP reconstructed image from 720 views. (Right) FBP reconstructed images from 180 views, a 4-fold sub-sampling of the original CBCT dataset. The grayscale window is [0.,0.25] cm^{-1}.

matrices; i.e., both AB and BA are square non-symmetric matrices. The original linear system of interest, Eq. (1), cannot be directly solved with GMRES because A is not necessarily square for CT, but this equation can be modified to

$$ABy = b, \quad x = By, \tag{4}$$

where the unknown vector y has the same length as the projection data b and the resulting $m \times m$ matrix AB is square. Also, the normal equations in Eq. (3) can be modified by replacing $A^\top$ with B

$$BAx = Bb, \tag{5}$$

and again the resulting $n \times n$ matrix BA is square. See [8] for details. Modeling CT with Eq. (4) is similar to the use of natural pixels [9]–[11] as the image is expressed as the back-projection of a sinogram.

We refer to the GMRES algorithms for solving Eqs. (4) and (5) as AB-GMRES and BA-GMRES, respectively. The pseudo-code for both algorithms is given in Ref. [7], and we briefly describe the algorithms here. Similar to CGLS, GMRES is a Krylov subspace method, where the basis vectors of the subspace are generated by choosing an initial vector and repeatedly applying the coefficient matrix (AB or BA) to obtain new linearly independent basis vectors. For AB-GMRES or BA-GMRES with a zero initial vector, the first basis vector is b or Bb, respectively, and subsequent basis vectors are generated by applying the matrix AB or BA, respectively. The GMRES algorithm involves orthonormalization of the Krylov subspace vectors to obtain a orthonormal basis set that spans the subspace. At each GMRES iteration the dimension of the subspace is increased by one and the minimum residual that can be expressed by the Krylov basis set is found.

The computational burden of GMRES lies with the fact that the Krylov basis set must be stored and the number of basis vectors is the same as the number iterations. For AB-GMRES and BA-GMRES the size of one basis vector is the same as the size of a sinogram and image, respectively. In our implementation, the basis set is stored on the computer disk. Restart methods [12] can reduce the basis vector storage burden, but for this work we demonstrate the basic GMRES implementation.

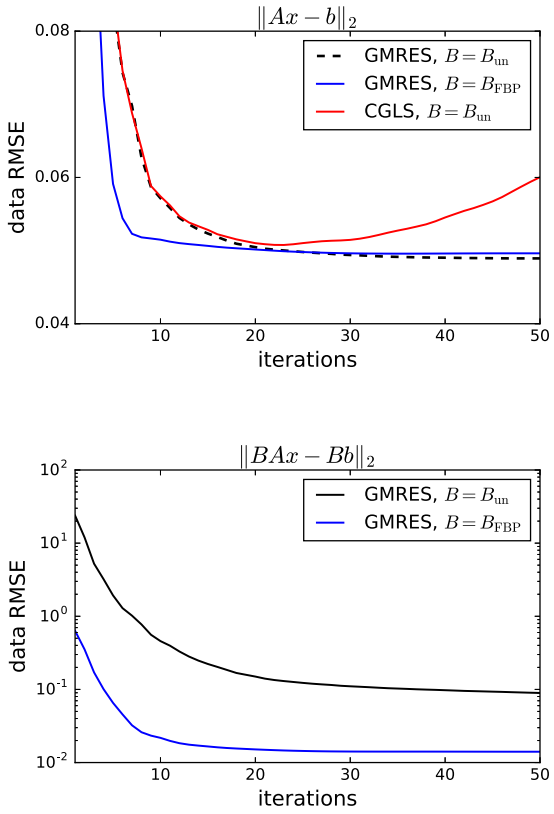


Fig. 2. Data RMSE in the form of (Top) $\|Ax-b\|_2$ and (Bottom) $\|BAx-Bb\|_2$. In the top graph unmatched CGLS is also shown to demonstrate divergence of the data RMSE with unmatched projector/back-projector pairs. The other two curves correspond to use of voxel-driven back-projection $B=B_{\text{un}}$, and FBP $B=B_{\text{FBP}}$. The projector A is a ray-driven implementation.

The AB-GMRES and BA-GMRES algorithms are guaranteed to minimize different data discrepancy measures. In the case of AB-GMRES, the algorithm minimizes

$$\|ABy-b\|_2,$$

while BA-GMRES minimizes

$$\|BAx-Bb\|_2.$$

Note that BA-GMRES is not necessarily minimizing

$$\|Ax-b\|_2.$$

In this work we focus on BA-GMRES and we demonstrate its use on cone-beam CT image reconstruction.

III. BA-GMRES applied to cone-beam CT image reconstruction

We apply BA-GMRES to a cone-beam CT (CBCT) data set acquired on an Epica Pegaso veterinary CT scanner. The particular scan configuration for the data set is 180 projections taken uniformly over one circular rotation. The detector size is 1088×896 detector pixels, where each pixel is $(0.278\text{mm})^2$ in size. The 180-view dataset is sub-sampled from a 720-view

scan of a quality assurance (QA) phantom. Image volumes are reconstructed onto a $1024 \times 1024 \times 300$ voxel grid, and a reference volume is generated by use of filtered back-projection (FBP) applied to the full 720-view dataset and shown in Fig. 1. Also shown in the figure is FBP applied to the 180-view sub-sampled dataset.

To demonstrate application of BA-GMRES to CBCT image reconstruction, we use a ray-driven cone-beam projector where the matrix elements for A are computed by the line-intersection method. We consider two implementations of B: (1) B_{un} voxel-driven back-projection using linear interpolation to determine the appropriate projection value on the detector, and (2) $B_{\text{FBP}} = B_{\text{un}}F$ filtered back-projection, where F represents the ramp filter. See [2, Chapter 9] for details about these discretization models. The first BA-GMRES implementation tests unmatched back-projector/projector pairs where $B_{\text{un}} \approx A^{\top}$, and the second implementation includes the additional ramp-filtering step for preconditioning. Use of B_{un} and A is also shown for CGLS, which requires that $B = A^{\top}$.

The data root-mean-square-error (RMSE) curves for both forms of BA-GMRES and CGLS using B_{un} and A are shown in the top panel of Fig. 2. The CGLS result initially shows convergence of the data RMSE, but after 20 iterations the data RMSE begins to diverge with increasing iteration number, as expected, since this algorithm is not designed to work for unmatched matrix transpose implementations. The corresponding BA-GMRES result does show a decreasing data RMSE with iteration number. For the preconditioned form of BA-GMRES, the decrease in data RMSE is even more rapid. The decreasing trends in $\|Ax - b\|_2$ for BA-GMRES occur even though this algorithm is not guaranteed to reduce this data norm. Also shown in Fig. 2 is the data RMSE curves for $\|BAx - Bb\|_2$, which is guaranteed to decrease with iteration number and they do indeed show decreasing trends for BA-GMRES. These issues are elaborated in [7].

The mid-slice images for BA-GMRES using both B implementations are shown in Fig. 3 at different iteration numbers. Preconditioning has a clear effect on the convergence as all the phantom structures are clearly visible in the early iterations and the gray-level is stabilized already at the fifth iteration. The BA-GMRES result without preconditioning is also fairly efficient as the main features of the QA phantom are visible at 20 iterations.

One measure of image quality is to compare the reconstructed volumes to a ground truth image. Employing the 720-view FBP reconstructed volume as a surrogate for the ground truth, the image RMSE is plotted in Fig. 4 for both versions of BA-GMRES. The BA-GMRES implementation with $B = B_{\text{un}}$ achieves a minimum image RMSE of 0.0201 at iteration 29, while the preconditioned version with $B = B_{\text{un}}$ achieves a minimum image RMSE of 0.0210 at iteration 4. For comparison the 180-view FBP result has an image RMSE of 0.0347. To appreciate the various image qualities, ROI images of the mid-slice are shown at the minimum image RMSE iteration numbers in Fig. 5.

That the image RMSE has a minimum at finite iteration number is a well-known phenomenon in iterative image reconstruction and it is known as semi-convergence [2, Chapter 11].

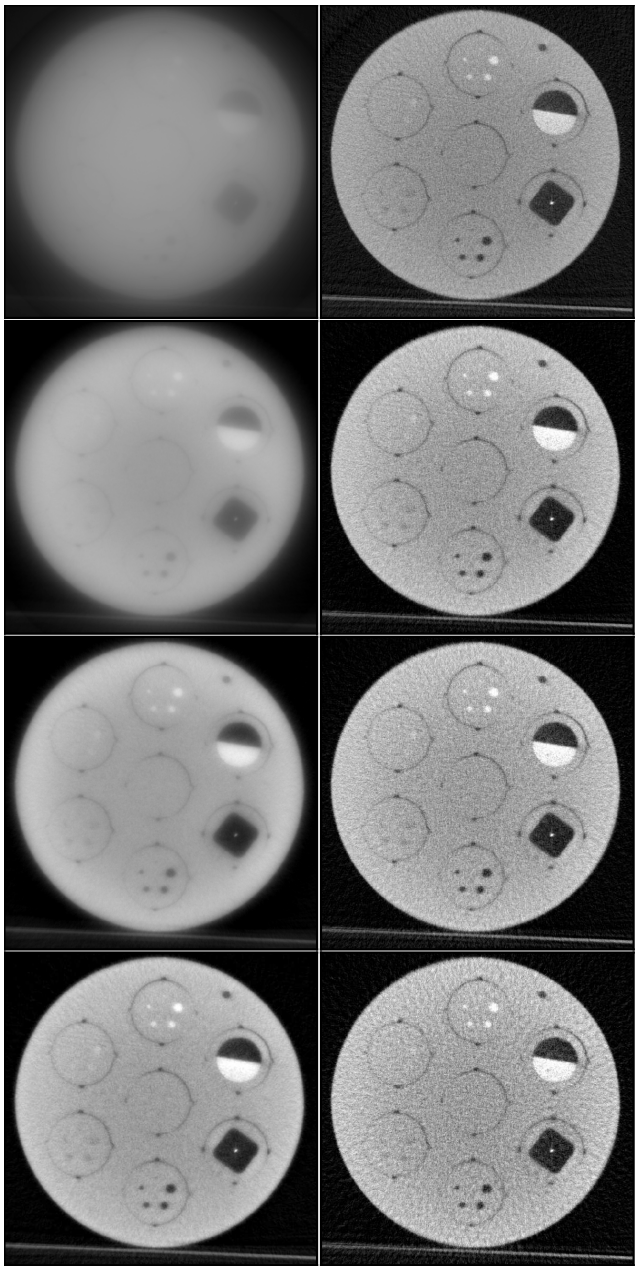


Fig. 3. Mid-slice BA-GMRES images for voxel-driven back-projection $B = B_{\text{un}}$ (Left column) and FBP $B = B_{\text{FBP}}$ (Right column). The shown iteration numbers are 2, 5, 10, and 20 going from the top row to bottom row. The grayscale window is [0.,0.25] cm^{-1}.

Early stopping in such algorithms is a form of regularization because the components associated with large singular values of A converge fast, while the unwanted noisy components associated with smaller singular values – that cause strong image artifacts – appear after more iterations. Semi-convergence is observed in the image RMSE curves of Fig. 4 and visually in the preconditioned BA-GMRES series of Fig. 3 where the image at 20 iterations clearly shows strong artifacts from iterating too far. The semi-convergence issue also presents a practical dilemma for preconditioning. With the shown pre-

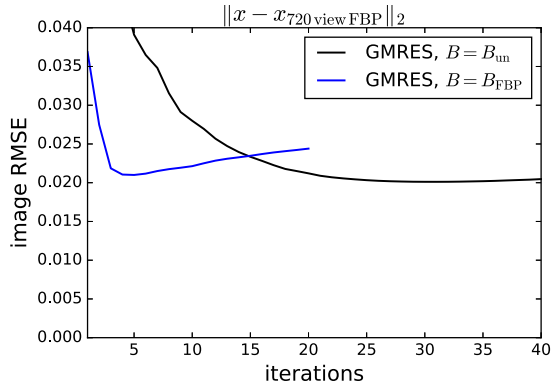


Fig. 4. Using the 720-view reconstructed volume (see mid-slice image on the left of Fig. 1 as a reference), the BA-GMRES reconstructed image RMSE is plotted as a function of iteration number for $B = B_{\text{un}}$ and $B = B_{\text{FBP}}$.

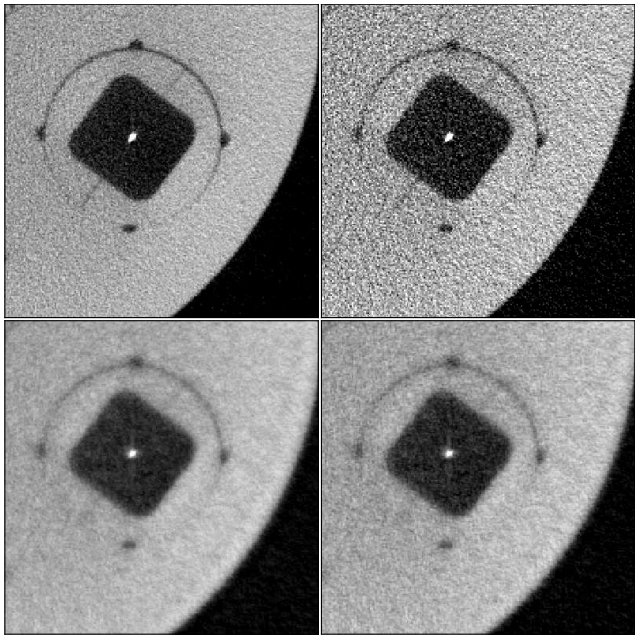


Fig. 5. Mid-slice ROI images of QA phantom. (Top, Left) FBP reconstructed image from 720 views. (Top, Right) FBP reconstructed images from 180 views. (Bottom, Left) BA-GMRES image for $B = B_{\text{un}}$ at iteration 29. (Bottom, Right) BA-GMRES image for $B = B_{\text{FBP}}$ at iteration 4. The grayscale window is [0.,0.25] cm^{-1}.

conditioned BA-GMRES results, the minimum image RMSE result is obtained already at the fourth iteration; thus the iteration number provides only coarse control over its image quality. The un-preconditioned BA-GMRES implementation achieves its image RMSE minimizer at the 29th iteration, which is computationally less efficient, but on the other hand the iteration number provides a finer control over the image quality. In any case, the BA-GMRES framework provides a flexible means for implementing back-projectors or preconditioning schemes, and optimizing the B implementation and iteration number will depend on the imaging task of interest.

IV. Conclusion

This work presents an iterative image reconstruction framework for linear CT problems that allows for the use of unmatched back-projector/projector pairs in a straight-forward manner. This possibility is convenient for implementation of efficient back-projectors, linear modeling of complex physics, and preconditioning. Also, because it is clear what equation is being solved when $B \neq A^\top$, BA-GMRES can be used for solving linear sub-problems that may arise in non-linear iterative image reconstruction. The BA-GMRES algorithm does present a challenge for computer memory because the Krylov basis set needs to be stored during the iteration, but the present demonstration on CBCT image reconstruction does show that BA-GMRES can be applied to large-scale CT systems of clinical interest.

Acknowledgment

The authors wish to thank Holly Stewart and Christopher Kawcak from Colorado State University for providing the QA phantom data. This work was supported in part by the Grayson-Jockey Club Research Foundation, NIH Grant Nos. R01-EB026282 and R01-EB023968, and a Villum Investigator grant (no. 25893) from The Villum Foundation. The contents of this article are solely the responsibility of the authors and do not necessarily represent the official views of the National Institutes of Health.

References

[1] J. A. Fessler, "Penalized weighted least-squares image reconstruction for positron emission tomography," *IEEE Trans. Med. Imag.*, vol. 13, pp. 290–300, 1994.

[2] P. C. Hansen, J. S. Jørgensen, and W. R B. Lionheart, Eds., *Computed Tomography: Algorithms, Insight, and Just Enough Theory*, SIAM, Philadelphia, 2021.

[3] S. Tilley, M. Jacobson, Q. Cao, M. Brehler, A. Sisniega, W. Zbijewsk, and J. W. Stayman, "Penalized-likelihood reconstruction with high-fidelity measurement models for high-resolution cone-beam imaging," *IEEE Trans. Med. Imag.*, vol. 37, pp. 988–999, 2017.

[4] I. A. Elbakri and J. A. Fessler, "Statistical image reconstruction for polyenergetic X-ray computed tomography," *IEEE Trans. Med. Imag.*, vol. 21, pp. 89–99, 2002.

[5] G. L. Zeng and G. T. Gullberg, "Unmatched projector/backprojector pairs in an iterative reconstruction algorithm," *IEEE Trans. Med. Imag.*, vol. 19, pp. 548–555, 2000.

[6] Y. Dong, P. C. Hansen, M. E. Hochstenbach, and N. A B. Riis, "Fixing nonconvergence of algebraic iterative reconstruction with an unmatched backprojector," *SIAM J. Sci. Comput.*, vol. 41, pp. A1822–A1839, 2019.

[7] P. C. Hansen, K. Hayami, and K. Morikuni, "GMRES methods for tomographic reconstruction with an unmatched back projector," 2021, arxiv preprint 2110.01481.

[8] T. Elfving and P. C. Hansen, "Unmatched projector/backprojector pairs: perturbation and convergence analysis," *SIAM J. Sci. Comput.*, vol. 40, pp. A573–A591, 2018.

[9] M. H. Byonocore, W. R. Brody, and A. Macovski, "A natural pixel decomposition for two-dimensional image reconstruction," *IEEE Trans. Biomed. Eng.*, pp. 69–78, 1981.

[10] C. Riddell, "Tomographic reconstruction through view-based implementation of least-square algorithms," in *Proc. 1nd Intl. Mtg. on image formation in X-ray CT*, 2010, pp. 220–223.

[11] S. D. Rose, E. Y. Sidky, and X. Pan, "TV constrained CT image reconstruction with discretized natural pixels," in *2016 IEEE Nuclear Science Symposium, Medical Imaging Conference and Room-Temperature Semiconductor Detector Workshop (NSS/MIC/RTSD)*, 2016, pp. 1–3.

[12] R. B. Morgan, "GMRES with deflated restarting," *SIAM J. Sci. Comp.*, vol. 24, pp. 20–37, 2002.

Deep Learning-Based Detector Row Upsampling for Clinical Spiral CT

Jan Magonov, Julien Erath, Joscha Maier, Eric Fournié, Karl Stierstorfer, and Marc Kachelrieß

***Abstract*—Due to longitudinal undersampling multislice spiral computed tomography (MSCT) scans may suffer from windmill artifacts in reconstructed images. To fulfill the sampling condition and achieve double sampling in z-direction, some CT scanners use the z-flying focal spot (zFFS) technique, a hardware-based solution that effectively doubles the number of detector rows. To obtain a software-based solution we developed a convolutional neural network that is trained in a supervised manner with clinical projection raw data that were acquired with zFFS enabled. We presented this approach as the row interpolation with deep learning (RIDL) network. In this work we simplified the network architecture, extended the clinical dataset and generated an experimental synthetic dataset consisting of two-dimensional projection data. We were able to observe a reduction in windmill artifacts for both datasets used for training. Especially the synthetic dataset is very promising as we could observe an increased reduction of artifacts with this dataset.**

I. Introduction

Multislice spiral computed tomography (MSCT), also known as multidetector CT, has become an integral part of modern medical imaging after the theoretical introduction of spiral CT in 1989 [1]. The most common application of these systems is spiral scanning, in which the patient is continuously moved through the gantry, resulting in shorter scan times and higher temporal resolution [2]. Nevertheless, artifacts can occur with this modality that degrade quality of reconstructed images. The windmill artifact is an image distortion in the axial plane whose appearance is characterized by bright streak-like patterns emerging from high contrast structures along the longitudinal axis [3]. When scrolling through the reconstructed slices these streaks appear to rotate. The cause of this artifact can be attributed to inadequate data sampling in the z-plane resulting in not satisfying the Nyquist condition and thus leading to aliasing [3], [4].

A hardware-based method to fulfill the sampling condition and reduce windmill artifacts is provided by the z-flying focal spot (zFFS) [2], [4]. This technique doubles the effective number of detector rows acquired during the scan by periodically deflecting the X-ray focal spot in longitudinal direction. The resulting higher sampling rate in z-direction reduces the occurrence of windmill artifacts. Figure 1 shows a scan acquired without zFFS compared to a corresponding scan with zFFS enabled.

Jan Magonov and Julien Erath are with the German Cancer Research Center (DKFZ), Heidelberg, Germany, with Siemens Healthcare GmbH, Forchheim, Germany and with the Heidelberg University, Heidelberg, Germany.

Eric Fournié and Dr. Karl Stierstorfer are with Siemens Healthcare GmbH, Forchheim, Germany.

Dr. Joscha Maier and Prof. Dr. Marc Kachelrieß are with the German Cancer Research Center (DKFZ), Heidelberg, Germany and with the Heidelberg University, Heidelberg, Germany.

Corresponding author: Jan Magonov (jan.magonov@dkfz.de)

Fig. 1: Reduction of windmill artifacts by using zFFS. The left image was taken without zFFS (32×0.6 mm collimation, pitch 1.4) while for the right reconstructed image ($2 \cdot 32 \times 0.6$ mm collimation, pitch 1.4) zFFS was enabled for acquisition (C = 0 HU, W = 200 HU).

However, this method also has some drawbacks, as it is technically complex and thus prevents the use of zFFS in CT systems that do not meet these requirements. Previous works, such as in [5], focus on the reduction of windmill artifacts in the image domain. In contrast we try to solve the problem in projection domain. In [6] we presented the row interpolation with deep learning (RIDL) network, which was similar to the zFFS designed to double the effective number of acquired raw detector rows in projection domain. The network was based on the SRResNet presented in [7] to compute super-resolution images, i.e. very high-resolution images. In this paper we simplified the network architecture in order to reduce complexity of training process while maintaining existing results. Furthermore, the clinical dataset used for network training was extended and an experimental synthetic dataset was generated. Two separate networks were trained with the individual datasets and the network predictions were compared in image domain by reconstructing two clinical spiral CT scans.

II. Methods and Material

A. Clinical Data Preparation

For the clinical dataset, we selected raw projection data from a total of 40 clinical CT scans from different patients. The scans covered different body regions such as head, thorax and abdomen and were acquired with Somatom Flash and Somatom Force dual source CT scanners (Siemens Healthineers, Forchheim, Germany) with zFFS enabled. The dataset was split into two disjoint subsets so that 32 of the scans were used as training dataset and 8 scans served as validation dataset. It

was ensured that the different body regions and CT systems used in the images were equally distributed in both datasets. Before training, some preprocessing steps were performed, i.e. instead of using the complete projection data of the scans as a whole, randomized image patches were generated to simplify the training process.

B. Synthetic Data Preparation

In addition to the clinical dataset, the acquisition of synthetic data for training the RIDL network was investigated. An advantage of using synthetic data would be that any number of data could be generated without requiring a CT scanner with zFFS. For the simulation we used the software package CT_SIM which is based on the deterministic ray propagation simulation software Deterministic Radiological Simulation (DRASIM). These tools allow to simulate the radiation properties in a defined X-ray imaging setup through geometrically defined phantoms [8]. In our first experimental setup, we generated two-dimensional projection data containing projections of overlapping water spheres with varying densities (0.5 - 3.0 $\mathrm{g/cm^3}$) and diameters (1 - 20 cm). Each of these water spheres was overlaid with a smaller water sphere of density 1.0 $\mathrm{g/cm^3}$, resulting in narrow circular edges with a width of 0.3 to 2 mm. These structures are particularly difficult to interpolate. Figure 2 shows an example representation of a projection from the synthetic dataset. A total of 200,000 projections with 80 detector rows and 800 channels containing randomly arranged water spheres were simulated. Comparable to real clinical projection data acquired with zFFS, the detector rows were simulated overlapped. No noise was added to the data. The dataset was split into a training dataset with 160,000 projections and a validation dataset with 40,000 projections. In addition, the range of values of the synthetic projection data was linearly scaled to the value range of the clinical data. Similar to the clinical dataset, random image patches were selected from the projection data for network training, as we will describe in more detail below.

C. Row Interpolation with RIDL-CNN

In the previous approach of our work, a neural network was trained that received raw projection data and generated a prediction of the input with interpolated rows to effectively double the number of rows. The clinical projection data used to train the network were obtained after the rebinning, which is the rearrangement of the measured fan-beam data to parallel-beam geometry. These projections were then divided into alternative rows so that projections containing all rows (acquired with zFFS) were used as the desired output y, and every other row from the corresponding projections was used for the network input x in training. In order to predict an upsampled version of the input data using the network, a so-called subpixel convolutional layer [9] was used, which essentially performs an upsampling of the generated feature maps within the network by a specific type of image reshaping. However, this procedure is time-consuming in network training, as well as in the subsequent use of the trained network for the prediction of rows.

Fig. 2: Example projection from the synthetic dataset with different sized water spheres consisting of 80 overlapped detector rows and 800 channels.

In further experiments, we could observe that a much simpler convolutional neural network without subpixel convolution is able to produce results comparable to the RIDL-SRResNet. In the following, we will refer to this network architecture as RIDL-CNN. Similar to the previous architecture, random patches with the size of $64 \times 32 \times 1$ pixels were generated from the underlying projection data to train the network. Also in this case, every other row from these patches serves as network input so that it has a size of $32 \times 32 \times 1$. The desired output has the same dimension and is obtained from the intermediate rows in the generated patches. Before network training, all patches were linearly normalized to a value range in the interval from 0 to 1. Slope and offset were set according to the minimum and maximum value of the clinical dataset. After network prediction, the input and output rows have to be interlaced to obtain corresponding interpolated projections. Furthermore, the value range of these projections must be denormalized to the original range of the clinical data. In total, the RIDL-CNN consists of an input layer followed by 12 convolutional layers with 128 filters and 3×3 kernels. The network output is computed by a final convolution. The number of trainable parameters is 1,625,857.

D. Implementation and Training

The RIDL-CNN was trained on both the clinical and synthetic dataset, resulting in two separately trained networks. For both datasets, 500,000 examples were selected from the corresponding training dataset and 125,000 from the corresponding validation dataset. For the training we used the Adam optimizer and a combined loss function that is described by:

$$L_{\text{comb}}(y, \hat{y}) = \alpha \cdot L_{\text{MS-SSIM}}(y, \hat{y}) + (1 - \alpha) \cdot L_{\text{MAE}}(y, \hat{y})$$

This loss function was proposed in [10] and takes into account the pixel-wise computed error between the network output $\hat{y}$ and ground truth y by the mean absolute error (MAE) but also the structural similarity between the two images by the multi-scale structural similarity index (MS-SSIM). The weighting factor was empirically determined as $\alpha = 0.84$. The networks were trained with a batch size of 256 and an initial learning rate set to 1×10^{-5}, which was halved if the error could not be minimized after 25 consecutive epochs.

E. Evaluation and Validation

In order to evaluate and validate the results, two scans of a skull phantom with real human bones were acquired with a Somatom Force system. For the first scan, a basic scan mode with a collimation of 96×0.6 mm and activated zFFS was used. For the second scan we used a scan mode of the CT

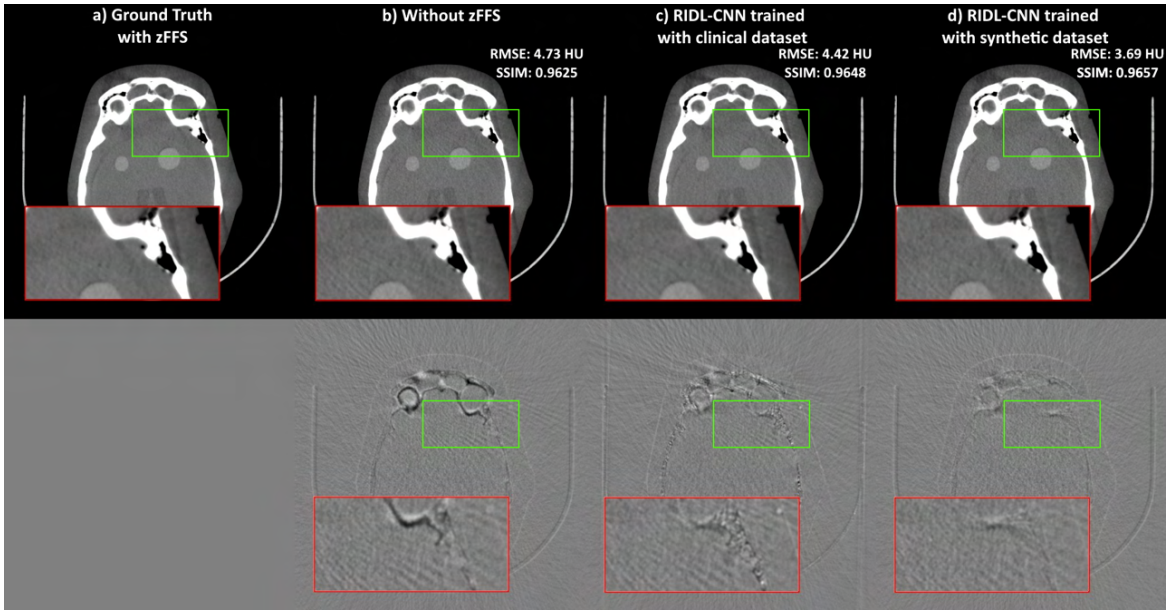


Fig. 3: Qualitative and quantitative comparison of a reconstructed slice (1. scan) without zFFS, with the RIDL-CNN trained with the clinical dataset and the RIDL-CNN trained with the synthetic dataset compared to the ground truth scan with zFFS (C = 60 HU, W = 360 HU). Below the difference images to the ground truth are shown (C = 0 HU, W = 150 HU).

system with a collimation of 48×1.2 mm. In this mode, no zFFS can be enabled and the acquired images show very strong windmill artifacts due to the lower sampling in the z-direction. In both scans the pitch factor was set to 1, since especially scans with pitch values in this range suffer from windmill artifacts [2], [3]. A summary for both scan settings can be found in table I.

Scan	Collimation	Pitch	zFFS	Reconstructed slices
1	96×0.6 mm	1.0	yes	1.0 mm
2	48×1.2 mm	1.0	no	1.5 mm

TABLE I: Summary of the settings for the two scans used to evaluate and validate the results.

As in our previous work, the trained networks were applied to reconstruct these clinical CT scans. For this purpose, the plugin we developed for the Siemens-specific reconstruction software was used to adjust the raw projection data after the rebinning. In the first scan, every second row, i.e. the zFFS-generated rows were replaced by rows predicted by the RIDL networks. In addition, a reconstruction with linear interpolated rows was performed, which should correspond to an acquisition without zFFS. For all reconstructions, the error measures RMSE and SSIM were calculated in relation to the ground truth reconstruction with zFFS enabled. Since no zFFS can be used in the acquisition setting employed in the second scan, there is no ground truth data. The results in these reconstructions can therefore only be evaluated qualitatively. In this case, the number of rows in the raw data was doubled by extending them with predictions from the RIDL networks.

III. Results

A. *Scan with* 96×0.6 *mm Collimation*

Figure 3 shows reconstructions of a specific slice with differently modeled projection row data from the first scan. Difference images are calculated to the ground truth data acquired with zFFS. Comparing the reconstruction without zFFS to the result of the RIDL-CNN trained on the clinical dataset, only a very slight reduction of the windmill artifacts can be seen in the image domain. Furthermore, there are noisy structures noticeable in the difference image in Figure 3c in the area of the bones. However, MSE and SSIM indicate a quantitatively slightly better result compared to omitting the zFFS. Looking at the reconstruction with the network trained with synthetic data (see Figure 3d), we find an improvement in the image quality both in the image domain and in the difference image. Especially the problem with the noisy structures in the bone areas does not occur. With regard to the error measures, this reconstruction also provides the best result quantitatively.

B. *Scan with* 48×1.2 *mm Collimation*

Figure 4 compares the results for two reconstructed slices from the second scan. In both slices reconstructed with WFBP without zFFS, very dominant windmill artifacts can be observed. Comparing these slices with the results of the network trained with clinical data, a slight reduction of the artifacts can be seen qualitatively. The results obtained with the network trained with the synthetic dataset can most effectively reduce the occurring windmill artifacts and provide superior image quality compared to the network trained with clinical data. The comparison of the reconstructions can only be performed qualitatively due to missing ground truth data, since the applied scan mode does not allow for enabling zFFS.

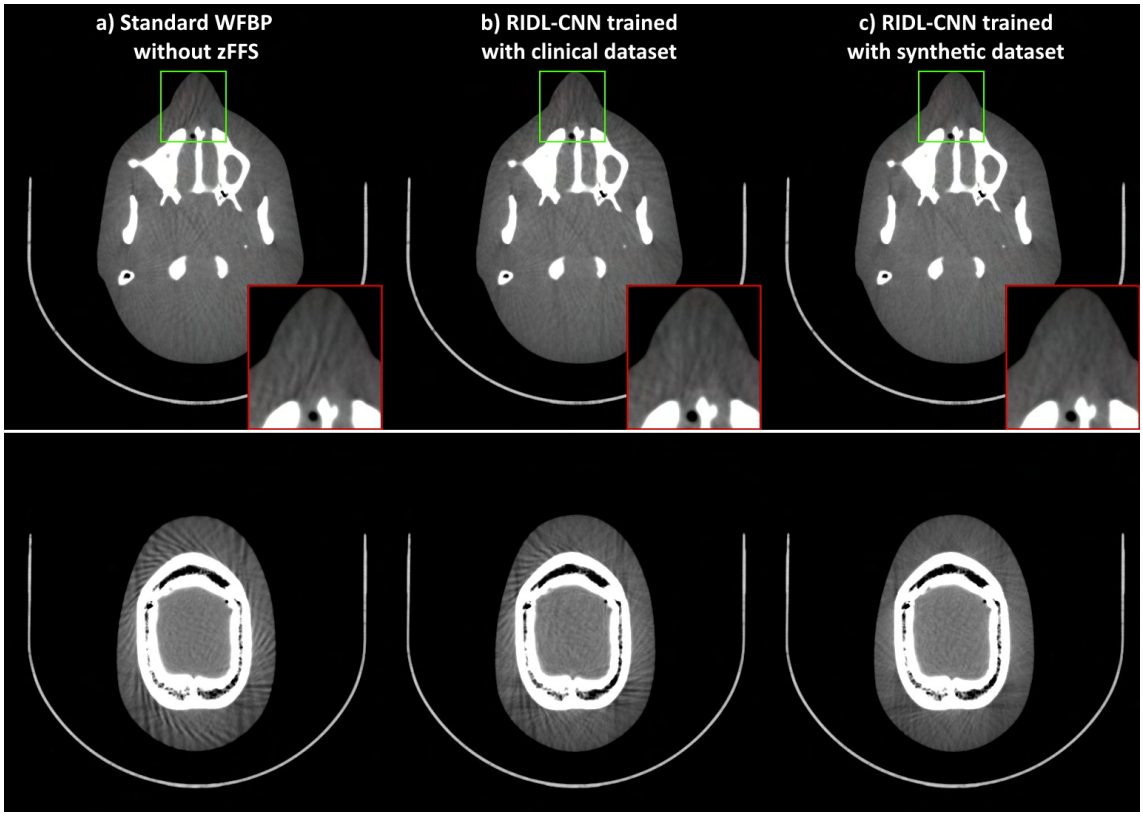


Fig. 4: Qualitative comparison of two selected slices (2. scan) without zFFS, with the RIDL-CNN trained with the clinical dataset and the RIDL-CNN trained with the synthetic dataset (C = 60 HU, W = 360 HU).

IV. Discussion and Conclusion

In this work, we further adapted our RIDL network and simplified the network architecture. In addition, we extended the clinical dataset and generated an experimental synthetic dataset. This was done by simulating two-dimensional raw data containing different sized overlapping spherical structures. In our experiments presented here, we observed that the results with the synthetic data are very promising. Although no clinical data were included in this dataset, windmill artifacts were reduced more effectively than with the RIDL-CNN trained with the current setup of clinical data. This observation suggests that training with clinical data can still be optimized. One problem could be the noise in clinical projection data. Denoising the clinical data before network training could be considered. However, it is valuable that training with synthetic data can address the problem of windmill artifacts, without having to rely on raw clinical projection data acquired with a CT system that supports zFFS. The next step is to investigate how the synthetic dataset can be adapted more efficiently to our task. In addition, data with a concrete CT system geometry will be simulated. Furthermore, we will optimize the training with clinical data and investigate whether the results can be improved by a combination of synthetic and clinical data.

References

[1] W. A. Kalender, W. Seissler, E. Klotz, and P. Vock, "Spiral volumetric CT with single–breath–hold technique, continuous transport, and continuous scanner rotation," *Radiology*, vol. 176, no. 1, pp. 181–183, Jul. 1990.

[2] T. Flohr, K. Stierstorfer, R. Raupach, S. Ulzheimer, and H. Bruder, "Performance evaluation of a 64-slice CT system with z-flying focal spot," *RöFo: Fortschritte auf dem Gebiete der Röntgenstrahlen und der Nuklearmedizin*, vol. 176, pp. 1803–10, 2005.

[3] M. D. Silver, K. Taguchi, I. A. Hein, B. Chiang, M. Kazama, and I. Mori, "Windmill artifact in multislice helical CT," *SPIE Medical Imaging Proc.*, vol. 5032, pp. 1918 – 1927, 2003.

[4] M. Kachelrieß, M. Knaup, C. Penßel, and W. A. Kalender, "Flying focal spot (FFS) in cone–beam CT," *IEEE Transactions on Nuclear Science*, vol. 53, no. 3, pp. 1238–1247, Jun. 2006.

[5] K. M. Brown and S. Žabic, "Method for reducing windmill artifacts in multislice CT images," *SPIE Medical Imaging Proc.*, vol. 7961, pp. 491 – 495, 2011.

[6] J. Magonov, M. Kachelrieß, E. Fournié, K. Stierstorfer, T. Buzug, and M. Stille, "Row interpolation in spiral CT with deep learning," *16th Virtual International Meeting on Fully 3D Image Reconstruction in Radiology and Nuclear Medicine*, pp. 376–380, Oct. 2021.

[7] C. Ledig, L. Theis, F. Huszár, J. Caballero, A. Cunningham, A. Acosta, A. Aitken, A. Tejani, J. Totz, Z. Wang, and W. Shi, "Photo-realistic single image super-resolution using a generative adversarial network," *IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, pp. 105–114, Jul. 2017.

[8] S. Kappler, D. Niederlohner, S. Wirth, and K. Stierstorfer, "A full-system simulation chain for computed tomography scanners," *IEEE Nuclear Science Symp. Conf. Record*, pp. 3433 – 3436, Dec. 2009.

[9] W. Shi, J. Caballero, F. Huszár, J. Totz, A. P. Aitken, R. Bishop, D. Rueckert, and Z. Wang, "Real-time single image and video super-resolution using an efficient sub-pixel convolutional neural network," *IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, Jun. 2016.

[10] H. Zhao, O. Gallo, I. Frosio, and J. Kautz, "Loss functions for image restoration with neural networks," *IEEE Transactions on Computational Imaging*, vol. 3, no. 1, pp. 47–57, Dec. 2016.

DL-Recon: Combining 3D Deep Learning Image Synthesis and Model Uncertainty with Physics-Based Image Reconstruction

Xiaoxuan Zhang, Pengwei Wu, Wojciech B. Zbijewski, Alejandro Sisniega, Runze Han, Craig K. Jones, Prasad Vagdargi, Ali Uneri, Patrick A. Helm, William S. Anderson, Jeffrey H. Siewerdsen

***Abstract*— High-precision image-guided neurosurgery – especially in the presence of brain shift – would benefit from intraoperative image quality beyond the conventional contrast-resolution limits of cone-beam CT (CBCT) for visualization of the brain parenchyma, ventricles, and intracranial hemorrhage. Deep neural networks for 3D image reconstruction offer a promising basis for noise and artifact reduction, but generalizability can be challenged in scenarios involving features previously unseen in training data.**

We propose a 3D deep learning reconstruction framework (termed "DL-Recon") that integrates learning-based image synthesis with physics-based reconstruction to leverage strengths of each. A 3D conditional GAN was developed to generate synthesized CT from CBCT images. Uncertainty in the synthesis image was estimated in a spatially varying, voxel-wise manner via Monte-Carlo dropout and was shown to correlate with abnormalities or pathology not present in training data. The DL-Recon approach improves the fidelity of the resulting image by combining the synthesized image ("DL-Synthesis") with physics-based reconstruction (filtered back-projection (FBP) or other approaches) in a manner weighted by uncertainty – i.e., drawing more from the physics-based method in regions where model uncertainty is high.

The performance of image synthesis, uncertainty estimation, and DL-Recon was investigated for the first time in real CBCT images of the brain. Variable input to the synthesis network was tested – including uncorrected FBP and precorrection with a simple (constant) scatter estimate – hypothesizing the latter to improve synthesis performance. The resulting uncertainty estimation was evaluated for the first time in real anatomical features not included in training (abnormalities and brain shift). The performance of DL-Recon was evaluated in terms of image uniformity, noise, and soft-tissue contrast-to-noise ratio in comparison to DL-Synthesis and FBP with a comprehensive artifact correction framework. DL-Recon was found to leverage the strengths of the learning-based and physics-based reconstruction approaches, providing a high degree of image uniformity similar to DL-Synthesis while accurately preserving soft-tissue contrast as in artifact-corrected FBP.

***Index Terms*—Cone-beam CT, deep learning, artifact correction, image-guided intervention, image synthesis**

I. Introduction

Neurosurgical approaches to cancer, trauma, or neuro-degenerative disease require a high degree of geometric precision to safely avoid vessels and eloquent brain and achieve effective treatment. The state of the art in intraoperative cone-beam CT (CBCT) is sufficient for visualization and registration of high-contrast objects (e.g., bone, surgical instruments), but it does not provide contrast resolution suitable to soft-tissue, brain parenchyma, or intracranial hemorrhage. Factors limiting CBCT image quality include image biases (e.g., scatter, beam hardening) and quantum and electronic noise.

This work was supported by NIH U01-NS-107133.

X Zhang, P Wu, W Zbijewski, A Sisniega, R Han, CK Jones, P Vagdargi, A Uneri, WS Anderson, JH Siewerdsen are with Johns Hopkins University, MD, USA. PA Helm is with Medtronic Inc, MA, USA.

Existing methods for improving CBCT image quality include artifact corrections [1] and model-based iterative reconstruction (MBIR) [2] that leverages physical knowledge of the imaging chain and image formation process. Recent developments in deep learning approaches provide another means of mitigating artifacts and reducing noise, including image synthesis from CBCT to approximate diagnostic-quality CT [3]. Such approaches offer improvements in computational runtime compared to MBIR, but the performance of image synthesis is subject to uncertainties arising from features not present in training (e.g., pathology, anatomical variations, and unmodeled imaging conditions). The fidelity of the synthesized image hence cannot be guaranteed [4].

Recognizing the potential pitfalls in generalizability of image synthesis to highly variable anatomical structures in image-guided surgery, we propose a deep learning reconstruction framework (referred to as "DL-Recon") that integrates image synthesis with physics-based reconstruction mediated by model uncertainty. Previous work [5] proposed a 2D U-Net for image synthesis and combined the result with FBP and MBIR reconstruction via model uncertainty in simulation studies. In this work, we developed a 3D generative adversarial network (GAN) for image synthesis and evaluated the performance of DL-Recon for the first time in real CBCT images, including anatomical abnormalities unseen in training data.

II. Methods

A. Image synthesis and uncertainty estimation

A 3D conditional GAN was developed for CBCT-to-CT image synthesis. For training (Section II.C), a high-fidelity, physics-based forward projection framework (including an accurate beam model, absorption / scatter characteristics, and model of the imaging chain) was used to generate simulated CBCT images from corresponding CT images. Two alternative inputs to the synthesis network were investigated: (i) an uncorrected FBP (μ_{uncorr}^{CBCT}), and (ii) a precorrected FBP ($\mu_{precorr}^{CBCT}$) for which a simple (constant) scatter correction was applied, hypothesizing that the precorrection to improve synthesis performance.

As illustrated in Fig. 1, a 3D GAN was implemented with a U-Net with a residual block at each level of the encoding / decoding path as the generator, and a convolutional pixel-wise classifier [6] as the discriminator. The objective function combined GAN and L1 loss as follows:

$$\hat{G} = \arg\min_{G}\max_{D} \mathcal{L}_{cGAN}(G,D) + \lambda\mathcal{L}_{L1}(G) \quad (1)$$

where

$$\mathcal{L}_{cGAN}(G,D) = \mathbb{E}\left[\log D\left(\mu^{CBCT},\mu^{CT}\right)\right] + \mathbb{E}\left[\log\left(1 - D\left(\mu^{CBCT},G\left(\mu^{CBCT}\right)\right)\right)\right] \quad (2)$$

$$\mathcal{L}_{L1}(G) = \mathbb{E}\left[\left\|\mu^{CT} - G\left(\mu^{CBCT}\right)\right\|_1\right] \quad (3)$$

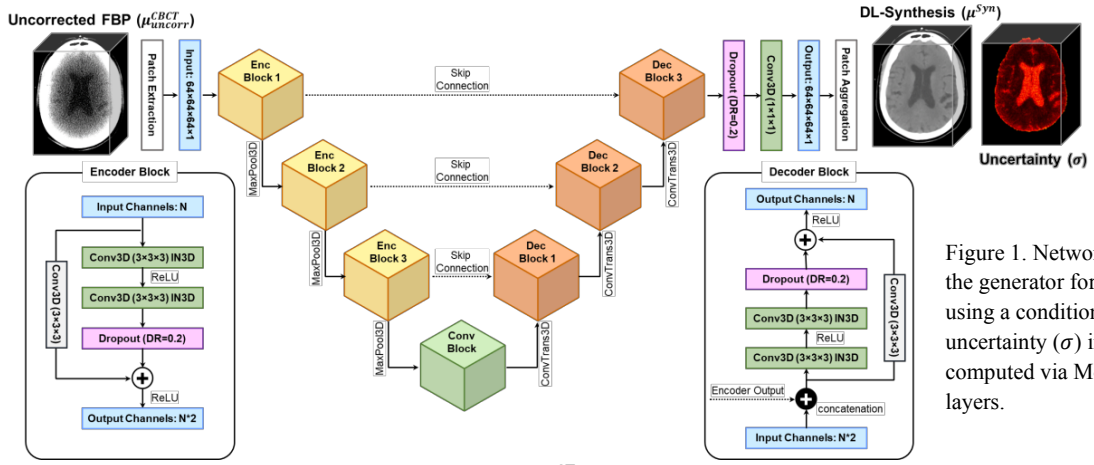


Figure 1. Network architecture of the generator for 3D image synthesis using a conditional GAN. Model uncertainty (σ) in the synthesis is computed via Monte-Carlo dropout layers.

G and D denote the generator and discriminator, and μ^{CT} and μ^{CBCT} represent paired CT and CBCT images. The L1 loss helps avoid over smoothing, and the balance between the GAN and L1 loss is controlled by λ.

As described in [7], dropout applied during network training is equivalent to a Bayesian approximation of the Gaussian process, and uncertainty in the model output can be estimated by computing the voxel-wise variance of multiple forward passes. Following such an approach, we added dropout layers (dropout rate = 0.2) prior to the skip connection in each encoder and decoder block and to the final output. Both training and inference were performed with dropout. The predictive mean computed from a collection of 8 network outputs yields the synthesized image (DL-Synthesis, μ^{Syn}), and the predictive variance (σ^2) serves as a proxy for model uncertainty.

B. *The DL-Recon framework*

The proposed method (termed DL-Recon) integrates 3D image synthesis with physics-based reconstruction via uncertainty associated with the synthesis model. The method involves three steps: (i) generation of a 3D synthetic CT image (μ^{Syn}) from a CBCT volume with estimation of model uncertainty (σ) as described above; (ii) physics-based 3D image reconstruction of projection data, including artifact corrections – for example, the pipeline described in [1] – to yield an artifact-corrected CBCT image (denoted μ_{corr}^{CBCT}); and (iii) voxel-wise combination of μ^{Syn} and μ_{corr}^{CBCT} weighted by the estimated uncertainty to yield the DL-Recon image (denoted $\mu^{DL-Recon}$). The resulting image is:

$$\mu^{DL-Recon} = [1 - \beta(\sigma)]\mu^{Syn} + \beta(\sigma)\mu_{corr}^{CBCT} \quad (4)$$

where uncertainty is contained within a spatially varying map (β, with values in the range [0, 1]) related by a sigmoid function:

$$\beta(\sigma) = \frac{1}{1 + e^{-(c_1 \cdot \sigma + c_2)}} \quad (5)$$

where c_1 and c_2 specify the range and level, respectively, of the sigmoid, and β controls the contribution of μ^{Syn} and μ_{corr}^{CBCT} in a voxel-wise manner. When predictive uncertainty is high, the β map draws more from the physics-based reconstruction.

The underlying premise in this approach is that the synthesis image (μ^{Syn}) carries particular benefits (e.g., uniformity and noise reduction) but may be subject to systematic error – for example, in structures unseen in the training data. The uncertainty map [$\sigma(x,y,z)$, alternatively $\beta(x,y,z)$] were shown previously in simulation studies [5] to correlate with deviations from ground truth. The "uncertainty map" therefore offers insight on where the synthesis image may be subject to error and where it is advantageous to draw more from the physics-based 3D image reconstruction (μ_{corr}^{CBCT}).

Note that the physics-based method incorporated in DL-Recon could be FBP or any particular form of MBIR, recognizing that the latter may invite disadvantages of computational load associated with conventional iterative optimization. Alternatively, the synthesis image could be incorporated as a prior within a penalized optimization, as in [5]. In any of these scenarios, the voxel-wise weighting of synthesis and physics-based image reconstructions is intended to leverage the strengths of each, mediated by the model uncertainty. In the work reported below, DL-Recon incorporates (artifact-corrected) FBP reconstruction as a practical implementation that may be compatible with the rapid runtime requirements of image-guided surgery, focusing here on intracranial neurosurgery.

C. *Training data generation*

To obtain a large training dataset of matched CT and CBCT images, CBCT projection data were simulated from 35 real, helical CT volumes of 35 healthy subjects using a high-fidelity forward projector [5]. CBCT system geometry and image acquisition were simulated to match data (~745 views over 360°) acquired from the O-arm (the O-arm™ "O2" imaging system, Medtronic) using nominal head scan protocols (100–120 kV and 75–240 mAs). Volumes were reconstructed with isotropic 0.7 mm voxels via FBP without artifact correction. Signal normalization linearly transformed the CBCT intensity histogram within the brain parenchyma to [-1, 1]. Volumetric patches (64×64×64 voxels) were stochastically sampled from the brain volume and fed to the network, and a total of 875 patches were used for training. The Adam optimizer (learning rate = 5×10^{-5}, $\beta_1 = 0.5$, $\beta_2 = 0.999$, L1 regularization λ=100, and batch size = 2) was used and early stopping at 800 epochs was applied.

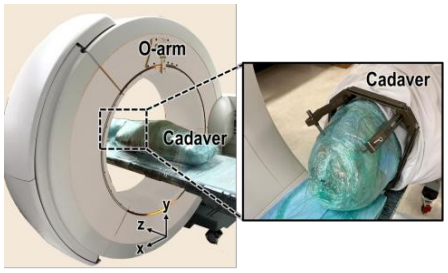


Figure 2. Experimental setup for cadaver studies using the O-arm.

D. Experimental studies

D1. Image synthesis of simulated and real brain CBCT images

The proposed image synthesis method was validated on both simulated and real CBCT data. Simulated CBCT projections of 5 test CT volumes were generated and reconstructed in the same manner as the training set. Intensity differences between synthesized images and ground truth were measured within the brain region for each volume. Experiments were conducted using the O-arm™ system illustrated in Fig. 2. Real projection data for 3 cadaveric heads (denoted below as cadaver #1-3) were collected at 120 kV and 150 mAs. Volumetric images were reconstructed on a grid of 320×320×280 voxels with isotropic 0.7 mm voxels. The runtime of DL-Synthesis was ~1 min per prediction (NVIDIA TITAN Xp). DL-Synthesis images were evaluated with uncorrected CBCT as input (denoted μ^{Syn}_{uncorr}) and with a basic (constant-scatter) precorrection (denoted $\mu^{Syn}_{precorr}$). Method performance was quantified in terms of image non-uniformity (NU), the difference in mean voxel value between region of interests (ROIs) in the parenchyma near the dural surface / sphenoid bone and about the lateral ventricles.

D2. Uncertainty estimation in real anatomical abnormalities

Previous work [5] has shown correlation between synthesis error and uncertainty for simulated lesions (not exist in the training cohort) of difference location, size, and contrast. In this work, the accuracy of uncertainty estimation was evaluated in cadaver images, including specimens exhibiting true abnormalities that were not present in the training data. Specifically, abnormalities included a large intraparenchymal calcification, a loss of cerebrospinal fluid, and brain shift in which the brain cortex collapsed from the interior surface of the cranium.

D3. Cadaver studies on an intraoperative CBCT system

Imaging performance was evaluated in terms of visual image quality as well as image uniformity, noise, and soft-tissue contrast-to-noise ratio (CNR) in cadavers imaged on the O-arm™ system (Fig. 2). FBP reconstructions were evaluated with and without artifact correction. DL-Recon was evaluated in comparison to FBP and DL-Synthesis, and uncertainty maps were displayed to understand how physics-based and deep learning-based approaches contributed to the final result.

III. Results

A. Performance of image synthesis

Fig. 3 shows results of image synthesis on simulated data (high-fidelity CBCT projections generated from CT). DL-Synthesis demonstrated good overall correspondence with the ground truth CT, yielding high image uniformity and reduced noise compared to the uncorrected FBP image. In 5 test volume images, DL-Synthesis exhibited a difference in overall mean intensity (in the brain) of less than 1 HU (compared to > 12 HU for FBP) to the ground truth, with residual differences owing mainly to image noise. The estimated uncertainty highlights regions with anatomical variations such as the lateral ventricles and sulci in the cerebral cortex, which is susceptible to error (e.g., contrast loss) in the synthesis mage.

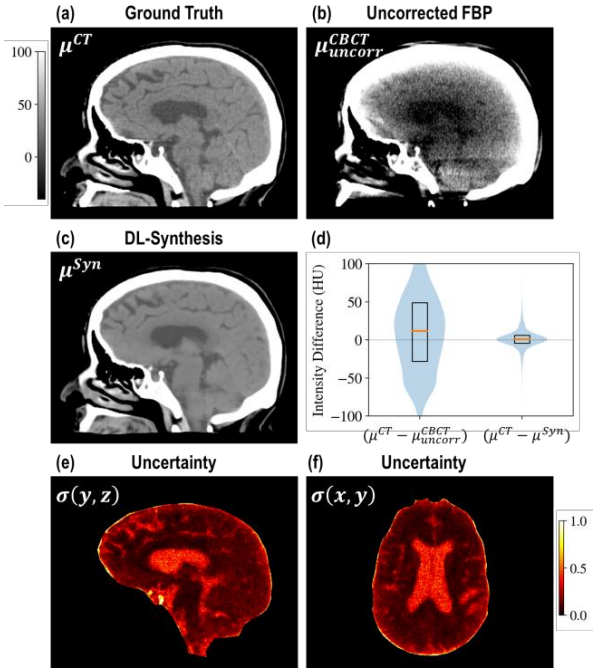


Figure 3. Synthesis performance in simulated CBCT data. (a) Sagittal slice of a test CT image volume. (b) Corresponding CBCT reconstruction (network input). (c) Resulting synthesized image. (d) Violin plot quantifying the respective difference in voxel values of uncorrected FBP and DL-Synthesis to the ground truth measured for 5 test data. (e) Sagittal and (f) axial slice of the estimated uncertainty.

Fig. 4 illustrates the performance of image synthesis on real data, in which the input to the synthesis network was either uncorrected or precorrected image data. DL-Synthesis acting on uncorrected FBP input exhibits performance degradation in regions affected by severe artifacts, yielding a higher degree of non-uniformity near the sphenoid bone (yellow arrow). A simple (constant) scatter correction was shown to partially account for biases that were not modeled by the forward projector (e.g., variation in bone density) and improve the overall image uniformity (2–4 HU). As a result, precorrected FBP yielded more

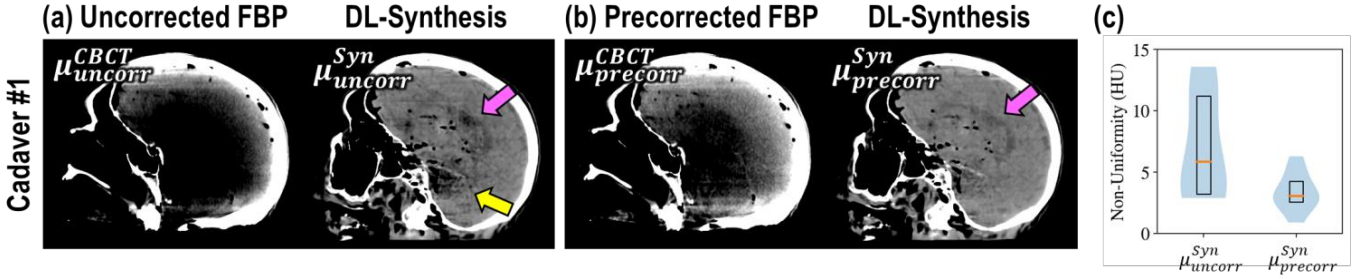


Figure 4. Synthesis performance for (a) uncorrected and (b) precorrected FBP of real CBCT images. (c) Boxplot quantifying image non-uniformity in synthesized images of 3 cadavers.

accurate synthesis, reducing image NU by ~50% compared to synthesis acting on uncorrected FBP. However, DL-Synthesis exhibited a loss in contrast in structures such as the lateral ventricles (cadaver #1, magenta arrows), demonstrating potential pitfalls in the generalizability of image synthesis to real and highly variable image data.

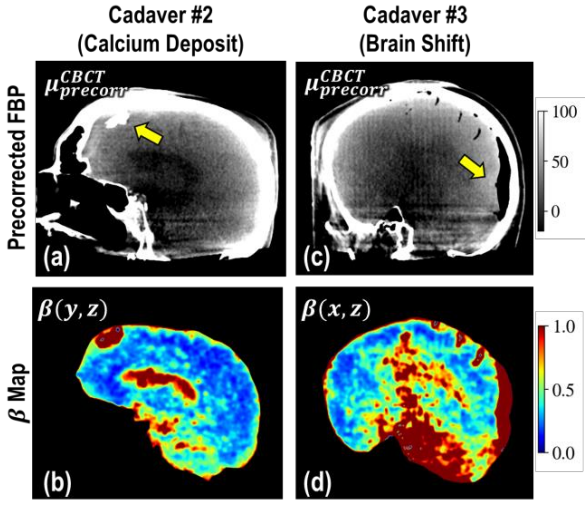


Figure 5. Uncertainty estimation in cadaver CBCT head images. Precorrected FBP and the estimated uncertainty (β map) within the brain parenchyma for (a-b) cadaver #1 with calcium deposit and (c-d) cadaver #3 with brain shift.

B. Uncertainty estimation in cadaver studies

Fig. 5 demonstrates the performance of uncertainty estimation on real data with unseen features (calcium deposit in cadaver #2 and brain shift in cadaver #3). For both cases, the uncertainty map highlights the location of the unseen structure as well as at the lateral ventricles, suggesting a lack of reliability in the synthesis result and the need for input from physics-based reconstruction.

C. Performance of DL-Recon

Fig. 6 shows reconstructed images from conventional methods (FBP and DL-Synthesis) and the proposed DL-Recon framework. As shown in Fig. 6(b), the comprehensive artifact correction pipeline reduced NU by 59%, but led to 38% increase in image noise. DL-Synthesis yielded the lowest NU value and noise but suffered from loss in soft-tissue contrast. In comparison, DL-Recon was able to reduce both NU and noise while preserving image contrast of the ventricles, providing ~15% increase in soft-tissue CNR compared to fully corrected FBP.

The intensity profile of a curve across the brain [yellow dashed curve shown in Fig. 6(b)] was plotted in Fig. 7 for fully corrected FBP, DL-Synthesis, and DL-Recon. Fully corrected FBP exhibited residual nonuniformity, especially just inside the cranium due to residual beam-hardening effects, as indicated by the nonuniform intensity profile between the ventricle and cranium. DL-Synthesis improved uniformity in these regions but reduced the contrast in the ventricle, similar to the effects shown above in relation to model uncertainty. By comparison, DL-Recon maintained the benefits of image uniformity from DL-Synthesis while achieving contrast in the ventricles similar to the fully corrected FBP.

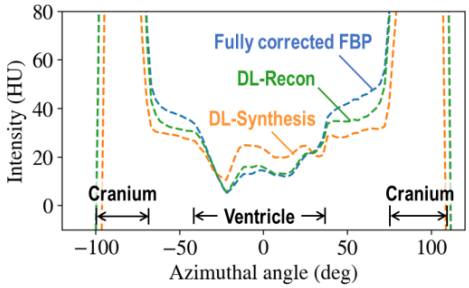


Figure 7. Intensity profiles of FBP, DL-Synthesis, and DL-Recon. The DL-Recon image leverages the improved uniformity of DL-Synthesis (region just inside the cranium) and the improved (accurate) contrast of fully corrected FBP (in the ventricles).

REFERENCES
[1] A Sisniega, et al. Physics in Medicine & Biology. 60(4), 2015.
[2] IA Elbakri and JA Fessler. IEEE TMI. 21(2), 2002.
[3] X Liang, et al. Physics in Medicine & Biology. 64(12), 2019.
[4] Q Yang, et al. IEEE TMI. 37(6), 2018.
[5] P Wu, et al. arXiv:2108.09229, 2021.
[6] P Isola, et al. IEEE CVPR pp. 1125-1134, 2017.
[7] Y Gal and Z Ghahramani., PMLR pp.1050-1059, 2016.

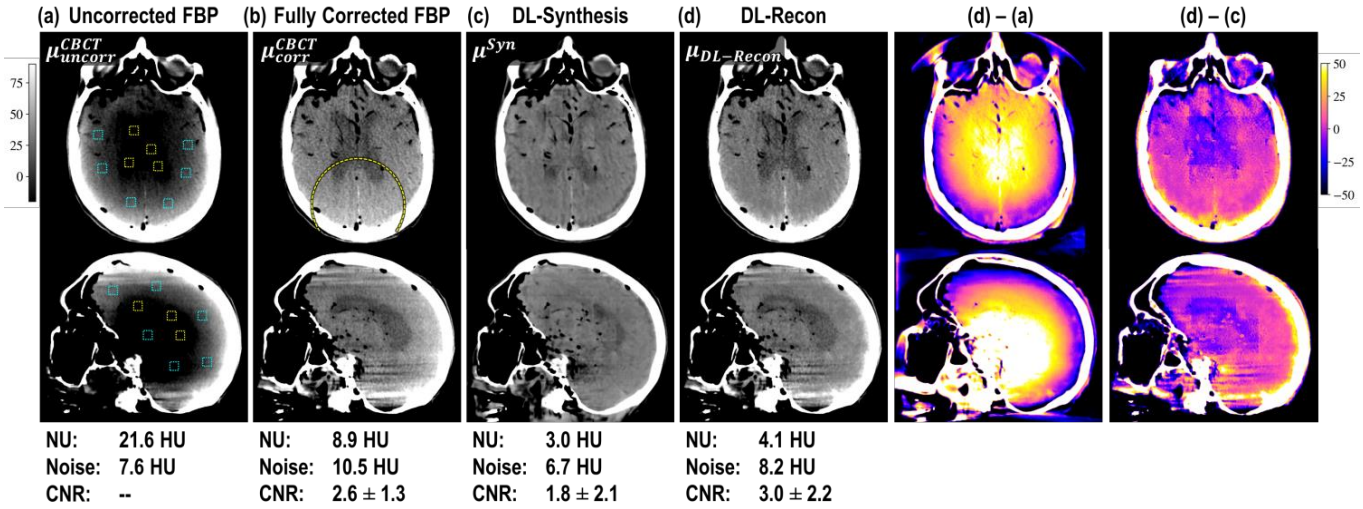


Figure 6. Example axial and sagittal slices of FBP, DL synthesis, and DL-Recon in cadaver CBCT data. Measurements of image non-uniformity (NU), noise, and contrast-to-noise ratio (CNR) of the lateral ventricles are listed below each image. Difference images show the contributions of the physics-based [(d)-(a)] and image synthesis [(d)-(c)] methods to the DL-Recon image (approximate Hounsfield Units).

Learned Cone-Beam CT Reconstruction Using Neural Ordinary Differential Equations

Mareike Thies, Fabian Wagner, Mingxuan Gu, Lukas Folle, Lina Felsner, and Andreas Maier, *Member, IEEE*

***Abstract*—Learned iterative reconstruction algorithms for inverse problems offer the flexibility to combine analytical knowledge about the problem with modules learned from data. This way, they achieve high reconstruction performance while ensuring consistency with the measured data. In computed tomography, extending such approaches from 2D fan-beam to 3D cone-beam data is challenging due to the prohibitively high GPU memory that would be needed to train such models. This paper proposes to use neural ordinary differential equations to solve the reconstruction problem in a residual formulation via numerical integration. For training, there is no need to backpropagate through several unrolled network blocks nor through the internals of the solver. Instead, the gradients are obtained very memory-efficiently in the neural ODE setting allowing for training on a single consumer graphics card. The method is able to reduce the root mean squared error by over 30% compared to the best performing classical iterative reconstruction algorithm and produces high quality cone-beam reconstructions even in a sparse view scenario.**

***Index Terms*—Inverse problems, computed tomography, iterative reconstruction, known operators.**

I. Introduction

EXTENDING analytical or iterative reconstruction algorithms for computed tomography (CT) by deep learning modules has shown to improve the quality of the reconstructed images, especially in challenging cases like high noise levels or insufficient projection data [1], [2]. For CT reconstruction, the input and output data of the problem are connected in a non-trivial geometrical manner. This is the reason why most deep-learning-based reconstruction approaches, instead of learning direct mapping from sinogram to image domain, incorporate knowledge about the physical operator connecting both domains and replace single components in the reconstruction pipeline by their learned counterpart, mostly operating in one domain only [3].

Unrolled iterative approaches seek to solve the reconstruction problem by loosely mimicking known iterative algorithms for inverse problems. This is done by unrolling a fixed number of iterations in depth as a deep learning architecture and incorporating learnable components in each step which are trained end-to-end. The exact architecture and the role of the trainable modules can vary giving rise to a number of approaches for MRI [4], [5] and CT [1], [6], [7].

While these unrolled iterative algorithms have achieved superior performance for the reconstruction of 2D images, their extension to the 3D case is challenging. Training requires gradient backpropagation through the entire unrolled sequence of trainable and known operators, thereby consuming a large amount of memory on the graphics card (GPU). In the 3D case, the amount of memory occupied by intermediate representations needed during backpropagation exceeds the memory of modern GPUs making the direct application of iterative approaches infeasible.

Previous work addressed the prohibitively high memory consumption of unrolled 3D models. Greedy training of each iteration independently reduces memory consumption and allows training on patches but does not result in an optimal joint weight configuration [8]. When increasing the volume resolution with network depth, memory consumption is dominated by the single final iteration on full scale, but image quality is coupled to the expressiveness of the last iteration [9]. Further, the use of invertible networks avoids storing intermediate representations which makes the memory requirement constant in depth but requires the network architecture to meet certain criteria for invertibility [10], [11].

In this work, we propose to interpret the series of unrolled iterations as a continuous residual process and formulate the problem in terms of an ordinary differential equation (ODE). This allows us to map the reconstruction problem onto an initial value problem which can be solved and trained memory-efficiently using recently proposed neural ODEs [12]. The key idea is that the memory requirement does not depend on the number of iterations, i.e., network depth. Instead, the forward pass is replaced by a call to an ODE solver and gradients are obtained by solving another adjoint ODE without storing the intermediate representations of the forward pass. Whereas a similar idea has been applied to MRI reconstruction [13], to the best of our knowledge we are the first ones to apply neural ODEs to CT reconstruction. We show that using this method we obtain 3D cone-beam CT reconstructions from few angles with superior image quality compared to classical analytic and algebraic algorithms.

II. Methods

A. Learned Iterative CT Reconstruction

The forward model of a CT acquisition can be written as

$$\mathbf{p} = \mathbf{A}\mathbf{x} + \epsilon \ , \tag{1}$$

where $\mathbf{x} \in \mathbb{R}^M$ is the volume, $\mathbf{p} \in \mathbb{R}^N$ is the projection data, $\mathbf{A} \in \mathbb{R}^{N \times M}$ is the forward operator defined by the imaging geometry (cone-beam in this case), and $\epsilon \in \mathbb{R}^N$ is additive noise. Typically, recovering the volume $\mathbf{x}$ from the measured data $\mathbf{p}$ is an ill-posed inverse problem meaning that $\mathbf{A}$ is

All authors are with the Pattern Recognition Lab, Friedrich-Alexander University Erlangen-Nuremberg, Erlangen, Germany (e-mail: mareike.thies@fau.de).

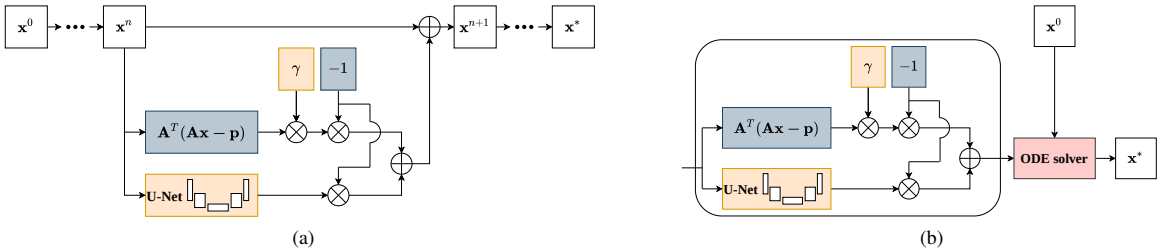


Fig. 1: Comparison of an unrolled network (a) and a neural ODE-based version (b). While the unrolled version explicitly repeats the same network block for a fixed number of steps, the ODE solver receives only one network block parameterizing the temporal derivative of the volume. The solver computes the numerical integration internally. All trainable parts of our architecture are highlighted in yellow, the fixed parts are gray.

not square and there are multiple solutions for $\mathbf{x}$ which are consistent with the measured data $\mathbf{p}$. Hence, the reconstruction is formulated as a regularized optimization problem

$$\mathbf{x}^* = \underset{\mathbf{x}}{\operatorname{argmin}}\{D(\mathbf{x},\mathbf{p}) + \mu R(\mathbf{x})\} \ . \tag{2}$$

Here, $D : \mathbb{R}^M \times \mathbb{R}^N \to \mathbb{R}$ is a function which measures the consistency of volume $\mathbf{x}$ and measured data $\mathbf{p}$. We choose the data consistency term in a least squares sense given as $D(\mathbf{x},\mathbf{p}) = \frac{1}{2}\|\mathbf{Ax}-\mathbf{p}\|_2^2$. $R : \mathbb{R}^M \to \mathbb{R}$ is a regularizer which helps finding a favorable solution $\mathbf{x}^*$ and is weighted against the data consistency term by a scalar $\mu \in \mathbb{R}$. Using a simple gradient descent optimization scheme, the resulting update formula is

$$\begin{aligned}\mathbf{x}^{n+1} &= \mathbf{x}^n - \lambda\nabla\{D(\mathbf{x},\mathbf{p}) + \mu R(\mathbf{x})\} \\ &= \mathbf{x}^n - \lambda(\mathbf{A}^T(\mathbf{Ax}-\mathbf{p}) + \mu\nabla R(\mathbf{x})) \ ,\end{aligned} \tag{3}$$

where $\mathbf{A}^T \in \mathbb{R}^{M\times N}$ is the adjoint operator of $\mathbf{A}$, $\lambda \in \mathbb{R}$ is a sufficiently small step size, and n is the iteration index. To learn a flexible regularizer from data, we replace its gradient by a network $N_\theta : \mathbb{R}^M \to \mathbb{R}^M$ with free parameters θ which allows to fit the regularizing component directly from data. The final update formula is given as

$$\mathbf{x}^{n+1} = \mathbf{x}^n - \lambda(\mathbf{A}^T(\mathbf{Ax}-\mathbf{p}) + \mu N_\theta(\mathbf{x})) \ . \tag{4}$$

If $\mathbf{x}$ represents a 2D image, this equation could inspire an unrolled network architecture (Fig. 1a) and the parameters θ can be trained from pairs of input projection data and ground truth reconstruction for a fixed number of unrolled iterations [1]. This is infeasible for 3D cone-beam data due to extremely high GPU memory requirements.

B. Neural Ordinary Differential Equations

Equation 4 has a residual form of the type

$$\mathbf{x}^{n+1} = \mathbf{x}^n + f_\theta(\mathbf{x},\mathbf{p}) \ , \tag{5}$$

with $f_\theta(\mathbf{x},\mathbf{p}) = -\lambda(\mathbf{A}^T(\mathbf{Ax}-\mathbf{p}) + \mu N_\theta(\mathbf{x}))$. The function f_θ describes how the volume $\mathbf{x}^n$ changes incrementally. Chen et al. [12] proposed to regard such residual neural network architectures as the numerical integration of some underlying continuous ordinary differential equation. Here, the continuous differential equation would be $\frac{d\mathbf{x}}{dt} = f_\theta(\mathbf{x}(\mathbf{t}),\mathbf{p})$. Following that idea, a full unrolled iterative reconstruction is similar to the solution of an initial value problem of the given ODE starting from an initial condition $\mathbf{x}^0$ integrated until some end time T

$$\mathbf{x}^* = \mathbf{x}^0 + \int_0^T f_\theta(\mathbf{x}(\mathbf{t}),\mathbf{p})\mathrm{dt} \ . \tag{6}$$

There exist a number of different numerical solvers to approximate solutions of such initial value problems. In this work, we use a fixed step-size Runge-Kutta solver of order 4 which integrates Eq. 6 by dividing the interval $[0,..,T]$ into a fixed number of steps S to solve the integral numerically. This highlights the analogy to residual networks of depth S.

To be able to combine this ODE-based problem formulation with a trainable network architecture, we need to compute a loss that is based on the output of the ODE solver and use its gradient to update the weights θ contained in f_θ. As demonstrated in [12], this gradient can be obtained without backpropagating through the internals of the solver. Instead, the solver is regarded as a black box and the gradient with respect to θ is computed by solving another ODE backward in time (adjoint sensitivity method). This allows to obtain gradients with a memory cost that is independent of the number of steps S taken by the solver. Coming back to the analogy with residual networks, we can unroll the reconstruction problem in many steps using neural ODEs without further increasing the memory cost.

C. Network Architecture

In the neural ODE setting, a neural network defines the dynamics of the system, i.e., its temporal derivative. Following the classical problem formulation in Eq. 4, we design this network using two branches: (1) A data consistency branch with no trainable parameters incorporating the system's forward and backward model as known operator and (2) a regularization branch which is trained from data. Figure 1b illustrates the proposed network architecture. The data consistency branch implements the term $\mathbf{A}^T(\mathbf{Ax}-\mathbf{p})$. Operators $\mathbf{A}$ and $\mathbf{A}^T$ are the CT forward and backprojection under the correct cone-beam geometry, respectively. We use the differentiable version of these operators described in [14] which computes analytical gradients and allows for a direct embedding of these operators in neural networks. For the regularization branch, we use a

standard 3D U-Net [15] with depth 4 and 8 feature maps on the first level which are doubled in each stage, ReLU activation function, and instance normalization. In total, this leads to a network with 255 000 free parameters. We further introduce an additional single trainable parameter γ (referred to as data consistency weight) which is multiplied to the output of the data consistency branch before adding the output of both branches together in order to enable a data-optimal weighting between the data consistency and regularization component. This network together with the initialization $\mathbf{x}^0$ is passed to the ODE solver.

III. Experiments

A. Data

The data set consists of 42 walnuts scanned under cone-beam CT geometry (cone angle: 40°) [16]. It contains the raw projection data and a corresponding ground truth reconstruction. The projection images are acquired on three circular trajectories on different heights along the walnuts' long axes with a full rotation of 360° divided into 1200 angular steps each. The ground truth reconstruction is computed iteratively from the full set of acquired projections. As input to our algorithm, we use projection data from only the central one of the three trajectories and downsample the projections by a factor of 10 in angular direction and 2 in the spatial directions. This results in 120 projection images of size 384×486 pixels with an angular increment of 3° covering a full rotation of 360°. Hence, the algorithm has much less projection data available than has been used for computing the ground truth reconstruction which serves as learning target and is downsampled by a factor of 2 resulting in volumes of size 251^3. We use walnut number 1 for validation and walnut number 2 for testing. The rest is used for training. The used data, its preprocessing and the train-test-splits are identical to [10].

B. Training Details

We use the neural ODE solver provided by [12]. Integration of Eq. 6 is performed from 0 to $T = 1$ with a fixed step size of 0.05. This results in 20 steps and 80 evaluations of the network per forward and backward pass as the fourth order Runge-Kutta solver takes four evaluations per step. The initial volume $\mathbf{x}^0$ is a Feldmann-Davis-Kress (FDK) reconstruction of the projection data. The last layer of the U-Net is initialized with zeros such that the regularizer has no influence upon initialization and the data consistency weight is initialized with $\gamma = 0.01$. The parameters are optimized by an Adam optimizer with learning rate 1×10^{-4} for the U-Net and 1×10^{-2} for the data consistency weight. Training is performed with batch size 1 for 80 epochs and an L1-Loss evaluated only inside the cylindrical scan field of view (FOV) captured in each projection. The model weights corresponding to the lowest validation loss during training are selected for further evaluation.

C. Reference Methods

The reconstruction of the proposed method is compared to (1) an FDK reconstruction, (2) a SIRT reconstruction with non-negativity constraint (500 iterations) [17] and (3) a total variation (TV) regularized reconstruction using gradient updates (300 iterations) [14].

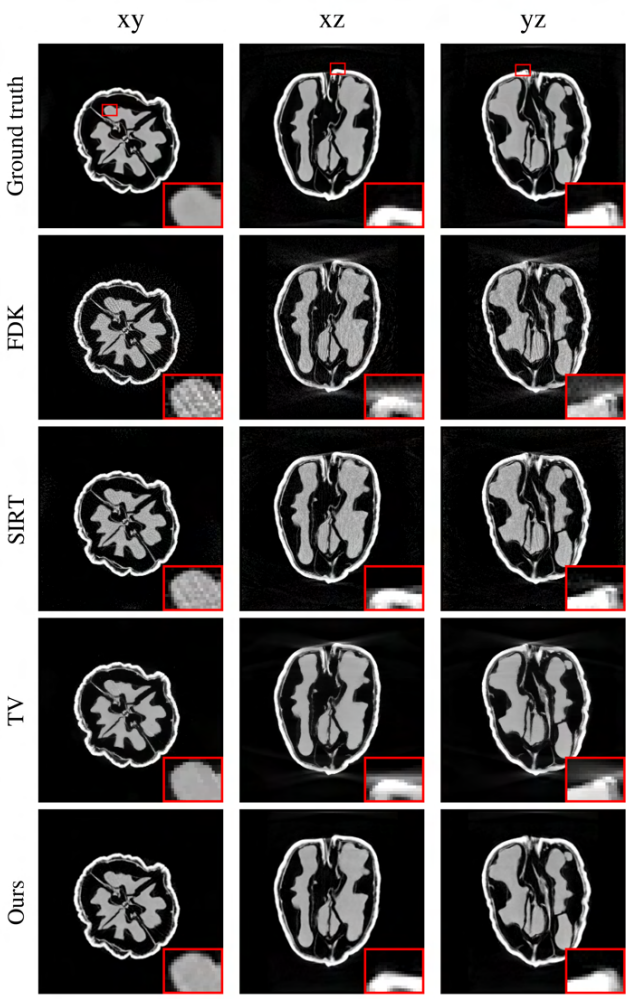


Fig. 2: Reconstructed center slices of the test walnut along each dimension using an analytical FDK algorithm, two classical iterative algorithms (SIRT, TV) and the proposed method. The gray value window is 0 to $0.06\,\mathrm{mm}^{-1}$. Zoomed regions are indicated in red.

IV. Results

The proposed model requires a maximum of 13 GB GPU memory during training. Reconstruction results of the test walnut are shown in Fig. 2. The FDK reconstruction exhibits strong streaking artifacts in the xy-plane as well as cone-beam artifacts at the top and bottom of the walnut visible in the xz- and yz-planes. The iterative SIRT algorithm only partly removes these artifacts. The TV regularized algorithm produces a very smooth image with homogeneous gray values for different structures inside the walnut. Nevertheless, cone-beam artifacts in the xz- and yz-plane are still visible. In contrast, our method achieves images with no noticeable cone-beam artifacts. Additionally, it performs well in removing

TABLE I: Quantitative results in terms of root mean squared error (RMSE), peak signal-to-noise ratio (PSNR), and structural similarity index measure (SSIM).

	FDK	SIRT	TV	Ours
RMSE $[\cdot 10^{-3}] \downarrow$	3.735	2.321	2.688	1.586
PSNR $\uparrow$	25.682	29.813	28.539	33.121
SSIM $\uparrow$	0.562	0.813	0.777	0.904

the streaks in the xy-plane and produces images which are visually closest to the ground truth. The proposed method also performs best regarding all quantitative metrics (Tab. I). Compared to the SIRT which is the second-best performing method, the RMSE is reduced by $31.7\,\%$ and PSNR and SSIM are increased by $11.1\,\%$ and $11.3\,\%$, respectively. All metrics have only been evaluated inside the cylindrical scan FOV captured in each projection. Concerning reconstruction time, our method lies between the two investigated iterative methods with a run time of $43.4\,\mathrm{s}$. The data consistency weight converges to a value of $\gamma = 0.036$.

V. Discussion

Our proposed method is able to train a network inspired from classical iterative reconstruction for 3D cone-beam data incorporating a trainable regularizer with a GPU memory consumption which is independent of the number of incremental update steps on the volume. We can reconstruct volumes of practically relevant size (251^3) while using only $13\,\mathrm{GB}$ of GPU memory during training. This is feasible with a single recent consumer graphics card.

The considered reconstruction problem is severely ill-posed due to the strong undersampling of projection data in angular direction. Hence, the analytical FDK reconstruction leads to strong artifacts in the reconstructed images. The trained algorithm removes the noise and streak artifacts successfully. While the classical TV-regularized iterative reconstruction also performs well in this regard, the proposed algorithm is the only one which is able to remove the cone-beam artifacts. We hypothesize that one main advantage of the learned regularizer parameterized by the U-Net over hand-crafted ones such as TV is its larger receptive field. It can suppress artifacts with non-local extent such as the cone-beam artifacts while TV depends only on the local gradient information in the image. A detailed comparison of our method to other learning-based approaches, such as [10], will be performed in future work.

Once trained, the reconstruction time of the presented method is comparable to that of classical iterative algorithms. Training time is rather high due to the high number of network evaluations. Potentially, using an adaptive ODE solver instead of the fixed step size solver can shorten training and inference times by adaptively adjusting the step size and hence the number of network evaluations to a given tolerance.

VI. Conclusion

This paper presents a method which uses neural ODEs to train a cone-beam reconstruction algorithm inspired by iterative reconstruction schemes. It ensures consistency with the measured data by incorporating a data consistency branch which exploits analytical knowledge about the physical operator connecting sinogram and image domain along with a trained regularizer. The proposed method outperforms well-known FDK and iterative reconstruction algorithms on the used walnut data set and is able to remove artifacts with non-local extent such as the cone-beam artifacts while being tractable concerning GPU memory.

Acknowledgment

The research leading to these results has received funding from the European Research Council (ERC) under the European Union's Horizon 2020 research and innovation program (ERC Grant No. 810316).

References

[1] H. Chen, Y. Zhang, Y. Chen, J. Zhang, W. Zhang, H. Sun, Y. Lv, P. Liao, J. Zhou, and G. Wang, "LEARN: Learned experts' assessment-based reconstruction network for sparse-data CT," *IEEE Trans. Med. Imag.*, vol. 37, no. 6, pp. 1333–1347, 2018.

[2] K. H. Jin, M. T. McCann, E. Froustey, and M. Unser, "Deep Convolutional Neural Network for Inverse Problems in Imaging," *IEEE Trans. Image Process.*, vol. 26, no. 9, pp. 4509–4522, 2017.

[3] J. Adler, "Learned Iterative Reconstruction," *Handbook of Mathematical Models and Algorithms in Computer Vision and Imaging: Mathematical Imaging and Vision*, pp. 1–22, 2021.

[4] K. Hammernik, T. Klatzer, E. Kobler, M. P. Recht, D. K. Sodickson, T. Pock, and F. Knoll, "Learning a variational network for reconstruction of accelerated MRI data," *Magn. Reson. Med.*, vol. 79, no. 6, pp. 3055–3071, 2018.

[5] F. Gadjimuradov, T. Benkert, M. D. Nickel, and A. Maier, "Robust partial Fourier reconstruction for diffusion-weighted imaging using a recurrent convolutional neural network," *Magn. Reson. Med.*, 2021.

[6] V. Vishnevskiy, R. Rau, and O. Goksel, "Deep Variational Networks with Exponential Weighting for Learning Computed Tomography," in *Proc. MICCAI*. Springer International Publishing, 2019, pp. 310–318.

[7] J. Adler and O. Öktem, "Learned Primal-Dual Reconstruction," *IEEE Trans. Med. Imag.*, vol. 37, no. 6, pp. 1322–1332, 2018.

[8] D. Wu, K. Kim, and Q. Li, "Computationally efficient deep neural network for computed tomography image reconstruction," *Med. Phys.*, vol. 46, no. 11, pp. 4763–4776, 2019.

[9] A. Hauptmann, J. Adler, S. Arridge, and O. Öktem, "Multi-scale learned iterative reconstruction," *IEEE Trans. Comput. Imag.*, vol. 6, pp. 843–856, 2020.

[10] J. Rudzusika, B. Bajic, O. Öktem, C.-B. Schönlieb, and C. Etmann, "Invertible Learned Primal-Dual," in *Proc. NeurIPS*, 2021.

[11] M. Kellman, K. Zhang, E. Markley, J. Tamir, E. Bostan, M. Lustig, and L. Waller, "Memory-efficient learning for large-scale computational imaging," *IEEE Trans. Comput. Imag.*, vol. 6, pp. 1403–1414, 2020.

[12] R. T. Chen, Y. Rubanova, J. Bettencourt, and D. Duvenaud, "Neural ordinary differential equations," in *Proc. NeurIPS*, 2018, pp. 6572–6583.

[13] E. Z. Chen, T. Chen, and S. Sun, "MRI Image Reconstruction via Learning Optimization Using Neural ODEs," in *Proc. MICCAI*. Springer, 2020, pp. 83–93.

[14] C. Syben, M. Michen, B. Stimpel, S. Seitz, S. Ploner, and A. K. Maier, "PYRO-NN: Python reconstruction operators in neural networks," *Med. Phys.*, vol. 46, no. 11, pp. 5110–5115, 2019.

[15] A. Wolny, L. Cerrone, A. Vijayan *et al.*, "Accurate and versatile 3D segmentation of plant tissues at cellular resolution," *eLife*, vol. 9, p. e57613, 2020.

[16] H. Der Sarkissian, F. Lucka, M. van Eijnatten, G. Colacicco, S. B. Coban, and K. J. Batenburg, "A cone-beam X-ray computed tomography data collection designed for machine learning," *Sci. Data*, vol. 6, no. 1, pp. 1–8, 2019.

[17] W. van Aarle, W. J. Palenstijn, J. Cant, E. Janssens, F. Bleichrodt, A. Dabravolski, J. D. Beenhouwer, K. J. Batenburg, and J. Sijbers, "Fast and flexible X-ray tomography using the ASTRA toolbox," *Opt. Express*, vol. 24, no. 22, pp. 25 129–25 147, Oct 2016.

$LaBr_3$:Ce and Silicon Photomultipliers: Towards the Optimal Scintillating Photon-Counting Detector

Stefan J. van der Sar, David Leibold, Stefan E. Brunner, and Dennis R. Schaart

***Abstract*: We investigate ultrafast silicon photomultiplier (SiPM)-based scintillation detectors for (medical) X-ray photon-counting applications, e.g., photon-counting computed tomography (CT). Such detectors may be an alternative to CdTe/CdZnTe (CZT) and Si detectors, which face challenges related to cost-effective growth of detector-grade material and detection efficiency, respectively. Here, we experimentally study energy response and count rate performance of a 1 mm × 1 mm single-pixel detector consisting of the commercially available $LaBr_3$:Ce scintillator and a fast SiPM prototype.**

We used three radio-isotopes and an X-ray tube for the experiments. Raw detector signals were processed by a second-order low-pass filter with a cut-off frequency f_c equal to 25 MHz or 100 MHz.

The detector pulse height was shown to be proportional to photon energy. We measured FWHM energy resolutions of 20% (f_c=25 MHz) and 22% (f_c=100 MHz) at 60 keV. The measured X-ray tube spectra showed signs of the expected features of such spectra. The best count rate performance was achieved using f_c=100 MHz. In case of paralyzable-like counting and a 30 keV counting threshold, the maximum observed count rate (OCR) was 10.5 Mcps/pixel. For nonparalyzable-like counting and the same threshold, the OCR appeared to approach an asymptotic value greater than 20 Mcps/pixel. These numbers are close to those of CdTe/CZT detectors highly optimized for photon-counting CT.

In conclusion, we show promising spectral X-ray photon-counting performance of an $LaBr_3$:Ce scintillation detector with SiPM readout. Depending on the application-specific requirements, miniaturization of the pixel size may be necessary, for which we discuss potential dose-efficient implementations.

***Keywords:* Count rate performance, energy response, scintillator, silicon photomultiplier (SiPM), X-ray photon-counting.**

I. Introduction

PHOTON-COUNTING detectors (PCDs) for X-ray imaging, e.g., for medical computed tomography (CT), are heavily investigated [1,2]. A PCD counts X-ray photon-induced detector pulses and registers them in one of at least two energy bins. This enables energy-resolved X-ray imaging. However, the task at hand is formidable, because the impinging photon fluence rate may exceed 10^8 $mm^{-2}s^{-1}$ [3], so pulse pile-up is likely to distort the measurement of counts and energies.

The main detector concept under consideration at the moment is based on the mechanism of direct conversion, i.e., each X-ray photon absorbed in a semiconductor is directly converted into a number of electron-hole (e-h) pairs. This number is proportional to the absorbed energy. Under the influence of an electric field, the e-h pairs are separated and guided towards (pixelated) electrodes on which they induce a current pulse. The pulse processing chain then outputs pulses of which the height is a measure of the energy of the X-ray photon. The semiconductors used in the existing prototype CT scanners are CdTe [4], CdZnTe (CZT) [5], and Si [6].

Although such a detector outputs short pulses and allows for pixel size miniaturization (both help to reduce pile-up), the cost-effective growth of CdTe and CZT detectors with a sufficiently low density of charge trapping centers (necessary to guarantee stable and reliable performance over time) remains an issue [2], [7]. On the other hand, Si detectors face challenges related to low mass density (ρ) and atomic number (Z). Thus, it is not clear what the best choice of detector is, leaving room for developing other detector concepts.

We are investigating fast scintillation detectors with silicon photomultiplier (SiPM) readout for X-ray photon-counting applications. This detector concept is based on the mechanism of indirect conversion, i.e., an X-ray photon is first converted into a pulse of optical scintillation photons, which is then converted into a current pulse by an SiPM (see Fig. 1a). The light pulse incident on the SiPM as a function of time t_1 after the X-ray interaction may be described as $A_1\exp(-t_1/\tau_d)$, with amplitude A_1 depending on deposited energy and τ_d the scintillation decay time constant. In order to minimize pile-up, shorter τ_d than those of CsI (1 µs) and GOS (2.5 µs), the scintillators used in integrating detectors, are needed [8].

Each scintillator pixel must be coupled to its own SiPM. This light sensor consists of a two-dimensional array of single-photon avalanche diodes (SPADs, see Fig. 1b). The absorption of a single optical photon in a SPAD creates an e-h pair that can trigger an avalanche multiplication process providing a gain in the order of 10^6. The time profile of the SPAD current pulse as a function of time t_2 after the detection of an optical photon may be described as $A_2\exp(-t_2/\tau_r)$, with amplitude A_2 and recharge time constant τ_r. Since the SPADs on an SiPM are connected in parallel, the time profile of an X-ray photon-induced pulse is essentially a convolution of $A_1\exp(-t_1/\tau_d)$ and $A_2\exp(-t_2/\tau_r)$ (see Fig. 2). The high internal gain assures that the signal due to a single X-ray photon exceeds the noise level of the electronics (difficult to accomplish using conventional photodiodes) and allows for a processing chain relying only on current-to-voltage conversion and pulse height discrimination.

Since the detector is based on transport of photons rather than charges and scintillators with high ρ and Z exist, this X-ray photon-counting detector concept may provide a solution

This manuscript was submitted on 23 January 2022. The SiPMs and the scintillation crystals were provided for free by Broadcom Inc and Saint Gobain Crystals, respectively.

S.J. van der Sar, D. Leibold, and D.R. Schaart are with the Department of Radiation Science and Technology, at Delft University of Technology, 2629JB Delft, the Netherlands (e-mails: s.j.vandersar@tudelft.nl, d.leibold@tudelft.nl, d.r.schaart@tudelft.nl).

S.E. Brunner is with Broadcom Inc, 93049, Regensburg, Bavaria, Germany (e-mail: stefan.brunner@broadcom.com).

for the aforementioned issues of direct conversion detectors.

Here, we experimentally investigate the count rate performance of an ultrafast SiPM-based scintillation detector irradiated by a 120 kVp X-ray beam. The scintillator used in this study is $LaBr_3$:Ce, a commercially available scintillator that we previously identified as having favorable properties for high-rate, energy-resolved X-ray photon-counting [8]. We compare the results with data from CdTe- and CZT-based PCDs and study the energy response of the detector.

II. MATERIALS AND METHODS

$LaBr_3$:Ce (5% Ce-doping, Saint Gobain Crystals) combines high X-ray detection efficiency (ρ=5.1 g cm^{-3}, Z_{La}=57, and Z_{Br}=35) with a short and intense scintillation pulse (τ_d=16 ns and a high light yield of 63 photons per keV). Using optically transparent glue, we coupled a 1 mm×1 mm $LaBr_3$:Ce crystal to a 1.0 mm×1.0 mm prototype SiPM (provided by Broadcom Inc., τ_r=7 ns). The $LaBr_3$:Ce crystal was 3.5 mm thick (equivalent to 2.5 mm CdTe in terms of detection efficiency in the diagnostic energy range). We also covered the crystal in reflective powder. Since $LaBr_3$:Ce is a hygroscopic scintillator, we applied an epoxy coating to our single-pixel detector, such that it can be used outside the moisture-free glovebox in which it was built. Note that the hygroscopic nature of a scintillator does not limit its applicability. The commonly used NaI:Tl scintillator is hygroscopic, and applied in detectors for medical SPECT scanners, for example.

The current signals from the SiPM were converted into voltage signals by a trans-impedance amplifier (gain=10) before being digitized by a TeledyneLeCroy HDO9404 oscilloscope (sampling rate = 1 GHz, bandwidth = 200 MHz). Due to the finite number of optical photons in the scintillation pulse, the raw detector pulses are not perfect convolutions of $A_1\exp(-t_1/\tau_d)$ and $A_2\exp(-t_2/\tau_r)$, but show random fluctuations. In order to prevent double counting of single pulses due to these fluctuations, all digitized signals were first smoothened by a second order low-pass filter with a cut-off frequency f_c=25 MHz or f_c=100 MHz. The former is expected to provide better energy resolution at low rates, the latter should yield faster pulses (see Fig. 2a) and better count rate performance.

In order to calibrate (mean) pulse height as a function of energy, we recorded pulses while exposing the detector to three low-activity radioactive sources with the following five photon emissions: 14 keV (Co-57), 32 keV (Ba-133), 60 keV (Am-241), 81 keV (Ba-133) and 122 keV (Co-57).

We performed X-ray tube experiments using an Yxlon Y.TU 320-D03 tube, which had a tungsten target and an anode angle of 20°. We selected a tube voltage of 120 kVp. The beam was filtered by 3.0 mm Be and 7.5 mm Al in total. The resulting spectrum ranges from 20 keV to 120 keV. We performed a tube current sweep starting at 0.5 mA and ending at 20 mA at a fixed source-detector distance. For each value of the tube current, we recorded ten pulse trains of 100 ms each.

We determined the number of counts for counting thresholds at 15 keV and 30 keV, because such low-energy thresholds usually limit the count rate capability of PCDs. Additionally, we determined the number of counts for two counting algorithms. The first one is paralyzable-like (p-like) counting. Here, every positive threshold crossing leads to a count and the maximum signal before the subsequent negative threshold crossing is considered a measure of the energy. The second one is nonparalyzable-like (np-like) counting. In this case, we determine, after a fixed time period τ_{np} following the registration of a count, whether or not the signal from the detector is still above threshold. If yes, a second count will be registered, and so on. If not, the next count will only be registered when the next positive threshold crossing occurs. The maximum signal within the time period τ_{np} is considered a measure of the energy. The value of τ_{np} should exceed the time-over-threshold of the highest-energy pulses in order to prevent double counting of pulses. We therefore used the pulses caused by 122 keV photons from the Co-57 source to determine proper values of τ_{np} for the detector (see Fig. 2b).

III. RESULTS AND DISCUSSION

Fig. 3a and Fig. 3b show pulse height spectra obtained by exposing the detector to the Am-241 source and setting f_c to 25 MHz and to 100 MHz, respectively. From right to left, we observe: The main photopeak corresponding to the full absorption of 60 keV photons in the crystal. It overlaps with a Br K-escape peak at approximately 48 keV. A La K-escape peak at about 26 keV is also visible, which overlaps with a peak caused by an 11-22 keV X-ray emission from Am-241.

We used Gaussian fits to determine the mean pulse heights corresponding to the five main peaks (photon energies) in the spectra of the three radioactive sources. The results are summarized in Fig. 3c and indicate that the detectors have a proportional response in the energy range of interest. As Fig. 3a and Fig. 3b show, a double Gaussian fit with a constraint on the distance between both peaks was used if a K-escape peak clearly overlapped with the photopeak. We derived the energy resolutions at 60 keV from the fits and found 19.9% FWHM for f_c=25 MHz and 22.3% FWHM for f_c=100 MHz.

Based on linear interpolation of the data points in Fig. 3c, we determined that the 15 keV and 30 keV thresholds are at 31.0 mV and 58.4 mV, respectively, for f_c =25 MHz. In case f_c=100 MHz, these thresholds are at 41.8 mV and 77.2 mV.

Based on the mean shape of the pulses due to the 122 keV photons from Co-57 (see Fig. 2b), the time-over-the x keV threshold for f_c=y MHz ($ToT_{x,y}$) was found to be as follows: $ToT_{15,25}$=68 ns, $ToT_{30,25}$=55 ns, $ToT_{15,100}$=53 ns, and $ToT_{30,100}$=42 ns. Thus, we used these values of τ_{np}: $\tau_{np,15,25}$=70 ns, $\tau_{np,30,25}$=60 ns, $\tau_{np,15,100}$=55 ns, and $\tau_{np,30,100}$=45 ns.

Fig. 4 shows the observed count rate (OCR) as a function of tube current I_{tube}. We estimated an incident count rate (ICR) for each value of I_{tube} by taking into account that ICR and I_{tube} are proportional to each other and by assuming that ICR=$OCR_{np,15,100}$ for I_{tube}=0.5 mA (the selected source-detector distance was such that the pile-up level was limited for I_{tube}=0.5 mA). The estimated ICRs are displayed on the top horizontal axes in Fig. 4 and indicate that we characterized the detector up to ICR≈31 Mcps/pixel. As expected, the curves for p-like counting feature a maximum OCR, which lies between 6 Mcps/pixel and 7.5 Mcps/pixel for f_c=25 MHz, and between

8 and 10.5 Mcps/pixel for f_c=100 MHz, depending on the threshold value. The curves for np-like counting should approach asymptotes defined by OCR=$1/\tau_{np}$ for high values of I_{tube}, i.e., they should approach OCRs of 14.3-16.7 Mcps/pixel for f_c=25 MHz and 18.2-22.2 Mcps/pixel for f_c=100 MHz, again depending on the threshold value. However, Fig. 4 shows that, especially for f_c=100 MHz, we did not reach these asymptotic values at the maximum available tube current of 20 mA yet. Thus, we plan to take measurements at shorter source -detector distances and show the results at the conference.

To put the results into perspective, existing CdTe and CZT detectors that have been highly optimized for medical photon-counting CT aim for a maximum OCR of 10-15 Mcps/pixel in case of p-like counting (e.g., the CZT detector in the prototype scanner of Philips [5]) or an OCR approaching 25-30 Mcps/pixel in case of np-like counting (e.g., the CdTe detector in the prototype scanner of Siemens [4]). Our prototype $LaBr_3$:Ce detector already approaches these numbers closely.

The rate capability per mm^2 can be further improved by having multiple pixels of 0.5 mm × 0.5 mm or smaller on a square millimeter. However, the reflective isolation between pixels (see Fig. 1a) then starts to become a relatively large X-ray insensitive area that limits the achievable OCR and dose efficiency. However, Canon managed to manufacture an energy-integrating detector for its Aquilion Precision CT scanner that has scintillator pixels of such small dimensions with very thin reflective septa. Even dead area-free options exist, such as a columnar microstructure of the scintillator (cf. the CsI scintillator in flat panel detectors) [9], and laser-induced optical barriers [10].

Lastly, examples of X-ray tube spectra measured using the $LaBr_3$:Ce detector under low fluence rate conditions (I_{tube}=0.5 mA) are shown in Fig. 5. The typical shape of such spectra is clearly present in the data, including signs of the characteristic X-rays from the tungsten target of the tube. For photon energies exceeding 80 keV, the spectral intensity gradually decreases towards the maximum energy of 120 keV, with some overflow to higher energies because of the finite energy resolution of the detector and some pulse pile-up.

IV. Conclusions

We are investigating ultrafast scintillation detectors with SiPM readout as alternatives to direct-conversion detectors for (medical) X-ray photon-counting applications. In this work, we built a 1 mm×1 mm single-pixel detector consisting of the fast and commercially available $LaBr_3$:Ce scintillator and an ultrafast SiPM prototype. We measured energy resolutions in the range 20%-22% at 60 keV, depending on the cut-off frequency of the filter used in the pulse processing. Moreover, we were able to measure an X-ray tube spectrum with signs of its characteristic features. We also measured count rate curves and found a maximum observed count rate (OCR) of 10.5 Mcps/pixel for paralyzable-like counting, a threshold of 30 keV, and a cut-off frequency of 100 MHz. In case of nonparalyzable-like counting, the OCR seemed to approach an asymptotic value exceeding 20 Mcps/pixel. This means the performance of our prototype detector already approaches that of CdTe/CZT detectors optimized for photon-counting CT.

References

[1] S.S. Hsieh et al., "Photon counting CT: Clinical …", *IEEE TRPMS,* vol. 5, no. 4, pp. 441-452, Aug. 2020. doi: 10.1109/TRPMS.2020.3020212.

[2] T. Flohr et al., "Photon-counting CT review", *PhysMed,* vol. 59, pp. 126-136, Nov. 2020. doi: 10.1016/j.ejmp.2020.10.030.

[3] M. Persson et al., "Upper limits of the photon fluence rate on CT…", *MedPhys,* vol. 43, no. 7, pp. 4398-4411, Jul. 2016. doi: 10.1118/1.4954008.

[4] S. Kappler et al., "Photon counting CT at elevated X-ray tube currents: Contrast …", *ProcSPIE* 90331C, Mar 2014. doi: 10.1117/12.2043511.

[5] R. Steadman et al., "ChromAIX2: A large area, high count-rate energy-...", *NIMA*, vol. 862, pp. 18-24, Aug. 2017. doi: 10.1016/j.nima.2017.05.010.

[6] J. da Silva et al., "Resolution characterization of a silicon-based, …", *JMI* vol. 6, no. 4, 043502, Oct. 2019. doi: 10.1117/1.JMI.6.4.043502.

[7] U.N. Roy et al., "Role of selenium addition to the CdZnTe matrix for…", *SciRep*, vol. 9, 1620 Feb. 2019. doi: 10.1038/s41598-018-38188-w.

[8] S.J. van der Sar et al., "Silicon photomultiplier-based scintillation…", *MedPhys*, vol. 48, no 10, pp. 6324-6338, Oct 2021. doi: 10.1002/mp.14886.

[9] H.B. Bhandari et al., "Large-area crystalline microcolumnar $LaBr_3$:Ce …", *IEEE TNS,* vol. 60, no. 1, pp. 3-8 Feb 2013. doi: 10.1109/TNS.2012.2213612.

[10] H. Sabet et al. "A sub-mm spatial resolution LYSO:Ce detector for small …", *IEEE NSS-MIC,* Nov. 2015. doi: 10.1109/NSSMIC.2015.7582201

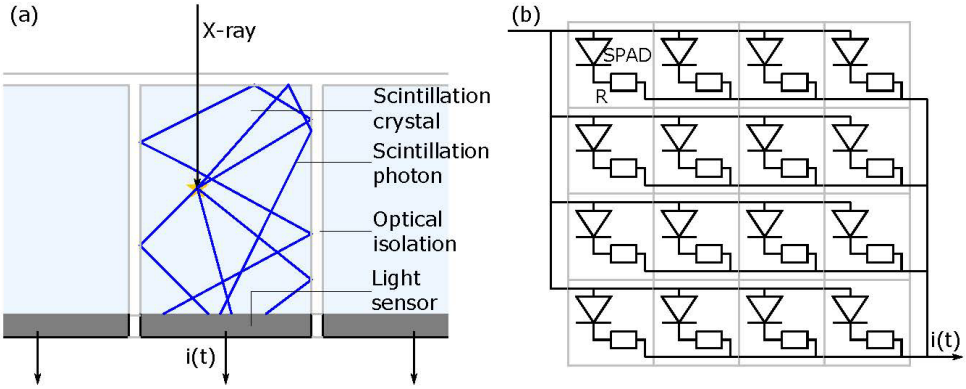


Fig. 1. a) Schematic side view of a scintillation detector. Each pixel consists of a scintillation crystal, in which an X-ray photon is converted into a pulse of optical scintillation photons, one-to-one coupled to a light sensor, which converts the light pulse into a current pulse $i(t)$. The isotropically emitted photons are guided towards the light sensor by a reflective optical isolation around the pixel, which also prevents light sharing among pixels (cf. charge sharing in semiconductor detectors). b) Schematic top view of the light sensor in our case, i.e., a silicon photomultiplier (SiPM). An SiPM is a 2D array of single photon avalanche diodes (SPADs) connected in parallel. When a SPAD detects an optical photon, an avalanche multiplication starts, which the resistor R quenches [8].

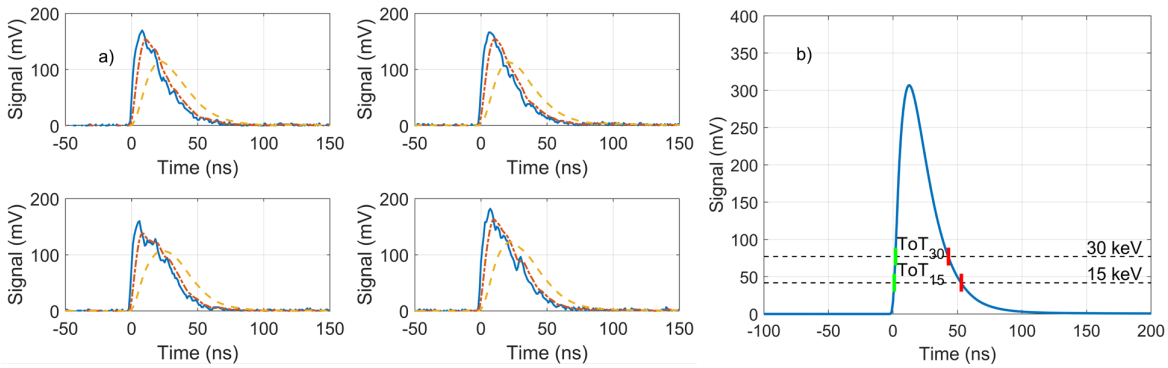


Fig. 2. a) Four examples of pulses due to 60 keV photons. The solid curves are the raw detector pulses, whereas the dashed and dash-dotted curves describe these pulses after they have been smoothened by second-order low-pass filters with cut-off frequencies f_c of 25 MHz and 100 MHz, respectively. b) The mean pulse shape for 122 keV photons and f_c=100 MHz. Start and end points of the time-over-threshold (ToT) at 15 and 30 keV used to determine τ_{np} are visualized, too.

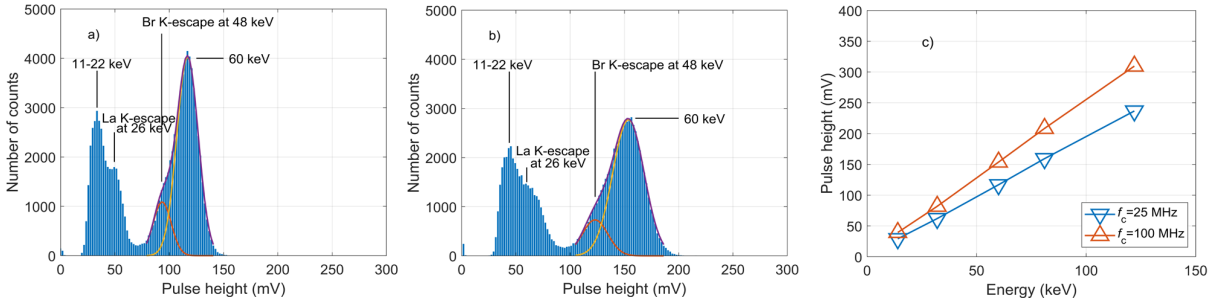


Fig. 3. a) An Am-241 pulse height spectrum measured using the $LaBr_3$:Ce detector and a second order low-pass filter with cut-off frequency f_c=25 MHz. The double Gaussian fit accounts for the K-escape peak at approximately 48 keV in order to determine the FWHM pulse height resolution at 60 keV. b) The same Am-241 spectrum measured using the same detector, but with f_c=100 MHz. c) Results of the pulse height calibration, showing highly proportional behavior.

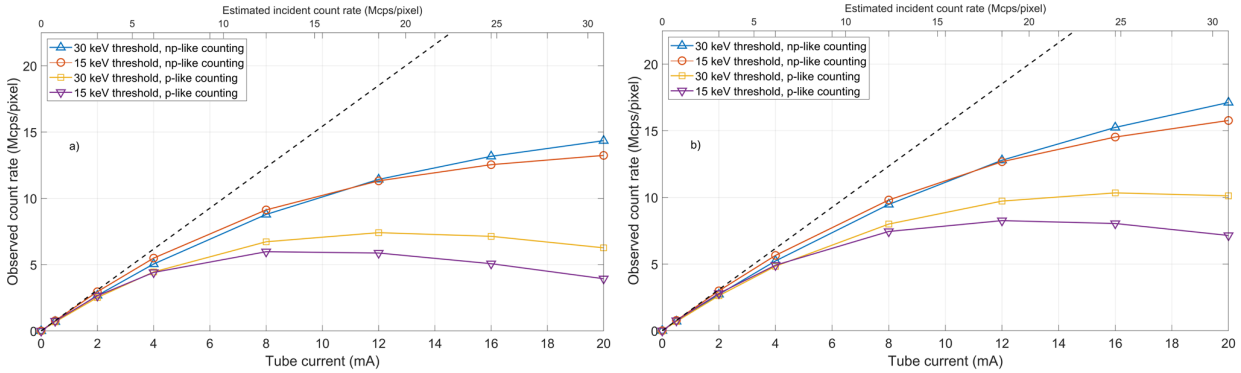


Fig. 4. Observed count rate as a function of tube current, counting algorithm, and low-energy threshold for a) the $LaBr_3$:Ce detector and a second order low-pass filter with cut-off frequency f_c=25 MHz and b) the same detector, but a filter with f_c=100 MHz. The top horizontal axes display our estimates of the incident count rate corresponding to each value of the tube current. The dashed lines represent ideal counting behavior, i.e., the observed rate equaling the incident rate.

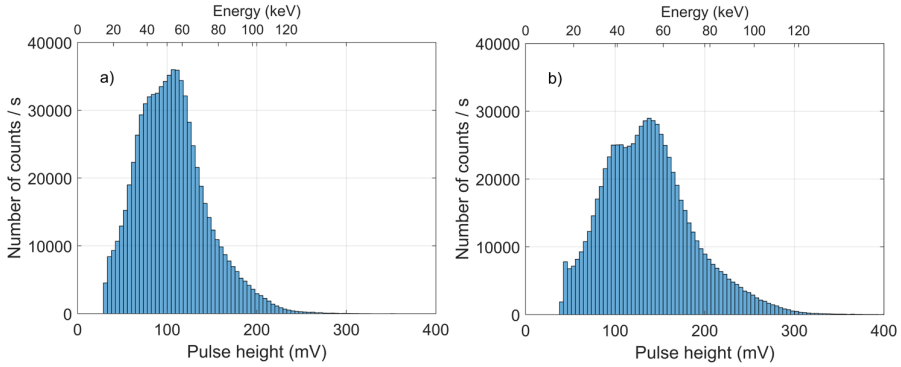


Fig. 5. X-ray tube spectra (120 kVp, tungsten anode, anode angle of 20°, 3.0 mm Be and 7.5 mm Al filtration) measured using a) the $LaBr_3$:Ce detector and a second order low-pass filter with cut-off frequency f_c=25 MHz and b) the same detector, but a filter with f_c=100 MHz, for p-like counting and a 15 keV threshold under low fluence rate conditions (i.e., a tube current of 0.5 mA). The energy-axes are based on the pulse height calibrations shown in Fig. 3c).

Preliminary Investigations of a Novel Dynamic CT Collimator

J. Webster Stayman, Nir Eden, Yiqun Q. Ma, Grace J. Gang, Allon Guez

***Abstract*—In this work we describe a new dynamic x-ray collimator that may be used to collect sparse computed tomography projection data. Data sparsity may be user-specified and controlled both angularly and radially - allowing a broad range of acquisition strategies. We consider protocols that have fully sampled projection data for a volume-of-interest with a sparsely sampled background. Model-based reconstruction methods are adapted to process the non-uniformly sampled projections. We demonstrate the ability of a CT system with this novel dynamic collimator to provide user controllable regional image quality and dose reduction in a set of phantom experiments.**

***Index Terms*—Dynamic Bowtie Filter, X-ray Modulation, Dose Reduction, Region-of-Interest Imaging**

I. Introduction

Image quality and radiation exposure are closely related in x-ray computed tomography (CT). In general, higher exposures are used to produce higher quality images. This is particularly true when image quality is largely driven by quantum noise associated with the incident x-ray beam. Thus control over the intensity of x-ray beam provides an important control over the balance between radiation dose and image quality. Modern clinical CT systems provide control over the overall beam intensity through tube current modulation. Dynamic current control can be optimized to provide minimum noise variance in images.[1] The spatial distribution of fluence may also be shaped - typically through bowtie filters that attenuate the beam more at the periphery of the patient where less fluence is required. Current clinical systems often have the capability to select between a small number of static filters.

There have been many research efforts to construct modulators to permit dynamic control of the spatial distribution of the x-ray beam. Such devices are often described as dynamic bowtie filters and have taken many forms. Several designs based on variable beam attenuation have included actuated split filter designs, fluid-filled filters[2][3], piece-wise linear filters with independently controlled leaves[4], and grid-like structures that shape the beam based on angle of incidence[5]. Other designs have leveraged binary filters that block x-rays on a fine scale (below system resolution)[6] or that reduce fluence at a position based on dwell time.[7][8]

With the advent of advanced reconstruction algorithms, another widely researched strategy for altering the dose-image quality trade-off is sparse sampling. Angular undersampling has been widely explored and is straightforward to implement in pulsed cone-beam CT systems[9]. Combined angular and radial undersampling has been explored with the use of moving, structured collimators[10] and with alternate system geometries (e.g. a moving line detector with a pulsed source[11]).

Such dynamic beam control permits increased control over radiation dose and image quality. For example, the beam profile may be dynamically controlled to avoid radiation sensitive organs or to provide a customized regional dose deposition[12]. Customized regional image quality may also be specified or optimized for particular imaging tasks or anatomical locations [Gang]. This includes specialized volume-of-interest (VOI) scans that focus on a particular organ or region (e.g. cardiac, spine, etc.). With some acquisition approaches, only a small region-of-interest is exposed leading to increased complexity in data processing/local tomography due to truncated projections. Other approaches are able to provide very low exposure outside the VOI avoiding full truncation of projection data.

This work considers a new dynamic modulator that incorporates many of the ideas mentioned above. Specifically, the device uses a multitude of individually actuated ”fingers” that locally block the x-ray beam. This enables various acquisition protocols including a fully sampled VOI surrounded by a background with angularly and radially sparse projection data. In this paper we describe the construction of the device and its installation in an x-ray imaging test-bench. Model-based image reconstruction is adapted and applied to the sparse projection data acquired from the experimental bench in two phantom experiments.

II. Methods

A. Dynamic Collimator Design and Integration

A photograph of the proposed dynamic collimator is shown in Figure 1A. The device consisted of 26 independently actuated beam blockers. Each ”finger” is made of a lead compound and is approximately 3 mm wide and 1.5 mm thick. This thickness stops $> 99\%$ of x-rays to provide a nearly binary beam profile. The blockers are connected to a linear solenoid which is computed controlled. The array of blockers is contained in a 3D-printed housing with a lead-shielded entrance and exit slot (approximately 6 mm in height) on either side of the housing.

The collimator was mounted on an experimental x-ray test-bench comprised of a flat-panel detector (Varex 4343CB, Salt Lake City, UT), a radiographic/fluoroscopic x-ray tube (Varex Rad-94, Salt Lake City, UT), and a motion stage (Physik Instrumente, Auburn, MA). The dynamic collimator was placed with the blockers at $\sim$30 cm from the x-ray focal spot. This permitted full lateral coverage of the 43 cm detector in a CT system with 785 mm source-to-axis distance and

Web Stayman, Yiqun Ma, and Grace Gang are from the Department of Biomedical Engineering, Johns Hopkins University, Baltimore, MD.

Nir Eden is from Art Cybe, Kafr Sabba, Israel

Allon Guez is from Philadelphia, PA.

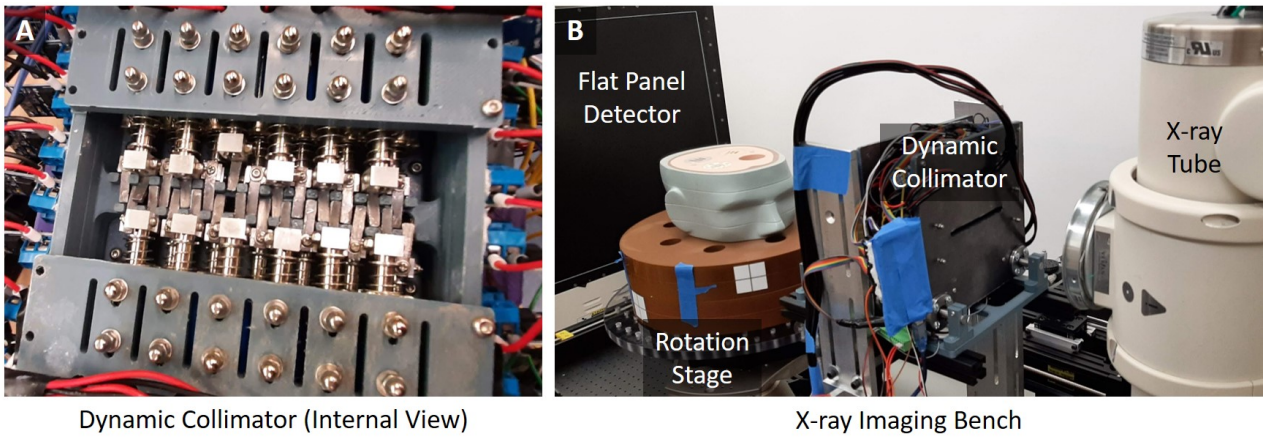


Fig. 1. (A) Photograph of the fabricated dynamic collimator with 26 independently controlled blockers. (B) Integration of the dynamic collimator on an experimental x-ray imaging test-bench.

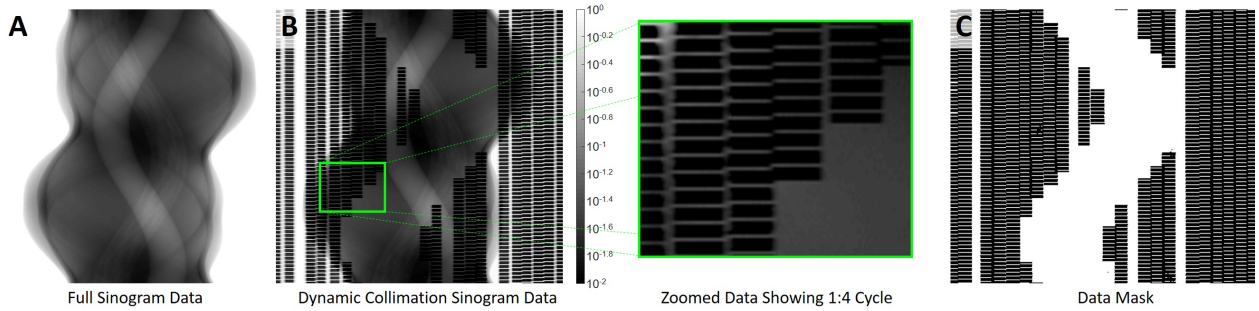


Fig. 2. Illustration of fully sampled and dynamically controlled sampling. (A) Full sinogram data. (B) Dynamically collimated data with full sampling in an off-center VOI and 1:4 sampling outside the VOI. (C) A data mask generated for model-based reconstruction of the dynamically collimated data. (Note that there are four faulty blockers in this data which are stuck open.)

1050 mm source-to-detector distance. The x-ray tube was fitted with 2.1 mm of aluminum and 0.2 mm of copper filtration without any bowtie filter.

Computer controls for the collimator have been developed and integrated within the x-ray test-bench so that blocker motion occurs (with rotary stage motion) between acquisition frames in a step-and-shoot acquisition. Control software permits arbitrary positioning of blockers (e.g. open or closed) for every projection frame.

B. Data Acquisition and Processing

While the dynamic collimator permits arbitrary actuation of each blocker over each data frame, in this work we have focused on fully sampling a VOI surrounded by an under-sampled background. This acquisition should provide image quality within the VOI that is comparable to fully sampled data and (generally) lower image quality in the background. We note that the sparse background sampling should help to avoid the complexity of local tomography and fully truncated data. Moreover, even if the background is lower image quality, this provides additional context for clinical tasks that would be absent from fully truncated data.

In these preliminary studies, we have opted to apply a duty cycle of 1:4 with respect to the background sampling. That is, outside the VOI, blockers were open for 1 frame followed by 4 closed frames. Moreover, open blockers were shifted frame-to-frame providing a combination of sparsity both angularly and radially. Thus, the exposure and sampling outside the VOI should be approximately 1/5 of the fully sampled case. This protocol for data collection is illustrated in Figure 2.

To reconstruct data acquired using this protocol, we used a modified penalized weighted least-squares (PWLS) reconstruction approach. Specifically, the mean measurement model was

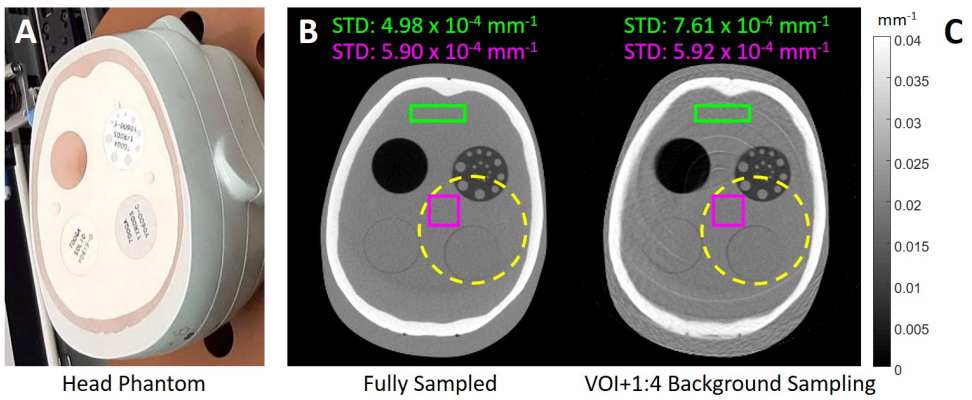
$$\bar{\mathbf{y}}(\mathbf{x}) = \mathbf{I}_0 \exp\Big(-\mathbf{A}\mu\Big), \tag{1}$$

where $\mathbf{I}_0$ denotes the incident x-ray intensity, $\mathbf{A}$ is the system matrix representing forward projection, and μ denotes the vector of attenuation coefficients we wish to estimate. The PWLS objective is

$$\hat{\mu} = \arg\min_{\mu} \Big(\mathbf{A}l - \mu\Big)^{\mathbf{T}} \mathbf{W}\Big(\mathbf{A}l - \mu\Big) + \beta\Big|\Big|\Psi\mu\Big|\Big| \tag{2}$$

$$l = -\log(y/\mathbf{I}_0) \tag{3}$$

where l denotes an estimate of the line integrals from the measurements y. The weighting matrix $\mathbf{W}$ is a modified version of the standard PWLS weighting by the inverse of



C

	Fully Sampled	
Raw Measure-ments	Center	0.72 mGy
	Peripheral	3.92 mGy
CTDI	2.85 mGy	

	VOI + 1:4 Background	
Raw Measure-ments	Center	0.70 mGy
	Front	2.21 mGy
	Left	2.23 mGy
	Back	1.91 mGy
	Right	2.05 mGy
CTDI	1.63 mGy (57% of Full)	

Fig. 3. Summary of head phantom experiments. (A) Photograph of head phantom with three inserts. (B) Comparison of fully sampled and VOI+1:4 sparse sampling reconstructions. Two rectangular regions are identified over which sample noise variance is computed. The VOI is indicated by a dashed yellow circle. (C) Dose measurements for each acquisition protocol.

the variance. A total variation penalty with weighting β and first-order difference operator Ψ was applied.

To account for the sparse sampling pattern, a projection mask was estimated from a system gain scan. Specifically, the same actuation pattern was applied with no object in the scanner. A mask was formed using simple thresholding followed by a dilation operation to increase the size of mask. Dilation was applied only in the radial direction of the sinogram. A modified diagonal weighting matrix was formed such that $\mathbf{W} = \mathbf{D}\{y \odot \text{mask}\}$. In essence this informs the PWLS algorithm that blocked measurements should be ignored for reconstruction.

For all experiments in this work, a single central row of flat-panel data was used to form a sinogram. This projection data was comprised of 2x2 binned pixels (0.556 mm) and 360 projections. Reconstructions were formed using 200 iterations of the separable paraboloidal surrogates method[13] using 20 angular subsets and 0.75 mm voxels.

C. Physical Experiments

Two different physical experiments were conducted. Both use the system geometry and processing scheme described above. Acquisitions were conducted using an x-ray technique of 120 kVp and 20 mA. Exposures of 1 ms yield 0.02 mAs/frame. With 360 frames the total exposure was 7.2 mAs.

The first experiment used an ATOM head phantom (CIRS, Norfolk, VA) and two sampling strategies: 1) fully sampled data, and 2) the VOI plus 1:4 background sampling scheme described above. Data were reconstructed using the above processing with $\beta = 1$ for fully sampled data and $\beta = 1.3$ for the dynamically collimated data. Noise was measured at location both inside and outside the VOI.

Radiation dose was measured for each scan using a standard 16 cm CTDI head phantom and a 0.6 cc Farmer chamber (RadCal, Monrovia, CA). Center and peripheral dose was measured for the fully sampled scan, and all five dosimeter locations were used for the dynamically collimated scan. Total dose was computed using $1/3$ center plus $2/3$ peripheral averages.

A second experiment was performed using a thorax phantom (Kyoto Kagaku, Japan). Both fully sampled and dynamically collimated data were acquired. A VOI around the spine was selected and scanned using the VOI plus 1:4 sparsity as well as a VOI plus 2:4 sparsity. The latter method used open pairs of adjacent blockers.

III. Results

Results of the head phantom experiment are summarized in Figure 3. Note that within the circular VOI the noise levels (as computed in the region identified by the magenta square) are nearly identical suggesting a very similar level of image quality. Outside the VOI, the noise levels (computed in green rectangle) are approximately 1.5 times higher in the VOI scan. We also note some ring artifacts associated with the boundaries of the blockers. We conjecture that this is the result of incomplete gain correction, gaps in radial sampling due to overlap of blockers (e.g. a whole column of missing data in the sinogram), and other uncorrected physical effects like scatter, lag, etc. Dosimetry results are summarized in Figure 3C. We see that the dynamic collimation scan was acquired with 57% of the dose of the fully sampled scan - indicating a significant potential for dose reduction.

Results from the thorax scan experiment are shown in Figure4. In comparing the fully sampled scan with the two dynamic collimation protocols we see a similar level of image quality in the VOI that was chosen around the spine for all methods. As one would expect the 1:4 protocol has increased noise over both the 2:4 protocol and the fully sampled data. Interestingly, the ring artifacts are significantly reduced in the 2:4 protocol. We conjecture that the pairwise actuation of blockers eliminates the systematic omission of radial bins that are present in the singlet blocker opening, which reduces systematic ring artifacts. The image quality outside the VOI is better in the 2:4 protocol (over the 1:4 case) with its increased sampling.

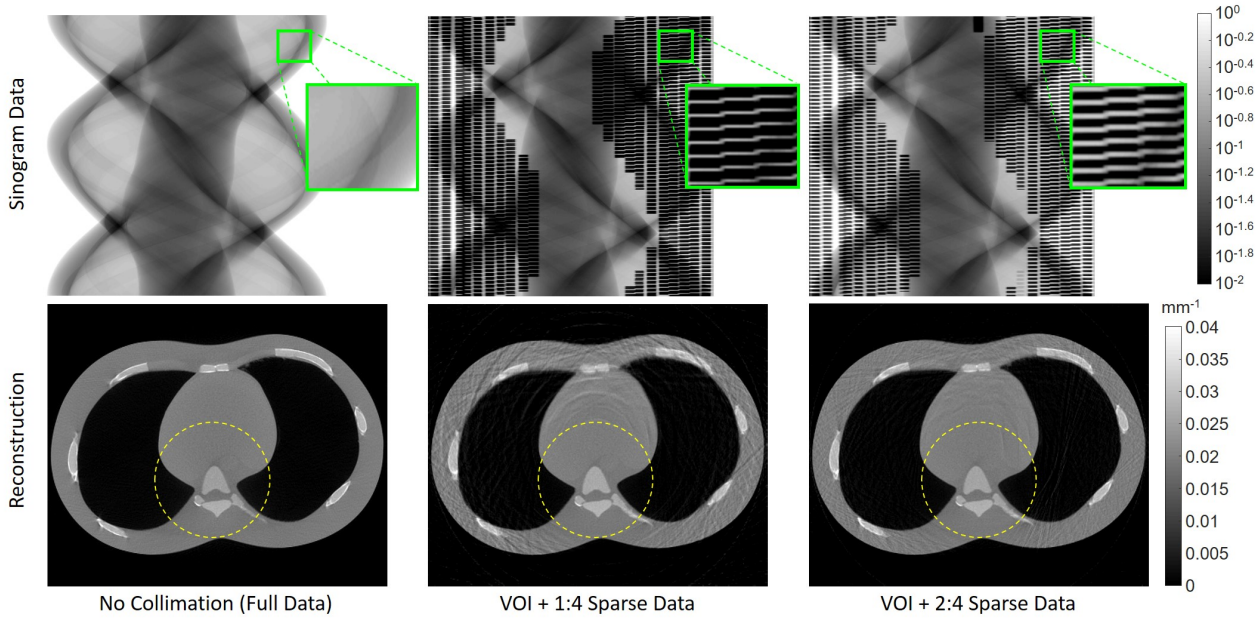


Fig. 4. Summary of the thorax phantom experiment including fully sampled data and two different sparse sampling protocols: VOI+1:4 and VOI+2:4. Both sinogram data and the associated PWLS reconstruction are shown for each case. The circular VOI is shown as a yellow dashed circle in all volumes.

IV. Conclusion

In this work we have demonstrated a novel dynamic collimation approach based on the independent actuation of beam blockers placed along the radial direction of a CT scanner. The device was constructed, calibrated, and used to acquire projection data on a experimental test-bench. A model-based approach was developed to reconstruct data from acquisition protocols with variable sampling. We demonstrated the ability to maintain image quality within a VOI while significantly reducing radiation dose.

These preliminary studies show one potential use of the dynamic collimator to control image quality and limit dose. Much more general dynamic sampling patterns are possible. For example, we have illustrated how pairwise actuation can limit certain artifacts. A wide range of other image quality and dose objectives could also be applied including smoother variations in image quality (spatially), organ-by-organ specification, task-driven sampling, etc.

There are numerous engineering details that would need to be addressed to translate this approach to a clinical system including strict constraints on the motion (accelerations, velocity), overall size, control and communication, etc. However, these initial studies suggest that the underlying technology has potential to provide a new way to control and balance the image quality-dose trade-off.

References

[1] M. Gies, W. A. Kalender, H. Wolf, C. Suess, M. Gies, W. A. Kalender, H. Wolf, C. Suess, M. T. Madsen, M. Gies, W. A. Kalender, H. Wolf, and C. Suess, "Dose reduction in CT by anatomically adapted tube current modulation. I. Simulation studies." *Medical physics*, vol. 26, no. 11, pp. 2248–2253, nov 1999.

[2] P. Shunhavanich, S. S. Hsieh, and N. J. Pelc, "Fluid-filled dynamic bowtie filter: a feasibility study," *Proc. SPIE*, vol. 9412, p. 94121L, 2015.

[3] T. P. Szczykutowicz and J. Hermus, "Fluid dynamic bowtie attenuators," *Proc. SPIE*, vol. 9412, p. 94120X, 2015.

[4] S. S. Hsieh and N. J. Pelc, "The feasibility of a piecewise-linear dynamic bowtie filter." *Medical physics*, vol. 40, no. 3, p. 031910, feb 2013.

[5] S. M. Huck, G. S. K. Fung, K. Parodi, and K. Stierstorfer, "Technical Note: Sheet-based dynamic beam attenuator – A novel concept for dynamic fluence field modulation in x-ray CT," *Medical Physics*, jul 2019.

[6] G. J. Gang, A. Mao, W. Wang, J. H. Siewerdsen, A. Mathews, S. Kawamoto, R. Levinson, and J. W. Stayman, "Dynamic fluence field modulation in computed tomography using multiple aperture devices," *Physics in Medicine and Biology*, vol. 64, no. 10, may 2019.

[7] T. P. Szczykutowicz and C. A. Mistretta, "Design of a digital beam attenuation system for computed tomography: part I. System design and simulation framework." *Medical physics*, vol. 40, no. 2, p. 021905, 2013.

[8] T. P. Szczykutowicz and C. a. Mistretta, "Design of a digital beam attenuation system for computed tomography. Part II. Performance study and initial results." *Medical physics*, vol. 40, no. 2, p. 021906, 2013.

[9] J. Bian, J. Wang, X. Han, E. Y. Sidky, L. Shao, and X. Pan, "Optimization-based image reconstruction from sparse-view data in offset-detector CBCT." *Physics in medicine and biology*, vol. 58, no. 2, pp. 205–30, jan 2013.

[10] B. Chen, E. Kobler, M. J. Muckley, A. D. Sodickson, T. O'Donnell, T. Flohr, B. Schmidt, D. K. Sodickson, and R. Otazo, "SparseCT: System concept and design of multislit collimators," *Medical Physics*, vol. 46, no. 6, pp. 2589–2599, jun 2019.

[11] A. Leong, G. J. Gang, A. Sisniega, W. Wang, J. Wu, S. B. Bambot, and J. W. Stayman, "An Investigation of Slot-scanning for Mammography and Breast CT," *Proceedings of SPIE–the International Society for Optical Engineering*, vol. 11312, p. 25, mar 2020.

[12] S. Bartolac, S. Graham, J. H. Siewerdsen, and D. A. Jaffray, "Fluence field optimization for noise and dose objectives in CT," p. S2, 2011.

[13] H. Erdogan and J. A. Fessler, "Ordered subsets algorithms for transmission tomography." *Physics in medicine and biology*, vol. 44, no. 11, pp. 2835–51, nov 1999.

CT imaging with truncated data over limited-angular ranges

Dan Xia[1], Zheng Zhang[1], Buxin Chen[1], Emil Y. Sidky[1], and Xiaochuan Pan[1,2]
[1]*Department of Radiology, The University of Chicago, Chicago, USA*
[2]*Department of Radiation and Cellular Oncology, The University of Chicago, Chicago, USA*

***Abstract*—In certain CT applications such as dental CT imaging, a scanning configuration with an offset-detector is often used for extending the field of view (FOV) of the system. While data are truncated on one-side of the detector, it remains possible to accurately reconstruct an image from the truncated data collected over a full-angular range (FAR) of 360° by use of existing analytical-based algorithms such as the FDK algorithm. However, there also exist interests in scanning configurations that collect data only over limited-angular ranges (LARs) for practical considerations, and existing algorithms generally yield reconstructions with significant artifacts from LAR data collected with an offset-detector. It has been demonstrated recently that, for non-truncated data, the directional-total-variation (DTV) algorithm can reconstruct images with significantly reduced artifacts from LAR data. In this work, we developed and tailored the DTV algorithm for image reconstruction from truncated LAR data collected with a scanning configuration employing an offset-detector. We carried out a study on image reconstruction for a number of LAR scanning configurations with an offset-detector of practical interest. The study results demonstrate that the DTV algorithm can be tailored to yield, from truncated LAR data, images with significantly reduced artifacts that are observed otherwise in images obtained with existing analytical-based algorithms.**

***Index Terms*—limited-angular range (LAR), truncated data, offset detector, directional total variation (DTV), primal-dual algorithm**

I. INTRODUCTION

In certain CT applications, a scanning configuration with an offset-detector is used often for effectively increasing its field of view (FOV) [1], [2]. While data collected with an offset-detector at each view are truncated on one side of the detector, accurate images can be reconstructed from the truncated data collected over a full-angular range (FAR) of 360°. It is also recognized that a reduced scanning angular range can be exploited for potentially lowering radiation dose and scanning time and for avoidance of the collisions between the gantry and scanned objects, e.g., in C-arm cone-beam CT (CBCT). Therefore, it is of interest and potential value to investigate scanning configurations with offset-detectors for data collection over limited-angular ranges (LARs) that are considerably shorter than the FAR of 360°. Clearly, existing algorithms, such as the FDK algorithm and its variations widely used for image reconstruction from FAR data collected with an offset-detector, will yield images with significant artifacts when applied to LAR data collected with the same offset-detector.

It has been shown recently that the directional-total-variation (DTV) algorithm can accurately reconstruct images from LAR data without any truncation [3], [4]. In this work, we developed and tailored the DTV algorithm for image reconstruction from truncated LAR data collected with scanning configurations in which an offset-detector is employed. The study results demonstrate that the DTV algorithm can be tailored to yield, from truncated LAR data, images with significantly reduced artifacts that are observed otherwise in images obtained with existing analytical algorithms.

II. MATERIAL AND METHODS

A. LAR scanning configuration with an offset-detector

In the work, the X-ray transform [5] in a discrete form is used as the imaging model, and image reconstruction is equivalent to inverting the imaging model. Clearly, the ill-conditionedness of the imaging model depends upon the LAR extent and distribution over which knowledge of the discrete X-ray transform (DXT), i.e., the imaging model, is available. Without loss of generality, we consider a circular fan-beam configuration with an offset-detector, as shown in Fig. 1a. From the head image in Fig. 1b, truncated data generated with the configuration over an FAR of 360° are shown in Fig. 1c. On the other hand, an LAR scanning configuration with an offset-detector is shown in row 1 of Fig. 2a, which consists of two orthogonal arc components, separated by 90°, indicated by arcs α_1 and α_2, where α_1 and α_2 indicate their angular ranges. We refer to this configuration as a two-orthogonal-arc (TOA) LAR configuration, and thus as a single-arc (SA) LAR configuration if $\alpha_1 = 0$ or $\alpha_2 = 0$ as displayed in row 2 of Fig. 2a. We investigated image reconstruction from data, as shown in Figs. 2b-2d, collected over the TOA and SA scanning configurations, of the same total angular range, with an offset-detector.

In the circular fan-beam configuration, the distances from the X-ray source to the detector and to the rotation center are 60 and 40 cm, respectively. The detector consists of 754 elements with a size of 0.2 mm. A detector-offset L = 30 mm is used for yielding an extended FOV. The head phantom shown in Fig. 1b consists of 200×200 pixels of size 0.8×0.8 mm^2.

Using the head phantom, we first generate data, as shown in Fig. 1c over an FAR of 360° with an angular interval of 0.5° per view. For each view, the generated data are truncated on one side of the detector as a result of the detector-offset.

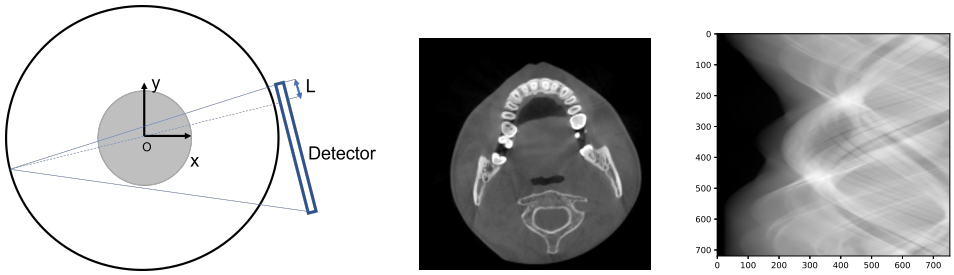


Fig. 1. (a) A circular fan-beam scanning configuration with an offset-detector in which O and L denote the rotation center and the length of the detector-offset. (b) The numerical head phantom, and (c) data collected over an FAR of 360°. It can be observed that the data are truncated in the right-hand side of the data space in (c). Display windows are [0.15, 0.3] cm^{-1} and [0, 3], respectively.

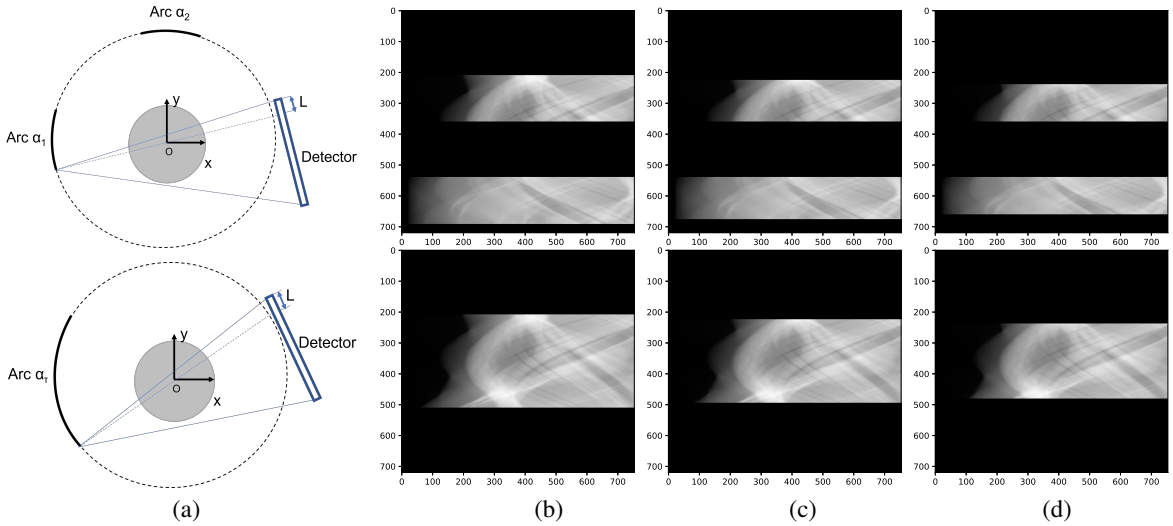


Fig. 2. (a) Two-orthogonal-arc (row 1) and single-arc (row 2) scanning configurations with an offset-detector. Data generated with the two-orthogonal-arc scanning configurations (row 1) over LARs (α_1, α_2) = $(75^\circ, 75^\circ)$ (b), $(67.5^\circ, 67.5^\circ)$ (c), and $(60^\circ, 60^\circ)$ (d), and with the single-arc scanning configurations (row 2) over LARs α_τ=150° (b), 135° (c), and 120° (d). Display window is [0, 3].

We consider TOA scanning configurations with LARs (α_1, α_2) = $(75^\circ, 75^\circ)$, $(67.5^\circ, 67.5^\circ)$, and $(60^\circ, 60^\circ)$, and use them to generate data from the head phantom, as shown in row 1 of Figs. 2b-2d. For comparison, we have also investigated image reconstruction from data of the head phantom generated over SAs with LARs $\alpha_\tau = \alpha_1 + \alpha_2 = 150^\circ$, 135°, 120°, as shown in row 2 of Figs. 2b-2d.

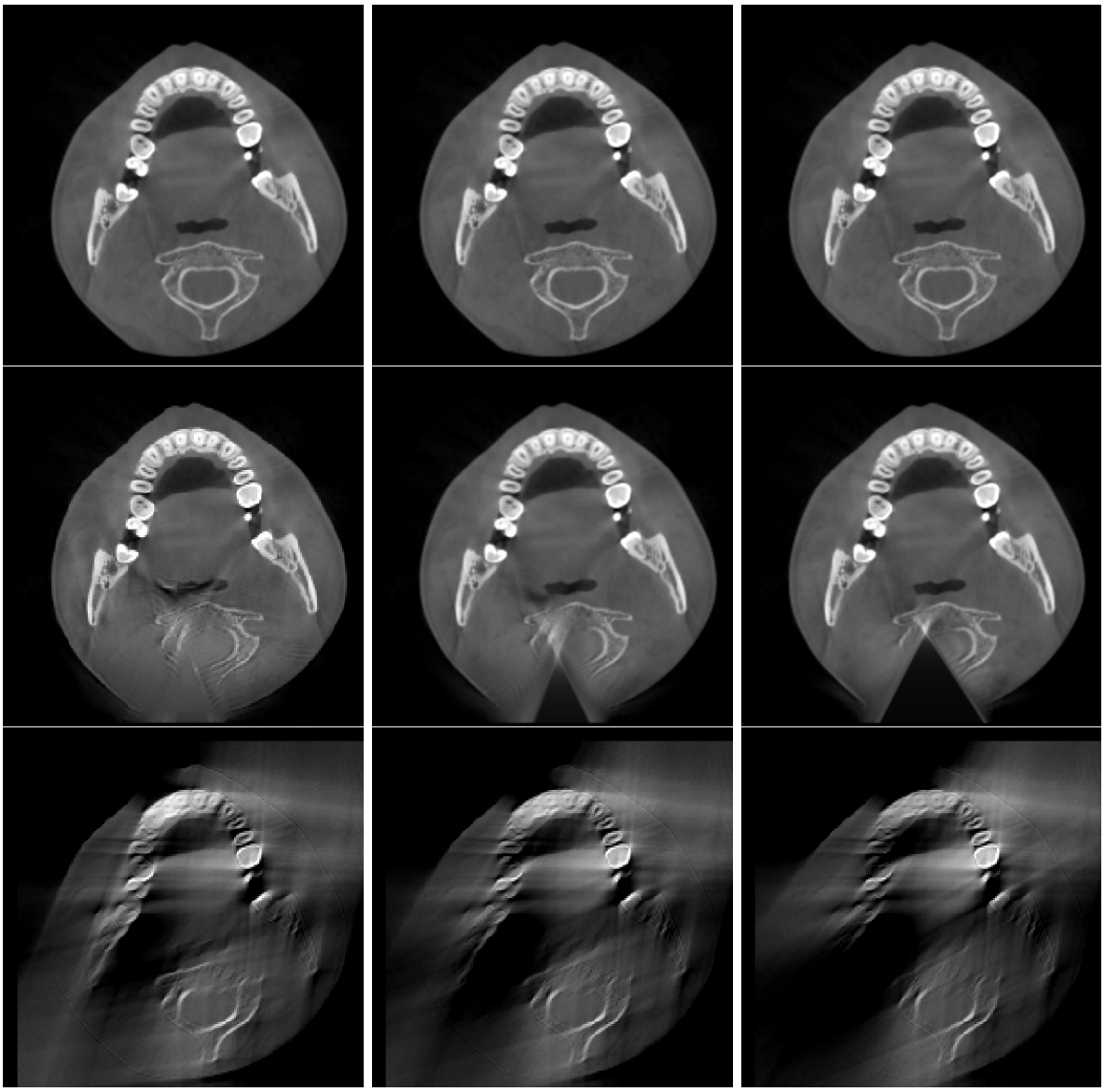
In addition, using noiseless LAR data generated above, we created noisy LAR data by adding the Poisson noise, corresponding to 10^7 noise equivalent quanta (NEQ) per detector bin in air scans, and we subsequently performed image reconstructions from the noisy LAR data.

B. Image reconstruction

For CT scans over TOA or SA LAR with an offset-detector, the image reconstruction can be formulated as the solution to a constrained optimization program in which a weighted data-ℓ_2 fidelity is minimized under image-DTV constraints along the x- and y-directions. The optimization program is given by

$$\mathbf{f}^\star = \underset{\mathbf{f}}{\operatorname{argmin}} \left\{ \frac{1}{2} \parallel W(\mathcal{H}\mathbf{f} - \mathbf{g}^{[\mathcal{M}]}) \parallel_2^2 \right\}$$
$$\text{s.t.} \quad ||\mathcal{D}_x\mathbf{f}||_1 \leq t_1, \ ||\mathcal{D}_y\mathbf{f}||_1 \leq t_2, \ \text{and} \ f_i \geq 0, \qquad (1)$$

where vector $\mathbf{g}^{[\mathcal{M}]}$ of size M denotes discrete measured data acquired with two-orthogonal-arc or single-arc LAR scanning configuration; vector $\mathbf{f}$ of size N is a 2D discrete image; f_i is the entry i of $\mathbf{f}$; $\mathcal{H}$ the system matrix of size $M \times N$ representing the X-ray transform, with element $h_{j,i}$ representing the contribution of pixel i to ray j; W is an $M \times M$ diagonal matrix in which each diagonal element represents a positive weighting factor for a corresponding X-ray measurement; $||\cdot||_p$ indicates the ℓ_p-norms of the input vector for $p = 1$ and 2; matrices $\mathcal{D}_x$ and $\mathcal{D}_y$ of size $N \times N$ denote two-point differences along the x- and y-axes, respectively; vectors $\mathcal{D}_x\mathbf{f}$ and $\mathcal{D}_y\mathbf{f}$ are of size N; and parameters t_1 and t_2 depict the upper bounds on the DTV constraints along x- and y-directions.



(a) (b) (c)

Fig. 3. Row 1: images reconstructed by use of the DTV algorithm from noiseless data of TOA scanning configurations over LARs (α_1, α_2) = $(75^\circ, 75^\circ)$ (a), $(67.5^\circ, 67.5^\circ)$ (b), and $(60^\circ, 60^\circ)$ (c); row 2: images reconstructed by use of the DTV algorithm from noiseless data of SA scan configurations over LARs α_τ=150° (a), 135° (b), and 120° (c); and row 3: images reconstructed by use of the FDK algorithm from noiseless data of TOA scanning configurations used for obtaining images in row 1. Display window is [0.0, 0.6] cm^{-1}.

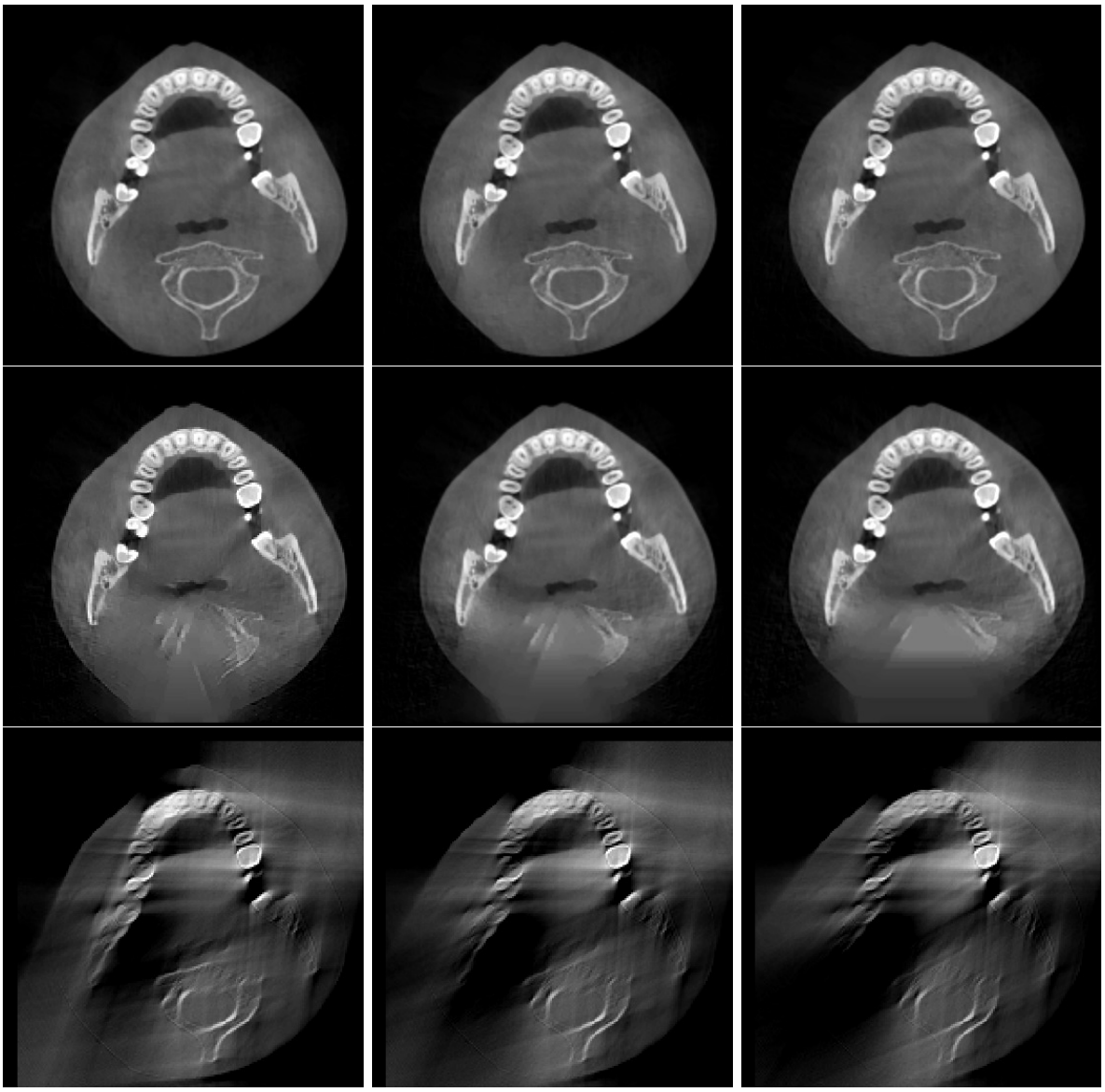
Weighting matrix W is introduced for controlling the numerical effect of the discontinuity of truncated data on image reconstruction. Different designs of weighting matrix W may result in different reconstruction programs, leading to different reconstructions when data are inconsistent with the imaging model.

Based upon the general primal-dual optimization framework [6], [7], a DTV algorithm was developed previously for solving the constrained optimization program minimizing data-ℓ_2 fidelity without weighting matrix W under DTV constraints [3], [4], [8], [9]. In this work, we tailored the DTV algorithm to solve Eq. (1) including a weighting factor W for image reconstruction from truncated data acquired with the offset-detector.

III. Results

A. Reconstruction from noiseless data

Using the tailored DTV algorithm, we first reconstruct images of the head phantom from noiseless data generated over a number of TOAs with LARs (α_1, α_2) = $(75^\circ, 75^\circ)$, $(67.5^\circ, 67.5^\circ)$, and $(60^\circ, 60^\circ)$, (i.e., data in row 1 of Figs. 2b-2d), and display the reconstructed images in row 1 of



(a) (b) (c)

Fig. 4. Row 1: images reconstructed by use of the DTV algorithm from noisy data of two-orthogonal-arc scanning configurations over LARs (α_1, α_2) = $(75°, 75°)$ (a), $(67.5°, 67.5°)$ (b), and $(60°, 60°)$ (c); row 2: images reconstructed by use of the DTV algorithm from noisy data of single-arc scanning configurations over LARs α_τ=150° (a), 135° (b), and 120° (c); and row 3: images reconstructed by use of the FDK algorithm from noisy data of two-orthogonal-arc scanning configurations used for obtaining images in row 1. Display window is [0.0, 0.6] cm^{-1}.

Fig. 3. In an attempt to compare the differences between TOA and SA LAR reconstructions from the offset-detector data (i.e., truncated data), we also perform reconstructions from noiseless data collected over several SAs of LARs α_τ = 150°, 135°, 120° (i.e., data in row 2 of Fig. 2b-2d), and show the reconstructed images in row 2 of Fig. 3. The FDK reconstructions from the truncated TOA LAR data are shown in row 3 of Fig. 3 for demonstrating the artifacts as the combined result of LAR and data truncation issues.

It can be observed that FDK reconstructions contain significant LAR artifacts that overwhelm soft-tissue contrast of interest, and tooth structures in the reconstructed images are difficult to be discerned. DTV reconstructions from both TOA and SA data appear visually comparable to the phantom image and the LAR artifacts observed in the corresponding FDK reconstructions are significantly reduced. Although the DTV reconstructions for SA scanning configurations show clearly the tooth structures without much of the artifacts in the FDK images, there remain residual artifacts around the spine region. As the LAR α_τ is reduced from 150° to 120°, the artifacts become more significant, caused by the combination of data-truncation and LAR. Conversely, the DTV reconstructions for

the TOA scanning configurations remain visually identical to the phantom image and largely free of the artifacts.

B. Reconstruction from noisy data

In order to investigate the robustness of the DTV reconstruction from truncated TOA and SA LAR data, we repeat the study in Sec. III-A with noisy data and show the results in Fig. 4. Again, for noisy LAR data considered, the DTV reconstructions for both the TOA and SA scans yield images with significantly reduced LAR artifacts as compared to the FDK reconstructions. Also, the TOA configurations appear to yield images with minimum artifacts as compared to that obtained for the SA configurations, consistent with the observation made above.

IV. DISCUSSION

In this work, we have investigated accurate image reconstruction from truncated data acquired with TOA and SA scanning configurations of LARs. The investigation is enabled by tailoring the primal-dual-based DTV algorithm to reconstruct images through solving an optimization program that includes a weighted data-ℓ_2 fidelity and DTV image constraints. The study results suggest that the tailored DTV algorithm can significantly reduce the LAR artifacts that are observed otherwise in reconstructions obtained with the existing algorithms. Moreover, it is also revealed that TOA scanning configurations generally appear to be more effective than the SA scanning configurations in terms of reducing the combined LAR and data-truncation artifacts. Further investigation will focus on the impact of the extent of detector-offset, weighting matrix W, and cone-beam effect on image reconstruction from LAR data.

V. ACKNOWLEDGMENT

This work was supported in part by NIH R01 Grant Nos. EB026282 and EB023968, and NIH R21 Grant No. 1R21CA263660-01A1. The contents of this article are solely the responsibility of the authors and do not necessarily represent the official views of the National Institutes of Health.

REFERENCES

[1] P. S. Cho, R. H. Johnson, and T. W. Griffin, "Cone-beam CT for radiotherapy applications," *Phys. Med. Biol.*, vol. 40, pp. 1863–1883, 1995.

[2] W. Chang, S. Loncaric, G. Huang, and P. Sanpitak, "Asymmetric fan transmission CT on SPECT systems," *Phys. Med. Biol.*, vol. 40, pp. 913–928, 1995.

[3] Z. Zhang, B. Chen, D. Xia, E. Sidky, and X. Pan, "Directional-TV algorithm for image reconstruction from limited-angular-range data," *Med. Image Anal.*, vol. 70, p. 102030, 2021.

[4] ——, "Image reconstruction from data over two orthogonal arcs of limited-angular ranges," *Med. Phys.*, p. In Press, 2022.

[5] A. K. Louis and F. Natterer, "Mathematical problems of computerized tomography," *Proceedings of the IEEE*, vol. 71, no. 3, pp. 379–389, 1983.

[6] A. Chambolle and T. Pock, "A first-order primal-dual algorithm for convex problems with applications to imaging," *J. Math. Imag. Vis.*, vol. 40, pp. 1 – 26, 2011.

[7] E. Y. Sidky, J. H. Jørgensen, and X. Pan, "Convex optimization problem prototyping for image reconstruction in computed tomography with the Chambolle–Pock algorithm," *Phys. Med. Biol.*, vol. 57, no. 10, pp. 3065–3091, 2012.

[8] B. Chen, Z. Zhang, D. Xia, E. Sidky, and X. Pan, "Dual-energy CT imaging with limited-angular-range data," *Phys. Med. Biol.*, vol. 66, p. 185020, 2021.

[9] ——, "Dual-energy CT imaging over non-overlapping, orthogonal arcs of limited-angular ranges," *J. X-Ray Sci. Technol.*, vol. 29, pp. 975–985, 2021.

Extension of the cone-beam CT field-of-view using two short scans with displaced centers of rotation

Gabriele Belotti, Simon Rit, Guido Baroni

***Abstract*—A robotic cone-beam computed tomography (CT) scanner has inherent advantages. In particular, it enables source and detector trajectories capable of extending the field-of-view (FOV) of the reconstructed CT images, where the FOV is defined as the region in the source trajectory plane for which all ray lines are acquired (at any position and direction in this plane). Previous extensions of the FOV used an offset detector or a displaced center of rotation and a single full 360° scan. However, due to limitations in the conventional range of motion inside the treatment room to avoid collisions, some systems can only perform short scans. This paper investigates a new approach to FOV extension for an existing system by adapting the FOV extension to two complementary short scans with displaced centers of rotation. We validate this approach on numerical simulations of the Forbild thorax phantom.**

***Index Terms*—Cone-beam computed tomography (CT), short scan, filtered backprojection, Feldkamp-Davis-Kress (FDK) algorithm, displaced center of rotation, tilted detector.**

I. Introduction

Cone-beam computed tomography (CBCT) is a common solution to provide in-room imaging in various forms of radiotherapy. Several commercial solutions are available, often as an imager mounted on the gantry of a linear accelerator. At the same time, custom solutions exist as well, in particular for hadrontherapy [1]. The National Center for Oncological Hadrontherapy (CNAO, Pavia, Italy) employs such a scanner: a custom CBCT scanner mounted on a Kawasaki robotic manipulator [2]. It provides high geometric reproducibility and is employed for both individual X-ray radiographs and CBCT acquisitions inside the treatment room before the delivery of the treatment fraction. The device was primarily intended for correcting the patient setup based on registration of the bony anatomy and cannot produce a sufficiently wide field-of-view (FOV) to avoid data truncation when targeting wider locations, e.g. the abdomen or the thorax. In this work, the FOV is defined as the region in the source trajectory plane for which all ray lines are acquired (at any position and direction). Adaptive radiotherapy using the acquired CBCT images is being investigated, but it is of clinical interest to accurately reconstruct complete axial slices in wide locations for qualitative and dosimetric assessments of interfractional variations. Like most CBCT scanners with a fixed flat panel detector (FPD), the *conventional geometry* of the system is such that for all source positions, the midline, defined by the point source and the center of the detector, is orthogonal to the rotation axis and intersects it at the center of the source trajectory [3]. Axial FOV enlargement is typically achieved through a technique called half fan, suitable for filtered backprojection (FBP) reconstruction [4]. Mechanically, half fan requires a detector capable of transverse displacements and an adjustable X-ray source collimation. Both features are not available in the aforementioned CNAO CBCT scanner. A feasible alternative is to perform a complete displaced center of rotation (DCoR) scan with a fixed tilt of the detector. This consists of a 360° scan where the midline is offset w.r.t. the axis of rotation. This offset can be realized by the CNAO CBCT scanner while the relative positions of the source and detector are unchanged and the same collimation can be used. Previous works dealt with DCoR and aimed at correcting it as a misalignment [5]. Instead, we would like to use this geometry to enlarge the reconstructed FOV, as in [3]. Such a trajectory is compatible with robotic manipulators as the one of the CNAO CBCT scanner and is easily programmable in the robot controller. However, bulky hardware in the treatment room limits the sampling arc of the C-arm to 220°. The objective of this work is to exploit a couple of short scans [6] with complementary offset centers of rotation to obtain the same FOV as a single full 360° DCoR scan would produce.

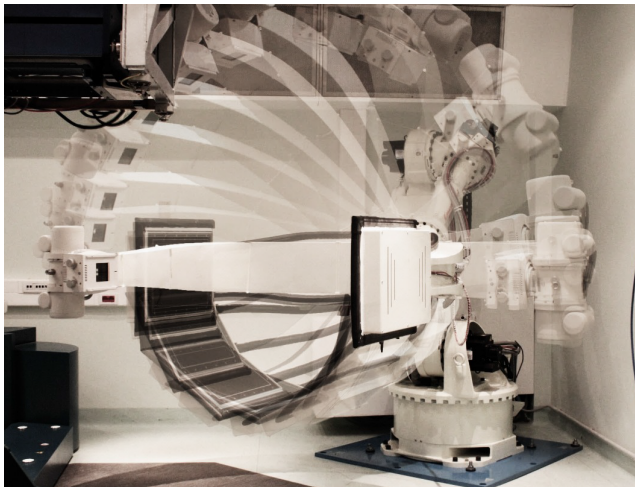


Fig. 1. Short time-lapse depicting conventional motion of custom in-room CBCT imaging system at CNAO. On the bottom left, one can observe the base of the robotic couch, which prevents the full rotation of the C-arm.

G. Belotti is with Politecnico di Milano, CartCasLab; Department of Electronics, Information and Bioengineering, Italy. S. Rit is with the Université de Lyon, CREATIS; CNRS UMR5220; Inserm U1294; INSA-Lyon; Université Lyon 1; Centre Léon Bérard, France. G. Baroni is with Politecnico di Milano, CartCasLab; Department of Electronics, Information and Bioengineering and head of the Bioengineering Unit, CNAO Foundation, Italy. This work was supported by CNAO Foundation (Pavia, Italy) in the project framework "Image guidance-Lateral Room" (reference number BAA9CONV01, 12/03/2019)
Corresponding author email: gabriele.belotti@polimi.it

II. Materials and Methods

A. Circular geometry with displaced center of rotation

We examine the scanner geometry in the central plane $z = 0$ containing the trajectory of the source S. The source and detector are attached to a C-arm which rotates around the isocenter I. We note SID and SDD the source-to-isocenter and source-to-detector distances which are fixed by construction. In the conventional operation, the source and detector rotate on a circle centered on I which is static. We note β the gantry angle between the y-axis and the midline (defined by the source S and the isocenter I).

The DCoR geometry makes use of the possibility given by the robotic manipulator to modify the position of I during the system rotation. With respect to the conventional operation, both the source and the FPD are offset from the center O in a direction orthogonal to the midline (and therefore parallel to the detector) as shown in Figure 2. We note τ the angle between the source-center line and the midline which is also the tilt angle between a conventional detector (orthogonal to the source-center line) and the DCoR detector. Like in [5] and unlike [3], τ is constant w.r.t. the gantry angle β. The angle τ is positive when I is on the positive side of the x-axis at gantry angle $\beta = 0$ as in Figure 2. The offset of the isocenter w.r.t. to the conventional geometry, i.e. the isocenter-to-center distance, is $R_I = \text{SID}|\tan\tau|$.

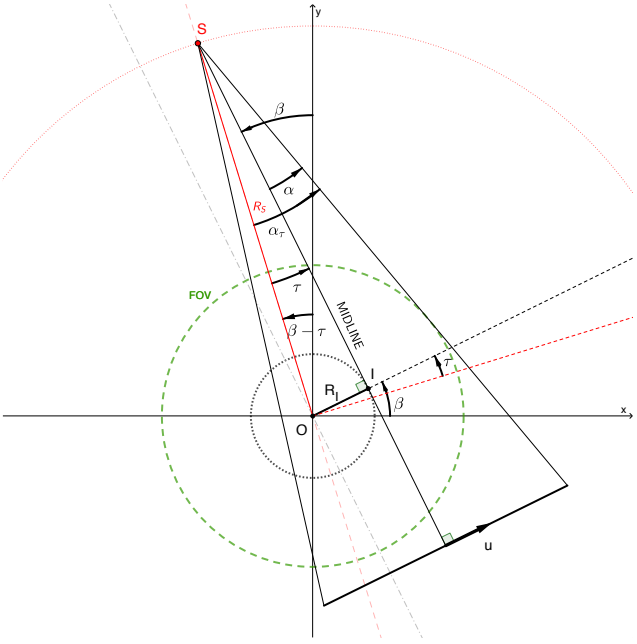


Fig. 2. Circular geometry defined as *displaced center-of-rotation* (DCoR) in the central plane. The FPD and source S are bound together while S travels the circle with radius R_S. The midline is offset from the center O of the source trajectory.

The source trajectory is still a circle, centered on O and defined by $S(\tau,\beta) = R_S(-\sin(\beta-\tau), \cos(\beta-\tau), 0)$, where $R_S = \sqrt{\text{SID}^2 + R_I^2}$ is the radius of the circle. The isocenter trajectory is also a circle of center O such that $I(\beta) = R_I(\cos\beta, \sin\beta, 0)$. The conventional geometry is when $\tau = 0$, i.e. when $R_I = 0$ and $R_S = \text{SID}$.

We note $\alpha_\tau = \alpha + |\tau|$ the large fan angle captured by the detector in the central plane with $\alpha = \arctan(N_u\Delta_u/2\ \text{SDD})$ the fan angle captured by the detector w.r.t. the midline (Figure 2), N_u the number of pixels of the detector in the transaxial direction and Δ_u their spacing. If $|\tau| < \alpha$, the source-to-center line hits the detector and the DCoR strategy could achieve exact reconstruction in the central slice of the FOV with radius $R_{\text{FOV}} = R_S \sin\alpha_\tau$. This is normally done with a full 360° scan but we propose to combine it with short scans.

B. Short scan

Let g_τ be the set of cone-beam projections acquired for a given DCoR geometry with an angle τ defined by the line integral

$$g_\tau(\beta,u,v) = \int_{\mathcal{L}(\tau,\beta,u,v)} f(x)\mathrm{d}x \tag{1}$$

where $\mathcal{L}(\tau,\beta,u,v)$ is the line defined at gantry angle β and tilt angle τ by the source position $S(\tau,\beta)$ and the position of the detector pixel with coordinates (u,v).

If the projections were untruncated with a full fan $\alpha_u \in [-\alpha_\tau, \alpha_\tau]$ with $\alpha_u = \tau + \arctan(u/\text{SDD})$, a short scan of $\pi + 2\alpha_\tau$ would be sufficient to reconstruct from divergent rays in the central plane using Parker pre-reconstruction weights [6]

$$w_P(\beta,u,v) = \begin{cases} \sin^2\left(\frac{\pi}{4}\frac{\beta}{\alpha_\tau - \alpha_u}\right) & \text{if } 0 \le \beta \le 2\alpha_\tau - 2\alpha_u, \\ \sin^2\left(\frac{\pi}{4}\frac{\pi + 2\alpha_\tau - \beta}{\alpha_\tau + \alpha_u}\right) & \text{if } \pi - 2\alpha_u \le \beta \le \pi + 2\alpha_\tau, \\ 1 & \text{otherwise.} \end{cases} \tag{2}$$

In [3], the authors provide an adaptation of the FDK algorithm [7] to a full scan DCoR geometry and we note $\mathcal{R}_\tau$ its application to a short scan:

$$f \simeq \mathcal{R}_\tau\left\{w_P g_\tau\right\}. \tag{3}$$

This short scan DCoR reconstruction is only exact in the central slice and approximate elsewhere due to the cone-beam artifact stemming from incomplete data with a circular source trajectory.

C. Complementary short scans

We note that for two scans with opposite offsets τ_1 and $\tau_2 = -\tau_1$, the source rotates on the same circle since $S(\tau_1,\beta) = S(\tau_2, \beta - 2\tau_1)$. We therefore acquire the corresponding cone-beam projections g_1 and g_2 for the same arc of source trajectory using the two sets of gantry angle $B_1 = [\tau_1, \pi + 2\alpha_{\tau_1} + \tau_1]$ and $B_2 = [\tau_2, \pi + 2\alpha_{\tau_2} + \tau_2]$ which measure the same line integrals in the central plane

$$g_1(\beta,u,0) = g_2(\beta - 2\tau_1, \text{SDD}\tan(\alpha_u - \tau_2), 0). \tag{4}$$

We refer to this composition of trajectories as *complementary displaced centers-of-rotation* (C-DCoR). The detectors are however in different positions with a small overlap (Figure 3).

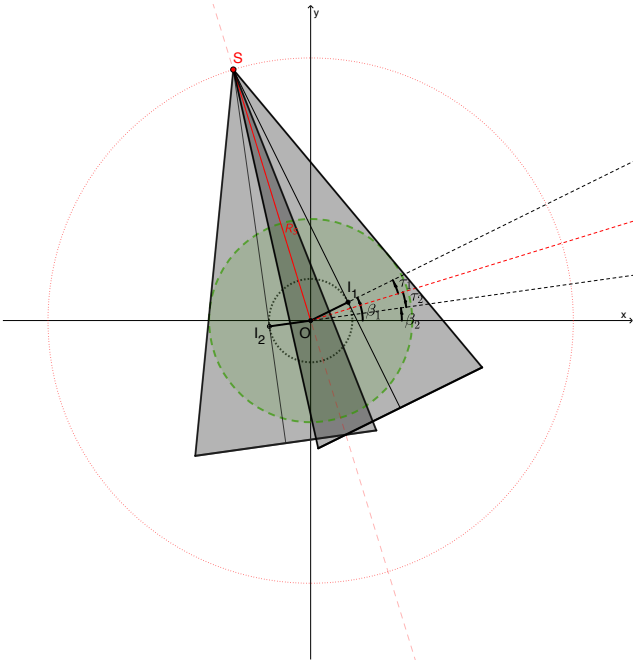


Fig. 3. Illustration of two projections of two complementary scans with τ_1 and $\tau_2 = -\tau_1$, with gantry angle $\beta_2 = \beta_1 - 2\tau_1$ such that the source is at the same position $S(\tau_1, \beta_1, 0) = S(\tau_2, \beta_2, 0)$ in the two arcs.

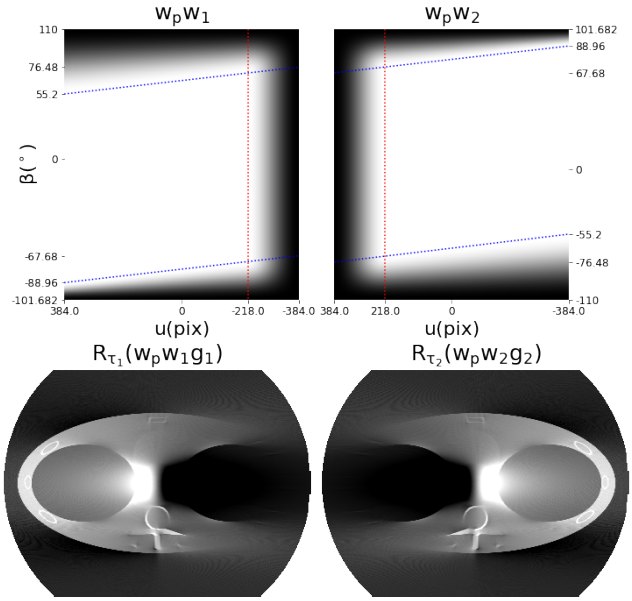


Fig. 4. Top row shows the weight maps $w_p w_1$ and $w_p w_2$ (grayscale range $[0, 1]$) for the two fan-beam sinograms of the central slice. Vertical dotted red lines indicate the limit for the offset detector weights w_1 and w_2 (Equation 5). Slanted blue dotted lines indicate the limits for Parker short scan weights (Equation 2). Bottom row shows partial reconstructions $\mathcal{R}_{\tau_1}(w_p w_1 g_1)$ and $\mathcal{R}_{\tau_2}(w_p w_2 g_2)$, the sum of which gives the axial slice shown at the left of the second row in Figure 5. The grayscale range is $[-740, 920]$ HU.

We follow the approach of [4] and [6] to manage redundancies with the weights

$$w_i(\beta, u, v) = \begin{cases} \frac{1}{2}\left(\operatorname{sign}\tau_i \sin\left(\frac{\pi\alpha_u}{2(\alpha - |\tau_i|)}\right) + 1\right) & \text{if } |\alpha_u| < \alpha - |\tau_i|, \\ 1 & \text{otherwise,} \end{cases} \tag{5}$$

which are derived from the displaced detector weights of [8]. We obtain the reconstruction formula

$$f \simeq \mathcal{R}_{\tau_1}\{w_P w_1 g_1\} + \mathcal{R}_{\tau_2}\{w_P w_2 g_2\}. \tag{6}$$

It can be shown that the reconstruction is exact in the central plane, as in (3). The result of (3) is however different in other planes but previous investigations have shown that the effect of the tilt is limited compared to the cone-beam artifacts in these planes [3].

III. Experiments

The proposed C-DCoR trajectories have been tested on simulated projections of a Forbild thorax phantom using RTK [9]. Geometrical parameters of the simulations mimick the CNAO scanner with SID = 1100 mm and SDD = 1600 mm. The detector has isometric spacing $\Delta_u = \Delta_v = 0.388$ mm and size $N_u \times N_v = 768 \times 1024$ pixels; we therefore have $\alpha = 5.32°$. We set $\tau_1 = -\tau_2 = 4.159°$ which results in $R_I = 80$ mm and $R_S = 1102.91$ mm. The same arc of the circle trajectory is covered with $B_1 \simeq [-102°, 110°]$ and $B_2 \simeq [-110°, 102°]$, which is larger than $180° + 2\alpha_\tau \simeq 199°$ and accounted for in the redundancy weights w_s (by replacing α_τ by half the scan range minus π in (2)).

The reconstructed DCoR scans have a FOV diameter of 363 mm instead of 204 mm for the conventional geometry. The Forbild phantom is isometrically rescaled 0.8 times such that the thorax ellipse semi-minor and semi-major axes are 80 mm and 160 mm and fit in the DCoR FOV. The phantom is centered on O and 400 projections are generated along each complementary scan. The reconstructed scans have size $400 \times 400 \times 400$ pixels with a resolution of $1 \times 1 \times 1$ mm.

IV. Results

The weight maps $w_P w_1$ and $w_P w_2$ and the partial reconstructions $\mathcal{R}_{\tau_1}\{w_P w_1 g_1\}$ and $\mathcal{R}_{\tau_2}\{w_P w_2 g_2\}$ are displayed in the central slice along with the corresponding weight maps in Figure 4. The resulting phantom reconstruction is qualitatively compared to the reference phantom in Figure 5. The images are masked to the FOV of the C-DCoR scan. In the latter, artifacts caused by strong attenuation in high density vertebrae along the z axis are visible in the sagittal plane. Intensity profiles are extracted from the middle of each slice along the colored lines in Figure 5 in the same figure. The difference between the reference image and the reconstructed image resulted in a mean absolute error (MAE) of $20.5(\pm 24.5)$ HU inside the FOV. Slightly worse results are obtained when simulating and reconstructing a complete 400 projections half fan acquisition with a 130 mm detector offset with a MAE of $27.6(\pm 25.5)$ HU (image not shown). Finally, we find equivalent results when simulating and reconstructing a conventional geometry acquisition with a large panel (1536×1024 pixels)

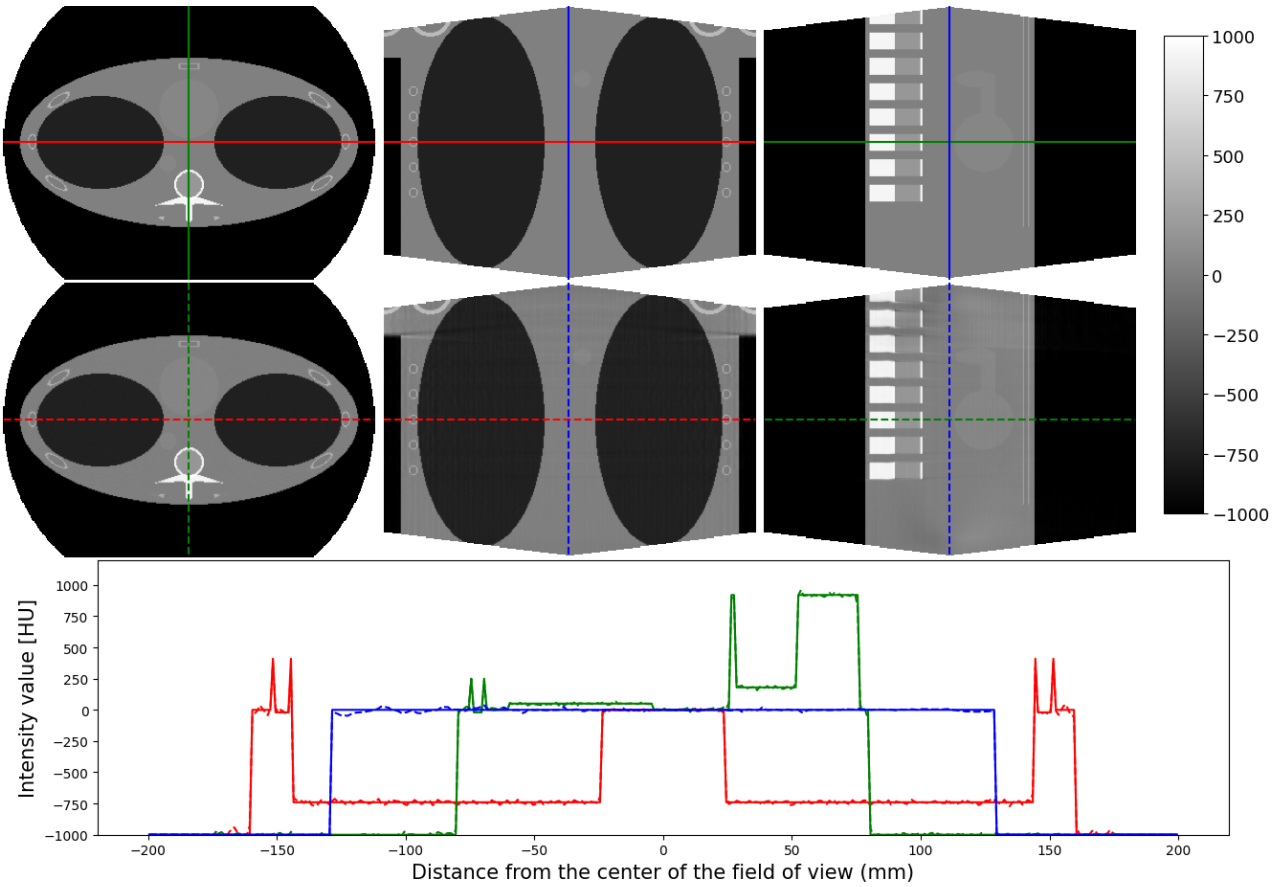


Fig. 5. Top row shows central slices (axial, coronal, sagittal) extracted from the Forbild reference volume, with grayscale in Hounsfield Units. Central row shows the same slices from a C-DCoR image with an offset $R_I = 80$ mm. Grayscale level and window are set to 0 and 2000 HU, where -740 HU is the value in the lung and 920 HU is the vertebrae density. Note that axial slices are cropped in the antero-posterior direction for optimal display. Colored lines indicate where the intensity profiles (bottom row) have been evaluated.

on the same source trajectory with a MAE of $20.6(\pm 24.0)$ HU (image not shown).

V. Discussion and Conclusions

We successfully developed a reconstruction algorithm to extend the FOV of the CNAO scanner. The C-DCoR acquisition can nearly double the FOV diameter. The method requires two short scans with a short movement of the isocenter and detector between the scans to ensure that the same source trajectory arc is imaged twice with different detector positions. With these arcs, we are able to produce CBCT images of similar quality as conventional short scan or displaced detector CBCT images. Most discrepancies in the reconstructed C-DCoR are caused by the phantom vertebrae and seem to be linked to the cone angle of incident X-rays. Physical effects such as scatter or geometrical fluctuations remain to be assessed on real data.

References

[1] G. Landry and C. H. Hua, "Current state and future applications of radiological image guidance for particle therapy," *Medical Physics*, vol. 45, no. 11, pp. e1086–e1095, 2018.

[2] G. Fattori, M. Riboldi, A. Pella, M. Peroni, P. Cerveri, M. Desplanques, G. Fontana, B. Tagaste, F. Valvo, R. Orecchia *et al.*, "Image guided particle therapy in CNAO room 2: implementation and clinical validation," *Physica Medica*, vol. 31, no. 1, pp. 9–15, 2015.

[3] S. Rit, R. Clackdoyle, P. Keuschnigg, and P. Steininger, "Filtered-backprojection reconstruction for a cone-beam computed tomography scanner with independent source and detector rotations," *Medical Physics*, vol. 43, no. 5, pp. 2344–2352, 2016.

[4] P. S. Cho, A. D. Rudd, and R. H. Johnson, "Cone-beam CT from width-truncated projections," *Computerized Medical Imaging and Graphics*, vol. 20, no. 1, pp. 49–57, 1996.

[5] G. T. Gullberg, C. R. Crawford, and B. M. Tsui, "Reconstruction algorithm for fan beam with a displaced center-of-rotation," *IEEE Transactions on Medical Imaging*, vol. 5, no. 1, pp. 23–29, 1986.

[6] D. L. Parker, "Optimal short scan convolution reconstruction for fan beam CT," *Medical Physics*, vol. 9, no. 2, pp. 254–257, 1982.

[7] L. A. Feldkamp, L. C. Davis, and J. W. Kress, "Practical cone-beam algorithm," *Journal of the Optical Society of America A*, vol. 1, no. 6, pp. 612–619, 1984.

[8] G. Wang, "X-ray micro-CT with a displaced detector array," *Medical Physics*, vol. 29, no. 7, pp. 1634–1636, 2002.

[9] S. Rit, M. V. Oliva, S. Brousmiche, R. Labarbe, D. Sarrut, and G. C. Sharp, "The Reconstruction Toolkit (RTK), an open-source cone-beam CT reconstruction toolkit based on the Insight Toolkit (ITK)," in *Journal of Physics: Conference Series*, vol. 489, no. 1. IOP Publishing, 2014, p. 012079.

X-ray CT Data Completeness Condition for Sets of Arbitrary Projections

Gabriel Herl, Andreas Maier, *Senior Member, IEEE,* and Simon Zabler

***Abstract*—X-ray tomography reconstruction requires a set of projections that provides sufficient information for the examined region. Commonly, to ensure mathematically complete reconstruction, first, a continuous curve (trajectory) that fulfils the Tuy conditions is chosen. Second, this curve is sampled based on the Nyquist-Shannon sampling theorem. This two-step approach is efficient for most standard X-ray tomography scanning scenarios. For agile X-ray tomography systems, e.g. robot-supported computed tomography systems, choosing a set of projections based on a continuous curve is often not useful. Instead, sets of projections from arbitrary views might be necessary.**

This work combines the Tuy-Smith condition with conclusions from the Nyquist-Shannon sampling theorem. In particular, the maximal pixel size and a requirement for the arrangement of projections are formulated depending on the smallest relevant object feature and integrated into the Tuy-Smith condition. We derive a comprehensive condition for data completeness that can assess the completeness of any set of arbitrary projections, e.g. for complex scanning scenarios with robot-supported X-ray tomography systems.

***Index Terms*—Data completeness condition, arbitrary scan geometry, robot-supported CT, Tuy conditions, Nyquist-Shannon sampling theorem.**

I. Introduction

X-RAY computed tomography (CT) allows digitisation of three-dimensional inner and outer structures for many applications in medicine [1] and in industry [2].

In a standard industrial CT, the examined object is placed on a turntable between an X-ray source and a detector. By rotating the object, projection views from a circular trajectory around the object are generated. In addition to these classical systems, twin robotic CT systems have been developed where the source and the detector are mounted on individual robots. This setup allows agile movement of source and detector in order to create projections from arbitrary views around the object. This agility of robot-supported CT systems enables new, more complex trajectories which can be utilised in numerous ways, e.g. to scan large-scale objects [3], to reduce metal artefacts [4], [5] and to reduce the scan time [6].

However, this agility of robot-supported CT systems complicates the CT scan process. In order to enable a true, mathematically complete reconstruction, it has to be ensured that the available projections generate sufficient information. In the standard CT process, first, the CT user chooses a continuous curve (commonly a circle or a helix) around a region of interest. Second, by choosing the number of projections, this continuous curve is sampled into individual, equidistantly placed views for the generation of individual projections.

A two-step approach is performed to ensure that the resulting set of projections provides sufficient information for mathematically complete reconstruction. First, the Tuy-Smith condition [7] assesses whether the continuous trajectory can provide enough information. Second, the sampling is validated based on the Nyquist-Shannon sampling theorem. This two-step process is efficient and sufficient for most scenarios for standard CT systems. However, especially when using agile CT systems, this process has weaknesses. The two-step approach requires a continuous curve as a base trajectory. However, in many scenarios, the requirement of a continuous curve is impractical. Fig. 1 shows an example from [4] of two sets of views for a region of interest scan of a defect in a motorcycle. A continuous curve would not be a useful basis for both sets of views.

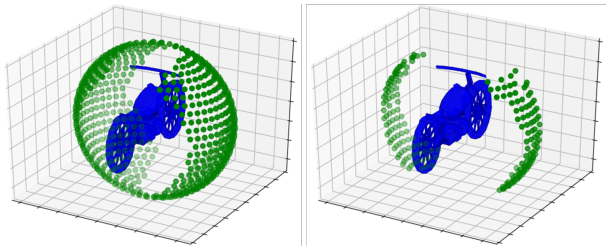


Fig. 1: Visualisation of two sets of views on a motorcycle. The green dots represent individual source positions.

This work extends the Tuy-Smith condition with the Nyquist-Shannon sampling theorem to create a single condition that indicates data completeness for a mathematically complete reconstruction. For the first time, this condition allows the assessment of data completeness for arbitrary sets of views.

II. State of the Art of CT Data Completeness

The X-ray attenuation process can be depicted as the Radon transform which maps every hyperplane of the spatial domain to a point in the Radon space. Let $\mathbf{f}$ be the density function of the examined object, $\mathbf{u}$ be the normal vector of a hyperplane and s be the distance of this plane to the origin. The Radon transform can be written as

$$\mathbf{Rf}(\mathbf{u},\mathbf{s}) = \int_{\mathbf{x}\cdot\mathbf{u}=s} \mathbf{f}(\mathbf{x})dx. \tag{1}$$

In 1917 [8], Radon proved that if $\mathbf{f}(\mathbf{u},\mathbf{s})$ is continuous, there exists a unique inverse. Therefore if the Radon space

G. Herl and S. Zabler are with the Deggendorf Institute of Technology, e-mail: gabriel.herl@th-deg.de.

A. Maier is with the Department of Pattern Recognition, Friedrich-Alexander University Erlangen-Nürnberg.

Manuscript received ??; revised ??.

is continuous and known, complete CT reconstruction can be ensured. However, it is not trivial whether a specific trajectory or a specific set of arbitrary X-ray projections can be applied to measure the Radon space sufficiently.

A. Trajectory Requirements

In 1983 [9], Tuy published data completeness conditions for continuous curves that ensure mathematically correct reconstruction. Following Tuy, a curve is a continuous function $\Phi : \Lambda \to \mathbb{R}^3$ where Λ is an interval in $\mathbb{R}$. Let $\mathbb{S} \subset \mathbb{R}^3$ be the unit sphere and $\Omega \subset \mathbb{R}^3$ be a compact region that contains the complete object with density function $\mathbf{f}$.

Tuy implicitly used several assumptions about the measuring and reconstruction processes to ensure mathematically correct reconstruction. This includes the following two (unrealistic) assumptions:

1) **Detector resolution assumption:** Tuy assumes that the exact position of impact of each measured photon is known. This corresponds to a detector with infinitely small pixels.
2) **Source assumption:** Tuy assumes that X-rays are emitted at every position of the trajectory of the source. This assumption corresponds to an infinite number of projections on the trajectory.

Tuy conditions
Using assumptions 1 and 2, an object in region Ω can be reconstructed from projection data generated on a curve Φ if the following three conditions are valid:

1) *The curve Φ is outside of the region Ω.*
2) *The curve Φ is bounded, continuous and almost everywhere differentiable.*
3) *For all $(\mathbf{x}, \mathbf{u})$ in $\Omega \times \mathbb{S}$, there exists $\lambda \in \Lambda$, such that $\mathbf{x}^\top \mathbf{u} = \Phi(\lambda)^\top \mathbf{u}$ and $\Phi'(\lambda)^\top \mathbf{u} \neq 0$.*

In 1985 [7], Smith proved that Tuy's third condition is sufficient:

Tuy-Smith condition
Using assumptions 1 and 2, an object in region Ω can be reconstructed from projection data generated on a curve Φ if, for all $(\mathbf{x}, \mathbf{u})$ in $\Omega \times \mathbb{S}$, there exists $\lambda \in \Lambda$, such that $\mathbf{x}^\top \mathbf{u} = \Phi(\lambda)^\top \mathbf{u}$.

Descriptively, the Tuy-Smith condition states that every plane through the region of interest must intersect the source trajectory. As every point in the Radon space corresponds to one plane in the spatial domain, the Tuy-Smith condition ensures a full sampling of the Radon space.

B. Sampling Requirements

As both assumptions, the detector resolution assumption and the source assumption of Section II-A, are impossible in practice, the projections as well as the continuous trajectory need to be sampled. To nevertheless ensure sufficient information, the Nyquist-Shannon sampling theorem can be applied.

A detailed derivation of the maximal pixel size and minimal number of projections is presented by Buzug [10]. Let k be the minimal magnification factor and f_{min} be the smallest relevant feature. Based on the Nyquist-Shannon sampling theorem, the maximal pixel size $\triangle\xi$ is given by

$$\triangle\xi < \frac{k}{2} f_{\text{min}} \tag{2}$$

Let r be the radius of the measuring field. Following Buzug [10], the maximal angular gap $\triangle\gamma$ between projections is given by

$$\triangle\gamma < \frac{f_{\text{min}}}{2r}. \tag{3}$$

As an example for parallel-beam geometry, for an equiangular sampling of a semi-circular trajectory, the minimum number of projections follows by

$$n = \frac{\pi}{\triangle\gamma}. \tag{4}$$

III. New data completeness condition for sets of arbitrary views

To directly assess the data completeness of a set of arbitrary projections, we combine the Tuy-Smith condition with the presented conclusions of the Nyquist-Shannon sampling theorem. This means that, first, the pixel size and, second, the maximal angular gap of projections are integrated in the Tuy-Smith condition.

The maximal pixel size can directly be applied to create a realistic new assumption:

1b. **Adapted detector resolution assumption:** The detector has a maximal pixel size

$$\triangle\xi < \frac{k}{2} f_{\text{min}}$$

based on the smallest magnification k of any of the used projections and the size f_{min} of the smallest feature that should be detectable.

To integrate the maximal angular gap into the Tuy-Smith condition, we extend the estimations of Maier *et al.* [11]. We assume parallel-beam scanning geometry to allow more intuitive and straightforward phrasing. The conclusions remain for cone-beam CT. Let $\tilde{\mathbf{u}} \in \mathbb{S}$ be a normal vector that represents an arbitrary plane in Radon space and $\mathbf{D}$ the set of the normal vectors of all measured planes in Radon space. According to the Nyquist-Shannon sampling theorem, the angular distance between neighbouring planes in Radon space does not have to be zero, but only must be smaller than the specified maximum angular gap $\triangle\gamma$ of (3). This means, for any possible plane in Radon space, there has to be a measured plane that is tilted less than $\triangle\gamma$. The cosine angle between two vectors equals the inner product of the corresponding unit vectors. Thus, this condition can be written as

$$\forall \tilde{\mathbf{u}} \in \mathbb{S}\ \exists \mathbf{d} \in \mathbf{D}\ :\ |\cos^{-1}(\mathbf{d}^\top \tilde{\mathbf{u}})| \leq \triangle\gamma. \tag{5}$$

Let $\mathbf{u} \in \mathbb{S}$ be a vector perpendicular to vector $\tilde{\mathbf{u}}$. Vector $\tilde{\mathbf{u}}$ being tilted less than $\triangle\gamma$ according to vector $\mathbf{d}$ equals vector $\mathbf{u}$

being perpendicular to vector $\mathbf{d}$ apart from an angle $\triangle\gamma$. Two vectors are perpendicular apart from an angle $\triangle\gamma$ if $|\mathbf{d}^\top \mathbf{u}| \leq \sin(\triangle\gamma)$. Equation (5) thus is equivalent to

$$\forall \mathbf{u} \in \mathbb{S}\ \exists \mathbf{d} \in \mathbf{D}\ :\ |\mathbf{d}^\top \mathbf{u}| \leq \sin(\triangle\gamma). \tag{6}$$

The normal vector of a measured plane in Radon space corresponds to the directional vector of the projection that measured this plane in Radon space. Let $\mathbf{L}$ be the set of all source positions of a set of projections and $\mathbf{x} \in \Omega$ any point in the region of interest. Defining directional vectors of projections by $\mathbf{d}_{\mathrm{x,l}} := \frac{\mathbf{x}-\mathbf{l}}{||\mathbf{x}-\mathbf{l}||_2}$, Equation (6) can be written as

$$\forall (\mathbf{x}, \mathbf{u}) \in \mathbf{\Omega} \times \mathbb{S}\ \exists \mathbf{l} \in \mathbf{L}\ :\ |\mathbf{d}_{\mathrm{x,l}}^\top \mathbf{u}| \leq \sin(\triangle\gamma). \tag{7}$$

This equation ensures that sufficient projections for a complete reconstruction are available. In total, by integrating the conclusions of the Nyquist-Shannon sampling theorem (Assumption 1b and Equation (6)) into the Tuy-Smith condition, the following combined condition can be derived:

Data completeness condition for sets of arbitrary projections
An object in region $\mathbf{\Omega}$ can be reconstructed based on projections with corresponding source positions $\mathbf{L}$ if

1) *the maximal pixel size of the projections is $\triangle\xi < \frac{k}{2} f_{min}$ depending on the smallest magnification factor k and the smallest relevant feature f_{min},*
2) *for all $(\mathbf{x}, \mathbf{u}) \in \mathbf{\Omega} \times \mathbb{S}$, there exists a projection with corresponding source location $\mathbf{l} \in \mathbf{L}$ such that*

$$|\mathbf{d}_{x,l}^\top \mathbf{u}| \leq \sin(\triangle\gamma)$$

with $\triangle\gamma = \frac{f_{min}}{2r}$ as the maximal sufficient angular sampling rate depending on the radius r of the measuring field.

IV. Examples

As example, let $f_{\min} := 0.03$ cm be the smallest relevant feature of an examined object with maximal radius $r := 1cm$ and $k := 10$ be the magnification (due to cone-beam CT). Using the presented equations, we can calculate the maximal detector pixel size $\triangle\xi$ and the maximal angular gap $\triangle\gamma$ for data complete reconstruction:

$$\triangle\xi < \frac{k}{2} f_{\min} = 0.15 \text{ cm} \tag{8}$$

$$\triangle\gamma < \frac{f_{\min}}{2r} = 0.015. \tag{9}$$

This means, a set of projections provides complete data if, first, the pixel size is smaller than 0.15 cm and, second, for all positions $\mathbf{x}$ in the region of interest and possible vectors $\mathbf{u}$ through the region of interest, there exists a projection with a normal vector $d_{x,l}$ so that the maximal angular gap $\gamma_{\max} = \mathbf{d}_{x,l}^\top \mathbf{u} \leq \sin(0.015) \approx 0.015$.

For the following examples, we chose a sufficiently small pixel size of 0.12 cm and tested different sets of views. In the first example, we chose to reconstruct a spherical object with concentric spherical shells in order to visualise cone-beam and aliasing artefacts. Fig. 2, 3 and 4 show examples of different sets of views and slices of the corresponding reconstructions. With each a maximal angular gap of $\gamma_{\max} = 0.308$ at the bottom and the top of the sphere, both circular trajectories of Fig. 2 and 3 do not fulfil the presented data completeness conditions. Cone-beam artefacts appear at the bottom and the top due to a lacking sampling. Aliasing appears in Fig. 2 due to too few projections. Fig. 4 shows a set of views with 300 arbitrary projections that does fulfil the presented condition with a maximal angular gap of $\gamma_{\max} = 0.013$. It contains no aliasing or cone-beam artefacts as the corresponding Radon space is sampled sufficiently.

Fig. 5 shows two additional sets of views for a plastic test specimen that all fulfil the presented data completeness condition. The right image contains highly attenuating metal blocks. This example demonstrates that our condition can be applied to assess data completeness in scenarios that require complex sets of views.

V. Conclusion

We have presented a CT data completeness condition that can be used to assess the completeness of any set of projections. This condition is not based on continuous curves, but can be applied directly to assess the completeness of data for any arbitrary set of projections. Thereby, this work introduces a method for evaluating sets of projections even for complex scanning scenarios, e.g. for robotic CT systems and scenarios with strongly attenuating components and spatial restrictions.

References

[1] G. D. Rubin, "Computed tomography: revolutionizing the practice of medicine for 40 years," *Radiology*, vol. 273, no. 2S, pp. S45–S74, 2014.
[2] S. Zabler, M. Maisl, P. Hornberger, J. Hiller, C. Fella, and R. Hanke, "X-ray imaging and computed tomography for engineering applications," *tm-Technisches Messen*, vol. 88, no. 4, pp. 211–226, 2021.
[3] W. Holub, F. Brunner, and T. Schön, "Roboct-application for in-situ inspection of join technologies of large scale objects," in *International Symposium on Digital Industrial Radiology and Computed Tomography*, 2019.
[4] G. Herl, J. Hiller, and A. Maier, "Scanning trajectory optimisation using a quantitative tuybased local quality estimation for robot-based x-ray computed tomography," *Nondestructive Testing and Evaluation*, vol. 35, no. 3, pp. 287–303, 2020.
[5] G. Herl, J. Hiller, M. Thies, J.-N. Zaech, M. Unberath, and A. Maier, "Task-specific trajectory optimisation for twin-robotic x-ray tomography," *IEEE Transactions on Computational Imaging*, 2021.
[6] F. Bauer, M. Goldammer, and C. U. Grosse, "Scan time reduction by fewer projections-an approach for part-specific acquisition trajectories," in *20th World Conference on Non-Destructive Testing*, 2020.
[7] B. D. Smith, "Image reconstruction from cone-beam projections: necessary and sufficient conditions and reconstruction methods," *IEEE transactions on medical imaging*, vol. 4, no. 1, pp. 14–25, 1985.
[8] J. Radon, "Über die bestimmung von funktionen längs gewisser mannigfaltigkeiten. sächsische gesellschaft der wissenschaften math," *Phys. Klasse, Leipzig*, vol. 69, pp. 262–277, 1917.
[9] H. K. Tuy, "An inversion formula for cone-beam reconstruction," *SIAM Journal on Applied Mathematics*, vol. 43, no. 3, pp. 546–552, 1983.
[10] T. M. Buzug, *Einführung in die Computertomographie: mathematisch-physikalische Grundlagen der Bildrekonstruktion*. Springer-Verlag, 2003.
[11] A. Maier, P. Kugler, G. Lauritsch, and J. Hornegger, "Discrete estimation of data completeness for 3d scan trajectories with detector offset," in *Bildverarbeitung für die Medizin 2015*. Springer, 2015, pp. 47–52.

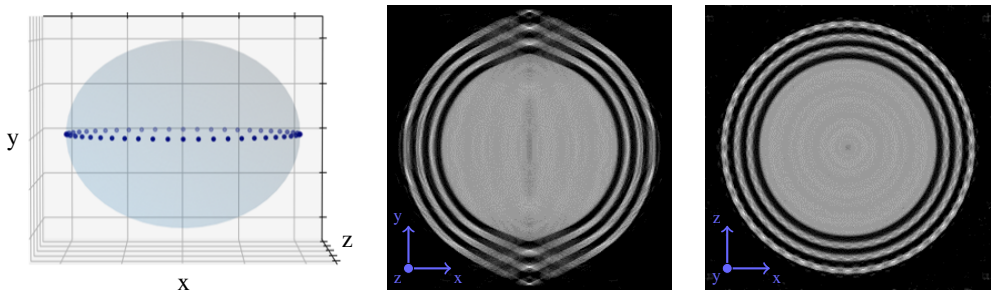


Fig. 2: Circular trajectory with 50 projections: Visualisation of the source positions and two reconstruction slices of a spherical object. This set of projections is not data complete due to missing information outside of the centre plane and too few projections.

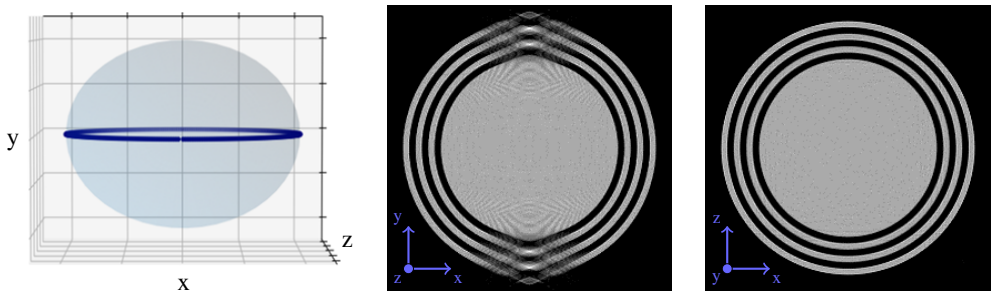


Fig. 3: Circular trajectory with 300 projections: Visualisation of the source positions and two reconstruction slices of a spherical object. This set of projections is not data complete due to missing information outside of the centre plane.

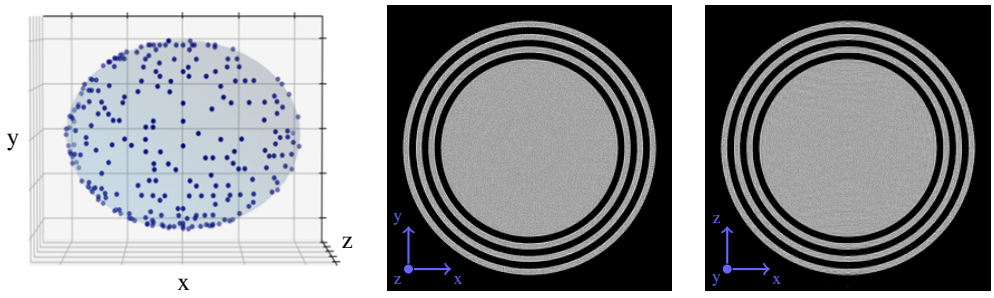


Fig. 4: Data complete set of 300 arbitrary projections: Visualisation of the source positions and two reconstruction slices of a spherical object. This set of projections is data complete.

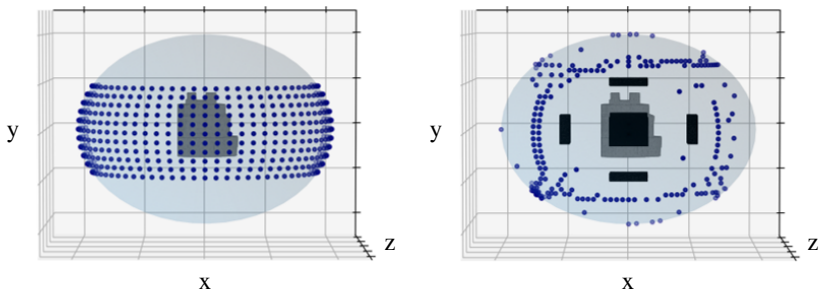


Fig. 5: Two data complete sets of views: Visualisation of the source positions and a test specimen. The right scenario also contains highly attenuating metal blocks. The presented set of projections mostly avoids X-rays through these blocks, generating complete data for the reconstruction of the test specimen without metal disturbance.

First results on Compton camera system used for X-ray fluorescence computed tomography

Chuanpeng Wu and Liang Li, *Senior Member, IEEE*

***Abstract*—Because of the ability to present molecular and functional information in organisms, nuclear medical imaging (NMI) is attracting more and more attention. Among NMI modalities, X-ray fluorescence computed tomography (XFCT) has the advantage that the tracers used in XFCT are not spontaneously decayed. The synthesis, storage of contrast agents is more convenient, the price of XFCT is much lower as well. However, XFCT usually has mechanical collimation to tell the incident photon direction, which results in the reduction of the detection efficiency. The Compton camera is an imaging modality, which does not need mechanical collimators in its structure, which makes Compton cameras have high detection efficiency. Therefore, it is a great idea to use Compton camera-based imaging systems to realize X-ray fluorescence (XF) imaging. In this work, the first XFCC imaging system in the laboratory environment is established, which consists of a 150keV X-ray tube and a single-layer Compton camera system based on the Timepix3 photon-counting detector (PCD). The element Gd (43keV) is used as the XF element. The first imaging reconstruction results of the XFCC system are represented.**

***Index Terms*—Compton camera, image reconstruction, X-ray fluorescence computed tomography, Timepix3, photon counting detector**

I. Introduction

BECAUSE of the ability to present molecular and functional information in organisms, nuclear medical imaging (NMI) is attracting more and more attention. Compared with traditional NMI modalities like single photon emission computed tomography (SPECT) and positron emission tomography (PET), the X-ray fluorescence computed tomography (XFCT) has the advantage that the tracers used in XFCT are not spontaneously decayed[1, 2]. The synthesis, storage of contrast agents is more convenient, the price of XFCT is much lower as well. SPECT and XFCT, which can use a variety of traces, usually have mechanical collimation to tell the incident photon direction, which results in the reduction of the detection efficiency.

The Compton camera (CC) is an imaging modality that does not need mechanical collimators in its structure[3]. This makes Compton cameras have high detection efficiency. Compton camera has been widely used in astronomical detection, environmental radiation detection, proton therapy, and medical imaging[4-6]. So it is a great idea to use Compton camera-based imaging systems to realize X-ray fluorescence (XF) imaging. However, the imaging of the X-ray fluorescence Compton camera (XFCC) is not easy. A large number of scattered photons enter the detector together with X-ray fluorescence photons, and it is tough to distinguish them. Besides, the commonly used elements of XF imaging are usually below 100keV. The high-resolution reconstruction of Compton cameras is difficult in this energy range.

There are few studies about Compton cameras for XF imaging. Vernekohl et al. carry out the Monte Carlo simulation with the incidence of ideal 82keV monochromatic X-rays and gold nanoparticle (GNPs) solution as the XF element[7]. In this work, the first XFCC imaging system is established in the laboratory environment, which consists of a 150keV X-ray tube and a single-layer Timepix3 detector Compton camera. The element Gd (43keV) is used as the XF element. The first imaging reconstruction results of our XFCC system are represented.

This work was partially funded by grants from NNSFC 12027811 and the National Key Research and Development Program of China 2018YFC0115502.

Chuanpeng Wu, and Liang Li are with the Department of Engineering Physics, Tsinghua University, Beijing, 100084, China & National Engineering Laboratory for Dangerous Articles and Explosives Detection Technologies.

Corresponding Author email: lliang@tsinghua.edu.cn

II. Methods

A. Compton camera

In Compton camera imaging, photons are incident on the first layer of the detector and interact with the detector atoms. The scattering detector will record the interaction position and the deposition energy of recoil electrons. The scattered photons are then emitted from the first lie detector and absorbed by the second layer detector. The absorption detector records the absorption position and the deposited energy. The Compton scattering angle θ is according to the Compton effect formula:

$$cos\theta = 1 - m_e c^2 \left(\frac{1}{E_2} - \frac{1}{E_0}\right) \quad (1)$$

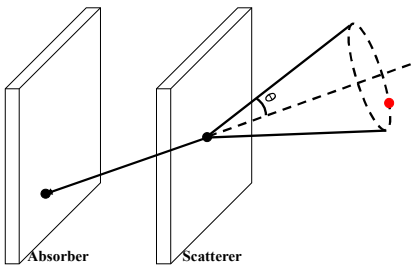


Fig. 1. The principle of the Compton camera.

where E_0 is the energy of the incident photon, E_2 is the energy of the scatted photon, θ is the Compton scattering angle, m_e is the rest mass of the electron.

After calculating the scattering angle θ, we are still not sure where the specific incident direction of the incident photon is. But we can build a cone surface with r_1r_2 as the axis and θ as the cone angle, on which we can find the incident direction. When enough Compton scattering events are detected, each event can be inversely calculated to a cone surface. The intersection of these cone surfaces is theoretically the spatial position of the radioactive source, which is the reconstruction principle of the simple back-projection (SBP) algorithm.

B. LM-MLEM reconstruction algorithm

In the SBP algorithm with cone intersection, each incident photon event will be back-projected to all points on a cone surface. So the spatial positions where the source is not located also obtain the weight by mistake, which makes the reconstruction result inaccurate.

The most commonly used high-resolution reconstruction algorithm is the list-mode maximum likelihood expectation maximization (LM-MLEM) algorithm. The iteration formula of LM-MLEM is as follows[8]:

$$\lambda_j^{(l+1)} = \frac{\lambda_j^{(l)}}{s_j} \sum t_{ij} \frac{1}{p_i^{(l)}}, \tag{2}$$

$$with\ p_i^{(l)} = \sum_{k=1}^{M} t_{ik} \lambda_k^{(l)} \tag{3}$$

where $\lambda_j^{(l)}$ is the reconstruction image. t_{ij} is the element of the system matrix, indexed on the events i and the voxels j. s_j is the element of the sensitivity matrix. The system matrix is obtained by our proposed numerical calculation method with detector resolution and Doppler broadening correction. While the sensitivity matrix is obtained by the Monte Carlo simulation method.

C. Timepix3 data processing

The Compton camera system established by us is based on the single-layer Timepix3 detector[9]. Timepix3 is an advanced photon-counting detector (PCD) with high spatial resolution, high time resolution, and fast readout speed.

The raw data of Timepix3 is the time of arrival (ToA) and time-over-threshold (ToT), which is corresponding to the time and energy information of the interaction event after calibration. The array of 256*256 pixels can give 2-dimensional position information. After time clustering and spatial clustering, the data belonging to the same incident photon can be clustered together, and the charge sharing effect is eliminated.

The information of the z-direction can be obtained by the carrier drift time, which is calibrated by the muon track. Due to the constant carrier drift velocity, the depth of the interaction is proportional to the carrier drift time. After calibration of the carrier drift velocity, the depth difference of the scattering position and the absorbing position can be calculated from the drift time difference. So that all information needed for the Compton camera reconstruction is obtained, including the energy information and the three-dimensional position information. Then coincidence operations are used for the detected interaction signals. And the valid Compton scattering events in the single-layer Timepix3 detector are selected for the SBP and LM-MLEM reconstruction of the proposed Compton camera.

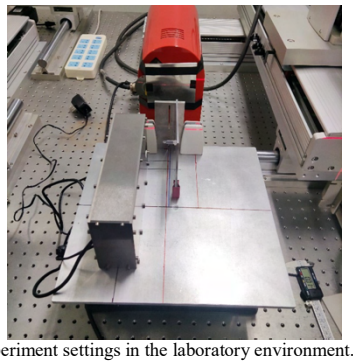


Fig. 2. The experiment settings in the laboratory environment.

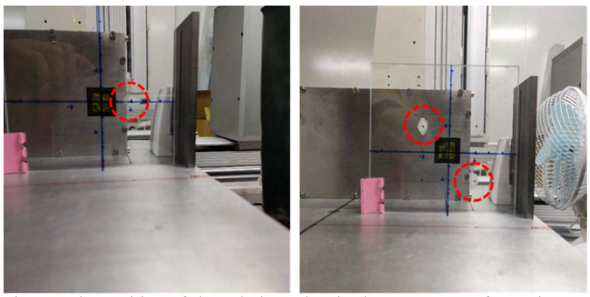


Fig. 3. The position of the solution tubes in the two group of experiments respectively.

III. Experiment Settings

The experimental design is shown in Figure 2. The incident X-ray is emitted by a 150kV microfocus X-ray tube with a copper filter. The X-ray is collimated to the fan beam by a tungsten slit and then and then irradiated on the Gd solution. The Timepix3 detector is placed at the location of 90° from the X-ray incident direction.

In all experiments, the X-ray tube is set to 150keV and 0.5mA. The concentration of the Gd solution is 100mg/ml. The Timepix3 detector has 256×256 pixels. The size of the detector unit is equal to 55 μm and the size of the detector is 14.08× 14.08 mm^2. The detector is with a bias voltage equal to 200V and an energy threshold equal to 3.02keV. The detector temperature is maintained at 26-28°C to keep the carrier drift velocity stable.

There are two groups of experiments, as shown in Figure3.

One has a single solution pipe, and another has two solution pipes. The diameter of all the solution pipes is equal to 5mm. The distance between the detector plane and the position where X-ray fluorescence photons emit is 50mm. For each experiment, the acquisition time is 100s. The imaging space is set to 60×60 ×20 pixels and the pixel size is equal to 4mm. Therefore, the field of view (FOV) is 240×240×80 mm^3. The central slice of the FOV is 50mm far away from the detector plane, and the center of the FOV is on the z-axis, which is passing through the center point of the detector and perpendicular to the detector plane.

IV. Results and Discussions

After data acquisition of the Compton camera system, the Compton scattering events detected by the Timepix3 detector

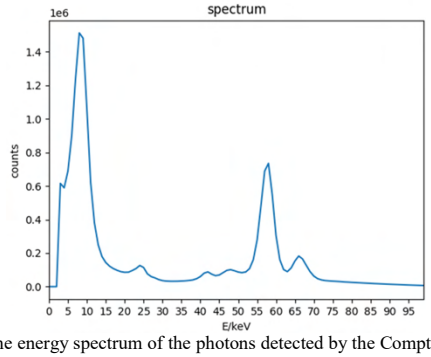


Fig. 4. The energy spectrum of the photons detected by the Compton camera system.

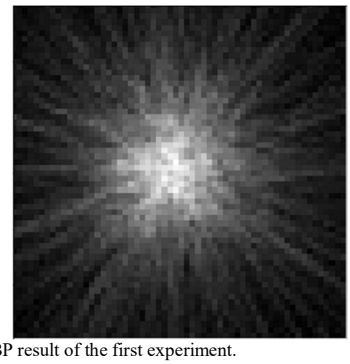


Fig. 5. The SBP result of the first experiment.

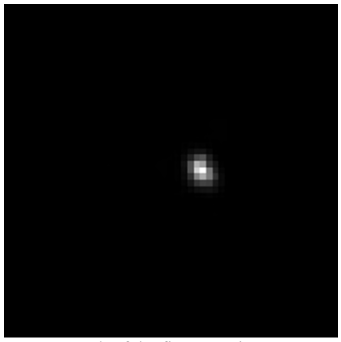


Fig. 6. The LM-MLEM result of the first experiment.

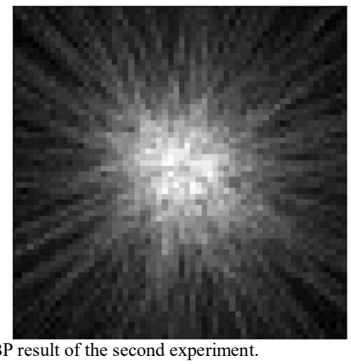


Fig. 7. The SBP result of the second experiment.

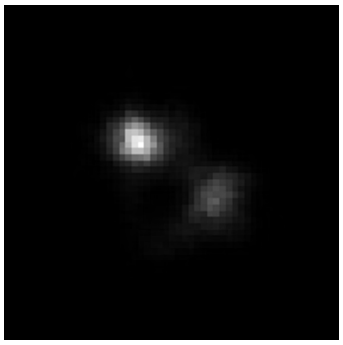


Fig. 8. The LM-MLEM result of the second experiment.

are selected. These events are used for SBP reconstruction and LM-MLEM reconstruction. The iteration number is 50 and the initial image of the LM-MLEM iteration is the SBP image.

Figure 4 shows the energy spectrum of all photons detected by the XFCC system. The spectrum peak below 20keV represents the photons that incompletely deposit their energy. The two peaks in the range of 40~50keV are the K_α (42.280 keV & 42.983 keV) peak and K_β (48.718 keV & 49.961 keV) peak of Gd, respectively. The peak around 60keV is the K_α peak of W, while the peak around 65keV is K_β. This is because many photons go through the tungsten slit and excite these characteristic gammas. From the spectrum, we can know that the proportion of the X-ray fluorescence photons is relatively low, which needs more improvement in our further works.

The reconstruction results of the first experiment prove the correct imaging reconstruction of the XFCC system. The solution pipe is placed at (20,0), which is the out center of the FOV. It is obvious that the SBP algorithm cannot realize a high-resolution reconstruction of the XFCC system at all. The single solution pipe can be clearly distinguished with the LM-MLEM algorithm. The spatial resolution is about 10mm for the single solution pipe.

For the second experiment, the reconstruction of the double solution pipes is also completed with LM-MLEM and there are two circle areas in the result image. However, the voxel value of the two areas is not very consistent, which may be due to the

imperfection of system geometric correction and the lack of Compton events. Besides, the reconstruction resolution becomes worse than the single pipe experiment. Multi-source reconstruction is a classic problem in the field of Compton cameras, which is to be overcome in the future. Besides, it is important to help increase the proportion of X-ray fluorescence photons.

V. Conclusions

In this work, we proposed the first XFCC imaging system in the laboratory environment, which consists of a 150keV X-ray tube and a single-layer Timepix3 detector Compton camera. The element Gd (43keV) is used as the XF element. The first results of the XFCC system are represented. Experiments with both single solution pipes and double solution pipes are carried out, and the LM-MLEM reconstruction result images illustrate the spatial resolution of about 10mm. For the double solution pipes experiment, the interaction between multi-sources makes the result not satisfactory enough. Besides, the problem of increasing the proportion of XF photons is still another important problem to be overcome in the future.

References

[1] L. Li, S. Zhang, R. Li, and Z. Chen, "Full-field fan-beam x-ray fluorescence computed tomography with a conventional x-ray tube and photon-counting detectors for fast nanoparticle bioimaging," Optical Engineering, vol. 56, no. 4, Apr 2017, Art. no. 043106.

[2] S. Zhang, L. Li, J. Chen, Z. Chen, W. Zhang, and H. Lu, "Quantitative Imaging of Gd Nanoparticles in Mice Using Benchtop Cone-Beam X-ray Fluorescence Computed Tomography System," International Journal of Molecular Sciences, vol. 20, no. 9, May 1 2019, Art. no. 2315.

[3] R. W. Todd, J. M. Nightingale, and D. B. Everett, "Proposed gamma camera," Nature, vol. 251, no. 5471, pp. 132-134, 1974 1974.

[4] V. Schonfelder, A. Hirner, and K. Schneider, "Telescope for SOFT Gamma-Ray Astronomy," Nuclear Instruments & Methods, vol. 107, no. 2, pp. 385-394, 1973 1973.

[5] S. Aldawood et al., "Development of a Compton camera for prompt-gamma medical imaging," Radiation Physics & Chemistry, p. S0969806X17300981, 2017.

[6] Y. F. Zhu and Z. He, "Performance of Larger-Volume 40 x 40 x 10- and 40 x 40 x 15-mm(3) CdZnTe Detectors," (in English), Ieee Transactions on Nuclear Science, Article vol. 68, no. 2, pp. 250-255, Feb 2021.

[7] D. Vernekohl, M. Ahmad, G. Chinn, and L. Xing, "Feasibility study of Compton cameras for x-ray fluorescence computed tomography with humans," Physics in Medicine and Biology, vol. 61, no. 24, pp. 8521-8540, Dec 21 2016.

[8] S. J. Wilderman, N. H. Clinthorne, J. A. Fessler, and W. L. Rogers, "List-mode maximum likelihood reconstruction of Compton scatter camera images in nuclear medicine," (in English), 1998 IEEE Nuclear Science Symposium Conference Record. 1998 IEEE Nuclear Science Symposium and Medical Imaging Conference (Cat. No.98CH36255), Conference Paper pp. 1716-20 vol.3, 1998 1998.

[9] D. Turecek, J. Jakubek, E. Trojanova, and L. Sefc, "Single layer Compton camera based on Timepix3 technology," Journal of Instrumentation, vol. 15, no. 1, Jan 2020, Art. no. C01014.

Iterative grating interferometry-based phase-contrast CT reconstruction with a data-driven denoising prior

Stefano van Gogh, Subhadip Mukherjee, Michał Rawlik, Zhentian Wang, Jinqiu Xu, Zsuzsanna Varga, Carola-Bibiane Schönlieb, Marco Stampanoni

***Abstract*—Breast cancer is the most common malignancy in women. Unfortunately, even though screening programs have helped to increase survival rates, the number of false positives and false negatives remains high. phase-contrast X-ray CT is a promising imaging technique which could improve breast cancer diagnosis by combining the high three-dimensional resolution of conventional CT with higher soft-tissue contrast. Grating Interferometry CT (GI-CT) arguably has the highest chance to make the transition to clinical practice. Unfortunately though, obtaining high-quality images is challenging. Grating fabrication defects and photon starvation lead to high noise amplitudes in the measured data. Moreover, the highly ill-conditioned differential nature of the GI-CT forward operator renders the inversion from corrupted data even more cumbersome. In this article we report on a novel regularized iterative reconstruction algorithm with a powerful data-driven regularization strategy to tackle this challenging inverse problem. In particular, we present an algorithm that combines the L-BFGS optimization scheme with a Plug-and-Play denoiser parameterized by a deep neural network and empirically show that the proposed method achieves high quality images, both on simulated data as well as on real measurements.**

***Index Terms*—Breast Imaging, Computed Tomography, Iterative Reconstruction, Data-driven Prior, Deep Learning**

I. Introduction

MALIGNANCIES of the breast still represent the most prevalent cancer in women [1]. Unfortunately, none of the currently used breast imaging techniques (mammography, breast ultrasound, breast MRI and absorption-based breast CT and tomosynthesis [2], [3]) is able to provide fully three-dimensional images with sufficiently high isotropic resolution and soft-tissue contrast necessary to identify critical breast cancer imaging biomarkers [4]. Therefore, better imaging modalities are needed to improve early detection and increase survival rates. X-ray phase-contrast CT could potentially offer a solution by combining the high three-dimensional resolution, which comes with CT, with superior soft-tissue contrast.

When X-ray waves interact with matter, their amplitude and phase are modified according to the refractive index of the material they interact with. The refractive index of a material is given by $n = 1-\delta+i\beta$. The real part δ dictates the change in the beam's phase Φ as

$$\Phi = \int \delta(x, y, z)\mathrm{d}z. \tag{1}$$

S. van Gogh, M. Rawlik, J. Xu and M. Stampanoni are with the Paul Scherrer Institut and ETH Zürich.
Z. Wang is with Tsinghua University.
Z. Varga is with the University Hospital Zürich.
S. Mukherjee and C.B. Schönlieb are with the University of Cambridge.

From this, the refraction angle α can be computed using

$$\alpha = \frac{\lambda}{2\pi}\frac{\partial\Phi}{\partial x}. \tag{2}$$

The imaginary part β is directly linked to the attenuation coefficient via $\mu = 4\pi\beta/\lambda$, which can then be used to compute the beam's attenuation by using the Beer-Lambert law.

It is widely known that soft tissues are characterized by similar β's [5], which makes it difficult to distinguish different tissue types in conventional absorption-based CT. Conversely, larger differences in δ's [5] can theoretically yield higher soft tissue contrast in reconstructed phase-contrast CT volumes.

The X-ray's phase must be computed indirectly and many approaches to achieve this have been proposed over the years. Grating interferometry [6], [7], [8] arguably has the highest chance of making the transition to clinical practice. In fact, it has non-restrictive requirements in terms of temporal and spatial coherence of the X-ray beam, it can be operated at large fields-of-view (FOV) and it has a comparably high mechanical robustness [6].

Grating interferometry detects the X-ray's refraction angle α by exploiting a peculiar interference pattern called Talbot carpet [9]. When an X-ray beam is refracted, this results in a lateral shift in the interference pattern. Therefore, by measuring this shift, the wavefront's change in phase can be easily obtained by integrating (2). To obtain this Talbot carpet, three gratings are positioned between the source and the detector [7]. The first grating (source grating or G0) is composed of a highly absorbing material such as gold and is placed immediately in front of the X-ray tube to improve beam coherence. The second grating (phase grating or G1), which is not designed to absorb photons, imposes a significant phase-shift to the X-ray beam and creates the interference pattern. To measure the lateral shift of this interference pattern induced by the sample, which is in the µm range, a highly resolving detector would be required. Unfortunately, to date no such detectors exist. Therefore, to circumvent this problem, a highly absorbing third grating (analyzer grating or G2) is placed in front of the detector. By moving one of the gratings with respect to the others in x-direction, it is then possible to obtain an interferogram called phase stepping curve [10], from which it is in turn possible to compute the lateral shift of the interference pattern.

The interferogram is modeled as

$$I_k = I_0 T \cdot [1 + V_0 D \cdot \cos(k + \Phi_0 - \varphi)] \tag{3}$$

where I_0, V_0, and Φ_0 are the flat-field intensity, visibility and phase maps, respectively, and k is the k-th phase step. The transmission T, the dark-field D and the differential phase sinograms are given by:

$$T = \exp\left[-\int \mu(x,y,z)\mathrm{d}z\right], \tag{4}$$

$$D = \exp\left[-\int \epsilon(x,y,z)\mathrm{d}z\right], \tag{5}$$

$$\varphi = \frac{\lambda d_2}{g_2}\frac{\partial}{\partial x}\int \delta(x,y,z)\mathrm{d}z \tag{6}$$

with λ being the wavelength, g_2 the pitch of the G2 grating and d_2 the distance between the origin and G2.

The interference pattern's shift φ, which is directly linked to the beam's refraction and thus phase, can then be retrieved with Fourier analysis. The same holds for the absorption signal, which is related to the average intensity of the curve, and for the dark-field signal, which is related to the curve's amplitude. In this work though we will focus exclusively on phase. When combined with a CT acquisition protocol, grating interferometry (GI) naturally extends to GI-CT.

In an attempt to bring GI-CT to clinical practice, our group has embarked in a long term effort to build a first-of-its-kind Grating Interferometry Breast CT (GI-BCT) prototype. As such, deposited radiation dose, scanning times and patient comfort must be compatible with clinical standards.

With these constraints in place, high quality images with grating interferometry CT represent a challenging objective. In fact, to date, the successful use of grating interferometry phase-contrast CT has been limited to synchrotron beamlines [11] and laboratory setups [12] where high image quality is achieved by a high X-ray flux or by long scanning times.

There are two main reasons why it is so challenging to obtain high-quality images with GI-CT. First, an intrinsic noise amplification takes place during signal retrieval [13]. Second, the differential nature of the phase-contrast forward operator causes the inverse problem to be more ill-conditioned as compared to conventional CT.

The main figure of merit which determines the quality of a grating interferometer is its visibility, i.e. the amplitude of the interference pattern in a flat-field scan. The higher the visibility, the more precisely one can compute the interference pattern's lateral shift and, consequently, the X-ray beam's refraction. Visibilities of 30% have been reported for polychromatic 46keV setups [14]. Unfortunately, high visibilities are challenging to achieve for systems which are shorter and/or have higher sensitivity, which requires the grating structures to be smaller.

Given the highly noisy sinogram data, it is important to reconstruct the tomograms with a stable inversion algorithm. A pseudo-inverse such as the filtered backprojection (FBP) algorithm could be applied in conjunction with the Hilbert filter [15] to solve this task. However, it is widely accepted in the CT community that iterative reconstruction algorithms are better suited to deal with highly ill-conditioned problems.

Typically, in iterative reconstruction, we need to define a variational loss function comprising a data-fidelity term and a regularization functional which incorporates prior knowledge about the expected reconstruction. Minimization of this loss with an optimizer of choice then allows to reconstruct an image which is simultaneously consistent with the measurements as well as with the prior knowledge.

The most widely used prior in image reconstruction is the total variation (TV) prior [16], which promotes homogeneous regions separated by sharp edges in the solution by assuming piece-wise constant signals. While TV is still considered to be a powerful baseline algorithm to regularize ill-conditioned inverse problems, recent years have witnessed the rise of data-driven algorithms which outperform traditional methods and thus constitute the current state-of-the-art in the field.

Two main data-driven approaches that draw inspiration from classical variational optimization schemes have been proposed in the literature. The first approach comprises end-to-end methods which unroll iterative schemes, thereby transforming each iteration of the iterative reconstruction algorithm into a distinct layer of a neural network [17], [18]. Since the imaging physics is embedded into the network, these models are generally believed to be more robust to noise and adversarial perturbations compared to pure black-box neural networks. In the second approach, the idea is to learn the regularizer a-priori on a representative training set, and to then use the trained regularizer in conjunction with the data-fidelity term in a classical variational optimization framework [19], [20], [21], [22], [23].

The first type of approach tends to yield superior results and deliver faster reconstructions [18]. However, it has the important disadvantage that no convergence guarantees can be derived and that it needs explicit supervision. Algorithms in the second category tend to be slower and potentially yield slightly inferior results. However, they have several advantages. The algorithms can be trained independently of the forward operator, they are more data-efficient (and possibly unsupervised), and they are amenable to stability and convergence analysis [21]. Since it is of utmost importance in the medical field to reliably reconstruct the tomograms, we propose a novel algorithm that fits into the second category.

Many different ideas have been proposed to learn a regularizer in a data-driven manner. Some notable ones are adversarial regularization [20] and its convex counterpart [21], score matching networks [24], regularization by denoising (RED) [19], network Tikhonov (NETT) [25] and data-driven Plug-and-Play denoisers [22], [23].

To address some of the challenges in GI-BCT and to be able to reconstruct phase-contrast tomograms from highly corrupted measurement data, we propose an iterative reconstruction algorithm which leverages the power of deep learning to regularize the highly ill-conditioned tomographic inversion problem. In particular, we propose an algorithm that alternates between data updates governed by the L-BFGS algorithm [26] and regularization steps performed with a denoising deep neural network in a Plug-and-Play fashion [27]. We apply the proposed approach to both simulated data and real measurements and show that it achieves excellent results.

II. Methods

The optimization problem we aim to solve is

$$\text{argmin}_\delta \frac{1}{2}\|A\delta - \varphi\|_2^2 \tag{7}$$

where δ is the image containing the real part of the index of refraction, φ is the retrieved differential phase-contrast sinogram, and A is a linear forward operator modeling equation (6).

Unfortunately, (7) is a highly ill-conditioned problem. Since an unregularized optimization of (7) will converge to a highly unstable solution, a powerful regularization strategy is necessary to stabilize the reconstruction.

As mentioned in the introduction, data-driven regularizers have emerged as the new state-of-the-art in the field and now routinely outperform classical regularization strategies such as total variation (TV) [16]. Among the many proposed strategies, we found data-driven Plug-and-Play regularizers to work best. Therefore, we here propose to learn a regularization network which is able to remove both noise and artefacts from the image iterates as they converge to the final reconstruction.

Given corrupted and clean images, we wish to train a network f_θ which maps the corrupted image δ to its clean counterpart δ^*. The objective function to achieve this is

$$\mathcal{L} = \frac{1}{n}\sum_{i=1}^{n} \|f_\theta(\delta_i) - \delta_i^*\|_2^2 \tag{8}$$

where f_θ is an deep neural network with sufficient expressive power and $\{\delta_i, \delta_i^*\}$ are training data pairs.

We parameterized f_θ with a 7 million parameter, bias-free U-net [28]. We removed the biases because it has been shown that this leads to 1) more interpretable denoising performance, and 2) that biasless networks generalize better to different noise amplitudes [29]. The higher interpretability comes from the fact that we can regard the denoising process as being locally linear [29]. Therefore, computing the Jacobian shows how the pixel neighborhood is used for denoising a particular pixel. The higher robustness is important in our case since 1) the noise encountered during image reconstruction might slightly differ from iteration to iteration, and 2) it makes it easier to train the network as we do not have to find the perfect amount of noise to train on.

To solve (7) we use the L-BFGS optimization scheme proposed by [26] and apply a denoising step after every k-th iteration. The proposed method is summarized in Algorithm 1.

By removing noise and artefacts, the denoising step can be interpreted as a projection of the current image iterate to the data manifold to which clean reconstructions belong. Iteratively alternating between data-fidelity optimization and projection to the data manifold thereby allows to get close to the measured data, while staying close to the data manifold of clean reconstructions.

Algorithm 1: L-BFGS optimization with data-driven Plug-and-Play denoiser

input : $i = 0$; $\delta_0 = 0$; $k_{\max} = 20$; $\epsilon = 10^3$;
while $\frac{1}{2}\|A\delta_k - \varphi\|_2^2 > \epsilon$ **do**
  $k = 0$; $y = 0$; $s = 0$;
  if $i > 0$ **then**
    $\delta_k = \delta_{\text{reg}}$;
  end
  while $k < k_{\max}$ **do**
    $\mathcal{L}_k = \frac{1}{2}\|A\delta_k - \varphi\|_2^2$;
    $\nabla_{\delta_k}\mathcal{L}_k = A^T(A\delta_k - \varphi)$;
    if $k > 0$ **then**
      $s[k-1] = \delta_k - \delta_{k-1}$;
      $y[k-1] = \nabla_{\delta_k}\mathcal{L}_k - \nabla_{\delta_{k-1}}\mathcal{L}_{k-1}$;
    end
    $\delta_{k+1} = \delta_k - \text{LBFGS}(\nabla_{\delta_k}, s, y)$;
    $k = k + 1$;
  end
  $i = i + 1$;
  $\delta_{\text{reg}} = f_\theta(\delta_{k_{\max}})$;
end
output: δ_{reg}

III. Results

A. Simulated data

To quantitatively assess the performance of our method we applied it to simulated breast phantoms developed in-house.

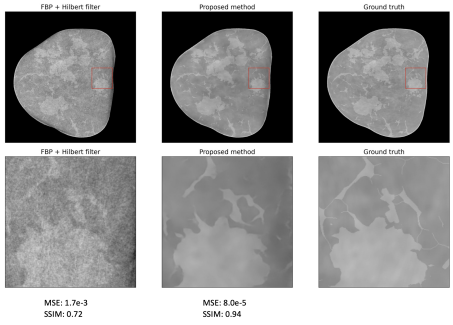


Fig. 1. Reconstruction results on simulated breast phantoms. First column: full slice, second column: the part inside the red region enlarged.

We simulated 600 projections and added Poisson noise that matched real measurements. The results in Figure 1 show that the proposed method achieves excellent results and clearly outperforms analytical reconstruction, both qualitatively as well as in terms of structural similarity index (SSIM) and mean squared error (MSE).

B. Real data

To demonstrate the effectivenss of our method on real data, we scanned a fixed mastectomy obtained at the University Hospital Zürich on our GI-BCT prototype. We acquired 600 projections under continuous circular rotation. A total of 10 scans have been averaged to compensate for the low visibility (thus high noise amplitudes) we are currently working with.

We applied exactly the same algorithm as for the simulated data. Importantly, we used the neural network that has been trained on simulated data. Figure 2 shows that the proposed method once again achieves excellent results and clearly outperforms analytical reconstruction. We want to emphasize that

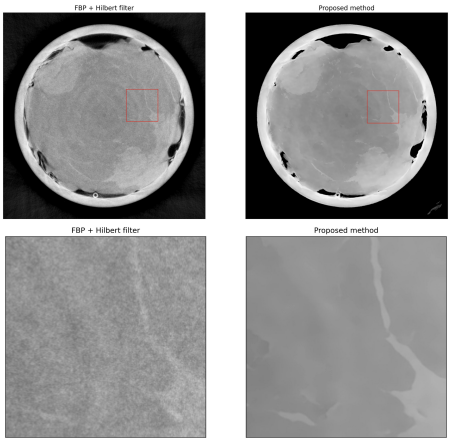


Fig. 2. Reconstruction results on mastectomy data. First column: full slice, second column: the part inside the red region enlarged.

the applicability of a denoiser, that was trained on simulated data, to real-world data, is a crucial advantage of our method since it is very difficult to obtain high-quality real data in a medical setting.

IV. Conclusion

In this article we have proposed a novel iterative reconstruction algorithm with a data-driven denoising prior and have shown that it is able to produce excellent results. Importantly, the regularizer can be trained on simulated data and later be applied to real measurements. Ongoing work focuses on the theoretical convergence guarantees of the proposed algorithm.

References

[1] N. Harbeck and M. Gnant, "Breast Cancer," *Lancelet*, vol. 389, no. 10074, pp. 1134–1150, 2017. [Online]. Available: http://dx.doi.org/10.1016/S0140-6736(16)31891-8

[2] W. A. Kalender, D. Kolditz, C. Steiding, V. Ruth, F. Lück, A. C. Rößler, and E. Wenkel, "Technical feasibility proof for high-resolution low-dose photon-counting CT of the breast," *European Radiology*, vol. 27, no. 3, pp. 1081–1086, 2017.

[3] S. Shim, N. Saltybaeva, N. Berger, M. Marcon, H. Alkadhi, and A. Boss, "Lesion Detectability and Radiation Dose in Spiral Breast CT With Photon-Counting Detector Technology: A Phantom Study," *Investigative radiology*, vol. 55, no. 8, pp. 515–523, 2020. [Online]. Available: https://doi.org/10.1097/RLI.0000000000000662

[4] C. J. D'Orsi, E. A. Sickles, E. B. Mendelson, E. A. Morris, W. E. Creech, P. F. Butler, P. G. Wiegmann, M. B. Chatfield, L. W. Meyer, and P. A. Wilcox, "ACR BI-RADS Atlas, Breast Imaging Reporting and Data System," *American College of Radiology*, 2013.

[5] S. A. Zhou and A. Brahme, "Development of phase-contrast X-ray imaging techniques and potential medical applications," *Physica Medica*, vol. 24, no. 3, pp. 129–148, 2008.

[6] T. Weitkamp, A. Diaz, C. David, F. Pfeiffer, M. Stampanoni, P. Cloetens, and E. Ziegler, "X-ray phase imaging with a grating interferometer," *Opt. Express*, vol. 13, no. 16, pp. 6296–6304, aug 2005. [Online]. Available: http://www.opticsexpress.org/abstract.cfm?URI=oe-13-16-6296

[7] F. Pfeiffer, T. Weitkamp, O. Bunk, and C. David, "Phase retrieval and differential phase-contrast imaging with low-brilliance X-ray sources," *Nature Physics*, vol. 2, no. 4, pp. 258–261, 2006.

[8] A. Olivo and R. Speller, "A coded-aperture technique allowing x-ray phase contrast imaging with conventional sources," *Applied Physics Letters*, vol. 91, no. 7, 2007.

[9] H. Talbot, "LXXVI. Facts relating to optical science. No. IV," *The London, Edinburgh, and Dublin Philosophical Magazine and Journal of Science*, vol. 9, no. 56, pp. 401–407, 1836.

[10] A. Momose, S. Kawamoto, I. Koyama, Y. Hamaishi, K. Takai, and Y. Suzuki, "Demonstration of x-ray Talbot interferometry," *Japanese Journal of Applied Physics*, vol. 42, no. 7B, pp. L866–L868, 2003.

[11] R. Longo, F. Arfelli, D. Bonazza, U. Bottigli, L. Brombal, A. Contillo, M. A. Cova, S. Donato, D. Dreossi, V. Fanti, C. Fedon, B. Golosio, G. Mettivier, P. Oliva, S. Pacile, A. Sarno, L. Rigon, P. Russo, A. Taibi, M. Tonutti, F. Zanconati, and G. Tromba, "Advancements towards the implementation of clinical phase-contrast breast computed tomography at Elettra," *Journal of Synchrotron Radiation*, vol. 26, no. 4, pp. 1343–1353, 2019.

[12] J. Vila-Comamala, L. Romano, K. Jefimovs, H. Dejea, A. Bonnin, A. C. Cook, I. Planinc, Z. Wang, M. Stampanoni, and M. Cikes, "High Sensitivity X-ray Phase Contrast Imaging by Laboratory Grating-based Interferometry at High Talbot Order Geometry," *Optics Express*, 2021.

[13] V. Revol, C. Kottler, R. Kaufmann, U. Straumann, and C. Urban, "Noise analysis of grating-based x-ray differential phase contrast imaging," *Review of Scientific Instruments*, vol. 81, no. 7, 2010.

[14] K. Willer, A. A. Fingerle, W. Noichl, F. De Marco, M. Frank, T. Urban, R. Schick, A. Gustschin, B. Gleich, J. Herzen, T. Koehler, A. Yaroshenko, T. Pralow, G. S. Zimmermann, B. Renger, A. P. Sauter, D. Pfeiffer, M. R. Makowski, E. J. Rummeny, P. A. Grenier, and F. Pfeiffer, "X-ray dark-field chest imaging for detection and quantification of emphysema in patients with chronic obstructive pulmonary disease: a diagnostic accuracy study," *The Lancet Digital Health*, vol. 3, no. 11, pp. e733–e744, 2021. [Online]. Available: http://dx.doi.org/10.1016/S2589-7500(21)00146-1

[15] Z. F. Huang, K. J. Kang, Z. Li, P. P. Zhu, Q. X. Yuan, W. X. Huang, J. Y. Wang, D. Zhang, and A. M. Yu, "Direct computed tomographic reconstruction for directional-derivative projections of computed tomography of diffraction enhanced imaging," *Applied Physics Letters*, vol. 89, no. 4, 2006.

[16] L. I. Rudin, S. Osher, and E. Fatemi, "Nonlinear total variation based noise removal algorithms," *Physica D: Nonlinear Phenomena*, vol. 60, no. 1-4, pp. 259–268, 1992.

[17] K. Hammernik, T. Klatzer, E. Kobler, M. P. Recht, D. K. Sodickson, T. Pock, and F. Knoll, "Learning a variational network for reconstruction of accelerated MRI data," *Magnetic Resonance in Medicine*, vol. 79, no. 6, pp. 3055–3071, 2018.

[18] J. Adler and O. Öktem, "Learned Primal-Dual Reconstruction," *IEEE Transactions on Medical Imaging*, vol. 37, no. 6, pp. 1322–1332, 2018.

[19] Y. Romano, M. Elad, and P. Milanfar, "The little engine that could: Regularization by Denoising (RED)," *SIAM Journal on Imaging Sciences*, vol. 10, no. 4, pp. 1804–1844, 2017.

[20] S. Lunz, O. Öktem, and C. B. Schönlieb, "Adversarial regularizers in inverse problems," *Advances in Neural Information Processing Systems*, vol. 2018-Decem, no. NeurIPS, pp. 8507–8516, 2018.

[21] S. Mukherjee, S. Dittmer, Z. Shumaylov, S. Lunz, O. Öktem, and C. B. Schönlieb, "Learned convex regularizers for inverse problems," *arXiv*, pp. 1–10, 2020.

[22] J. Hertrich, S. Neumayer, and G. Steidl, "Convolutional Proximal Neural Networks and Plug-and-Play Algorithms," pp. 1–34, 2020. [Online]. Available: http://arxiv.org/abs/2011.02281

[23] R. Cohen, M. Elad, and P. Milanfar, "Regularization by denoising via fixed-point projection (RED-PRO)," *arXiv*, pp. 1–33, 2020.

[24] Z. Ramzi, B. Remy, F. Lanusse, J.-L. Starck, and P. Ciuciu, "Denoising Score-Matching for Uncertainty Quantification in Inverse Problems," 2020. [Online]. Available: http://arxiv.org/abs/2011.08698

[25] H. Li, J. Schwab, S. Antholzer, and M. Haltmeier, "NETT: Solving inverse problems with deep neural networks," 2018.

[26] J. Nocedal, "Updating Quasi-Newton Matrices with Limited Storage," *Mathematics of Computation*, vol. 35, no. 151, p. 773, 1980.

[27] S. V. Venkatakrishnan, C. A. Bouman, and B. Wohlberg, "Plug-and-Play priors for model based reconstruction," *2013 IEEE Global Conference on Signal and Information Processing, GlobalSIP 2013 - Proceedings*, pp. 945–948, 2013.

[28] O. Ronneberger, P. Fischer, and T. Brox, "U-net: Convolutional networks for biomedical image segmentation," *Medical Image Computing and Computer-Assisted Intervention – MICCAI*, vol. 9351, pp. 234–241, 2015.

[29] S. Mohan, Z. Kadkhodaie, E. P. Simoncelli, and C. Fernandez-Granda, "Robust and interpretable blind image denoising via bias-free convolutional neural networks," pp. 1–22, 2019. [Online]. Available: http://arxiv.org/abs/1906.05478

A scatter correction method of CBCT via CycleGAN and forward projection algorithm

Tianxu Tang, Wei Zhang, and Weiqi Xiong

***Abstract*—Scatter artifacts is one of the limitations of cone-beam computed tomography(CBCT) image quality, this paper proposed a novel scatter correction method for CBCT via combining the deep learning and forward projection algorithm. This method can be mainly divided follow steps: Firstly, raw projections were used to reconstruct raw CBCT via FDK algorithm, and then the raw CBCT was processed by CycleGAN network to generate synthetic CT, after that the synthetic CT was used to forward projection based on the Beer's laws to generate scatter free primary projections. The raw scatters can be estimated via subtracting the primary projections from the raw projections, and then a median and low-pass Gaussian filter was used to smooth the raw scatters. Finally, The scatter corrected projections can be acquired by subtracting the filtered scatters form raw projections. The study results of pelvis and chest validated the effective of proposed correction method in reducing scatter artifacts of CBCT. Compared with uncorrected CBCT and CBCT corrected by scatter kernel superposition(SKS) method, the proposed method can more effectively improve the quality of reconstructed CBCT, reducing the CT number errors and increasing contrast-to-noise(CNR) and so on. All results show that this method has strong scatter artifacts restriction ability, so it has significant promising for clinical application.**

***Index Terms*—CBCT, scatter correction, CycleGAN, forward projection.**

I. Introduction

CONE-beam CT (CBCT) has been widely used in image-guide radiation therapy, it can provided accurate patient position, and also can track the tumor positon in therapy. But as the existing of scatter signals in CBCT, it will lead to strong noise and CT number errors. This is one of reasons to limit the application of CBCT in dose calculation, adaptive radiation therapy and so on. Scatter correction technology for CBCT has been widely study, those methods can be mainly divided three categories: software-based[1-4], hardware-based[5,6] and software-hardware hybrid methods[7-11]. In earlier study, many hardware-based methods been proposed, such as anti-scatter grid method, air gap method and so on. Software-based mainly contains Monte Carlo, deep learning and SKS methods. Furthermore, there are also some methods combined the hardware and software, those study usually include beam stop array, primary modulation method and so on. Above scatter correction methods mainly via direction measurement or theoretical calculation to estimate scatter, considering the characteristic of existing planning CT in radiation therapy, there are also methods estimate scatter signals via planning CT forward projection and then subtracted it from raw projections[12-15]. Nui et al[12] via rigid registration of the planning CT with CBCT, and then forward projection process be used in planning CT to acquire primary projections, so the scatter signals can be estimated easily and be smoothed. To further decrease the estimated scatter errors, they also used some improvement technologies such as deformable registration, gas pocket matching in this method. As the mismatching of CBCT with planning CT, it will leads serious errors for scatter estimation, Cui et al[15] used local filtration technique for the estimated scatter signals to improvement the accuracy of scatter estimation furthermore.

Nui and Cui's method just can used in those situation there are planning CT and it significantly rely on the accuracy of registration. In this paper, We proposed a novel scatter correction method via combine the deep learning and forward projection technique. A CycleGAN deep learning network was used to convert the CBCT to synthetic CT, and then use synthetic CT to generate primary projections, and estimating scatter signals. Our method has significant advantages that it didn't need any prior information, so the application regions are more extensive, and proposed method also avoided the scatter estimation errors bring by registration.

Tianxu Tang, Wei Zhang and Weiqi Xiong are with United Imaging Healthcare, shanghai, 201807, china. (E-mail: tangtianxu0815@gmail.com)

II. Methods

A. Generating Synthetic CT

We used CycleGAN network[16] as basic deep learning network to convert raw CBCT to synthetic CT, it used two generators and discriminators to convert images between two cycle networks. We used a U-net architecture for generators and a patchGAN for discriminators. Generators are used to generate synthetic CT, while discriminators are used to distinguish real or fake data, Fig2 shows the architecture of CycleGAN network in this paper, it mainly include two cycles. In cycle one, CBCT is converted to synthetic CT via generator G_1, and discriminator D_2 is used to discriminate synthetic CT and real fan-beam computed tomography(FBCT) to restrain generator G_1, generator G_2 is used to generator synthetic CBCT from synthetic CT. In cycle two, FBCT is firstly converted to synthetic CBCT by generator G_2, while discriminator D_2 discriminates between synthetic CBCT and real CBCT, and the same time synthetic CBCT is converted to synthetic CT using generator G_2. We used 24 patients dataset as training, 5 patients dataset as validation, and 2 patients dataset as test to train CycleGAN network. Every patient contains 270 CBCT slices with 512x512 axial pixel and thickness 1mm, the corresponding FBCT also was resampled to same size with CBCT.

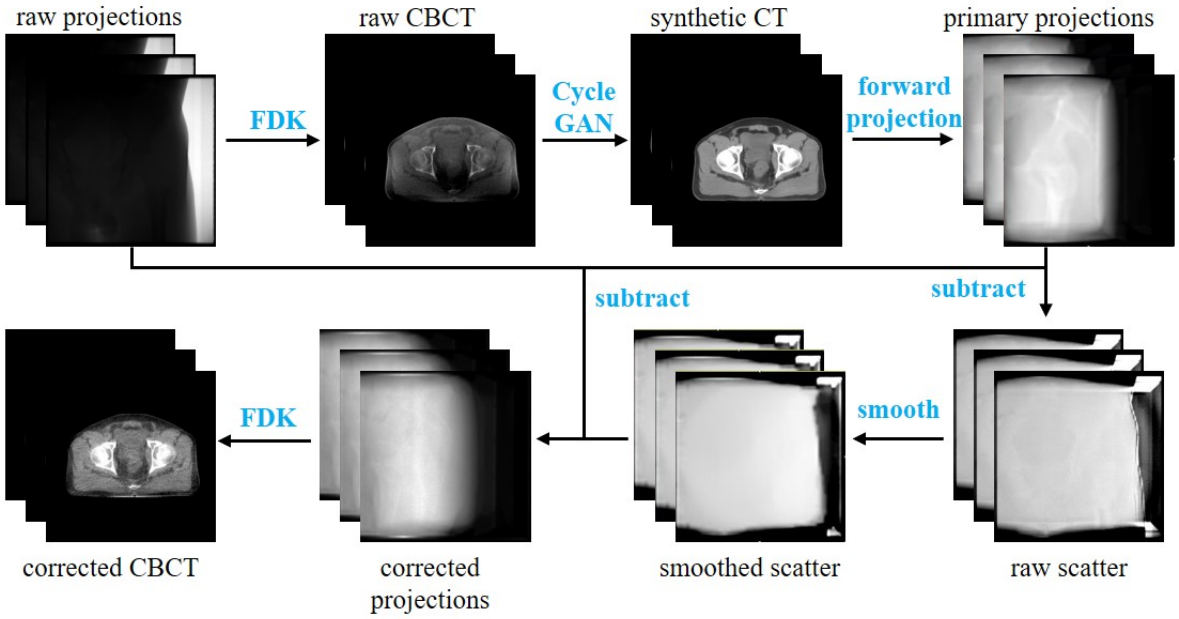


Fig. 1. The workflow of proposed scatter correction method.

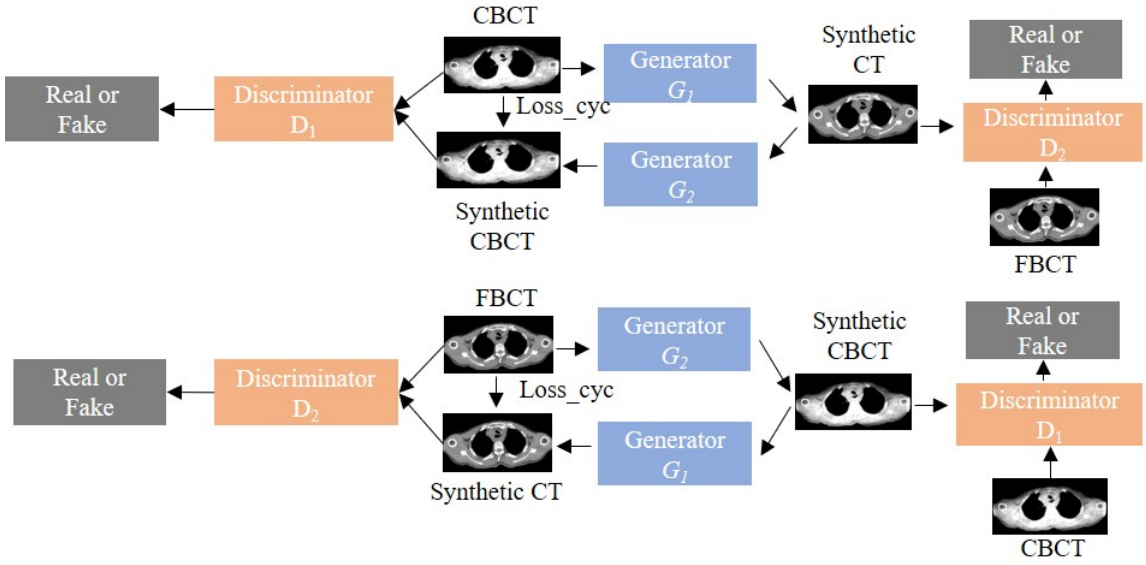


Fig. 2. The architecture of CycleGAN network in this paper.

B. Scatter Estimation

The work flow of proposed scatter correction method as show in Fig1. Raw CBCT images were reconstructed firstly using standard FDK algorithm, and then the CBCT images were used to generate synthetic CT images via CycleGAN network. In forward projection procedure, HU number of synthetic CT firstly was transformed to corresponding water density, scatter free primary projections were acquired via forward projection of synthetic CT images based on Beer's law. In traditional scatter correction using deep learning directly, the high-frequency errors can not be tolerant, but in our case only the low-frequency signals of synthetic CT are used, so our method has some tolerance for high-frequency errors of synthetic CT. To improve the calculation times and save memory space, the primary projections and raw projections were downsampled from 1024×1024 to 256×256. When subtracting estimated scatter form raw projections, the scatter signals will be upsampled firstly back to 1024×1024. For every projection angle, subtraction of primary projections from raw CBCT projections gives the raw scatter projection

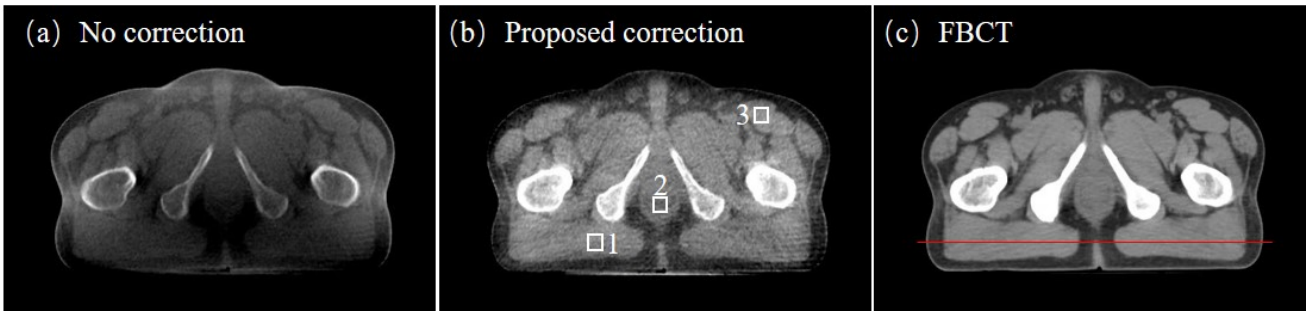


Fig. 3. The pelvis study with different way: (a)no correction, display window:[-200 400]; (b)with proposed correction, display window:[40 400]; (c)ground truth FBCT, display window:[40 400].

images. Synthetic CT maybe has some slight errors, so there were some mismatching between raw CBCT projections and primary images of forward projection, and the low frequency characterize of scatter signals, a median filter(size 45-by-45 pixels) and a low-pass Gaussian filter(size 51-by-51 pixels) was used to smooth raw scatter projection images. Finally, subtracting smoothed scatter from raw projections, and then using subtracting results to reconstructed via FDK algorithm.

C. Evaluation Metrics

To evaluate the performance of proposed correction method, using contrast and contrast-to-noise (CNR) as the evaluation metrics, the contrast calculated as:

$$Contrast = HU_r - HUs \tag{1}$$

the CNR calculated as:

$$CNR = \frac{HU_r - HUs}{\sqrt{\sigma}} \tag{2}$$

here HU_r is the mean of selected region, HU_s is the mean of surrounding area, σ is the standard deviation of selected region.

III. Results and Discussion

We performed our study on United-Imaging uRT-linac 506c, it contains an advanced MV-CBCT and FBCT system. The reconstructed FBCT images had a pixel size of 512×512×90 and voxel size of 0.97mm in axial plane and 3mm in longitudinal direction, while the CBCT images had a pixel size of 512×512×270 and voxel size of 0.78mm in axial plane and 1mm in longitudinal direction. To enlarge the field of view of MV-CBCT imaging, the flat detector was shifted by 90mm, and the reconstruction system of uRT-linac 506c contains strong denoising process. Fig3 shows a pelvis study, it contains without correction CBCT, proposed correction CBCT and ground truth FBCT. In the pelvis case, the HU number of CBCT corrected by proposed method has significant increase than without correction CBCT, and the serious artifacts in without correction images had been improved via proposed method. The red line profile showed in Fig4 also demonstrate the effective of our method, and it is more nearing the FBCT compare with without correction. To quantitatively evaluate the performance of our method, three ROIs were chosen to calculate the CNR and contrast number as show in table1. In the selected ROIs, the CNR and contrast of three areas had obvious increasing. To further evaluate proposed scatter correction method, Fig5 shows the CBCT images of a chest data corrected by proposed method and SKS correction method. SKS scatter correction has been wildly study and used in commerce, because it has good performance and fast calculation, and don't need any other hardware equipment. From the chest performance at axial, coronal and sagittal plane, we can see that the contrast and HU uniformity of CBCT corrected by proposed correction has significant improvement compared with SKS correction.

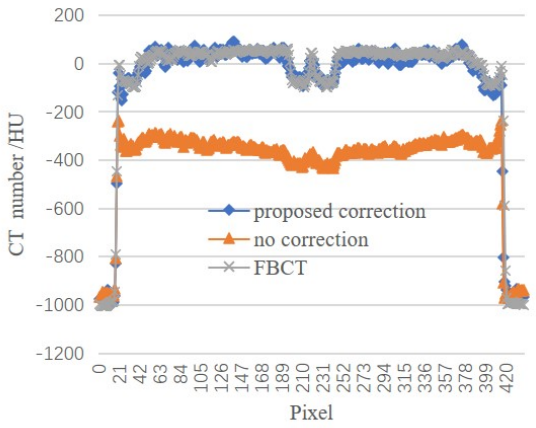


Fig. 4. 1D horizontal profile along the solid line drawn in Fig3(c).

TABLE I
Contrast and CNR of selected regions

Method	Metrics	ROI1	ROI2	ROI3
Proposed correction	Contrast	104	80	121
	CNR	25.22	17.45	31.4
Without correction	Contrast	52	48	11
	CNR	18.44	13.9	3.5

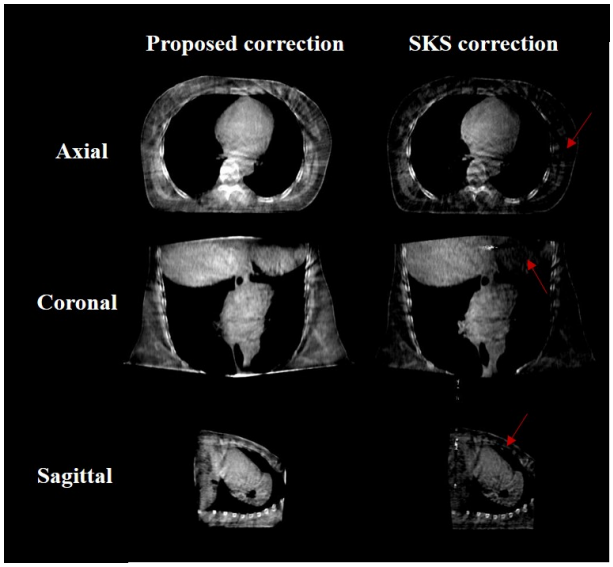


Fig. 5. The chest study with proposed and SKS correction. Left column is proposed correction, right column is SKS correction.From top to bottom is axial, coronal and sagittal plane. All images with display window:[40 400].

IV. Conclusion

An effective CBCT scatter correction method was proposed in this paper, this method combined the advantage of deep learning and forward projection. In traditional scatter correction using deep learning directly, there is a big problem is that can't tolerate high-frequency errors, but in our method slight high-frequency errors are can be tolerated, because we just used the low-frequency signals of synthetic CT to estimate scatters, and a median filter and low-pass Gaussian filter was used to smooth the raw scatter images. Previous correction method for CBCT via forward projection all most based on planning CT, while the mismatching and registration errors between CBCT and planning CT will affect the correction effects, out method don't need planning CT, so it can avoid above disadvantages bring by forward projection of planning CT. Whatever compared with raw CBCT or SKS correction CBCT with subjective feeling and quantitative analysis, the quality of CBCT corrected by proposed method all is better, and it is very close to FBCT. But there are also some problems, when the CBCT appears truncation for large body, the performance of proposed method is not good for those case. And at present, we just used a simple 2D median and Gaussian filter for raw scatters, more effective or precision filtering algorithms could be study in the future.

References

[1] Jarry G, Graham S A, Moseley D J, et al. Characterization of scattered radiation in kV CBCT images using Monte Carlo simulations[J]. Medical Physics, 2006, 33(11):4320-4329.

[2] Maslowski A, Wang A, Sun M, et al. Acuros CTS: A fast, linear Boltzmann transport equation solver for computed tomography scatter–Part I: Core algorithms and validation[J]. Medical Physics, 2018;45(5):1899-1913.

[3] Xiao, Liang, Liyuan, et al. Generating synthesized computed tomography (CT) from cone-beam computed tomography (CBCT) using CycleGAN for adaptive radiation therapy[J]. Physics in medicine and biology, 2019, 64(12):125002-125002.

[4] Niu T, Sun M, J Star-Lack, et al. Shading correction for on-board cone-beam CT in radiation therapy using planning MDCT images[J]. Medical Physics, 2010, 37(10):5395-5406

[5] Shen S Z, Bloomquist A K, Mawdsley G E, et al. Effect of scatter and an antiscatter grid on the performance of a slot-scanning digital mammography system[J]. Medical Physics, 2006, 33(4):1108-1115.

[6] Siewerdsen, J, H, et al. The influence of antiscatter grids on soft-tissue detectability in cone-beam computed tomography with flat-panel detectors[J]. Medical Physics, 2004.31(12):3506-3520.

[7] Ning R, Tang X, Conover D. X-ray scatter correction algorithm for cone beam CT imaging[J]. Medical Physics, 2004, 31(5):1195-1202

[8] Min, J., Pua, R., Kim, C., Park, M., et al. A weighted rebinned backprojection-filtration algorithm from partially beam-blocked data for a singlecan cone-beam ct with hybrid type scatter correction. Medical Physics, 2019.46(3):1182-1197.

[9] Choi Y W, Choi J G, Kim Y S, et al. Scatter characterization using a beam-stop-array algorithm for digital breast tomosynthesis[J]. Journal of the Korean Physical Society, 2013, 46(3):1182-1197.

[10] Gao H, Zhu L, Fahrig R. Virtual scatter modulation for X-ray CT scatter correction using primary modulator[J]. Journal of X-ray science and technology, 2017, 25(6):1-17.

[11] Gao H, Zhu L, Fahrig R. Modulator design for x-ray scatter correction using primary modulation: Material selection[J]. Medical Physics, 2010, 37(8):4029-4037.

[12] Niu T, Sun M, J Star-Lack, et al. Shading correction for on-board cone-beam CT in radiation therapy using planning MDCT images[J]. Medical Physics, 2010, 37(10):5395-5406.

[13] Andersen A G, Park Y K, Elstrm U V, et al. Evaluation of an a priori scatter correction algorithm for cone-beam computed tomography based range and dose calculations in proton therapy[J]. Physics and Imaging in Radiation Oncology, 2020, 16:89-94.

[14] Yang C, Wu P, Gong S, et al. Shading correction assisted iterative cone-beam CT reconstruction[J]. Physics in Medicine and Biology, 2017, 62(22):8495-8520.

[15] Cui H, Jiang X, Fang C, Zhu L, Yang Y. Planning CT-guided robust and fast cone-beam CT scatter correction using a local filtration technique. Med Phys. 2021 Nov;48(11):6832-6843.

[16] Zhu J Y, Park T, Isola P, et al. Unpaired Image-to-Image Translation using Cycle-Consistent Adversarial Networks[J]. IEEE, 2017.

Design and Optimization of 3D VSHARP® Scatter Correction for Industrial CBCT using the Linear Boltzmann Transport Equation

Kevin Holt[1], Devang Savaliya[2], Amy Shiroma[2], Martin Hu[1], David Nisius[1], Steve Hoelzer[1], Mingshan Sun[2], Don Vernekohl[2], Josh Star-Lack[2]

[1]Varex Imaging, 3835 Carnation St, Franklin Park, IL 60131
[2]Varex Imaging, 683 River Oaks Parkway, San Jose, CA 95131

***Abstract*— We have developed VSHARP®, a suite of scatter correction solutions that have been incorporated into the commercially available cone-beam software development toolkit, CST (Varex Imaging, Salt Lake City, UT) enabling scatter correction to be applied as part of an entire CBCT reconstruction pipeline. The suite includes 2D VSHARP®, a deconvolution correction using asymmetric Gaussian kernels, 2D VSHARP-ML, a U-NET machine-learning correction, and 3D VSHARP®, a correction using a rapid finite-element Linear Boltzmann Transport Equation (LBTE) solver to estimate scatter in a manner similar to traditional stochastic Monte Carlo (MC) simulations. Of the three corrections, 3D VSHARP is the most accurate and flexible since it can be readily applied to arbitrary scanner geometries, protocols, and scan parts while the 2D VSHARP models may need to be regenerated for each configuration. On the other hand, 3D VSHARP is inherently slower since a minimum of two reconstruction passes are needed and the LBTE solver, while much faster than traditional MC, is still computationally intensive. The goal of this work was to minimize LBTE run times for (typically large) industrial datasets by optimizing parameter settings, particularly the choice of the sampling grid dimensions. This was achieved by applying a multi-objective genetic algorithm to find the Pareto front characterizing the tradeoff between speed and accuracy and identifying key operating points on the curve. Testing with 720 frames of 3720x3720 projection data to make a reconstruction volume of size 500x500x600, we found that excellent image quality can be obtained by using a coarse scatter grid size of 27x27x32 volume and 44x44 detector and a primary grid size of 246x246 x295 volume and 295x295 detector, both over 42 frames for a grand total of 21 seconds LBTE computation time. We show the Pareto characterization, as well as demonstrations of 3D VSHARP image quality with significantly reduced scatter-induced artifacts such as streaking and shading.**

Keywords— CBCT, Scatter Correction, Linear Boltzmann Transport Equation (LBTE), Genetic Algorithms, Pareto Optimization

I. INTRODUCTION

Recent algorithmic and computational advances have made MC-like scatter correction approaches much more practical. The Acuros platform [1,2] rapidly solves the linear Boltzmann Transport Equation (LBTE) using finite element methods to determine the scatter distribution directly rather than stochastically as in conventional MC. Recently, we unveiled 3D VSHARP [3,4] that uses Acuros to achieve accuracies comparable to MC methods in a tiny fraction of the time. Acuros' accuracy and run-time are both highly dependent on the choice of sampling grid used for the finite element solution. If the grid is too coarse then results are inaccurate, and if the grid is too fine then time is wasted. Wang et al [2] have addressed this issue for medical use by optimizing the Pareto front over the set of sampling parameters. In this work we perform a similar Pareto optimization for an industrial case of an aluminum motorcycle cylinder head.

II. METHODS

A. CST Framework

The correction utilizes CST, Varex's CT reconstruction SDK, which allows for flexible connection of modular plugins to perform a reconstruction. CST includes over 30 bundled plugins, including those necessary to implement 2D VSHARP [5], 2D VSHARP-ML [6], and 3D VSHARP. CST also includes a Physics Library that contains user-selectable cross-sections as well as x-ray spectra and detector response files required by 3D VSHARP.

B. Pipeline with 3D VSHARP and 2-Pass FDK Reconstruction

An example pipeline with 3D VSHARP is shown in Figure 1.

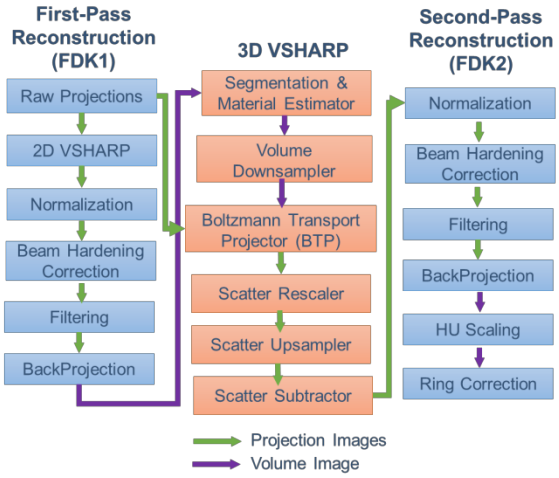


Figure 1. Two-pass pipeline for FDK reconstruction with 3D VSHARP. The data flow green lines indicate projection data and purple lines indicate volume data.

Processing for 3D VSHARP is in 6 basic stages:

1. Acquisition and Pre-processing: Read data from disk or from the detector, and perform pre-processing operations such as offset correction, bad-pixel correction, or lag correction.

2. 2D VSHARP: Perform 2D (kernel-based) scatter corrections for scatter from the detector housing and from the scanned object.
3. FDK1: Complete a first pass reconstruction using the 2D VSHARP scatter-corrected projection data.
4. 3D VSHARP contains six main components:

   a. Segmentation and Material Estimation: The FDK1 volume is segmented into different regions corresponding to different materials, and a density is assigned to each material-voxel.

   b. Volume Downsampling: To reduce BTP computation time and GPU memory requirements, the volume is downsampled to make a lower resolution volume sampling grid.

   c. The Boltzmann Transport Projector (BTP) runs the LBTE solver for a subset of projection angles. In addition to using a lower resolution volume sampling grid, the BTP detector matrix size is typically smaller than the original detector matrix size.

   d. Scatter Rescaling: The simulated scatter output from BTP is rescaled to be at a comparable signal level to the measured data. To help determine the scaling factor, the simulated primary output signal is used as a reference since the measured data should be proportional to the sum of the primary and scatter.

   e. Scatter Upsampling: The scatter signal is upsampled in 3 dimensions (detector matrix U,V and projection angle) so that sampling matches the measured data.

   f. Scatter Subtraction: The subtraction also includes a scatter-fraction smoothing and clipping step to ameliorate noise amplification from the subtraction.
5. FDK2: Perform a second pass reconstruction. Apply post-processing operations such as ring correction.

D. Data Acquisition and Reconstruction

A motorcycle cylinder head was scanned and reconstructed with the parameters shown in Table 1.

Acquisition Parameters	
Number of Detector Pixels (Dexels)	3072 x 3072
Projection Detector Pixel (Dexel) Size	139 μm x 139 μm
Detector Size	427 mm x 427 mm
Number of Projection Frames	720
Source-Axis Distance	1084 mm
Source-Imager Distance	1302 mm
Tube Spectrum	450 kV + 2 mmCu
Detector Model	4343HE with DRZ Plus
Reconstruction Parameters	
Number of Voxels	500 (x) x 500 (y) x 600 (z)
Voxel Size	0.5 mm x 0.5mm x 0.5 mm

Table 1: Acquisition and Reconstruction Parameters.

Reconstruction was performed on a PC with 2 Intel Xeon ES-2637v4 chips each containing 8 cores at 3.5 GHz, and an NVIDIA Titan RTX GPU with 4608 cores at 1.35 GHz.

E. Pareto Optimization

To characterize the processing time-vs-error tradeoff, we used the NSGA2 algorithm [7], a genetic algorithm (GA) for discovering Pareto fronts in multi-objective problems. The two objectives were time and error. To measure error, a "golden" reconstruction was performed with all parameters set for maximum accuracy, then for each operating point the root mean square (RMS) error versus the golden reconstruction was measured.

The total search space included: 1. Primary Volume Matrix Size, 2 Scatter Volume Matrix Size, 3. Primary Detector Matrix Size, 4. Scatter Detector Matrix Size, 5. Number of Primary Projections, 6. Number of Scatter Projections.

The constraints on the search were:

1. All voxels were isotropic and completely filled the prescribed reconstruction field-of-view (FOV) in the *x*-, *y*- and *z*-directions. As the z-axis FOV was somewhat larger than the transaxial FOV -- because the first pass reconstruction was extrapolated to better capture second-order scattering events in the top and bottom portions of the object -- the GA algorithm chose values for the number of voxels in the *x*-*y* direction, denoted as *Primary NumVoxelsXY* and *Scatter NumVoxelsXY.* and then automatically compute the corresponding number of primary or scatter voxels that spanned the entire z-axis FOV.
2. All detector pixels (**dexels)** were isotropic and completely filled the (square) detector extent in the U and V directions. The number of scatter dexels, *Scatter NumDexelsUV,* was chosen by the GA but the number of primary dexels depended on the number of primary voxels (*Primary NumVoxelsXY)* so that the Primary Dexel Size (mm) was the Magnification * Primary Voxel Size = SID/SAD * Primary Voxel Size = 1.2 x Primary Voxel Size.
3. The primary and scatter projection angles were equally spaced. The relevant native search parameter was termed the *Downsampling Factor* from which an *AngularIncremen*t value, quantized to multiples of 0.5° was computed. The number of frames then was then equal to Floor (720/*AngularIncrement*). The scatter frame rate downsampling factor (*Scatter FrameRate Downampling*) was an integer multiple of the *Primary FrameRate Downsampling*.

Table 2 shows the native search space used by the GA.

Search Parameter	**Search Range (Integers Only)**
Scatter NumVoxelsXY	25 to 90
Primary NumVoxelsXY	25 to 500
Scatter NumDexelsUV	10 to 42
Primary FrameRate Dowsampling	6 to 60
Scatter FrameRate Downsampling	*PrimaryFrameRateDownsampling × (1,2,3)*

Table 2: Native search space used by the GA.

We ran 25 generations of NSGA2 with a population of 50, then another 25 generations with a population of 75.

III. RESULTS

A. Motorcyle Cylinder Head Reconstructions

Figure 2 shows example reconstructions of a sagittal slice including A) an "Uncorrected" reconstruction, B) the first pass reconstruction (from the unoptimized 2D VSHARP), and two 3D VSHARP reconstructions: C) the "Golden" reconstruction performed at maximum LBTE resolution for a run time of about 6 hours, and D) the "Operating Point F" reconstruction using a coarser LBTE grid, requiring only 21 seconds of BTP time. Significant improvements in crispness and homogeneity are seen in the 3D VSHARP images with Operating Point F retaining similar image quality to the Golden image.

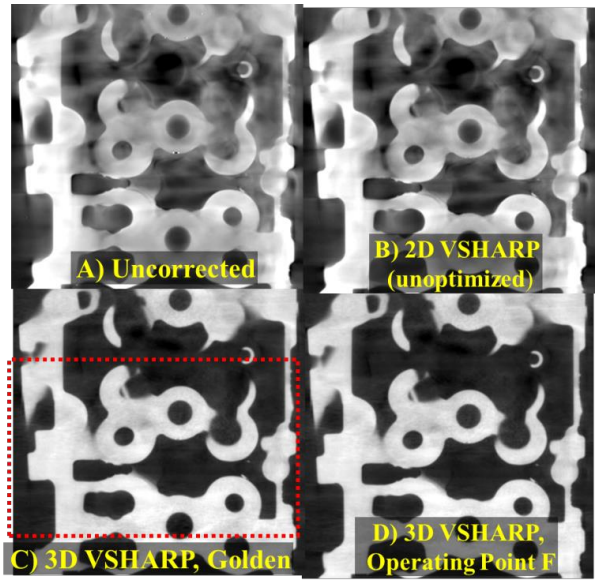


Figure 2:Central sagittal slice. The window for each is chosen so that the 10% and 90% gray values within the window correspond to the 5th and 95th percentiles of the voxel values. The red outline shows the zoomed-in region in Figure 4.

B. Pareto Results

Figure 3 shows results from all generations of the NSGA2 run. Each operating point is shown as a blue dot, the Pareto front is shown as a red line, the convex hull of all the operating points is shown in green, and the set of points that were further studied is indicated by labeled circles.

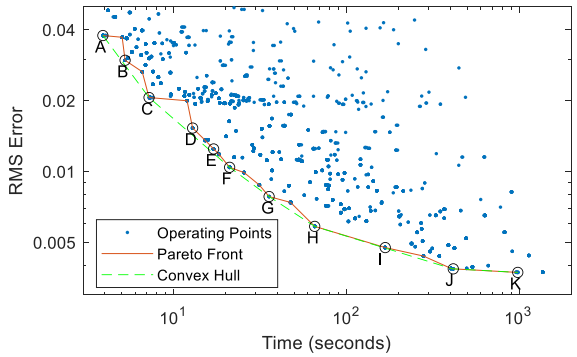


Figure 3: Log-Log plot of Pareto Optimization Results.

Table 2 shows search parameter and objective results for several operating points (A,D,F,I). As expected, RMS error decreases as 3D VSHARP runtime increases as do the sampling grid dimensions in general. Of note, is the exception that *Scatter NVoxelsXY* is relatively constant. This may partially reflect that its lower bound was 25. Also of note, *Scatter NDexelsUV* is more than 1.5x larger than *ScatterNVoxelsXY* which is greater than the 1.2x magnification one might have assumed. This may reflect the relatively high fidelity and angular resolution of each voxel's computed scatter distribution which is enabled by the use of Legendre polynomials to describe the profile and propagate scatter across the grid [1] which, in turn, might allow for coarser volume resolution than detector resolution. Finally of note is that *FrameRateDownsampling* for each operating point was the same for the primary and scatter computations even though higher voxel and dexel resolution was required for the primary estimate.

Sampling Grid Search Parameter	**GA results for selected Operating Points**			
	A	**D**	**F**	**I**
Scatter NVoxelsXY	27	27	27	25
Primary NVoxelsXY	85	119	246	449
Scatter NDexelsUV	13	44	44	57
Number Primary Projections	14	37	42	120
Number Scatter Projections	14	37	42	120
Objectives	**A**	**D**	**F**	**I**
RMS Error	.038	.015	.010	.005
3DVSHARP time (s)	3.9	12.9	21.1	168

Table 2: Results for selected 0perating points matching Figure 3.

C. Example Images Along the Pareto Front

Figure 3 shows zoomed-in images, corresponding to the red outlined region in Figure 2c, for the Operating Points in Table 3. The red arrows point to inhomogeneities in the form of streaks or shading. The top image (Operating Point A) takes the least amount of BTP time, 3.9 seconds, but does show artifacts. Moving down the figure, artifacts decrease as execution time increases. While homogeneity steadily improves with increased BTP time, we note that for many applications, images D or F may be perfectly acceptable, requiring 13 and 21 seconds BTP time respectively.

D. Contrast Analysis

The histograms of different reconstructed volumes are shown in Figure 4. In the Uncorrected Image, the air and aluminum peaks are poorly separated, with slightly better separation occurring in the 2D VSHARP image and good separation in 3D VSHARP images.

To quantify the separation of air and aluminum, the contrast-to-noise ratio in each histogram was computed by segmenting the images into "Air" or "Object" voxels and using the equation $\mathrm{CNR} = \frac{(\mu_{Obj}-\mu_{Air})\sqrt{2}}{\sqrt{\sigma_{Obj}^2+\sigma_{Air}^2}}$ where μ_{air} and μ_{obj} are the respective typical linear attenuations of the Air and Object voxels and σ_{air} and σ_{obj} are the respective standard deviations. We used the histogram peaks for μ, and their midpoints as the segmentation thresholds, shown respectively by an o and an x in Figure 4.

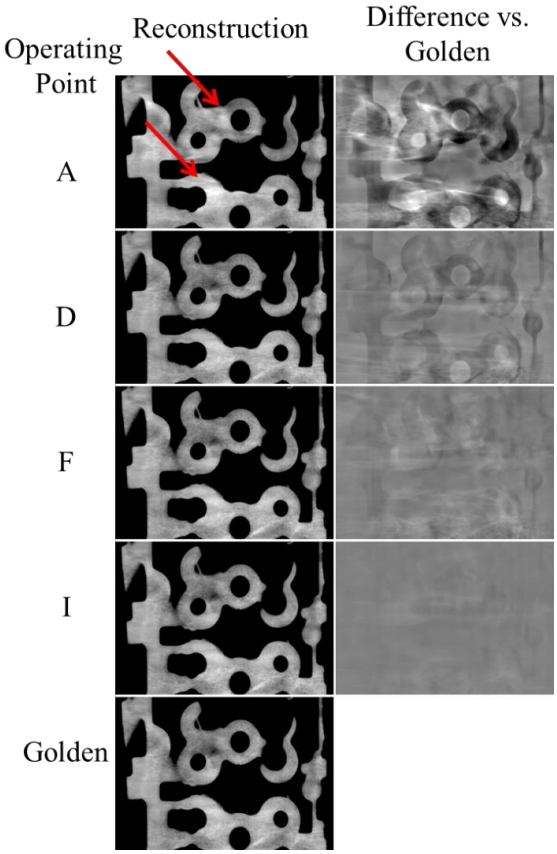


Fig. 3. Zoom-in on the central sagittal slice from various reconstructions. The window for each reconstruction is $[\mu_{Obj}-3\sigma_{Obj}, \mu_{Obj}+2\sigma_{Obj}]$. The window for each difference image is [-0.02 mm^{-1}, +0.02 mm^{-1}].

The resulting CNRs for operating Points A, D, F, I and "Golden" were 8.1, 8.9, 9.0, 9.0, 9.0 respectively. The Uncorrected CNR of 4.7 was the lowest while the 2D VSHARP CNR was slightly improved at 5.3. Points F and I have CNRs comparable to the golden image, which again suggests that there is no significant benefit to spending more than 10-20 seconds on the LBTE solution.

IV. DISCUSSION

MC or pseudo-MC scatter correction methods such as 3D VSHARP can produce highly accurate scatter estimates and, in fact, are used as a gold standard for training ML-scatter correction methods. A main advantage of MC methods is that they are versatile since all that is needed at runtime is the geometric specification of the CBCT system and a physics library which characterizes it. However, MC or even pseudo-MC methods are generally not as fast as Machine Learning or Kernel methods since a first pass reconstruction is required and the scatter transport calculation is computationally intensive.

For this data set, it was found that 10 to 20 second LTBE run times are sufficient if using an optimized sampling grid. We expect this result to be somewhat problem dependent, and may change with object size, complexity, or material, as well as with scanner geometry. However, it is interesting to note that our optimal time is roughly in line with the results of [2].

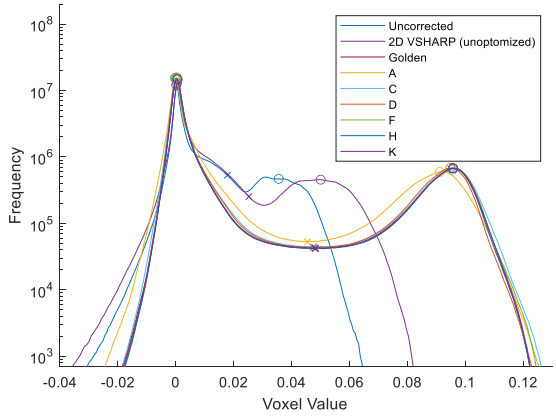


Fig. 4. Histograms of reconstructed volumes

Of note is that the 2D VSHARP calibration was not optimized for this setup. Although proper tuning may improve 2D VSHARP image quality, we chose to leave it unoptimized to show that the segmentation algorithm is fairly forgiving.

For future work, there are still many interesting parameters left to optimize including looking into non-uniform angular sampling [2], optimizing interpolation and segmentation methods, and optimizing intrinsic LBTE parameters such as energy grouping.

V. CONCLUSION

3D VSHARP was shown to significantly reduce scatter artifacts and produce excellent results with 10-20 seconds of computation time. Although a first-pass reconstruction is still needed, for the second pass reconstruction the results show that the additional time added by 3D VSHARP is minimal especially given that CST permits the LBTE computation to be performed in parallel with other FDK operations such as filtering and backprojection.

V. REFERENCES

[1] A. Maslowski et. al., "Acuros CTS: A fast, linear Boltzmann transport equation solver for computed tomography scatter– Part I: Core algorithms and validation", Med. Phys. 45(5) (2018) 1899-1913.

[2] A. Wang et. .al, "Acuros CTS: A fast, linear Boltzmann transport equation solver for computed tomography scatter –Part II: System modelling, scatter correction, and optimization", Med. Phys. 45(5) (2018) 1914-1925.

[3] A. Shiroma et al. "Scatter Correction for Industrial Cone-Beam Computed Tomography (CBCT) Using 3D VSHARP, a fast GPU-Based Linear Boltzmann Transport Equation Solver", 9th Conference on Industrial Computed Tomography (iCT) 2019, 13-15 Feb, Padova Italy.

[4] Star-Lack, et. al. "3D VSHARP®, a general-purpose CBCT scatter correction tool that uses the Linear Boltzmann Transport Equation." In *Medical Imaging 2021: Physics of Medical Imaging*, vol. 11595, International Society for Optics and Photonics, 2021.

[5] M. Sun, J. Star-Lack, "Improved scatter correction using adaptive scatter kernel superposition", Phys. Med. Biol. 55 (2010) 6695-6720.

[6] Maier, et. al. "Real-time scatter estimation for medical CT using the deep scatter estimation: Method and robustness analysis with respect to different anatomies, dose levels, tube voltages, and data truncation." *Medical physics* 46, no. 1 (2019): 238-249.

[7] K. Deb, et. al. "A Fast Elitist Non-dominated Sorting Genetic Algorithm for Multi-objective Optimization: NSGA-II", Lecture Notes in Computer Science, vol 1917, Springer, 2000

Motion Correction Image Reconstruction using NeuralCT Improves with Spatially Aware Object Segmentation

Zhennong Chen[1], Kunal Gupta[1], Francisco Contijoch[1,2]

***Abstract*— NeuralCT [1] has been recently proposed as an implicit neural representation-based image reconstruction that can produce time-resolved images from CT sinograms and reduce motion artifacts, even when undergoing complex motions. NeuralCT does not require the prior motion model or estimation of object motion. Instead, it utilizes a network to implicitly represent the time-varying object boundary by singed distance function and optimizes the network via differentiable rendering. In this work, we modify the NeuralCT framework to reconstruct scenes that have multiple moving objects with distinct attenuation levels. We show that the performance of NeuralCT reconstruction depends on the quality of the initialization of the network (in this case, object segmentation in motion corrupted FBP image). We show how spatially aware object segmentation can improve motion-corrected reconstruction in moving objects with multiple attenuation levels despite high angular motion and complex topological changes.**

***Index Terms*— Motion Correction, Implicit Neural Representation, Differentiable Rendering**

I. INTRODUCDTION

Cardiac computed tomography (CT) has emerged as a noninvasive method to evaluate the coronary artery disease and assess the cardiac function. However, image quality can be limited by motion of cardiac structures. For example, even slow coronary vessel motion (~15mm/s) can cause significant blurring of vessels [2]. Improved hardware such as faster gantry rotation or dual source designs can avoid/reduce motion artifacts but further improvement appears limited by physical constraints. Machine learning algorithms [3], [4] have been used to correct motion artifacts in reconstructed images. However, current approaches are limited by the need as a true motion vector field (for training) is unavailable in clinical data.

Recently, implicit neural representations (INR) [5] have been used to improve reconstruction of medical images [6], [7]. Gupta et al. [1] recently developed an INR-based framework to improve reconstruction of CT data corrupted by object motion. This framework, called "NeuralCT", takes CT sinograms as the input and produces time-resolved images and was shown to correct motion artifacts. A key benefit of NeuralCT is that it does not impose a motion model nor require estimates of the object motion. An overview is shown in **Fig 1**.

NeuralCT utilizes a neural network to implicitly represents (neural representation) the moving object boundary via the signed distance function (SDFs). Concretely, the INR maps the spatiotemporal domain of the moving object (a point at a particular position and time) to SDF value domain (the real-time relative position of this point with respect to the object boundary). In this work, the neural representation was initialized using intensity-based segmentation of the motion corrupted Filtered Backprojection (FBP) result. The representation was then optimized via differentiable rendering (DR) [5], a technique used to identify the shape of an object that best "explains" its acquired projection. Thus, NeuralCT aims to identify the optimal time-varying shape of moving object such that the resultant projection agrees with the CT sinogram (ground truth projections). We emphasize that NeuralCT is not a learning task that requires training and testing datasets as such approaches depend on data driven priors which have a tendency to introduce bias in the reconstruction. Instead NeuralCT builds on work where INR problems are solved via optimization. In this case, NeuralCT performs optimized reconstruction by forward rendering the moving object to acquire projection estimates, calculating the error between projection estimates and the true sinogram, and then updating the reconstruction by backpropagating the error via gradient descent.

In the initial description of NeuralCT, Gupta et al. showed *high-quality* motion-correction for a single foreground object with high angular motion (up to 200° displacement per gantry rotation) as well as complex topological deformation [1]. However, clinical CT scans are not composed of a single foreground class. Therefore, the core contribution to this study is to extend NeuralCT to successfully correct motion artifacts in scenes with multiple (i.e., different intensity) moving objects. In particular, we observed that imaging multiple moving objects with different attenuations can limit the accuracy of intensity-based segmentation and consequently decrease the reconstruction performance. As a result, we incorporate spatial information into the segmentation and compare our improved reconstruction result with the initial NeuralCT and FBP.

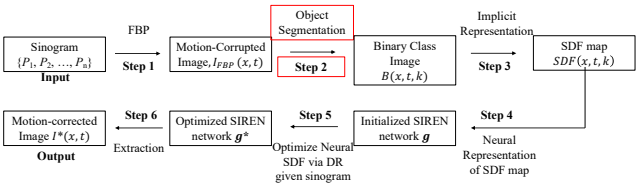


Fig. 1. NeuralCT framework. FBP = filtered backprojection, SDF = signed distance function, DR = Differentiable Rendering. In this study we proposed a new segmentation (red box) to extend NeuralCT to more complicated scenes.

This work is supported by NIH grant HL 143113.

[1]Department of Bioengineering, [1]Department of Radiology, University of California, San Diego, La Jolla, CA, 92093

II. METHODS

A. NeuralCT Framework

The NeuralCT framework is described in **Fig. 1** and the full description can be found in Ref [1]. The CT sinogram is the input and a time-resolved attenuation map $I^*(x,t)$ (motion-corrected image) is the output. The steps of the algorithm are:

Step 1: FBP images are created via backprojection of the sinogram P (comprised of a set of projections $\{P_1, P_2, \ldots, P_n\}$ for n gantry positions). This results in a series of motion-corrupted attenuation images $I_{FBP}(x,t)$.

Step 2: Segmentation *Seg* is used to identify different foreground objects from $I_{FBP}(x,t)$. The choice of *Seg* will be further discussed in Section II.B and II.C. Segmentation results in a binary time-varying images $B(x,t,k)$ where the k^{th} channel corresponds to the k^{th} foreground object.

Step 3: The time-varying scene of k binary images $B(x,t,k)$ is implicitly represented using the signed distance function (SDF). Specifically, $SDF(x,t,k)$ is generated to represent the position of the boundary as the signed distance of a point at location x in space at a particular time t to the boundary of the k^{th} object.

Step 4: For each location $x \in R^N$ where N is the number of spatial dimensions, the temporal evolution of an object's SDF was represented by Fourier Features (FF) using Fourier coefficients $\{A_0, A_1, \ldots, A_M, B_0, B_1, \ldots, B_M\}$:

$$SDF(x,t,k) \triangleq \frac{1}{M}\sum_{i=0}^{M} A_i(x,k)sin(2\pi\omega_i t) + B_i(x,k)cos(2\pi\omega_i t) \quad (1)$$

Here, ω_i are M randomly sampled frequencies. In our work, we approximated the SDF map $SDF(x,t,k)$ by a SIREN neural network [8] (an efficient framework to capture high frequency information). This neural network $\boldsymbol{g}(x,k;\boldsymbol{w})$, where $\boldsymbol{w}$ are weights in the network, was trained to output correct Fourier coefficients $\{A_i, B_i\}$ in Eqn. 1: $A_i(x,k;\boldsymbol{g})$ and $B_i(x,k;\boldsymbol{g})$.

The weights $\boldsymbol{w}$ were initialized randomly, and then updated by the standard gradient descent,

$$\boldsymbol{w} \leftarrow \boldsymbol{w} - \alpha\nabla\mathcal{L} \quad (2)$$

where $\mathcal{L} = \mathcal{L}_{SDF} + \lambda\mathcal{L}_E$. $\mathcal{L}$ is the total loss; $\mathcal{L}_{SDF}$ is the mean difference of the true SDF map (derived from FBP) versus $SDF(x,t,k;\boldsymbol{g})$ for all x,t,k; $\mathcal{L}_E$ is the Eikonal constraint computed as the mean value of absolute value of $\|\nabla_x SDF(x,t,k;\boldsymbol{g})\|_2$ -1 for all position x. λ is the regularization factor.

To conclude, after Steps 1-4 a SIREN neural network $\boldsymbol{g}$ is created that implicitly approximates the SDF map of the motion corrupted FBP images so $\boldsymbol{g}$ contains motion artifacts present after FBP.

Step 5: Differentiable Rendering (DR) is used to optimize $\boldsymbol{g}$ such that it represents a scene that is consistent with the acquired sinogram. Specifically, DR was used to identify the optimized shape $S^*$of an object that minimizes the projection loss $\mathcal{L}_P$ between the true projections (P_i) and the projections obtained via rendering of the estimated shape S:

$$\mathcal{L}_P = \sum_{i=0}^{n} |P_i - DR(S;\theta_i)| \quad (3)$$

Here, $DR(S;\theta_i)$ is the differentiable rendering operator; in CT, it represents the projection of an object shape S from "spatiotemporal attenuation space" $I(x,t)$ to the "projection space" P_i by the line integral of attenuation along the x-ray path u traversing through the scene at a gantry position θ_i:

$$DR(I(x,t);\theta_i) = \int_u I(x,t)\mathcal{R}_{\theta_i}(t)du \quad (4)$$

where $\mathcal{R}_\theta(t)$ is the time-varying rotation matrix describing the gantry rotation with angle θ_i.

Spatiotemporal attenuation maps $I(x,t)$ in Eqn. 4 were obtained from the SIREN SDF ($SDF(x,t,k;\boldsymbol{g})$) by first converting the SDF to an occupancy map $\mathcal{E}$ (where negative SDF value means the pixel is occupied) and then multiplying $\mathcal{E}$ with the object's attenuation $a(k)$ (Eqn. 5). $a(k)$ was approximated as the median attenuation of the k^{th} segmented object in the FBP image.

$$I(x,t) = \sum_k a(k) \times \mathcal{E}(SDF(x,t,k;\boldsymbol{g})) \quad (5)$$

Combining Eqn. 3-5, this approach enables the loss $\mathcal{L}_P$ to be defined as a differentiable function of $\boldsymbol{g}$**.** Additional loss terms – $\mathcal{L}_E$ (Eikonal constraint), $\mathcal{L}_{TVS}$ and $\mathcal{L}_{TVT}$ (total variances computed as the gradient of the SDF with respect to x and t) were added to constrain the result, leading to a total loss $\mathcal{L} = \mathcal{L}_P + \lambda_1\mathcal{L}_E + \lambda_2\mathcal{L}_{TVS} + \lambda_3\mathcal{L}_{TVT}$ where λ_1 to λ_3 serve as regularization weighting parameters.

Step 6: After optimization, the result $SDF(x,t,k;\boldsymbol{g}^*)$ was convert to the motion-corrected image $I^*(x,t)$ (i.e., the final product of NeuralCT reconstruction) via Eqn. 5.

B. NeuralCT with Intensity-based Segmentation

As outlined above, a key step in the NeuralCT framework is the initialization described in Step 4 where SIREN $\boldsymbol{g}$ aims to approximate the SDF map of the scene of interest. Gupta et al. [1] used a Gaussian Mixture Model (GMM) [9] that was solely based on the intensity histogram in $I_{FBP}(x,t)$. GMM fits a finite number of Gaussian distributions to the intensity histogram and assigns pixels with intensity from the same Gaussian distribution as the same class. After excluding the background, the top k classes with the most pixel were used to identify foreground objects. As shown in [1], this segmentation method, hereafter referred to as Seg_{GMM}, worked well in the scenes with a single foreground object – as it readily separates the object from the background, despite motion artifacts.

C. NeuralCT with Spatially Aware Segmentation

However, when Seg_{GMM} is applied to a scene with multiple moving objects, each with different attenuations, it becomes difficult to differentiate objects based solely on the intensity distribution. **Fig. 2** shows a failure of Seg_{GMM} when analyzing the FBP reconstruction of two moving dots with two different attenuations (top = 0.7, bottom = 0.2). Based on the histogram, GMM identifies the top two intensity values with the most

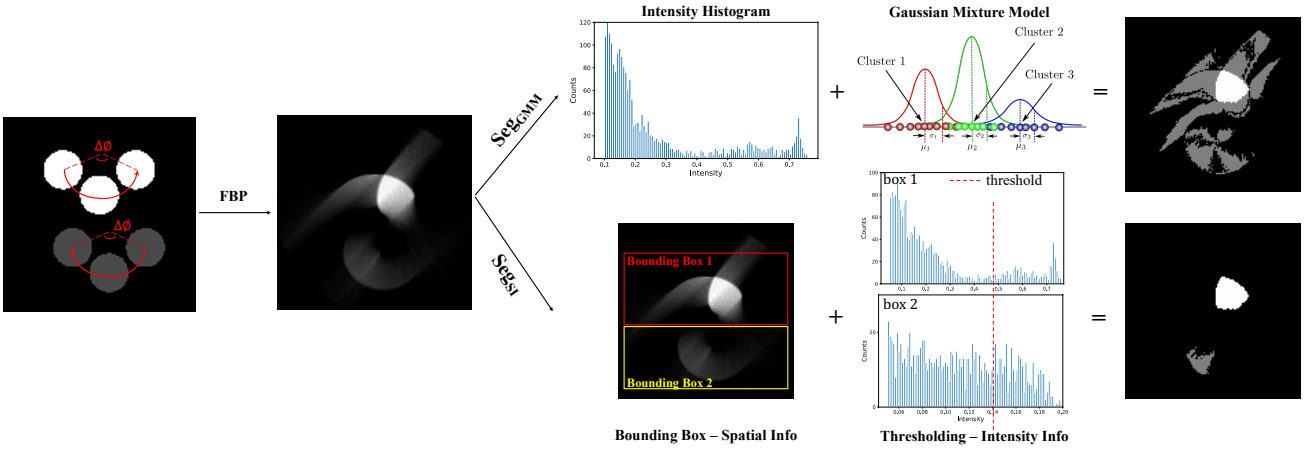


Fig.2. Two different object segmentation approaches used in NeuralCT. The first image shows the ground truth motion of two dots (top intensity = 0.7, moving from left to right, bottom intensity = 0.2, moving from right to left). ΔØ is the angular displacement per gantry rotation. *Seg$_{GMM}$*: Gaussian mixture model incorrectly assigned the motion artifacts and the bottom dot as the same class. *Seg$_{SI}$*: Spatially aware segmentation utilized both spatial info (by setting bounding box in this example) and intensity info (thresholding) and led to correct detection of both top and bottom dots.

pixel counts from the distribution. However, this results in incorrect labeling of two dots as one bright foreground class and a second dimmer object spread throughout the image.

The core contribution of this study is to improve NeuralCT performance in the case of multiple intensity objects by resolving this segmentation error. We did so by applying a spatially-aware segmentation approach *Seg$_{SI}$* which incorporated both the Spatial (S) and Intensity (I) information of each object in the FBP image. *Seg$_{SI}$* aims to assign different classes to objects with different spatial positions and be aware of the different intensities between the real object and the motion artifacts. This can be achieved using various approaches such as Region-Of-Interest (ROI) definition plus thresholding or data-driven methods (e.g., deep learning segmentation). Here, we focus on demonstrating that this improvement in segmentation leads to improvements in NeuralCT performance. In **Fig. 2**, we show a simple approach to add spatial information. Specifically, bounding boxes were used to guide thresholding-based segmentation. Each bounding box was defined to only contain one moving dot such that we assigned one individual class to each box. In the box, we defined an intensity threshold = $\gamma \times I_{max}$ where I_{max} is the maximum intensity in the scene in each box to capture the real object. $\gamma = 0.7$ was set empirically.

Given that artifacts will always be present in the initial FBP images, we *highlight* here that the goal with this new segmentation is not to achieve a perfect segmentation but rather to provide a segmentation that is not so poor that it precludes improvement by the NeuralCT framework. We hypothesize that by improving the initial segmentation, we will avoid overt failures and improve image quality obtained with NeuralCT.

III. Experiments and Results

We performed two experiments to demonstrate the impact of the segmentation on the subsequent result and evaluate the improvement associated with our new segmentation approach.

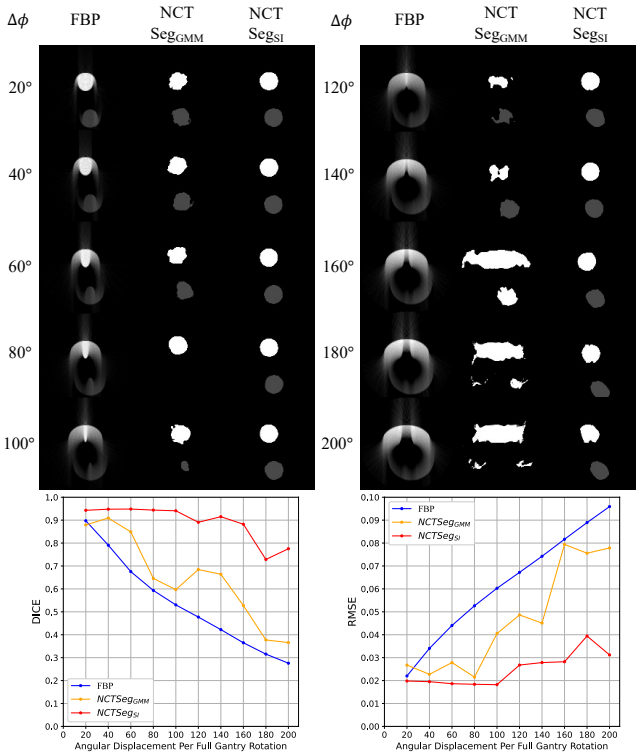


Fig. 3. NCT-*Seg$_{SI}$* accurately depicts the moving dots with two attenuations and high angular displacements. FBP suffers from motion artifacts for all ΔØ; NCT-*Seg$_{GMM}$* failed the reconstructions for high ΔØ (>60); Only NCT-*Seg$_{SI}$* maintained high-quality motion-corrected reconstruction for all ΔØ with higher DICE and lower RMSE when compared with FBP and NCT-*Seg$_{GMM}$*. ΔØ = angular displacement per gantry rotation.

Experiment 1: Angular Displacement of Two Dots

As shown in **Fig. 2**, two circular dots which translate with angular displacement ΔØ per full gantry rotation were imaged. The two dots had different attenuation levels (top = 0.7, bottom = 0.2), mimicking the difference between contrast-enhanced vessels and the myocardium in cardiac CT. Background = 0. The image resolution was set to 128×128 and a parallel beam

CT geometry was used with 720 gantry positions per rotation. Two NeuralCT frameworks were then evaluated – intensity-based segmentation (NCT-Seg_{GMM}) and spatially aware segmentation (NCT-Seg_{SI}) – across a range of ΔØ (from 20° to 200° per gantry rotation). Performance was evaluated using root-mean-square-error (RMSE) and DICE coefficients relative to the ground truth image.

As shown by the images and metrics in **Fig. 3**, FBP motion artifacts increased at higher ΔØ. Reconstruction with NCT-Seg_{GMM} was limited when ΔØ > 60°. In contrast, NCT-Seg_{SI} maintained high-quality motion-corrected reconstructions for all ΔØ and achieved low RSME (<0.028) and high (>0.89) DICE for ΔØ up to 160°.

Experiment 2: Complex Deformation of Letters

In experiment 2, we evaluated the ability of NCT-Seg_{SI} to improve reconstruction of scenes with complex topological changes. As shown in **Fig. 4**, in this case, we simulated CT imaging during transformation of letters. The top letter transformed from "A" to "B" to "A" (attenuation = 0.7) while the bottom letter transformed from "B" to "A" to "B" (attenuation = 0.4). NCT-Seg_{SI} (red line) significantly reduced the severity of artifacts observed with FBP (blue) and NCT-Seg_{GMM} (orange), especially during transformation periods (2^{nd}-3^{rd} and 5^{th}-6^{th} columns). Quantitatively, median RMSE of NCT-Seg_{SI} (median = 0.050 [0.042-0.061]) was significantly lower ($p<0.05$) than NCT-Seg_{GMM} (0.090 [0.076-0.096]) and FBP (0.069 [0.047-0.085]). Median DICE for NCT-Seg_{SI} (0.89 [0.86-0.93]) was significantly higher ($p<0.05$) than NCT-Seg_{GMM} (0.72 [0.69-0.76]) and FBP (0.72 [0.64-0.87]). Lastly, NCT-Seg_{SI} increased the percentage of the frames with RMSE< 0.05 (NCT-Seg_{SI}: 45.7%, NCT-Seg_{GMM}: 0%, FBP: 28.0%) as well as with DICE > 0.85 (NCT-Seg_{SI}: 89.6%, NCT-Seg_{GMM}: 0%, FBP: 27.6%).

IV. Summary

Reconstruction of moving scenes using a neural implicit representation-based framework (NeuralCT) can improve image quality the need for a prior motion model or estimation. Here, we show that when imaging scenes with multiple moving objects, performance of NeuralCT can be limited by poor segmentation of motion-corrupted FBP images. Using a spatially aware object segmentation method that incorporates both spatial and intensity information can result in an NeuralCT solution which maintains high reconstruction performance for moving objects with multiple attenuation levels despite high angular motion and complex topological changes.

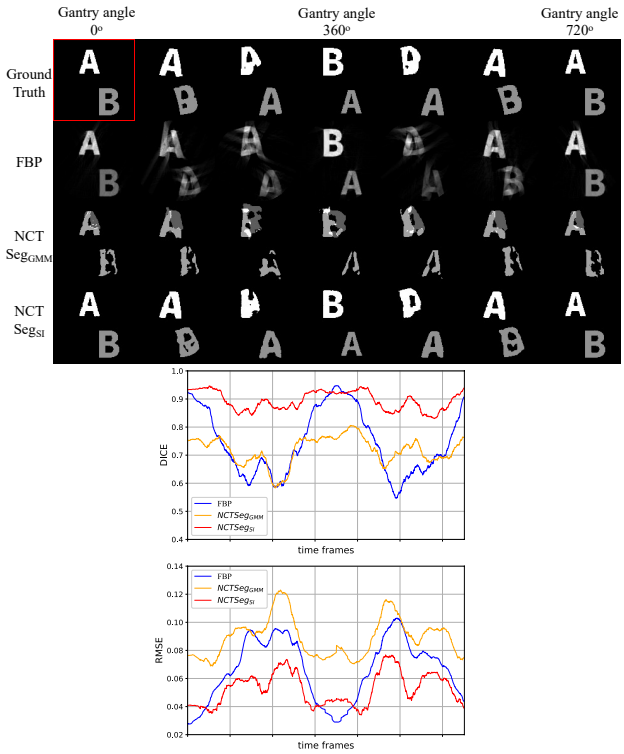


Fig. 4. NCT-Seg_{SI} accurately depicts the complex topological change with multiple attenuations. The ground truth image (red box) contains two letters that transform over two gantry rotations. Seven frames including three stationary phases (column 1,4, 7) and four intermediate transformation phases (column 2-3, 5-6) are displayed. Both reconstructed images and the quantitative metrics indicates that NCT-Seg_{SI} improved the imaging of a complex scene.

References

[1] K. Gupta, B. Colvert, and F. Contijoch, "Neural Computed Tomography,", Jan. 2022, Available: http://arxiv.org/abs/2201.06574

[2] Z. Chen *et al.*, "Precise measurement of coronary stenosis diameter with CCTA using CT number calibration," *Med. Phys.*, vol. 46, no. 12, pp. 5514–5527, Dec. 2019, doi: 10.1002/mp.13862.

[3] T. Lossau *et al.*, "Motion estimation and correction in cardiac CT angiography images using convolutional neural networks," *Comput. Med. Imaging Graph.*, vol. 76, p. 101640, Sep. 2019, doi: 10.1016/j.compmedimag.2019.06.001.

[4] Y. Ko, S. Moon, J. Baek, and H. Shim, "Rigid and non-rigid motion artifact reduction in X-ray CT using attention module," *Med. Image Anal.*, vol. 67, p. 101883, Jan. 2021, doi: 10.1016/j.media.2020.101883.

[5] P. Wang, L. Liu, Y. Liu, C. Theobalt, T. Komura, and W. Wang, "NeuS: Learning Neural Implicit Surfaces by Volume Rendering for Multi-view Reconstruction," Dec. 2021 Available: http://arxiv.org/abs/2106.10689

[6] L. Shen, J. Pauly, and L. Xing, "NeRP: Implicit Neural Representation Learning with Prior Embedding for Sparsely Sampled Image Reconstruction," Aug. 2021. Available: http://arxiv.org/abs/2108.10991

[7] Q. Wu *et al.*, "An Arbitrary Scale Super-Resolution Approach for 3-Dimensional Magnetic Resonance Image using Implicit Neural Representation," Oct. 2021. Available: http://arxiv.org/abs/2110.14476

[8] V. Sitzmann *et al.*, "Implicit Neural Representations with Periodic Activation Functions," Jun. 2020. Available: http://arxiv.org/abs/2006.09661

[9] D. Reynolds *et al.*, "Gaussian Mixture Models," *Encyclopedia of Biometrics*, 2009, pp. 659-664. doi: 10.1007/978-0-387-73003-5_196.

Photon-Counting X-ray CT Perfusion Imaging in Animal Models of Cancer

Darin P. Clark, Alex J. Allphin, Yvonne M. Mowery, and Cristian T. Badea

***Abstract*—Photon-counting detector (PCD) CT promises to improve routine CT imaging applications with higher spatial resolution, lower levels of noise at fixed dose, and improved image contrast while providing spectral information with every scan. We propose and demonstrate a novel application of PCD imaging in a preclinical model of head and neck squamous cell carcinoma: spectral perfusion imaging of cancer. To handle the high dimensionality of our data set (3D volumes at 12 perfusion time points times 4 energy thresholds), we update our previously proposed multi-channel iterative reconstruction algorithm to handle the perfusion reconstruction problem, and we propose an extension which adds patch-based singular value thresholding (pSVT) along the perfusion dimension. Adding pSVT reduces noise by an additional 45% relative to our standard algorithm, which itself reduces noise by 2-7 times relative to analytical reconstruction. Preliminary analysis suggests that the addition of pSVT does not negatively impact material decomposition accuracy or image spatial resolution. Notable weaknesses of this preliminary study include relatively high contrast agent dose (0.5 mL *ISOVUE-370* over 10 seconds), ionizing radiation dose (~570 mGy), and computation time (2.9 hours, no pSVT; 11 hours with pSVT); however, following from our past work, our reconstruction algorithm may be an ideal source of training labels for supervised deep learning applied to computationally cheap analytical reconstructions.**

***Index Terms*—X-ray CT, photon counting, multi-energy CT, perfusion, preclinical, head and neck cancer**

I. Introduction

Photon-counting detector (PCD) technology brings significant advancements to traditional X-ray CT imaging applications, including higher spatial resolution, reduced electronic noise, and improved image contrast [1, 2]. Furthermore, replacing traditional energy-integrating detectors with PCDs in routine imaging applications will intrinsically provide spectral information for artifact reduction and post-processing. Fully exploiting this spectral information will enable novel applications of multi-contrast imaging and even functional targeting of CT contrast [3, 4]. Practically, a new generation of data processing algorithms is required to take full advantage of spectral information at current dose levels because of the increased noise associated with photon binning and multi-material decomposition.

In this work, we demonstrate a new application of PCCT, spectral perfusion imaging of cancer, as a future alternative to contrast-enhanced CT imaging protocols like those used to stage head and neck squamous cell carcinoma (HNSCC) in humans [5]. Specifically, we present a preclinical experiment using a syngeneic transplant oral cavity model of HNSCC (MOC2) [6] initiated in C57BL/6J mice by implantation of cells into the buccal mucosa. To handle high levels of noise associated with PCCT and preclinical imaging, we extend our previously proposed iterative reconstruction algorithm from spectral [7] and cardiac [8] applications to multi-channel perfusion reconstruction. New to this work we employ an updated version of our rank-sparse kernel regression (RSKR, [7]) algorithm which includes localized noise estimation for handling spatial variance and patch-based singular value thresholding [9] for improved temporal redundancy.

II. Materials and Methods

A. Data Acquisition

Our photon counting micro-CT system uses a SANTIS 1604 detector (DECTRIS Ltd.). The detector is constructed with a 1 mm thick CdTe sensor, 150-μm isotropic pixels (1035x257), and four independent energy thresholds, here set to 20, 25, 34, and 50 keV to bracket the K-edge of iodine. The projection data set was acquired with a G297 X-ray tube (Varian Medical Systems; fs = 0.3/0.8 mm; tungsten rotating anode; filtration: 0.1 mm Cu; 80 kVp, 2.5 mA, 90 ms exposure / projection). The source-to-detector and source-to-object distances were 821 mm and 674 mm, respectively. To minimize ring artifacts in our reconstructions, we scanned using a helical trajectory with 3 rotations (45 seconds/rotation) and 1.25 cm of total translation during scanning, acquiring a total of 1500 projections per threshold. This acquisition protocol was repeated 4 times, consecutively, yielding a total of 6000 projections over 540 seconds. The absorbed radiation dose associated with imaging was ~570 mGy. To image perfusion, we injected 0.5 ml of *ISOVUE-370* (Bracco Diagnostic Inc.) over 10 secs via a tail vein catheter using a computer-controlled injector (*Nexus 3000*; Chemyx Inc.), starting 22.5 seconds after beginning the scan.

The animal scan conducted for this work followed protocols approved by the Duke University Institutional Animal Care and Use Committee. The animal model was a syngeneic transplant oral cavity squamous cell carcinoma model (MOC2) [6] initiated in a C57BL/6J mouse by implantation of 30,000 cells

All work was performed at the Quantitative Imaging and Analysis Lab of Duke University with support from the NIH National Cancer Institute (U24 CA220245) and the National Institute on Aging (RF1AG070149-01).

D. P. Clark, A. J. Allphin, and C. T. Badea are with the Quantitative Imaging and Analysis Lab, Duke University, Dept. of Radiology, Durham, NC 27710 USA (e-mail: cristian.badea@duke.edu). Y. M. Mowery is with the Dept. of Radiation Oncology at Duke University, Durham, NC 27710 USA.

into the buccal mucosa (median time to tumor onset: 2.2 weeks). The model exhibits metastasis to the cervical lymph nodes, similar to the natural history of human HNSCC. Imaging was performed when caliper measurements of the primary tumor fell in the range of 50-75 mm^3.

B. Reconstruction

To handle the channels of photon-counting perfusion data (3D volumes with n_x voxels, n_t time points, and n_e energy thresholds), we organize the reconstructions into a 3D array:

$$X \coloneqq \begin{bmatrix} [\mathbf{x}_{t=1} & \mathbf{x}_{t=2} & \cdots & \mathbf{x}_{t=n_t}]_{e=1} \\ & & \cdots & \\ [\mathbf{x}_{t=1} & \mathbf{x}_{t=2} & \cdots & \mathbf{x}_{t=n_t}]_{e=n_e} \end{bmatrix}, \quad (1)$$

where, e.g., $\mathbf{x}_{t=1,e=2}$ denotes a column vector of the reconstruction at perfusion time point 1 and energy threshold 2. Each energy threshold slice corresponds to a 2D matrix. Following from this notation (array, A; matrix, A; vector, $\mathbf{a}$; scalar, a; index, a), algebraic reconstruction minimizes the reconstruction error relative to log-transformed projection data, $\mathbf{y}$, under an L2 norm penalty applied separately to each channel:

$$X = \underset{X}{\arg\min} \frac{1}{2}\|RX - \mathbf{y}\|_Q^2 \coloneqq \underset{X}{\arg\min} \tfrac{1}{2}\sum_{t=1}^{n_t}\sum_{e=1}^{n_e}\left[\left(R\mathbf{x}_{t,e} - \mathbf{y}\right)^T Q_{t,e}\left(R\mathbf{x}_{t,e} - \mathbf{y}\right)\right]. \quad (2)$$

R is the system-specific projection matrix, which is universal for all channels when working with co-registered photon counting projection data. The weight array Q is constructed from weight matrices, $Q_{t,e}$, the product of diagonal matrices indicating the target energy threshold and perfusion phase.

We use multi-channel regularization to improve reconstruction performance. Specifically, we perform iterative reconstruction using the split Bregman optimization framework [10], and we employ an extended version of our RSKR regularizer to enforce gradient sparsity and data consistency between channels [7]. Our iterative reconstruction is governed by the following objective function:

$$X = \underset{X}{\arg\min} \tfrac{1}{2}\|RX - \mathbf{y}\|_Q^2 + \lambda_{*1}\left\|X_{n_x n_t \times n_e}\right\|_* + \lambda_S\left\|X_{n_x n_t \times n_e}\right\|_{BTV} + \lambda_{*2}\sum_{p=1}^{n_p}\left\|P_p\left(X_{n_x n_e \times n_t}\right)\right\|_*. \quad (3)$$

$X_{n_x n_t \times n_e}$ denotes the unfolding of the array X into a matrix with $n_x n_t$ rows and n_e columns ($X_{n_x n_e \times n_t}$: $n_x n_e$ rows, n_t columns). These unfoldings and their corresponding regularization terms enable penalization of the global spectral rank ($\|.\|_*$, nuclear norm; λ_{*1}), the spectrally joint intensity gradient sparsity (BTV: bilateral total variation, [7]; λ_S), and the patch-wise temporal rank (λ_{*2}; n_p total patches; P_p: patch extraction operator for patch p). Within the split Bregman framework, it is feasible to split each regularization term into a separate regularization sub-step (each with its own set of auxiliary variables, D and V) to incrementally solve this objective function. Practically, however, the computational cost of the data fidelity updates (Fig. 1, step 6; here, evaluated on 48 volumes) and the often slow rate of convergence of shrinkage methods (singular value thresholding to reduce the nuclear norm, [11]) make this approach intractable.

To reduce the number of global Bregman iterations (data fidelity update steps) required to achieve convergence (here, 3), we address all of the regularization terms in a single regularization step (step 4). This one step is itself solved by a series of inner Bregman iterations (3-4), which perform image-domain denoising with our RSKR algorithm (steps 4a-4i). RSKR jointly minimizes the spectral rank (λ_{*1} term) and the spectral gradient sparsity (λ_S term) by performing joint bilateral filtration on singular vectors computed along the energy dimension (step 4b). As in our previous implementations of RSKR, we use the median absolute deviation (MAD, [12]) to estimate the noise level in each channel. These noise estimates are used to weight the singular value decomposition such that high singular values correspond to singular vectors with proportionately high signal-to-noise ratios (steps 4a, 4b). As in past work, a similar noise estimation strategy is used to calibrate the joint bilateral filter parameters (jBF, [7]; step 4e), reducing the number of free parameters which must be determined by the user to achieve robust performance.

New to this work, we have improved our implementation of BF to use local noise estimates (vs. one global noise estimate per energy) to deal with spatially variant noise levels in CT data. Our GPU-based implementation of jBF now computes the MAD of image gradients in 32x32 patches with a stride of 16 voxels in axial planes. Overlapping estimates are averaged to prevent sharp transitions in denoising performance, while the noise level is estimated in axial planes, rather than volumetrically, to improve computational efficiency.

Also new to this work, we perform patch-based singular value thresholding (pSVT, [9]) along the time dimension of the singular vectors. Notably, pSVT is applied to the singular values of a second singular value decomposition performed on $3^3 \times n_t$ patch matrices extracted from L (step 4f). Given a vector of these singular values, $\boldsymbol{\varepsilon}$, computed from patch matrix p, thresholding is applied to the singular values (indexed by i; evaluated separately per energy) as follows:

$$\varepsilon'_{p,i,e} = \max\left(\varepsilon_{p,i,e} - \tau_{i,e}\left(\frac{\varepsilon_{p,i,e}}{\varepsilon_{p,0,e}}\right)^{\nu-1}, 0\right). \quad (4)$$

These thresholded singular values, $\boldsymbol{\varepsilon}'$, are then used to reverse the SVD for each patch matrix, and the central voxel for each time point is stored as the regularized output.

Previously, we have used kernel smoothing to overcome inconsistent denoising performance at high contrast edges when denoising with jBF [13]. Now, we apply pSVT ($\nu = 0.7$). pSVT similarly improves image consistency at edges, but also improves consistency along the time dimension and may better preserve spatial resolution. pSVT addresses the final cost term of the Objective function (Eq. 3; Fig. 1; λ_{*2}). Like with jBF, MAD-based noise estimates are used to calibrate the threshold values, $\tau_{i,e}$, used during each inner Bregman iteration. Specifically, the SVD is performed on zero-mean Gaussian noise patch matrices with noise standard deviations matching those measured globally in the corresponding energy channel. Here, the average singular values from 125 noise realizations were used to set the threshold values per energy.

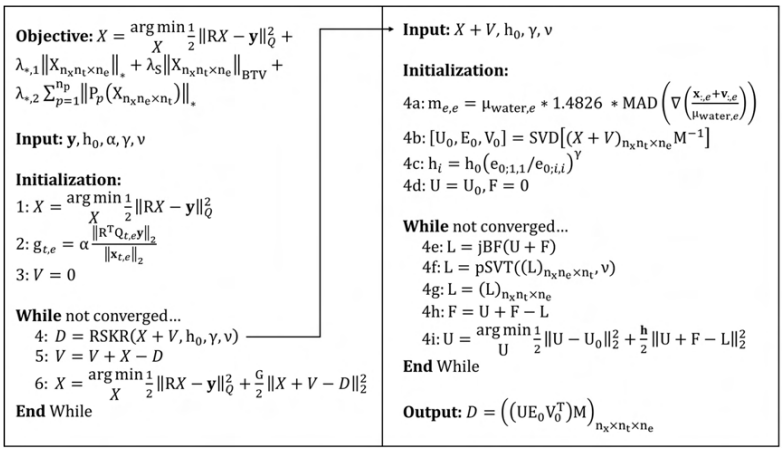


Fig. 1. The split Bregman method applied for iterative reconstruction of photon-counting CT perfusion data. Starting from the left column, an Initialization procedure estimates the reconstructed data, X, and the regularization parameters, G (here $\alpha = 0.007$, $h_0 = 1.5$, $\gamma = 0.5$, $\nu = 0.7$). Outer Bregman iterations approximately solve the Objective function through regularization (step 4, D), residual (5, V), and data fidelity (6, X) updates. More details on the inner Bregman iterations for regularization are provided in the text (steps 4a-4i). Six iterations of the Bi-CGSTAB(2) solver (algorithm 3.1 in [14]) are used to update X at each time point and energy during Initialization (step 1) and for each data fidelity update (step 6). Additional notation: $\mu_{water,e}$, the expected attenuation of water at energy threshold e; ∇, compute high-pass only component of the 3D Haar wavelet transform.; i, singular value, vector index (n_e total); $\boldsymbol{x}_{:,e} + \boldsymbol{v}_{:,e}$, time indices assumed equivalent for noise estimation.

C. Spectral Processing and Analysis

We compare three sets of reconstruction results: analytical reconstructions of each time point and energy (WFBP algorithm, [15]), iterative reconstruction without pSVT (Fig. 1, steps 4f and 4g skipped), and iterative reconstruction with pSVT (Fig. 1, all steps). Notably, the reference iterative reconstruction (no pSVT) is similar to spectral reconstruction results we have quantitatively validated in previous work [7] (with the exception of localized noise estimation now performed during jBF). Following reconstruction, we perform non-negative material decomposition using sensitivity matrix inversion and a sub-space projection with iodine (red), photoelectric (PE, green), and Compton scattering (CS, gray) basis functions [3]. We characterize the convergence of our iterative reconstructions (Fig. 1, step 6) by measuring the relative error after the final outer Bregman iteration for each perfusion phase and energy threshold (generically, $\|\mathbf{Ax} - \mathbf{b}\|_2 / \|\mathbf{b}\|_2$; $\mathbf{Ax}$: variable terms; $\mathbf{b}$: constant terms). We compare our reconstruction results through standard deviation measurements, residual images, material decomposition errors in reference solutions, and through modulation transfer function measurements.

D. Computation

All reconstruction and denoising operations were performed with our custom multi-channel GPU-based reconstruction toolkit [16]. For the mouse data set, a total of 6000 projections were acquired over 12, 360° rotations. The temporal basis functions (*rect* functions, $Q_{t,e}$) assigned equal weight to every projection in each subset of 500 projections, allowing reconstruction of 12 non-overlapping perfusion phases spanning 45 seconds each. Combined, 12 perfusion time points at 4 four energy thresholds were reconstructed (48 total volumes) with a volume size of 360x360x384 voxels and 125-micron, isotropic voxels. For the analytical reconstruction results, a ramp filter was used. For algebraic reconstruction (Fig. 1, steps 1 and 6) six iterations of the Bi-CGSTAB(2) solver were used [14]. Reconstructions were performed on an Ubuntu Linux workstation (v18.04) with four NVIDIA RTX8000 GPUs, 256 GB of system RAM, and two Intel Xeon W-2295 CPUs. Iterative reconstruction of all 48 volumes took 2.9 hours without pSVT (3 outer, 3 inner Bregman iterations) and 11 hours with pSVT (3 outer, 4 inner Bregman iterations).

III. Results

Fig. 2 summarizes the perfusion reconstruction results along the energy dimension. Specifically, analytical WFBP, iterative reconstruction without pSVT, and iterative reconstruction with pSVT results are compared at perfusion phase 6. The final data fidelity update step of the outer Bregman iterations (Fig. 1, step 6) converged to within 4% error with pSVT and to within 3% error without pSVT for each individual energy threshold and perfusion phase, suggesting robust algorithm convergence. The noise standard deviation was measured in a water vial included in the scan. As expected, the noise level increases from 123 to 362 HU with increasing energy threshold (reduced photon counts). Our standard iterative reconstruction method (no pSVT), which includes data adaptation (Fig. 1, steps 2 and 4a), reduces the noise level by 2 to 7 times and equalizes the noise standard deviation across energy channels. Even with identical reconstruction hyperparameters (Fig. 1, "Input"), adding pSVT to the reconstruction further reduced noise by ~45%. Iterative reconstruction with and without pSVT yielded similar spatial resolution (20 keV energy threshold; 50% cutoff of the MTF: ~2.5 lp/mm in both cases).

Fig. 3 summarizes the reconstruction results along the prefusion dimension ("P6": perfusion phase 6), comparing the phase of minimum enhancement (phase 1) with phases of intermediate (2) and peak enhancement (6). The prefusion results are shown as material decomposition maps to emphasize the value of energy information in quantitatively separating soft tissue (CS), calcium (PE), and iodine (I). Additional CT images demonstrate the value of iterative reconstruction in enforcing redundancy between energy channels to counteract noise amplification associated with material decomposition and analytical reconstruction. Based on material calibration vials included in the scan (12 mg/mL I; 1.0 × water for PE, CS), material decomposition with and without pSVT yields similar mean material decomposition accuracy, but pSVT has lower noise (pSVT, no pSVT mean ± standard deviation: 11.67 ± 0.18, 12.08 ± 0.52 I; 0.68 ± 0.11, 0.68 ± 0.32 PE; 0.93 ± 0.03, 0.93 ± 0.08 CS). This is in stark contrast to the WFBP results, which are less accurate for I and CS, and have an order of magnitude larger standard deviations. Yellow arrows in Fig. 3 highlight an artifact pattern that is removed with the addition of pSVT; however, blue arrows point to a cervical lymph node

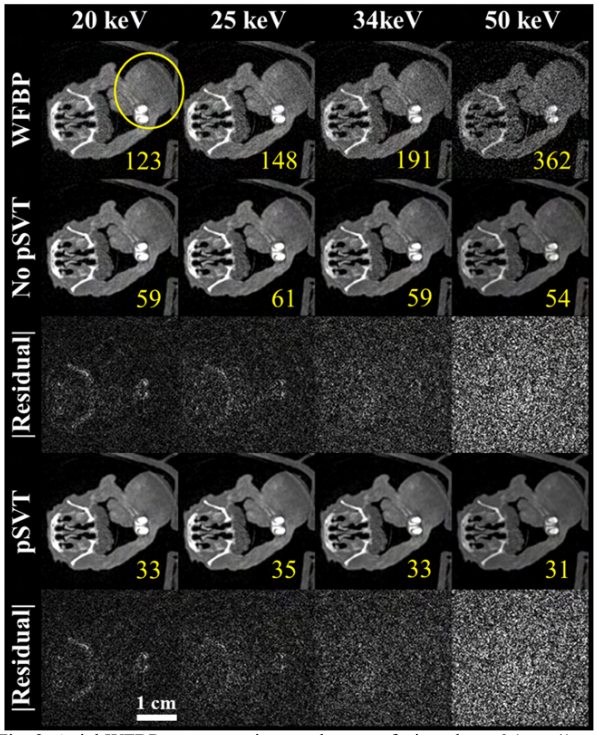


Fig. 2. Axial WFBP reconstruction results at perfusion phase 6 (row 1) are compared with matching iterative reconstruction results excluding ("No pSVT", row 2) and including ("pSVT", row 4) the pSVT regularizer. A yellow circle indicates the location of the primary tumor. Yellow text indicates noise standard deviation values measured in water for each energy threshold and reconstruction algorithm (HU; water vial not shown). CT window: [0.0,0.80] cm^{-1}. Absolute residual window: [0.0,0.16] cm^{-1}.

metastases where the iodine map magnitude appears reduced when pSVT is added to the reconstruction. Further investigation is needed to characterize the trade-offs between spatial, spectral, and temporal resolution as a function of pSVT hyperparameters.

IV. Discussion and Conclusions

In this work, we have demonstrated the future of perfusion imaging with X-ray CT and PCD technology. Given the increases in noise associated with photon binning, data acquisition and hardware constraints, and spectral post-processing, a new class of algorithms will be required which take full advantage of the inherent structure in multi-channel CT data. We repurpose our existing multi-channel iterative reconstruction algorithm for compatibility with time- and energy-resolved perfusion data and propose a novel extension with pSVT. pSVT removes additional noise by exploiting data consistency along the temporal perfusion dimension.

The current work has several limitations. The imaging dose and contrast agent dose are non-trivial and unsuitable for longitudinal studies. Furthermore, the computation time associated with batched SVD operations for pSVT is too significant for routine use. In previous work [17], we have demonstrated that similar challenges can be overcome for multi-channel cardiac CT reconstruction using deep learning. Extending this approach, we propose to use our current reconstruction algorithms and data acquisition protocol to generate high quality reference data. We could then artificially degrade these data sets and use a supervised learning paradigm to learn to map computationally inexpensive analytical reconstructions to our high-fidelity iterative reconstructions.

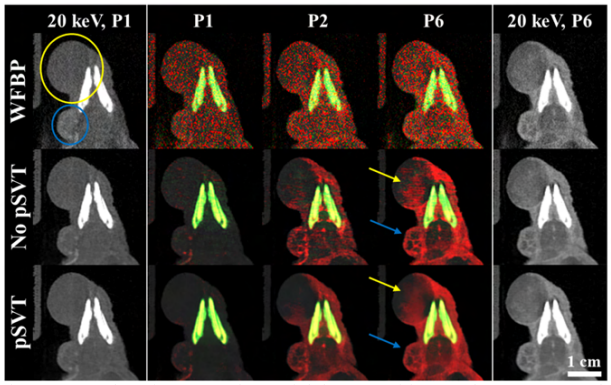


Fig. 3. Coronal CT images and material decomposition results are compared between WFBP reconstruction (row 1) and regularized iterative reconstruction excluding ("No pSVT", row 2) and including ("pSVT", row 3) pSVT. A yellow circle indicates the location of the primary tumor, while a blue circle indicates the location of a metastatic tumor. Results are shown for perfusion phases 1, 2, and 6 ("P6"). Arrows are referred to in the text. CT window: [0.0,0.80] cm^{-1}. Iodine (red) window: [1.0,6.0] mg/mL. PE (green) window: [1.0,15.0] × water. CS (gray) window: [0.0,8.0] × water.

References

[1] M. J. Willemink, M. Persson, A. Pourmorteza, N. J. Pelc, and D. Fleischmann, "Photon-counting CT: technical principles and clinical prospects," *Radiology,* vol. 289, no. 2, pp. 293-312, 2018.

[2] R. Gutjahr *et al.*, "Human Imaging With Photon Counting–Based Computed Tomography at Clinical Dose Levels: Contrast-to-Noise Ratio and Cadaver Studies," *Investigative Radiology,* vol. 51, no. 7, pp. 421-429, 2016.

[3] C. T. Badea, D. P. Clark, M. Holbrook, M. Srivastava, Y. Mowery, and K. B. Ghaghada, "Functional imaging of tumor vasculature using iodine and gadolinium-based nanoparticle contrast agents: a comparison of spectral micro-CT using energy integrating and photon counting detectors," *Physics in medicine and biology,* 2019.

[4] J. R. Ashton, J. L. West, and C. T. Badea, "In vivo small animal micro-CT using nanoparticle contrast agents," *Frontiers in pharmacology,* vol. 6, p. 256, 2015.

[5] H. Lewis-Jones, S. Colley, and D. Gibson, "Imaging in head and neck cancer: United Kingdom national multidisciplinary guidelines," *The Journal of Laryngology & Otology,* vol. 130, no. S2, pp. S28-S31, 2016.

[6] N. P. Judd *et al.*, "ERK1/2 regulation of CD44 modulates oral cancer aggressiveness," *Cancer research,* vol. 72, no. 1, pp. 365-374, 2012.

[7] D. P. Clark and C. T. Badea, "Hybrid spectral CT reconstruction," *PLOS ONE,* vol. 12, no. 7, p. e0180324, 2017.

[8] D. Clark, M. Holbrook, C. Lee, and C. Badea, "Photon-counting cine-cardiac CT in the mouse," *PLoS ONE,* vol. 14, no. 9, p. e0218417, 2019.

[9] K. Kim *et al.*, "Sparse-view spectral CT reconstruction using spectral patch-based low-rank penalty," *IEEE Transactions on Medical Imaging,* vol. 34, no. 3, pp. 748-760, 2015.

[10] H. Gao, H. Yu, S. Osher, and G. Wang, "Multi-energy CT based on a prior rank, intensity and sparsity model (PRISM)," *Inverse Problems,* vol. 27, no. 11, p. 115012, 2011.

[11] E. J. Candès, X. Li, Y. Ma, and J. Wright, "Robust principal component analysis?," *Journal of the ACM (JACM),* vol. 58, no. 3, p. 11, 2011.

[12] D. L. Donoho and I. M. Johnstone, "Adapting to unknown smoothness via wavelet shrinkage," *Journal of the American Statistical Association,* vol. 90, no. 432, pp. 1200-1224, 1995.

[13] D. P. Clark and C. T. Badea, "Spectral diffusion: an algorithm for robust material decomposition of spectral CT data," *Physics in Medicine and Biology,* vol. 59, no. 21, p. 6445, 2014.

[14] G. L. Sleijpen and M. B. Van Gijzen, "Exploiting BiCGstab (ℓ) strategies to induce dimension reduction," *SIAM journal on scientific computing,* vol. 32, no. 5, pp. 2687-2709, 2010.

[15] K. Stierstorfer, A. Rauscher, J. Boese, H. Bruder, S. Schaller, and T. Flohr, "Weighted FBP—a simple approximate 3D FBP algorithm for multislice spiral CT with good dose usage for arbitrary pitch," *Physics in Medicine & Biology,* vol. 49, no. 11, p. 2209, 2004.

[16] D. P. Clark and C. T. Badea, "GPU-Based Tools for Multi-Channel X-ray CT Reconstruction," in *The Fifth International Conference on Image Formation in X-Ray Computed Tomography*, Salt Lake City, Utah, F. Noo, Ed., 2018.

[17] D. P. Clark and C. T. Badea, "Convolutional regularization methods for 4D, x-ray CT reconstruction," presented at the Proceedings of SPIE Medical Imaging, 2019 Paper #10948-81.

Undersampled Dynamic Tomography with separated spatial and temporal regularization

Xiufa Cao, Yinghui Zhang, Ran An, Hongwei Li*

***Abstract*—Dynamic tomography finds its usage in certain important applications. The reconstruction problem could be cast in the Bayes framework as a time series analysis problem, such that the Kalman filter(KF) could come into play. A series of drawback of such an approach lies in its high computational and storage complexity. Dimension reduction Kalman filter (DR-KF) method has been proposed in the literature to relieve such a pain. However, our tests show that DR-KF results in heavy ringing artifacts. To solve this dilemma, in this paper, we propose to approximate the regularization term involving the precision matrix with a spatial regularization term plus a temporal regularization term, such that the space and time complexity are greatly reduced. Besides, to get the original Kalman filter into play, we propose a blocked KF for backward smoothing, i.e. split the reconstructed image slices into small overlapping blocks, and the KF is applied on each block. The blocked KF has much smaller computational and storage complexity, and fits well for parallel computations. Numerical experiments show that the proposed approach achieves better reconstructions while calling for much smaller computational resources.**

***Index Terms*—Dynamic X-ray tomography, Kalman filter, dimension reduction, ringing artifacts, TV regularization.**

I. Introduction

COMPUTED tomography (CT), as a noninvasive method, is widely used in many applications to reveal the inner structure of a target. In biomedical applications, however, the target is usually non-stationary so that the projection measurements are time-dependent. That is, the target changes during the scanning, a case commonly referred as dynamic tomography. Such changes can be periodic (e.g., the beating of a heart) or aperiodic (e.g., the flow of contrast agent in a blood vessel), in each case, the reconstruction problem becomes severe ill-posed. If traditional static CT reconstruction methods (such as FBP, ART, IART, etc.) are directly applied, the reconstructions will be degraded seriously (e.g., being blurred or getting other artifacts), even if with small motions. A number of approaches have been proposed for dynamic CT reconstructions. For very slow or periodic movements, e.g. heart and lung imaging, gating techniques can be used [1], [9]. That is, the scanning is executed at specific time during the periodic motion, or the acquired projection data are selected according to the periodicity. For general motions, motion compensation techniques can be used to improve the reconstruction quality [2], [7], [3], [4], [5]. By motion compensation, prior information about the underlying dynamic process can be utilized by the CT reconstruction algorithms. However, this technique usually assumes simple or known motion, which is not the case for a wide variety of complex motions.

The authors are with the school of Mathematical Sciences, Capital Normal University, Beijing, 100048, China.

E-mail: hongwei.li91@cnu.edu.cn.

Dynamic CT reconstruction aims to reconstruct a series of image slices, while each image slice undergoes very limited resources, i.e. undersampled scanning. To reduce artifacts and improve reconstruction quality, many methods have been proposed in the literature by incorporating various priors. One of the main approach is to cast the reconstruction problem under the Bayesian framework. Then, the priors could be naturally encoded by the prior distribution, and by employing the Kalman filter, priors concealed between image slices could also be easily modelled and brought into the reconstruction process. A serious drawback of this approach is that the precision matrices, which are used to encode the needed priors, are usually non-sparse huge ones. Besides, the inversion of these large-scale matrices are needed during the reconstruction process, which is not applicable for real applications. To address this issue, the dimension reduction Kalman filter (DR-KF) was introduced in [6] to reduce the computational complexity. This is actually an effective method to serve the purpose. However, images reconstructed by this method are often smeared by "ringing" artifacts. In our experiments, even a small portion of dimension reduction leads to clear sharp ringing.

To combat the ringing artifacts and reduce the computational complexities, in this paper, we will introduce a splitting technique to approximate the priors encoded by the precision matrices with a spatial prior plus a temporal prior. This approximation helps to greatly reduce the computational complexity and totally avoid the ringing artifacts. Besides, to further improve the reconstruction quality, we propose a block-wise Kalman filter for the backward smoothing pass, which calls for much less computations and achieves very close performance compared to the original Kalman filter.

The remainder of this paper is organized as follows. In Section II, we will briefly introduce the DRKF method for dynamic X-Ray tomography and analyze its "ringing" artifacts. Then, the proposed method shall be described in details. Numerical experiments are performed in Section III to validate the proposed method. Finally, we conclude our paper in Section IV.

II. Materials and methods

A. Dynamic X-Ray Tomography

In dynamic X-ray tomography, the attenuation variable $\vec{x}$ is a function of the "time" $k \in N$. In a discrete setting, $\vec{x}(k)$ is

a two-dimensional image, which is related to the observation $\vec{y}_k$ by the following model

$$\vec{y}_k = A_k \vec{x}_k + \vec{\varepsilon}_k. \tag{1}$$

In tomography, $\vec{y}_k$ is the measurement vector called sinogram, $\vec{x}_k$ is the vectorized image, A_k is the measurement matrix and $\vec{\varepsilon}_k$ represents noise distribution. Thus, in dynamic X-ray tomography, the aim is to reconstruct a set of images changing over time.

B. Dimension Reduction Kalman filtering

In Bayesian framework, prior probability distributions are introduced to bring into prior knowledge. Suppose $\vec{x}_k \sim N(\vec{\mu}_k, \Sigma_k)$, $\vec{\varepsilon}_k \sim N(0, R_k)$, then the posterior density can be written as

$$p\left(\vec{x}_k \mid \vec{y}_{1:k}\right) \propto \exp\left(-\frac{1}{2}\left(\left\|\vec{y}_k - A_k\vec{x}_k\right\|^2_{R_k} + \left\|\vec{x}_k - \vec{\mu}_k\right\|^2_{\Sigma_k}\right)\right).$$

In linear Kalman filtering, a linear operator is introduced to model the state transfer process. The KF is generally divided into two steps: prediction and update, which could be described as

$$\begin{aligned} \vec{x}_k &= M_k\vec{x}_{k-1} + \vec{\xi}_k \\ \vec{y}_k &= A_k\vec{x}_k + \vec{\varepsilon}_k \end{aligned},$$

where the matrix M_k moves the previous state $\vec{x}_{k-1}$ to $\vec{x}_k$. Take the transferred state as a prediction, the complete process can be written as

$$p\left(\vec{x}_k \mid \vec{y}_{1:k}\right) \propto \exp\left(-\frac{1}{2}\left\|\vec{y}_k - A_k\vec{x}_k\right\|^2_{R_k} + \left\|\vec{x}_k - \vec{x}^p_k\right\|^2_{C^p_k}\right),$$

where

$$\begin{aligned} \vec{x}^p_k &= M_k\vec{x}^{est}_{k-1} \\ C^p_k &= M_kC^{est}_{k-1}M^T_k + Q_k \end{aligned}, \tag{2}$$

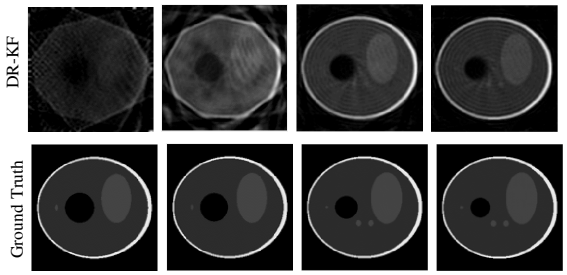
$$\begin{aligned} \vec{x}^{est}_k &= \vec{x}^p_k + C^{est}_k A^T_k R^{-1}_k\left(\vec{y}_k - A_k\vec{x}^p_k\right) \\ C^{est}_k &= \left(A^T_k R^{-1}_k A_k + \left(C^p_k\right)^{-1}\right)^{-1} \end{aligned}. \tag{3}$$

Direct application of the the above procedure is very time consuming. Suppose the image slices are size of $N \times N$. The precision matrix $(C^p_k)^{-1}$, which are size of $N^2 \times N^2$ dense matrices, are involved in the computations, which is not affordable even for small images like $N = 128$. To reduce the computational complexity, the dimension reduction Kalman filter is introduced for dynamic tomography in [8]. It starts by parameterizing the state $\vec{x}_k = \vec{\mu}_k + P_{k,r}\vec{\alpha}_k$, where $P_{k,r}$ is constructed by the first r largest singular values of C^p_k and the corresponding singular vectors. By setting a small r, the computational complexity shall be sharply reduced since the matrices involved in the computations are now size of $r^2 \times r^2$. However, the SVD decompositions, which operate on the original matrices are required for each state updating, which is still very time consuming. This is solved by assuming that all variations $\vec{x}_k = \vec{\mu}_k - \vec{\mu}_k$ live in the same fixed subspace, i.e. fixing $P_{k,r} = P_{0,r} \triangleq P_r$, which can be built by performing SVD decomposition once on the initial covariance matrix $C^p_0 = \Sigma$. Let $\Sigma = USU^T$, then $P_r = U_rS^{1/2}_r$, and (2) and (3) can be rewritten as:

$$\begin{aligned} \vec{x}^p_k &= M_k(\vec{x}^p_{k-1} + P_r\vec{\alpha}^{\text{est}}_k) \\ C^p_k &= (M_kP_r)\Psi^{\text{est}}_k(M_kP_r)^T + Q_k \end{aligned},$$

where

$$\begin{aligned} \vec{\alpha}^{\text{est}}_k &= \Psi^{\text{est}}_k(A_kP_r)^T R^{-1}_k\left(\vec{y}_k - A_k\vec{x}^p_k\right) \\ \Psi^{\text{est}}_k &= \left((A_kP_r)^T R^{-1}_k(A_kP_r) + P^T_r\left(C^p_k\right)^{-1}P_r\right)^{-1}. \end{aligned}$$



(a) $t = 1$ (b) $t = 5$ (c) $t = 15$ (d) $t = 20$

Fig. 1. Illustration of ringing artifacts from DR-KF. Slices of phantom3d reconstruction at time step $t = 1, t = 5, t = 15, t = 20$.Dimension reduction ($r = 1000$) methods with KF (5 angles).

It has been mentioned in [6] that the reconstructed images from the dimension reduction KF might suffers from "ringing" artifacts. This is demonstrated in the following experiment.

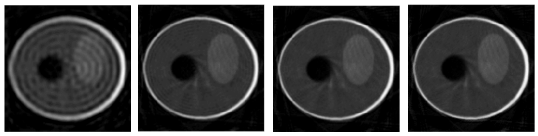
Let the image slices are size of 128×128, and assume the standard prior Gaussian covariance matrix [6], which is then size of 16384×16384 with the (i, j)th element defined as

$$\Sigma_{i,j} = \sigma^2 \exp\left(-\frac{d\left(x_i, x_j\right)^2}{2l^2}\right),$$

where σ^2 and l are configuration parameters, and $d\left(x_i, x_j\right)$ denotes the Euclidean distance between pixels x_i and x_j.

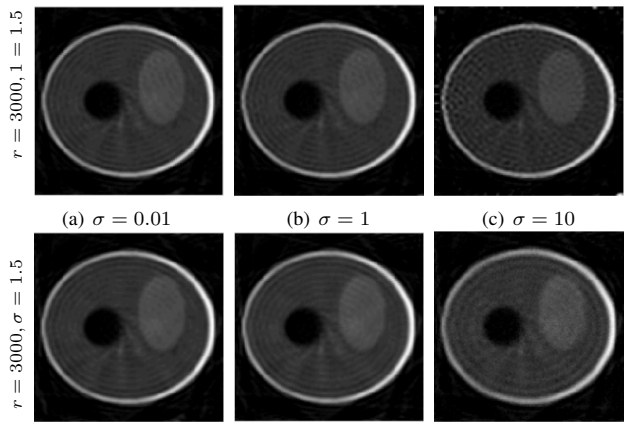
Both observation and model error covariance matrices are set to $\sigma^2 I$ and $M_k = I, \sigma = 0.1, l = 1.5$. The phantom3d function in MATLAB is employed to construct the dynamic images with 33 slices. The open-source library ASTRA is utilized to perform the forward and backward projections. There are only 5 projection views at each moment. The reconstruction results from the dimension reduction Kalman filter (DR-KF) with $r = 1000$ are shown in Fig.1. As shown in the last two columns, heavy ringing artifacts presents for the time-frame $t = 15$ and 20. At the early stage, the image slices suffer from severe undersampling artifacts and the ringing artifacts are buried inside. When the image slices getting better reconstructed, the ringing artifacts also show up and then persist in subsequent reconstructions.

To further verify the ringing artifacts, another experiment is performed to demonstrate the influence of the ratio of dimension reduction. As shown in Figure 2, by increasing the dimension parameter r, the ringing artifacts are suppressed. However, even for $r = 12000$, which amounts to a reduction ratio $12000/16384 \approx 73\%$, the ringing artifacts are still observable. For $r = 5000$, which amounts to a ratio $3/10$, heavy ringings present. Clearly, such a reduction ratio is far



(a) $r = 1000$ (b) $r = 5000$ (c) $r = 8000$ (d) $r = 12000$

Fig. 2. The influence of the ratio of dimension reduction on ringing artifacts. Phantom3d reconstruction at time step $\mathrm{t} = 15, \sigma = 0.1, l = 1.5$. Dimension reduction methods with $r = 1000, 5000, 8000, 12000$, KF(5 angles).



(a) $\sigma = 0.01$ (b) $\sigma = 1$ (c) $\sigma = 10$

(d) $1 = 3$ (e) $1 = 5$ (f) $1 = 7$

Fig. 3. The influence of configuration parameters of Gaussian prior on ringing artifacts. Slices of phantom3d reconstruction at time step $t = 1, t = 5t = 15, t = 20$. Dimension reduction methods with $r = 1000$ KF(5 angles).

from satisfactory, since the computational complexity is still very high.

One may argue that the ringing artifacts could be introduced by improper choice of the configuration parameters for the Gaussian prior. So, another experiment is performed with various parameter settings. As shown in Figure 3, the parameters σ and l do have some influence on the ringing artifacts. However, they just have certain influence on the ringing patterns rather than on the magnitude of the ringings.

C. Our Method

1) Decomposition of the prior: The computational complexity can be attributed to the regularization term

$$\Phi(\vec{x}_k - \vec{x}_k^p) = \|\vec{x}_k - \vec{x}_k^p\|^2_{C_k^p} \tag{4}$$

which is used to encode the prior information. The complexity of the covariance matrix ${C_k}^p$ results in a complex regularizer. This motivates us to approximate the above term with simpler, easier handled terms. The function Φ could always be decomposed as

$$\Phi(\vec{x}_k - \vec{x}_k^p) = \psi(\vec{x}_k) + \chi(\vec{x}_k^p) + \varrho(\vec{x}_k, \vec{x}_k^p).$$

The first two terms can be though of being encoding the priors regarding to $\vec{x}_k$ and $\vec{x}_k^p$, respectively, while the last term expresses the couplings between adjacent image slices. Since the prior regarding to $\vec{x}_k^p$ is not relevant for the reconstruction of $\vec{x}_k$, the second term could be removed. So, we propose the following posterior distribution

$$p\left(\vec{x}_k \mid \vec{y}_{1:k}\right) \propto \exp\left(-\frac{1}{2}\left\|\vec{y}_k - A_k\vec{x}_k\right\|^2_{R_k} + \psi(\vec{x}_k) + \varrho(\vec{x}_k, \vec{x}_k^p)\right).$$

The regularization term $\psi(\vec{x}_k)$ represents the prior knowledge regarding to the current ideal image $\vec{x}_k$. This prior is usually available, e.g. CT images are usually assumed to be piecewise constant and thus possess sparse gradients. The regularization term $\varrho(\vec{x}_k, \vec{x}_k^p)$ models the continuity along the temporal axis. A simple strategy is to just take $\vec{x}_k^p$ as a prior reference image.

So, we are proposing to decompose the regularization term involving the covariance matrix into a spatial regularization term plus a temporal regularization term. By choosing these two regularizers properly, the computations could be greatly reduced and ringing artifacts could also be completely avoided.

2) The online forward pass: By choosing total variation regularization for the spatial regularizer, we propose the following model for dynamic tomography reconstruction

$$\begin{aligned} \vec{x}_k^p &= M_k\vec{x}_{k-1}^{est} \\ \vec{x}_k^* &= \arg\min_{\vec{x}_k} \frac{1}{2}\left(\left\|\vec{y}_k - A_k\vec{x}_k\right\|^2_{R_k} + \lambda\left\|\vec{x}_k - \vec{x}_k^p\right\|_2^2 + \beta\left|\nabla\vec{x}_k\right|\right), \end{aligned} \tag{5}$$

where the parameter λ weights the importance of the prior image $\vec{x}_k^p$, while β weights the spatial prior. To solve the above minimization problem, the primal-dual based Chambolle-Pock (CP) algorithm is employed.

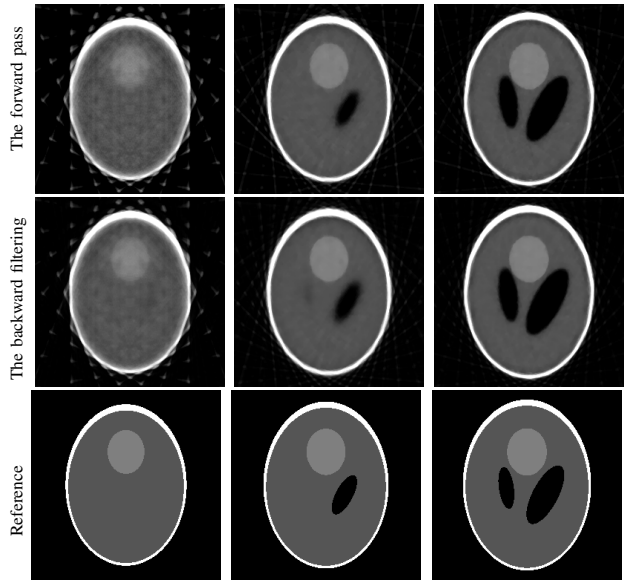
3) Block Kalman Filter: In the Kalman filtering framework, a backward smoothing procedure is usually applied for offline reconstructions. When all the projection data have been acquired and all image slices have been reconstructed, then the Kalman filter could be applied once more reversely starting from the last image slice. The backward smoothing procedure could significantly further improve the reconstruction quality.

However, the backward filtering suffers from the same drawbacks as the forward filtering. To take advantage of the backward smoothing, we propose two approximations to reduce the computational complexity. The first one is to replace the system matrix A_k by identity matrices. This is reasonable since all the image slices have been reconstructed. The second one is to decompose the filtering into blocks, i.e. decomposing the image slices into overlapping blocks and applying the Kalman filter on each block. The final results are synthesized from the filtered blocks.

With much less computations, the block Kalman filter could attain almost the same smoothing results as the non-block version does. For example, for an image size of 256×256, we can divide it into 81 overlapping blocks such that each block is size of 32×32. In the block Kalman filtering process, the covariance matrix size of 65536×65536 is replaced by 81 image blocks size of 1024×1024. Clearly, storage and computational complexities shall be dropped considerably.

III. Numerical experiments

Similarly, in the phantom3d simulated data, which is size of $256 \times 256 \times 256$, we take 50 consecutive slices in the vertical direction. The projection data are acquired with 10



(a) $t = 1$ (b) $t = 10$ (c) $t = 30$

Fig. 4. Different reconstructions of the shepp-logan phantom with the forward pass(top), the block-KF smoothing(middle), and the reference imagse at time step $t = 1, 10, 30$.

TABLE I
PSNR OF DIFFERENT METHODS FOR DIFFERENT FRAMES

Method \Time	$t = 1$	$t = 10$	$t = 30$
The forward pass	19.0807	19.2380	20.5397
KF smoothing	17.5414	19.7271	20.7552
Block-KF smoothing	19.4299	19.8850	21.1628

equally distributed scanning angles for each slice. We set $R_k = I, Q_k = 0.1 * I$. The experiments are carried out on a Linux server with two Intel(R) Xeon(R) Gold 6132 CPUs @ 2.60GHz and 256G memory. The server is also equipped with 8 Nvidia GeForce RTX 2080 Ti GPU cards, and ASTRA uses one of them for applying the forward and backward projectors.

The reconstruction results are illustrated in Fig. 4. As shown in the first row, the ringing artifacts have been completely avoided in the image slices reconstructed by the proposed forward pass, i.e. the online reconstruction procedure. The backward smoothing could help to reduce noise, as shown in the second row. However, structuring information could also be wrongly propagated back during the backward smoothing pass. In fact, we also implemented the non-block version of the Kalman filtering for the backward smoothing pass, and the results of block version is a little better than those of the non-block. This can be told from the quantitative indices listed in Table I. It should be mentioned that, to process one image, the block-KF takes 7s for one image, while the non-block KF takes about 4000s.

IV. Conclusion

To reduce the computational complexity of the Kalman filtering approach for dynamic X-ray tomography, we propose to approximate the regularization term involving the covariance matrix with two simpler regularization terms, one accounts for spatial correlation while the other one accounts for temporal correlations. The resulting method calls for much less storage and computational complexity while achieves competitive reconstructions. Besides, a block version of the Kalman filter is proposed for the backward smoothing procedure, which achieves similar smoothing behavior compared to the non-block one while again requires much less computational resources.

In our numerical experiments, the matrix M_k is always set as the identity matrix. In fact, we can also use the optical flow method to compute a better M_k such that the reconstruction quality could be further improved. This shall be investigated soon.

In order to further improve the computation speed, we also plan to use GPU and deep learning methods to approximate the forward reconstruction pass.

Acknowledgment

Thanks for the support of the National Natural Science Foundation of China (NSFC) (61971292, 61827809 and 61871275).

References

[1] Stephan Achenbach, Tom Giesler, Dieter Ropers, Stefan Ulzheimer, Hans Derlien, Christoph Schulte, Evelyn Wenkel, Werner Moshage, Werner Bautz, Werner G Daniel, et al. Detection of coronary artery stenoses by contrast-enhanced, retrospectively electrocardiographically-gated, multi-slice spiral computed tomography. *Circulation*, 103(21):2535–2538, 2001.

[2] Christophe Blondel, Régis Vaillant, Grégoire Malandain, and Nicholas Ayache. 3d tomographic reconstruction of coronary arteries using a pre-computed 4d motion field. *Physics in Medicine & Biology*, 49(11):2197, 2004.

[3] Bernadette Hahn. Reconstruction of dynamic objects with affine deformations in computerized tomography. *Journal of Inverse and Ill-posed Problems*, 22(3):323–339, 2014.

[4] Bernadette N Hahn. Null space and resolution in dynamic computerized tomography. *Inverse Problems*, 32(2):025006, 2016.

[5] Bernadette N Hahn and Eric Todd Quinto. Detectable singularities from dynamic radon data. *SIAM Journal on Imaging Sciences*, 9(3):1195–1225, 2016.

[6] Janne Hakkarainen, Zenith Purisha, Antti Solonen, and Samuli Siltanen. Undersampled dynamic x-ray tomography with dimension reduction kalman filter. *IEEE Transactions on Computational Imaging*, 5(3):492–501, 2019.

[7] T Li, Eduard Schreibmann, Y Yang, and L Xing. Motion correction for improved target localization with on-board cone-beam computed tomography. *Physics in Medicine & Biology*, 51(2):253, 2005.

[8] Antti Solonen, Tiangang Cui, Janne Hakkarainen, and Youssef Marzouk. On dimension reduction in gaussian filters. *Inverse Problems*, 32(4):045003, 2016.

[9] Jun Zhao, Yang Lu, Tiange Zhuang, and Ge Wang. Overview of multi-source ct systems and methods. In *Developments in X-Ray Tomography VII*, volume 7804, page 78040H. International Society for Optics and Photonics, 2010.

Full-Spectrum-Knowledge-Aware Unsupervised Network for Photon-counting CT Imaging

Danyang Li, Zheng Duan, Dong Zeng, Zhaoying Bian, and Jianhua Ma

***Abstract*—Deep learning (DL) based methods have been widely adopted in computed tomography (CT) field. And they also show a great potential in photon-counting CT (PCCT) imaging field. They usually require a large quantity of paired data to train networks. However, it is time-consuming and expensive to collect such large-scale PCCT dataset. In addition, lots of energy-integrating detector (EID) data, which is more easily to be obtained, are not yet included in the DL-based PCCT reconstruction network training. In this work, to address the issue of limited PCCT data and take advantage of labeled EID data, we propose a novel unsupervised full-spectrum-knowledge-aware DL-based network (FSANet), which contains supervised and unsupervised networks, to produce high-quality PCCT images. Specifically, the supervised network is trained based on paired EID dataset and serves as the prior knowledge to regularize the unsupervised PCCT network training. Moreover, a data-fidelity term for characterizing the PCCT image characteristics is constructed as a self-supervised term. Finally, we train the PCCT network with the prior knowledge and self-supervised terms following an unsupervised learning strategy. Numerical studies on synthesized clinical data were conducted to validate and evaluate the performance of the presented FSANet method, qualitatively and quantitatively. The experimental results demonstrate that presented FSANet method significantly improves the PCCT image quality in the cased of limited photon counts.**

***Index Terms*—Photon-counting CT, full-spectrum-knowledge-aware, unsupervised learning.**

I. Introduction

COMPARING with the conventional computed tomography (CT), photon-counting CT (PCCT) can obtain multiple measurements of the scanned object at multi-energy bins and provide abundant energy-dependent material-specific information. Due to the energy discrimination capability, PCCT can effectively improve contrast-to-noise ratio, increase the dose efficiency and reduce electronic noise [1], [2].

However, the collected PCCT data is corrupted by serious quantum noise [3], because the photon counts decrease in the narrow energy bin. And then, the PCCT image quality degrades obviously due to the limited photons. To solve this problem, many statistical iteration reconstruction (SIR) methods have been proposed in the past decades. The main idea of the SIR is to construct a reconstruction model with data fidelity and regularization terms, where the first term incorporates the statistical property of X-ray photons and the second term provides the prior information of the desired PCCT images. For example, Rigie et al. introduced a total nuclear variation regularization to leverage similar gradient information and improve the image quality [4]. Kim et al. developed a patch-based low-rank regularization to maintain the image structures and reduce noise [5]. Semerci et al. combined a tensor nuclear norm and a total variation regularization to suppress noise [6]. Zhang et al. proposed to deliver inner spectrum correlation information and constructed a tensor-based dictionary learning strategy [7]. Niu et al. considered the self-similarity of the spectral CT images and proposed a non-local low-rank and sparse matrix decomposition method [3]. Wu et al. proposed to encourage the similarity of spectral CT images by utilizing a cube-based tensor regularization [8]. Recently, Zeng *et al.* proposed a full-spectrum-knowledge-aware tensor by imposing the global correlation, piecewise smooth and latent full-spectrum properties of PCCT images [9]. These methods have been shown great potential in preserving image details and suppressing noise. However, there remains some challenges in practice. First, the SIR methods usually utilize fan-beam geometry and the computation costs will be a burden for cone-beam geometry. Second, SIR methods are sensitive for the hyper-parameters, and appropriate parameters selection is needed for different clinical applications.

Recently, deep learning (DL) has been widely adopted in CT imaging field. In spectral CT imaging field, Lu *et al.* utilized DL-based method for material decomposition [10]. Fang *et al.* proposed to remove ring artifacts for PCCT data by using DL-based method [11]. Wu *et al.* employed a DL-based method for reconstructing PCCT images [12]. It is shown that the DL-based methods achieved competing results compared with the SIR methods. However, the current DL-based methods need a large quantity of paired data (i.e., noisy and high-quality data) to obtain a desired model by supervised training strategy. However, collecting large-scale spectral CT data is time-consuming, and the PCCT data in clinics is hard to be obtained. In addition, a number of energy-integrating detector (EID) data, which is easily to be obtained, is not yet included in training the DL-based method for PCCT imaging.

Therefore, we present an unsupervised DL-based method in the image domain by utilizing the prior information of the paired EID dataset (i.e., low-dose images/high-quality ones). Specifically, we first initialize two denoising networks for EID and PCCT images, respectively. Then, the supervised network for conventional CT is trained on a well paired EID data and serves as the prior knowledge to regularize the unsupervised PCCT network training. Moreover, a data-fidelity term for

This work was supported in part by the NSFC under Grant U21A6005 and Grant U1708261. *(Corresponding author: Zhaoying Bian and Jianhua Ma.)*

D. Li, Z. Duan, and D. Zeng are with School of Biomedical Engineering, Southern Medical University, Guangzhou 510515, China.

Z. Bian is with School of Biomedical Engineering, Southern Medical University, Guangzhou 510515, China (e-mail: zybian@smu.edu.cn).

J. Ma is with School of Biomedical Engineering, Southern Medical University, Guangzhou 510515, China (e-mail: jhma@smu.edu.cn).

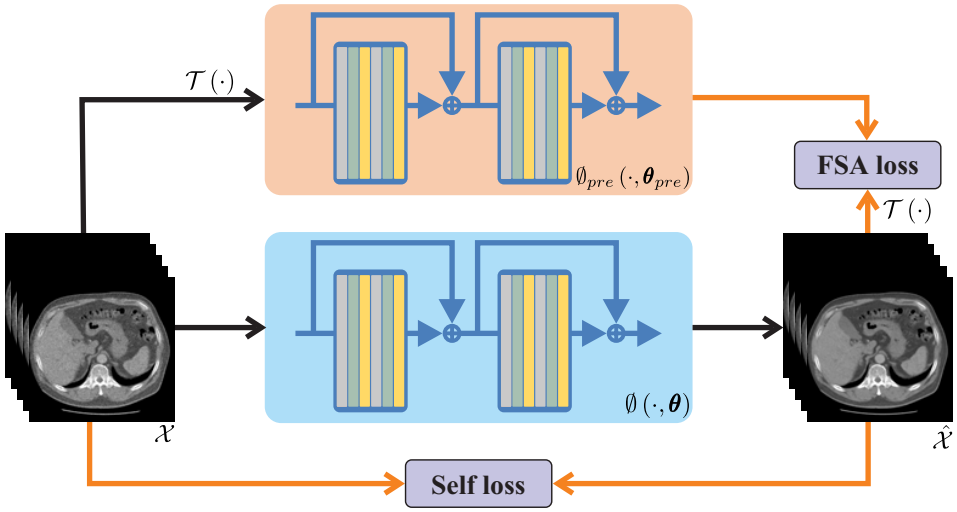


Fig. 1. Illustration of the presented FSANet method for PCCT image recovery. The pipeline with black arrows denote the data flow of the PCCT images. The pipeline with orange arrows denote calculations of loss functions. The supervised and unsupervised networks have the same architecture. It should be noted that the optimization for the parameters of the unsupervised network follows an unsupervised strategy, and only noisy PCCT images are involved during its training period.

characterizing the PCCT image characteristics is constructed as the self-supervised loss. Finally, with the prior knowledge and self-supervised terms, we can train the network for PCCT images following an unsupervised learning strategy. We call the presented DL-based model as full-spectrum-knowledge-aware DL-based method, shorten as "FSANet". We evaluated the presented FSANet and other reconstruction methods on synthesized clinical data. Experimental results demonstrated that the presented FSANet outperforms the competing methods in terms of qualitative and quantitative metrics.

II. Methods

Considering the spatial and energy dimensions of the spectral data, the PCCT imaging model can be expressed as follows:

$$\mathcal{Y} = \mathcal{A}\mathcal{X}^* + \varepsilon, \tag{1}$$

where $\mathcal{Y} = \{y_n, n \leq N\}$ and $\mathcal{X}^* = \{x_n^*, n \leq N\}$ are the measurements and desired PCCT images along the multi-energy bins, and N is the total number of the energy bins. $\mathcal{A}$ is the linear projection operator for PCCT imaging, and ε denotes the noise in the projection domain.

The images reconstructed from $\mathcal{Y}$ suffer from noise and artifacts. To improve image quality, DL-based methods are feeded with the PCCT images and produce the denoised ones. It be expressed as follows:

$$\hat{\mathcal{X}} = \emptyset_{DL}(\mathcal{X}, \boldsymbol{\theta}_{DL}), \tag{2}$$

where $\emptyset_{DL}$ represents the network of the DL-based method with parameters $\boldsymbol{\theta}_{DL}$. $\hat{\mathcal{X}}$ are the estimated PCCT images by the network. $\mathcal{X} = \left(\mathcal{A}^{\mathcal{T}}\mathcal{A}\right)^{-1}\mathcal{A}^{T}\mathcal{Y}$ are the network inputs which are directly reconstructed by the filtered back projection (FBP) algorithm [13]. Followed by the supervised strategy, the network parameters $\boldsymbol{\theta}_{DL}$ are optimized by minimizing the loss function between the target images $\mathcal{X}^{target}$ and $\hat{\mathcal{X}}$, and the loss function can be expressed as follows

$$\hat{\boldsymbol{\theta}}_{DL} = \arg\min_{\boldsymbol{\theta}_{DL}} \mathcal{L}\left(\hat{\mathcal{X}}, \mathcal{X}^{target}\right), \tag{3}$$

where $\mathcal{L}$ is the user defined loss function and $\mathcal{X}^{target}$ are the FBP-reconstructed images from the high-dose measurements.

In order to utilize the prior information of the network pre-trained on paired EID dataset, we present a full-spectrum-aware (FSA) loss function, as follows:

$$\mathcal{L}_{FSA}(\mathcal{X}) = \left\|\mathcal{T}\left(\emptyset\left(\mathcal{X}, \boldsymbol{\theta}\right)\right) - \emptyset_{pre}\left(\mathcal{T}\left(\mathcal{X}\right), \boldsymbol{\theta}_{pre}\right)\right\|_1, \tag{4}$$

where $\emptyset_{pre}$ denotes a pre-trained network for EID images with parameters $\boldsymbol{\theta}_{pre}$. $\emptyset(\cdot, \boldsymbol{\theta})$ is the network for PCCT images with parameters $\boldsymbol{\theta}$. $\mathcal{T}(\cdot)$ is an operator to transform the PCCT images to the EID images. $\|\cdot\|_1$ is L_1 norm. Moreover, we also construct a self-supervised loss to encourage the data fidelity of the PCCT images, as follows:

$$\mathcal{L}_{Self}(\mathcal{X}) = \|\emptyset(\mathcal{X}, \boldsymbol{\theta}) - \mathcal{X}\|_1. \tag{5}$$

In summary, the total loss function of the presented model is expressed as follows:

$$\mathcal{L}(\mathcal{X}) = \mathcal{L}_{Self}(\mathcal{X}) + \alpha\mathcal{L}_{FSA}(\mathcal{X}), \tag{6}$$

where α is a hyper-parameter of the loss function. It can be seen that the presented model engages an unsupervised learning strategy where only noisy PCCT images is involved in the training stage. Fig. 1 illustrates the presented FSANet method for PCCT image recovery. Finally, we utilize Adam [14] to optimize the parameters in the network.

III. Results

A. Implementation details

The network $\emptyset_{pre}$ was trained on the paired EID data collected from conventional CT in local hospital with WGAN

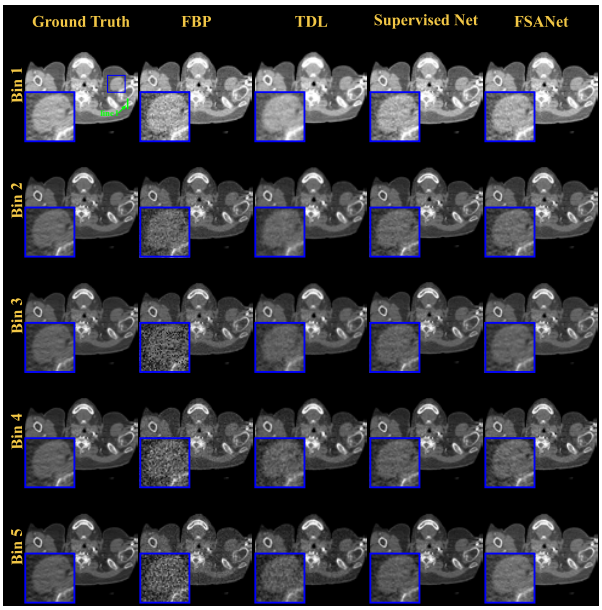


Fig. 2. Results of the presented and compared methods on Case 1. The display windows from Bin 1 to 5 are [0.005, 0.008], [0.0028, 0.005], [0.002, 0.004], [0.0015, 0.0035] and [0.0015,0.0028] mm^{-1}, respectively. Zoomed ROIs indicated by the blue box are displayed for better visualization.

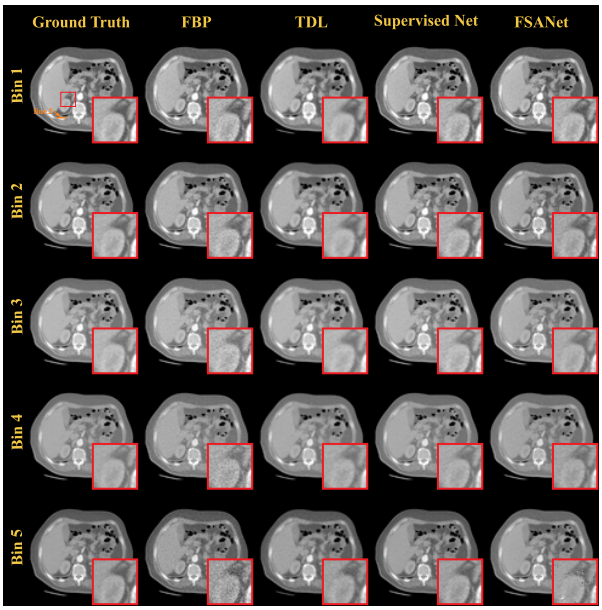


Fig. 3. Results of the presented and compared methods on Case 2. The display windows from Bin 1 to 5 are [0.0015, 0.0025], [0.0009, 0.0018], [0.0007, 0.0015], [0.0006, 0.0012] and [0.0005, 0.0010] mm^{-1}, respectively. Zoomed ROIs indicated by the red box are displayed for better visualization.

loss [15]. The PCCT images datasets involved in this study were simulated from EID images by segmenting the soft and bone tissues. The noisy data was generated by adding Poisson noise in the projection domain. The X-ray spectrum with 120 kVp tube voltage and 1.6 mm Al filtration was generated by SPEKTR toolbox [16]. Five energy bins are determined by the five thresholds: 25, 50, 60, 70 and 85 keV. The simulation imaging parameters are listed as follows: 816 parallel X-ray beams and 1160 projection views over 360^o are adopted, and source-to-detector and source-to-center distances are 1040.0 and 570.0 mm, respectively. The network trained by supervised strategy, called "Supervised Net", serves as the upper-bound of the presented FSANet. Morover, the filtered back projection (FBP) method with a ramp filter and a tensor-based dictionary learning regularization (TDL) method are the compared methods for the presented method.

In experiments, we simulated 3000 cases to established the whole dataset, and randomly selected 2000, 500 and 500 cases for training, validation and testing datasets, respectively. In training period, learning rate is $1e^{-4}$, batch size is 6, and training epoch is 2000. A modified residual network [17] is selected as the backbone network for the pre-trained and the presented FSANet networks. Both networks are implemented in Python with PyTorch package on a NVIDIA Tesla K40c GPU.

B. Qualitative analysis

Fig. 2 illustrates the visual comparisons of the presented and compared methods on Case 1. The images at normal dose are chosen as the ground truth. It can be observed that FBP algorithm suffers from noise. TDL can effectively remove the noise-induced artifacts and improve the image quality, but losses the image resolution. On the contrary, the

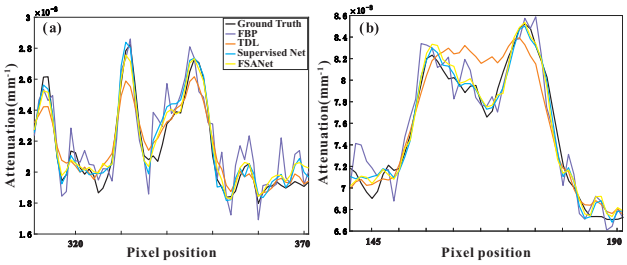


Fig. 4. Profiles of the green and orange lines in Bin 1 for Case 1 and 2, respectively. (a) are the profiles of the results from different methods on Case 1, respectively. (b) are the profiles of the results from different methods on Case 2, respectively.

presented FSANet produces more remarkable results closing to the Supervised Net and the ground truth in terms of noise reduction and structure preservation. Moreover, zoomed in regions-of-interest (ROIs) by the blue box in Fig. 2 is selected for better visual inspection. It can be seen that TDL smooths the structure edges, but the Supervised Net and presented FSANet methods maintain the image details.

Fig. 3 shows the results of different methods on Case 2. Similar to Case 1, FBP induces noise to the images, TDL is prone to produce blurry results, and the presented FSANet avoids over smoothing and preserves structure details. The zoomed in ROIs indicated by the red box in Fig. 3 also demonstrate the advanced performance of the presented FSANet method. Fig. 4(a) and (b) show the profiles indicated by the green and orange lines in the Fig. 2 and Fig. 3, respectively. From the results, we can observe that the presented FSANet produces the closet results to the ground truth compared with the FBP and TDL methods.

C. Quantitative analysis

In this study, peak signal-to-noise (PSNR) and root-mean-square-error (RMSE) are utilized to quantify the performances of different methods. Table I lists the quantitative measurements of the results from different methods on the whole testing dataset. From the results, it can be seen that the presented FSANet achieves better results among all metrics, compared with FBP and TDL methods. Therefore, the qualitative and quantitative results demonstrate that the presented FSANet method achieves superior results to the FBP and TDL methods.

TABLE I
QUANTITATIVE MEASUREMENTS ON THE RECONSTRUCTION RESULTS ON THE TESTING DATASET FROM THE DIFFERENT METHODS

	PSNR (dB)	RMSE ($\times 10^{-5}$)
FBP	28.36 ± 2.88	12.5 ± 0.73
TDL	35.37 ± 5.70	7.62 ± 4.324
Supervised Net	$\mathbf{38.14 \pm 2.60}$	$\mathbf{3.78 \pm 0.433}$
GMM-3DTV	$\mathbf{37.44 \pm 2.25}$	$\mathbf{4.46 \pm 0.788}$

IV. Discussion and Conclusion

DL-based methods have been shown promising performance in conventional CT imaging, and it has also inspired the application in PCCT imaging. However, most of them are supervise-based and need a large quantity of paired training dataset, which is hard to be obtained for PCCT. To address this intrinsic limitation, in this work, we presented an DL-based PCCT denoising method with an unsupervised learning strategy, called "FSANet". Specifically, we first trained a denoising network on paired EID dataset and served it as a prior for PCCT images. Then, we constructed this prior network as an training loss to regularize the PCCT network learning in an unsupervised manner. Moreover, an self-supervised loss of the noisy PCCT images is introduced to promote data-fidelity of the PCCT images. Finally, with the mentioned two loss terms, we can obtained the presented FSANet method. Simulation experiments demonstrated the feasibility and effectiveness of the presented FSANet method. In the future, clinical patient studies would be involved to further demonstrate the denosing performance of the presented FSANet method.

References

[1] M. J. Willemink, M. Persson, A. Pourmorteza, N. J. Pelc, and D. Fleischmann, "Photon-counting CT: Technical principles and clinical prospects," *Radiology*, vol. 289, no. 2, pp. 293-312, 2018.

[2] K. Taguchi, J. S. Iwanczyk, "Vision 20/20: Single photon counting x-ray detectors in medical imaging," *Medical Physics*, vol. 40, no. 10, p. 100901, 2013.

[3] S. Niu, G. Yu, J. Ma, and J. Wang, "Nonlocal low-rank and sparse matrix decomposition for spectral CT reconstruction,"*Inverse Problems*, vol. 34, no. 2, p. 024003, 2018.

[4] D. S. Rigie and P. J. La Rivire, "Joint reconstruction of multi-channel, spectral CT data via constrained total nuclear variation minimization," *Physical in Medical & Biology*, vol. 60, no. 5, pp. 1741-1762, 2015.

[5] K. Kim and *et al.*, "Sparse-view spectral CT reconstruction using spectral patch-based low-rank penalty," *IEEE Transactions on Medical Imaging*, vol. 34, no. 3, pp. 748-760, 2015.

[6] O. Semerci, N. Hao, M. E. Kilmer, and E. L. Miller, "Tensor-based formulation and nuclear norm regularization for multienergy computed tomography," *IEEE Transactions on Image Processing*, vol. 23, no. 4, pp. 1678-1693, 2014.

[7] Y. Zhang, X. Mou, G. Wang, and H. Yu, "Tensor-based dictionary learning for spectral CT reconstruction," *IEEE Transactions on Medical Imaging*, vol. 36, no. 1, pp. 142-154, 2017.

[8] W. Wu, F. Liu, Y. Zhang, Q. Wang, and H. Yu, "Non-local low-rank cube-based tensor factorization for spectral CT reconstruction," *IEEE IEEE Transactions on Medical Imaging*, vol. 38, no. 4, pp. 1079-1093, 2019.

[9] D. Zeng, L. Yao, Y. Ge, S. Li, Q. Xie, H. Zhang, and *et al.*, "Full-spectrum-knowledge-aware tensor model for energy-resolved CT iterative reconstruction," *IEEE Transactions on Medical Imaging*, vol. 39, pp. 2831-2843, 2020.

[10] Y. Lu, M. Kowarschik, X. Huang, Y. Xia, J. Choi, S. Chen, and *et al.* "A learning-based material decomposition pipeline for multi-energy X-ray imaging," *Medical physics*, vol. 46, pp. 689-703, 2019.

[11] W. Fang, L. Li, and Z. Chen, "Removing ring artefacts for photon-counting detectors using neural networks in different domains," *IEEE Access*, vol. 8, pp. 42447-42457, 2020.

[12] W. Wu, D. Hu, C. Niu, L. Broeke, A. Butler, P. Cao, and *et al.*, "Deep learning based spectral CT imaging," *Neural Networks*, vol. 144, pp. 342-358, 2021.

[13] G. L. Zeng, "Medical image reconstruction: a conceptual tutorial," Springer, 2010.

[14] D. P. Kingma and J. Ba, "Adam: A method for stochastic optimization," *arXiv preprint*, arXiv: 1412.6980, 2014

[15] M. Arjovsky, S. Chintala, and L. Bottou, "Wasserstein GAN." [Online]. Available: https://arxiv.org/abs/1701.07875

[16] J. Punnoose, J. Xu, A. Sisniega, W. Zbijewski, and J. H. Siewerdsen, "Technical note: Spektr 3.0-A computational tool for X-ray spectrum modeling and analysis," *Medical Physics*, vol. 43, no. 8Part1, pp. 4711-4717, 2016.

[17] K. He, X. Zhang, S. Ren, *et al.*, "Deep residual learning for image recognition," *Proc. IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, pp. 770-778, 2016.

Soil matrix study using a hybrid a-Se/CMOS pixel detector for CT scanning

Akyl Swaby[1], Adam S. Wang[2], Michael G. Farrier[3], Weixin Cheng[1], and Shiva Abbaszadeh[1]

***Abstract*—Computed tomography (CT) is a non-invasive means of localizing a region of interest within an object, which enables the investigation of soil distributions and localized flow processes within soil pore systems. CT scanning allows for cross-sectional successions that provide visibility within the environment of plant samples. Knowledge of the characteristics of the soil pore system is essential for evaluating various processes that take place between root-soil interactions. In this study, we investigate the potential application of a high-resolution amorphous selenium (a-Se) direct conversion detector on complementary metal-oxide-semiconductor (CMOS) readouts for micro-CT scanning of a soil matrix to image the status of aggregation and networks of pore spaces within intact soil. The combination of the intrinsic high spatial resolution of a-Se and small pixel CMOS readouts provide detailed visualization in the soil aggregates of the plant samples. The X-ray energy and plant soil thickness were varied during the investigation for imaging the root-soil. A 10 μm spatial resolution and noise limited performance of 8 photons/pixel at 20 keV were achieved. High attenuation of X-rays in thick soil poses challenges however fine details are observable in thinner samples and care should be taken when choosing soil thickness and container material.**

***Index Terms*—amorphous selenium, computed tomography, radiation imaging, soil properties.**

I. Introduction

X-ray computed tomography (CT) is implemented to characterize and visualize the environment of plant samples for the investigation of the physical and hydrological properties related to root-soil interactions. The use of CT scanning for characterization of the impact of pore spaces within a soil matrix has been previously studied [1-3]. These studies have demonstrated the advantages of using CT scanning for the characterization of soil aggregate properties such as volume, surface area, and sphericity enabled by the non-destructive quantification of soil structures as three dimensional (3D) images. Imaging the soil structures, which are described as the aggregation or distribution and networks of pore spaces, provides visibility into the root-soil interactions that affect the pore structure within the rhizosphere. The rhizosphere is the site of interaction between the plant root and the soil where water and nutrients are absorbed by the roots and where photosynthates are distributed [4]. CT scanning presents many advantages including the ability to rotate the 3D images and to view their cross-sectional slices, making it efficient to locate the regions of interest in the rhizosphere of the plant. Visibility within the sample and the relatively high spatial resolution of CT images combined with positron emission tomography (PET) allow for the investigation of the temporal changes occurring in the plant, soil, and root tips due to the transport of a radiation tracer, such as Carbon-11 (^{11}C) attached to CO_2 [4].

Previous investigations using CT for studying the root-soil interactions have highlighted challenges such as [5-10]:

(i) scale integration from the micron scale to the ecosystem scale of the rhizosphere; and

(ii) a lack of dynamic observations/measurements of soil structural changes in response to disturbances at various temporal scales.

To enable a low cost, high spatial resolution X-ray detector towards real-time CT scanning of plants combined with dynamic PET imaging, we are investigating replacing the indirect conversion scintillator detector with a direct conversion photoconductive layer made of amorphous selenium (a-Se). Easily processed as a uniform thick layer over large areas, a-Se has an atomic number (Z = 34) sufficient for high absorption of X-ray imaging (20-100 keV), a k-edge energy of 12.66 keV, low dark current, high charge collection efficiency, and high inherent spatial resolution [11, 12]. In this study, we investigate the impact of using a hybrid a-Se/CMOS coupled to an active pixel array for X-ray imaging.

II. Methods

A preliminary experiment was conducted at Stanford University using the Siemens Artis Zeego [13] with a tube voltage range of 40 – 125 kV, a 210 mA current, and 11.5 ms pulse width. The X-ray tube within the system is a Megalix Cat 125/15/40/80 three-focus high-performance X-ray tube assembly. The X-ray detector has an indirect conversion layer composed of amorphous silicon (a-Si) with a cesium iodide (CsI) scintillation material. The detector consists of an area of 30 × 40 cm^2 with a pixel size of 154 μm [14]. The CT projection images were collected at 30 images per second, for a total of 496 projections.

A sample root-soil system (Fig. 1) was imaged within an acrylic container (10 × 10 × 12 cm^3; L × W × H). At the time of

We acknowledge support from the US Department of Energy (DOE), Office of Biological and Environmental Research (BER) under Award Number DE-SC0021975. Detector development was also funded through DOE, Office of Science, phase I SBIR program, grant No. DE-SC0019626.

University of California, Santa Cruz[1], Stanford University[2], and Farrier Microengineering LLC[3]

the imaging, the Calypso beans (*Phaseolus vulgaris*, cultivar 'Calypso') were grown for three weeks after germination with soil moisture ranging from 50% to 80% field capacity.

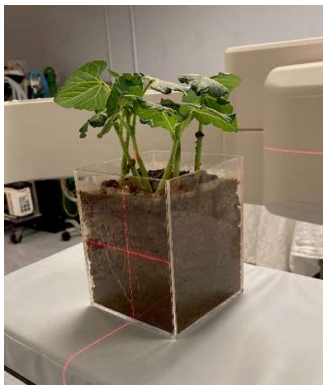


Fig. 1. A sample of a root-soil system imaged to study soil aggregates.

In comparison to the images from the Artis Zeego CT system, the 1-Megapixel (1Mp) X-ray detector can acquire images of the plant sample with finer detailed information of the roots and soil structure. A performance summary of the detector is listed in Table 1 [15]:

Table 1. Summary of the a-Se/CMOS direct conversion X-ray detector performance specifications.

Detector features	Value
Pixel size	7.8 µm
Array size	1000 × 1000 pixels
Field of view	7.8 mm × 7.8 mm
Frame rate max	5 fps
Full well capacity	877 ke$^-$
RMS noise	180 e$^-$ @ 5fps
Conversion gain	762 e$^-$/ph @ 63 keV, 120 e$^-$/ph @ 20 keV 5 V/µm
System conversion	90.6 e$^-$/DN (theoretical)
MTF (Modulation transfer function)	50% @ 0.5 Nyq. (32 lp/mm) @ 63 keV @ 5 V/µm
Detector field (HV)	4 to > 10 V/µm
Image Lag	1% to 3 % @ 1s

The test bench for investigating the hybrid a-Se detector included a microfocus X-ray source (Microbox, Micro X-ray Inc. (MXR), Santa Cruz) and the 1Mp X-ray detector. The MXR microfocus X-ray source has a varying focal spot size ranging from 5 µm to 10 µm as the source output power changes from 7.5 to 15 W. Developed by KA Imaging, the 1Mp X-ray detector is an a-Se/CMOS hybrid structure with 1000 × 1000 pixels (Fig. 2) [16]. The key technology consists of a monolithic hybrid X-ray detector built by layering an a-Se film directly deposited on each 7.8 µm pixel of the CMOS active pixel readout array [15]. This a-Se/CMOS direct X-ray detector technology has demonstrated micron scale resolution as well as an order of magnitude increase in detection efficiency from typical indirect X-ray detectors at energies ranging from 15 keV to > 63 keV [16].

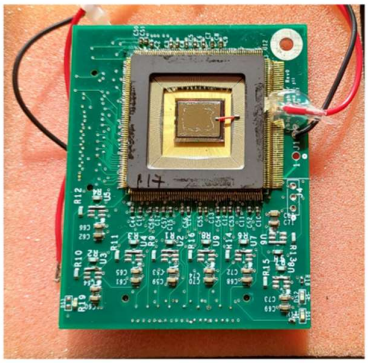
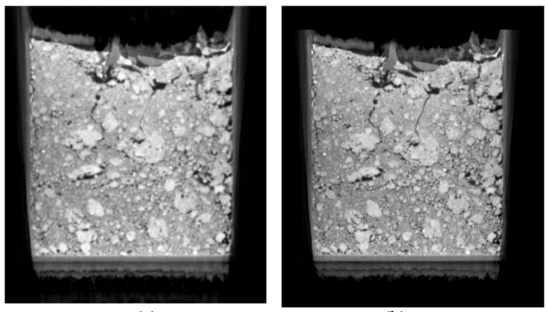


Fig. 2. Photo of the 1Mp a-Se/CMOS Hybrid Detector with a 7.8 µm pixel pitch. The 1Mp detector package and analog daughter board are soldered to the test board.

The CMOS readout integrated circuit (ROIC) design of the 1Mp detector has the potential to become scalable in array size, using reticle stitching IC fabrication technology, to achieve an array greater than 8000 × 8000 = 64 megapixels, with FOV greater than 63 × 63 mm^2. Due to the limited size of the chip, in order to scan a larger sample, repeated scans of the sample will be necessary as well as combining the acquisition images as the FOV would be limited by the size of the pixel array.

III. Results

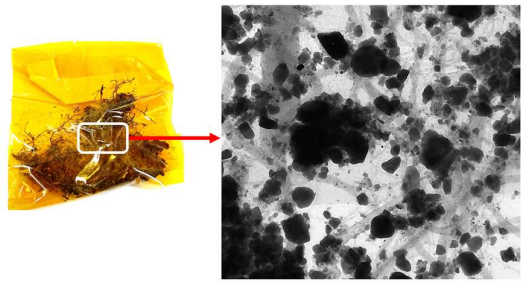
Figure 3 shows the CT images taken at Stanford University, using the Siemens Artis Zeego, at an average energy of 97 kV vs 109 kV, 512 × 512 voxel array, and 0.49 vs 0.25 mm voxel sizes. The Artis Zeego C-arm was positioned with respect to the orientation of the potted plant, using the laser traces to align the sample. Due to its design, the C-arm has a non-continuous axis of rotation that was accounted for in the reconstruction. Parameters that were set for acquisition and reconstruction included tube voltage, dose, automatic exposure control (AEC) field, VOI (volume of interest) size, reconstruction kernel, and slice matrix, while parameters that were automatically selected included tube current and pulse width.



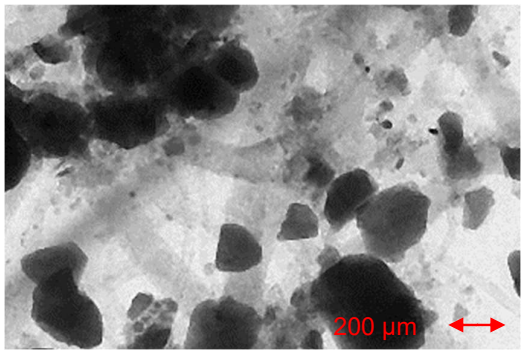
(a) (b)

Fig. 3. Image reconstruction of the Calypso sample with (a) average tube voltage of 97 keV and 0.49 voxel size [mm] and (b) average 109 keV and 0.25 voxel size [mm].

Using the 1Mp detector and MXR micro focus X-ray source, images were acquired at an energy of 20 kV using a 900 × 900 pixel array. The microfocus X-ray source emits at a spot size that is matched to the detector pixel pitch to minimize penumbral blurring. Figure 4 shows an image acquired from grass roots with a thickness of 0.5 cm to 0.75 cm at an energy of 20 kV with a 1:1 magnification. To improve contrast of the roots in the soil, soil-root samples were contained between Kapton films to visualize the biomass. At low kV energy, significant structural detail in the grass roots can be visualized and detected at a 15 µm to 23 µm resolution.



(a)



(b)

Fig. 4. (a) Image of grass roots in soil with a sample thickness of 0.5 cm to 0.75 cm. The image was acquired using a 900 × 900 pixel array from the 1Mp detector, equivalent to an active area of 7 × 7 mm^2. (b) A detail of the grass root sample using 2 to 3 pixels, from the detector, provides a resolution of 15 µm to 23 µm at 1:1 magnification.

Further imaging of plant root samples demonstrates significant structural detail visible at a 10 µm to 15 µm resolution (Fig. 5). However, attenuation in soil substantially obscures the low-density biomass structure. As it can be seen from Fig. 6, the low-density target made of food-grade polypropylene (PP) plastic emulating the low density of the biomass materials was obscured by soil material. A 5 mm thickness of soil absorbs more than 80% of X-ray photons at 20 kV. Figure 7 shows the X-ray penetration vs. soil thickness at 60 kV. It can be seen that 10 mm of soil absorbs 60% of X-ray photons at 60 kV.

IV. Discussion

Details in the rhizosphere provided by a higher resolution image may allow for modelling of representative interacting volumes of root hairs and soil particles [17, 18]. The percentage of porous area can be calculated from the images shown in Fig. 3. A high resolution image can be utilized to determine the correlation between where roots continue to grow and the porosity of the soil. Compared to the pixel size of 154 µm in the Artis Zeego detector, the 7.8 µm pixel size of the a-Se detector provides improved resolution which is needed for studying the finer structure of microaggregates (with diameters ranging from 10 µm to 250 µm) and the micropores inside microaggregates. More importantly, many crucial functions provided by plant roots, microorganisms, and soil aggregates normally operate at this finer resolution.

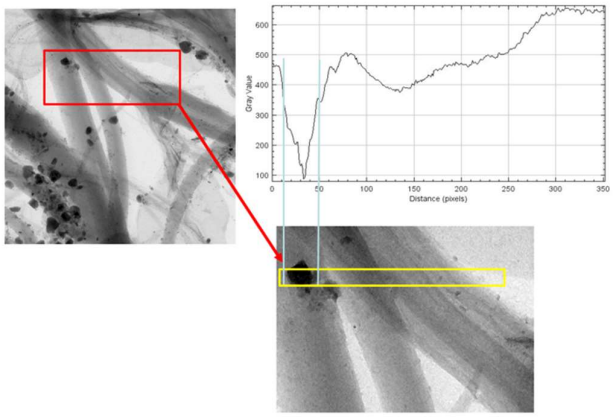


Fig. 5. Image of a live plant root sample used to study the attenuation of X-ray signal through soil material. The attenuation in soil substantially obscures the low-density biomass structure. The graph shows the gray scale of the acquired image versus the distance equivalent to the number of pixels of the detector.

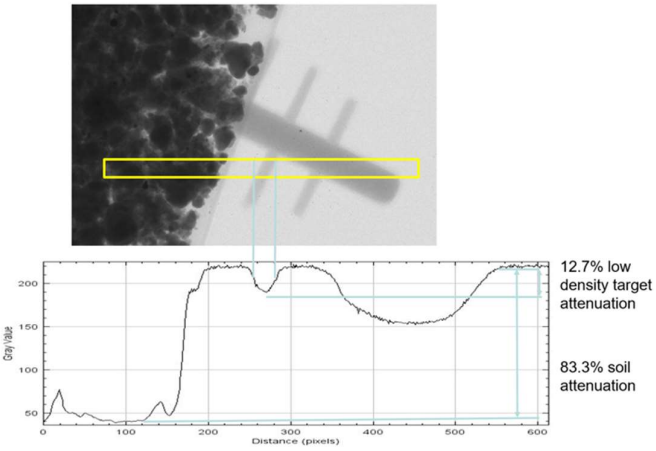


Fig. 6. Image of soil material obscuring a low-density target (food-grade polypropylene (PP) plastic) with a tube voltage of 20 kV. The plastic target was used to simulate a low-density biomass target which demonstrated that 5 mm of soil absorbs more than 80% of X-ray photons at 20 kV. The graph shows the gray scale of the acquired image versus the distance equivalent to the number of pixels of the detector.

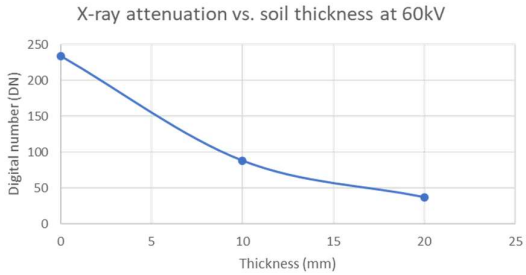


Fig. 7. X-ray attenuation study vs. soil thickness. 10 mm of soil absorbs 60% at 60 kV.

From initial imaging of plant samples within an acrylic container, it was observed that the acrylic container in combination with soil is absorbing X-ray at low energies, causing limited image quality of the rhizosphere. As previously investigated, considering X-ray CT with energy information, the X-ray energy distribution can be calculated using a mathematical model, such as a response function, to obtain reference measurements of the optimal thickness for the X-ray path length in an acrylic container [19].

V. Conclusion

Based on previous investigations related to the study of root-soil characteristics using X-ray CT scanning [5-10], we have investigated the use of a direct conversion CMOS ROIC, the 1Mp detector, to improve the image quality and resolution of the pixel detector currently implemented in the Artis Zeego. CMOS direct charge sensors may be capable of improving the limitations caused by the ratio between sample size and resolution (voxel size), as well as the smaller resolution for larger sample sizes [20]. The combination of high spatial resolution and high quality image reconstruction may allow for improved detail in the image quality of the 3D acquisition of the rhizosphere. From this preliminary work, the X-ray source emitted at a spot size that is matched to the detector pixel pitch to minimize penumbral blurring. A 10 µm spatial resolution and noise limited performance of 8 photons/pixel at 20 keV were achieved. We believe that microCT of live plants, in situ, would require:

i) A high fluence, high keV X-ray source that would generate images of voids in the soil where biomass in growing

ii) A form of radio dense metallic dopant to be taken up in the root system

The combined application of PET and CT scanning may provide simultaneous spatial and temporal data on root morphology and architecture.

Acknowledgment

We acknowledge support from the US Department of Energy (DOE), Office of Biological and Environmental Research (BER) under Award Number DE-SC0021975. Detector development was also funded through DOE, Office of Science, phase I SBIR program, grant No. DE-SC0019626.

References

[1] Grevers, M. C. J., Jong, E. D., & St. Arnaud, R. J. (1989). The characterization of soil macroporosity with CT scanning. Canadian Journal of Soil Science, 69(3), 629-637.

[2] Warner, G. S., Nieber, J. L., Moore, I. D., & Geise, R. A. (1989). Characterizing macropores in soil by computed tomography. Soil Science Society of America Journal, 53(3), 653-660.

[3] Garbout, A., Munkholm, L. J., & Hansen, S. B. (2013). Temporal dynamics for soil aggregates determined using X-ray CT scanning. Geoderma, 204, 15-22.

[4] Garbout, A., Munkholm, L. J., Hansen, S. B., Petersen, B. M., Munk, O. L., & Pajor, R. (2012). The use of PET/CT scanning technique for 3D visualization and quantification of real-time soil/plant interactions. Plant and soil, 352(1), 113-127.

[5] Hou, L. H., Gao, W., Weng, Z. H., Doolette, C. L., Maksimenko, A., Hausermann, D., ... & Kopittke, P. M. (2022). Use of X-ray tomography for examining root architecture in soils. Geoderma, 405, 115405.

[6] Naveed, M., Moldrup, P., Schaap, M. G., Tuller, M., Kulkarni, R., Vogel, H. J., & Wollesen de Jonge, L. (2016). Prediction of biopore-and matrix-dominated flow from X-ray CT-derived macropore network characteristics. Hydrology and Earth System Sciences, 20(10), 4017-4030.

[7] Peyton, R. L., Haeffner, B. A., Anderson, S. H., & Gantzer, C. J. (1992). Applying X-ray CT to measure macropore diameters in undisturbed soil cores. Geoderma, 53(3-4), 329-340.

[8] Singh, J., Singh, N., & Kumar, S. (2020). X-ray computed tomography–measured soil pore parameters as influenced by crop rotations and cover crops. Soil Science Society of America Journal, 84(4), 1267-1279.

[9] Sun, X., Li, X., Zheng, B., He, J., & Mao, T. (2020). Study on the progressive fracturing in soil and rock mixture under uniaxial compression conditions by CT scanning. Engineering Geology, 279, 105884.

[10] Helliwell, J. R., Sturrock, C. J., Mairhofer, S., Craigon, J., Ashton, R. W., Miller, A. J., ... & Mooney, S. J. (2017). The emergent rhizosphere: imaging the development of the porous architecture at the root-soil interface. Scientific reports, 7(1), 1-10.

[11] Allec, N., Abbaszadeh, S., Fleck, A., Tousignant, O., & Karim, K. S. (2012). K-edge imaging using dual-layer and single-layer large area flat panel imagers. IEEE Transactions on Nuclear Science, 59(5), 1856-1861.

[12] Hellier, K., Benard, E., Scott, C. C., Karim, K. S., & Abbaszadeh, S. (2021). Recent Progress in the Development of a-Se/CMOS Sensors for X-ray Detection. Quantum Beam Science, 5(4), 29.

[13] Siemens Healthineers | Corporate Home. Accessed December 10, 2021. https://www.siemens-healthineers.com/

[14] System Description. zeego@Stanford Lab. Accessed December 10, 2021. https://med.stanford.edu/zeegolab/instrument/system.html

[15] Scott, C. C., Farrier, M., Li, Y., Laxer, S., Ravi, P., Kenesei, P., ... & Karim, K. S. (2021). High-energy micrometre-scale pixel direct conversion X-ray detector. Journal of Synchrotron Radiation, 28(4).

[16] Farrier MG, Scott C, Karim KS, Con C, Li Y. High Energy Micron Scale Pixel Hybrid Detector. Farrier Microengineering LLC, Boyne City, MI; 2019. doi:10.2172/1576188

[17] Keyes, S. D., Daly, K. R., Gostling, N. J., Jones, D. L., Talboys, P., Pinzer, B. R., ... & Roose, T. (2013). High resolution synchrotron imaging of wheat root hairs growing in soil and image based modelling of phosphate uptake. New Phytologist, 198(4), 1023-1029.

[18] Daly, K. R., Keyes, S. D., Masum, S., & Roose, T. (2016). Image-based modelling of nutrient movement in and around the rhizosphere. Journal of experimental botany, 67(4), 1059-1070.

[19] Minami, Y., Imamura, R., Kanno, I., Ohtaka, M., Hashimoto, M., Ara, K., & Onabe, H. (2011). Using x-ray energy information in CT measurement of a phantom with an Al region. Journal of nuclear science and technology, 48(1), 108-112.

[20] Pajor, R., Falconer, R., Hapca, S., & Otten, W. (2010). Modelling and quantifying the effect of heterogeneity in soil physical conditions on fungal growth. Biogeosciences, 7(11), 3731-3740.

The Reason of Why Dynamic Dual-Energy CT is Better than Multi-Energy CT in Reducing Statistical Noise

Yidi Yao, Liang Li, *Senior Member, IEEE* and Zhiqiang Chen

***Abstract*—Multi-energy CT conducted by photon-counting detectors has a wide range of applications, especially in multiple contrast agent imaging. However, multi-energy CT imaging suffers from higher statistical noise because of increased energy bin numbers. Our team has proposed the dynamic dual-energy CT imaging mode and the corresponding iterative imaging algorithms to solve this problem. The multi-energy projections and reconstructions calculated from the dynamic dual-energy CT data are less noisy than the static multi-energy CT, which has been verified by sufficient numerical simulations and experiments. However, a rigorous mathematical derivation has not been conducted to explain why dynamic dual-energy CT is better than static multi-energy CT in reducing statistical noise. In this work, we drive the noise model of the dynamic dual-energy CT to explain the reason. The reason is: compared to the multi-energy projections that are directly measured from a static multi-energy CT, the multi-energy projections, which are calculated from the dynamic dual-energy CT data, have the same expectation, but the variance is lower.**

***Index Terms*— spectral CT, multi-energy CT, CT reconstruction, photon-counting CT, dynamic dual-energy**

I. Introduction

PHOTON counting CT has made significant progress in both technique and clinical application in recent years[1]. Compared to energy-integrating CT, photon-counting CT has advantages in spatial resolution, radiation dose, equal weighting for all photons, etc. Multi-energy imaging is one of the major characteristics of photon counting CT, enabling simultaneous imaging of multiple contrast agents and the future of functional imaging. However, as the number of energy bins increases, the photon counts in each energy bin decrease, which leads to higher statistical noise of the projection data and bad quality of the reconstruction images.

To reconstruct less-noisy multi-energy CT images, we proposed the dynamic dual-energy (DDE) CT to reduce the statistical noise of the multi-energy CT data in our previous works [2-4]. The dual-energy CT data are obtained by applying an adjustable energy threshold in the photon-counting detector. The reconstruction and decomposition results calculated from the DDE CT data are less noisy than the static-energy-threshold multi-energy (SME) CT results, which have been verified through sufficient numerical simulations and experiments. However, why DDE CT outperforms SME CT in reducing statistical noise has not been explained in theory. Because DDE CT can utilize fewer data to reconstruct less-noisy multi-energy CT images is challengeable, giving a mathematical explanation is crucial to make DDE CT more acceptable.

In this work, we analyze DDE CT from the perspective of statistics. The mathematical foundation under DDE CT is: the variance of a Poisson random variable is larger than the variance of a random variable if this variable is calculated from a Poisson random variable with a larger expectation. Applying this principle to DDE CT, the conclusion is: if the ratios among the transmitted photons of different energy bins are accurately known, the noise of the multi-energy projections calculated from the DDE CT data is less than the noise of the multi-energy projections that are directly measured from the SME CT. With the convergence analysis of the simulation results, we further show that the ratios among the transmitted photons of different energy bins can be accurately calculated. In conclusion, we explain why DDE CT is better than SME CT in reducing statistical noise.

The structure of this paper is organized as follows. Since DDE CT is still a new concept, we first briefly review the DDE CT mode and the iterative DDE algorithm for multi-energy CT imaging in Section 2. Section 3 introduces the noise model of the DDE CT. Section 4 presents the convergence analysis of the DDE CT. Section 5 is the conclusion.

II. Review of the Dynamic Dual-Energy CT

A. Dynamic Dual-Energy CT Mode

The SME CT diagram is shown in Figure 1a, and the DDE CT diagram is shown in Figure 1b. In SME CT, there are N_k

This work was supported by the grants from NNSFC 11775124 and 12027811.

Y. Yao is with Department of Engineering Physics, Tsinghua University, Beijing 100084, China, and also with the Key Laboratory of Particle and Radiation Imaging, Tsinghua University, Ministry of Education, Beijing 100084, China.

L. Li is with Department of Engineering Physics, Tsinghua University, Beijing 100084, China, and also with the Key Laboratory of Particle and Radiation Imaging, Tsinghua University, Ministry of Education, Beijing 100084, China (e-mail: lliang@tsinghua.edu.cn).

Z. Chen is with the Department of Engineering Physics, Tsinghua University, Beijing, 100084 China, and also with the Key Laboratory of Particle and Radiation Imaging, Tsinghua University, Ministry of Education, China (e-mail: czq@tsinghua.edu.cn).

energy thresholds for N_k-energy CT imaging. However, in DDE CT, there are only two energy thresholds. The low-energy threshold is unchanged and fixed, while the high-energy threshold changes among different preset values. When a photon is injected into the detector, it will be counted either in the low-energy bin or in the high-energy bin.

The high-energy thresholds change randomly for different detector pixels and different scan views. The preset values for the high-energy thresholds are set according to the requirement of multi-energy imaging. N_k-energy imaging requires N_k -1 preset values for the high-energy thresholds. These values are the same as the values of energy thresholds in an SME CT that aims for the same N_k-energy imaging.

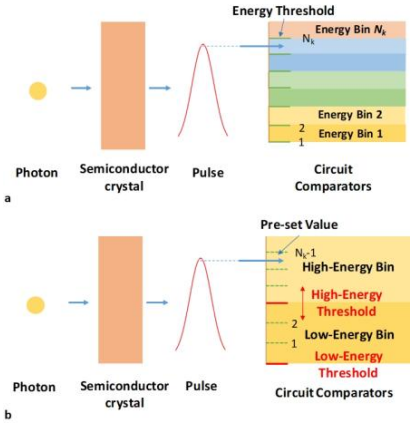


Fig 1. Diagrams of the SME CT (up) and DDE CT (down) [4].

B. Iterative Dynamic Dual-Energy CT Algorithm

In a DDE CT scan, there are two measurements $i_{low,j}$ and $i_{high,j}$ for the transmitted photons of the j_{th} ray. Supposing the measurements follow a Poisson distribution, the corresponding random variables of the measurements $I_{low,j}$ and $I_{high,j}$ are equal to:

$$\begin{aligned} I_{low,j} &= Poisson\{\sum_{k=1}^{N_{th,j}} I_{0,k}\cdot\exp(-[A\cdot X_k]_j)\} \\ I_{high,j} &= Poisson\{\sum_{k=N_{th,j}+1}^{N_k} I_{0,k}\cdot\exp(-[A\cdot X_k]_j)\} \end{aligned} \tag{1}$$

In the above formula, $I_{0,k}$ are the incident photons of the k_{th}-energy bin, which is the same for all X-rays. $A\in\mathbb{R}^{N_j\times N_i}$ is the system matrix. N_j is the number of X-rays. N_i is the number of pixels in a reconstruction image. X_k is the accurate, noise-free reconstruction image of the k_{th} energy bin. $[AX_k]_j$ represents the line integral of the j_{th} X-ray. N_k is the number of total energy bins. $N_{th,j}$ is the number of preset values for the high-energy threshold at the j_{th} ray.

The iterative DDE CT algorithm for calculating the multi-energy projections from the DDE CT data can be concluded to the following two steps in one iteration:

Step 1: Update the multi-energy transmitted photons $\bar{i}_{k,j}$ using the last updated multi-energy CT results $\bar{X}_k$ and the low- and high-energy measurements $i_{low,j}$ and $i_{high,j}$.

Step 2: Update the multi-energy CT images $\bar{X}_k$ with the multi-energy transmitted photons $\bar{i}_{k,j}$.

For step 1, the multi-energy transmitted photons $\bar{i}_{k,j}$ can be calculated according to the following formula:

$$\bar{i}_{k,j}=\begin{cases} \dfrac{I_{0,k}\cdot\exp(-[A\bar{X}_k]_j)}{\sum_{k=1}^{N_{th,j}} I_{0,k}\cdot\exp(-[A\bar{X}_k]_j)}\cdot i_{low,j}, & 1\le k\le N_{th,j} \\ \dfrac{I_{0,k}\cdot\exp(-[A\bar{X}_k]_j)}{\sum_{k=N_{th,j}+1}^{N_k} I_{0,k}\cdot\exp(-[A\bar{X}_k]_j)}\cdot i_{high,j}, & N_{th,j}<k\le N_k \end{cases} \tag{2}$$

For step 2, the SIRT [5] algorithm is used for CT reconstruction of all energy bins. The initial values for the multi-energy CT image $\bar{X}_k$ are set to the values of the mono-energetic CT image, which can be reconstructed from the mono-energetic projections $i_{low,j}+i_{high,j}$.

III. The Noise Model of Dynamic Dual-Energy CT

A. The Property of Poisson Random Variable

In this section, we first discuss a property of the Poisson random variable. Considering a Poisson random variable X , the probability density function for the random variable X is:

$$\mathrm{P}(X=k)=\frac{\lambda^k e^{-k}}{k!} \tag{3}$$

According to (3), the expectation and variance of the random variable X are equal to:

$$\mathrm{E}(X)=\sum_{k=0}^{\infty}k\cdot\frac{\lambda^k e^{-k}}{k!}=\lambda \tag{4}$$

$$\begin{aligned} \mathrm{Var}(X) &= \mathrm{E}(X^2)-(\mathrm{E}(X))^2 \\ &= \sum_{k=0}^{\infty}k^2\cdot\frac{\lambda^k e^{-k}}{k!}-\lambda^2 \\ &= \lambda(\lambda+1)-\lambda^2=\lambda \end{aligned} \tag{5}$$

Consider another Poisson random variable Y , where the expectation of Y is $t\lambda$ ($0<t<1$). According to (4) and (5), the expectation and variance of Y are:

$$\mathrm{E}(Y)=\mathrm{Var}(Y)=t\lambda \tag{6}$$

Consider a new random variable Z . Its definition is:

$$Z=tX,\quad 0<t<1 \tag{7}$$

The probability density function for Z is:

$$\mathrm{P}(Z=tk)=\frac{\lambda^k e^{-k}}{k!} \tag{8}$$

The expectation and variance of Z are:

$$\mathrm{E}(Z)=\sum_{k=0}^{\infty}tk\cdot\frac{\lambda^k e^{-k}}{k!}=t\lambda \tag{9}$$

$$\begin{aligned} \mathrm{D}(Z) &= \mathrm{E}(Z^2)-(\mathrm{E}(Z))^2 \\ &= \sum_{k=0}^{\infty}t^2k^2\cdot\frac{\lambda^k e^{-k}}{k!}-t^2\lambda^2 \\ &= t^2\lambda(\lambda+1)-t^2\lambda^2 \\ &= t^2\lambda \end{aligned} \tag{10}$$

Random variables Y and Z have the same expectation, while the variance of Z is smaller than the variance of Y . The conclusion for the above derivation is: the variance of a Poisson random variable (Y) is larger than the variance of a random variable (Z), if this variable (Z) is calculated from the Poisson random variable with a larger expectation (X).

B. The Noise Model of Dynamic Dual-Energy CT

In this section, we drive the noise model of the DDE CT data and compare it to the SME CT data. Considering an N_k-energy SME CT, the number of transmitted photons in the k_{th}-energy bin I_k is a Poisson random variable. The number of transmitted photons of all energy bins I_{total} is also a Poisson random variable, and they satisfy the following relationship:

$$I_{total}=\sum_{k=1}^{N_k} I_k \tag{11}$$

Therefore, the expectations of these random variables satisfy the following relationship:

$$\mathrm{E}[I_k]=t_k\cdot\mathrm{E}[I_{total}],\ \text{where } 0<t_k\le 1,\ \sum_k t_k=1 \tag{12}$$

If the expectation of the number of total transmitted photons is λ_0, according to (6), the expectations and variances for the number of transmitted photons in different energy bins are:

$$\mathrm{E}[I_k]=\mathrm{Var}[I_k]=t_k\cdot\lambda_0 \tag{13}$$

Now consider the DDE CT data. In DDE CT, the number of transmitted photons of the low-energy bin and the high-energy bin I_{low} and I_{high} are Poisson random variables and satisfy the following relationships:

$$I_{total}=I_{low}+I_{high} \tag{14}$$

$$\begin{aligned} I_{low}&=\sum_{k=1}^{N_{th}} I_k \\ I_{high}&=\sum_{k=N_{th}+1}^{N_k} I_k \end{aligned} \tag{15}$$

Because the number of transmitted photons measured in the low-energy bin and the high-energy bin in a DDE CT are Poisson random variables, according to (13) and (15), their expectations and variations are equal to:

$$\begin{aligned} \mathrm{E}[I_{low}]&=\mathrm{Var}[I_{low}]=(\sum_{k=1}^{N_{th}} t_k)\cdot\lambda_0 \\ \mathrm{E}[I_{high}]&=\mathrm{Var}[I_{high}]=(\sum_{k=N_{th}+1}^{N_k} t_k)\cdot\lambda_0 \end{aligned} \tag{16}$$

In DDE CT, the number of transmitted photons of multiple energy bins is not obtained from direct measurement but is calculated from the low- and high-energy bin data:

$$\bar{I}_k=\begin{cases} \dfrac{t_k}{\sum_{k=1}^{N_{th}} t_k}\cdot I_{low}, & 1\le k\le N_{th} \\ \dfrac{t_k}{\sum_{k=N_{th}+1}^{N_k} t_k}\cdot I_{high}, & N_{th}<k\le N_k \end{cases} \tag{17}$$

$\bar{I}_k$ is still a random variable and represents the number of transmitted photons of the k_{th} energy bin. According to (9), (10), and (16), the expectation and variance of $\bar{I}_k$ are:

$$\mathrm{E}[\bar{I}_k]=t_k\cdot\lambda_0 \tag{18}$$

$$\mathrm{Var}[\bar{I}_k]=\begin{cases} t_k\cdot\dfrac{t_k}{\sum_{k=1}^{N_{th}} t_k}\cdot\lambda_0, & 1\le k\le N_{th} \\ t_k\cdot\dfrac{t_k}{\sum_{k=N_{th}+1}^{N_k} t_k}\cdot\lambda_0, & N_{th}<k\le N_k \end{cases} \tag{19}$$

Comparing the expectation and variance of the number of multi-energy transmitted photons in SME CT and DDE CT, we can obtain the following relationship:

$$\mathrm{E}[\bar{I}_k]=\mathrm{E}[I_k] \tag{20}$$

$$\mathrm{Var}[\bar{I}_k]=\begin{cases} \dfrac{t_k}{\sum_{k=1}^{N_{th}} t_k}\cdot\mathrm{Var}[I_k], & 1\le k\le N_{th} \\ \dfrac{t_k}{\sum_{k=N_{th}+1}^{N_k} t_k}\cdot\mathrm{Var}[I_k], & N_{th}<k\le N_k \end{cases} \tag{21}$$

For all k, the following inequalities always satisfy:

$$\frac{t_k}{\sum_{k=1}^{N_{th}} t_k}<1,\quad \frac{t_k}{\sum_{k=N_{th}+1}^{N_k} t_k}<1 \tag{22}$$

Therefore, we obtain the following conclusion: compared to the multi-energy CT transmitted photons I_k that are directly measured from an SME CT, the multi-energy transmitted photons $\bar{I}_k$, which are calculated from the DDE CT data, have the same expectation, but the variance is lower.

C. Verification

To verify the theory proposed in Section 3B, we simulate a large amount of data. We calculate the variances of the multi-energy CT transmitted photons for these simulation results and verify whether the relationship between the variances of the SME CT and the DDE CT is consistent with the theory.

A 20 cm diameter water cylinder is scanned in the simulation. The number of energy bins is 8. The variance images of the third-energy-bin SME CT transmitted photons and the third-energy-bin DDE CT transmitted photons are calculated from 1000 samples and shown in Figures 2a and 2b, respectively. If the proposed theory is correct, the following formula should be satisfied:

$$\mathrm{Ratio}_{k,j}\cdot\mathrm{Var}[prj_{k,j}]/\mathrm{Var}[\overline{prj}_{k,j}]=1, \tag{23}$$

$$\text{where } \mathrm{Ratio}_{k,j}=\begin{cases} \dfrac{t_{k,j}}{\sum_{k=1}^{N_{th,j}} t_{k,j}}, & 1\le k\le N_{th,j} \\ \dfrac{t_k}{\sum_{k=N_{th,j}+1}^{N_k} t_{k,j}}, & N_{th,j}<k\le N_k \end{cases}$$

k is the number of energy bins. j is the number of X-rays. The ratio image of the third energy bin is shown in Figure 2c. The third-energy bin calculation results are shown in Figure 2d. The mean value of Figure 2d is 1.0022, which is extremely close to 1. This result verifies the correctness of the noise model proposed in Section 3B.

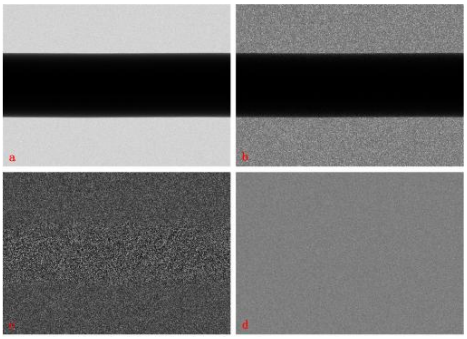


Fig 2. The variance images and calculation results of the third energy bin. 2a and 2b are variance images of the SME CT and DDE CT, respectively. 2c is the ratio image. 2d is the calculation result. The display windows for 2d is [0.5, 1.5].

IV. The Noise Model of Dynamic Dual-Energy CT

A. Calculating Accurate t_k in the Dynamic Dual-Energy CT

Section 3 proves that the noise of projection calculated from DDE CT is lower than that calculated from SME CT. However, there is a premise for this conclusion: the ratios between the number of transmitted photons of a specific energy bin and all energy bins are accurately known, e.g., t_k is known and accurate for any k. In theory, t_k can be accurately calculated only if the noise-free transmitted photons of different energy bins are known. t_k cannot be calculated directly from the original DDE CT data.

The iterative DDE CT algorithm, which is described in Section 2B, is proposed to calculate accurate t_k. When the algorithm converges, (2) will be rewritten as:

$$\bar{i}_{k,j,converge}=\begin{cases}\dfrac{t_{k,j}}{\sum_{k=1}^{N_{th,j}} t_{k,j}}\cdot i_{low,j}, & 1\le k\le N_{th,j}\\ \dfrac{t_{k,j}}{\sum_{k=N_{th,j}+1}^{N_k} t_{k,j}}\cdot i_{high,j}, & N_{th,j}<k\le N_k\end{cases},\quad t_{k,j}=\frac{E[I_{k,j}]}{\sum_{k=1}^{N_k}E[I_{k,j}]} \tag{24}$$

$\bar{i}_{k,j,converge}$ is the convergent value for the multi-energy transmitted photons calculated from DDE CT and is an observation of the random variable defined in (17). Therefore, according to the conclusion in Section 3B, $\bar{i}_{k,j,converge}$ has lower statistical noise than $i_{k,j}$, which is the transmitted photon measured from the SME CT. In other words, the multi-energy projections calculated from the DDE CT data are less noisy than the SME CT if the iterative DDE algorithm converges.

B. Convergence of the Iterative Dynamic Dual-Energy Algorithm

The convergence of the iterative DDE algorithm is verified through simulation. The XCAT thorax phantom [6] is scanned in the simulation. The simulation configurations are shown in Table 1.

TABLE I
CONFIGURATIONS OF THE SIMULATION

Parameter	Value
Scan Method	2D fan-beam
Source Voltage	120 kV
Distance between Source and Detector	100 cm
Distance between Source and Gantry Center	50 cm
Views over 360 Degrees	720
Number of Detectors	1024
Detector Length	1 m
Size of Reconstruction Image	512 * 512
Reconstruction Pixel Size	1 mm * 1 mm
Number of Energy Bins	8
Number of Incident Photons for Each Ray	$5*10^6$

The noise of the projections is measured in terms of the relative error, which is defined as:

$$\text{RelativeError}(prj)=\frac{\|prj-prj_{truth}\|^2}{\|prj_{truth}\|^2}\times 100\% \tag{25}$$

prj_{truth} is the ground truth of the line integrals of multi-energy CT, which can be obtained from the noise-free SME CT. In the analysis of the convergence, the relative error is calculated for line integrals of three types: the multi-energy line integrals measured in SME CT, the multi-energy line integrals calculated from DDE CT data, and the convergent values for the multi-energy line integrals from DDE CT. If the algorithm converges, the second type of relative error will decrease to the same value as the third type of relative error.

The convergence curve is plotted in Figure 3, which shows that the algorithm converges. This illustrates that the iterative DDE algorithm can accurately calculate t_k. Moreover, the relative errors of the multi-energy line integrals from the DDE CT are much smaller than the relative errors of the multi-energy line integrals from the SME CT, which indicates that a less-noisy multi-energy projection can be calculated from the DDE CT.

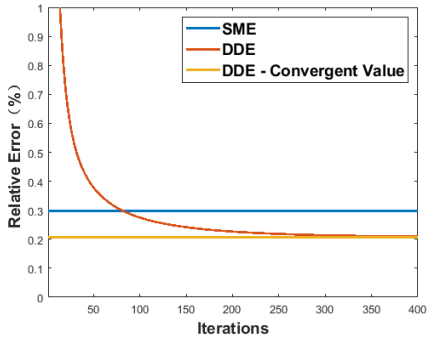


Fig 3. The convergence curve.

V. Conclusion

In this work, we derive the noise model of DDE CT to explain why DDE CT can reduce statistical noise better than SME CT. Based on this derivation, we further analyze the convergence of DDE CT and show that the multi-energy projection calculated from DDE CT is less noisy than that calculated from SME CT.

References

[1] M. Danielsson, M. Persson, and M. Sjolin, "Photon-counting x-ray detectors for CT," *Phys. Med. Biol.,* vol. 66, no. 3, Feb 7 2021, Art no. 03tr01, doi: 10.1088/1361-6560/abc5a5.

[2] L. Li, Z. Chen, W. Cong, and G. Wang, "Spectral CT modeling and reconstruction with hybrid detectors in dynamic-threshold-based counting and integrating modes," *Ieee T Med Imaging,* vol. 34, no. 3, pp. 716-728, 2015.

[3] Y. D. Yao, L. Li, and Z. Q. Chen, "Dynamic-dual-energy spectral CT for improving multi-material decomposition in image-domain," (in English), *Phys. Med. Biol.,* Article vol. 64, no. 13, p. 22, Jul 2019, Art no. 135006, doi: 10.1088/1361-6560/ab196d.

[4] Y. Yao, L. Li, and Z. Chen, "Iterative dynamic dual-energy CT algorithm in reducing statistical noise in multi-energy CT imaging," *Physics in Medicine & Biology,* 2021. [Online]. Available: http://iopscience.iop.org/article/10.1088/1361-6560/ac459d.

[5] J. Gregor and T. Benson, "Computational analysis and improvement of SIRT," *Ieee T Med Imaging,* vol. 27, no. 7, pp. 918-924, Jul 2008, doi: 10.1109/tmi.2008.923696.

[6] W. P. Segars, G. Sturgeon, S. Mendonca, J. Grimes, and B. M. W. Tsui, "4D XCAT phantom for multimodality imaging research," (in English), *Medical Physics,* Article vol. 37, no. 9, pp. 4902-4915, Sep 2010, doi: 10.1118/1.3480985.

Cone-Beam X-ray Luminescence Computed Tomography Reconstruction Based on Huber Markov Random Field Regularization

Tianshuai Liu, Junyan Rong, Wenqin Hao, Hongbing Lu
Department of Biomedical Engineering, Fourth Military Medical University, Xi'an, Shaanxi 710032, China

ABSTRACT

In recent years, cone-beam X-ray luminescence computed tomography (CB-XLCT) has drawn much attention with the development of X-ray excited nanophosphors. Compared with traditional bio-optical imaging modalities such as bioluminescence tomography (BLT) and fluorescence molecular tomography (FMT), CB-XLCT can effectively improve imaging sensitivity and depth because of the reduction of background fluorescence and the high penetrability of X-rays. However, due to high degree of scattering of light through biological tissues, the reconstruction of CB-XLCT is inherently ill-conditioned. To solve the ill-posed inverse problem, appropriate priors or regularizations are needed to facilitate the reconstruction. Based on the fact that adjacent pixels generally have the same or similar concentration and in order to further balance the degree of regional smoothness and edge sharpening, a prior information model based on Huber Markov Random Field (HuMRF) was established to constrain the reconstruction process of CB-XLCT. Mice experiments indicate that compared with the traditional ART and ADAPTIK method, the proposed method could improve the image quality of CB-XLCT significantly in terms of target shape, localization accuracy and image contrast.

Keywords: X-ray luminescence computed tomography, image reconstruction techniques, Huber Markov Random Field, mice experiments

1. INTRODUCTION

With the advances of X-ray excitable nanophosphors, cone-beam X-ray luminescence computed tomography (CB-XLCT) has attracted more attention for its promising performance[1]. In CB-XLCT, X-ray excitable nanophosphors are used as imaging probes and emit visible or near-infrared (NIR) light when irradiated by X-rays. The photons excited by X-rays arrive at the surface of the imaging object and can be measured by sensitive photon detectors. By solving an inverse problem using an appropriate imaging model of X-ray and photon transport, the three-dimensional (3-D) distribution of the nanophosphors in the imaged object can be resolved. Compared with traditional bio-optical imaging modalities such as bioluminescence tomography (BLT) and fluorescence molecular tomography (FMT), CB-XLCT can effectively improve imaging sensitivity and depth because of the reduction of background fluorescence and the high penetrability of X-rays[2]. However, due to high degree of scattering of light through biological tissues, the reconstruction of CB-XLCT is inherently ill-conditioned. In order to improve the reconstruction quality of CB-XLCT, the priori information is needed to constrain the reconstruction process.

In this study, we propose a reconstruction approach based on Huber Markov Random Field (HuMRF) for the CB-XLCT reconstruction. The remainder of this paper is organized as follows. In Section 2, the proposed method is described in detail. In Section 3, the mice experiments design and results are described for the performance evaluation of the proposed reconstruction approach. Finally, conclusion is given in Section 4.

2. METHODS

2.1 Forward model of XLCT and Inverse Problem based on the proposed GHuMRF algorithm

Based on the forward model of CB-XLCT[3]:

$$Wn = \Phi_{meas} \tag{1}$$

In practical application of XLCT, noise of the XLCT imaging system needs to be considered, and equation (1) becomes:

$$y = \Phi_{meas} + \varsigma = Wx + \varsigma \tag{2}$$

where $y = [y_1, y_2, \cdots, y_M]^T$ represents the actual fluorescence signals measured on the surface of the imaging object, $\varsigma = [\varsigma_1, \varsigma_2, \cdots, \varsigma_M]^T$ is the noise of the system, W is the weight matrix, $x = n$ represents the unknown distribution of nanophosphors in the imaging object.

Based on Bayes theory, the maximum a posteriori (MAP) estimation of the unknown distribution of nanophosphors in the imaging object can be expressed as[4]:

$$\hat{x}_{MAP} = \arg\max_x \{\log p(x/y)\} = \arg\max_x \{\log p(y/x) + \log p(x)\} \tag{3}$$

where $p(x/y)$ represents the posterior probability density function, $p(y/x)$ represents the conditional probability function (measurement model), $p(x)$ represents the priori probability density function (priori model).

Since the working temperature of the EMCCD camera is very low, the measurement model is constructed based on a shot-noise model, which assumes the independent measurement noise can be described by a Gaussian distribution:

$$p(y/x) = \frac{1}{(\pi\kappa)^M \left|\Lambda_y\right|} \exp\left[-\frac{1}{\kappa} \left\| y - Wx \right\|^2_{\Lambda_y^{-1}} \right] \tag{4}$$

where M represents the number of measurement points，κ is the unknown hyperparameters related to noise variance， Λ_y is the covariance matrix of the fluorescence measurement signal.

Based on the adjacent pixels generally have the same or similar concentration and in order to further balance the degree of regional smoothness and edge sharpening, the priori model is constructed based on Huber Markov Random Field (HuMRF):

$$p(x) = \frac{1}{Z\sigma^N} \exp\left[-\frac{1}{2\sigma^2} \sum_{j=1}^{N} \sum_{k \in \partial j} b_{j-k} \phi(\Delta) \right] \tag{5}$$

where $\Delta = x_j - x_k$ ， $\phi(\Delta) = \begin{cases} \Delta^2 & |\Delta| \le \delta \\ 2\delta|\Delta| - \delta^2 & |\Delta| > \delta \end{cases}$ ，

δ is an adjustable parameter to balance the smoothness of the region and the sharpness of the boundary, b_{j-k} is the weight coefficients between the jth and kth pixels which is inversely proportional to the distance between the pixels.

2.2 Quantitative evaluation

The quality of reconstructed CB-XLCT images was evaluated quantitatively by several indexes including the location error (*LE*), dice similarity coefficient (*DICE*) and contrast-to-noise ratio (*CNR*) [3].

LE evaluates the localization accuracy of the reconstructed target, which is defined as the Euclidean distance error between the centers of true and reconstructed targets:

$$LE = \left\| \mathbf{L_r} - \mathbf{L_t} \right\|_2 \tag{6}$$

where $\boldsymbol{L_r}$ and $\boldsymbol{L_t}$ denote the centers of the reconstructed and true targets, respectively.

DICE reflects the similarity of the true and reconstructed targets and can be calculated by:

$$DICE = \frac{2\left|\mathbf{ROI_r} \cap \mathbf{ROI}_t\right|}{\left|\mathbf{ROI_r}\right| + \left|\mathbf{ROI_t}\right|} \tag{7}$$

where $\boldsymbol{ROI_t}$ and $\boldsymbol{ROI_r}$ denote the regions of true and reconstructed targets, respectively, and $|\cdot|$ defines the number of voxels in a region.

CNR is used for quantitative evaluation of noise and artifacts in reconstructed images, as shown below:

$$CNR = \frac{\left|\mu_{ROI} - \mu_{BCK}\right|}{(w_{ROI}\sigma^2_{ROI} + w_{BCK}\sigma^2_{BCK})^{1/2}} \tag{8}$$

where *ROI* and *BCK* denote the target and background regions of the imaged object, w_{ROI} and w_{BCK} are weighting factors determined by the relative volumes of the target and background, μ_{ROI} and μ_{BCK} are the mean intensity values of the *ROI* and *BCK*, and σ^2_{ROI} and σ^2_{BCK} represent the variances of the ROI and BCK, respectively.

3. EXPERIMENTAL DESIGN AND RESULTS

Mice experiments were performed to evaluate the performance of the proposed method based on the custom-developed CB-XLCT system in our laboratory. All experiments were conducted in compliance with the provisions of the Animal Ethics Review Committee of the Air Force Military Medical University. For comparison, two traditional methods, algebra reconstruction technique (ART), and adaptive tikhonov regularization (ADAPTIK) were also implemented to reconstructed the image.

A female BALB/c nude mice was used in this experiment. A small glass tubes (3mm in diameter) filled with Y_2O_3: Eu^{3+} (60mg/ml) was embedded into the abdominal cavity of mice to serve as the nano-probe in XLCT imaging. During imaging experiments, the mice were fixed on the rotation stage. The voltage and current of the X-ray source were set as 50kVp and 1mA, respectively. The mice was rotated from 0° to 360° and the optical images were obtained every 15° by the EMCCD camera. The exposure time of the EMCCD camera was set as 2s, with the EM gain set as 260.

The XLCT tomographic images were reconstructed with different algorithms in the mouse experiments, as shown in Fig. 1. All the reconstruction results are normalized based on their maximum values. Fig. 1.（c），（e），（g）show the reconstruction results based on ART, ADAPTIK and the proposed methods respectively. Fig. 1. (d), (f), (h) are the fusion results of XLCT and XCT reconstructed images. It can be seen that compared with the traditional ART and ADAPTIK method, the proposed method could improve the image quality of CB-XLCT significantly in terms of target shape, localization accuracy and image contrast.

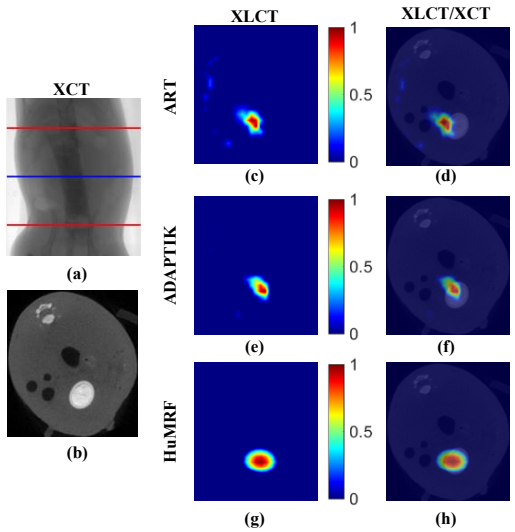


Fig. 1. The tomographic images were reconstructed based on ART, ADAPTIK and the proposed methods.

The quality of reconstructed CB-XLCT images is given in Table 1. For the Mice experiments, the reconstruction results based on the proposed the proposed GHuMRF algorithm yield the highest Dice and CNR with lowest LE. The results indicate that proposed GHuMRF algorithm performs better in target location, shape recovery and image contrast, when compared to the conventional reconstruction methods of ART and ADAPTIK, which further confirm the observation in Fig. 1.

Table 1 Quantitative Evaluation on Mice experiments reconstructions using different methods

	LE(mm)	DICE	CNR
ART	2.5	0.51	2.63
ADAPTIK	1.8	0.6	3.5
GHuMRF	0.9	0.9	5.28

4. DISCUSSION AND CONCLUSIONS

In this study, a reconstruction approach based on HuMRF regularization is proposed for the CB-XLCT inverse problem. Mice experiments indicate that compared with the traditional ART and ADAPTIK method, the proposed method could improve the image quality of CB-XLCT significantly in terms of target shape, localization accuracy and image contrast.

It should be noted that Both Gaussian noise model [5]and Poisson noise model [6] can be used to simulate the measurement noise of EMCCD camera. Generally, Poisson distribution tends to Gaussian distribution with the increase of sample size. In this paper, the Gaussian noise model is used to simulate the measurement noise of EMCCD camera. In the further in vivo imaging research the Poisson noise model can be used to construct the measurement model, because the number of photons collected by EMCCD camera is relatively lower.

In the mouse experiment, the mice were placed in deep anesthesia. When the turntable rotates, it uses a low speed of 6°/s while maintaining a uniform rotation speed, to minimize the adverse effects of internal organ movement in mice during data collection. Moreover, the mouse was monitored throughout the data acquisition process to ensure that its position would not change. Therefore, in this paper, for mouse experiments, motion correction was not performed during reconstruction.

It is necessary to accurately segment the organs of the imaging object, because of giving corresponding optical parameters to different organs is important to obtain high-quality reconstructed images. Therefore, a more accurate image segmentation algorithm needs to be used for organ segmentation in mice

In this paper, because CT reconstruction requires more projections, XCT projection data and optical detection signals are collected twice. In future experiments, the simultaneous acquisition of XCT projection data and optical detection signals can reduce the scanning time, radiation dose and the impact of organ movement between two scans on the reconstruction results.

In summary, based on Bayes theory, a CB-XLCT reconstruction method based on GHuMRF algorithm is proposed in this paper. Compared with the traditional ART and ADAPTIK method, the GHuMRF algorithm add spatial constraints between adjacent pixels during reconstruction, which improves the image quality of CB-XLCT reconstruction. In the further study, more prior information constraints can be added to CB-XLCT reconstruction based on Bayes theory to improve the image quality of CB-XLCT.

Funding

National Natural Science Foundation of China (NSFC) (11805274)

National Key Research and Development Program of China (2017YFC0107400, 2017YFC0107403)

Key Research and Development Program of Shaanxi Province (2020SF-214, 2020SF-208)

Disclosures

The authors declare that there are no conflicts of interest related to this article.

References

[1] G. Pratx, C. Carpenter, C. Sun, and L. Xing, "X-ray luminescence computed tomography via selective excitation: a feasibility study," *IEEE Trans Med Imaging,* vol. 29, pp. 1992-9, 2010.
[2] T. Liu, J. Rong, P. Gao, H. Pu, W. Zhang, X. Zhang*, et al.*, "Regularized reconstruction based on joint L 1 and total variation for sparse-view cone-beam X-ray luminescence computed tomography," *Biomedical Optics Express,* vol. 10, pp. 1-17, 2019.
[3] T. Liu, J. Rong, P. Gao, W. Zhang, W. Liu, Y. Zhang*, et al.*, "Cone-beam x-ray luminescence computed tomography based on x-ray absorption dosage," *J Biomed Opt,* vol. 23, pp. 1-11, 2018.
[4] G. Zhang, F. Liu, J. Liu, J. Luo, Y. Xie, J. Bai*, et al.*, "Cone beam x-ray luminescence computed tomography based on Bayesian method," *IEEE transactions on medical imaging,* vol. 36, pp. 225-235, 2017.
[5] D. Chen, S. Zhu, H. Yi, X. Zhang, D. Chen, J. Liang, et al., "Cone beam x-ray luminescence computed tomography: a feasibility study," Med Phys, vol. 40, p. 031111, 2013.
[6] W. Cong, H. Shen, and G. Wang, "Spectrally resolving and scattering-compensated x-ray luminescence/fluorescence computed tomography," J Biomed Opt, vol. 16, p. 066014, 2011.

Dual-domain network with transfer learning for reducing bowtie-filter induced artifacts in half-fan cone-beam CT

Sungho Yun, Uijin Jeong, Donghyeon Lee, Hyeongseok Kim, and Seungryong Cho

***Abstract*—In a cone-beam CT system, the use of bowtie-filter may induce artifacts in the reconstructed images. Through a Monte-Carlo simulation study, we confirm that the bowtie filter causes spatially biased beam energy difference thereby creating beam-hardening artifacts. We also note that cupping artifacts in conjunction with the object scatter and additional beam-hardening may manifest. In this study, we propose a dual-domain network for reducing the bowtie-filter induced artifacts by addressing the origin of artifacts. In the projection domain, the network compensates for the filter induced beam-hardening effects. In the image domain, the network reduces the cupping artifacts that generally appear in cone-beam CT images. Also, transfer learning scheme was adopted in the projection domain network to reduce the total training costs and to increase utility in the practical cases while maintaining the robustness of the dual-domain network. Thus, the pre-trained projection domain network using simple elliptical cylinder phantoms was utilized. As a result, the proposed network shows denoised and enhanced soft-tissue contrast images with much reduced image artifacts. For comparison, a single image domain U-net was also implemented as an ablation study. The proposed dual-domain network outperforms, in terms of soft-tissue contrast and residual artifacts, a single domain network that does not physically consider the cause of artifacts.**

I. Introduction

In a clinical cone-beam CT system, the bowtie-filter is often used to homogenize projection data across the field-of-view thereby better utilizing detector response characteristics and partly to reduce the amount of object scatter. However, it can induce eclipse shape artifacts that typically have a bright ring and dark shade in the reconstructed image as shown in Fig 1. Artifacts from the bowtie-filter may build up in various forms depending on the geometric shape of the filter and the scanning system conditions. Since the artifacts can severely degrade the soft-tissue contrast in CT images, there have been studies for reducing these artifacts and for clarifying their physical causes [1-3].

Due to the elliptic shape nature of human anatomy in transverse plane, the bowtie-filter gradually thickens towards the outside from the principal ray projection position. Therefore, the energy of the incident x-ray from the source can be hardened when it passes through the ticker part of the filter. This spatially varying spectral incident beam, together with the nonlinear detector energy-response characteristics, can cause data inconsistencies from the linear imaging model resulting in peculiar beam-hardening artifacts such as eclipse artifacts. M. Cai et al. [3] introduced a decoupling technique that decomposes the artifacts into bowtie filter-induced beam-hardening and object-induced cupping artifacts. The decoupling scheme is valid from the physical perspectives in that the artifacts originated from the bowtie-filter can be separated from the object-originated ones. However, the suggested method in [3] requires heuristic parameter optimization in each case and has to handle the correction mismatch in an iterative reconstruction framework that requires heavy computational cost. A neural-network-based artifact correction method can be an alternative; a single domain network (e.g., in image-domain) may not be a suitable candidate though since physical factors are hardly incorporated in such a network, possibly resulting in residual artifacts and structural distortions. In recent studies, a dual-domain network has been introduced to partly incorporate physical factor issues in the projection domain and shown promising results in metal artifacts reduction [4].

In this study, we propose a dual-domain network for reducing the bowtie-filter induced artifacts by addressing the causes of artifacts in the respective domains efficiently. In the projection domain, the network compensates for the filter induced beam-hardening effects. In the image domain, the network solves the cupping artifacts that are more associated with the imaged object. Also, transfer learning scheme was adopted in the projection domain training to reduce the total training costs and to increase utility in the practical cases.

II. Method

A. Decoupling of Bowtie Artifacts

Through a Monte-Carlo simulation study, we have confirmed that the bowtie filter causes spatially biased beam energy distribution and creates an eclipse shape of beam-hardening artifacts in the image on top of the object-induced cupping artifacts. Figure 1 shows example slice images of the reconstructed uniform cylinder phantom. The two cases represent different shapes of the bowtie-filters. Overall the

This research was supported in part by Korean National Research Foundation [NRF-2020R1A2C2011959, NRF-2021M3I1A1097909]

S. Yun, U. Jeong, D. Lee and S. Cho are with the Department of Nuclear and Quantum Engineering, Korea Advanced Institution of Science and Technology, Yuseong-gu, Daejeon, 34141 Korea (e-mail: ysh3399@kaist.ac.kr). H. Kim is with KI Artificial Intelligence, KAIST, Korea. S. Cho is in correspondence of this paper. (e-mail: scho@kaist.ac.kr).

images are subject to cupping due to the object scatter and beam-hardening, and the central ring-shaped blooming is due to the bowtie-filter. Thus, we aim at correcting for the bowtie-filter artifacts in the projection domain and by doing so decoupling the bowtie-filter artifacts from the object related ones.

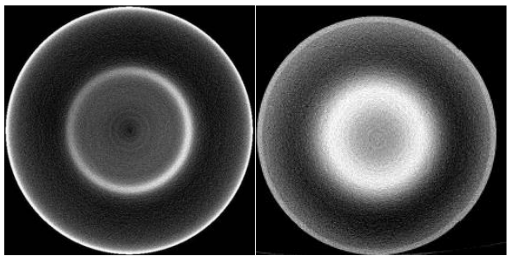


Fig. 1. Bowtie filter artifacts under two different bowtie-filter designs. The eclipse artifacts occurred in the images that have a bright ring and dark shade inside on top of the global cupping.

B. *Dual-domain network with transfer-learning*

The transfer learning can be applied when the size of datasets is small but when there exists a pre-trained network that performs similar tasks. In our case, the transfer-learning scheme was adopted in the projection domain network that intend to do the filter-induced beam-hardening correction. This is due to the limited availability of patient projection data under specific geometric conditions (filter shapes or system geometry). Therefore, simple elliptical cylinder phantoms that are easy to implement were used to prepare a pretrained network. After the projection-domain network, the image domain network addresses the remaining cupping artifacts, which can be interpreted as bottleneck feature in the transfer learning scheme.

C. *Cone-beam CT system using Monte-Carlo simulation*

The system geometry was designed according to Nano Focus Ray (inc)'s Phion v2.0 CBCT system and we used GPU-based Geant4 Monte-Carlo simulation tool [5]. The system uses a half-fan scan mode with a half-bowtie filter (aluminum based filter). The source to detector distance (SDD) and source to object distance (SOD) are 835mm and 480mm, respectively. In addition, an array size of 256x240 detector was used for data acquisition. The tube voltage was set to be 110kVp. The detailed scanning parameters are summarized in TABLE I

TABLE I
GEOMETRY TABLE

SDD	835mm
SOD	480mm
Detector pitch	1.16x1.16mm
Detector resolution	256x240
Thickness of detector	20mm
Tube voltage	110kVp
Angular step size	1 degree

D. *Projection domain training*

For projection domain training, 6 sets of elliptical cylinder water phantoms with its height of 400 mm were prepared with different ellipticities. They have a fixed major axis diameter of 265mm (the value that nearly fulfills the FOV under the given geometry) and only the minor axis diameter changes from 115mm to 241mm, which corresponds to an ellipticity ranging from 0.4 to 0.9. The patient shape is assumed to be a kind of elliptical cylinder and its ellipticity would range in the targeted range above. 360 projections were acquired for each phantom over a full rotation and half of them were used for training data considering redundant information of the elliptic phantoms placed in the isocenter. A total of 1080 projections from six phantoms were used for training. The projection data with the bowtie filter were used as inputs and the data without the bowtie filter were used as labels. We used projection image of the bowtie filter alone, i.e., without imaged object, as an additional input. The data were divided into 810 pairs for the training set and 270 pairs for the validation set for the network training.

One thing we would like to note is that we used a frequency splitting technique when preparing the input considering the drastic difference of the simple phantoms and the patient data. In order to make the input patient data more consistent with the elliptic cylinder phantoms, low-frequency information was extracted and fed into the network for removing the bowtie-filter artifacts which are also low-frequency dominated. A Gaussian filter was used for frequency splitting and only the low-frequency of data was given as an input to the network. After the bowtie-filter network, the high-frequency information was added to form the corrected projection data.

By doing so, beam-hardening correction can be effectively done by low-frequency matching high-frequency components of the patient projection data are preserved. For the network model, residual U-net was used with mean squared error loss.

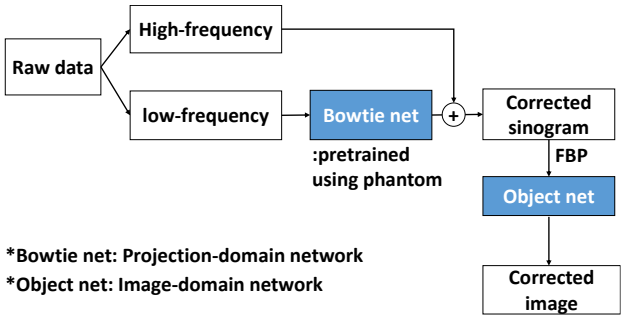


Fig. 3. Total workflow of study. The frequency split technique was applied before projection domain training for preventing loss of high-frequency information of input.

E. *Image domain Training*

For image domain training, 5 clinical patient CT volume data from Mayo clinic (AAPM low-dose CT challenge dataset) were used. The abdominal part of clinical data was segmented into 30 different materials with different densities based on HU unit to create ground-truth material maps. Then, polychromatic forward projection of the maps with known source energy spectrum was performed to acquire projection data and they

were reconstructed to create ground-truth images for training. Also, to create bowtie artifact corrupted patient projection image, the bowtie filter inserted Monte-Carlo simulation was conducted. These data were passed to the pre-trained projection-domain network as explained above and reconstructed to use as input data for the image-domain network training. The reconstructed volume size is 256x256x240 for each patient.

Among total 1200 paired data, 720 pairs of data were used for the training set, 240 pairs of data were used for the validation, and 240 pairs of data were used for the test. For the network model, residual U-net was used with mean square error loss.

III. Results and Discussion

A. Results of projection domain network

The network was trained to 24k epochs with 5e-6 learning rate with Adam optimizer. The validation loss was almost saturated over 16k epoch and it converged to 5.50e-5. The network parameter weight at the 23kth epoch was used for this study. The training loss of that point is 3.03e-6 and the validation loss is 5.49e-5. In Fig 4 (a), (b), and (c), the results of validation images are shown. The compensation was successfully done and also line-profile of projection was almost recovered as label data

Then, the patient projection data are also given to this pre-trained network. The beam-hardening due to the bowtie-filter was well recovered, and the line profile shows such recovery as shown in Fig 4. (d) and (e). Please note that there is no ground-truth for the patient projection in this case. These results are reconstructed and compared with the original image in Fig 5.

As shown in Fig. 5, the eclipse artifact was clearly removed, and only cupping artifact remains. Also, due to the frequency split technique, the high-frequency structures are well preserved. We believe that removing these eclipse artifacts would help the image-domain network remove the remaining cupping artifacts. A single image-domain network would have to process the compounding artifacts without incorporating different physical factors.

B. Results of image domain network

The network trained to 1.4k epochs with 1e-4 learning rate with Adam optimizer. The training loss was 1.58e-7, validation loss was 1.76e-7 and test loss was 1.78e-7. To show the robustness of our dual-domain network, the results were compared with a single image-domain U-net. The compared network was trained with the bowtie artifact corrupted patient image and label patient image in the image domain.

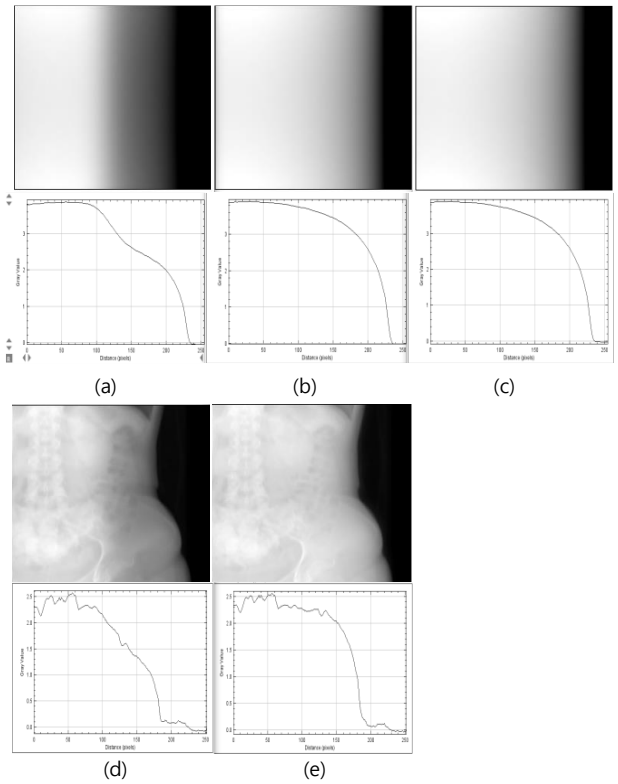


Fig. 4. (a) is input image with line-profile, (b) is output image with line-profile, (c) is label image with line-profile. (d) is input image with line-profile of patient projection data, (e) is output image with line-profile of patient projection data. The network successfully compensates the beam-hardening artifacts and recovered the line-profile. The window level is [0.00 3.98] for (a),(b) and (c) and [0.00 2.82] for (d) and (e)

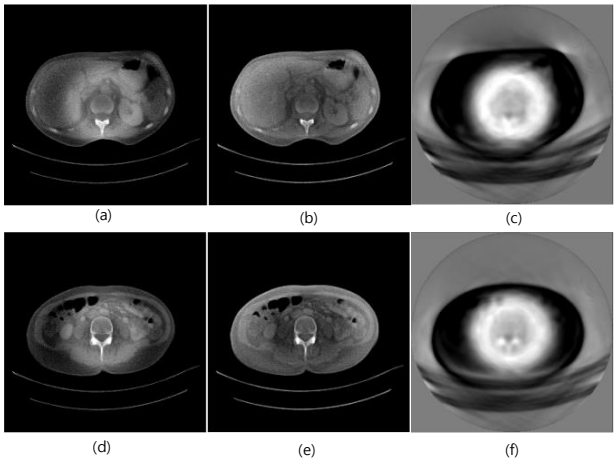


Fig. 5. (a) is original image that corrupted by bowtie artifacts, (b) reconstructed output image of projection domain network, (c) subtracted image between (a) and (b), second row show another body part results with same sequence

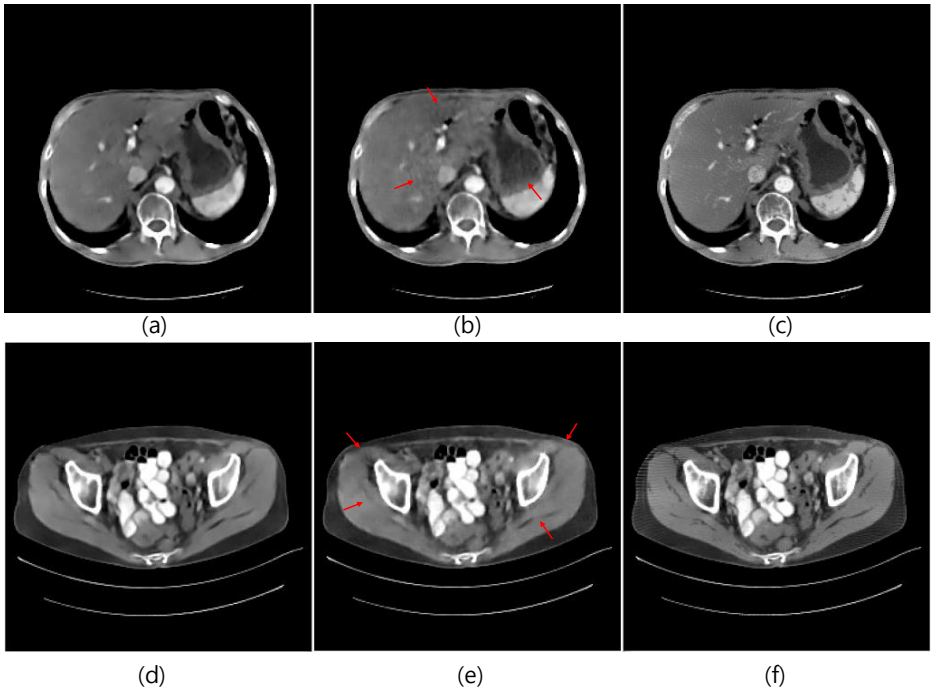


Fig. 6 (a) is output of proposed network, (b) output of single domain U-net, (c) ground-truth image and second row show another body part with same sequence. The window level is [0.017 0.03]

In the single image-domain network case, the training was performed to 900 epoch and a slight overfitting was detected around 500th epoch. Thus, the weight at 470th epoch was used for comparison. Meanwhile, 1e-4 learning rate was used and the same network model and optimizer used as the proposed network. The training loss was 2.05e-7, validation loss was 2.10e-7 and test loss was 2.12e-7. The compared test results are shown in Fig 6.

In Fig 6, the proposed network clearly removes the remaining cupping artifacts. Furthermore, it shows enhanced soft-tissue contrast and denoised outputs. It is observed that the output images are smoother than the label image. We would like to note that this is due to the label image nature where scattering is neglected. In the Monte Carlo generated data, scatter would naturally come into play and the resulting image would not be as sharp as the label image. The compared network outputs also have enhanced contrast and denoised properties, however, some residual artifacts are still observed where the eclipse artifacts originally existed. Moreover, structure distortion was also observed in the results of the compared network (It is highlighted with red arrow in Fig. 6.)

For quantitative analysis, mean square error (MSE) loss, root mean square error (RMSE), and similarity structure index (SSIM) value were evaluated and presented in Table 2. The proposed dual-domain network outperforms a single domain network that does not physically consider the cause of artifacts.

TABLE II
QUANTITATIVE ANALYSIS

MSE(loss)	Training	Validation	Test
Proposed network	1.58e-7	1.76e-7	1.78e-7
Single domain U-net	2.05e-7	2.10e-7	2.12e-7

RMSE	Training	Validation	Test
Proposed network	3.97e-4	4.19e-4	4.21e-4
Single domain U-net	4.52e-7	4.57e-4	4.60e-4

SSIM	Training	Validation	Test
Proposed network	0.963	0.949	0.967
Single domain U-net	0.958	0.946	0.961

IV. CONCLUSION

In this study, we proposed a dual-domain network to reduce bowtie-filter induced image artifacts in cone-beam CT. The promises of the dual-domain network has been successfully shown. Experimental validation of the proposed method is under our research and will be presented in the near future.

REFERENCES

[1] H. Zhang, V. Kong, K. Huang, and J. Y. Jin, "Correction of Bowtie-Filter Normalization and Crescent Artifacts for a Clinical CBCT System," *Technol. Cancer Res. Treat.*, vol. 16, no. 1, 2017.

[2] Y. Cao, T. Ma, S. F. de Boer, and I. Z. Wang, "Image artifacts caused by incorrect bowtie filters in cone-beam CT image-guided radiotherapy," *J. Appl. Clin. Med. Phys.*, vol. 21, no. 7, 2020.

[3] M. Cai, M. Byrne, B. Archibald-Heeren, P. Metcalfe, A. Rosenfeld, and Y. Wang, "Decoupling of bowtie and object effects for beam hardening and scatter artefact reduction in iterative cone-beam CT," *Phys. Eng. Sci. Med.*, vol. 43, no. 4, 2020.

[4] W. A. Lin, H. Liao, C. Peng, X. Sun, J. Zhang, J. Luo, R. Chellappa, and S. K. Zhou, "DuDoNet: Dual domain network for CT metal artifact reduction," in *Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition*, 2019, vol. 2019-June.

[5] J. Bert, H. Perez-Ponce, Z. El Bitar, S. Jan, Y. Boursier, D. Vintache, A. Bonissent, C. Morel, D. Brasse, and D. Visvikis, "Geant4-based Monte Carlo simulations on GPU for medical applications," *Phys. Med. Biol.*, vol. 58, no. 16, 2013.

[6] S. M. Lee, T. Bayaraa, H. Jeong, C. M. Hyun, and J. K. Seo, "A direct sinogram correction method to reduce metal-related beam-hardening in computed tomography," *IEEE Access*, vol. 7, 2019.

Organ-Specific vs. Patient Risk-Specific Tube Current Modulation in Thorax CT Scans Covering the Female Breast

Laura Klein, Lucia Enzmann, Achim Byl, Chang Liu, Stefan Sawall, Andreas Maier, Joscha Maier, Michael Lell, and Marc Kachelrieß.

***Abstract*—An important goal in modern CT imaging is reducing the dose delivered to patients especially to risk-relevant organs. This work compares the clinically applied dose reduction techniques mAs-minimizing tube current modulation (mAsTCM) and a typical organ-specific TCM (osTCM, here: X-Care, Siemens Healthcare) with a novel radiation risk-minimizing tube current modulation (riskTCM) with a focus on the dose delivered to the female breast. The mAsTCM minimizes the mAs product as a surrogate parameter for the patient dose but does not consider the different organs' risks. In contrast, osTCM aims to minimize the dose delivered to the female breast by reducing the tube current for anterior projections. The riskTCM minimizes the patient risk by minimizing the effective dose to the patient, which is done by taking accounting for the organ doses. In this study, the dose reduction effect of the TCM techniques is compared by simulations based on clinical CT Scans.**

Results: riskTCM reduces the effective dose by up to 35% in comparison to mAsTCM, and by up to 30% in comparison to osTCM depending on the anatomical region and the patient.

***Index Terms*—CT, dose reduction, tube current modulation, chest imaging.**

I. Introduction

SINCE the invention of the first CT scanner, the scanners improved greatly and the amount of performed scans increased from 3 million in the early 1980s to 67 million in 2006 [1]. As ionizing radiation is used for CT scans and it is well known that ionizing radiation involves the risk of damaging the DNA, which can cause cancer, it is critical to reduce the patient's risk by decreasing the dose delivered to the patient but maintaining image quality on the other hand. In modern CT imaging, there are several dose-saving methods, for instance, automatic exposure control (AEC) [2], automated tube potential selection [3], and adaptive dose shields [4]. This study is focused on the tube current modulation part of AEC that varies the tube current during gantry rotation and along the z-direction [5].

A commonly used tube current modulation (TCM) technique is the mAs-minimizing TCM (mAsTCM) that minimizes the mAs product and reaches an mAs reduction of $20-40\%$ for the scanned body depending on the region that is investigated [6][7]. The mAs is used as a surrogate value for the patient's dose in this method. An issue with the mAsTCM is that it is based on a physical quantity and does not consider the radiation sensitivity of different organs. The influence of ionizing radiation on the human body, and therefore the risk to induce cancer, is well known. The international commission on radiological protection (ICRP) provides a guideline of protection for people against the effects of radiation exposure and created weighting factors that represent the risk of radiation to induce cancer for single organs[8]. These weighting factors allow for calculating the effective dose that can be chosen as a stochastic parameter of the patient's risk [9]. Since 2007, the radiation sensitivity of the female breast is stated in the ICRP Publication of 2007 [8] to be higher than expected. Therefore, several vendors implemented organ-specific tube current modulation techniques to allow to reduce the exposure at certain organs. Here, we are interested in minimizing the dose to the breast and thus osTCM here refers to a reduction of the tube current in the anterior position. Our implementation of osTCM mimics the X-Care algorithm by Siemens Healthcare. It reduces the tube current for the anterior projections (120° in front of the patient) and increases it for the remaining projections, as specified in reference [10]. Today's osTCM implementations only account for the exposure to a specific organ and do not consider every organs' risks. The radiation-risk minimizing TCM (riskTCM), in contrast, that is detailed in reference [11] and that was proposed in references [12][13][14], minimizes the effective dose by considering all dose-sensitive organs in its cost function.

This work aims at evaluating the dose reduction achievable with riskTCM in comparison to mAsTCM and osTCM. A special focus is laid on the evaluation of the dose delivered to the female breast as this is the organ accounted for by osTCM in thorax CT scans.

II. Material and Methods

A. Dose per view estimations

To calculate the tube current modulation curves and the resulting changed effective doses, the effective dose before TCM needs to be estimated for each view and organ. For this a previous CT reconstruction is necessary. Since this study is a simulation study, CT reconstructions are already provided. In practice, an approach to receive an estimated CT scan of

L. Klein, L. Enzmann, and A. Byl are with X-Ray Imaging and CT, German Cancer Research Center, Heidelberg, Germany, and with the Department of Physics and Astronomy, Heidelberg University, Heidelberg, Germany.

C. Liu and A. Maier are with the Pattern Recognition Lab, Friedrich-Alexander-University Erlangen-Nürnberg, Erlangen, Germany.

S. Sawall, and M. Kachelrieß are with are with X-Ray Imaging and CT, German Cancer Research Center, Heidelberg, Germany, and with the Medical Faculty, Heidelberg University, Heidelberg, Germany.

M. Lell is with the Klinikum Nürnberg, Paracelsus Medical University, Nürnberg, Germany.

Corresponding author: Marc Kachelrieß (marc.kachelriess@dkfz.de)

a simulation study, CT reconstructions are already provided. In practice, an approach to receive an estimated CT scan of a patient is to use a deep-learning model which performs a reconstruction using the available topograms [15]. The dose is calculated from the existing CT reconstructions by using the deep dose estimation (DDE) algorithm that is detailed in reference [16] and was first proposed in reference [17]. DDE reproduces 3D Monte Carlo dose simulations with a two-channel input from a CT reconstruction and a first-order dose estimation. With the calculated dose distribution the effective dose can be calculated by weighting the dose with the organ-specific factor that is defined by the ICRP [8] and sum the weighted dose over all organs:

$$D_{\text{eff}}(\alpha) = \sum_T \int d^3r\, w_T(r) D(\alpha, r). \tag{1}$$

The normalized organ-specific weighting factors w_{T} are listed in Table I.

TABLE I: Tissue weighting factors as in [8].

Tissue	w_T	$\sum w_T$
Bone-marrow (red), colon, lung, stomach, breast, remainder tissues	0.12	0.72
Gonads	0.08	0.08
Bladder, oesophagus, liver, thyroid gland	0.04	0.16
Bone surface, brain, salivary glands, skin	0.01	0.04
	Total	1.00

Equation (1) is the effective dose normalized to a constant tube current so this effective dose needs to be weighted by the tube current curve and integrated over all views:

$$D_{\text{eff}} = \int d\alpha\, I(\alpha) D_{\mathit{eff}}(\alpha). \tag{2}$$

B. Prerequisites

For the simulation study, polychromatic attenuation values are assumed so this yields a projection value of

$$q(L) = -\ln \int dE\, w(E) e^{-\int dL\, \mu(r,E)} \tag{3}$$

$$= -\ln \int dE\, w(E) e^{-p(L)\psi(E)}. \tag{4}$$

This dependency can be written as $q = Q(p)$ and can be inverted by $p = P(q)$, with $P(q)$ as water precorrection function. In order to calculate the noise of an CT Image, a Poisson distribute signal before water precorrection and log is considered. Therefore the variance and thus also noise of the signal is proportional to

$$I e^{-q}. \tag{5}$$

It results:

$$\text{Var}\; q \propto \frac{e^q}{I}. \tag{6}$$

The variance of the projection value can now be calculated propagating the error through the water precorrection function as

$$\text{Var}\; p \propto \frac{e(p)}{I} \tag{7}$$

with $e(p)$ as polychromatic exponential function which reduces to a simple exponential function for monochromatic scans.

In order to simulate the TCMs, two surrogates are needed. First, the average of the projection value over all detector rows is calculated:

$$p(\alpha, \beta) = \frac{1}{B} \int_{-B/2}^{B/2} db\, p(\alpha, \beta, b) \tag{8}$$

In this case, $B = 64 \times 0.6$ mm which is a collimation of about 40 mm.

The second surrogate is the 90th percentile of $p(\alpha, \beta)$:

$$p(\alpha) = p_{90\%}(\alpha, \beta). \tag{9}$$

C. TCM Approaches

1) Minimizing the Tube Current Time Product: The mAs-minimizing TCM takes into account that for non-circular regions, for instance, pelvis or shoulders, the attenuation varies for lateral and anterior/posterior views [6][7]. To calculate the optimal tube current for minimizing the mAs product a central ray approximation is considered that uses the projection data $p(\alpha)$ and interprets this value as central ray value ($\beta = 0$) which is backprojected into the isocenter of the scanner. The mAs minimizing can be formulated as cost function

$$C = \int d\alpha \left(\frac{e(p(\alpha))}{I(\alpha)} + \lambda(I(\alpha) - \text{const}) \right), \tag{10}$$

which allows to either keep the noise constant and minimizes the mAs product or keep the mAs product constant and minimizes noise.

2) organ–specific TCM: Here, osTCM mimics the X-Care algorithm [10]. It is an organ-specific tube current modulation so that especially the female breast could be prevented from high exposure. The algorithm also reduces the dose exposed to the thyroid gland and eyes but here we focus on the breast.

To protect the female breast from radiation exposure the tube current is reduced for anterior projections and for posterior projections it is increased to achieve the same image quality. The anterior projections are defined as the projections within an angle of 120° in front of the patient and the posterior within an angle of 240° on the back of the patient. It should be noted that osTCM is applied for the complete scan even if it covers more regions than just the breast.

3) Risk-minimizing TCM averaging case: Equation (10) can be modified by considering that the effective dose is also dependent on the projection angle α:

$$C = \int d\alpha (\text{Var}\; p(\alpha) + \lambda(I(\alpha) D_{\text{eff}}(\alpha))). \tag{11}$$

This function could also be minimized by differentiating and yields

$$I^2_{\text{riskTCMavg}}(\alpha) \propto \frac{e(p(\alpha))}{D_{\text{eff}}(\alpha)}. \tag{12}$$

This TCM is called riskTCMavg because it minimizes the effective dose but performs an averaging of the projection values for complementary ray directions.

4) Risk-minimizing TCM optimal case: In the optimal risk-minimizing TCM the difference of the effective dose of two complementary rays is considered. Consequently, the variance can be received by statistical optimal weighting for example the inverse-variance weighting:

$$V_{\text{opt}}(r) \propto \frac{1}{I + I_{\text{c}}} \tag{13}$$

with I_{c} as complementary tube current. This distinction of complementary rays leads to major changes in the cost function because it is no longer possible to use the central ray approximation. An assumption that can be done in this case is neglecting the cone-beam nature of the beam and just assuming a fan beam geometry because the cone angle is much smaller than the fan beam angle. The image noise can be described by summation over the square-root of all variances:

$$N_{\text{opt}}(I) = \int dxdy\, w(r) \sqrt{\int d\vartheta \frac{e(p(\alpha(\vartheta, r), \beta(\vartheta, r)))}{I(\alpha(\vartheta, r)) + I(\alpha_{\text{c}}(\vartheta, r))}}. \tag{14}$$

The optimization problem can now be formulated as

$$I = \arg\min\, N_{\text{opt}}(I) \tag{15}$$

with

$$\int d\alpha\, D_{\text{eff}}(\alpha) I(\alpha) = const. \tag{16}$$

The constant can be chosen so either the effective dose is constant or the image noise. As there is no analytical solution to this optimization problem, the solution has to be found numerically. Details are given in reference [11].

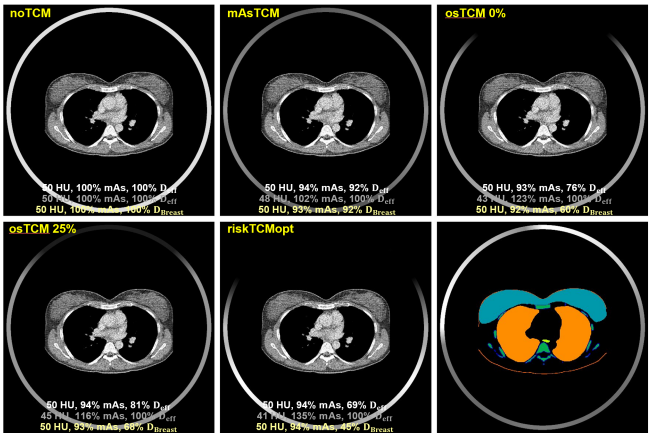
D. Materials

For the evaluation the reconstructed volumes of seven CT scans are used. For the simulation of the rawdata the geometry of a Somatom Definition Flash CT scanner with a collimation of $B = 64 \times 0.6$ mm is assumed. These CT images are forward projected with a 2D fan-beam forward projection to obtain rawdata. To reduce the noise in these rawdata a boxcar filter of 15 mm width was applied to the forward projected rawdata. Now, the rawdata can be regarded as being (almost) noise-free and we can add noise corresponding to the desired TCM curves. These noisy rawdata are then reconstructed by FBP. The tube current curves were scaled in a way to either obtain the same image noise for all TCM algorithms, or to obtain the same D_{eff} for all TCM approaches. This allows for an easy comparison of either the resulting effective dose values or the resulting image noise values.

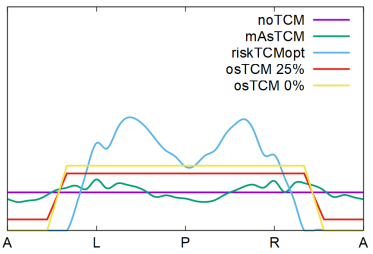
In the following, four tube current modulations are simulated: mAsTCM, riskTCMopt, osTCM with a low value of 25% and osTCM with a low value of 0%. The resulting effective dose values are compared to a scan with constant tube current which is called noTCM.

III. Results and Discussion

The TCMs can be evaluated for different anatomical regions, e.g. thorax, abdomen and pelvis. In Figure 1, the region of the breast is shown with an image noise of 50 HU.



(a)



(b)

Fig. 1: a: noTCM, mAsTCM, osTCM 0%, osTCM 25% and riskTCMopt images with $C = 25$ HU and $W = 400$ HU with an additional segmentation image. The circular density plot is the TCM curve for the first five images and it is the effective dose $D_{\text{eff}}(\alpha)$ for the last image. b: TCM curves $I(\alpha)$ as a function of angular position. The letters under the abscissa indicate the anterior, left, posterior and right tube position.

The mAsTCM reduces the effective dose by 8%, riskTCMopt by 31%, osTCM (0%) by 24%, and osTCM (25%) by 19%. This region is circular shaped so mAsTCM results in a mainly constant tube current. osTCM just accounts for the radiation sensitivity of the breast and therefore reduces the tube current for anterior projections and increases it for the remaining projections. RiskTCMopt reduces the tube current for anterior projections more than for posterior projections because of the risk of the breast. Furthermore, the organ dose to the breast is calculated (highlighted in yellow). For mAsTCM the dose reduction of the breast is 8%, for riskTCMopt 55%, for osTCM (0%) 40%, and for osTCM (25%) 32%. The dose reduction for riskTCMopt is higher than for osTCM because RiskTCMopt minimizes a cost function with detailed anatomical knowledge and on the other hand osTCM assumes the anterior region of the patient independently of the patient's anatomy as 120° in front of the patient.

The results for all regions (Thorax, Abdomen, Pelvis) and patients are shown in table II. For the evaluation only osTCM with 25% as a low tube current value is shown since this case is the conventional one. The whole body effective doses are calculated supplementary and listed in table III.

In the final analysis, also the dose of the breast is calculated and listed in table IV. In comparison to noTCM reduces

mAsTCM the effective dose around 2% to 20%, riskTCMopt around 35% to 65%, and osTCM (25%) around 12% to 39%.

TABLE II: Effective doses for different anatomical regions and patients of different tube current modulations. The effective doses are calculated relative to the effective dose with constant tube current. The image quality is kept constant.

Region	Pat.	noTCM	mAsTCM	osTCM (25%)	riskTCMopt
Thorax	1	100%	76%	70%	51%
	2	100%	89%	86%	71%
	3	100%	86%	86%	70%
	4	100%	87%	79%	65%
	5	100%	89%	88%	71%
	6	100%	90%	84%	73%
	7	100%	87%	79%	66%
	Avg	100%	$(86 \pm 5)\%$	$(82 \pm 6)\%$	$(67 \pm 7)\%$
Abd	1	100%	79%	78%	62%
	2	100%	98%	98%	88%
	3	100%	93%	91%	78%
	4	100%	87%	84%	65%
	5	100%	97%	90%	73%
	6	100%	98%	94%	63%
	7	100%	91%	84%	63%
	Avg	100%	$(92 \pm 7)\%$	$(88 \pm 6)\%$	$(70 \pm 10)\%$
Pelvis	1	100%	75%	88%	70%
	4	100%	74%	77%	60%
	7	100%	72%	82%	64%
	Avg	100%	$(74 \pm 2)\%$	$(82 \pm 6)\%$	$(65 \pm 5)\%$

TABLE III: Whole body effective doses with TCMs.

Pat.	noTCM	mAsTCM	osTCM (25%)	riskTCMopt
1	100%	78%	78%	62%
2	100%	91%	89%	74%
3	100%	88%	88%	74%
4	100%	85%	81%	64%
5	100%	92%	91%	77%
6	100%	94%	91%	77%
7	100%	87%	83%	66%
Avg	100%	$(88 \pm 5)\%$	$(86 \pm 5)\%$	$(71 \pm 6)\%$

TABLE IV: Organ doses for the breast tissue with TCMs.

Pat.	noTCM	mAsTCM	osTCM (25%)	riskTCMopt
1	100%	80%	61%	35%
2	100%	96%	73%	49%
3	100%	93%	69%	37%
4	100%	92%	69%	43%
5	100%	98%	88%	65%
6	100%	93%	71%	47%
7	100%	96%	71%	40%
Avg	100%	$(93 \pm 6)\%$	$(72 \pm 8)\%$	$(45 \pm 10)\%$

IV. Conclusion

The risk-minimizing TCM that is evaluated in this work reduces the risk to a patient by minimizing the effective dose numerically while maintaining the image quality. The resulting effective dose of riskTCM is reduced up to 35% in comparison to mAsTCM and up to 30% in comparison to osTCM depending on the anatomical region and the patient. Even by comparing the effect of riskTCM on the effective dose of the breast with the effect of osTCM, which is a dose-saving algorithm especially for the breast, riskTCM reduces the effective dose up to 30% more than osTCM because riskTCM accounts for the patient's anatomy.

Acknowledgment

Parts of the reconstruction and simulation software were provided by RayConStruct® GmbH, Nürnberg, Germany. The CT data were provided by Prof. Dr. Michael Lell, Klinikum Nürnberg, Germany. This work was supported by the Deutsche Forschungsgemeinschaft (DFG) under grant KA 1678/24, LE 2763/3 and MA 4898/15.

References

[1] D. A. Schauer and O. W. Linton, "National council on radiation protection and measurements report shows substantial medical exposure increase," *Radiology*, vol. 253, no. 2, pp. 293–296, 2009.

[2] M. Söderberg and M. Gunnarsson, "Automatic exposure control in computed tomography – an evaluation of systems from different manufacturers," *Acta Radiologica*, vol. 51, no. 6, pp. 625–634, 2010.

[3] A. Winklehner, R. Goetti, S. Baumueller, C. Karlo, B. Schmidt, R. Raupach, T. Flohr, T. Frauenfelder, and H. Alkadhi, "Automated attenuation-based tube potential selection for thoracoabdominal computed tomography angiography: improved dose effectiveness," *Invest Radiol*, vol. 46, no. 12, pp. 767–73, 2011.

[4] P. D. Deak, O. Langner, M. Lell, and W. A. Kalender, "Effects of adaptive section collimation on patient radiation dose in multisection spiral CT," *Radiology*, vol. 252, no. 1, pp. 140–147, 2009.

[5] J. N. Althén, "Automatic tube-current modulation in CT—a comparison between different solutions," *Radiation Protection Dosimetry*, vol. 114, no. 1-3, pp. 308–312, 2005.

[6] W. A. Kalender, H. Wolf, C. Suess, M. Gies, H. Greess, and W. A. Bautz, "Dose reduction in CT by on-line tube current control: principles and validation on phantoms and cadavers," *European Radiology*, vol. 9, no. 2, pp. 323–328, 1999.

[7] M. Gies, W. A. Kalender, H. Wolf, C. Suess, and M. T. Madsen, "Dose reduction in CT by anatomically adapted tube current modulation. i. simulation studies," *Medical Physics*, vol. 26, no. 11, pp. 2235–2247, 1999.

[8] ICRP, "The 2007 recommendations of the international commission on radiological protection," *ICRP Publication 103*, vol. 36, no. 2-4, 2007.

[9] W. Huda, K. M. Ogden, and M. R. Khorasani, "Effect of dose metrics and radiation risk models when optimizing CT x-ray tube voltage," *Physics in Medicine and Biology*, vol. 53, no. 17, pp. 4719–4732, 2008.

[10] D. Ketelsen, M. Buchgeister, M. Fenchel, B. Schmidt, T. G. Flohr, R. Syha, C. Thomas, I. Tsiflikas, C. D. Claussen, and M. Heuschmid, "Automated computed tomography dose-saving algorithm to protect radiosensitive tissues: estimation of radiation exposure and image quality considerations," *Invest Radiol*, vol. 47, no. 2, pp. 148–52, 2012.

[11] L. Klein, C. Liu, J. Steidel, L. Enzmann, S. Sawall, A. Maier, M. Lell, J. Maier, and M. Kachelrieß, "Patient-specific radiation risk-based tube current modulation for diagnostic CT," *Medical Physics*, 2022, in press.

[12] J. Maier, L. Klein, S. Sawall, C. Liu, A. Maier, and M. Kachelrieß, "Radiation risk minimizing tube current modulation (rmTCM) for x-ray computed tomography," ECR 2021 Book of Abstracts, Insights into Imaging 12(Suppl. 1):104-105, RPS 213-3, June 2021.

[13] L. Klein, C. Liu, J. Steidel, S. Sawall, A. Maier, M. Lell, J. Maier, and M. Kachelrieß, "A novel CT tube current modulation technique that minimizes patient risk," program of the 107th Scientific Assembly and Annual Meeting of the RSNA: PH03-A5, November 2021.

[14] L. Klein, C. Liu, J. Steidel, L. Enzmann, S. Sawall, J. Maier, A. Maier, M. Lell, and M. Kachelrieß, "Risk-minimizing tube current modulation (riskTCM) for CT – potential dose reduction across different tube voltages," ECR 2022.

[15] A.-A.-Z. Imran, S. Wang, D. Pal, S. Dutta, B. Patel, E. Zucker, and A. Wang, "Personalized CT organ dose estimation from scout images," ser. Medical Image Computing and Computer Assisted Intervention – MICCAI 2021. Springer International Publishing, Conference Proceedings, pp. 488–498.

[16] J. Maier, L. Klein, E. Eulig, S. Sawall, and M. Kachelrieß, "Real-time estimation of patient-specific dose distributions for medical CT using the deep dose estimation," *Medical Physics*, 2022, in press.

[17] J. Maier, E. Eulig, S. Dorn, S. Sawall, and M. Kachelrieß, "Real-time patient-specific CT dose estimation using a deep convolutional neural network," in *2018 IEEE Nuclear Science Symposium and Medical Imaging Conference Proceedings (NSS/MIC)*, Conference Proceedings, pp. 1–3.

An Analytical Prj2CH Covariance Estimation Method for Iterative Reconstruction Methods

Xiaoyue Guo, Li Zhang, Yuxiang Xing

***Abstract*—Image mean and covariance required for a model observer are usually calculated by statistical method using image samples that are hard to acquire in reality. Although some analytical methods are proposed to estimate image covariance from a single projection, these methods are of high computational cost for large-dimensional images (e.g., 512×512), and images of large size are commonly used. Considering the covariance used for CHO calculation is the covariance of the channel response, whose dimension is much smaller than the image covariance, we aim to obtain the small-dimensional channel covariance directly from projection. The analytical projection to channel (Prj2CH) covariance estimation method is derived from the analytical projection to image (Prj2Img) covariance estimation method by reducing the dimension of matrices in Prj2Img method. We validate the feasibility and utility of the proposed Prj2CH method by simulations. For 128×128 reconstruction, the SNR error of Prj2CH method is about 1%, which is comparable to the SNR variation error of statistical method, demonstrating that the channel covariance estimated by the Prj2CH method is reasonable. For 512×512 reconstruction, the SNR of the Prj2CH method is comparable to that of the statistical method with 350 samples, which shows that the proposed method can be applied for CHO calculation. Meanwhile, the Prj2CH method can be used for different systems, objects, and reconstruction algorithms.**

***Index Terms*—Covariance Estimation, Iterative Reconstruction, Model Observer, Dimension Reduction**

I. Introduction

Model observers are used to mimic human observers despite that they require the knowledge of image mean and covariance, which is difficult to achieve in reality. Analytical projection to image (Prj2Img) covariance estimation methods are proposed to estimate image covariance from a single projection for commonly used iterative reconstruction algorithms. Iteration-based and fixed-point methods are two ways to analytically estimate covariance from projection to image. For iteration-based methods, the covariance estimation is updated with iteration formula [1]. For fixed-point methods, the covariance estimation is derived from the converged point of an objective function [2]. Li et.al. [3] study the difference and consistence of these two methods. Meanwhile, analytical Prj2Img methods for iterative reconstruction with a quadratic regularization is studied by [4], while that with a non-quadratic regularization is studied by [5]. Although analytical Prj2Img methods can yield reasonable covariance estimations, it is computationally expensive for large-dimensional images (pixels > 128×128). Usually, images with larger size, e.g., 512×512, are more preferred. Fessler et.al. [6] give a fast variance estimation method for the quadratic penalized weighted least square (WLS) algorithm. Fast covariance estimation methods have not been studied yet.

In fact, the covariance used for channelized Hotelling observers (CHOs) is the covariance of the channel response [7] that has a much smaller dimension than its image covariance matrix. Typically, we can use an analytical projection to image to channel (Prj2Img2CH) covariance estimation method for CHO calculation, which estimates image covariance from a single projection by the Prj2Img method first, and then channel covariance from the image covariance. However, the high computation cost of Prj2Img method still exists in the Prj2Img2CH method. Therefore, we propose an analytical projection to channel (Prj2CH) covariance estimation method for non-quadratic penalized WLS algorithms in this work. Rather than calculating the covariance of the channel response from the image covariance, we directly estimate the channel covariance from a single projection by one step. The relationship between the channel and projection covariance is derived from the fixed-point method for analytical Prj2Img covariance estimation. Moreover, we apply the proposed method to a CHO of large-dimensional images evaluation for validation.

II. Methods

In this section, we briefly introduce the analytical Prj2Img covariance estimation method for non-quadratic penalized WLS studied in our previous work [8], then describe the CHO, and finally deduce the proposed analytical Prj2CH covariance estimation method in detail.

A. Analytical Prj2Img Covariance Estimation Method

The cost function $\Phi(\cdot)$ of a penalized WLS reconstruction can be expressed as:

$$\hat{\boldsymbol{\mu}} = \arg\min_{\boldsymbol{\mu}} \Phi(\boldsymbol{\mu}) = \arg\min_{\boldsymbol{\mu}} \frac{1}{2}\left\|\mathbf{H}\boldsymbol{\mu} - \mathbf{p}\right\|_{\mathbf{W}}^{2} + \alpha R(\boldsymbol{\mu}) \qquad (1)$$

This work is supported by National Natural Science Foundation of China (Grant No. 62031020 and 61771279)

Xiaoyue Guo is with the Department of Engineering Physics, Tsinghua University, Beijing, 100084, China (e-mail: guoxy15@mails.tsinghua.edu.cn).

Li Zhang is with the Department of Engineering Physics, Tsinghua University, Beijing, 100084, China (e-mail: xingyx@mail.tsinghua.edu.cn).

Yuxiang Xing is with the Department of Engineering Physics, Tsinghua University, Beijing, 100084, China (e-mail: zli@mail.tsinghua.edu.cn).

where $\hat{\boldsymbol{\mu}} \in \mathbb{R}^{N\times 1}$ is the linear attenuation image reconstructed from its projection $\mathbf{p} \in \mathbb{R}^{M\times 1}$. $\mathbf{H} \in \mathbb{R}^{M\times N}$ denotes the system matrix and $\mathbf{W} \in \mathbb{R}^{M\times M}$ the noise model with $\mathbf{W}_{mm} = \exp(-p_m)$. $R(\cdot)$ is the penalty function and α the penalty parameter. The fixed-point method makes use of convergence condition:

$$\left.\frac{\partial \Phi(\boldsymbol{\mu})}{\partial \boldsymbol{\mu}}\right|_{\boldsymbol{\mu}=\hat{\boldsymbol{\mu}}} = \mathbf{0} \tag{2}$$

Plugging Eq. (1) into Eq. (2) and finding its covariance, we have:

$$\begin{aligned}&(\mathbf{H}^T\mathbf{W}\mathbf{H}+\alpha\mathbf{A})\mathrm{Cov}(\hat{\boldsymbol{\mu}})(\mathbf{H}^T\mathbf{W}\mathbf{H}+\alpha\mathbf{A}^T)\\ &\simeq \mathbf{H}^T\mathbf{W}\mathrm{Cov}(\mathbf{p})\mathbf{W}\mathbf{H}\end{aligned} \tag{3}$$

where $\mathbf{A} \in \mathbb{R}^{N\times N}$ is a coefficient matrix approximates $\nabla R(\hat{\boldsymbol{\mu}})$:

$$\nabla R(\hat{\boldsymbol{\mu}}) \simeq \mathbf{A}\hat{\boldsymbol{\mu}} + \mathbf{c} \tag{4}$$

Taylor expansion is a commonly used approximation method with $\nabla R(\hat{\boldsymbol{\mu}}) \simeq \nabla^2 R(\bar{\hat{\boldsymbol{\mu}}})\hat{\boldsymbol{\mu}} + \mathbf{c}$, thus the Taylor approximation based method (TAM) for covariance estimation is:

$$\begin{aligned}&\left(\mathbf{H}^T\mathbf{W}\mathbf{H}+\alpha\nabla^2 R(\bar{\hat{\boldsymbol{\mu}}})\right)\mathrm{Cov}(\hat{\boldsymbol{\mu}})\left(\mathbf{H}^T\mathbf{W}\mathbf{H}+\alpha\nabla^2 R(\bar{\hat{\boldsymbol{\mu}}})\right)\\ &\simeq \mathbf{H}^T\mathbf{W}\mathrm{Cov}(\mathbf{p})\mathbf{W}\mathbf{H}\end{aligned} \tag{5}$$

Besides, considering that the second derivative of $R(\cdot)$ may be singular or nonexistent, we proposed a linear approximation based method (LAM) in our previous work with $\nabla R(\hat{\boldsymbol{\mu}}) = \mathbf{L}(\hat{\boldsymbol{\mu}})\hat{\boldsymbol{\mu}}$ that estimates the covariance as:

$$\begin{aligned}&\left(\mathbf{H}^T\mathbf{W}\mathbf{H}+\alpha\mathbf{L}(\bar{\hat{\boldsymbol{\mu}}})\right)\mathrm{Cov}(\hat{\boldsymbol{\mu}})\left(\mathbf{H}^T\mathbf{W}\mathbf{H}+\alpha\mathbf{L}(\bar{\hat{\boldsymbol{\mu}}})\right)\\ &\simeq \mathbf{H}^T\mathbf{W}\mathrm{Cov}(\mathbf{p})\mathbf{W}\mathbf{H}\end{aligned} \tag{6}$$

We adopt the WLS penalized with the total variance (TV-WLS) as well as the qth generalized Gaussian Markov random field (qGGMRF-WLS) as two representatives in this work.

B. Channelized Hotelling Observer (CHO)

A CHO is calculated as:

$$\lambda = [\mathbf{S}_c^{-1}(\bar{\hat{\boldsymbol{\mu}}}_{sc} - \bar{\hat{\boldsymbol{\mu}}}_{bc})]^T\hat{\boldsymbol{\mu}}_c \equiv \boldsymbol{\omega}^T\hat{\boldsymbol{\mu}}_c \tag{7}$$

where λ denotes the decision variable and $\boldsymbol{\omega}$ the template of CHO. The image of channel response is denoted by a subscript c:

$$\hat{\boldsymbol{\mu}}_c = \mathbf{V}\hat{\boldsymbol{\mu}} \tag{8}$$

with $\mathbf{V} \in \mathbb{R}^{C\times N}$ being the channel matrix consists of C channel profiles and $C \ll N$. Meanwhile, the $\bar{\hat{\boldsymbol{\mu}}}_{sc}$ and $\bar{\hat{\boldsymbol{\mu}}}_{bc}$ in template $\boldsymbol{\omega}$ represent the mean of signal present and signal absent images respectively, and the intra-class channel scatter matrix $\mathbf{S}_c = (\mathbf{K}_{sc} + \mathbf{K}_{bc})/2$, where the channel covariance $\mathbf{K}_{sc} = \mathbf{V}\mathbf{K}_s\mathbf{V}^T$ and $\mathbf{K}_{bc} = \mathbf{V}\mathbf{K}_b\mathbf{V}^T$ with $\mathbf{K} = \mathrm{cov}(\hat{\boldsymbol{\mu}})$ being the image covariance. For a given threshold t, if the decision variable satisfies $\lambda > t$, we consider the target image to contain the signal; otherwise, we consider it not. Meanwhile, signal-to-noise ratio [9] is adopted to measure the performance of a CHO:

$$\mathrm{SNR} = \sqrt{(\bar{\boldsymbol{\mu}}_{sc} - \bar{\boldsymbol{\mu}}_{bc})^T\mathbf{S}_c^{-1}(\bar{\boldsymbol{\mu}}_{sc} - \bar{\boldsymbol{\mu}}_{bc})} \tag{9}$$

C. Analytical Prj2CH Covariance Estimation Method

Note that the analytical Prj2Img method in Eq. (3) involves a matrix inverse operation which leads to high computational cost when the dimension of the reconstructed image is large. As described in section Ⅱ.B, the CHO only requires a channel covariance of small dimension that $C \ll N$. Therefore, we construct a relationship between the channel covariance $\mathbf{K}_c = \mathbf{V}\mathbf{K}\mathbf{V}^T$ and its projection covariance:

$$\begin{aligned}&\mathbf{V}(\mathbf{H}^T\mathbf{W}\mathbf{H}+\alpha\mathbf{A})\mathrm{Cov}(\hat{\boldsymbol{\mu}})(\mathbf{H}^T\mathbf{W}\mathbf{H}+\alpha\mathbf{A}^T)\mathbf{V}^T\\ &= \mathbf{V}(\mathbf{H}^T\mathbf{W}\mathbf{H}+\alpha\mathbf{A})\mathbf{V}^T\mathbf{X}[\mathbf{V}\mathbf{K}\mathbf{V}^T]\mathbf{X}^T\mathbf{V}(\mathbf{H}^T\mathbf{W}\mathbf{H}+\alpha\mathbf{A}^T)\mathbf{V}^T\\ &\simeq \mathbf{V}\mathbf{H}^T\mathbf{W}\mathrm{Cov}(\mathbf{p})\mathbf{W}\mathbf{H}\mathbf{V}^T\end{aligned} \tag{10}$$

where the transition matrix $\mathbf{X} = \mathbf{V}(\mathbf{V}^T\mathbf{V})^{-1}(\mathbf{V}^T\mathbf{V})^{-1}\mathbf{V}^T$. The channel covariance is then calculated as:

$$\mathbf{K}_c = \mathbf{Y}^{-1}[\mathbf{V}\mathbf{H}^T\mathbf{W}\mathrm{Cov}(\mathbf{p})\mathbf{W}\mathbf{H}\mathbf{V}^T](\mathbf{Y}^T)^{-1} \tag{11}$$

where $\mathbf{Y} = [\mathbf{V}(\mathbf{H}^T\mathbf{W}\mathbf{H}+\alpha\mathbf{A})\mathbf{V}^T]\mathbf{X} \in \mathbb{R}^{C\times C}$. According to different approximations of $\mathbf{A}$, we have Prj2CH-LAM and -TAM methods.

However, matrix $\mathbf{V}^T\mathbf{V}$ is not full rank and hence it is irreversible, and the transition matrix $\mathbf{X} \in \mathbb{R}^{C\times C}$ becomes unknow. Note that $\mathbf{X}$ is theoretically independent of the system, phantom, and reconstruction algorithm, and only depends on the channel. Given a Prj2Img covariance estimation of a small size image, we can calibrate $\mathbf{X}$ by the known $\mathbf{K}_c$. Thus, we build an optimization problem for calibration:

$$\hat{\mathbf{X}} = \arg\min_{\mathbf{X}} \frac{1}{2}\left\|\mathbf{Y}\mathbf{K}_c\mathbf{Y}^T - \mathbf{V}\mathbf{H}^T\mathbf{W}\mathrm{Cov}(\mathbf{p})\mathbf{W}\mathbf{H}\mathbf{V}^T\right\|_2^2 + \frac{\gamma}{2}\|\mathbf{X}\|_2^2 \tag{12}$$

We split Eq. (12) into two optimization problems to make it easier to get a reasonable solution:

$$\hat{\mathbf{Y}} = \arg\min_{\mathbf{Y}} \frac{1}{2}\left\|\mathbf{Y}\mathbf{K}_c\mathbf{Y} - \mathbf{V}\mathbf{H}^T\mathbf{W}\mathrm{Cov}(\mathbf{p})\mathbf{W}\mathbf{H}\mathbf{V}^T\right\|_2^2 \tag{13.1}$$

$$\hat{\mathbf{X}} = \arg\min_{\mathbf{X}} \frac{1}{2}\left\|\hat{\mathbf{Y}} - [\mathbf{V}(\mathbf{H}^T\mathbf{W}\mathbf{H}+\alpha\mathbf{A})\mathbf{V}^T]\mathbf{X}\right\|_2^2 + \frac{\gamma}{2}\|\mathbf{X}\|_2^2 \tag{13.2}$$

Since $\mathbf{X}$ is an underdetermined matrix, the optimization problem in Eq. (13) is locally convergent, and we choose start points as:

$$\begin{aligned}\mathbf{Y}^{(0)} &= [\mathbf{V}(\mathbf{H}^T\mathbf{W}\mathbf{H}+\alpha\mathbf{A})\mathbf{V}^T][\mathrm{pinv}(\mathbf{V}^T)\,\mathrm{pinv}(\mathbf{V})]\\ \mathbf{X}^{(0)} &= \mathbf{I}\end{aligned}$$

with pinv(·) being the Moore-Penrose generalized inverse operation. Plugging the calibrated $\hat{\mathbf{X}}$ into Eq. (11), we can finally obtain the channel covariance estimation $\hat{\mathbf{K}}_c$ under arbitrary conditions.

III. EXPERIMENTS

Simulations are carried out for validation. We select Gabor channels are selected for the CHO:

$$\begin{aligned}v(x,y) = &\exp[-4\ln 2((x-x_0)^2+(y-y_0)^2)/\omega_s^2]\\ &\cdot\cos[2\pi f_c((x-x_0)\cos\theta+(y-y_0)\sin\theta)+\beta]\end{aligned} \tag{14}$$

here the parameter configuration of Gabor channels is similar to

that in [10], where the channel width $\omega_s = 56.48, 28.24, 14.12$, the channel frequency $f_c = 3/128, 3/64, 3/32$, the orientation $\theta = 0, 2\pi/5, 4\pi/5, 6\pi/5, 8\pi/5$, and the phase $\beta = 0, \pi/2$.

The system geometry of 128 × 128 and 512 × 512 reconstructed images is listed in Table I. Phantoms are generated from the Grand challenge dataset of Mayo clinic. Set the number of incident photons $I_0 = 5\times10^3$, we acquire a noisy projection from a random 128×128 region of interest (ROI) extracted from the dataset. We reconstruct its image by qGGMRF-WLS from the noisy projection, and estimate image covariance by TAM. With these known conditions, the calibrated $\hat{\mathbf{X}}$ is obtained by Eq. (13). The hyper-parameter γ in Eq. (13.2) is set to be 10^5 to balance the value between its fidelity and regularization.

TABLE I
SYSTEM GEOMETRY OF SIMULATIONS

System Geometry	128×128	512×512
Source to origin distance (mm)	200	595
Origin to detector distance (mm)	200	490.6
Detector pixel size (mm)	0.5	1.2858
Detector pixel number	240	736
Image pixel size (mm)	0.3	0.7422
Image pixel number	128	512
Rotating angle	2π	2π

Another 128×128 ROI chosen from the dataset is used to validate the calibrated $\hat{\mathbf{X}}$ with $I_0 = 5\times10^4$. Both TV- and qGGMRF-WLS algorithms are used for reconstruction, and Prj2CH-LAM as well as -TAM are used for channel covariance estimation. We adopt an absolute percent error (APE) to quantitatively measure the performance of the proposed Prj2CH method:

$$\mathrm{APE} = \left|\frac{\mathrm{SNR} - \mathrm{SNR}}{\mathrm{SNR}}\right| \times 100\% \qquad (15)$$

here SNR is calculated according to Eq. (9) where its $\mathbf{S}_c$ is estimated by Prj2Img2CH method, i.e., the ground truth. Meanwhile, the $\mathbf{S}_c$ for SNR calculation is estimated by Prj2CH method. Since we focus on the effect of covariance estimation, we take the noise-free image as the image mean in Eq. (9). Besides, the coefficient of variation (CV) of statistical covariance estimation method is used as a reference:

$$\mathrm{CV}_R = \frac{\sigma_{\mathrm{SNR}_R^*}}{u_{\mathrm{SNR}_R^*}} \times 100\% \qquad (16)$$

where u and σ are the mean and standard deviation of SNR^*. The SNR_R^* is calculated by the image covariance $\mathbf{K}^*$ estimated statistically from image samples:

$$\mathbf{K}_R^* = \frac{1}{R}\sum_{r=1}^{R}(\hat{\boldsymbol{\mu}}_r - \bar{\hat{\boldsymbol{\mu}}})(\hat{\boldsymbol{\mu}}_r - \bar{\hat{\boldsymbol{\mu}}})^T$$

with R being the total number of image samples. Once $\mathrm{APE} \le \mathrm{CV}$, we think the proposed Prj2CH method is feasible.

In addition, the same validation procedure as that for the 128 ×128 ROI is performed for a 512×512 phantom selected in the dataset with $I_0 = 10^6$. The SNR in Eq. (9) is also used as the measurement for performance evaluation of the Prj2CH method. Considering there is no ground truth for Prj2CH method with the phantom of 512×512, we use $u_{\mathrm{SNR}_R^*} \pm \sigma_{\mathrm{SNR}_R^*}$ as the reference.

IV. EXPERIMENTAL RESULTS

Both APE and CV from the phantom of 128×128 pixels are listed in Table II. We find that the APE between SNRs of Prj2Img2CH and Prj2CH methods is less than 2%, which at least comparable to the CV of SNRs statistically calculated by 300 image samples, indicating that the error of about 1% for the Prj2CH method is acceptable. The proposed Prj2CH method is feasible for different phantoms and reconstruction algorithms with a same system geometry.

TABLE II
COMPARISON BETWEEN APE OF SNRS OBTAINED BY ANALYTICAL COVARIANCE ESTIMATION METHODS AND CV OF SNRS OBTAINED FROM IMAGE SAMPLES IN CASE OF 128×128

Covariance Estimation Method \ Number of samples		100	200	300	400	1
TV-WLS	Image samples	5.14%	3.14%	1.98%	1.10%	/
	LAM	/	/	/	/	**0.87%**
	TAM	/	/	/	/	**1.71%**
qGGMRF-WLS	Image samples	5.15%	3.24%	1.96%	1.39%	/
	LAM	/	/	/	/	**0.99%**
	TAM	/	/	/	/	**1.09%**

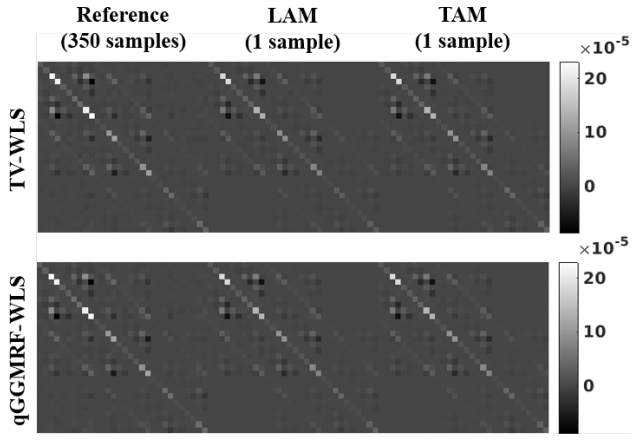
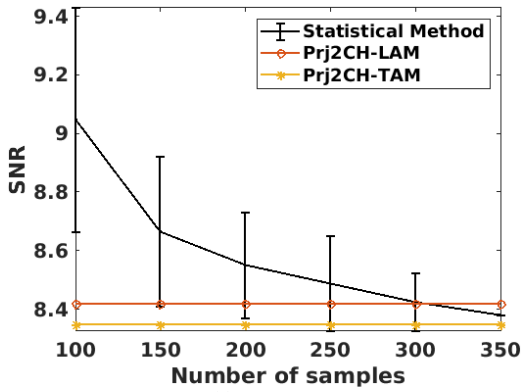
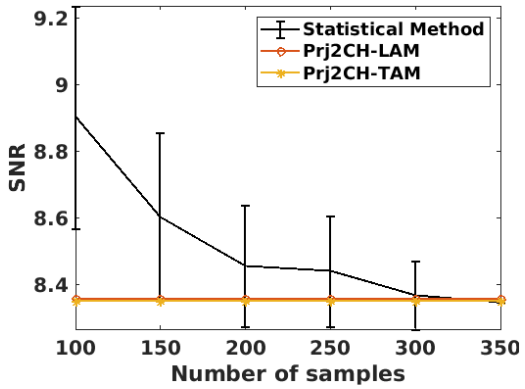


Fig. 1 $\mathbf{S}_c$ from image samples and analytical Prj2CH methods in case of 512 ×512 size phantom. Column 1: $\mathbf{S}_c$ calculated by statistical method from 350 image samples. Column 2 and 3: $\mathbf{S}_c$ estimated by analytical Prj2CH-LAM and -TAM method from 1 projection and its reconstruction samples.

Images of $\mathbf{S}_c$ are shown in Fig. 1. The estimated $\mathbf{S}_c$ by the Prj2CH method has similar shape and values to that by statistical method. It shows that the channel covariance estimation is reasonable for large-dimensional images even with different scanning systems.



(a) TV-WLS



(b) qGGMRF-WLS

Fig. 2 SNR in case of 512×512. (a) the results from TV-WLS algorithms, and (b) the results from qGGRMF-WLS algorithms. The black solid line reprensents $u_{\mathrm{SNR}_R^*}$ with error bar $\sigma_{\mathrm{SNR}_R^*}$, the red solid line with circle markers is the SNR of Prj2CH-LAM from 1 projection, and the orange solid line with asterisk markers is the SNR of Prj2CH-TAM from 1 projection.

As is plotted in Fig. 2, SNRs for Prj2CH method are within the variation range of the SNR from 300 image samples in case of 512×512. The SNR of Prj2CH method is close to that from 350 samples, and hence the CHO obtained by Prj2CH method has better performance than that obtained from commonly used 100 samples.

V. Discussion And Conclusion

In this work, we propose an analytical method to estimate the covariance of channel response in CHO from a single projection. We build this analytical projection to channel (Prj2CH) covariance estimation method based on the fixed-point method that estimates image covariance from projection analytically. The proposed Prj2CH method avoids large-dimensional matrix inversion operation in the projection to image (Prj2Img) covariance estimation method, which enables the channel covariance estimation of large-dimensional reconstructed images, and further contributes to the calculation of CHO. Results show that, for ROI of size 128, the APE of Prj2CH method is about 1%, which is acceptable since the CV of SNR from 400 samples is also about 1%. Therefore, we think the proposed Prj2CH method is feasible for channel covariance estimation. For application to images with 512×512, we find the proposed Prj2CH method is comparable to the statistical method with 350 image samples, and has much better performance than the statistical method with commonly used 100 samples. The proposed Prj2CH method can be applied to different scanning systems, objects, and reconstruction algorithms, and is also reasonable for CHO applications.

For our future work, we are to compare the performance of the CHO from the proposed Prj2CH method with that the human observer.

Acknowledgment

This work is supported by National Natural Science Foundation of China (Grant No. 62031020 and 61771279).

References

[1] J. Qi, “A unified noise analysis for iterative image estimation,” *Phys. Med. Biol.*, vol. 48, no. 21, pp. 3505–3519, 2003, doi: 10.1088/0031-9155/48/21/004.

[2] J. A. Fessler, “Mean and variance of implicitly defined biased estimators (such as penalized maximum likelihood): Applications to tomography,” *Biomed. Imaging V - Proc. 5th IEEE EMBS Int. Summer Sch. Biomed. Imaging, SSBI 2002*, vol. 5, no. 3, pp. 493–506, 2002, doi: 10.1109/SSBI.2002.1233984.

[3] Y. Li, “Noise propagation for iterative penalized-likelihood image reconstruction based on Fisher information,” *Phys. Med. Biol.*, vol. 56, no. 4, pp. 1083–1103, 2011, doi: 10.1088/0031-9155/56/4/013.

[4] S. M. Schmitt, M. M. Goodsitt, and J. A. Fessler, “Fast variance prediction for iteratively reconstructed ct images with locally quadratic regularization,” *IEEE Trans. Med. Imaging*, vol. 36, no. 1, pp. 17–26, 2017, doi: 10.1109/TMI.2016.2593259.

[5] A. A. Sánchez, “Estimation of noise properties for TV-regularized image reconstruction in computed tomography,” *Phys. Med. Biol.*, vol. 60, no. 18, pp. 7007–7033, 2015, doi: 10.1088/0031-9155/60/18/7007.

[6] Y. Zhang-O’Connor and J. A. Fessler, “Fast variance predictions for 3D cone-beam CT with quadratic regularization,” in *Medical Imaging 2007: Physics of Medical Imaging*, 2007, vol. 6510, p. 65105W, doi: 10.1117/12.710312.

[7] X. He and S. Park, “Model observers in medical imaging research,” *Theranostics*, vol. 3, no. 10, pp. 774–786, 2013, doi: 10.7150/thno.5138.

[8] X. Guo, Y. Xing, and L. Zhang, “Analytical covariance estimation for iterative CT reconstruction methods,” in *16th International Meeting on Fully 3D Image Reconstruction in Radiology and Nuclear Medicine*, 2021, pp. 236–241.

[9] A. Wunderlich and F. Noo, “Image covariance and lesion detectability in direct fan-beam x-ray computed tomography,” *Phys. Med. Biol.*, vol. 53, no. 10, pp. 2471–2493, 2008, doi: 10.1088/0031-9155/53/10/002.

[10] S. Leng, L. Yu, Y. Zhang, R. Carter, A. Y. Toledano, and C. H. McCollough, “Correlation between model observer and human observer performance in CT imaging when lesion location is uncertain,” *Med. Phys.*, vol. 40, no. 8, pp. 1–9, 2013, doi: 10.1118/1.4812430.

Material Decomposition from Photon-Counting CT using a Convolutional Neural Network and Energy-Integrating CT Training Labels

Rohan Nadkarni, Alex Allphin, Darin P. Clark, and Cristian T. Badea

***Abstract*—Although photon counting detectors (PCD) can offer numerous benefits for CT imaging, it is difficult to generate accurate material decompositions from photon counting (PC) CT images due to spectral distortions. In this work, we present a deep learning (DL) approach for material decomposition from PCCT. To produce training and testing data for this study, we scanned 2 *ex-vivo* mice using a PCD scan protocol with a dose of 36 mGy and a multi-EID scan protocol with a dose of 296 mGy. PCD images were reconstructed using filtered backprojection. EID images were reconstructed using an iterative algorithm to reduce noise, and decomposed into iodine (I), Compton scattering (CS), and photoelectric effect (PE) material maps by a matrix inversion approach. We then trained a convolutional neural network with a 3D U-net structure using PCD images as inputs and multi-EID material maps as labels, and evaluated its performance. The 3D U-net predictions provided substantially lower RMSE compared to decomposition from PCD images using a matrix inversion approach. Measurements from iodine vials in the test set showed that 3D U-net predictions gave mean values within 0.6 mg/mL of the mean values from the multi-EID material maps and much lower standard deviation than PCD material map measurements. Our results show that the trained 3D U-net enables low-noise, quantitatively accurate material decomposition from a low dose PCD scan.**

***Index Terms*—Material decomposition, photon-counting CT, CNN, deep learning**

I. Introduction

PHOTON counting detectors (PCDs) offer enormous potential to improve CT imaging. A PCD counts the incident photons and bins them using energy thresholds. Each threshold image records only those photons with energy greater than a user-defined threshold. Therefore, spectral CT with a PCD can offer improved dose efficiency and spectral resolution compared to a conventional scan with an energy-integrating detector (EID) [1, 2]. Unfortunately, PCDs also suffer from effects such as K-escape, charge sharing, and pulse pileup that distort their spectral response [1]. These distortions make it difficult to perform accurate material decomposition in PCCT. Therefore, finding methods for improving material decomposition is critical to maximize the benefits of PCCT.

This work presents a method for generating low noise and accurate material decompositions from PCCT images by compensating for distortions with a convolutional neural network (CNN). Several recent works have shown great success in using deep learning (DL) for material decomposition from EID images as well as PCD images [3-6]. However, our approach differs from past work by using material decompositions derived from multi-EID CT scans as training labels. Our EID scan data is a good source for training labels because they are not affected by the spectral distortions described earlier and they have low noise due to high x-ray dose. The CNN is trained using PCCT images as inputs and material maps derived from multi-EID scans of the same specimens as labels. The trained network can then be used to generate accurate material decomposition maps from PCD images without the need to pass corresponding EID scans into the network. Our implementation and results based on micro-CT scans are presented in the following sections.

II. Methods

A. Image Acquisition

All image data were acquired on our dual-source, micro-CT imaging system that has one EID (*Dexela 1512* CMOS x-ray detector with a CsI scintillator, 75 μm pixel size) and one PCD (Dectris *Santis 1604*, 150 μm pixel size, 4 energy thresholds) [7]. For this study, we scanned 2 mice that were injected with soft tissue sarcomas and liposomal-based iodine contrast agent. For both PCD and EID protocols, we acquired separate *ex-vivo* micro-CT scans for the upper and lower body.

To perform material decomposition from multi-EID micro-CT data, we acquired three different EID scans with the following parameter settings: i) 80 kVp, 40 mA, 10 ms/exposure, ii) 50 kVp, 50 mA, 16 ms/exposure, and iii) 40 kVp, 50 mA, 16 ms/exposure. Each scan used a helical trajectory, with 900 projections over 1070 degrees rotation and 1.25 cm vertical translation.

For our PCD scan, we used an X-ray tube voltage of 80 kVp, current of 2 mA, integration time of 200 ms/exposure, and energy thresholds of 25, 34, 50, and 60 keV. Just like the EID scans, this helical scan was acquired with 900 projections, 1070 degrees rotation, and 1.25 cm vertical translation. The estimated dose for this PCD scan was 36 mGy, which is much lower than the estimated dose of 296 mGy for the 3 EID scans described above.

B. Image Reconstruction

For the multi-EID micro-CT data, we performed an iterative reconstruction using the split Bregman method with the add-

R. Nadkarni, A. Allphin, D. P. Clark, and C. T. Badea are with the Quantitative Imaging and Analysis Lab, Dept. of Radiology at Duke University, Durham, NC 27710. Contact: cristian.badea@duke.edu. This work was supported by the NIH (R01 CA196667, U24 CA220245, RF1 AG070149). We acknowledge the support provided by Drs. Y.M. Mowery, Y.Qi and K. Ghaghada for our experiments.

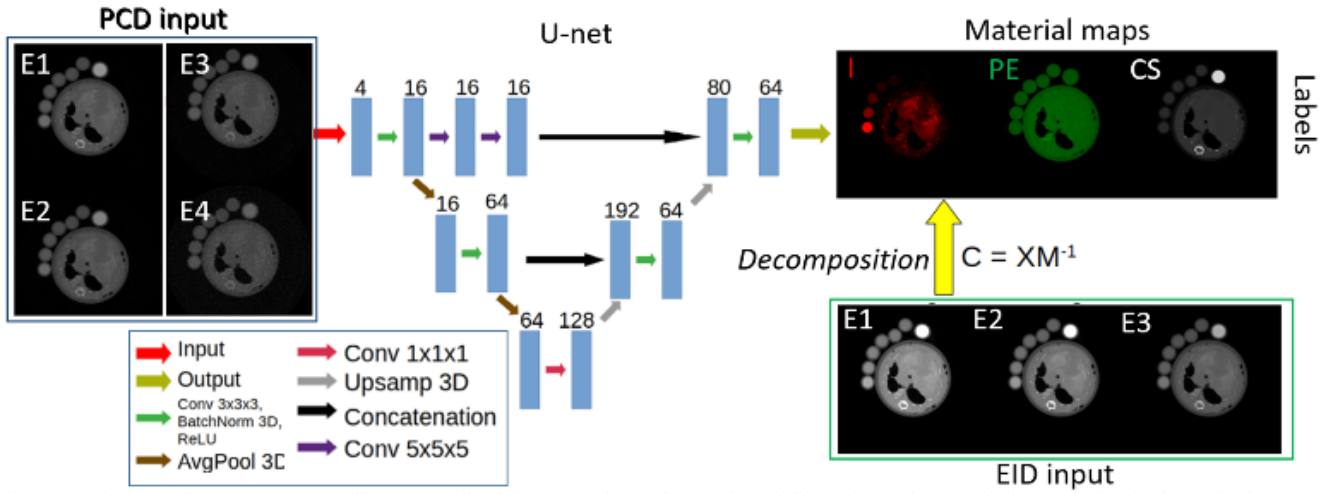


Figure 1. Diagram of our 3D U-net architecture and training procedure. The number of filters for each convolution operation is shown in the diagram. The network is trained using 3D patches of the PCD images as inputs and 3D patches of the multi-EID material maps as labels.

residual-back strategy [8] and rank-sparse kernel regression regularization (RSKR) [9], solving the following optimization problem:

$$\mathrm{X} = \arg\min_{X} \frac{1}{2} \sum_e || RX_e - Y_e ||_2^2 + \lambda || X ||_{BTV}.$$

Thus, we solve iteratively for the vectorized, reconstructed data, the columns of X, for each energy threshold simultaneously (indexed by *e*). The reconstruction for each threshold minimizes the reprojection error (R, system projection matrix) relative to the log-transformed projection data acquired at each threshold (the columns of Y). To reduce noise in the reconstructed results, this data fidelity term is minimized subject to the bilateral total variation (BTV) measured within and between energies via RSKR. Unlike the multi-EID data, the PCD micro-CT images were reconstructed using filtered back projection (FBP) with a Ram-Lak filter to explore the potential of our CNN to perform denoising and material decomposition simultaneously.

C. *Material Decomposition*

We performed image-based material decomposition on the EID and PCD micro-CT reconstructions using the approach of Alvarez and Macovski [10]. We chose to decompose CT images into iodine (I), Compton scattering (CS), and photoelectric effect (PE) maps. Thus, we performed a post-reconstruction spectral decomposition with I, PE and CS as bases: $\mathrm{X}(e) = C_{PE} M_{PE}(e) + C_{CS} M_{CS}(e) + C_I M_I(e)$. The matrix inversion spectral decomposition was performed by solving the following linear system at each voxel: $\mathrm{C} = \mathrm{XM}^{-1}$. In this formulation, *C* is the least-squares solution for the concentration of I (mg/mL) and the fractions of PE and CS relative to water. *M* is a matrix of material sensitivities (attenuation/mg/mL for iodine) at each energy that was computed using vials of water and known concentrations of iodine and calcium that were included in the scans. Finally, *X* is the attenuation coefficient of the voxel under consideration at energy *e*. Orthogonal subspace projection was used to prevent negative concentrations. PCD micro-CT data was also decomposed via matrix inversion for comparison .

D. *Animal Experiments*

We used a transplant model of soft tissue sarcoma that resembles human undifferentiated pleomorphic sarcoma. A sarcoma cell line was generated from an autochthonous soft tissue sarcoma (p53/MCA model) induced in C57BL/6 wild type mice by intramuscular injection of adenovirus expressing Cas9 endonuclease and sgRNA to *Trp53* gene (Adeno-sgp53-Cas9; Viraquest) followed by intramuscular injection of the carcinogen 3-methylcholenthrene (MCA) [11]. Liposomal-based contrast agents containing iodine (Lip-I) were fabricated similar to methods described previously [12]. Two mice with sarcomas were intravenously injected with Lip-I (1.32 mg I/kg body weight) and shortly after, were euthanized. The mice were immersed in formalin for 5 days, then scanned with our PCD micro-CT and multi-EID micro-CT. For this preliminary work, we used two mice with sarcomas, but more mice will be used in the near future.

E. *Network Training and Testing*

Using the Pytorch library, we implemented a CNN with a 3D U-net architecture [13] for material decomposition. The details of network architecture and our training scheme are shown in Figure 1. Three matching sets of PCD and EID micro-CT scans from two mice were used for training and one set from the upper body of one mouse was held out for testing. The CNN was trained using partially overlapping 3D patches (size 38 x 62 x 62, stride 26 x 50 x 50) from PCD micro-CT images as input and smaller (size 26 x 50 x 50, stride 26 x 50 x 50), non-overlapping 3D patches of the corresponding I, CS, and PE material maps obtained via multi-EID micro-CT as labels. Each PCD energy threshold is treated as a separate input channel. Since the CNN does not use zero padding for convolutions, it crops input patches down to the size of the labels, which prevents patch artifacts when the predicted material map patches are stitched back into a complete image. The 5 x 5 x 5 convolutions ensure consistent dimensions of feature maps that are being concatenated. We trained our CNN using a mean square error (MSE) loss function with a batch size of 32, Adam

optimizer [14] with learning rate of 5 x 10^{-5}, and 500 training epochs. Once training was complete, we used the 3D U-net to predict I, CS, and PE maps with partially overlapping 3D patches of PCD micro-CT images from the test set as inputs.

F. Performance Evaluation

We evaluated performance on the test set using both qualitative and quantitative methods. The material maps derived from multi-EID images via matrix inversion served as a surrogate for ground truth. For quantitative comparisons, we measured the root mean square error (RMSE) over the whole volume for PCD with matrix inversion decomposition and for 3D U-net decomposition. In addition, we compared measurements (mean and standard deviation) from I and Ca vials in the I and PE maps for all 3 decomposition approaches. For PCD with matrix inversion-based decomposition, we used the FBP reconstruction as the input because this is also what the 3D U-net takes as an input.

Finally, we generated modulation transfer functions (MTF) derived from iodine maps for the EID decomposition and the 3D U-net decomposition to assess how well the CNN preserves spatial resolution.

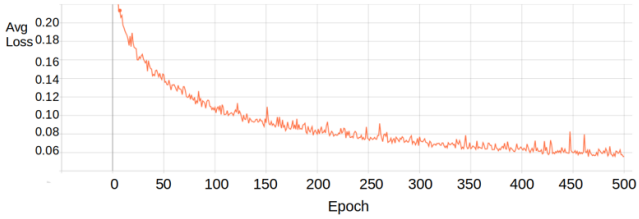


Figure 2. Average training MSE loss. There were a total of 500 epochs.

III. Results

The training loss curve (Figure 2) indicates a decrease in error over the 500 epochs. An axial slice with decompositions provided by the multi-EID data, the PCD micro-CT via matrix inversion and the predictions from 3D U-net are shown in Figure 3. Although the CNN results are the closest to the multi-EID, some smoothing is also noticeable. The 3D U-net prediction appears superior to the PCD decomposition via matrix inversion, which shows higher levels of noise.

Figure 4 shows the RMSE of I, CS and PE material maps for both the 3D U-net prediction and the standard PCD decomposition via matrix inversion. The largest error in the standard PCD decomposition is for the I map. Using the trained 3D U-net reduced the RMSE by 2.597 mg/mL in the I map, 0.221 in the CS map, and 0.370 in the PE map. Note that the expected CS and PE values in water were normalized to 1.0 for the EID data.

Quantitative measurements from I and Ca vials included in the test set scans are shown in Figure 5. Compared to multi-EID decomposition, the PCD decomposition via matrix inversion shows mean measurements that overestimate I concentrations and underestimate PE values in the Ca vial.

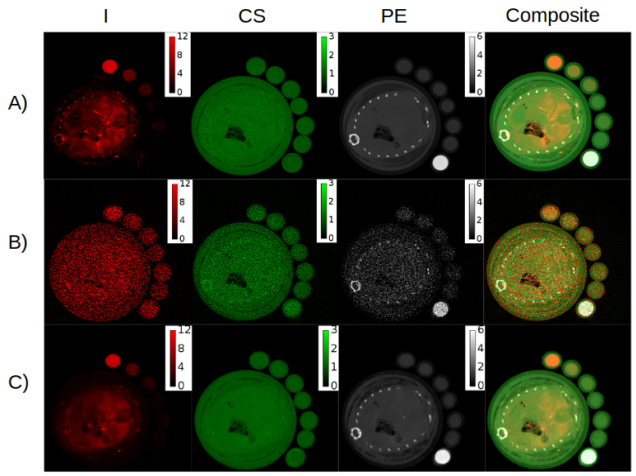


Figure 3. I, CS, PE, and composite material maps in an axial slice from the test set for A) decomposition using matrix inversion from multi-EID scans B) decomposition using matrix inversion from the PCD scan and C) decomposition via prediction by the 3D U-net.

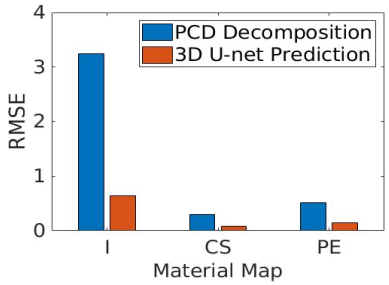


Figure 4. RMSE of I, CS, and PE material maps in the test set. RMSE values for both PCD decomposition via matrix inversion and 3D U-net predicted decomposition are computed relative to the multi-EID decomposition via matrix inversion.

The measurements from 3D U-net prediction are more accurate since they are much closer to the multi-EID ground truth. The mean value from the 3D U-net predicted material map is within 0.6 mg/mL of the mean value from the multi-EID material map for both iodine vials. Measurements from the 3D U-net prediction consistently provide better precision as indicated by smaller standard deviation than measurements from PCD decomposition via matrix inversion. This result is consistent with the previous observation of reduced noise from 3D U-net decomposition.

Figure 6 shows an axial slice from the absolute difference maps between 3D U-net prediction and multi-EID decomposition for each material. These maps indicate that the 3D U-net predictions tend to have the greatest error at edges and in highly enhancing structures such as the bone. The MTFs for the multi-EID decompositions and 3D U-net predictions are in Figure 7. At 10% MTF, the prediction from 3D U-net has a spatial resolution of 257.5 μm, while the multi-EID decomposition has a spatial resolution of 231.5 μm. This confirms the blurring added by the CNN decomposition, and is consistent with the edge errors observed in the absolute difference maps from Figure 6.

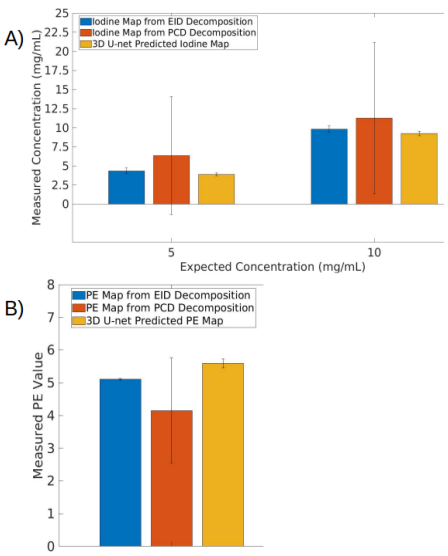


Figure 5. Mean and standard deviation of intensity values measured in vials of known concentration from multi-EID decomposition, PCD decomposition, and 3D U-net predicted material maps. A) Measurements in I map from I vials, with the expected concentrations shown on the x-axis. B) Measurements in PE map from a vial with 50 mg/mL of calcium in water.

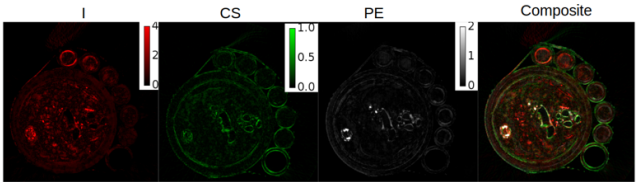


Figure 6. Absolute difference between multi-EID decomposition and 3D U-net predicted material maps in an example axial slice from the test set.

IV. Conclusions

In this work, we implemented a unique approach for DL material decomposition from PCD images by using multi-EID material maps as training labels. Our results on the test set show that a 3D U-net trained with this method can perform material decomposition from a PCD scan with greater quantitative accuracy and precision than the conventional matrix inversion-based approach. There are a few limitations that merit further discussion. The material maps predicted by the 3D U-net produced overly smooth images, resulting in errors at edges. It is possible that these errors can be mitigated by implementing a more sophisticated loss function than MSE such as gradient correlation loss [15]. In addition, like other DL approaches, our trained CNN can only be used on PCD images with similar scan and reconstruction protocols to the training data. Despite these issues, our CNN-based approach is a powerful method for compensating for spectral distortion during material decomposition and exploiting the improved dose efficiency offered by PCCT imaging.

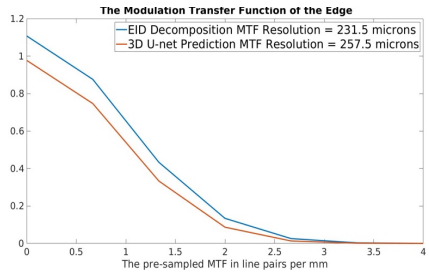


Figure 7. MTF from multi-EID decomposition and from 3D U-net prediction.

References

[1] K. Taguchi and J. S. Iwanczyk, "Vision 20/20: Single photon counting x-ray detectors in medical imaging," *Med Phys,* vol. 40, no. 10, p. 100901, Oct 2013.

[2] M. J. Willemink, M. Persson, A. Pourmorteza, N. J. Pelc, and D. Fleischmann, "Photon-counting CT: Technical Principles and Clinical Prospects," *Radiology,* vol. 289, no. 2, pp. 293-312, Nov 2018.

[3] J. F. P. J. Abascal, N. Ducros, V. Pronina, *et al.*, "Material Decomposition in Spectral CT Using Deep Learning: A Sim2Real Transfer Approach," *IEEE Access,* vol. 9, pp. 25632-25647, 2021.

[4] D. P. Clark, M. Holbrook, and C. T. Badea, "Multi-energy CT decomposition using convolutional neural networks," presented at the SPIE Medical Imaging, 2018.

[5] H. Gong, S. Tao, K. Rajendran, W. Zhou, C. H. McCollough, and S. Leng, "Deep-learning-based direct inversion for material decomposition," *Med Phys,* vol. 47, no. 12, pp. 6294-6309, Dec 2020.

[6] W. Zhang *et al.*, "Image domain dual material decomposition for dual-energy CT using butterfly network," *Med Phys,* vol. 46, no. 5, pp. 2037-2051, May 2019.

[7] M. D. Holbrook, D. P. Clark, and C. T. Badea, "Dual source hybrid spectral micro-CT using an energy-integrating and a photon-counting detector," *Phys Med Biol,* vol. 65, no. 20, p. 205012, Oct 21 2020.

[8] H. Gao, H. Yu, S. Osher, and G. Wang, "Multi-energy CT based on a prior rank, intensity and sparsity model (PRISM)," *Inverse Probl,* vol. 27, no. 11, p. 115012, Nov 1 2011.

[9] D. P. Clark and C. T. Badea, "Hybrid spectral CT reconstruction," *PLOS ONE,* vol. 12, no. 7, p. e0180324, 2017.

[10] R. E. Alvarez and A. Macovski, "Energy-selective reconstructions in X-ray computerized tomography," *Phys Med Biol,* vol. 21, no. 5, pp. 733-44, Sep 1976.

[11] A. J. Wisdom *et al.*, "Single cell analysis reveals distinct immune landscapes in transplant and primary sarcomas that determine response or resistance to immunotherapy," *Nat Commun,* vol. 11, no. 1, p. 6410, Dec 17 2020.

[12] S. Mukundan, Jr. *et al.*, "A liposomal nanoscale contrast agent for preclinical CT in mice," *AJR Am J Roentgenol,* vol. 186, no. 2, pp. 300-7, Feb 2006.

[13] O. Ronneberger, P. Fischer, and T. Brox, "U-Net: Convolutional Networks for Biomedical Image Segmentation," Cham, 2015: Springer International Publishing, in Medical Image Computing and Computer-Assisted Intervention – MICCAI 2015, pp. 234-241.

[14] D. P. Kingma and J. L. Ba, "Adam: A Method for Stochastic Optimization," *International Conference on Learning Representations,* 12/22 2014.

[15] M. D. Holbrook, D. P. Clark, and C. T. Badea, " Deep learning based spectral distortion correction and decomposition for photon counting CT using calibration provided by an energy integrated detector " *SPIE Medical Imaging,* vol. 11595, 15 February 2021 2021.

Using Tissue-Energy Response to Generate Virtual Monoenergetic Images from Conventional CT for Computer-aided Diagnosis of Lesions

Shaojie Chang, Yongfeng Gao, Marc J. Pomeroy, Ti Bai, Hao Zhang, and Zhengrong Liang

***Abstract*— Based on the X-ray physics in computed tomography (CT) imaging, the linear attenuation coefficient (LAC) of each human tissue is described as a function of the X-ray photon energy. Different tissue types (i.e. muscle, fat, bone, and lung tissue) have their energy responses and bring more tissue contrast distribution information along the energy axis, which we call tissue-energy response (TER). In this study, we propose to use TER to generate virtual monoenergetic images (VMIs) from conventional CT for computer-aided diagnosis (CADx) of lesions. Specifically, for a conventional CT image, each tissue fraction can be identified by the TER curve at the effective energy of the setting tube voltage. Based on this, a series of VMIs can be generated by the tissue fractions multiplying the corresponding TER. Moreover, a machine learning (ML) model is developed to exploit the energy-enhanced tissue material features for differentiating malignant from benign lesions, which is based on the data-driven deep learning (DL)-CNN method. Experimental results demonstrated that the DL-CADx models with the proposed method can achieve better classification performance than the conventional CT-based CADx method from three sets of pathologically proven lesion datasets.**

***Index Terms*—CT image analysis, Computer-aided diagnosis, Machine learning, Malignant and benign differentiation**

I. Introduction

In computed tomography (CT) imaging, different tissue types can be represented by the linear attenuation coefficients (LACs) [1]. Based on the well-established X-ray physics inside the human tissues [2], Fig. 1(a) shows the LACs as a function of the X-ray energy for four important human body tissues, i.e., bone, muscle, fat, and lung, and water as the reference. Fig. 1(b) shows the difference between these tissues in terms of CT values. Fig. 1(c) is a zoomed version of Fig. 1(b) to emphasize the differences among muscle, fat, lung and water. It is clearly seen that different tissue types have their energy responses along the energy axis, which we call tissue-energy response (TER) in this study. The different TERs shall bring more tissue contrast distribution information in a series of monoenergetic images than that of the single image reconstructed from the conventional CT, which uses a wide spectrum and could not take the response into consideration. More tissue contrast distribution information shall enhance tissue characterization and, therefore, improve lesion diagnosis. If the energy-independent fraction of each tissue can be obtained, a series of virtual monoenergetic images (VMIs) is able to be generated by the TER curve. Hence, photon-counting CT (PCCT) [3] and multiple energy spectral CT (MECT) [4, 5] are recently developed to reconstruct the energy-independent fractions with multiple energy measurements and then generate a series of VMIs. However, they require the use of expensive photon counting detection technology and/or sophisticated image reconstruction methods. Following the physics behind the TER, this study explores an alternative approach to use the TER to

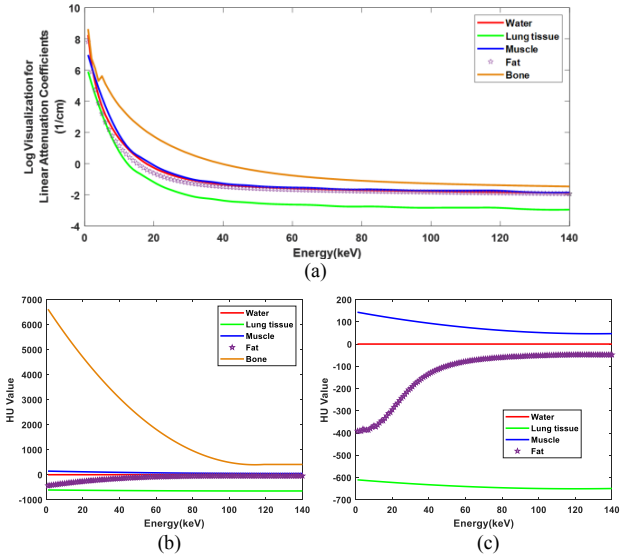


Fig. 1: Tissue energy response curve. (a) Log visualization for linear attenuation coefficients curve; (b) HU curve; (c) Zoomed HU curve.

This work was partially supported by the NIH/NCI grant #CA206171.

Shaojie Chang and Yongfeng Gao are with the Department of Radiology, Stony Brook University, Stony Brook, NY 11794, USA

Marc J. Pomeroy, and Zhengrong Liang are with the Departments of Radiology and Biomedical Engineering, Stony Brook University, Stony Brook, NY 11794, USA (e-mail: Jerome.Liang@SUNYSB.EDU).

Ti Bai is with the Department of Radiation Oncology, University of Texas Southwestern Medical Centre, Dallas, TX 75390, USA

Hao Zhang is with the Department of Medical Physics, Memorial Sloan Kettering Cancer Center, NY 10065, USA

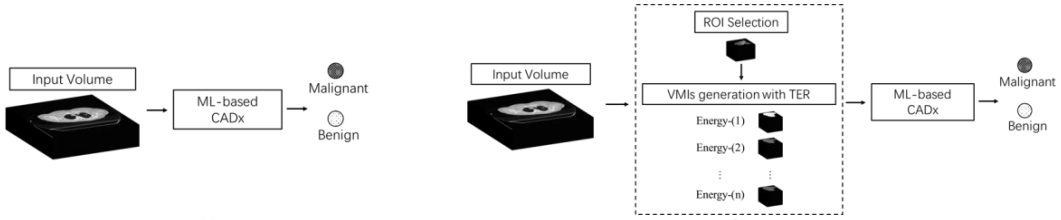


Fig. 2: Illustration and comparison between traditional and our proposed method for lesion diagnosis. (a) Traditional ML-based CADx pipeline. (b) Our proposed ML-based CADx pipeline with the VMIs generated by the use of TER.

generate a series of VMIs from a conventional CT image to further extract contrast textures for lesion diagnosis. An assumption is made that the conventional CT image is obtained from the effective energy of the energy spectrum used by the conventional CT. For example, if conventional CT operates on a 120kVp X-ray tube voltage, the effective photon energy would be 75keV. With the identified location in the TER curve, a series of VMIs are obtained by the relative factor along the energy axis.

With the above VMIs, a machine learning (ML)-based CADx model is proposed to differentiate malignant from benign lesions by exploiting the energy-enhanced material features, which uses automatically extracted features by deep learning (DL) technics. DL-based CADx algorithms, more specifically convolutional neural network (CNN), have achieved noticeable successes in the differentiation of malignant and benign lesions [6-10]. Due to the advanced feature learning power, a multi-channel 3D CNN-based CADx model is developed in this study to help recognize energy-specific features to differentiate malignant lesions from benign ones. The final classification result will reflect the CADx performance with the explored spectral information.

The remainder of this paper is organized as follows. Section II will describe the proposed computer-aided diagnosis framework and the overall workflow. Section III presents the experiment design and results. Discussion and conclusions are drawn in Sections IV and V.

II. Methods

An overall illustration of our proposed lesion diagnosis pipeline can be found in Fig.2. Fig.2(a) shows the traditional diagnosis pipeline, the conventional CT data is directly used in the ML-based CADx model. As shown in Fig.2(b), our proposed framework is composed of two components: one is virtual monoenergetic images (VMIs) generation by the use of TER and the other is ML-based CADx model for classification. The details of each are described as follows.

A. CT Image-based Virtual Monoenergetic Images Generation by the Use of TER

In a CT image, a linear attenuation coefficient can be represented by *R* types of tissues, e.g. muscle, fat, lung, bone, and so on. The LAC function $\mu_j(\varepsilon)$ at the j^{th} pixel of the image is decomposed as:

$$\mu_j(\varepsilon) = \sum_{r=1}^{R} \mu_r(\varepsilon) f_{rj}, \quad (1)$$

where $\mu_r(\varepsilon)$ denotes linear attenuation coefficient of tissue r at energy ε, as shown in Fig.1(a). Notation f_{rj} is a unitless tissue fraction that quantifies the contribution of tissue r to attenuation in pixel j. Once the tissue fractions are identified, a series of VMIs can be generated.

1) Region of interest (ROI) selection

For lesion diagnosis, an ROI containing the lesion is firstly selected from the whole CT image. An example of ROI selection from one patient slice for lung nodule diagnosis is shown in Fig. 3.

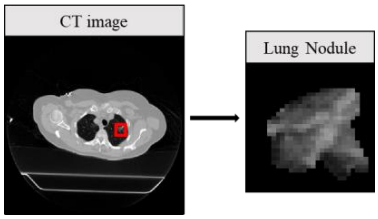


Fig. 3: An example of ROI selection from the CT image for lung nodule diagnosis. The display window is [0,0.35] cm^{-1}.

2) Tissue fractions identification

Due to the polychromatic X-ray source, a CT ROI image μ^{E_1} reconstructed by filtered backprojection (FBP) method reflects the attenuation coefficients at the effective energy with the tube voltage E_1 kVp as follows,

$$\mu_j^{E_1} = \sum_{r=1}^{R} \mu_r\left(E_1^{eff}\right) f_{rj}, \quad (2)$$

where E_1^{eff} denotes the effective energy the selected tube voltage E_1. To identify the tissue fractions, the CT image is first segmented into four tissue types including lung tissue, fat, bone and muscle by a threshold method. And then, for each tissue region, the tissue fraction f_{rj} can be obtained by $\frac{\mu_j^{E_1}}{\mu_r\left(E_1^{eff}\right)}$. In this work, we assumed $E_1^{eff} = 75$ when a conventional CT scan at $E_1 = 120$ kVp. And $\mu_r\left(E_1^{eff}\right)$ can be directly found with TER as shown in Fig.1(a).

3) Virtual monoenergetic CT images generation

Based on the identified tissue fractions, a series of VMIs at selected n energies are generated with the corresponding tissue LACs as follows.

$$\mu_j^{virtual}(\varepsilon) = \sum_r \mu_r(\varepsilon) f_{rj}, \quad (6)$$

In this work, $n = 10$, the energy values are used by 5, 8, 10, 12, 15, 20, 25, 30, 35, 40 and 45 keV as an example. It is because that as shown in Fig. 1(c), the HU values of different tissues have the maximum differences in this energy range and the contrast features could be efficiently enhanced.

B. Machine Learning (ML)-based CADx

1) Deep learning (DL)-based CNN model:

For the DL-based model, a 3D CNN architecture with a multi-channel input is developed, which uses each energy

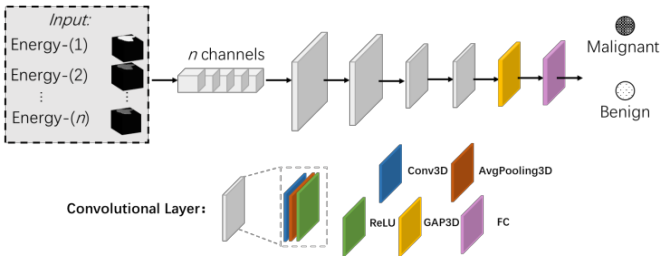


Fig. 4: The 3D CNN architecture of CADx for lesion diagnosis.

TABLE I: DETAILS OF NETWORK DESIGN

Layer	Type	Channels	Kernel Size	Padding	Stride	Activation
1	Conv3D	32	(7,7,7)	(1,1,1)	(1,1,1)	ReLU
2	Avgpooling3D	-	(2,2,2)	-	(2,2,2)	-
3	Conv3D	64	(5,5,5)	(1,1,1)	(1,1,1)	ReLU
4	Avgpooling3D	-	(2,2,2)	-	(2,2,2)	-
5	Conv3D	128	(3,3,3)	(1,1,1)	(1,1,1)	ReLU
6	Avgpooling3D	-	(2,2,2)	-	(2,2,2)	-
7	Conv3D	128	(3,3,3)	(1,1,1)	(1,1,1)	ReLU
8	Avgpooling3D	-	(2,2,2)	-	(2,2,2)	-
9	GAP3D	-	-	-	-	-
10	FC	128	-	-	-	-

image as one input channel as shown in Fig. 4. First, four convolutional layers are used to extract the features, which are then pooled together with a global average pooling (GAP3D) layer such that the final features have a global receptive field. At last, a fully connected (FC) layer-based classifier is adopted to distinguish the malignant and benign lesions. Specifically, each convolutional layer consists of three operators: 3D convolution (Conv3D), 3D average pooling (Avgpolling3D), and rectified linear unit (ReLU). And the binary cross-entropy loss is utilized to train the model. The details of the CNN model are listed in Table I.

Based on the above model, the final classification results will show the lesion diagnosis performance with the VMIs generated by the TER.

III. EXPERIMENTS AND RESULTS

A. Datasets

In this study, three pathologically proven clinical datasets consisting of colon polyp and lung nodules are used to evaluate our proposed method. All the patients were recruited to this study under informed consent after approval by the Institutional Review Board. Details are presented in Table II.

TABLE II: DATASETS INFORMATION

Dataset	Total Number	Benign (0)	Malignant (1)	Pathological Report
Dataset 1	63	31	32	✓
Dataset 2	67	18	49	✓
Dataset 3	114	50	64	✓

1) Dataset 1

In dataset 1, 59 patients were scanned by a conventional CT at 120 kVp with automatic exposure control at the University of Wisconsin, USA. A total of 63 colon polyp masses were found and resected by the clinical examination. The pathological reports indicate 31 benign and 32 malignant polyps.

2) Dataset 2

In dataset 2, 66 patients were scheduled for CT-guided lung nodule needle biopsy at 120 kVp with automatic exposure control at Stony Brook University Hospital, USA. With the pathological reports, a total of 67 lung nodules with 18 benign and 49 malignant were confirmed.

3) Dataset 3

In dataset 3, 114 patients were scheduled for CT-guided lung nodule needle biopsy with X-ray exposure of clinical dose at 120 kVp, 100 mAs in Stony Brook University Hospital, USA. With the pathological report, a total of 114 lung nodules with 50 benign and 64 malignant were confirmed.

B. CNN Training Implementation

For the input to the CNN-based implementation, we first converted each *n*-energy data with the resolution of 64×64×64 voxels. And these converted energy volumetric images were fed into the multi-channel 3D CNN as shown in Fig. 4 for training. And the target is the results from the pathological reports of the malignant and benign lesions. The k-fold (k=5) cross-validation was implemented to test the robustness and avoid data bias. The procedure is as follows. We firstly shuffled the dataset randomly and split it into 5 folds. For each fold, we randomly divided the dataset into training and testing datasets. And then we trained a model on the training dataset and evaluated it on the testing dataset. Finally, we retained the evaluation score for each fold and the average score was calculated. In this study, the CNN model was trained for 100 epochs with a learning rate of 0.001 and batch size of 8 using Adam optimizer [11].

C. Classification Performance

TABLE III: MEAN AUC VALUES FOR DL-BASED CNN DIAGNOSIS MODEL

Data Input	Dataset 1	Dataset 2	Dataset 3
Conventional CT (120kVp)	74.20	52.42	59.54
VMIs with TER	**80.71**	**69.33**	**71.97**

For the DL-based CNN model, the conventional CT and the enhanced 10-energy VMI data generated by the use of TER were incorporated into our 3D-CNN network, respectively. We

calculated the mean values of AUC scores, which are shown in Table III. For both Dataset 2 and Dataset 3, the lung nodules are undifferentiable to be benign or malignant by human experts who recommended biopsy on these nodules. By the reports of their biopsies as the ground truth, the CADx with 52.42% / 59.54% AUC scores performed slightly better than human experts, whose score would be random or 50%. The results from the CADx on the VMIs with TER demonstrated that the VMIs data with the TER achieve higher mean AUC values than the conventional 120 kVp data, which verifies the effectiveness of the contrast enhancement brought from the VMIs. The AUC values of each dataset can be improved 6.51%, 16.91%&12.55% for lesion characterization, respectively. This is powerful proof that our proposed CADx model could benefit from the energy spectral information in VMIs.

IV. Discussion

This study aims at exploring the energy spectral information from a conventional CT image by using the TER to generate VMIs for lesion diagnosis. The proposed CADx framework with the VMIs achieved improved diagnosis performance than the traditional CADx pipeline with the conventional CT data. As we have demonstrated in this work, the VMIs at different energy bins show significant effectiveness for lesion characterization. How to choose the energy range and energy number of the VMIs to capture the meaningful lesion features is still a promising direction. With the analysis from the above, this work will also have a great potential in guiding the energy selection in PCCT imaging for diagnosis. Meanwhile, to compare the PCCT images from the practical measurements with the VMIs by the use of TER from the conventional CT would be a very interesting research topic in the future. And last but not least, clinical evaluations with more disease data sets are needed to test the robustness of the proposed method.

V. Conclusions

In conclusion, we proposed to use tissue-energy response to generate the virtual monoenergetic images from the conventional CT for CADx of lesions. In this framework, each tissue contrast distribution along the energy axis is fully enhanced, which brings richer information to ML-based CADx. Experimental results demonstrated that the VMIs generated with the use of TER from the conventional CT applied to the ML-based CADx obtain better performance than the traditional CADx pipeline in lesion classification.

References

[1] C. H. McCollough, S. Leng, L. Yu, and J. G. Fletcher, "Dual-and multi-energy CT: principles, technical approaches, and clinical applications," Radiology, vol. 276, no. 3, pp. 637-653, 2015.

[2] Y. Gao, Y. Shi, W. Cao, S. Zhang, and Z. Liang, "Energy enhanced tissue texture in spectral computed tomography for lesion classification," Visual Computing for Industry, Biomedicine, and Art, vol. 2, no. 1, pp. 1-12, 2019.

[3] S. Leng et al., "Photon-counting detector CT: system design and clinical applications of an emerging technology," Radiographics, vol. 39, no. 3, pp. 729-743, 2019.

[4] H. Machida et al., "Dual-energy spectral CT: various clinical vascular applications," Radiographics, vol. 36, no. 4, pp. 1215-1232, 2016.

[5] R. Forghani, B. De Man, and R. Gupta, "Dual-energy computed tomography: physical principles, approaches to scanning, usage, and implementation: part 1," Neuroimaging Clinics, vol. 27, no. 3, pp. 371-384, 2017.

[6] A. A. A. Setio et al., "Pulmonary nodule detection in CT images: false positive reduction using multi-view convolutional networks," IEEE transactions on medical imaging, vol. 35, no. 5, pp. 1160-1169, 2016.

[7] H. Jiang, H. Ma, W. Qian, M. Gao, and Y. Li, "An automatic detection system of lung nodule based on multigroup patch-based deep learning network," IEEE journal of biomedical and health informatics, vol. 22, no. 4, pp. 1227-1237, 2017.

[8] H. Wang et al., "A hybrid CNN feature model for pulmonary nodule malignancy risk differentiation," Journal of X-ray Science and Technology, vol. 26, no. 2, pp. 171-187, 2018.

[9] J. Tan et al., "3D-GLCM CNN: A 3-dimensional gray-level Co-occurrence matrix-based CNN model for polyp classification via CT colonography," IEEE transactions on medical imaging, vol. 39, no. 6, pp. 2013-2024, 2019.

[10] S. Zhang et al., "An investigation of CNN models for differentiating malignant from benign lesions using small pathologically proven datasets," Computerized Medical Imaging and Graphics, vol. 77, p. 101645, 2019.

[11] D. P. Kingma and J. Ba, "Adam: A method for stochastic optimization," arXiv preprint arXiv:1412.6980, 2014.

Detruncation of Clinical CT Scans Using a Discrete Algebraic Reconstruction Technique Prior

Achim Byl, Michael Knaup, Magdalena Rafecas, Christoph Hoeschen, and Marc Kachelrieß

***Abstract*—Successful image reconstruction in computed tomography (CT) relies on the completeness of the projections. If the patient does not fit in the field of measurement, the projections are truncated causing cupping artifacts in the image and a diminished field of view (FOV). In order to restore the CT values and extend the FOV, the projections have to be completed, for example via an extrapolation. The discrete algebraic reconstruction technique (DART) has shown its efficacy in reconstructing discrete images from insufficient raw data. In this work, we use DART images as a prior for projection completion of clinical CT scans. We compare our method to the conventional adaptive detruncation (ADT) and evaluate the RMSE inside and outside the FOV along with the Dice score.**

I. Introduction

Image reconstruction in computed tomography (CT) depends heavily on the quality of the available raw data. However, in many instances the acquisition parameters lead to raw data that are insufficient for a conventional method such as filtered back-projection (FBP). Typical problems are a low number of projections, small angular range, or low tube current. This work focuses on the issue of truncated projections, i.e. the case where the patient does not fully fit in the maximum field of view (FOV) of the scanner. In clinical practice, truncation most often occurs with obese patients, patients that are not centered properly on the table, or when using C-arm systems, which generally feature small FOVs.

Sinogram truncation is twofold problematic. On the one hand, voxels inside of the FOV suffer from cupping artifacts. On the other hand, there is a demand of artificially extending the FOV. Voxels in the extended FOV (eFOV) are of interest for several algorithms, including metal artifact reduction and beam hardening correction in CT, and attenuation and scatter correction in PET [1], [2].

In order to reduce artifacts within the FOV, several data extrapolation methods have been developed [3]–[5]. These conventional algorithms are able to restore the CT values within the FOV to acceptable values, such that the cupping artifacts are removed. However, voxels in the eFOV are typically not reconstructed accurately. Recently, there have also been deep learning methods developed for the purpose of extending the FOV, showing promising results [6], [7].

Another method that has shown promise for reconstructing truncated raw data is the discrete algebraic reconstruction technique (DART) [8]–[10]. DART is a well-known algorithm for reconstructing discrete CT volumes, e.g. in the context of non-destructive testing of homogeneous objects. It was specifically developed for scans with few projections or limited angular range [11]–[13]. Utilizing the assumption that the object consists of very few attenuation coefficients, DART is capable of image reconstruction with severely limited data.

In this work, we apply DART to clinical CT data to obtain a prior image for detruncation, i.e. to fill the missing projections. Although clinical CT volumes are not strictly homogeneous, a typical CT scan consists only of soft tissue, bone, and air. In addition, for the purposes of the projection completion, the patient is often approximated by a single attenuation value [3]–[5]. Although DART was developed for discrete CT, it internally computes real valued images, which we use for detruncation. To evaluate our method, it is compared to the established adaptive detruncation (ADT) [4].

II. Methods

We define the detruncation problem as follows. We have access to truncated parallel beam raw data $p_{\mathrm{t}}(\vartheta, \xi)$ with M_{t} pixels of size $\triangle\xi$. Let $\xi_{\mathrm{t,max}} = \frac{1}{2}(M_t - 1)\triangle\xi$, such that the detector range is $-\xi_{\mathrm{t,max}} \leq \xi \leq \xi_{\mathrm{t,max}}$. In order to reconstruct an artifact free image, the full detector should have $M_f > M_t$ pixels, corresponding to $\xi_{\mathrm{f,max}} = \frac{1}{2}(M_f - 1)\triangle\xi$. Thus, the required extrapolation includes pixels within $\xi_{\mathrm{t,max}} < |\xi| \leq \xi_{\mathrm{f,max}}$. In image domain, this expands the original FOV with radius $R_{\mathrm{M,t}} = \xi_{\mathrm{t,max}}$ to the eFOV with radius $R_{\mathrm{M,f}} = \xi_{\mathrm{f,max}}$. Due to the limitations of the scanner, the eFOV cannot be larger than the bore size. Figure 1 shows our strategy to fill the missing projections: First, generate a prior image using the DART algorithm. Second, forward-project the prior image and fill in the missing data in the original sinogram using the prior sinogram.

A. Generating a DART Prior Image

DART employs a combination of discretization and real-valued algebraic reconstruction [11], [12]. Figure 1 shows the basic DART scheme on the left. During each DART iteration, only a subset of pixels is changed, while the rest of the pixels are set to constant values. In order to determine the fixed pixels, we first segment the current image estimate into air and

Achim Byl is with the German Cancer Research Center (DKFZ), Heidelberg, Germany, and with Heidelberg University, Heidelberg, Germany. Email: achim.byl@dkfz.de.

Dr. Michael Knaup is with the German Cancer Research Center (DKFZ), Heidelberg, Germany.

Prof. Dr. Magdalena Rafecas is with the University of Lübeck, Lübeck, Germany.

Prof. Dr. Christoph Hoeschen is with the University Magdeburg, Magdeburg, Germany.

Prof. Dr. Marc Kachelrieß is with the German Cancer Research Center (DKFZ), Heidelberg, Germany, and with Heidelberg University, Heidelberg, Germany.

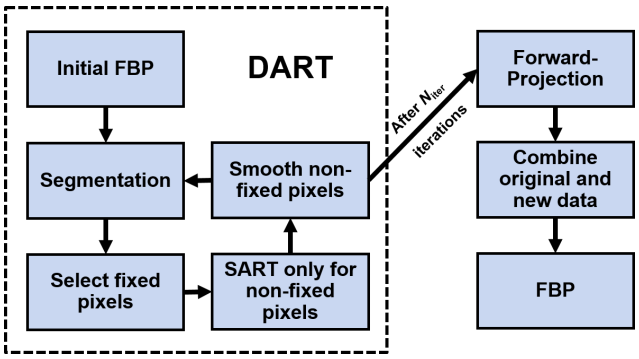


Fig. 1. Scheme of the proposed DART detruncation. The dashed box highlights the DART reconstruction, the right side shows the data completion step.

tissue by thresholding at -500 HU. Then, every pixel that is fully surrounded by pixels of the same class is fixed and set to a value of -1000 HU or 100 HU for air or tissue, respectively. Each pixel is given an additional probability of 65% to be classified as non-fixed, so that values inside homogeneous patches can be corrected. Subsequently, we perform five SART iterations where only the non-fixed pixels are considered [14]. One SART iteration is defined as

$$f_{\text{new}} = f + \lambda \frac{1}{X^T 1} X^T \left(\frac{p - Xf}{X1} \right), \tag{1}$$

where f is the current image estimate, f_{new} is the new estimate, p are the raw data, λ is a relaxation factor, X is the forward-projection and X^T is the back-projection.

Since the fixed pixels contain no noise, all projection noise will be distributed over the non-fixed pixels. Therefore, the last step is a smoothing with a Gaussian filter with a standard deviation of 0.5 pixels. This finishes one iteration and the algorithm continues with the segmentation step. Since there is no formal stopping criterion, we set the maximum number of iterations N_{iter} to 5000. In most cases, a few hundred iterations were sufficient. To initialize the DART algorithm, we perform a simple cosine detruncation followed by an FBP. Note that only the original raw data are used for the DART iterations.

In conventional applications, the DART result will again be segmented at the end of the algorithm, as a discrete volume is desired. In this work, DART is used with clinical data. Therefore, we omit the final thresholding and instead use the real valued image as the prior image.

B. Projection Completion

In order to acquire a complete sinogram, the information in the prior image and the original raw data p_{t} have to be combined. First, forward-project the prior image to obtain the prior sinogram p_{DART}. Then, the detruncated sinogram p_{d} incorporates the original data p_{t} wherever possible and p_{DART} otherwise. Thus, only the outer regions are filled with the prior sinogram. However, simply copying the pixel values might cause artifacts at the edge between original and new data. To amend this, a factor $c(\vartheta, \xi)$ scales the new data to match the original data at the boundary. This factor is calculated separately for each projection and side. Finally, the DART-detruncated image is reconstructed from p_{d} via FBP.

C. CT Data

All data sets were obtained with a SOMATOM Force CT scanner (Siemens Healthineers, Forchheim, Germany). Images were acquired at 70 kV, and reconstructed on a 512×512 matrix with 0.6 mm slice thickness and 0.8 mm $\times$ 0.8 mm pixel size.

We generate full and truncated sinograms by monochromatically forward-projecting in parallel beam geometry. The simulations are performed with $N = 256$ projections with an angular range of 0 to $180°$. The full sinogram has $M_{\text{f}} = 1024$ detector pixels with a pixel size of 0.5 mm. We investigate two levels of truncation with $M_{\text{t}} = 682,\ 372$ detector pixels, which correspond to a relative detector size of $\frac{2}{3}$ and $\frac{1}{3}$, respectively. Note that our method can easily be adjusted to other geometries, e.g. fan-beam or cone-beam.

D. Analysis

To quantify how well the algorithm can correct CT values inside the original FOV, we calculate the root mean squared error (RMSE) within the FOV with respect to the image reconstructed from the full detector. The RMSE for an image $f(i,j)$ with respect to the ground truth image GT is defined as

$$\text{RMSE} = \sqrt{\frac{1}{N_x N_y} \sum_{0 \le i,j \le N_x, N_y} (f(i,j) - \text{GT}(i,j))}. \tag{2}$$

Furthermore, we compute the RMSE within the eFOV and Dice score of the whole image to compare the ability of the detruncation to restore the CT values and patient outline outside of the original FOV. For the latter, the images are segmented into air and soft tissue with a threshold of -500 HU. The Dice score is defined as

$$\text{Dice} = \frac{2\text{TP}}{2\text{TP} + \text{FP} + \text{FN}}, \tag{3}$$

where TP are the true positives, FP the false positives and FN the false negatives. To compare the DART detruncation with a conventional method, we also apply the adaptive detruncation (ADT) [4] to the data.

III. Results

Figure 2 shows the first patient, including detruncation results from cosine detruncation, ADT, DART, and our method. In the left column, the scan suffers from light truncation. This causes minor cupping artifacts at the left and right side of the patient in the uncorrected FBP reconstruction. All methods have have successfully removed the cupping and restored CT values to voxels in the eFOV. After cosine detruncation, artifacts remain at the edge of the FOV and the tissue in the eFOV is too dark. In the ADT result, voxels on the right side are slightly brighter than in the other ground truth. For the case of strong truncation, displayed in the right column of Figure 2, there is a considerable difference between ADT and

TABLE I
RMSE AND DICE VALUES FOR BOTH PATIENTS AND LEVELS OF TRUNCATION. RMSE IS GIVEN FOR THE ORIGINAL AND FOR THE EXTENDED FOV.

		Truncated		Cosine Detruncation		ADT		DART		Our Method	
		$M_t = 678$	$M_t = 372$	$M_t = 678$	$M_t = 372$	M=678	$M_t = 372$	$M_t = 678$	$M_t = 372$	$M_t = 678$	$M_t = 372$
Patient 1	$RMSE_{FOV}$ [HU]	82.01	1005.6	28.38	92.56	15.64	38.53	124.25	141.07	**9.67**	**35.57**
	$RMSE_{eFOV}$ [HU]	145.31	788.32	72.75	275.07	42.20	152.89	108.16	163.23	**28.34**	**35.57**
	Dice	0.977	0.65	0.991	0.859	0.996	0.968	0.996	0.973	**0.999**	**0.975**
Patient 2	$RMSE_{FOV}$ [HU]	97.29	912.14	27.94	114.15	14.83	**26.79**	150.42	178.30	**12.72**	45.92
	$RMSE_{eFOV}$ [HU]	162.53	699.16	73.03	294.24	42.77	**127.81**	132.38	194.03	**34.47**	161.48
	Dice	0.9703	0.673	0.988	0.835	0.996	**0.965**	0.993	0.960	**0.999**	0.963

our method. While ADT is able to restore CT values inside the original FOV very well, the tissue outside of the FOV is too bright and the patient outline does not agree with the ground truth. In contrast, the DART-detruncated image is very close to the full-view reconstruction, although fat has been converted into water at the bottom of the image and some bone anatomy is incorrect. These findings are supported by the numerical evaluation in Table I. In all cases, our method outperforms ADT in terms of RMSE and Dice score. Notably, the DART prior image has relatively high RMSE values (above 100 HU in all cases) but good Dice scores, indicating high image noise. Figure 3 shows the projections for the case of strong truncation. Although the ADT is able to estimate the projections approximately, the DART prior and DART detruncation are noticeably smoother and more accurate.

Figure 4 displays the results of the second patient. Again, both ADT and DART detruncation perform well when the data are only mildly truncated. As for the first case, the ADT image features slightly too high CT values on the right side. The DART prior image and DART-detruncated image show a dark streak along the edge of the FOV. For strong truncation, the ADT again yields CT values that are too high on the left and right side of the FOV and does not provide a clear patient outline. Still, ADT clearly outperforms the cosine detruncation. While the DART reconstruction gives more accurate CT values in general, there are gaps in the soft tissue of the patient. This artifact is also evident in the RMSE of the eFOV and the Dice score, which are lower than ADT for both the DART and DART-detruncated image. In the case of mild truncation, our method still yields superior numerical results.

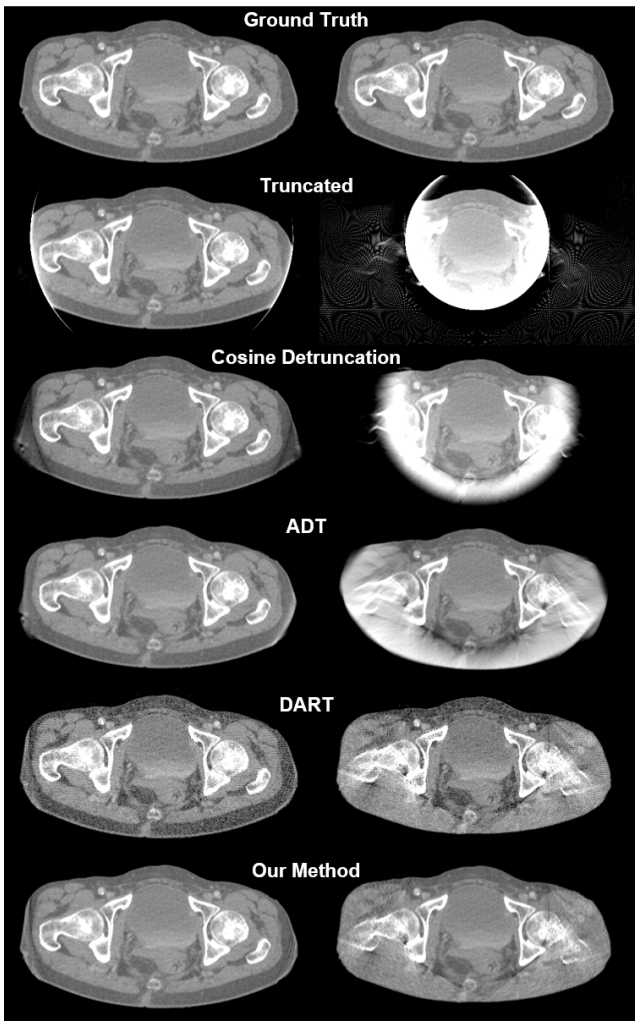


Fig. 2. Detruncation results for Patient 1. Left column shows mild truncation ($M_t = 682$), right column strong truncation ($M_t = 372$). Top to bottom: ground truth, truncated FBP, cosine detruncation, ADT, DART prior image, DART detruncation. $C = 0$ HU, $W = 1000$ HU.

IV. DISCUSSION & CONCLUSION

In this work, we propose a method of CT raw data detruncation based on the DART reconstruction. The DART detruncation was capable of reducing cupping artifacts in the FOV, as well as restoring CT values in the eFOV, for two levels of truncation. Compared to the conventional ADT, our method produced superior visual and numerical results. However, for patient 2 with strong truncation, some anatomy was incorrectly reconstructed as air, yielding worse numerical results than ADT. These artifacts would likely be prevented with a better initial estimate of the image, along with additional iterations of DART.

This work uses the original DART algorithm [11], [12] with a single, manually determined threshold. In the future, a more refined version of DART, e.g. with automatic parameter optimization or improved performance for noisy projections [13], [15], [16] should be preferred, in order to reduce computation time and image noise.

ACKNOWLEDGMENT

This work is supported by the German Federal Ministry for the Environment, Nature Conservation, Nuclear Safety and Consumer Protection (BMUV) under grant 67KI2036B.

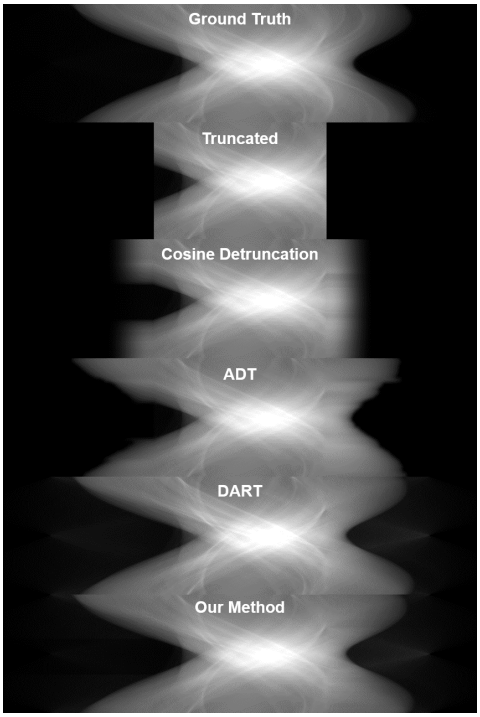


Fig. 3. Projections for Patient 1 with strong truncation after detruncation. Top to bottom: ground truth, truncated FBP, cosine detruncation, ADT, DART prior image, DART detruncation.

Parts of the reconstruction software were provided by RayConStruct®GmbH, Nürnberg, Germany.

References

[1] L. Shi, N. R. Bennett, A. Shiroma, M. Sun, J. Zhang, R. Colbeth, J. StarLack, M. Lu, and A. S. Wang, "Single-pass metal artifact reduction using a dual-layer flat panel detector," *Medical Physics*, vol. 48, no. 10, pp. 6482–6496, 2021.

[2] S. Žabić, L. He, and Y. Bao, "A method for reduction of axial truncation artifacts," in *CT Meeting 2020 Proceedings*, 2020, pp. 158–161.

[3] J. Hsieh, E. Chao, J. Thibault, B. Grekowicz, A. Horst, S. McOlash, and T. Myers, "A novel reconstruction algorithm to extend the CT scan field-of-view," *Medical Physics*, vol. 31, no. 9, pp. 2385–2391, 2004.

[4] K. Sourbelle, M. Kachelrieß, and W. A. Kalender, "Reconstruction from truncated projections in CT using adaptive detruncation," *European Radiology*, vol. 15, no. 5, pp. 1008–1014, 2005.

[5] J. S. Maltz, S. Bose, H. P. Shukla, and A. R. Bani-Hashemi, "CT truncation artifact removal using water-equivalent thicknesses derived from truncated projection data," in *2007 29th Annual International Conference of the IEEE Engineering in Medicine and Biology Society*. IEEE, 2007, pp. 2907–2911.

[6] G. P. Fonseca, M. Baer-Beck, E. Fournie, C. Hofmann, I. Rinaldi, M. C. Ollers, W. J. van Elmpt, and F. Verhaegen, "Evaluation of novel AI-based extended field-of-view CT reconstructions," *Medical Physics*, 2021.

[7] Y. Huang, L. Gao, A. Preuhs, and A. Maier, "Field of view extension in computed tomography using deep learning prior," in *Bildverarbeitung für die Medizin 2020*. Springer, 2020, pp. 186–191.

[8] G. Van Gompel, M. Defrise, and K. J. Batenburg, "Reconstruction of a uniform star object from interior x-ray data: Uniqueness, stability and algorithm," *Inverse Problems*, vol. 25, no. 6, p. 065010, 2009.

[9] G. Van Gompel, *Towards Accurate Image Reconstruction from Truncated X-Ray CT Projections*. Universiteit Antwerpen, Faculteit Wetenschappen, Departement Fysica, 2009.

[10] H. Banjak, M. Costin, C. Vienne, R. Guillamet, and V. Kaftandjian, "Iterative CT reconstruction on limited angle trajectories applied to robotic inspection," in *AIP Conference Proceedings*, vol. 1806, no. 1. AIP Publishing LLC, 2017, p. 020009.

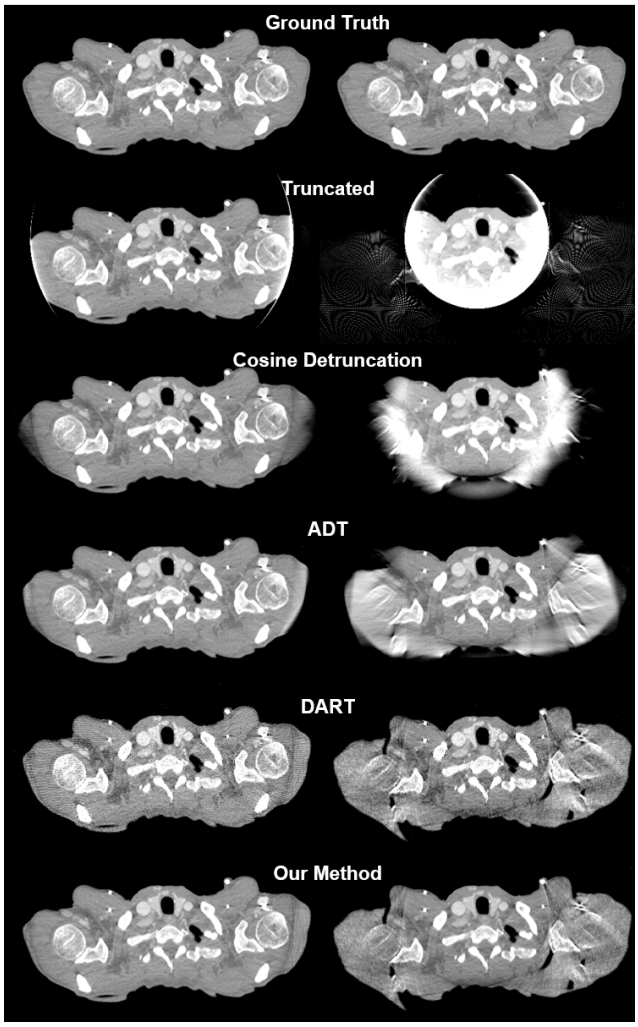


Fig. 4. Detruncation results for Patient 2. Left column shows mild truncation ($M_t = 682$), right column strong truncation ($M_t = 372$). Top to bottom: ground truth, truncated FBP, cosine detruncation, ADT, DART prior image, DART detruncation. $C = 0$ HU, $W = 1000$ HU.

[11] K. J. Batenburg and J. Sijbers, "DART: a fast heuristic algebraic reconstruction algorithm for discrete tomography," in *2007 IEEE International Conference on Image Processing*, vol. 4. IEEE, 2007, pp. IV133–IV136.

[12] ——, "DART: a practical reconstruction algorithm for discrete tomography," *IEEE Transactions on Image Processing*, vol. 20, no. 9, pp. 2542–2553, 2011.

[13] F. Yang, D. Zhang, K. Huang, Z. Gao, and Y. Yang, "Incomplete projection reconstruction of computed tomography based on the modified discrete algebraic reconstruction technique," *Measurement Science and Technology*, vol. 29, no. 2, p. 025405, 2018.

[14] A. H. Andersen and A. C. Kak, "Simultaneous algebraic reconstruction technique (SART): a superior implementation of the ART algorithm," *Ultrasonic Imaging*, vol. 6, no. 1, pp. 81–94, 1984.

[15] W. van Aarle, K. J. Batenburg, and J. Sijbers, "Automatic parameter estimation for the discrete algebraic reconstruction technique (DART)," *IEEE Transactions on Image Processing*, vol. 21, no. 11, pp. 4608–4621, 2012.

[16] X. Zhuge, W. J. Palenstijn, and K. J. Batenburg, "TVR-DART: A more robust algorithm for discrete tomography from limited projection data with automated gray value estimation," *IEEE Transactions on Image Processing*, vol. 25, no. 1, pp. 455–468, 2016.

Deep Learning based Respiratory Surrogate Signal Extraction

Jean Radig, Pascal Paysan, Stefan Scheib

***Abstract*—We present a feasibility study on extracting the respiratory surrogate signal (RSS) from cone-beam computed tomography (CBCT) projections using a supervised convolutional neural network (CNN) model. Determining the intrinsic RSS instead of using an external surrogate signal, provided by optical tracing hardware such as the Real-time Position Management (RPM) system, has the advantage that it spares patient setup time and hence permits faster 4D CBCT acquisition. Another convenience of such an approach is that it can be applied retrospectively without special preparation or equipment. For the implementation, we made use of the MONAI open source library. Our model is based on a modified version of the MONAI regressor class. We trained the model using CBCT projections with the corresponding RSS as recorded by an external marker block using the RPM system. The model is to deduce the RSS given the CBCT. Using a specific dataset with CBCT from anesthetized animals breathing with the help of a mechanical ventilator, results show a good correlation between the actual and predicted RSS. Unlike the Amsterdam Shroud algorithm, our method shows promising results to predict the normalized amplitude of the breathing signal. Further work could extend the model to permit RSS prediction for its online use in radiation therapy or detection of sudden motion deteriorating CBCT image quality. To conclude, we have made a first step towards proving the concept which consists in using a deep learning model to extract the RSS out of acquired CBCT projection images. The approach is promising but requires more work for robustness, i.e. for sufficient accuracy both in the frequency and normalized amplitude extraction.**

***Index Terms*—CBCT, respiratory phase, respiratory amplitude, convolutional neural network, MONAI**

I. Introduction

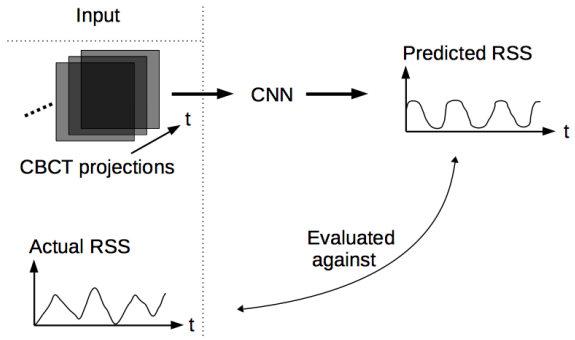
In medical imaging as well as in radiation therapy, patient motion is a challenge that needs to be dealt with. Knowing the internal and external motions of the patient permits taking them into consideration for image reconstruction and treatment planning. Of interest is for example the patient's chest motion, as it gives a proxy for the respiratory motion. Obtaining the patient's chest motion can be done by recording the position of a marker block, a device positioned onto the patient's chest. The position is recorded by optical tracing such as the Real-Time Position Management (RPM) system. The technique requires additional installation time, which we would like to avoid. To this end, we will focus on extracting the patient's chest motion, called for generality the respiratory surrogate signal (RSS), intrinsically from the patient's x-ray projections. In the following, we will focus on CBCT projection images. Selecting projections for the same bin of the RSS periodically at different angles around the patient permits their use for 4D CBCT reconstruction. The 4D CBCT can then be used e.g. for radiotherapy planning, patient setup, motion analyses, and treatment beam gating. Hence the reliable intrinsic determination of the RSS would permit efficient 4D CBCT acquisition and clinical use. Various methods to obtain the RSS directly from the CBCT projections have been proposed such as the Amsterdam Shroud (AS) algorithm [1], a Fourier transform-phase based method [2], intensity-based determination [3], a center-of-mass based [4] method, and the LPCA method [5], [6]. A comparison of the methods has been presented in [7]. Even methods using AI to segment the diaphragm [8] for tumor tracking and breathing phase extraction have been proposed. We compare the presently proposed method against our in-house Amsterdam Shroud implementation which shows good accuracy against recorded RSS from the RPM system but lacks the ability to recover any kind of motion amplitude and makes certain assumptions about expected frequencies. In this work, we would like to investigate if intrinsic RSS extraction could be obtained by using a 3D-CNN model. Such a method would be both fast and easy to use. As a first approach, we modified the MONAI regressor class to serve our purpose. MONAI [9] is an open-source framework for artificial intelligence in medical applications. We use supervised learning by giving both the complete CBCT projection images and the RSS as recorded by the RPM system to the training loop. The model learns to minimize the difference between the predicted RSS and the one as recorded by the RPM system, which we will denote as the "actual" RSS. In the evaluation loop, we would like to get the predicted RSS solely from the CBCT (Fig. 1).

In the following, we present the framework and the architecture of our model. We then show some preliminary results and finally conclude by discussing further development and use of the work.

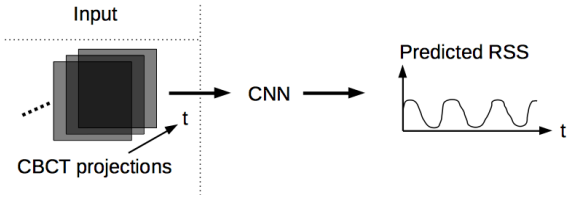
J. Radig, P. Paysan, and S. Scheib are employees of Varian, a Siemens Healthineers company, Taefernstrasse 7, 5405 Daettwil, Switzerland

II. Framework, Architecture, and Data

OUR CNN model is based on a modified version of the open-source MONAI regressor class implemented in the PyTorch Lightning framework. Given the CBCT raw data, we define the input of dimensions (batch, channels, number of projections, projection dimension along y, projection dimension along x) which in our case, using (by factor 2,4) down-sampled projections, read (1, 1, number of projections [variable in function of the dataset], 384, 256). We operate successive convolutions to collapse the x, y dimensions and preserve the dimension corresponding to the number of projections. Accordingly, the actual RSS for each projection is



(a) Scheme of the training loop



(b) Scheme of the evaluation loop

Fig. 1: Scheme of both the training and the evaluation loops

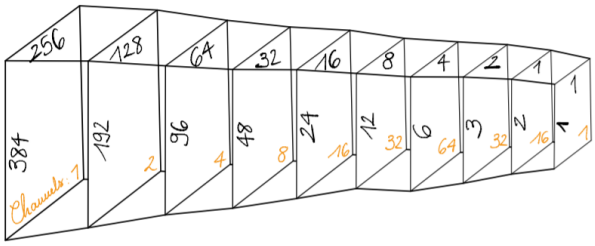


Fig. 2: Successive dimension reductions via 9 convolutions are shown for the projection dimensions. The number of projections stays the same throughout the process, only the dimensions in y and x are altered. The channels per convolution are evolving as (2, 4, 8, 16, 32, 64, 32, 16, 1). The final dimension (number of projections) is reflecting the dimension of the recorded RSS.

recorded during CBCT acquisition and given as input. The network consists of 9 convolutions with kernel size (3, 3, 3), padding (1, 0, 0), and strides (1, 2, 2). We set the number of channels for the 9 convolutions to (2, 4, 8, 16, 32, 64, 32, 16, 1).

We replaced the fully connected layer of the MONAI regressor model with a convolution layer to preserve the temporal relation between projections and the corresponding breathing amplitude. Additionally, this leads to a favorable reduction of the number of trainable parameters to 41,9 million for our current model.

To train the model we used as a first instance the CBCT Animal Motion Imaging Study (CAMIS) dataset. In this set, CBCT were taken from animals under general anesthesia using mechanical ventilation and under enforced breath-hold. This results in very regular RSS for the ventilated scans. The amount of data at our disposal was: 21 data sets comprising about 2000 projections and 20 breath-hold scans with about 900 projections each. In order to increase the variation in the training set, we divided the full CBCT scans into batches of subsets of projections. The data points (both RSS and projections) were normalized between 0 and 1 according to the 2% percentile of the complete set. To compensate for systematic baseline drifts in the actual RSS, we applied linear regression to correct for the estimated slope and offset. Doing so we lose information about the absolute breathing amplitude, which is anyhow strongly dependent on the placement of the marker block but preserves the normalized amplitude.

III. Results

We trained our model using the animal dataset. To train the model, we subdivided some of the CBCT into batches containing 128 projections along with the corresponding 128 points from the recorded RSS (Figure 3a). In this way, we increased the variance with respect to the angular range of the projections and phase shift of the RSS. After having trained the model, we evaluate it on complete sets of CBCT projection images separated from the training data (Figure 3b, Figure 3d). Note that the model can be applied to CBCT projection image data sets containing any arbitrary number of projections. In addition to the validation loss, we were also interested in the correlation between actual and predicted data points. To visualize the correlation we used scatter plots of the actual versus the predicted RSS. The closer the points to the identity line, the greater the correlation, (Figure 3c, Figure 3e). The Pearson correlation coefficient

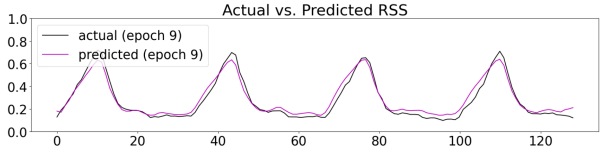
$$p = \frac{\mathrm{cov}(y, \hat{y})}{\mathrm{std}(y)\,\mathrm{std}(\hat{y})}$$

between the actual y and the predicted $\hat{y}$ RSS was applied as quantitative metric to compare results.

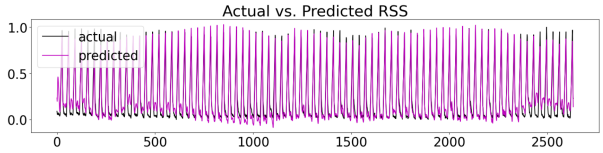
In a second training, we included batches of the enforced breath-hold CBCTs in the training set. We evaluated against the same data, i.e. on complete sets of 4D CBCT acquisitions of mechanically ventilated animals. The idea was to bring more variance into the training set and see in which way it would affect the evaluation performance (Table I).

Additionally, the model trained using breath-hold data was tested on enforced breath-hold CBCTs (Figure 4).

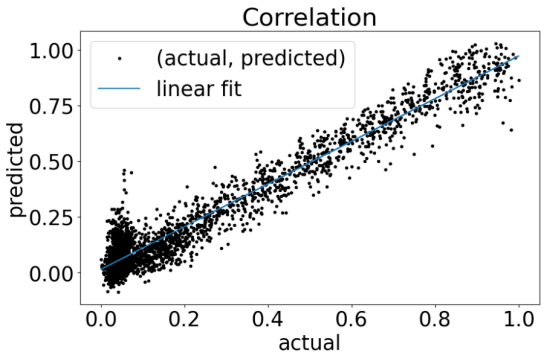
In order to perform a fair comparison with our in-house implementation of the Amsterdam Shroud (AS) algorithm, we base the comparison on the retrospectively calculated phase of the signals. The phase calculation is an in-house algorithm that finds the local maxima (end inhale) of the full signal and assigns the phases accordingly. The reason to compare phases is that the AS algorithm is not able to recover the amplitude of the signal (Figure 5a) but provides sufficient accurate peaks for the phase (Figure 5b) calculation. The retrospective calculated phase of the RPM signal serves as a ground truth referred to as the ”actual” phase. As presented in (Table II) the proposed method slightly outperforms the AS method in terms of Pearson Correlation for most cases. We examined the cases and found a slight phase shift of the extracted signal that needs further attention.



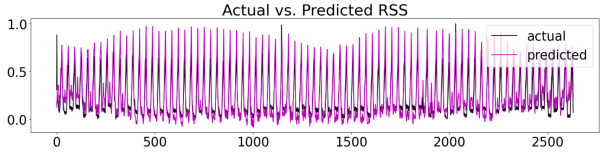
(a) Example of training data. The actual RSS in black and the predicted RSS in violet, over which we optimize.



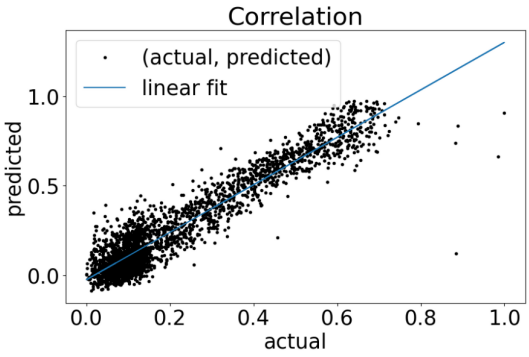
(b) Best evaluation data (w.r.t. the absolute difference between actual and predicted RSS) using a full CBCT projection image data set from the animal dataset.



(c) Correlation between actual and predicted RSS value as derived from the best evaluation data above.



(d) Worst evaluation data (w.r.t. the absolute difference between actual and predicted RSS) using a full CBCT projection image from the animal dataset.



(e) Correlation between actual and predicted RSS value as derived from the worst evaluation data above.

Fig. 3: a) An example for the training data. b-e) best and worst-case data evaluation and corresponding correlation plots.

TABLE I: Comparison of our model trained on two different sets. One set only contains ”free breathing” animal data while the other also contains ”breath-hold” animal data.

Patient	Free Breathing Model	Free Breathing / Breath-Hold Model
	Pearson correlation coefficient	Pearson correlation coefficient
A04	0.967	**0.977**
A05	0.943	**0.950**
A06	0.967	**0.973**
A07	0.948	**0.953**
A08	**0.977**	0.976
A09	0.944	**0.959**
A10	**0.981**	0.978
Average	0.961	**0.966**
	Validation loss	Validation loss
A04	**0.004**	0.018
A05	0.013	**0.012**
A06	**0.009**	0.012
A07	0.013	**0.007**
A08	**0.005**	0.009
A09	**0.011**	0.018
A10	**0.009**	0.010
Average	**0.009**	0.012

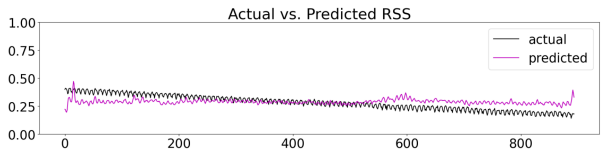


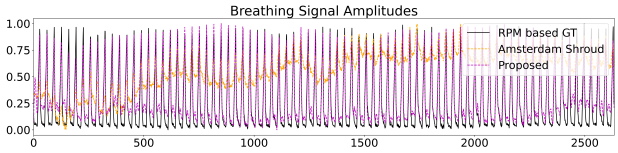
Fig. 4: Model evaluated on breath-hold data.

IV. Discussion

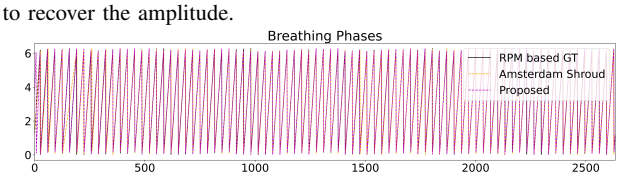
FROM Figure 3b and 3d we readily see that evaluating data from the animal dataset, i.e. on data with similar variance in phase, frequency, and amplitude yields acceptable results. In particular, the phase and frequency are matched accurately. We still observe variance concerning the matching of the amplitude. A good fitting on the normalized data will provide a good absolute amplitude fitting given we apply the inverse of the normalization process back. In both the best and worst cases the correlation between the actual and predicted RSS is linear, which comes to underline the good accuracy with respect to the frequency matching.

Looking in more detail at the performance of the model trained on free-breathing data versus the model trained on free-breathing plus breath-hold data (Table I), we remark that both give very similar results, the former having the advantage in the evaluation loss but the latter the one in the correlation between actual and predicted RSS. From these few data, it is not clear whether bringing variance in the training set has consistently a negative impact on evaluation, but this should nonetheless be looked after.

An advantage of our model as trained on breath-hold data (and free-breathing data), is that it systematically manages to correctly discern between breath-hold and free-breathing data (Figure 4). Here the proposed method is able to predict breath-hold signals, unlike the AS method which shows a strong bias towards finding a breathing rate in the expected range. The motivation is that this opens new applications such



(a) Comparison of RPM with the extracted breathing signals applying the AS and the proposed method. It can be seen that AS method fails to recover the amplitude.



(b) Retrospectively calculated phases show good agreement for all methods. In this case, only the AS method misses one breathing cycle around projection number 200.

Fig. 5: Comparison of the Amsterdam Shroud algorithm and our proposed model. The actual is given by the recorded RSS.

TABLE II: Pearson correlation coefficient of retrospectively calculated phases. Shown is the actual phase calculated based on the RPM signal versus the Amsterdam Shroud (AS) result and the result of the proposed method.

Patient	Pearson Correlation Actual vs. AS Phase	Pearson Correlation Actual vs. Proposed Phase
A04	0.899	**0.953**
A05	0.929	**0.935**
A06	**0.948**	0.882
A07	0.886	**0.944**
A08	0.867	**0.967**
A09	**0.948**	0.875
A10	0.896	**0.960**
Average	0.910	**0.931**

as monitoring of breath-hold compliance, automatic image quality estimation, or reconstruction method determination.

In Figure 5 and Table II are shown comparisons of results obtained using our model against results obtained from the Amsterdam shroud algorithm. We remark that the amplitude (Figure 5a), the phase shift (Figure 5b), and the correlation between actual and predicted RSS (Table 5) are better evaluated by our model.

The results need to be put in contrast with the goal we want to achieve. Evaluating the model trained on the animal dataset on human patient data would not yield concluding results with respect to frequency or amplitude fit. We would need to train our model on a large amount of patient data with a large variance in the RSS to be able to conclude the robustness of the model on patients.

Along with training with a more diverse and human patient-oriented dataset, we could also modify the network. Augmentation techniques (e.g. noise or virtual frame rate changes by skipping projections) can be applied to further increase the variance in the training set.

V. Conclusion

We presented our preliminary work concerning the automated extraction of the RSS given a full CBCT projection image data set via the use of a CNN model. We trained and tested our model on an animal (dog) dataset and obtained encouraging results w.r.t. phase, frequency, and normalized amplitude extraction. Concluding on the robustness of the presented method would require training and evaluation of a large amount of patient data. Along selecting a more diverse dataset to train our model we could change its architecture as well to allow for more precision in the amplitude extraction. The following steps would include the use of the predicted RSS for the reconstruction of 4D CBCT.

VI. Acknowledgment

The authors thank Prof. Michael Kent from the University of California, Davis, School of Veterinary Medicine for providing the in vivo canine CBCT data.

References

[1] L. Zijp, J.-J. Sonke, and M. van Herk, "Extraction of the respiratory signal from sequential thorax cone-beam x-ray images," in *International conference on the use of computers in radiation therapy*, 2004, pp. 507–509.

[2] I. Vergalasova, J. Cai, and F. Yin, "A novel technique for markerless, self-sorted 4D-CBCT: Feasibility study," *Medical Physics*, vol. 39, no. 3, pp. 1442–1451, 2012.

[3] A. Kavanagh, P. M. Evans, V. N. Hansen, and S. Webb, "Obtaining breathing patterns from any sequential thoracic x-ray image set," *Physics in Medicine and Biology*, vol. 54, no. 16, pp. 4879–4888, 2009.

[4] S. H. Bartling, J. Dinkel, W. Stiller, M. Grasruck, I. Madisch, H.-U. Kauczor, W. Semmler, R. Gupta, and F. Kiessling, "Intrinsic respiratory gating in small-animal CT," *European Radiology*, vol. 18, no. 7, p. 1375, 2008.

[5] H. Yan, X. Wang, W. Yin, T. Pan, M. Ahmad, X. Mou, L. Cerviño, X. Jia, and S. B. Jiang, "Extracting respiratory signals from thoracic cone beam CT projections," *Physics in Medicine and Biology*, vol. 58, no. 5, p. 1447, 2013.

[6] P. Tsai, G. Yan, C. Liu, Y. Hung, D. L. Kahler, J. Park, N. Potter, J. G. Li, and B. Lu, "Tumor phase recognition using cone-beam computed tomography projections and external surrogate information," *Medical Physics*, vol. 47, no. 10, pp. 5077–5089, 2020.

[7] R. Martin, A. Rubinstein, M. Ahmad, L. Court, and T. Pan, "Evaluation of intrinsic respiratory signal determination methods for 4d cbct adapted for mice," *Medical Physics*, vol. 42, no. 1, pp. 154–164, 2015.

[8] D. Edmunds, G. Sharp, and B. Winey, "Automatic diaphragm segmentation for real-time lung tumor tracking on cone-beam CT projections: a convolutional neural network approach," *Biomedical Physics & Engineering Express*, vol. 5, no. 3, p. 035005, 2019.

[9] M. Consortium, "Monai: Medical open network for ai," Mar. 2020. [Online]. Available: https://doi.org/10.5281/zenodo.5728262

Deep learning enabled wide-coverage high-resolution cardiac CT

Tzu-Cheng Lee[1]*, Jian Zhou[1], John Schuzer[1], Masakazu Matsuura[2], Takuya Nemoto[2], Hiroki Taguchi[2], Zhou Yu[1], Liang Cai[1]

***Abstract*— Wide-coverage detector CT and ultra-high-resolution (UHR) detector CT are two important features for current cardiac imaging modalities. The former one enables the scanner to finish a whole heart image scan in one bed position; the latter one gives superior resolution in fine structures such as stenoses, calcifications, implemented stents, and small vessel boundaries. However, no commercially available scanner has both these features in one scanner as of today. Herein, we propose to use existing UHR-CT data to train a super resolution (SR) neural network and apply the network in a wide-coverage detector CT system. The purpose of the network is to enhance the system resolution performance and reduce the noise while maintaining its wide-coverage feature without additional hardware changes. Thirteen UHR-CT patient datasets and their simulated-normal-resolution pairs were used for training a 3D residual-block U-Net. The modulation transfer function (MTF) measured from Catphan phantom scans showed the proposed super-resolution aided deep learning-based reconstruction (SR-DLR) improved the MTF resolution by relative ~30% and ~10% as compared to filtered-back projection and model-based iterative reconstruction approaches. In real patient cases, the SR-DLR images show better noise texture and enhanced spatial resolution along with better aortic valve, stent, calcification, and soft tissue features as compared to other reconstruction approaches.**

***Index Terms*— super resolution, wide coverage detector, ultra-high resolution (UHR) CT, cardiac imaging**

I. Introduction

WIDE-coverage detector CT, such as the 16-cm coverage area of the Canon Aquilion ONE system, improves the ability to obtain high-quality images for routine cardiac and chest scans. It takes only one rotation to acquire a whole heart scan with less dose and great z-axis uniformity [1, 2]. Moreover, the superior time resolution from single bed position scans reduces possibility for motion artifact due to patient movement. Ultra-high resolution (UHR) CT, on the other hand, equipped with a finer size detector and smaller x-ray focal spot source, provides diagnostic images with two times the spatial resolution compared to current standard resolution CT (0.25 vs. 0.50 mm detector size at isocenter, Canon Aquilion Precision system for example). Several studies have shown improvements in tumor classification and staging [3, 4].

Ideally, a wide-coverage detector UHR-CT scanner is an advanced solution for cardiac imaging which offers better dose efficiency, spatial resolution and motion control at the same time. However, to the best of our knowledge, this ideal hardware system is currently not commercially available. Such a scanner may also have issues with image reconstruction, data transfer and processing times, or the inevitably higher cost. Therefore, an alternative approach for wide-coverage detector UHR-CT imaging is desirable.

Super-resolution (SR) technology aims at recovering high-resolution information from low-resolution images. Recent research has showed that deep convolution neural network-based (DCNN-based) SR approaches produce superior image quality along with processing speeds that compare with conventional methods [5, 6]. Nevertheless, none of references have shown the performance advantages of combining both wide-coverage and UHR-CT imaging on real clinical cases. In this work, we propose a super-resolution aided deep-learning based reconstruction (SR-DLR) framework for achieving near UHR image quality on a wide coverage detector CT system; with both UHR and wide coverage features without hardware modifications.

II. Method

The proposed super-resolution DCNN directly uses the high-resolution CT images acquired on our UHR CT scanner as the training target. In order to simulate lower resolution CT images for network training input, we perform data-domain downsampling and then the downsampled data are reconstructed to obtain the lower resolution or normal resolution (NR) input. Note that we do not recommend direct image-domain downsampling as it may introduce spatial interpolation artifacts and create potentially unrealistic CT images. Multiple dose levels from low to high are also considered in the simulations which allow the network to learn not only the resolution enhancement but also noise reduction in particular when the input dose is low. The optimized network is then applied on the real wide-coverage CT data for achieving the wide-coverage UHR CT-like image. The detailed workflow for training data preparation and network inference is shown in Figure 1. Currently the proposed DCNN is an image-domain super-resolution network but extension to data-domain super

This work was supported by Canon Medical Research USA, Inc and Canon Medical System Corp. [1]Tzu-Cheng Lee, Jian Zhou, John Schuzer, Zhou Yu and Liang Cai are with Canon Medical Research USA, Inc, IL 60061 USA (*Correspondent e-mail: elee@mru.medical.canon).

[2]Masakazu Matsuura, Takuya Nemoto, Hiroki Taguchi are with CT Development Department, Canon Medical Systems Corporation, Tochigi 324-8550, Japan.

resolution is straightforward and will be investigated in our future work.

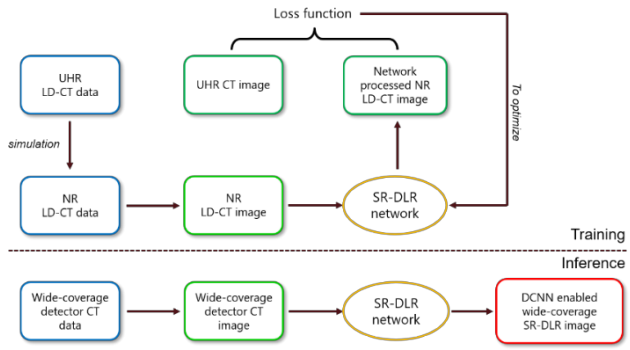


Figure 1 Training and inference processing flow of proposed DCNN enabled wide-coverage SR-DLR imaging.

A. *Data Description*

Thirteen high quality cardiac patient cases acquired from a UHR-CT system were used as the learning target. The acquisition energy and dose varied from 120 ~ 140 kVp and 150 ~ 190 mAs. Different reconstruction field of views and doses were applied and simulated for enriching data diversity. All target image dimensions reconstructed with model-based iterative reconstruction (MBIR) method were 1024x1024 pixels by 0.25 mm thickness. The acquired high-resolution data were first converted to the pre-log count domain, and then 2x2 data binning was applied to mimic the NR data. For more realistic data, additional white noise can be added at this step to compensate for the electronic noise. All simulated NR data were reconstructed with filtered-back projection (FBP) and standard ramp filter in the dimensions of 1024x1024 pixels and 0.25mm reconstruction pitch. The reconstructed SR-DLR images are compared with FBP and MBIR images. For quantitative comparison, the Catphan-600 phantom with different modules were used to evaluate the resolution improvement in terms of MTF (modulation transfer function) and line pairs (lp/cm) analysis. One true UHR-CT patient data (140kVp, 164mAs, 150mm FOV, 0.25mm reconstruction pitch) also used for comparison as shown in Figure 4. NR simulation process as stated in previous publication [7]. Another three real normal resolution (NR-CT) patient scans (*see* Table 1) were reconstructed to test the image quality through visual inspection. All images in result section reconstructed in default 512x512 pixels (SR-DLR in 1024x1024 pixels) with 240-mm field of view and 0.25 mm reconstruction pitch.

Table 1 Dose information for three representative patient cases

Patient	kVp	mAs	CTDI(mGy)	Remarks
A	100	63.3	2.5	Low-dose
B	120	159.5	9.4	Stent
C	120	206.3	14.3	High calcium

B. *Network Architecture, Training and Implementation*

The SR-DLR network adopts the U-NET structure with residual 3D convolutional blocks as basic building units, which enables deep structure without gradient vanishing and maintains proper receptive field for capturing large spatial features in the image (*see* Figure 2) [8, 9]. In network training we used small patches of size 32x128x128. Data argumentation such as patch flipping were also performed during the training. The training was mainly conducted using the TensorFlow-Keras (ver. 2.3) framework with Nvidia Titan RTX GPUs for acceleration. We chose the ADAM optimizer and the mean absolute error loss function. A total 200 epochs were run for effective convergence.

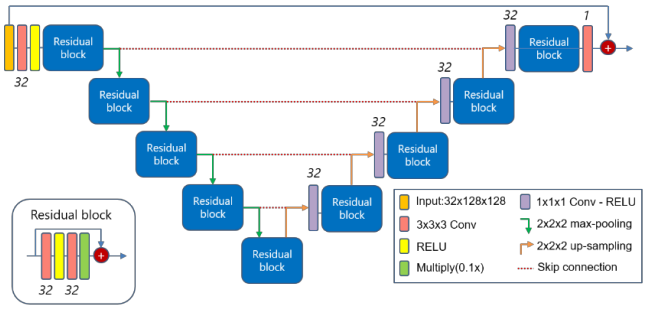


Figure 2 The proposed SR-DLR network which includes a 3D-UNET structure with residual blocks. The number beside the convolutional layer represents the number of filters.

III. RESULTS

A. *Quantitative Evaluation*

SR-DLR imaging shows superior 200mm FOV MTF performance as compared to FBP and MBIR reconstruction among all three materials. For low-contrast polystyrene (~-30HU), SR-DLR improves 10%-MTF to 1.04 lp/mm compares to 0.77 in FBP and 0.80 in MBIR. For mid-contrast Delrin (~330HU), 10%-MTF in SR-DLR is 1.15 lp/mm compares to 0.86 in FBP and 0.91 in MBIR. For high-contrast Teflon (~900HU), 10%-MTF in SR-DLR is 1.24 lp/mm compares to 0.86 in FBP and 1.11 in MBIR. (*see* Figure 3).

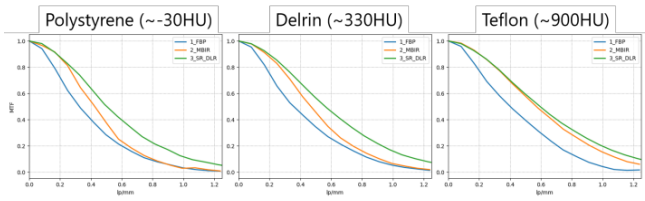


Figure 3 Modulation transfer function measurement as a function of line pair per millimeter with three different representative materials (polystyrene, Delrin and Teflon) in standard CATPHAN phantom (CTP404 module). Three different reconstruction methods (FBP in blue, MBIR in orange and proposed SR-DLR in green) listed from 0 to 1.25 lp/mm for comparison.

B. *Comparison with True UHR-CT Images*

Figure 4 compares the SR-DLR image to the closest-ground-truth image (e.g. true UHR-CT MBIR image) so that the resolution performance can be examined. The simulated-NR SR-DLR phantom image has better bar resolution compared to

the FBP image (~9 lp/cm vs. ~8 lp/cm). The image contrast as well as the bar intensity of SR-DLR is also closer to the ground truth image compared to the FBP reconstruction. The simulated NR SR-DLR patient image shows better resolution in stent and higher contrast in soft tissue compares to the FBP image. No artificial feature created when we compare NR SR-DLR to the true UHR-CT patient image.

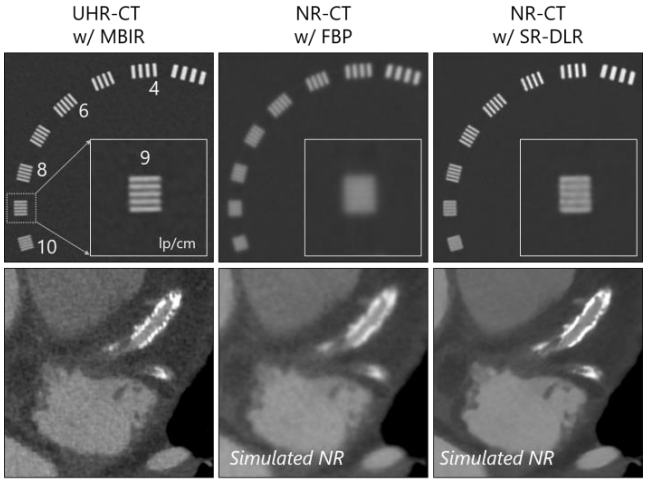


Figure 4 Resolution phantom module (CTP528 module) and patient examples. *Upper row*: two reconstruction methods with normal resolution (NR) data to the true UHR-CT MBIR image. The 9 lp/cm contrast bars are zoomed in the bottom-right small image. Window level and width: 1000 and 3000 HU. *Lower row*: two simulated NR data for FBP and SR-DLR reconstructions to the true UHR-CT MBIR patient image. Window level and width: 350 and 1500 HU

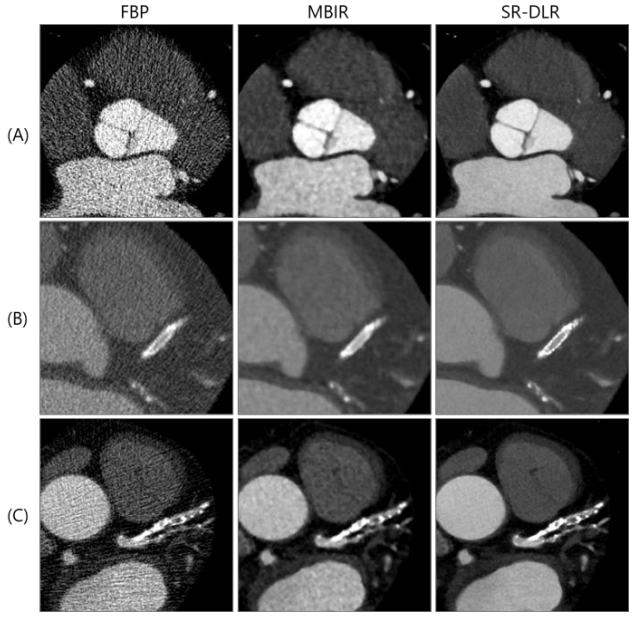


Figure 5 Three representative patient cases (case A: aortic valve; case B: stent; case C: calcium) with three different reconstruction methods. Window level and width for case A and C: 300 and 800 HU; for case B: 350 and 1500 HU.

C. Compare SR-DLR Image to Other Reconstructions

In Figure 5, we are able to see the noise and resolution differences when comparing SR-DLR images to FBP and MBIR reconstructions. Patient (A) images shows that the SR-DLR image has better resolution for the aortic valve and better contrast of small vessels. Patient (B) images further show cleaner microstructure of the implanted stent. Patient case (C) shows the SR-DLR image has sharper calcification contrast and boundaries compared to other two reconstruction methods. In addition, SR-DLR has lowest noise and sharpest soft tissue boundaries among these three methods.

IV. Conclusions

Testing on quantitative phantom data and clinical patient data shows that SR-DLR is a promising approach for enhancing resolution and suppressing noise. In summary, the proposed method has advantages over prior approaches: compared to current wide-coverage detector CT imaging, the proposed method improves resolution and noise performance; this improvement is gained from the UHR-CT trained network with finer reconstructed pixel sizes. Compared to UHR-CT system imaging, this method has wider-detector coverage data as the input, which benefits dose efficiency, image uniformity, and motion management. The larger acquisition detector pixel size from wide-coverage CT also benefits noise performance. Compared to a hypothetical wide-coverage UHR-CT system, the SR-DLR method has substantially lower cost and much less image processing complexity from both hardware and software perspectives.

Acknowledgment

Authors would like to thank Dr. Marcus Chen, M.D. and his group at Cardiovascular CT Program of NHLBI, NIH for data collection and their insightful clinical feedbacks.

References

[1] Lewis M, *et. al*, "Selecting a CT scanner for cardiac imaging: the heart of the matter" *BJR* 2016 (89):376

[2] Annoni A, *et. al*, "CT angiography prior to TAVI procedure using third-generation scanner with wide volume coverage: feasibility, renal safety and diagnostic accuracy for coronary tree" *BJR* 2018 (91):196

[3] Yanagawa M, *et. al*, "Subjective and objective comparisons of image quality between ultra-high-resolution CT and conventional area detector CT in phantoms and cadaveric human lungs" *European Radiology* 2018 (28):5060

[4] Hata A, *et. al*, "Effect of matrix size on the image quality of ultra-high-resolution CT of the lung" *Acad Radiol.* 2018 (25):869

[5] Yu, H. *et. al*,"Computed tomography super-resolution using convolutional neural networks," *2017 IEEE International Conference on Image Processing (ICIP)* 2017

[6] Umehara, K. *et. al*, "Application of super-resolution convolutional neural network for enhancing image resolution in chest CT," *J Digit Imaging* 2018; 31(4): 441-450

[7] Hernandez, A. et. al, "Validation of synthesized normal-resolution image data generated from high-resolution acquisitions on a commercial CT scanner" *Medical Physics*, 2020; 47(10):4776

[8] Lim, B. et. al, "Enhanced deep residual networks for single image super-resolution" 2017, arXiv:1707.02921

[9] Milletari, F. et. al, "V-Net: fully convolutional neural networks for volumetric medical image segmentation," 2016, arXiv:1606.04797

Preliminary study on image reconstruction for limited-angular-range dual-energy CT using two-orthogonal, overlapping arcs

Buxin Chen, Zheng Zhang, Dan Xia, Emil Y. Sidky, and Xiaochuan Pan

***Abstract*—Dual-energy CT (DECT) of limited-angular ranges (LARs) collects data from angular ranges smaller than π for low- and high-kVp scans, and thus may potentially be exploited for reducing scanning time and radiation dose and for avoiding collision between the imaged object and the moving gantry of the scanner. Image artifacts resulting from beam hardening (BH) and limited-angular range (LAR) can be suppressed by using the data-domain decomposition and the directional-total-variation (DTV) algorithm for image reconstruction. In this work, we investigate two-orthogonal-arc (TOA) scanning configuration with overlapping arcs for collecting LAR DECT data, in an effort to reduce LAR artifacts and improve quantitative accuracy of estimated physical quantities. The TOA configuration consists of two arcs, of equal LAR, whose centers are positioned 90° apart, and is designed to reduce the ill-conditionedness of the imaging system matrix. The data are decomposed into basis sinograms, from which basis images are reconstructed using the DTV algorithm. Visual inspection of the monochromatic images and quantitative estimation of the effective atomic numbers suggest that the TOA configuration, as compared to the single-arc (SA) configuration of the same total angular range, can help reduce remaining LAR artifacts and bias in the estimated atomic number relative to the reference values from the full-angular-range data of 360°.**

***Index Terms*—dual-energy CT, limited-angular range, two-orthogonal arc, directional total variation**

I. Introduction

Current dual-energy CT (DECT) typically collect data, of either low- or high-kVp spectrum, in a full-scan, or at least short-scan, rotation [1], [2]. Images are then reconstructed often by use of conventional algorithms, such as FBP, from kVp sinogram directly or decomposed basis sinogram. One-step algorithms have also been investigated for reconstructing basis images directly from full-scan data [3]. DECT of limited-angular ranges (LARs) [4], [5] collects data from angular ranges smaller than π for low- and high-kVp scans, and thus may be potentially useful for reducing scanning time and radiation dose and for avoiding collision between the imaged object and the moving gantry of the scanner. In LAR DECT, LAR artifacts in images are usually more dominant than other ones, such as beam hardening (BH), and are thus subject to more focus and effort in artifacts correction [6], [7]. Simultaneous correction of LAR and BH artifacts may help reduce image artifacts and improve quantitative accuracy in LAR DECT, especially for extremely small angular ranges. In this work, we aim to improve simultaneous correction for LAR and BH artifacts, and thus quantitative reconstruction accuracy, by investigating the two-orthogonal-arc (TOA) configuration for DECT with overlapping rays, which can help alleviate the ill-conditionedness in the system matrix.

Numerical studies are carried out with a suitcase phantom containing different materials. LAR scanning configurations are set up with overlapping rays from low- and high-kVp scans, i.e., the scanning arcs of the low- and high-kVp scans are identical. For either kVp scan, the scanning arcs consist of two LAR arcs, whose centers are separated by 90°. Data are first decomposed into basis sinograms, and basis images are then reconstructed by use of the directional-total-variation (DTV) algorithm. This primal-dual algorithm has been developed for solving a DTV-constrained, data-ℓ_2-minimization problem [4], [6]. Basis images are combined into monochromatic images for visual assessment, and then used for computing effective atomic numbers of different materials in the phantom. Results are compared with those from the single-arc (SA) configuration with the same total angular ranges.

B. Chen, Z. Zhang, E. Y. Sidky, and D. Xia are with the Department of Radiology, The University of Chicago, Chicago, IL 60637, USA.

X. Pan is with the Departments of Radiology & Radiation and Cellular Oncology, The University of Chicago, Chicago, IL 60637, USA.

II. Materials and methods

A. Two-orthogonal-arc configuration

A 2D circular fan-beam geometry is considered, as shown in Fig. 1, while the approach can readily be extended to non-circular fan-beam geometry and to circular/non-circular 3D cone-beam geometry. The source-to-rotation and source-to-detector distances are 100 cm and 150 cm, respectively, while the linear detector is 32 cm in length with 512 bins. The scanning arcs of low- and high-kVp spectra are identical, thus generating overlapping rays suitable for data-domain decomposition. With the TOA configuration, for each of the low- and high-kVp scan, there are two scanning arcs, separated by 90° from center to center, covering LARs of α_x and α_y (it is assumed that the two arcs are symmetric relative to the x and y axes in the image array, respectively). Given a total angular range $\alpha_\tau = \alpha_x + \alpha_y$, there are many different ways of distributing between α_x and α_y. In this work, we focus on the case with equal range, $\alpha_\tau/2$, i.e., $\alpha_x = \alpha_y = \alpha_\tau/2$, because studies have suggested that TOA with equal-range arcs might perform better than other distributions [8]. Data are also collected using a SA configuration with $\alpha_y = \alpha_\tau$ and $\alpha_x = 0$. A set of total angular ranges is studied with $14^\circ \leq \alpha_\tau \leq 180^\circ$,

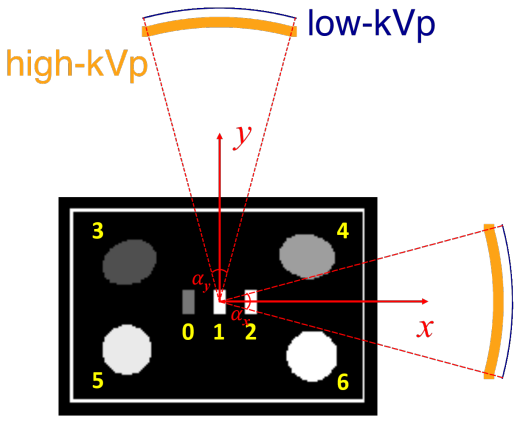


Fig. 1: The TOA scanning configuration with overlapping LAR arcs of low- and high-kVp spectra, for collecting dual-energy data from a suitcase phantom, containing different materials including C, Al, Ca, water, ANFO, teflon, and PVC, shown as ROIs 0-6 respectively.

with a fixed angular interval of $0.25°$ per view. In this study, we focus on the three smallest LARs of $\alpha_\tau = 20°$, and $30°$, which are the most challenging. The TOA configuration for LAR DECT can be readily implemented with existing DECT technologies, such as dual-source DECT, sandwiched detector, fast-kVp-switching X-ray tubes, and also sequential scans.

B. Dual-energy data

A digital suitcase phantom is used containing different materials that are challenging to be differentiated and identified with conventional CT, including circular and elliptical ROIs filled with water, ammonium nitrate and fuel oil (ANFO), teflon, and PVC, as well as three bar-shaped ROIs of single-element materials, C, Al, and Ca. The discrete phantom is set up on an 175×256 image array with 0.7-mm square pixels, while each pixel is labeled with a material type and thus associated with the material's linear attenuation coefficients from the NIST database. Dual-energy data are generated using a non-linear data model incorporating the spectral integral, such as Eq. (1) in Ref. [6], with simulated low- and high-kVp spectra at 80 kVp and 140 kVp using the TASMIC model [9]. Both noiseless and noisy data are generated, with Poisson noise added to the noiseless data, corresponding to 4.5×10^5 and 6.75×10^5 noise equivalent quanta (NEQ) per ray in the air scan for $\alpha_\tau = 30°$ and $20°$, respectively, such that the total NEQ is constant for different LARs.

Once data, namely, low- and high-kVp sinograms, are generated, they are decomposed into basis sinograms for BH correction using a well recognized data-domain decomposition method [10]. An interaction-based basis decomposition is used, where photoelectric effect (PE) and Compton scattering (KN) are the two bases with $1/E^3$ and the Klein-Nishina formula as their spectral responses. The low- and high-kVp spectra are assumed to be known exactly, therefore minimizing the impact of spectrum mismatch in BH correction.

C. Image reconstruction

With decomposed basis sinogram, $\mathbf{l}_{\rm PE}$ and $\mathbf{l}_{\rm KN}$ for photoelectric effect and Compton scattering, the reconstructions of basis images, $\mathbf{b}_{\rm PE}$ and $\mathbf{b}_{\rm KN}$, can be formulated into two separate convex optimization problems as

$$\begin{aligned} \mathbf{b}_k^\star = \underset{\mathbf{b}_k}{\operatorname{argmin}} \ & \frac{1}{2} \parallel \mathbf{l}_k - \mathcal{A}\,\mathbf{b}_k \parallel_2^2 \\ \text{s.t. } & ||\mathcal{D}_x \mathbf{b}_k||_1 \le t_{kx}, \ ||\mathcal{D}_y \mathbf{b}_k||_1 \le t_{ky}, \end{aligned} \tag{1}$$

where k = PE or KN; $\mathcal{A}$ is the discrete X-ray transform (DXT) of the TOA configuration; and $||\mathcal{D}_x \mathbf{b}_k||_1$ and $||\mathcal{D}_y \mathbf{b}_k||_1$ are ℓ_1 norms of the image partial derivatives along the x and y axes, respectively, also referred to as the image's DTVs. The separate DTV constraints along the image array's orthogonal directions have been shown to reduce effectively the directional artifacts in CT images reconstructed from LAR data. The two convex optimization problems in Eq. (1) for the two bases are solved by the DTV algorithm [4], which is based on a general primal-dual algorithm [11].

D. Evaluation

From basis images reconstructed, monochromatic images are formed as a linear combination of basis images and visually assessed for artifact reduction. Further, the effective atomic number, z, can be estimated from the interaction-based basis images, due to different orders of z-dependence for the photoelectric effect and Compton scattering interactions [1], [12]. In particular, a linear relationship in the log-log domain can be assumed between z and the ratio of the basis image values, where the slope and intercept can be fitted using the single-element materials in the phantom with known z values.

III. Results

We show in Fig. 2 monochromatic images at 40 keV from noiseless LAR data collected over SAs of $\alpha_\tau = 20°$ and $30°$, and over TOAs of equal total angular ranges. No visual LAR artifacts can be observed in the images. Quantitatively, the monochromatic images from the TOA scan are closer to the reference image from full-scan range data of $360°$ than those from the SA scan in terms of the normalized root-mean-square-error (nRMSE) (nRMSE = 4.88×10^{-5} and 1.05×10^{-4} for $30°$ and $20°$ over SA, respectively, and $5.29\times^{-6}$ and 6.76×10^{-6} over TOA.)

We show the same results with noisy data in Fig. 3. Horizontal shading artifacts can be observed in the images from data collected over SAs of the extremely small LARs under investigation. The circular and elliptical disks are distorted, and the horizontal edges of the suitcases are difficult to recover, since they are mostly parallel to the SAs and are characterized as "invisible boundaries" [13]. In images from data collected over TOAs, there is no significant shape distortion to the circular and elliptical disks, while the edges of the suitcase are recovered.

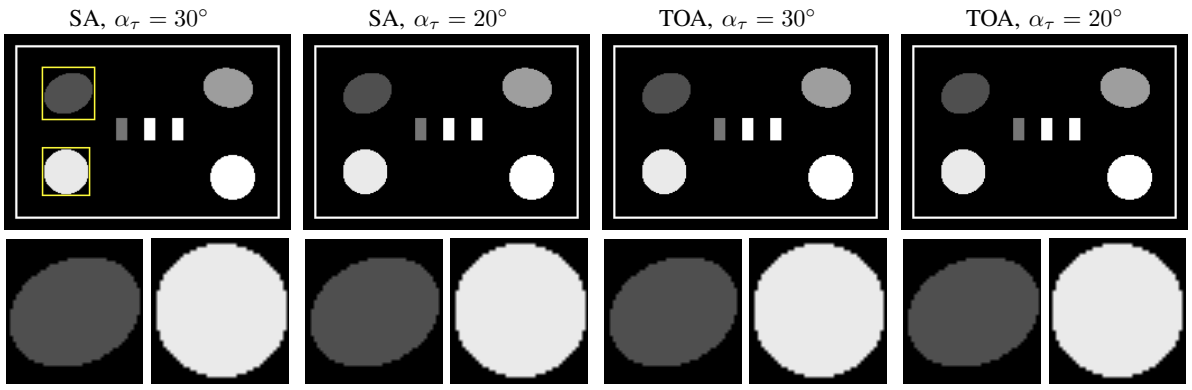


Fig. 2: Monochromatic images (top row) of the suitcase phantom at 40 keV obtained from noiseless data over SAs (columns 1 & 2) of $\alpha_\tau = 30°$ and $20°$, and TOAs (columns 3 & 4) of the same total angular range by use of the DTV algorithm, along with their respective zoomed-in views (bottom row). The zoomed-in regions are enclosed by the rectangular boxes including ROIs 3 and 5, as depicted in the top-left image. Display window: [0.1, 0.65] cm^{-1}.

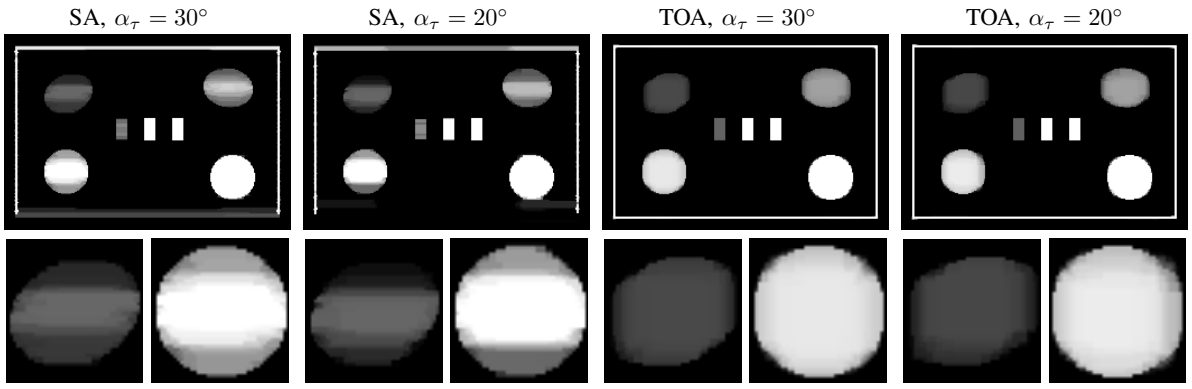


Fig. 3: Monochromatic images (top row) of the suitcase phantom at 40 keV obtained from noisy data over SAs (columns 1 & 2) of $\alpha_\tau = 30°$ and $20°$, and TOAs (columns 3 & 4) of the same total angular range by use of the DTV algorithm, along with their respective zoomed-in views (bottom row). The zoomed-in regions are enclosed by the rectangular boxes including ROIs 3 and 5, as depicted in the top-left image in Fig. 2. Display window: [0.1, 0.65] cm^{-1}.

From basis images of PE and KN, we can estimate the effective atomic numbers by assuming a linear relationship in the log-log domain between z and the ratio of the basis image values. The coefficients can be calibrated using single-element materials with known z number, such as C, Al, and Ca in the phantom. We show in Tables I and II the estimated z numbers of ROIs 3 (water) and 5 (teflon), respectively, from both noiseless and noisy data over SAs and TOAs. The reference values, 7.49 and 8.50, are calculated from the noiseless data collected over the full-scan range of $360°$. It can be observed that the TOA configuration can help reduce the bias in the quantitative estimation of effective atomic number, relative to the reference values, as compared to the SA configuration.

TABLE I: Estimated atomic number of ROI 3 (water) from noiseless and noisy data collected over SA and TOA of the same total angular coverage of $30°$ and $20°$. The reference value from full-scan $360°$ noiseless data is 7.49.

		$\alpha_\tau = 30°$	$\alpha_\tau = 20°$
noiseless	SA	7.49	7.49
	TOA	7.49	7.49
noisy	SA	7.97	7.36
	TOA	7.50	7.38

IV. Conclusion

In this work, we have proposed and investigated the two-orthogonal-arc scanning configuration with overlapping arcs, in combination with the DTV algorithm, for improving artifact correction and quantitative accuracy in DECT with LAR data. Numerical studies were carried out with a digital suitcase phantom, from which dual-energy data, with both SA and TOA configurations of the same total angular range, were generated using a non-linear data model. With overlapping arcs of low- and high-kVp scans, the kVp sinograms were decomposed into interaction-based basis sinograms by use of a data-domain decomposition method for BH correction.

TABLE II: Estimated atomic number of ROI 5 (teflon) from noiseless and noisy data collected over SA and TOA of the same total angular coverage of 30° and 20°. The reference value from full-scan 360° noiseless data is 8.50.

		$\alpha_\tau = 30^\circ$	$\alpha_\tau = 20^\circ$
noiseless	SA	8.50	8.50
	TOA	8.50	8.48
noisy	SA	9.31	9.04
	TOA	8.46	8.48

Basis images were then reconstructed from the basis sinograms by use of the DTV algorithm. Monochromatic images at 40 keV were combined for visual inspection and effective atomic numbers were estimated. Results suggest that the TOA configuration, as compared to the SA configuration of the same total angular range as low as 20°, can effectively reduce the remaining artifacts in monochromatic images obtained with the DTV algorithm, and also yield accurate estimation of effective atomic numbers relative to the reference values from the full-angular-range data of 360°.

Results of estimated effective atomic numbers for other ROIs and additional results with more LARs will be reported at the conference. Further investigation will focus on adding other physical factors, such as scatter, and using phantoms with different anatomies and structures.

ACKNOWLEDGMENT

This work was supported in part by NIH Grants Nos. R01-EB026282, R01-EB023968, and 1R21-CA263660-01A1. The contents of this article are solely the responsibility of the authors and do not necessarily represent the official views of the National Institutes of Health.

REFERENCES

[1] R. E. Alvarez and A. Macovski, "Energy-selective reconstructions in X-ray computerised tomography," *Phys. Med. Biol.*, vol. 21, no. 5, pp. 733–744, 1976.

[2] T. G. Flohr, C. H. McCollough, H. Bruder, M. Petersilka, K. Gruber, C. Sü, M. Grasruck, K. Stierstorfer, B. Krauss, R. Raupach, A. N. Primak, A. Küttner, S. Achenbach, C. Becker, A. Kopp, and B. M. Ohnesorge, "First performance evaluation of a dual-source CT (DSCT) system," *Eur. Radiol.*, vol. 16, no. 2, pp. 256–268, Dec. 2005.

[3] B. Chen, Z. Zhang, D. Xia, E. Y. Sidky, and X. Pan, "Non-convex primal-dual algorithm for image reconstruction in spectral CT," *Comput. Med. Imaging Graph.*, vol. 87, p. 101821, 2021.

[4] Z. Zhang, B. Chen, D. Xia, E. Y. Sidky, and X. Pan, "Directional-TV algorithm for image reconstruction from limited-angular-range data," *Med. Image Anal.*, vol. 70, p. 102030, 2021.

[5] H. Zhang and Y. Xing, "Reconstruction of limited-angle dual-energy CT using mutual learning and cross-estimation (mlce)," in *Proc. SPIE Med. Imag.: Phys. Med. Imag.*, vol. 9783. International Society for Optics and Photonics, 2016, p. 978344.

[6] B. Chen, Z. Zhang, D. Xia, E. Y. Sidky, and X. Pan, "Dual-energy ct imaging with limited-angular-range data," *Phys. Med. Biol.*, vol. 66, no. 18, p. 185020, 2021.

[7] ——, "Dual-energy ct imaging over non-overlapping, orthogonal arcs of limited-angular ranges," *Journal of X-ray Science and Technology*, no. Preprint, pp. 1–11, 2021.

[8] Z. Zhang, B. Chen, D. Xia, E. Y. Sidky, and X. Pan, "Image reconstruction from data over two orthogonal arcs of limited-angular ranges," *Med. Phys.*, vol. preprint, pp. 1–13, 2022.

[9] A. M. Hernandez and J. M. Boone, "Tungsten anode spectral model using interpolating cubic splines: Unfiltered X-ray spectra from 20 kv to 640 kv," *Med. Phys.*, vol. 41, no. 4, p. 042101, 2014.

[10] Y. Zou and M. D. Silver, "Analysis of fast kv-switching in dual energy CT using a pre-reconstruction decomposition technique," in *Proc. SPIE Med. Imag.: Phys. Med. Imag.*, vol. 6913, 2008, p. 691313.

[11] E. Y. Sidky, J. H. Jorgensen, and X. Pan, "Convex optimization problem prototyping for image reconstruction in computed tomography with the Chambolle-Pock algorithm," *Phys. Med. Biol.*, vol. 57, no. 10, pp. 3065–3091, 2012.

[12] Z. Ying, R. Naidu, and C. R. Crawford, "Dual energy computed tomography for explosive detection," *J. X-Ray Sci. Technol.*, vol. 14, no. 4, pp. 235–256, 2006.

[13] E. T. Quinto, "Artifacts and visible singularities in limited data X-ray tomography," *Sens. Imaging*, vol. 18, no. 1, p. 9, 2017.

Correcting spurious signal using an automated Deep Learning based reconstruction workflow

Matthew Andrew, Andriy Andreyev, Faguo Yang and Lars Omlor

***Abstract*—A technique for the automated training of deep-learning based tomographic reconstruction networks is shown and benchmarked using sintered ceramic and Ceramic Matrix Composite samples. Networks were trained to remove both image noise and sparse sampling artefacts, and the relative signature of both of these effects is discussed. Images reconstructed with deep learning show a significant reduction in both noise and artefact levels, even with up to 10 times fewer projections than in the original data. This technique has the potentially of greatly improving image quality and throughput, particularly for challenging high resolution imaging applications.**

***Index Terms*— Computed Tomography, Image Reconstruction, Deep Learning, AutoML.**

I. Introduction

ONE of the principal challenges for high resolution X-ray tomography for materials analysis are those arising from the presence of spurious reconstructed X-ray signal. While these are often termed as "noise", both in the technical X-ray and application community, in reality the spurious signal arises from a range of sources, principal among them being true Poisson (or "shot") noise, and sampling artefacts arising from an insufficient number of projections. This is particularly problematic in high (submicron) resolution imaging as the relatively limited flux available for laboratory source high resolution X-ray sources, coupled with the constraints placed on detector design means that image exposure times and consequently total acquisition time are typically long (in the order of tens of minutes to hours). While such a long acquisition time may be acceptable for certain scientific applications, many applications (such as *in situ* scanning, with a requirement for high absolute temporal resolution, or semiconductor failure analysis, with a requirement for rapid time-to-result) require faster acquisition times. To achieve this, users will often either reduce exposure time or total projection number, in turn degrading reconstructed image quality. Other applications require are extremely noise sensitive (such as distinguishing materials of extremely similar effective attenuation), while others are subject to extremely challenging experimental design (such as high resolution interior tomography), meaning high quality image acquisition may take many hours and be impractical.

The field of study broadly known as "AI" or "machine learning" has broadly revolutionized fields from stock market analysis to weather prediction, however their application to X-ray imaging, particularly high-resolution X-ray imaging for materials characterization is still in its infancy. This is largely because the range of samples, geometries and conditions that samples are imaged in, coupled with the particular constraints provided when imaging with high resolution X-ray sources, mean that it is extremely challenging to generalize single deep learning models (whether for classification or image recovery) across all samples. As expertise in deep learning applications and the details of their integration with sample workflows is extremely limited, this poses a significant barrier to adoption for deep learning based reconstruction and/or image recovery techniques. This has led to a range of application specific point solutions for image improvement, segmentation or analysis [1], [2], [3], however the development of high quality general purpose solutions have historically been a significant challenge.

In this work we use a commercially available reconstruction product (Zeiss DeepRecon, available from Carl Zeiss XRM, Dublin, CA) which offers a completely automated workflow for training new reconstruction networks from tomographic X-ray datasets [4]. The automation of network training circumvents to a great degree the issue associated with network generalizability, as it provides a robust procedure for new networks to be generated when a novel sample or system condition is encountered. This technique is benchmarked its performance across a range of a range of representative materials samples. This includes a synthetic rock sample which was used to separate the relative impact on image quality of sampling artefacts and true (random) image noise due to signal detection statistics, showing that high quality image recovery can only be achieved if both are considered. Samples showing the limitations of such an automated technique are also discussed.

II. Methods

Spurious signal in the volumetric domain can occur from a range of sources, including X-ray scatter (particularly when it is differentiated through a sample), poorly handled spectral effects, X-ray reflections and photon starvation. Particularly prominent in high resolution X-ray imaging for materials analysis is the combination of pure Poisson ("shot") noise (associated with the inherent randomness of X-ray detection)

The authors are with Carl Zeiss X-Ray Microscopy, Dublin, CA 94588 USA (e-mail: matthew.andrew@zeiss.com).

and the sampling artefacts (associated with imperfect data sampling in the frequency domain). It should be noted that step-and-shoot acquisition mode at high resolution imaging creates constraints particularly likely to exacerbate the presence of sampling artefacts in the reconstructed volume during practical user-led imaging scenarios. For example, the desire to minimize sample-motion blur artefacts, coupled with lengthy detector readout times, can lead to long projection-to-projection sample stage rotation time overheads. This can limit the effectiveness of trading off projection number with per-projection exposure time. The desire to reduce total acquisition time & dose remains, however, often leading users to reduce projection number and thereby introduce sampling artefacts.

For the purposes of this study, we consider the reconstructed image function as D as a combination of a true object structure function S, a noise function N and an artefact function A in the three spatial dimensions x, y, z:

$$D(x,y,z) = S(x,y,z) + N + A(S(x',y',z'))$$

It should be noted that while in theory the noise function N can be viewed as theoretically spatially white in the projection domain, it may lose significant contribution of high frequency information during the back projection process. The artefact function A at a position x, y, z is function of the structure elsewhere in the volume (denoted by x', y', z'). As both N and A must be removed for high quality image recovery, autocorrelative denoising techniques (e.g. [5]) which assume a spatial white signature in the spurious signal to be removed perform poorly. While modifications of this technique have been proposed (e.g. [6]) the extreme correlation present in the structure of spurious tomographic signal still makes such an approach ineffective.

In DeepRecon workflow image recovery networks were trained using a modification of noise-to-noise [7] techniques, with loss functions, network structures and data augmentation tailored to 3D tomography, modified to ensure rapid convergence and high performance even with early training stopping. Network structures were adapted from the UNet architecture [8]. Images used for training were constructed such that noise and sampling artefacts were uncorrelated between input and target sets. This forces the networks to learn to remove both the noise component (N) and the artefact component (A) from spurious reconstructed signal. Data was dynamically augmented for variations in intensity and orientation during training, improving model robustness.

Image training is performed in a Python-based TensorFlow environment, after which network weights were transferred into an .onnx format and inference performed using NVIDIA TensorRT runtime. An already trained model can also be applied to any other structurally similar sample if the acquisition conditions still match the conditions for which the model has been trained for. A range of experimental samples were characterized, comparing the results from traditional reconstruction (Feldkamp-Davis-Kress (FDK) reconstruction [9]) with those reconstructed using Deep learning based image recovery. All image data has been acquired using a Zeiss Xradia Versa 620 X-ray Microscope (Carl Zeiss XRM, Dublin, CA).

First a synthetic rock (sintered ceramic) sample is imaged twice, once when only the noise function is uncorrelated between the training and target images, and once when both noise and artefact functions are uncorrelated. As the structure function is the same between datasets, this allows for the noise and artefact functions to be independently identified. The sample was imaged with only 801 projections, leading to an extremely high level of spurious image signal. While this is not a representative application dataset, it is extremely informative when investigating the difference between the noise function N and the artefact function A.

A more typical sample application of a ceramic-matrix-composite (CMC) was then imaged using a 3,001 projection baseline acquisition. A series of volumes were reconstructed by sub-sampling the complete projection set by a range of sampling factors (1X, 4X, 8X, 10X), specifically training reconstruction networks for each sampling factor. The results are then analyzed visually and quantitatively to assess relative performance at various subsampling scales.

III. Results and Discussions

Images of the reconstructed volume from the sintered ceramic sample are shown in figure 1. The traditional filtered backprojection reconstruction is shown in 1A, the result of removing noise and artefact or only the noise fields are shown in 1B, the noise field and the artifact field are shown in one C, and the loss convergence curve is shown in 1D. The combined noise & artefact fields can be found by either looking at a region of the image known to be free of internal structure, or by looking at the difference between the reconstructed images with these fields removed (using Deep Learning) and the traditional reconstruction. They can be individually differentiated by looking at the result of a reconstruction network trained to remove only true noise and one trained to remove both noise and sampling artefacts.

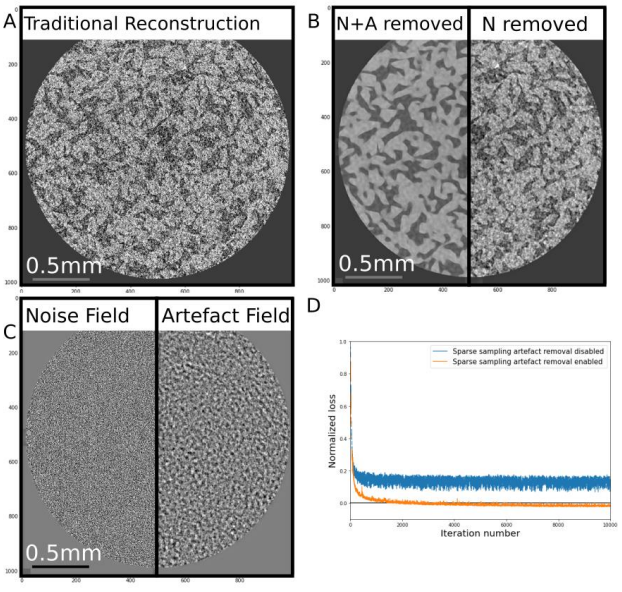


Figure 1: Deep Learning based image recovery on a sintered ceramic (synthetic rock) sample.

Correcting spurious signal using an automated Deep Learning based reconstruction workflow

The image reconstructed using traditional reconstruction techniques is extremely high in spurious signal. We can see that removing only the noise field (1B) results in an image which still has a high level of residual spurious signal. The noise and artefact fields (1C) show a strong contrast in structural autocorrelation. The noise field is normally distributed and spatially white (pixelwise random), whereas the artefact field shows a strong larger wavelength structure. A detailed frequency domain analysis of the noise and artefact structure is shown in figure 2. The true noise field is extremely white – there is very little bias across the spectrum to different energies. This is in contrast to the entire spurious signal (noise and artefact) field, which shows a significant bias to lower frequencies. The autocorrelated nature of the artefact field shows why denoising techniques designed based upon the assumption of non-correlated (spatially white) noise struggle when applied to tomographic data.

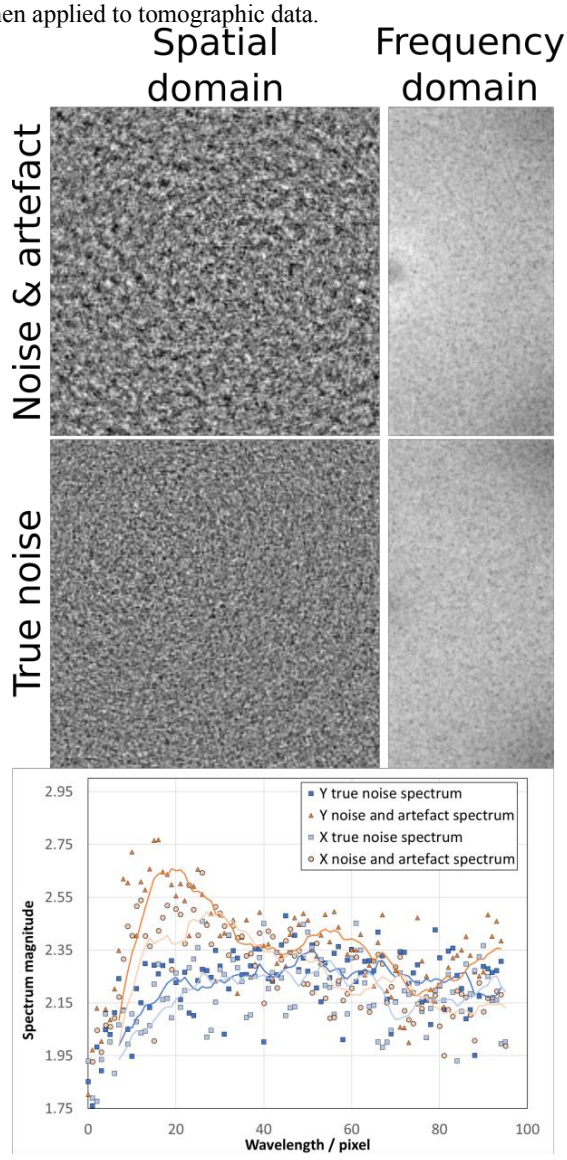


Figure 2: Artefact and noise analysis for sintered ceramic sample, contrasting the full spurious artefact signal with true noise.

Results from the ceramic matrix composite are shown below in figure 3. FDK reconstructions are shown for both the full projection set (3001 projections) and a 1:10 subsampled projection set (301 projections), and the deep learning recovered data using 301 projections is shown in comparison. The scale (standard deviation) of the total spurious signal for each subsampling fraction (1, 4, 8 and 10).

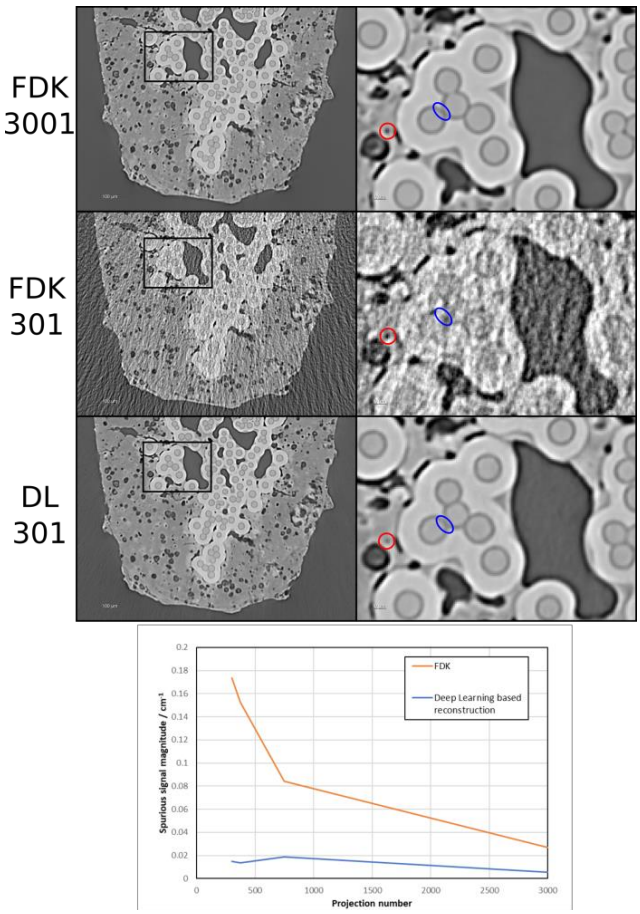


Figure 3: Ceramic matrix composite results. Small features are highlighted in red and blue color.

It should be noted that for many practical applications often noise and artefact trade-off against each-other in total scan time, at least to a certain extent, as they correspond to projection number and exposure time per projection. This means that other considerations (scanning overheads, photon starvation effects or data storage requirements) can govern user experimental design decisions.

As apparent acquisition time decreases the spurious signal level progressively increases when the data is reconstructed using FDK, whereas the increase is much more limited when data is reconstructed using deep learning. This is achieved without losing distinct small features, such as the small pore highlighted in red, or the interface between composite fibers, highlighted in blue.

The precise structure of the artefact field is highly dependent on the details of the sample structure (it is this structure that is causing the field as a byproduct of back-projection), and can be

variable throughout the sample. The frequency domain signature of isotropic poorly sampled structures should be expected to be equally isotropic as their spatial dimension structure. As anisotropy increases, we would expect these artefact structures to similarly increase in anisotropy, stretching into streaks in extreme cases. This variability with and across samples highlights the importance of sample specific network training to learn to remove the artefact structure specific to that sample.

The relatively isotropic sample structure of the synthetic rock sample (figure 1) leads to a comparatively uniform noise and artefact field (figure 2). This is contrasted with the CMC sample which has a much more anisotropic structure, with an aspect ratio of 2:1 to 3:1. An analysis of the spatial and frequency domain signature of the noise and artefact field from the CMC 301 projection dataset reconstructed with FDK is shown in figure 4.

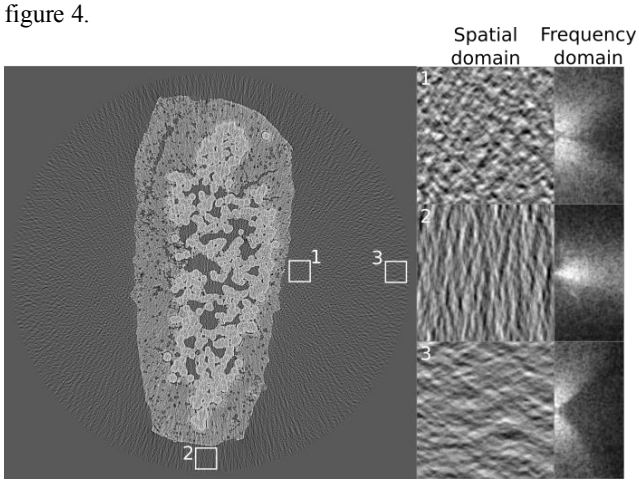


Figure 4: Spatial and frequency domain analysis of noise and artefact field

The frequency spectrum of the synthetic rock noise and artefact field is approximately isotropic – the *x* dimension and *y* dimension spectrum show approximately the same profile. In the CMC dataset, the frequency spectrum from three regions of the spurious signal field all show strongly anisotropic frequency domain signatures. The direction of anisotropy is spatially variable throughout the sample, with region 2 showing a strong structure in the *x* dimension, and region 3 showing strong structure in the dimension, and region 3 showing strong structure in the *y* dimension. The variation in field structure between samples highlights the importance of training deep learning networks to target the noise and artefact structures specific to a sample / dataset. The spatial variability of anisotropy orientation indicates some of the challenges traditional techniques face when addressing these issues – any approach needs to be able to handle spatially variable effects and cannot be universal.

Measurements of the scale of spurious signal field were acquired for each region of interest in the CMC sample (table 1). Datasets reconstructed using deep learning neural network successfully reduced the scale of the spurious signal field without removing small features of interest (as shown in figure 3).

	FDK (3001)	FDK (301)	Deep learning recon (301)
Region 1	0.028	0.180	0.014
Region 2	0.030	0.141	0.012
Region 3	0.021	0.167	0.014

Table 1: Measurements of the scale of the spurious signal field for a 3,001 projection dataset reconstructed using FDK, and a 301 projection dataset reconstructed using FDK and deep learning

IV. Conclusions

A new technique for the automated training of tomographic reconstruction networks is presented and benchmarked across multiple samples, including a synthetic rock and a ceramic matrix composite (CMC) sample. Its performance is discussed in relation to its ability to remove both noise from the randomness inherent in the X-ray detection process and sampling artefacts produced by a limited projection number. The relative impact and signature of these two effects is discussed, and differentiated on a synthetic rock sample. High quality image recovery is only achieved if both sampling artefacts and noise are considered.

The performance of the deep learning based reconstruction is then tested on a CMC sample, showing high quality image reconstruction (mitigating the noise and artefact levels in the reconstructed volume while maintaining small features of interest) with a data sampling ratio up to 10:1 relative to a baseline acquisition. Such a technology could have broad application in the fields of high resolution materials characterization where high spatial resolutions are required without sacrificing acquisition time.

References

[1] W. Yao, L. Chen, H. Wu, Q. Zhao, and S. Luo, "Micro-CT image denoising with an asymmetric perceptual convolutional network," Phys. Med. Biol., vol. 66, no. 13, p. 135018, Jul. 2021.

[2] Y. Da Wang, M. Shabaninejad, R. T. Armstrong, and P. Mostaghimi, "Physical Accuracy of Deep Neural Networks for 2D and 3D Multi-Mineral Segmentation of Rock micro-CT Images," Feb. 2020.

[3] L. Mosser, O. Dubrule, and M. J. Blunt, "Reconstruction of three-dimensional porous media using generative adversarial neural networks," Phys. Rev. E, 2017.

[4] M. Andrew, L. Omlor, A. Andreyev, Ravikumar Sanapala, and M. Samadi Khoshkhoo, "New technologies for x-ray microscopy: phase correction and fully automated deep learning based tomographic reconstruction," 2021.

[5] A. Krull, T. O. Buchholz, and F. Jug, "Noise2void-Learning denoising from single noisy images," in Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition, 2019, vol. 2019-June.

[6] C. Broaddus, A. Krull, M. Weigert, U. Schmidt, and G. Myers, "Removing Structured Noise with Self-Supervised Blind-Spot Networks," in Proceedings - International Symposium on Biomedical Imaging, 2020.

[7] J. Lehtinen et al., "Noise2Noise: Learning image restoration without clean data," in 35th International Conference on Machine Learning, ICML 2018, 2018, vol. 7.

[8] O. Ronneberger, P. Fischer, and T. Brox, "U-Net: Convolutional Networks for Biomedical Image Segmentation [2015; First paper exploring U-Net architecture.]," in International Conference on Medical image computing and computer-assisted intervention, 2015.

[9] L. A. Feldkamp, L. C. Davis, and J. W. Kress, "Practical cone-beam algorithm," J. Opt. Soc. Am. A, vol. 1, no. 6, p. 612, 1984.

Combining Deep Learning and Adaptive Sparse Modeling for Low-dose CT Reconstruction

Ling Chen, Zhishen Huang, Yong Long*, Saiprasad Ravishankar

***Abstract*—Traditional model-based image reconstruction (MBIR) methods combine forward and noise models with simple object priors. Recent application of deep learning methods for image reconstruction provides a successful data-driven approach to addressing the challenges when reconstructing images with measurement undersampling or various types of noise. In this work, we propose a hybrid supervised-unsupervised learning framework for X-ray computed tomography (CT) image reconstruction. The proposed learning formulation leverages both sparsity or unsupervised learning-based priors and neural network reconstructors to simulate a fixed-point iteration process. Each proposed trained block consists of a deterministic MBIR solver and a neural network. The information flows in parallel through these two reconstructors and is then optimally combined, and multiple such blocks are cascaded to form a reconstruction pipeline. We demonstrate the efficacy of this learned hybrid model for low-dose CT image reconstruction with limited training data, where we use the NIH AAPM Mayo Clinic Low Dose CT Grand Challenge dataset for training and testing. In our experiments, we study combinations of supervised deep network reconstructors and sparse representations-based (unsupervised) learned or analytical priors. Our results demonstrate the promising performance of the proposed framework compared to recent reconstruction methods.**

***Index Terms*—Low-dose X-ray CT, image reconstruction, deep learning, transform learning, optimal combination.**

I. Introduction

X-ray computed tomography (CT) is widely used in industrial and clinical applications. It is highly valuable to reduce patients' exposure to X-ray radiation during scans by reducing the dosage. However, this creates challenges for image reconstruction. The conventional CT image reconstruction methods include analytical methods and model-based iterative reconstruction (MBIR) methods [1]. The performance of analytical methods such as the filtered back-projection (FBP) [2] degrades due to the greater influence of noise in the low X-ray dose setting. MBIR methods aim to address such performance degradation in the low-dose X-ray computed tomography (LDCT) setting. MBIR methods often use penalized weighted least squares (PWLS) reconstruction formulations involving simple priors for the underlying object such as edge-preserving (EP) regularization that assumes the image is approximately sparse in the gradient domain. More recent dictionary learning-based methods [3] provide improved image reconstruction quality compared to nonadaptive MBIR schemes, but involve expensive computations for sparse encoding. Recent PWLS methods with regularizers involving learned sparsifying transforms (PWLS-ST [4]) or a union of learned transforms (PWLS-ULTRA [5]) combine both computational efficiency (cheap sparse coding in transform domain) and the representation power of learned models (transforms).

Data-driven (deep) learning approaches have also demonstrated success for LDCT image reconstruction (see [6] for a review). FBPConvNet [7] is a convolutional neural network (CNN) scheme that refines the quality of FBP reconstructed (corrupted) CT images to match target or ground truth images. Another approach WavResNet [8] learns a set of filters that are used in constructing the encoder and decoder of the convolutional framelet denoiser to refine crude LDCT images. However, deep learning methods often require large training sets for effective learning and generalization. Methods based on sparsifying transform learning typically require small training sets and have been shown to generalize reasonably to new data [5]. Hence, Ye et al. [9] proposed a unified supervised-unsupervised (referred to here as Serial SUPER) learning framework for LDCT image reconstruction that combined supervised deep learning and unsupervised transform learning (ULTRA) regularization for robust reconstruction. The framework alternates between a neural network-based denoising step and optimizing a cost function with data-fidelity, deep network and learned transform terms.

In this work, we propose an alternative repeated parallel combination of deep network reconstructions and transform learning-based reconstructions (dubbed Parallel SUPER) for improved LDCT image reconstruction. We show that the adaptive transform sparsity-based image features complement deep network learned features in every layer with appropriate weights to provide better reconstructions than either the deep network or transform learning-based baselines themselves. The proposed parallel SUPER method also outperforms the recent Serial SUPER scheme in our experiments.

II. Parallel SUPER Model and the Algorithm

The proposed parallel SUPER reconstruction model is shown in Fig. 1. Each layer of parallel SUPER is comprised of a neural network and a PWLS based LDCT solver with sparsity-promoting data-adaptive regularizers. The images in the pipeline flow in parallel through these two components in a layer and are combined together (with adapted weight). The

L. Chen, and Y. Long are with the University of Michigan Shanghai Jiao Tong University Joint Institute, Shanghai Jiao Tong University, Shanghai 200240, China (chen_ling@sjtu.edu.cn; yong.long@sjtu.edu.cn).

Z. Huang is with the Department of Computational Mathematics, Science and Engineering, East Lansing, MI 48824, USA (huangz78@msu.edu).

S. Ravishankar is with the Department of Computational Mathematics, Science and Engineering and the Department of Biomedical Engineering, Michigan State University, East Lansing, MI 48824, USA (ravisha3@msu.edu).

*Y. Long is the corresponding author.

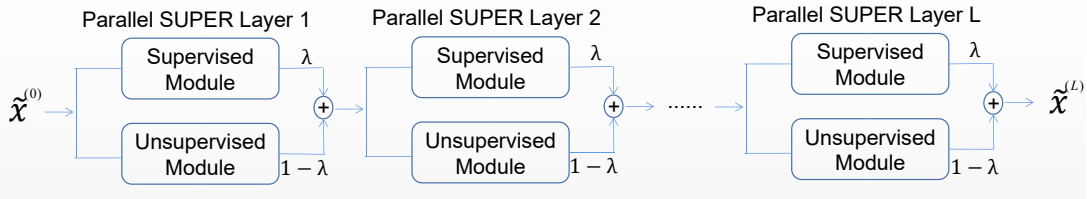


Fig. 1: Overall structure of the proposed Parallel SUPER framework.

framework consists of multiple such parallel SUPER layers to ensure empirical reconstruction convergence. In this work, we have used the FBPConvNet model in the supervised module and PWLS-ULTRA as the unsupervised reconstruction module with a pre-learned union of transforms. However, specific deployed modules in the parallel SUPER framework can be replaced with other parametric models or MBIR methods.

A. Supervised Module

The supervised modules are trained sequentially. We set the loss function during training to be the root-mean-squared error (RMSE) to enforce alignment between the refined images and ground truth images. In the l-th parallel SUPER layer, the optimization problem for training the neural network is:

$$\min_{\boldsymbol{\theta}^{(l)}} \sum_{n=1}^{N} \|G_{\boldsymbol{\theta}^{(l)}}(\widetilde{\mathbf{x}}_n^{(l-1)}) - \mathbf{x}_n^{\star}\|_2^2, \tag{1}$$

where $G_{\boldsymbol{\theta}^{(l)}}(\cdot)$ denotes the neural network mapping in the l-th layer with parameters $\boldsymbol{\theta}^{(l)}$, $\widetilde{\mathbf{x}}_n^{(l-1)}$ is the n-th input image from $(l-1)$-th layer, $\mathbf{x}_n^{\star}$ is the corresponding regular-dose (reference) image or the training label. Note that the neural networks in different layers have different parameters.

B. Unsupervised Module

For the unsupervised module of each layer, we solve the following MBIR problem to reconstruct an image $\mathbf{x} \in \mathbb{R}^{N_p}$ from the corresponding noisy sinogram data $\mathbf{y} \in \mathbb{R}^{N_d}$:

$$\min_{\mathbf{x} \geq 0} J(\mathbf{x}, \mathbf{y}) := \underbrace{\frac{1}{2}\|\mathbf{y} - \mathbf{A}\mathbf{x}\|_{\mathbf{W}}^2}_{:=L(\mathbf{A}\mathbf{x}, \mathbf{y})} + \beta \mathcal{R}(\mathbf{x}), \tag{2}$$

where $\mathbf{W} = \text{diag}\{w_i\} \in \mathbb{R}^{N_d \times N_d}$ is a diagonal weighting matrix with the diagonal elements w_i being the estimated inverse variance of y_i, $\mathbf{A} \in \mathbb{R}^{N_d \times N_p}$ is the system matrix of the CT scan, $L(\mathbf{A}\mathbf{x}, \mathbf{y})$ is the data-fidelity term, penalty $\mathcal{R}(\mathbf{x})$ is a (learning-based) regularizer, and the parameter $\beta > 0$ controls the noise and resolution trade-off.

In this work, we use the PWLS-ULTRA method to reconstruct an image $\mathbf{x}$ from noisy sinogram data $\mathbf{y}$ (measurements) with a union of pre-learned transforms $\{\boldsymbol{\Omega}_k\}_{k=1}^K$. The image reconstruction is done through the following nonconvex optimization problem:

$$\begin{aligned} \widehat{\mathbf{x}}^{(l)}(\mathbf{y}) = \arg\min_{\mathbf{x}} \Bigg\{ & \frac{1}{2}\|\mathbf{y} - \mathbf{A}\mathbf{x}\|_{\mathbf{W}}^2 + \\ & \min_{\mathcal{C}_k, \mathbf{z}_j} \sum_{k=1}^{K} \sum_{j \in \mathcal{C}_k} \left(\|\boldsymbol{\Omega}_k \mathbf{P}_j \mathbf{x} - \mathbf{z}_j\|_2^2 + \gamma^2 \|\mathbf{z}_j\|_0 \right) \Bigg\}, \end{aligned} \tag{3}$$

where $\widehat{\mathbf{x}}^{(l)}(\mathbf{y})$ is the reconstructed image by the unsupervised solver in the l-th layer, the operator $\mathbf{P}_j \in \mathbb{R}^{l \times N_p}$ extracts the j-th patch of l voxels of image $\mathbf{x}$ as $\mathbf{P}_j\mathbf{x}$, $\mathbf{z}_j$ is the corresponding sparse encoding of the image patch under a matched transform, and $\mathcal{C}_k$ denotes the indices of patches grouped into the k-th cluster with transform $\boldsymbol{\Omega}_k$. Minimization over $\mathcal{C}_k$ indicates the computation of the cluster assignment of each patch. The regularizer $\mathcal{R}$ includes an encoding error term and an ℓ_0 sparsity penalty counting the number of non-zero entries with weight γ^2. The sparse encoding and clustering are computed simultaneously. We apply the alternating minimization method from [5] (with inner iterations for updating $\mathbf{x}$) on the above optimization problem. The algorithm also uses a different (potentially better) initialization in each parallel SUPER layer, which may benefit solving the involved nonconvex optimization problem.

C. Parallel SUPER Model

The main idea of the Parallel SUPER framework is to combine the supervised neural networks and iterative model-based reconstruction solvers in each layer. Define $\mathcal{M}(\widetilde{\mathbf{x}}^{(l-1)}, \mathbf{y}; \Gamma)$ as an iterative MBIR solver with initial solution $\widetilde{\mathbf{x}}^{(l-1)}$, noisy sinogram data $\mathbf{y}$ and hyperparameter setting Γ to solve optimization problem (2). In the l-th layer, the parallel SUPER model is formulated as:

$$\begin{aligned} & \widetilde{\mathbf{x}}^{(l)} = \lambda \cdot G_{\theta^{(l)}}(\widetilde{\mathbf{x}}^{(l-1)}) + (1-\lambda) \cdot \widehat{\mathbf{x}}^{(l)}(\mathbf{y}) \\ & \text{s.t.} \begin{cases} \widehat{\mathbf{x}}^{(l)}(\mathbf{y}) & = \mathcal{M}(\widetilde{\mathbf{x}}^{(l-1)}, \mathbf{y}; \Gamma), \\ \boldsymbol{\theta}^{(l)} & = \min_{\boldsymbol{\theta}^{(l)}} \sum_{n=1}^{N} \|G_{\boldsymbol{\theta}^{(l)}}(\widetilde{\mathbf{x}}_n^{(l-1)}) - \mathbf{x}_n^{\star}\|_2^2, \end{cases} \end{aligned} \tag{P0}$$

where λ is the nonnegative weight parameter for the neural network output in each layer and it is selected and fixed in all layers. Each parallel super layer can be thought of as a weighted average between a supervised denoised image and a reconstructed low-dose image from the unsupervised solver. Repeating multiple parallel SUPER layers simulates a fixed-point iteration to generate an ultimate reconstructed image.

The Parallel SUPER training algorithm based on (P0) is shown in Algorithm 1. The Parallel SUPER reconstruction algorithm is the same except that it uses the learned network weights in each layer.

III. Experiments

A. Experiment Setup

In our experiments, we use the Mayo Clinics dataset established for "the 2016 NIH-AAPM-Mayo Clinic Low Dose CT

Algorithm 1 Parallel SUPER Training Alogorithm

Input: N pairs of reconstructed low-dose images and corresponding regular-dose reference images $\{(\widetilde{\mathbf{x}}_n^{(0)}, \mathbf{x}_n^\star)\}_{n=1}^N$, low-dose sinograms $\{\mathbf{y}_n\}$, weights $\mathbf{W}_n$, $\forall$ n, number of parallel SUPER layers L, weight of the supervised module λ,

Output: supervised module parameters $\{\boldsymbol{\theta}^{(l)}\}_{l=1}^L$.

for $l = 0, 1, 2, \ldots, L$ **do**

(1) Update $\widehat{\mathbf{x}}_n^{(l)}(\mathbf{y}_n)$: using $\widetilde{\mathbf{x}}_n^{(l-1)}$ as initial image, use the PWLS-ULTRA method with $\mathbf{W}_n$ [5] to obtain each $\widehat{\mathbf{x}}_n^{(l)}(\mathbf{y}_n)$.

(2) Update $\boldsymbol{\theta}^{(l)}$: with N paired images $\{(\widetilde{\mathbf{x}}_n^{(l-1)}, \mathbf{x}_n^\star)\}_{n=1}^N$, train the supervised model by solving problem (1) to obtain $\boldsymbol{\theta}^{(l)}$.

(3) Generate the output of l-th layer $\widetilde{\mathbf{x}}_n^{(l)}$: use formula in (P0) to obtain the output $\widetilde{\mathbf{x}}_n^{(l)}$ $\forall$ n.

end for

Grand Challenge" [10]. We choose 520 images from 6 of 10 patients in the dataset, among which 500 slices are used for training and 20 slices are used for validation. We randomly select 20 images from the remaining 4 patients for testing. We project the regular dose CT images $\mathbf{x}^\star$ to sinograms $\mathbf{y}$ by adding Poisson and additive Gaussian noise to them as follows:

$$y_i = -\log\left(I_0^{-1}\max\left(\text{Poisson}\{I_0 e^{-[\mathbf{A}\mathbf{x}^\star]_i}\} + \mathcal{N}\{0, \sigma^2\}, \epsilon\right)\right),$$

where the original number of incident photons per ray is $I_0 = 10^4$, the Gaussian noise variance is $\sigma^2 = 25$, and ϵ [11] is a small positive number to avoid negative measurement data when taking the logarithm.

We use the Michigan Image Reconstruction Toolbox to construct fan-beam CT geometry with 736 detectors $\times$ 1152 regularly spaced projection views, and a no-scatter mono-energetic source. The width of each detector column is 1.2858 mm, the source to detector distance is 1085.6 mm, and the source to rotation center distance is 595 mm. We reconstruct images of size 512×512 with the pixel size being 0.69 mm $\times$ 0.69 mm.

B. Parameter Settings

In the parallel SUPER model, we use FBPConvNet as the supervised module and PWLS-ULTRA as the unsupervised module. It takes about 10 hours for training the model for 10 layers in a GTX Titan GPU graphics processor. We train models for different values of the parameter λ (to then select an optimal value), including 0.1, 0.3, 0.5, 0.7, and 0.9. During the training of the supervised method, we ran 4 epochs (kept small to reduce overfitting risks) of the stochastic gradient descent (SGD) optimizer for the FBPConvNet module in each parallel SUPER layer. The training hyperparameters of FBPConvNet are set as follows: the learning rate decreases logarithmically from 0.001 to 0.0001; the batchsize is 1; and the momentum parameter is 0.99. The filters are initialized in the various networks during training with i.i.d. random Gaussian entries with zero mean and variance 0.005. For the unsupervised module, we have trained a union of 5 sparsifying transforms using 12 slices of regular-dose CT images (which are included in the 500 training slices). Then, we use the pre-learned union of 5 sparsifying transforms to reconstruct images with 5 outer iterations and 5 inner iterations of PWLS-ULTRA. In the training and reconstruction with ULTRA, we set the parameters $\beta = 5 \times 10^3$ and $\gamma = 20$. PWLS-EP reconstruction is used as the initialization $\widetilde{\mathbf{x}}^{(0)}$ of the input of network in the first layer.

We compare the proposed parallel SUPER model with the unsupervised method (PWLS-EP), standalone supervised module (FBPConvNet), standalone unsupervised module (PWLS-ULTRA), and the serial SUPER model. PWLS-EP is a penalized weighted-least squares reconstruction method with edge-preserving hyperbola regularization. For the unsupervised method (PWLS-EP), we set the parameters $\delta = 20$ and $\beta = 2^{15}$ and run 100 iterations to obtain convergent results. In the training of the standalone supervised module (FBPConvNet), we run 100 epochs of training to sufficiently learn the image features with low overfitting risks. In the standalone unsupervised module (PWLS-ULTRA), we use the pre-learned union of 5 sparsifying transforms to reconstruct images. We set the parameters $\beta = 10^4$ and $\gamma = 25$, and run 1000 alternations with 5 inner iterations to ensure good performance. In the serial SUPER model, we run 4 epochs of training when learning the supervised modules (FBPConvNet), and we use the pre-learned union of 5 sparsifying transforms and set the parameters $\beta = 5 \times 10^3$, $\gamma = 20$ and $\mu = 5 \times 10^5$ to reconstruct images with 20 alternations and 5 inner iterations for the unsupervised module (PWLS-ULTRA).

C. Results

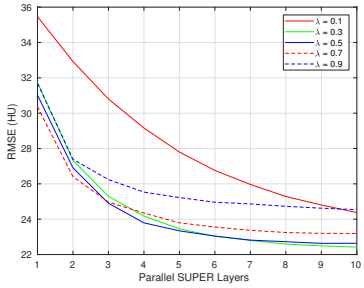


Fig. 2: RMSE (over 20 test slices) comparison with different choices of parameters.

To compare the performance quantitatively, we compute the RMSE in Hounsfield units (HU) and structural similarity index measure (SSIM) [12] for the reconstructed images. For a reconstructed image $\widehat{\mathbf{x}}$, RMSE is defined as $\sqrt{\sum_{j=1}^{N_p}(\widehat{x}_j - x_j^\star)^2 / N_p}$, where $x_j^\star$ denotes the reference regular-dose image intensity at the j-th pixel location and N_p is the number of pixels.

We train the parallel SUPER framework with different choices of the parameter λ including $0.1, 0.3, 0.5, 0.7$ and 0.9 to obtain the best choice. Fig. 2 shows the evolution of RMSE over layers for 20 validation slices with different λ choices. We can see that we obtain the best RMSE when $\lambda = 0.3$.

We have conducted experiments on 20 test slices (slice 20, slice 50, slice 100, slice 150 and slice 200 of patient L067, L143, L192, L310) of the Mayo Clinic data. Table I shows the averaged image quality of 20 test images with different methods. From Table I, we observe that Parallel

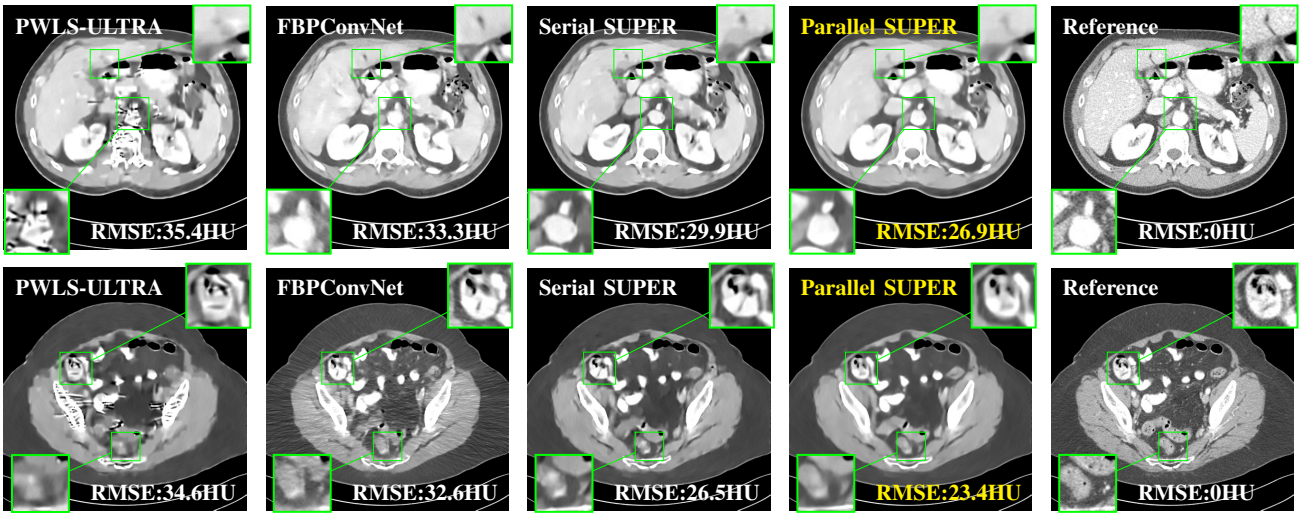


Fig. 3: Reconstruction of slice 50 from patient L067 and reconstruction of slice 150 from patient L310 using various methods. The display window is [800, 1200] HU.

SUPER significantly improves the image quality compared with the standalone methods. It also achieves 1.8 HU better average RMSE compared with Serial SUPER while its SSIM is comparable with Serial SUPER. Fig. 3 shows the reconstructions of L067 (slice 50) and L310 (slice 150) using PWLS-ULTRA, FBPConvNet, serial SUPER (FBPConvNet + PWLS-ULTRA), and parallel SUPER (FBPConvNet + PWLS-ULTRA), along with the references (ground truth). The Parallel SUPER scheme achieved the lowest RMSE and the zoom-in areas show that Parallel SUPER can reconstruct image details better.

TABLE I: The mean RMSE and SSIM values for 20 test images with PWLS-EP, PWLS-ULTRA, FBPConvNet, Serial SUPER, and the proposed Parallel SUPER.

	PWLS-EP	PWLS-ULTRA	FBPConvNet
RMSE	41.4	32.4	29.2
SSIM	0.673	0.716	0.688
	Serial SUPER	Parallel SUPER	
RMSE	25.0	**23.2**	
SSIM	0.748	**0.751**	

IV. Conclusions

This paper proposes the parallel SUPER framework combining supervised deep learning methods and unsupervised methods for low-dose CT reconstruction. We have experimented on a setting with the supervised model FBPConvNet and the unsupervised model PWLS-ULTRA. This framework demonstrates better reconstruction accuracy and faster convergence compared to individual involved modules as well as the recent serial SUPER framework.

Acknowledgment

The authors thank Dr. Cynthia McCollough, the Mayo Clinic, the American Association of Physicists in Medicine, and the National Institute of Biomedical Imaging and Bioengineering for providing the Mayo Clinic data.

References

[1] Z. Huang, S. Ye, M. T. McCann, and S. Ravishankar, "Model-based Reconstruction with Learning: From Unsupervised to Supervised and Beyond," *arXiv e-prints*, p. arXiv:2103.14528, Mar. 2021.

[2] L. A. Feldkamp, L. C. Davis, and J. W. Kress, "Practical cone-beam algorithm," *Journal of the Optical Society of America A*, vol. 1, no. 6, pp. 612–619, 1984.

[3] Q. Xu, H. Yu, X. Mou, L. Zhang, J. Hsieh, and G. Wang, "Low-dose X-ray CT reconstruction via dictionary learning," *IEEE Trans. Med. Imag.*, vol. 31, no. 9, pp. 1682–97, 2012.

[4] X. Zheng, Z. Lu, S. Ravishankar, Y. Long, and J. A. Fessler, "Low dose ct image reconstruction with learned sparsifying transform," *2016 IEEE 12th Image, Video, and Multidimensional Signal Processing Workshop (IVMSP)*, 2017.

[5] X. Zheng, S. Ravishankar, Y. Long, and J. A. Fessler, "PWLS-ULTRA: An efficient clustering and learning-based approach for low-dose 3D CT image reconstruction," *IEEE Trans. Med. Imag.*, vol. 37, no. 6, pp. 1498–510, Jun. 2018.

[6] S. Ravishankar, J. C. Ye, and J. A. Fessler, "Image reconstruction: From sparsity to data-adaptive methods and machine learning," *Proceedings of the IEEE*, vol. 108, no. 1, pp. 86–109, 2020.

[7] K. H. Jin, M. T. Mccann, E. Froustey, and M. Unser, "Deep convolutional neural network for inverse problems in imaging," *IEEE Transactions on Image Processing*, vol. PP, no. 99, pp. 4509–4522, 2016.

[8] E. Kang, W. Chang, J. Yoo, and J. C. Ye, "Deep Convolutional Framelet Denosing for Low-Dose CT via Wavelet Residual Network," *IEEE Trans. Med. Imaging*, vol. 37, no. 6, pp. 1358–1369, 2018.

[9] S. Ye, Z. Li, M. T. Mccann, Y. Long, and S. Ravishankar, "Unified supervised-unsupervised (super) learning for x-ray ct image reconstruction," *IEEE Transactions on Medical Imaging*, vol. 40, no. 11, pp. 2986–3001, 2021.

[10] C. Mccollough, "Tu-fg-207a-04: Overview of the low dose ct grand challenge," *Medical Physics*, vol. 43, no. 2, pp. 3759–60, 2016.

[11] S. Ye, S. Ravishankar, Y. Long, and J. A. Fessler, "Spultra: Low-dose ct image reconstruction with joint statistical and learned image models," *IEEE Transactions on Medical Imaging*, vol. 39, no. 3, pp. 729–741, 2020.

[12] W. Zhou, A. C. Bovik, H. R. Sheikh, and E. P. Simoncelli, "Image quality assessment: from error visibility to structural similarity," *IEEE Trans Image Process*, vol. 13, no. 4, 2004.

Dual-Energy Head Cone-Beam CT Using a Dual-Layer Flat-Panel Detector: Physics-Based Material Decomposition

Zhilei Wang, Hao Zhou, Shan Gu, Hewei Gao*

Department of Engineering Physics, Tsinghua University, Beijing 100084, China

***Abstract*—Flat panel detector (FPD) based cone-beam computed tomography (CT) has made tremendous progress these days, with many new medical and industrial applications keeping emerging from diagnostic imaging and image guidance for radiotherapy and interventional surgery. However, current cone-beam CT (CBCT) is still suboptimal for head scan whose requirement for image quality is extremely strict. Recently, the dual-layer flat panel detector technology is under development and is promising to further advance CBCT from qualitative anatomic imaging to quantitative dual-energy CT. Its potential of enabling head CBCT applications has yet been investigated.The relatively moderate energy separation from the dual-layer FPD and the overall low signal level especially for the bottom layer detector, raise significant challenges in performing high quality dual-energy material decomposition.**

In this work, we propose a physics based material decomposition algorithm that attempts to fully use the detected X-ray signals and prior-knowledge behind head CBCT using dual-layer FPD. Specifically, projection data from the two layers of detector are first adaptively combined to generate conventional CT images with reduced noise. A physics model based dual-layer multi-material spectral correction (dMMSC) is then developed to make the combined image reconstruction beam-hardening free. After a regular projection-domain material decomposition (MD) being conducted, the corresponding beam-hardening free projections from the dMMSC will be taken as a guidance to further enhance the dual-layer MD performance, leading to significantly improved robustness of MD and suppressed low-signal artifact in our preliminary results.

***Index Terms*—head cone-beam CT, material decomposition, dual energy imaging, dual-layer flat panel detector**

I. Introduction

FPD-based CBCT has many advantages, such as low cost, high spatial resolution, and 3D imaging in single rotation that greatly improves the utilization of x-ray photons and reduce the radiation dose to patients. However, the overall image quality of CBCT can be easily compromised by beam hardening, scatter, and cone-beam artifacts [1], impeding its further application in diagnostic imaging such as head scan which has strict requirements on the imaging performance [2]. The emergence of dual-layer flat-panel detector technology can enable dual-energy imaging without changing the conventional scanning protocol, providing extra means of improving CBCT image quality, with potentially material-specific information [3] [4]. However, besides X-ray scattering that is beyond the scope of this study, dual-layer flat panel detector based dual-energy CBCT has its own challenges due to its relatively moderate energy separation and the overall low signal level especially for the bottom layer detector. In this work, we attempt to explore the feasibility of having dual-energy head CBCT using a dual-layer FPD, and to develop a physics model guided material decomposition that fully uses the detected X-ray signals and the physics knowledge behind head CT imaging.

* Author to whom correspondence should be addressed. Email address: hwgao@tsinghua.edu.cn

II. Method

A. Data acquisition from a dual-layer flat panel detector prototype

With a dual-layer FPD, dual-energy data can be acquired simultaneously. As illustrated in fig. 1, a higher energy spectrum is obtained from the bottom layer which is essentially filtered by the top layer detector. Mathematically, low- and high-energy projections can be expressed as

$$P_L = -\ln\left(\frac{\int S(E)e^{-\int \mu(E)\mathrm{dl}}\eta_T(E)\mathrm{dE}}{\int S(E)\eta_T(E)\mathrm{dE}}\right)$$
$$P_H = -\ln\left(\frac{\int S(E)e^{-\sum \mu_i(E)D_i}e^{-\int \mu(E)\mathrm{dl}}\eta_B(E)\mathrm{dE}}{\int S(E)e^{-\sum \mu_i(E)D_i}\eta_B(E)\mathrm{dE}}\right), \quad (1)$$

where, $S(E)$ denotes the incident spectrum of source; $\eta_T(E) \propto (1-e^{-\mu_T(E)D_T})E$ and $\eta_B(E) \propto (1-e^{-\mu_B(E)D_B})E$ denote the detector responses of top layer and bottom layer, respectively;$e^{-\sum \mu_i(E)D_i}$ denotes the attenuation of top-layer

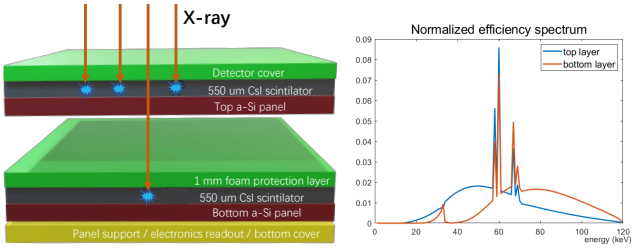


Fig. 1. Left: The detailed structure of a prototype dual-layer FPD used in our study.Right: the normalized effective spectrum for top and bottom layer

material and possible inter-layer filter; $S(E) \cdot \eta_T(E)$ and $S(E)e^{-\sum \mu_i(E)D_i} \cdot \eta_B(E)$ therefore indicate the overall detective spectra of low- and high-energy scan, respectively.

B. Reconstruction from combined dual-layer projections

On a dual-layer flat panel detector, the top layer is more likely to absorb low-energy x-ray photons, while relatively more high-energy x-ray photons are absorbed in the bottom layer. Naturally, by weighting two layers of projection data properly, noise reduction in the corresponding reconstruction will be achieved as more signals (x-ray photons) will be utilized. An optimal weight factor ω can be determined by minimizing the noise on reconstructed images from combined dual-layer projections as below.

$$P_t = \omega P_L + (1-\omega)P_H \tag{2}$$

C. Physics model-guided material decomposition

A high-level diagram of our proposed method can be found in Fig. 2. First, conventional head CT images with decreased noise are reconstructed from adaptively combined dual-layer projection data, followed by a dual-layer multi-material spectral correction (dMMSC) to generate beam hardening free images [5]. After a regular projection-domain material decomposition (MD) using dual-layer detector data that is usually quite sensitive to low-signal x-ray photon and energy separation, the dMMSC corrected projections are used as a physics-model based guidance to further enhance the dual-layer MD performance.

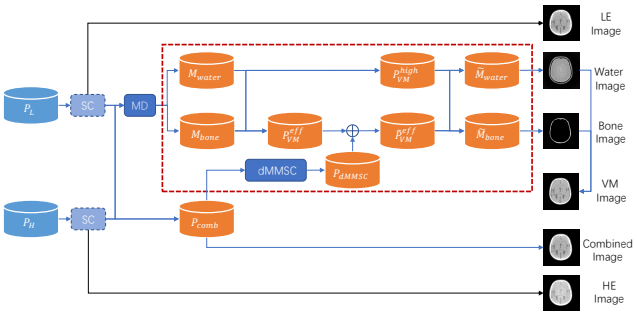


Fig. 2. An illustration of physics-based material decomposition. In our approach, a beam-hardening free projection generated by dual-layer multi-material spectral correction that is based on a physics model of the bone in head is taken as a guidance to improve the material decomposition performance from dual-layer detector.

1) First-pass projection-domain material decomposition: For the dual-layer FPD, since the acquired low- and high-energy projections (P_L, P_H) are quite consistent both spatially and temporarily, it is better to perform a material decomposition in the projection domain so that beam hardening can be better eliminated. For a head scan using a dual-layer FPD, water and bone can be chosen as the basis materials. In this study, we employed a five-order polynomial fitting [6] to generate the basis material projections.

$$\begin{aligned} M_1 &= \sum_{ij} a_{ij} P_L^i P_H^j \\ M_2 &= \sum_{ij} b_{ij} P_L^i P_H^j \end{aligned} \tag{3}$$

After image reconstruction from the basis material projections, two basis material CT images will be created. By weighting the two basis material images using their attenuation coefficients at a specific energy, a virtual monoenergetic (VM) image at that energy can be generated. Theoretically, the virtual monoenergetic images should be free from beam hardening artifacts. However, due to the sensitivity of MD to noise and energy separation, streaks can be easily observed in the VM images which significant degrade the dual-energy performance from the dual-layer FPD.

2) Dual-layer multi-material spectral correction: As we know, due to the significant bony structures in human head, a water correction is not enough to remove beam hardening artifact in head CT images. In the literature, multi-material spectral correction has been developed as a poster-reconstruction method for conventional CT scan to estimate and correct for bone's beam hardening impact. In this work, we extend this kind of method to dual-layer projection data, which consists of the following steps. First, an initial reconstruction after water correction is conducted to estimate the distribution of bony structures that can be easily segmented out from soft tissues. Then, projection of bony structures P_b can be computed using forward projection. With the help of the bone projection, a multi-material spectra-corrected projection that is beam-hardening free can be computed as

$$P_{dMMSC} = P_t + f(P_t, P_b), \tag{4}$$

where, $f(P_t, P_b)$ represents the bone-induced beam-hardening error and can be modeled and calculated in advance by simulations or measurements.

3) material decomposition enhancement using dMMSC: In order to improve the dual-energy performance from the dual-layer FPD, beam hardening free projections after dMMSC is adopted as a guidance to minimize the material decomposition errors and generate a high-quality head image. An optimal hybrid VM projections can be generated using an adaptive fusion of projection with dMMSC and VM projection at an efficient energy as

$$\tilde{P}_{VM}^{eff} = \lambda P_{VM}^{eff} + (1-\lambda)P_{dMMSC}, \tag{5}$$

where, P_{dMMSC} and P_{VM}^{eff} denote the spectra-corrected projection and VM projection, respectively. In practice, the adaptive weighting factor λ might be simplified based on the X-ray signal level.

$$\lambda = \begin{cases} 0 & I < I_l \\ \dfrac{(I - I_l)}{I_h - I_l} & I_l \le I \le I_h \\ 1 & I > I_h \end{cases} \tag{6}$$

Where, I denotes the intensity measured by top layer detector, I_l and I_h are the low and high thresholds, respectively. Basis material projections can also be re-fined by using the hybrid VM projection at optimal keV and a VM projection generated at higher keV (P_{VM}^{high}).

$$\begin{pmatrix} \tilde{M}_{water} \\ \tilde{M}_{bone} \end{pmatrix} = \begin{pmatrix} \mu_{water}(E_{eff}) & \mu_{bone}(E_{eff}) \\ \mu_{water}(E_{high}) & \mu_{bone}(E_{high}) \end{pmatrix}^{-1} \begin{pmatrix} \tilde{P}_{VM}^{eff} \\ P_{VM}^{high} \end{pmatrix} \tag{7}$$

If needed, iterations can be done by using the re-fined decomposition as the initial decomposed results.

III. RESULTS

A. Experimental system set up

Our study is based on a benchtop CBCT system equiped with a prototype dual-layer FPD in our lab.The prototype dual layer FPD (Varex imaging, Salt Lake City, UT, USA) consists of two amorphous silicon (a-Si) panels with pixel size of 150 μm, both deposited with a 550 μm-thick CsI scintillator, with no additional intermediate filter placed in between [Fig. 3].The source-to-axis distance was set to 750mm and the source-to-detector distance to 1184 mm. In our study, 720 projections over 360 degree were collected at 30 fps in a 3×3 binning mode. The X-ray source (Varex Imaging G-242, Salt Lake City, UT, USA) at 120 kV, 64 mA and 5 ms of pulse width. In our current study, a narrowed X-ray collimation was implemented to get rid of X-ray scatter's impact which is another big challenge to CBCT spectral imaging but is under a separate investigation. In our study, a 120 kV X-ray source is used. The estimated effective spectrum for the top and bottom layer is shown in Fig. 1 and the average energy separation between low- and high-energy spectra can reach a level of 17 keV without object in the beam.

B. Combined reconstruction from dual-energy projections

By combining the dual layer projection data using different weighting factor, an optimal weighting factor can be determined empirically by measuring the standard deviations of selected ROIs. As shown in Fig. 4, the measurements of signal to noise ratio (SNR) on reconstructed images from top-layer, bottom-layer and combined data are shown beside the selected ROIs. Our preliminary results suggest that the standard deviation can be reduced by roughly 10% when compared with that of the top layer alone.

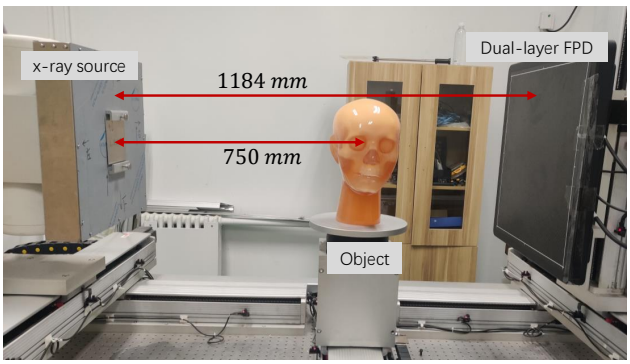


Fig. 3. The benchtop CBCT system using DL FPD.

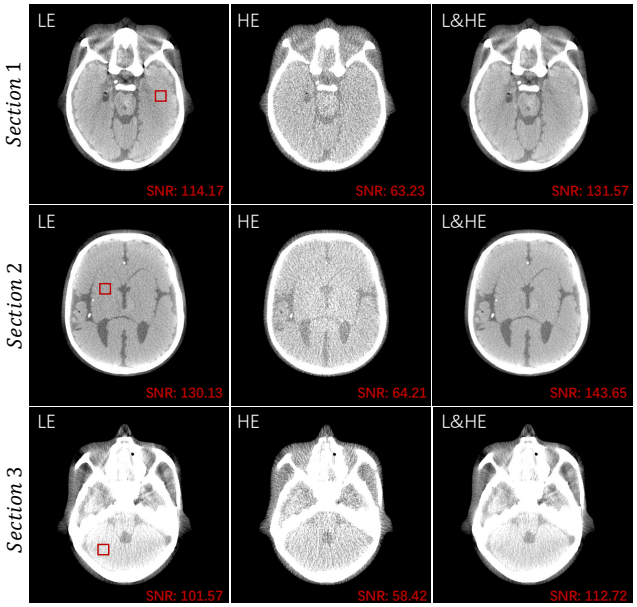


Fig. 4. Reconstruction from low- and high-energy projections and our proposed weighted combination of the projections. Display window: [-100, 100] HU (Hounsfield Units).

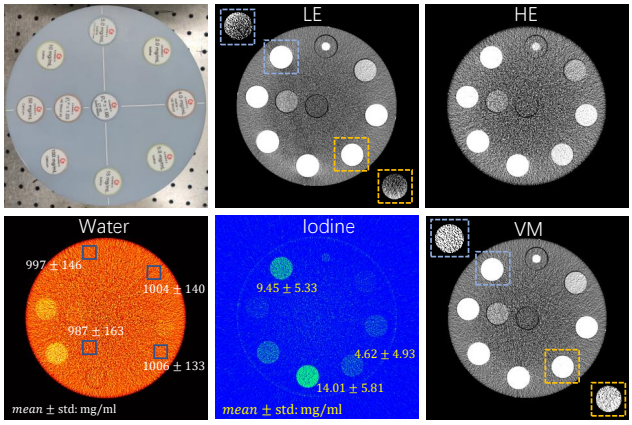


Fig. 5. Results of a multi-energy phantom on the DE-CBCT, including material decomposition and VM images. Display Windows for water and iodine images are [500, 1500] and [3, 14] mg/ml respectively; Display windows for LE, HE and VM images: [-100, 100] HU; Display windows for zoomed-in ROIs marked by blue and orange squares are [425, 525] and [290, 350] HU, respectively.

C. Quantitative evaluations on material decomposition

To quantitatively assess the dual-energy CT imaging performance, Fig. 5 shows reconstructed DE-CBCT images 0f a multi-energy phantom, decomposed results and the virtual monoenergetic (VM) image at 60 keV. According to our preliminary results, the estimated material densities are 14.01±5.81 mg/ml, 9.45±5.33 mg/ml and 4.62±4.93 mg/ml

for 15 mg/ml, 10mg/ml and 5mg/ml iodine, respectively. The decomposition errors are roughly 6% compared to ground truth, which may result from the spectral nonuniformity in the X-ray beam or detector response. The VM image generated from the combination of decomposed water and iodine has fewer artifacts and an improved contrast-to-noise ratio when compared with the regular single-energy image.

The results of basis material decomposition of a head phantom are shown in Fig. 6. The VM images at 63 keV was generated and it is seen that most beam hardening artifacts caused by the bony objects are removed. However, in some locations strong streak artifacts occur badly in the CT images, which can be well suppressed by using our proposed approach. Firstly, we use the spectra-corrected projection at 63 keV as a guidance of VM projection at the same keV to generate an improved VM projection which is combined with generated VM image at 90 keV to refine the water and bone image. By using our physics based material decomposition, we can see that most streak artifacts can be removed in the basis material images and VM images.

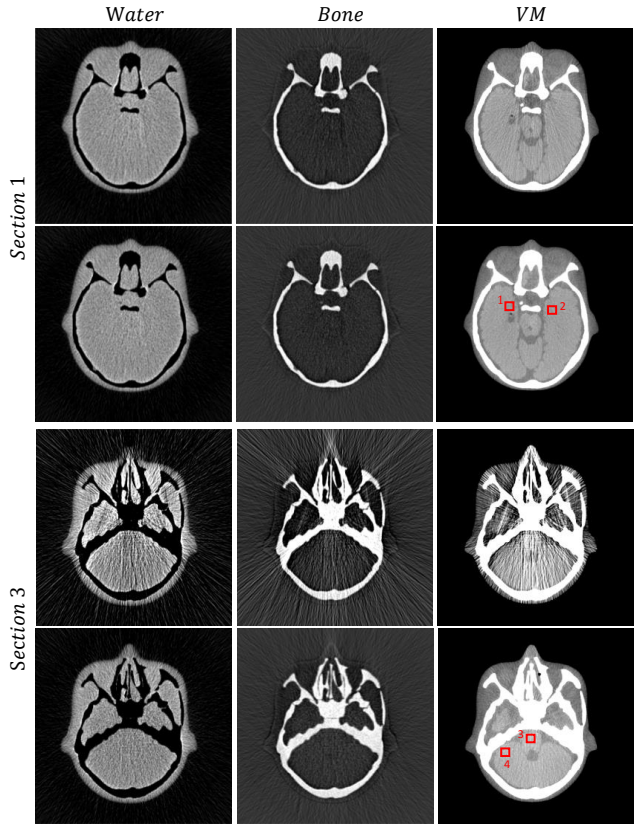


Fig. 6. Results of first-pass material decomposition (1st and 3rd rows) and physics model-guided material decomposition (2nd and 4th rows) at two head phantom sections. Display windows for the water, bone and VM images are [-500, 2000] mg/ml, [0, 2000] mg/ml, [-150, 150] HU, respectively.

To quantitatively evaluate the results of our proposed method, we select some regions of interest (ROIs) and measure the averaged CT numbers and their standard deviations. The measurements are summed in the TABLE I. We can see that the improved VM images by our proposed method show better image quality with more uniform CT values and smaller standard deviations.

TABLE I
MEASUREMENTS OF QUANTITATIVE ANALYSIS

		MEAN(HU)	STD(HU)
ROI1	combined recon	19.0	10.28
	first-pass VM image	35.6	13.5
	improved VM image	31.4	9.2
ROI2	combined recon	14.6	9.5
	first-pass VM image	34.4	13.8
	improved VM image	26.3	9.2
ROI3	combined recon	81.6	16.9
	first-pass VM image	123.5	59.4
	improved VM image	74.4	13.6
ROI4	combined recon	92.8	12.7
	first-pass VM image	22.5	37.8
	improved VM image	69.1	13.8

IV. CONCLUSION

In this work, we proposed a physics model guided material decomposition algorithm that is suitable for head CBCT using dual-layer FPD. Preliminary results using narrowed X-ray collimation on our benchtop dual-layer CBCT system showed its effectiveness of improving image quality in terms of SNR, CNR and low-signal artifact suppression.

Further investigations of our method include more rigorous and broader performance evaluations. Correlated scatter correction for dual-layer FPD is also under development, which will be utilized together with our proposed pMD algorithm here to fully assess the cone-beam CT spectral imaging capability in the near future.

V. ACKNOWLEDGEMENTS

This project was supported in part by grants from the National Natural Science Foundation of China (No. U20A20169 and No. 12075130).

REFERENCES

[1] Tang, X., Krupinski, E. A., Xie, H. and Stillman, A. E. (2018), On the data acquisition, image reconstruction, cone beam artifacts, and their suppression in axial MDCT and CBCT – A review. Med. Phys., 45: e761-e782.

[2] Wu, P., Sisniega, A., Stayman, J.W., Zbijewski, W., Foos, D., Wang, X., Khanna, N., Aygun, N., Stevens, R.D. and Siewerdsen, J.H. (2020), Cone-beam CT for imaging of the head/brain: Development and assessment of scanner prototype and reconstruction algorithms. Med. Phys., 47: 2392-2407.

[3] Wang, W., Ma, Y., Tivnan M, Li J, Gang, G. J., Zbijewski, W., Lu, M., Zhang, J., Star-Lack, J., Colbeth, RE., Stayman, JW. High-resolution model-based material decomposition in dual-layer flat-panel CBCT. Med. Phys. 2021 Jul 17.

[4] Shi, L., Lu, M., Bennett, N.R., Shapiro, E., Zhang, J., Colbeth, R., Star-Lack, J. and Wang, A.S. (2020), Characterization and potential applications of a dual-layer flat-panel detector. Med. Phys., 47: 3332-3343.

[5] Gao, H., Cohen, A., Imai, Y. Quantitative Uniformity of Iodinated Contrast Across the Z-Coverage of Large Cone-Angle CT. in Proceedings of 3rd CT meeting. 2014, pp. 220-223

[6] Stenner, P., Berkus, T. and Kachelriess, M. (2007), Empirical dual energy calibration (EDEC) for cone-beam computed tomography. Med. Phys., 34: 3630-3641.

X-ray Dissectography Enables Stereotography

Chuang Niu and Ge Wang

***Abstract*—X-ray imaging is the most popular medical imaging technology. While x-ray radiography is rather cost-effective, tissue structures are superimposed along the x-ray paths. On the other hand, computed tomography (CT) reconstructs internal structures but CT increases radiation dose, is complicated and expensive. Here we propose "x-ray dissectography" to extract a target organ digitally from few radiographic projections for stereographic and tomographic analysis in the deep learning framework. As an exemplary embodiment, we propose a general X-ray dissectography network, a dedicated X-ray stereotography network, and the X-ray imaging systems to implement these functionalities. Our experiments show that x-ray stereography can be achieved of an isolated organ such as the lungs in this case, suggesting the feasibility of transforming conventional radiographic reading to the stereographic examination of the isolated organ, which potentially allows higher sensitivity and specificity, and even tomographic visualization of the target. With further improvements, x-ray dissectography promises to be a new x-ray imaging modality for CT-grade diagnosis at radiation dose and system cost comparable to that of radiographic or tomosynthetic imaging.**

***Index Terms*—X-ray radiography, digital tomosynthesis, computed tomography, x-ray dissectography, x-ray stereography, artificial intelligence, deep learning**

I. Introduction

X-ray imaging is the first and still most popular modern medical imaging approach, which is performed by various kinds of systems. In the low-end, x-ray radiography takes a two-dimensional projective image through a patient, which is called a radiogram or radiograph. In the high end, many x-ray projections are first collected and then reconstructed into tomographic images transversely or volumetrically. Between these two extremes, digital tomosynthesis takes a limited number of projections over a relatively short scanning trajectory and infers three-dimensional features inside a patient. These x-ray imaging modes have their strengths and weaknesses. X-ray radiography is cost-effective but it produces a single projection, and has multiple oranges and tissues superimposed along x-ray paths, compromising the diagnostic performance. On the other hand, x-ray CT unravels structures overlapped in the projection domain into tomographic images in a 3D coordinate system but CT uses a much higher radiation dose, is complicated and expensive. Digital tomosynthesis is a balance between x-ray radiography and CT in terms of the number of needed projections, the information in resultant images, and the cost to build and operate the imaging system.

Reducing radiation dose and improving imaging quality and speed are the main tasks for the development of x-ray imaging technologies. As x-ray radiography has the lowest radiation dose, the fastest imaging speed, and the lowest price, researchers have been focusing on improving the radiogram quality. Currently, there are mainly two ways for this purpose: 1) suppressing interfered structures [1], [2], [3], [4] or enhancing related structures [5], [1], and 2) generating 3D volumes [6], [7], [8]. It is well known that superimposed anatomical organs in 2D radiographs significantly complicate signal detection, such as for diagnosis of lung diseases. In early studies [9], [10] model-based methods were developed to suppress ribs in chest radiographs, some of which require manually annotated bone masks. In recent years, deep learning methods [1], [2], [3], [4] were proposed for suppression of ribs by leveraging 3D CT prior. Instead of suppressing the ribs in CXR images, Gozes and Greenspan proposed to enhance lung structures by extracting the extracted lungs first and then adding the result back with a scaling factor [5]. Generating a 3D volume from a single or a few radiographs is another way to improve radiography. Ying et al. proposed X2CT that generates a 3D CT volume from a pair of orthogonal radiographs using a CycleGAN framework [6]. Recently, some methods [7], [8] proposed to generate a 3D volume from single or few radiograhps.

Among the above-surveyed methods, those methods that are for suppressing/enhancing specific structures are mainly intended to improve the performance of classification [1], [2], [3], [4] and detection [5] from a single radiograph without providing 3D information. On the other hand, although the rest of the existing methods that map 2D radiograms to 3D CT volumes achieved remarkable results, they cannot reconstruct structures accurately and reliably, since their clinical utilities have not been demonstrated so far. Particularly, it can be seen that almost all the above methods depend on the GAN framework and/or unpaired learning for 2D/3D image generation. A major potential problem is that GAN-based models tend to generate fake structures, which is a major concern in the medical imaging field.

In this study, we propose x-ray dissectography (XDT) in general and specialize it for x-ray stereotography to improve image quality and diagnostic performance. The essential idea is that we digitally extract a target organ from the original radiograph or radiogram which contains superimposed organs via deep learning, facilitating both visual inspection and quantitative analysis. Considering that radiograhs from different views contain complementary information, we design a physics-based XDT network to extract the multi-view features and transform them into a 3D space. In this way, the target organ can be synergistically analyzed in isolation from different projection angles. As a special yet important application of our XDT approach, we propose X-ray stereography that allows a reader immersively perceive the target organ from two dissected radiographs in 3D, synergizing machine

C. Niu and G. wang are with Department of Biomedical Engineering, Center for Biotechnology and Interdisciplinary Studies, Rensselaer Polytechnic Institute, Troy, NY USA, 12180. E-mail: niuc@rpi.edu; wangg6@rpi.edu.

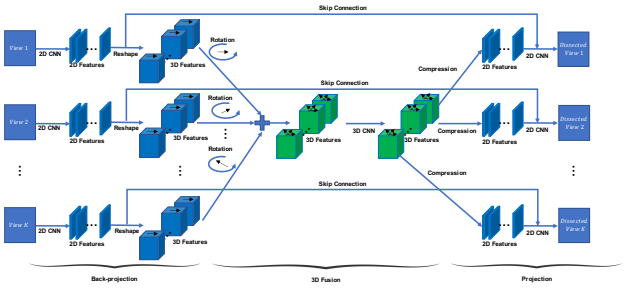


Fig. 1. X-ray dissectography network (XDT-Net).

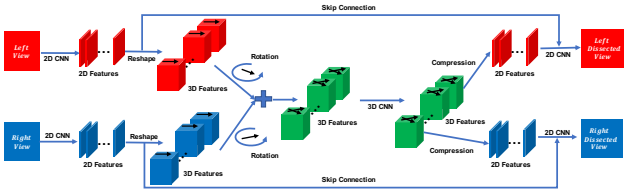


Fig. 2. X-ray stereoraphic imaging network (XST-Net).

intelligence and human intelligence, similar to what CT does. Biologically, stereo perception is based on binocular vision for the brain to reconstruct a 3D scene, and can be applied to see through dissected radiograms and form a 3D rendering in a radiologist in mind. In this work, we design an X-ray imaging system dedicated to this scenario. Different from our daily visual information processing, which senses surroundings with reflected light signals, radiograms are projective through an object to allow 3D concept of x-ray semi-transparent features.

To avoid fake structures, we optimize XDT neural networks in a supervised 2D-to-2D learning paradigm without using a GAN-like model. To obtain a 2D radiograph of a target organ without surrounding tissues, we can manually or automatically segment the organ in the associated CT volume first and then compute the ground-truth radiograph through projecting the dissected organ according to the system parameters. In other words, radiographs and CT images are obtained from the same patient and the same imaging system to avoid unpaired learning. We utilize a cutting-edge simulation platform, such as the popular academic [11] and industrial [**?**] simulators, for training XDT networks. These simulators can take either a clinical CT volume or a digital 3D phantom to compute a conventional x-ray radiograph, and then extracts a target organ digitally to produce the ground-truth radiograph of the organ. Our initial experimental results have shown that XDT indeed separates the lungs with faithful texture and structures, and we can perceive the extracted lungs via stereoscopic viewing with a pair of 3D glasses. Potentially, this approach can improve the diagnostic performance in lung cancer screening, COVID-19 follow-up, and other applications.

II. Methodology

A. General Workflow of X-ray Dissectography

X-ray dissectography (XDT) is dedicated to transform a conventional radiogram $\boldsymbol{x} = \sum_{i=1}^{B} \boldsymbol{y}_i + \boldsymbol{y}_t$ to a projection image $\boldsymbol{y}_t$, where $\boldsymbol{y}_t$ is a projection of only a target organ, and $\sum_{i=1}^{B} \boldsymbol{y}_i$ represents the superimposed image of the B anatomical components involved in the conventional radiogram. In fact, it is an extremely ill-posed problem as we can only observe the radiograph $\boldsymbol{x}$, and impossible to obtain any analytic solution in a general setting. Fortunately, a specific organ in the human body has a fixed relative location, a strong prior on material composition, and similar patterns (such as shapes, textures, and other properties). Given this kind of knowledge, a radiologist can identify different organs in the conventional radiogram. However, the superimposed organs challenge the human in visual inspection for the target one. Considering the great progresses in deep imaging [12], [13], [14], deep neural networks (DNNs) is used to learn such priors and extract purified radiographs as if x-rays go only through the target organ. Such DNNs can be trained with an individual or a specific population for quantitative accuracy and clinical utilities.

In this study, we propose a physics-based XDT network (XDT-Net) for separating a target organ from more than one views, as shown in Fig. 1. Note that various organs may be separated using this framework in different combinations, depending on specific applications. The XDT-Net consists of the three modules: 1) a back-projection module, 2) a 3D fusion module, and 3) a projection module. The back-projection module maps 2D radiographs to 3D features, like a tomographic back-projection process. It consists of k 2D convolutional neural networks (CNNs) followed by reshape operators, where k is the number of input views, each CNN is applied to a specific view, and different CNNs have the same architecture but trainable parameters may be optimized differently. The fusion module integrates the information from all views in the 3D feature space. It first aligns the 3D features of different views by rotation according to their projection angles, and then combines them by a 3D CNN. The projection module predicts each radiograph containing only the target organ. It first squeezes each 3D feature volume to a 2D feature map along a given angle, and then the 2D CNN takes the 2D features from both the squeeze operator and the back-projection module to predict the radiograph of the target organ.

B. Specific Embodiment for X-ray Stereography

We perceive the world in 3D thanks to binocular vision. Given binocular disparity, the human brain is capable of sensing the depth in the scene. Inspired by this amazing fact, here we investigate X-ray stereography (XST) with two radiograms of an isolated organ. When inspecting the human body with x-rays, organs with large linear attenuation coefficients will overwhelm the ones with small attenuation coefficients in radiograms. As a result, it is difficult to discern subtle changes in internal organs due to the superimposition of multiple organs, significantly compromising stereopsis. With our proposed XDT-Net and XST-Net, we can integrate machine intelligence for target organ dissection and human intelligence for stereographic perception so that a radiologist can perceive a target organ in 3D with details much more vivid than in 2D, potentially improving diagnostic performance.

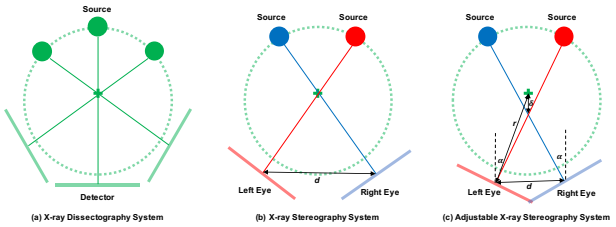


Fig. 3. X-ray imaging system configuration to facilitate radiographic, stereographic and tomographic analysis on a digitally isolated organ or tissue.

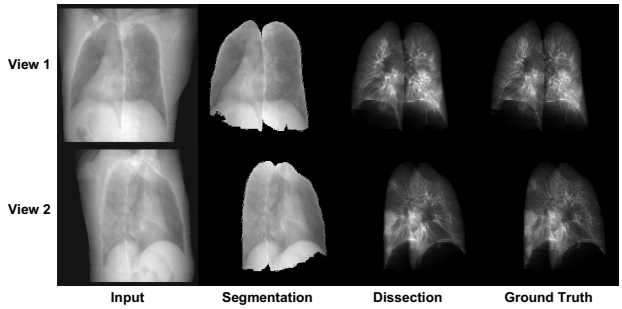


Fig. 4. XDT testing results from two new orthogonal views of the same patient.

To enable XST of a specific organ, we adapt the XDT-Net to the XST-Net as shown in Fig 2. The XST-Net also consists of the same three modules: the backprojection module, the 3D fusion module, and the projection module. Each module of the XST-Net shares the same network architecture as that of the XDT-Net but needs to be adapted for stereo viewing. First, the backprojection module of the XST-Net takes two radiographs as inputs, which are two images into our eyes. Second, in reference to the view angles of two eyes, the 3D fusion module uses a different rotation center to align 3D features from two branches appropriately. Our proposed XST imaging system is described in Subsection II-C. Third, the projection module translates the merged 3D feature first and then squeezes it to 2D feature maps according to the human reader's viewing angles. Finally, two dissected radiographs are respectively sent to the left and right eyes through a pair of 3D glasses for stereoscopy.

C. Design on X-ray Dissectography and Stereography

We assume that radiographs from sufficiently many different angles can be obtained in the network training stage such that image volumes can be reconstructed. The traditional cone-beam CT system, as shown in Fig. 3 (a), serves this purpose. Then, many pairs of conventional radiographs and the counterparts of the target-only radiographs can be obtained from a reconstructed CT volume and a segmented organ in the reconstructed CT volume respectively. In the testing stage, the same XDT system only needs to generate few radiographs at any angles for the trained XDT-Net to extract the corresponding radiograms of the target organ alone, without surround tissues, for much-improved visual inspection and quantitative analysis. To achieve x-ray stereopsis, we design an XST imaging system, as shown in Fig. 3 (b) and (c), where each source is regarded as an eye while the projection through the body is recorded on the opposite detector. In a simple setting, we directly take two radiograms from the XDT system so that the distance between eyes is d, as shown in Fig. 3 (b). In this case, the center X-rays from the source positions intersect at the object center. For the adaption to different applications and readers, we further design an adjustable XST system in Fig. 3 (c). There are two parameters of the system to control the offset between the two eyes and the viewing angle from a pre-specified principal direction. In Fig. 3 (c), red and blue dots denote the left and right eyes, the red and blue plates are the corresponding detectors, and green cross is viewed as the object center. The distance between two eyes is d, the distance between the source and the object center is r, and the angle between center X-ray and the pre-specified reference direction is α for both eyes. Thus, the intersection point of two center X-rays is translated from the object center along the vertical direction. The distance offset δ can be computed as $\delta = \frac{d}{2\tan(\alpha)} - \sqrt{r^2 - (d/2)^2}$, which is used to adjust the rotation center for XST-Net as introduced in Subsection II-B.

In practice, the XST system for inspecting different organs may require different geometric parameters. Both XDT and XST systems can be implemented in various ways such as with robotic arms [15] so that the geometric parameters can be easily set to match a reader's preference.

III. Experimental Results

A. XDT and XST Simulation

Here we used a clinical CT dataset, denoted as CT-lung, and simulated radiograms in cone-beam geometry. Specifically, 50 reconstructed CT volume of patients were selected, including 10 patients from [16] and 40 patients from [17]. The data from [16] provide the 3D lung masks. Hence, we can simulate the paired radiograms with and without lung masks [18]. Note that before performing the cone-beam projection, the patient bed in the CT volume was first masked out in a semi-automatic manner. Since the lung masks provided in [17] are not consistent with those provided in [16], we trained a semantic segmentation UNet with the annotated data in [16] to identify the body region for removal of the patient bed and segment the lung region slice-by-slice for the data in [17] consistently. When 2D radiograms were synthesized, we rotated the patient CT volume from 0° to 180°, where the angle of 0° is the frontal view. Totally, we obtained 9,000 pairs of radiograms, and the image size is 320×320.

B. XDT Results

We first evaluated the effectiveness of the proposed XDT-Net on the CT-lung dataset. In the current experiments, we focused on simultaneously dissecting two radiograms at orthogonal angles. To be more specific, the region of interest was first segmented from 2D radiograms before forwarding to the XDT-Net. In this way, the task of XDT-Net is purified to improve dissection results. For this purpose, we trained a segmentation model to identify the region of interest. The

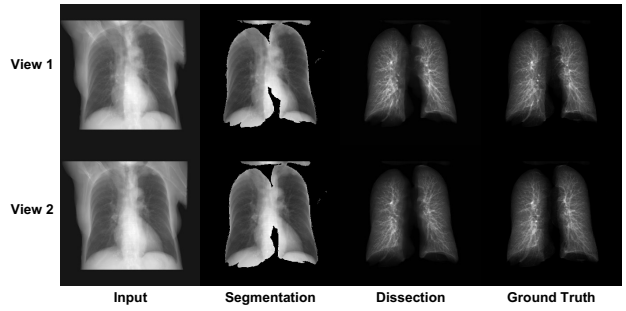


Fig. 5. XST testing results from two new stereo views of the same patient.

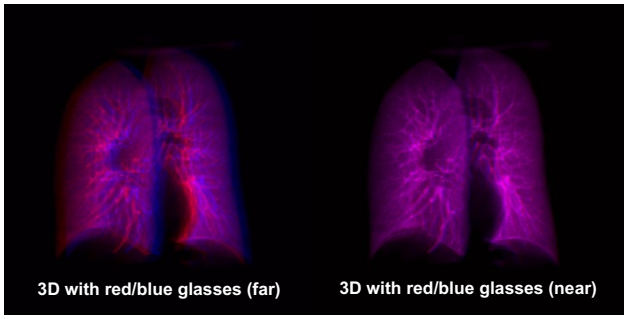


Fig. 6. XST image pair and rendering for stereo-viewing.

target mask for training this segmentation model can be easily obtained by thresholding the radiogram of isolated lungs, where the threshold was empirically set to 0.01 in the unit of linear attenuation coefficient.

The testing results are shown in Fig. 4, where radiographs of the same patient were collected from different angles. From the first to the fourth columns are the input radiograms, the segmentation results, the dissection results, and the ground-truth respectively. The first and the second rows present two orthogonal projections respectively. The visual inspection shows that the dissected radiograms are very close to the ground-truth in terms of detailed structures despite being slightly smoother. The blurring effect may be due to noise reduction [19]. Compared to normal radiograms, the dissected radiograms remove irrelevant surrounding structures, highlight the target organ, and potentially improve the diagnosis performance.

C. XST Results

Then, we evaluated the feasibility of the proposed XST-Net for X-ray stereoscopic imaging on the CT-lung dataset. We first evaluated the joint dissection results from two stereo radiograms collected at two new angles of the same patient and then generalized the stereo-imaging technology to different patients. Our representative results are shown in Fig. 5, showing XST-Net achieves very promising results for stereo views. In addition, we have found that the dissection networks are quite robust to segmentation results, geometric parameters, and image noise. Finally, we generated 3D perception by overlapping the left-eye and right-eye images in red and blue channels and then viewing both through a pair of red/cyan glasses. Fig. 6 shows stereo images and two 3D images adjusted with different geometric parameters, as discussed in Subsection II-C. Readers can enjoy watching the 3D lungs through red/cyan glasses (you may need some visual adaptation to see the 3D scene).

IV. Conclusion

In conclusion, we have proposed the x-ray dissectography (XDT) and x-ray stereography (XST) systems and methods for improving the utilities of conventional X-ray radiography and digital tomosynthesis. The proposed XDT and XST can dissect a target organ type from X-ray radiograms with deep learning. The experimental results clearly demonstrate the feasibility and potential utility of the proposed imaging technology. In the future, we will continue improving the network model and producing clinically relevant results systematically. Hopefully, the proposed XDT and XST techniques empowered by artificial intelligence may open a new door for traditional X-ray radiography to have new impacts on healthcare.

References

[1] Z. Li and et al., "Encoding ct anatomy knowledge for unpaired chest x-ray image decomposition," 2019.
[2] C. Peng and et al., "Xraysyn: Realistic view synthesis from a single radiograph through ct priors," 2020.
[3] H. Li and et al., "High-resolution chest x-ray bone suppression using unpaired ct structural priors," *IEEE TMI*, vol. 39, no. 10, pp. 3053–3063, 2020.
[4] L. Han and et al., "Gan-based disentanglement learning for chest x-ray rib suppression," 2021.
[5] O. Gozes and H. Greenspan, "Lung structures enhancement in chest radiographs via CT based FCNN training," *arXiv*, 2018.
[6] X. Ying and et al., "X2ct-gan: Reconstructing ct from biplanar x-rays with generative adversarial networks," in *CVPR*, 2019.
[7] L. Shen, W. Zhao, and L. Xing, "Patient-specific reconstruction of volumetric computed tomography images from a single projection view via deep learning," *Nature Biomedical Engineering*, vol. 3, no. 11, pp. 880–888, 2019.
[8] L. Shen and et al., "A geometry-informed deep learning framework for ultra-sparse 3d tomographic image reconstruction," 2021.
[9] L. Hogeweg and et al., "Suppression of translucent elongated structures: Applications in chest radiography," *IEEE TMI*, vol. 32, no. 11, pp. 2099–2113, 2013.
[10] D. Wu and et al., "A learning based deformable template matching method for automatic rib centerline extraction and labeling in ct images," 2012, pp. 980–987.
[11] E. Abadi and et al., "Dukesim: A realistic, rapid, and scanner-specific simulation framework in computed tomography," *IEEE TMI*, vol. 38, pp. 1457–1465, 2019.
[12] C. Niu and et al., "Low-dimensional manifold constrained disentanglement network for metal artifact reduction," *IEEE TRPMS*, pp. 1–1, 2021.
[13] C. Niu, M. Li, and G. Wang, "Multi-window learning for metal artifact reduction," in *Developments in X-Ray Tomography XIII*, vol. 11840, 2021, p. 1184015.
[14] C. Niu and et al., "Noise entangled gan for low-dose ct simulation," *arXiv preprint arXiv:2102.09615*, 2021.
[15] A. Fieselmann and et al., "Twin robotic x-ray system for 2D radiographic and 3D cone-beam CT imaging," in *Medical Imaging 2016: Physics of Medical Imaging*, vol. 9783, 2016, pp. 128 – 133.
[16] S. P. Morozov and et al., "Mosmeddata: Chest CT scans with COVID-19 related findings dataset," *CoRR*, 2020.
[17] P. Bilic and et al., "The liver tumor segmentation benchmark (lits)," *CoRR*, 2019.
[18] W. van Aarle and et al., "Fast and flexible x-ray tomography using the astra toolbox," *Opt. Express*, vol. 24, no. 22, pp. 25 129–25 147, 2016.
[19] C. Niu and G. Wang, "Noise2sim - similarity-based self-learning for image denoising," *CoRR*, vol. abs/2011.03384, 2020.

Mixed coronary plaque characterization with the first clinical dual-source photon-counting CT scanner: a phantom study

Thomas Wesley Holmes[1], Leening P. Liu[2,3], Nadav Shapira[2], Elliot McVeigh[3], Amir Pourmorteza[1,5], Peter B. Noël[2]

***Abstract*— Purpose: to investigate image quality of the ultra-high-resolution (UHR) mode of a dual-source photon-counting CT scanner in visualizing mixed (soft and hard) coronary artery plaques.**

Materials and methods: We scanned a custom-made phantom with 10 mixed plaques of various sizes and compositions. Each scan was repeated three times. Images were reconstructed with FBP, and model-based quantum iterative reconstruction (QIR). Image quality was investigated by measuring mean CT numbers, noise standard deviation (SD), and by line profiles.

Results: UHR mode provided sharper difference between soft and hard plaques, and the lumen by reducing blooming artifacts. Furthermore, it improved the true CT number of the values by reducing partial volume However, SD of noise increases by a factor of ~8 in FBP reconstructions at thinnest slice thickness (0.2 mm). Quantum iterative reconstruction algorithm reduced image noise x4 of the SR FBP without any apparent loss of spatial resolution.

Conclusion: UHR PCCT improves plaque characterization through improved spatial resolution which results in lower blooming artifacts and partial volume effects. The increase in image noise can be mitigated by using model-based iterative reconstruction algorithms without any loss of spatial resolution. Depending on the imaging task, further noise reduction can be achieved by reconstructing thicker slices. A more detailed investigation with noise power spectrum analysis and observer model studies is warranted.

***Index Terms*— photon-counting, computed tomography, cardiac CT, coronary plaque, plaque characterization, ultra-high-resolution imaging,**

1. Department of Radiology and Imaging Sciences and Winship Cancer Institute, Emory University, Atlanta, GA.
2. Department of Radiology, Perelman School of Medicine, University of Pennsylvania, Philadelphia, PA.
3. Department of Bioengineering, University of Pennsylvania, Philadelphia, PA.
4. Departments of Medicine, Radiology, and Biomedical Engineering, University of California San Diego, San Diego, CA.
5. Department of Biomedical Engineering, Georgia Institute of Technology,Atlanta,GA
Corresponding Author: Amir Pourmorteza (amir.pourmorteza@emory.edu)

I. Introduction

PHOTON-COUNTING CT (PCCT) has proven to be the next leap in CT imaging technology with many studies performed on prototype scanners developed by all major CT manufacturers [1], [2]. In this study, we investigate the image quality and ultra-high-resolution (UHR) performance of the first clinical dual-source PCCT scanner to characterize coronary artery plaques. It is well understood that increasing in spatial sampling frequency of a CT scanner results in wider image noise bandwidth while reducing blooming and partial-volume artifacts[3]. In this work we present preliminary results of imaging mixed, i.e. soft and hard coronary artery plaques with such a system using a custom-made CT phantom.

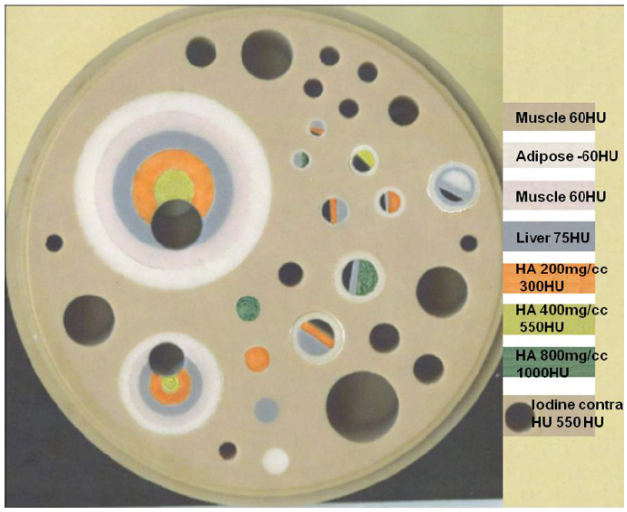


Fig. 1. Custom-made mixed plaque phantom with eight tissue types and ten simulated mixed plaques of various compositions. Contrast agents can be added to the empty holes at desired concentrations. The diameter of the phantom is 100 mm and it can be fit inside a standard QRM thorax phantom.

II. Methods

A. Test Object

We used a custom-made mixed plaque phantom with eight

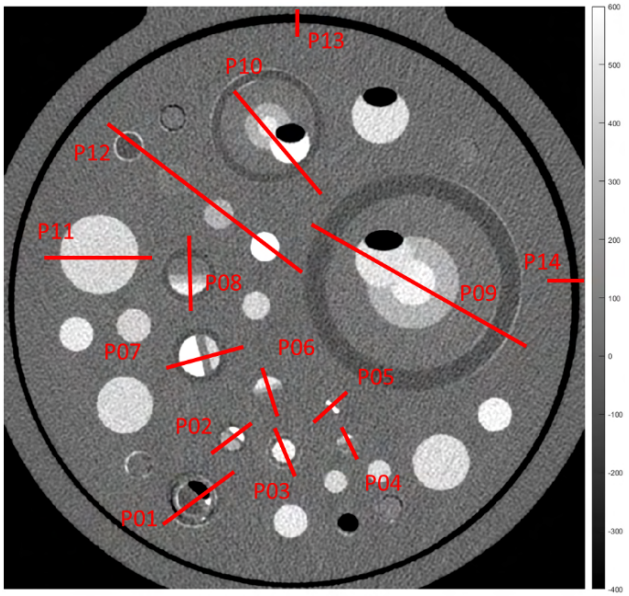


Fig. 2. The locations of line profiles in the mixed plaque phantom. (WC: 100, WW: 500). The black regions inside contrast material regions were caused by air bubbles.

tissue types and ten simulated plaques with different compositions of hard (calcium) and soft (cholesterol-rich) materials and contrast materials (Fig-1). The phantom has multiple empty holes that can be filled with desired concentrations of iodinated contrast material. We diluted the contrast material to have CT number of ~500 HU at 120 kVp, which is high enough to provide good contrast but low enough not to cause severe beam hardening artifacts.

Scanner Model	**NAEOTOM Alpha**
Tube Voltage	120 kVp
Revolution Time	0.5 sec
Spiral Pitch Factor	1
Slice Thickness	0.2, 0.4, 1.5 mm
Total Collimation Width	24 mm
Iterative Reconstruction	QIR: 0, 2, 4
Reconstruction Filter	Qr40, Qr72
Reconstruction Diameter	102 mm
Matrix Size	1024 x 1024
CTDIvol	15 mGy
Pixel Spacing	0.0996 mm

Table-1 – PCCT scan and reconstruction parameters.

B. Image Acquisition

The phantom was scanned on a clinical dual-source PCCT scanner with cadmium-telluride detectors (NAEOTOM Alpha, Siemens Healthcare, Germany). This scanner can operate in standard resolution mode, and UHR mode with apparent detector pixel sizes of 0.2 and 0.50 mm at the isocenter, respectively. The detailed description of the scanner can be found in [4]. Table-1 summarizes the scan and reconstruction parameters. Each scan was repeated three times. All values are reported as mean and SD of the three measurements. While the scanner is dual-source, the UHR scan mode was only available in single-source mode. We used the image reconstructed from all detected photons with energies > 25 keV (T3D) in this study. Hence no spectral analysis was possible in UHR mode.

C. Image Analysis

We investigated the effective spatial resolution of the scans by comparing line profiles drawn in different regions of the plaques as seen in Fig-2. Image noise was measured as the SD of large circular region of interest (ROI) with diameter=20 mm in the difference image of two repetitions of a scan divided by √2.

The noise was measured for images reconstructed FBP and QIR with thinnest (0.2mm) possible slice thickness, as well as 0.4, and 1.5 mm thicknesses, which are more commonly used in the clinic.

III. Results

Figure 3 compares the line profiles for standard (Qr40) and UHR (Qr72) images reconstructed with FBP and QIR. UHR acquisition mode provided sharper delineation of soft and hard plaques, and the lumen albeit with amplified noise. This can be attributed to the reduced blooming artifacts due to the doubling of spatial sampling frequency. Furthermore, UHR showed more accurate CT number values due to the reduced partial volume effects. For example, P13 and P14 pass through air and therefore should have a minimum at -1000HU. The standard mode failed to resolve the true CT number due to partial volume effect, while UHR images all showed the correct CT number. Sample closeup images and profile lines are shown in figures 4,5.

Image noise, as measured by SD of difference image, was approximately 8 time higher in UHR mode compared to standard resolution for images reconstructed with FBP for all slice thicknesses. Images reconstructed with different strengths of QIR reduced the image noise by approximately 2 folds in maximum strength (fig-6).

IV. Discussion

In this preliminary study we compared the spatial resolution and noise of UHR mode compared to standard resolution for imaging various shapes and compositions of calcified plaques. Our study had several limitations which warrant further detailed studies: 1- we used a stationary phantom; future experiments with a dynamic phantom are planned to assess the effect of motion on the effective spatial resolution of the system. 2- the phantom has relatively simple geometries; we plan to assess the image quality in a series of excised heart with various levels and shapes of coronary calcification. 3- We measured image noise as the SD of a large ROI; a more detailed study with noise power spectrum (NPS) measurements is warranted.

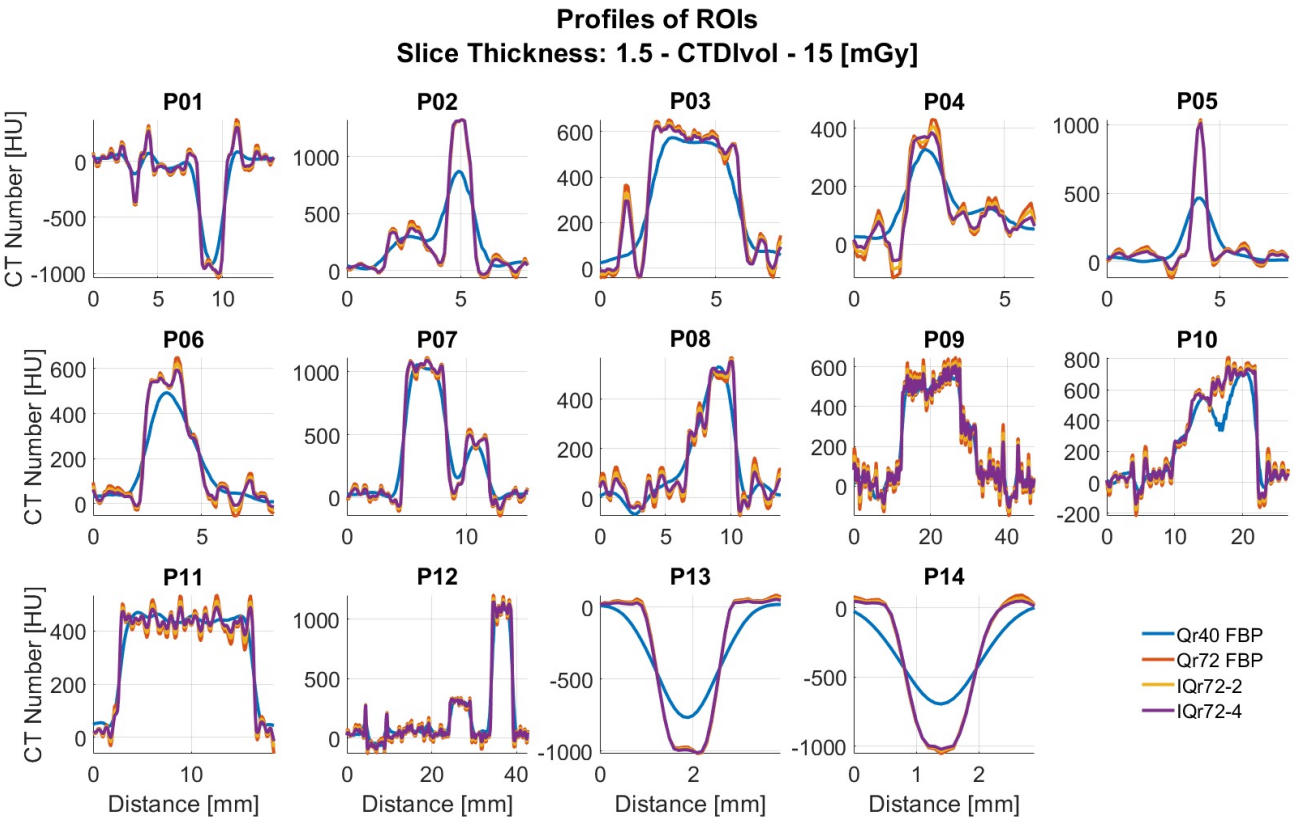


Figure-3- Comparison of line profiles drawn through the regions depicted in fig-2 for various scan and reconstruction modes. standard resolution mode: (Qr40 FBP), UHR mode with FBP (Qr72 FBP), and with QIR of strengths 2 and 4: IQr72-2, IQr72-4. Overall, UHR profiles showed sharper separation between different parts of the plaques due to reduced blooming artifacts. P13 and P14 pass through air and therefore should have a minimum at -1000HU. The standard mode failed to resolve the true CT number due to partial volume effect.

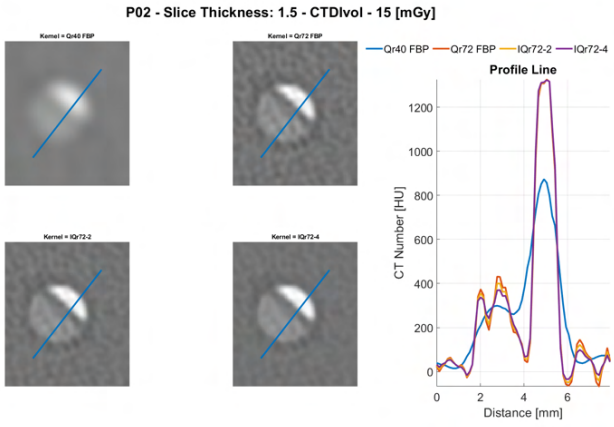


Figure-4- Sample closeup images of a mixed calcified plaque and their corresponding line profiles. Standard resolution image (top left), and UHR images with various strengths of QIR algorithm.

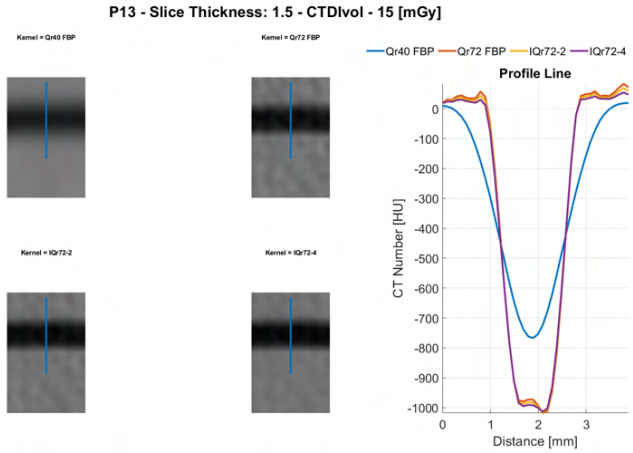


Figure-5- Sample closeup images of a mixed calcified plaque and their corresponding line profiles. Standard resolution image (top left), and UHR images with various strengths of QIR algorithm.

[1] Aug. 2007, Art. no. 061103.

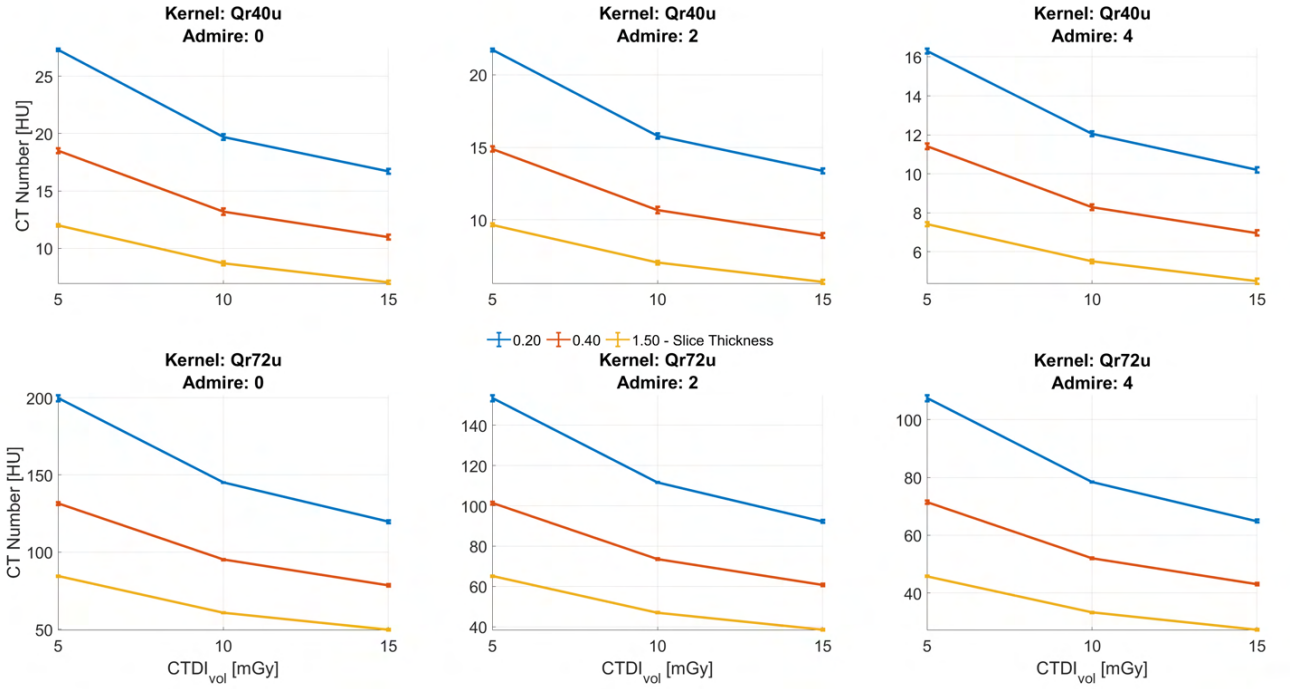


Figure-6- Image noise measured as SD of a uniform ROI with diameter = 20 mm for standard resolution and UHR mode images reconstructed with FBP (strength = 0) and QIR (strength = 2,4) at different slice thicknesses and radiation doses.

V. Conclusion

UHR PCCT improves plaque characterization through improved spatial resolution which results in lower blooming artifacts and partial volume effects. The increase in image noise can be mitigated by using model-based iterative reconstruction algorithms without any loss of spatial resolution. Depending on the imaging task, further noise reduction can be achieved by reconstructing thicker slices. A more detailed investigation with noise power spectrum analysis and observer model studies is warranted.

References

[1] V. Sandfort, M. Persson, A. Pourmorteza, P. B. Noël, D. Fleischmann, and M. J. Willemink, “Spectral photon-counting CT in cardiovascular imaging,” *J. Cardiovasc. Comput. Tomogr.*, 2020.

[2] M. J. Willemink, M. Persson, A. Pourmorteza, N. J. Pelc, and D. Fleischmann, “Photon-counting CT: Technical Principles and Clinical Prospects,” *Radiology*, vol. 289, no. 2, p. 172656, 2018, doi: 10.1148/radiol.2018172656.

[3] A. Pourmorteza, R. Symons, A. Henning, S. Ulzheimer, and D. A. Bluemke, “Dose Efficiency of Quarter-Millimeter Photon-Counting CT: First-in-human Results,” *Invest. Radiol.*, vol. epub ahead, 2018, doi: 10.1097/RLI.0000000000000463.

[4] M. Eberhard *et al.*, “Coronary calcium scoring with first generation dual-source photon-counting CT—first evidence from phantom and in-vivo scans,” *Diagnostics*, vol. 11, no. 9, p. 1708, 2021.

Tuesday, June 14

Invited Talks on Photon Counting CT

Spectral CT

Tuesday Poster Session

Artifacts and Sparse CT

Consistency-based auto-calibration of the spectral model in dual-energy CT

Jérôme Lesaint, Simon Rit

***Abstract*—We propose a consistency-based material decomposition algorithm. The method is free from any calibration procedure. The inverse spectral mixing model is approximated by a polynomial whose indeterminates are the raw-data values and whose coefficients are estimated by minimizing a consistency-based cost function. The consistency is in both the material sinograms and their mono-energetic combination. A small *a priori* on the object is incorporated in the minimization problem as a constraint. The method was evaluated on dual-energy simulations of a numerical phantom made of water and bone.**

***Index Terms*—data consistency conditions (DCCs), spectral CT, material decomposition.**

I. Introduction

This work is related to projection-domain material decomposition of energy-resolved X-ray projections, which aims to decompose energy-resolved projections onto a basis of material specific functions [1]. Since the early work of Alvarez and Macovski [2], it is known that the linear attenuation coefficient μ can be modeled as a linear combination

$$\mu(\vec{x}, E) = \sum_{m=1}^{M} a_m(\vec{x}) f_m(E), \tag{1}$$

of a small number M of energy-dependent basis functions f_m. In Equation 1, f_m can for example be the linear attenuation of the material m (expressed in cm^{-1}) and $a_m(\vec{x})$ the unitless proportion of material m at spatial position $\vec{x}$. In typical photon-counting detectors, several photon counters are maintained at different energy ranges, based on pulse height analysis. We denote B the total number of energy bins. Each detector pixel returns B measurements m_b, modeled by the Beer-Lambert's law:

$$m_b = \int_0^\infty I_b^0(E) \exp\left(- \sum_{m=1}^{M} A_m f_m(E) \right) \mathrm{d}E \tag{2}$$

where $I_b^0(E)$ is the effective spectrum of the bin b and $A_m = \int_{\mathcal{L}} a_m(\vec{x})\, \mathrm{d}\vec{x}$ is the line integral of the material map a_m along the X-ray path $\mathcal{L}$. In other words, A_m is the equivalent length of material m in the object μ along $\mathcal{L}$. In this work, the basis materials will be water and bone, so that $M = 2$ in the sequel.

This work was partially supported by grant ANR-17-CE19-0006 (ROIdoré project) from the Agence Nationale de la Recherche (France). This work was performed within the framework of the SIRIC LYriCAN INCa-INSERM-DGOS-12563 and the LABEX PRIMES (ANR-11-LABX-0063) of Université de Lyon, within the program "Investissements d'Avenir" (ANR-11-IDEX-0007) operated by the French National Research Agency (ANR).

J. Lesaint is with Univ. Grenoble Alpes, CNRS, Grenoble INP, TIMC-IMAG, 38000 Grenoble, France (e-mail : lesaint.jerome@gmail.com).

S. Rit is with Univ. Lyon, INSA-Lyon, UCB Lyon 1, UJM-Saint Etienne, CNRS, Inserm, CREATIS UMR5220, U1206, Centre Léon Bérard, F-69373, Lyon, France.

By material decomposition, we mean recovering the coefficients A_m from the measurements m_b (or from their log s_b, see Equation 4 below). Several approaches have been proposed. A rigourous and natural way is to inverse the forward model $(m_1, ..., m_B) = \Phi(A_1, ..., A_m)$ of Equation 2. This has been done with a maximum likelihood approach in [3] further regularized in [4] or with a regularized least-square approach in [5]. In all cases, the forward model needs to be known and the quality of the decomposition depends on the accuracy of the model. In particular, the effective spectra need to be calibrated, e.g. with a spectrometer for the source spectrum and monochromatic sources for the detector response. To avoid such a cumbersome procedure, it is possible to calibrate a parametric model either of the direct mapping Φ [2] or of the inverse mapping $(A_1, ..., A_M) = \Phi^{-1}(m_1, ..., m_B)$ (or $\Phi^{-1}(s_1, ..., s_B)$). In [6], the measured attenuations are related to the coefficients A_m via a polynomial model. The polynomial coefficients are learnt from a set of calibration measurements at various combinations of basis material lengths, which cover the range of length combinations that will be present in the imaged object. This procedure only requires a specific calibration phantom with known thicknesses of the basis materials but it is a time-consuming procedure. In [7], the authors introduce an empirical dual-energy material decomposition method. It is three-step: first, a calibration phantom, made of the basis materials, is scanned. Second, the reconstructed phantom image is segmented and regions of interest (ROI) of each material are selected. Third, the coefficients of a polynomial approximation of the inverse mapping Φ^{-1} are estimated so that the reconstruction obtained by applying the polynomial coefficients to the measures fits the segmented ROIs. The procedure is called *empirical* because the inverse mapping Φ^{-1} is indirectly estimated to retrieve the A_m from the m_b measurements without knowing I_b^0.

The aim of the project is to avoid the calibration scan in the material decomposition. The polynomial coefficients of the inverse mapping are estimated by enforcing data consistency conditions on the material-specific sinograms. Consistency conditions have been successfully used in a number of CT artefacts correction problems, e.g., geometric calibration, beam-hardening correction, scatter correction. In this paper, we consider 2D parallel geometry only and its corresponding set of consistency conditions known as Helgason-Ludwig consistency conditions. The proposed method does not require a calibration scan, only the scan of the object of interest. The decomposition of the sinograms is exclusively based on the raw data, plus a tiny a priori knowledge on the object.

II. Theory

The method minimizes a consistency-based cost function (subject to some constraints) which is described in this section. For simplicity, we focus on a 2D parallel scanning geometry. Projections are acquired over a 180 degree angular range. In a coordinate system (O, x, y), we denote the projection angle θ (due to discretization, θ is assumed to vary in a set of discrete values Θ, whose cardinal is denoted $|\Theta|$) and the corresponding unit vector $\vec{\theta} = (\cos\theta, \sin\theta)$. The latter indicates the direction of the 1D linear detector, which is placed perpendicular to the direction of the X-rays. Position along the detector is denoted p (again, due to discretization, we denote δp the detector spacing and P the finite set of all pixel positions). Without loss of generality, we assume that the center of the detector is at the origin O of the coordinate system, that it rotates around O and that the object of interest fits the resulting field of view. At projection angle θ, the X-ray line $\mathcal{L}(\theta, p)$ intercepted at position p of the detector has equation $\vec{x} \cdot \vec{\theta} = p$.

A. Photon-counts and attenuations

Assuming a photon counting detector with B bins and according to Equation 2, the photon counts may be corrupted with Poisson noise. The measures become

$$m_b^*(\theta, p) \sim \text{Poisson}(m_b(\theta, p)). \tag{3}$$

We only use $m_b(\theta, p)$ in the rest of the paper and explicitely indicate wether data are corrupted with noise. The projections are then log-transformed according to

$$s_b(\theta, p) = -\log\left(\frac{m_b(\theta, p)}{m_{0,b}(\theta, p)}\right), \tag{4}$$

where $m_{0,b}$ is the number of photons without object.

B. The polynomial model

We look for a a simplified model of the inverse mapping $(A_1, ..., A_m) = \Phi^{-1}(s_1, ..., s_B)$. We choose a polynomial model. Each material sinogram A_m is approximated by a polynomial $\psi_{m,D}$ of degree D in the variables $(s_1, ..., s_B)$. Formally,

$$A_m \approx \psi_{m,D}(s_1, ..., s_B) = \sum_{|k| \leq d} c_m^k s^k \tag{5}$$

where $k = (k_1, ..., k_B)$ is a multi-index, $|k| = k_1 + ... + k_B$ and $s^k = s_1^{k_1} ... s_B^{k_B}$. We define $N(D, B)$ the total number of coefficients of a polynomial of degree D in B variables, e.g., N(2,2)=6 and N(3,2)=10. Note also that the coefficients c_m^k must be determined for each basis material m. If D, B and M are fixed, $M \times N(D, B)$ coefficients have to be determined. For example, if $B = 2$, $M = 2$ and $D = 3$, we seek $2 \times N(3, 2) = 20$ coefficients. Note that the same polynomial is applied to all the pixels of the sinogram A_m, i.e., that the source spectrum and the detector response are uniform over the beam and the detector, respectively.

C. The consistency metric

In 2D parallel geometry, the sought A_m are the Radon transform of the material map a_m

$$A_m(\theta, p) = \int_{\mathcal{L}(\theta,p)} a_m(\vec{x})\,\mathrm{d}\vec{x} = \int_{\mathbb{R}} a_m(p\vec{\theta} + q\vec{\theta}^{\perp})\,\mathrm{d}q \tag{6}$$

where $\vec{\theta} = (\cos\theta, \sin\theta)$ and $\vec{\theta}^{\perp} = (-\sin\theta, \cos\theta)$ are perpendicular.

To account for the spectral nature of the decomposition problem, we combine the material sinograms A_m into mono-energetic sinograms C_n, in view of applying the consistency metric to them. We choose N energy levels E_n in the energy range of the source and form the mono-energetic sinograms C_n

$$C_n(\theta, p) = \sum_{m=1}^{M} f_m(E_n) A_m(\theta, p). \tag{7}$$

The coefficients $f_m(E_n)$ are known (see Equation 1).

A consistency condition of the Radon transform states that the integral of each projection (the order-0 moment) does not depend on the projection angle θ since each of these integrals equals the integral of the attenuation coefficient over the object. We define the moment $J_n(\theta)$ of the C_n by

$$J_n(\theta) = \int_{\mathbb{R}} C_n(\theta, p)\,\mathrm{d}p = \sum_{n=1}^{N} f_m(E_n) \int_{\mathbb{R}} A_m(\theta, p)\,\mathrm{d}p \tag{8}$$

We use the variance of J_n to evaluate if J_n is constant over Θ. The consistency function hence reads:

$$\ell_n(\mathbf{c}) = \frac{1}{|\Theta|} \sum_{\theta \in \Theta} (J_n(\theta) - \overline{J_n(\theta)})^2 \tag{9}$$

where $\overline{J_n(\theta)}$ denotes the mean of J_n over all θ. Note that the function ℓ_n depends on the polynomial coefficients $\mathbf{c} = \left(c_m^k\right)$ since all A_m do (see Equation 5). Finally, we define M consistency functions $\ell'_m(\mathbf{c})$ on the material sinograms A_m in a similar way. The final consistency metric accumulates the consistency of all computed mono-energetic sinograms and all material-specific sinograms

$$\ell(\mathbf{c}) = \sum_{n=1}^{N} \ell_n(\mathbf{c}) + \sum_{m=1}^{M} \ell'_m(\mathbf{c}) \tag{10}$$

The consistency metric would evaluate to zero on perfectly consistent material sinograms.

D. Minimization

Due to the hardening of the beam in each bin (see e.g. [8]), the measured attenuations s_b do not satisfy the consistency condition. The loss ℓ is minimized with respect to the coefficients $\mathbf{c}$ to achieve mono-energetic and material sinograms which are as consistent as possible.

Since we use only order-0 consistency conditions, a constant sinogram (i.e. $A_m(\theta, p) = \text{constant}$ for all θ and p) is perfectly consistent. To prevent the minimization to output such undesirable solution, we follow the idea of [8] and constrain the minimization by some known values. First, if there is no attenuation in all bins ($s_b = 0, \forall b \in \{1, ..., B\}$), we enforce 0 in all material and mono-energetic sinograms by

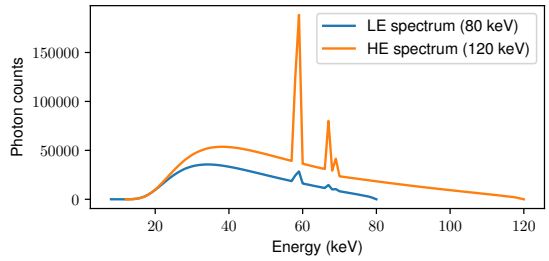


Fig. 1. The low-energy (LE) and high-energy (HE) spectra.

setting $c_m^0 = 0$ for all m. Second, there still is a trivial solution to the minimization of the consistency loss function: the null sinogram. We enforce, for each material m, a particular value in one voxel of each reconstructed material map a_m. To this end, a small sample of each material is placed in the field-of-view and a reconstruction from raw data is computed. The small samples are easily identifiable in the reconstruction. Let $\vec{x}_m$ be one voxel in each material sample and assume reconstructions are computed with a standard Filtered Back-projection (FBP) algorithm. Since FBP is a linear operation, one has

$$a_m(\vec{x}_m) = \text{FBP}\left(\sum_{|k|\leq d} c_m^k s^k\right)(\vec{x}_m) = \sum_{|k|\leq d} c_m^k \text{FBP}(s^k)(\vec{x}_m). \tag{11}$$

The values $\text{FBP}(s^k)(\vec{x}_m)$ are easily computed once, before the minimization. Each material map is constrained by exactly M relations, which take the form

$$\sum_{|k|\leq d} c_m^k \text{FBP}(s^k)(\vec{x}_{m'}) = \begin{cases} 0 & \text{if } m \neq m' \\ 1 & \text{if } m = m' \end{cases} \tag{12}$$

III. Numerical experiments

A. Simulation of data

Numerical experiments used a 2D phantom made of an outer water disc of diameter 32 mm and five bone inserts (with diameters ranging from 2 to 5 mm), placed inside the water disc (see Figure 2). Two tiny inserts of bone and water (1 mm in diameter each) were placed outside the phantom. One voxel in each insert was chosen for the constraints. The material sinograms A_m were analytically computed with RTK [9]. The simulated sinograms had 700 pixels with 0.05 mm spacing and 720 projections over a 180° angular range. Two effective spectra were used. The low-energy (LE) and high-energy (HE) spectra had a tube-voltage of 80 keV and 120 keV respectively (Figure 1). Without object, the detector received a total number of photons of 1.3×10^6 and 2.9×10^6 photons for the LE and HE spectra respectively. The photon counts m_b were computed by applying Equation 2, then log-transformed according to Equation 4.

The degree of the sought polynomials was fixed to $D = 3$ and the consistency metric Equation 10 was minimized under the constraints defined above, with the Sequential Least Squares Programming algorithm. The total number of estimated polynomial coefficients was 18. The initial guess was always set to zero for all coefficients.

B. Evaluation methods

Our method was compared to the calibration from a set of dedicated measurements with the same LE and HE spectra over a set of water and bone lengths. The set covered all combinations which were present in the phantom. More precisely, 100 equi-spaced lengths of each material were measured. Water lengths ranged from 0 to 32.6 mm and bone lengths ranged from 0 to 10.97 mm. All these combinations were irradiated with the same LE and HI spectra as the phantom. Then the polynomials $\tilde{\psi}_{m,D}$ were fitted to the calibration data and further applied to the phantom data to produce a reference polynomial decomposition.

We compared the poly-energetic reconstructions from raw-data with mono-energetic images computed from our DCC-based material maps and from the reference calibrated material maps.

IV. Results

A. Noise-less data

Results of the decomposition are shown in Figure 2. Water and bone are adequately separated. The profiles in Figure 2 (bottom row) indicate residual cross-talk between the two material maps. In the center of the water phantom, the low-constrast feature is visible.

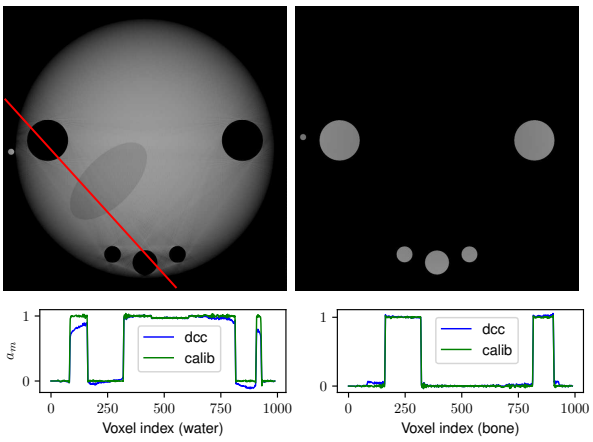


Fig. 2. *Top:* Material map obtained with the DCC-based decomposition. Grayscale is : 1 ± 0.3. The red line indicates the profile used in Figure 4. *Botoom:* Profiles along the red line for the DCC-based and the calibrated water (*left*) and bone (*right*) maps. The consistency metric at convergence was $\ell(\mathbf{c}) = 0.0208$. If evaluated on the calibrated sinograms, the consistency function was 0.0018.

Since the consistency is enforced on the mono-energetic images, Figure 3 compares poly-energetic images, DCC-based and calibration-based mono-energetic images. Poly-energetic images clearly suffer from severe beam-hardening, which is almost completely corrected on both mono-energetic images. The profiles presented in Figure 4 reveal that the DCC-based and calibrated 40 keV images can hardly be distinguished. A slight discrepancy between the 80 keV images still subsists though, especially in the vicinity of the border of the phantom.

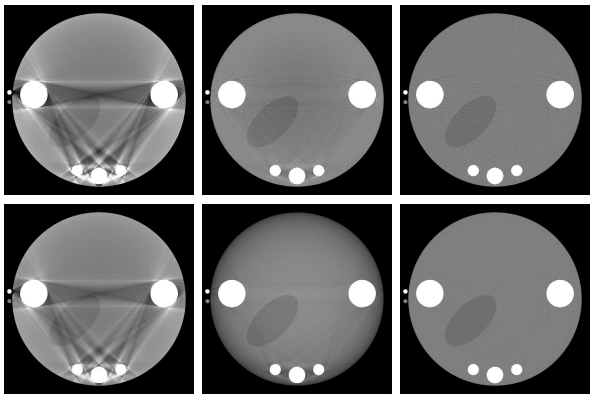


Fig. 3. Poly-energetic reconstructions (*left*) from LE (*top*) and HE (*bottom*) data. DCC-based (*middle*) and calibration-based (*right*) mono-energetic images at 40 keV (*top*) and 80 keV (*bottom*). Grayscale for all images : 0 ± 250HU.

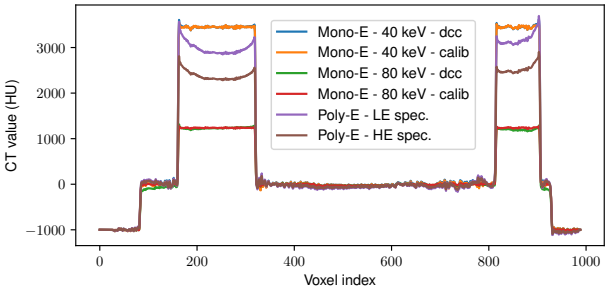


Fig. 4. Profile poly-energetic reconstructions from raw data and from mono-energetic images computed from DCC-based and calibrated material maps.

B. Robustness to noise

The photons count measurements m_b were corrupted with Poisson noise according to Equation 3. The reference photon flux is given by the spectra in Figure 1, i.e. 1.3×10^6 and 2.9×10^6 photons for the LE and HE spectra respectively. The noise level was set by reducing the total number of emitted photons by a factor 1, 10 and 100. The influence of noise on the material maps is presented in Figure 5. The quality of the images is significantly degraded. The consistency function value at convergence increases with the level of noise. It is 0.2108 for reference noise level (factor 1), 1.4738 at factor 10 and 12.815 at factor 100. We expect the choice of the reference voxel to play a critical role in the presence of heavy noise.

V. Discussion and Conclusion

We have demonstrated a consistency-based material decomposition, which does not require any calibration procedure. Only the raw data and a tiny *a priori* knowledge on the object are sufficient to produce material specific sinograms. This tiny *a priori* can be implemented in practice by placing inserts in the field-of-view of the scanner. The resulting sinograms are free from beam-hardening. The method achieves (on simulated data) results that are close to those obtained with a standard calibration-based decomposition method.

The influence of the choice of the reference voxels in the reconstruction may play a critical role and should be further investigated. By choosing a voxel in the reconstruction as a reference, we indirectly incorporate in the constraint all the projections values from lines passing through the voxel (with filtered back-projection, all the lines are incorporated but the ramp filter drops rapidly, so mainly the lines through the voxel are). If those lines "see" a wider range of length combinations, we expect that the decomposition is improved.

Finally, we used a parallel geometry for its simplicity. But order-0 DCC are also available for divergent beam 3D data. In [8], they use such DCC to correct beam-hardening in a circular acquisition. We expect that the decomposition method presented in this work generalizes to multi-energy divergent projections.

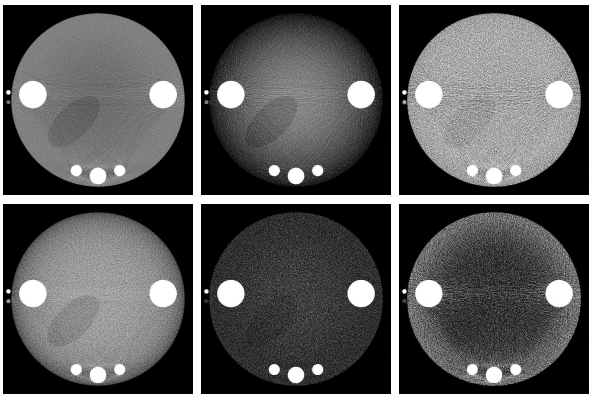


Fig. 5. Mono-energetic images at 40 keV (*top*) and 80 keV (*bottom*), for Poisson noise with photon counts donwscaled by a factor 1, 10 and 100 (from *left* to *right*). Grayscale for all images : 0 ± 250HU. The consistency function at convergence was $\ell(\mathbf{c}) = 0.2108, 1.4738, 12.815$ for factor 1, 10 and 100 respectively.

References

[1] S. Rit, C. Mory, and P. Noël, "Image Formation in Spectral Computed Tomography," in *Spectral, Photon Counting Computed Tomography*. CRC Press, jul 2020, no. 1, pp. 355–372.

[2] R. E. Alvarez and A. Macovski, "Energy-selective reconstructions in X-ray computerized tomography." *Physics in medicine and biology*, vol. 21, no. 5, pp. 733–44, 1976.

[3] E. Roessl and R. Proksa, "K-edge imaging in x-ray computed tomography using multi-bin photon counting detectors," *Physics in Medicine and Biology*, 2007.

[4] B. Brendel, F. Bergner, K. Brown, Kevin and T. Koehler, "Penalized likelihood decomposition for dual layer spectral CT," in *Proc. 4th international meeting on image formation in X-ray CT*, 2016, pp. 41-4.

[5] N. Ducros, J. F. P.-J. Abascal, B. Sixou, S. Rit, and F. Peyrin, "Regularization of nonlinear decomposition of spectral x-ray projection images," *Medical Physics*, vol. 44, no. 9, pp. e174–e187, sep 2017.

[6] R. E. Alvarez, "Estimator for photon counting energy selective x-ray imaging with multibin pulse height analysis," *Medical Physics*, 2011.

[7] P. Stenner, T. Berkus, and M. Kachelriess, "Empirical dual energy calibration (EDEC) for cone-beam computed tomography," *Medical Physics*, vol. 34, no. 9, pp. 3630–3641, aug 2007.

[8] T. Würfl, N. Maaß, F. Dennerlein, X. Huang, and A. K. Maier, "Epipolar Consistency Guided Beam Hardening Reduction-ECC 2," in *14th International Meeting on Fully Three-Dimensional Image Reconstruction in Radiology and Nuclear Medicine*, Xian, China, 2017, pp. 181–185.

[9] S. Rit, M. V. Oliva, S. Brousmiche, R. Labarbe, D. Sarrut, and G. C. Sharp, "The Reconstruction Toolkit (RTK), an open-source cone-beam CT reconstruction toolkit based on the Insight Toolkit (ITK)," *Journal of Physics: Conference Series*, vol. 489, no. 1, p. 12079, 2014.

Direct binning for photon counting detectors

Katsuyuki Taguchi and Scott S. Hsieh

I. Description of Purpose

Photon counting detector (PCD)-based computed tomography (CT) has great potential for many clinical applications [1-3]. However, one of the challenges is spectral distortion in the PCD due to pulse pileup (PP). In this work, we propose a modification of the conventional photon counting technique, *direct binning*, that makes it more resilient to PP. This study consists of two parts: the first on comparing direct binning to the existing scheme in the presence of PP; and the second on the agreement with counting statistics models.

Detection schemes: To handle intense x-ray flux, most of the current PCDs with multiple energy windows count x-ray photons using a relatively simple scheme: pulse height analysis (PHA) with threshold-and-subtract (TS). Each counter counts up-crossing events when pulses exceed the corresponding energy threshold. After a set time period such as 200 μs corresponding to a single view of the sinogram, the counters' data are read out and the outputs of the adjacent energy threshold are subtracted to produce the number of pulse peaks within the two-sided energy window.

As shown in this study, the conventional TS scheme does not function as intended when PP is severe. In this study, we report that a PHA with direct binning (DB) scheme has advantages over a PHA with TS scheme.

Counting statistics models: Accurate statistical models of PCD data are critical for statistical estimation methods. In previous studies, statistical models for the expectation and variance (and covariance) of total counts (and energy window outputs) have been derived for a few simple detection schemes such as non-paralyzable and paralyzable detection schemes. In this study, we assess the agreement between DB and TS schemes with the models.

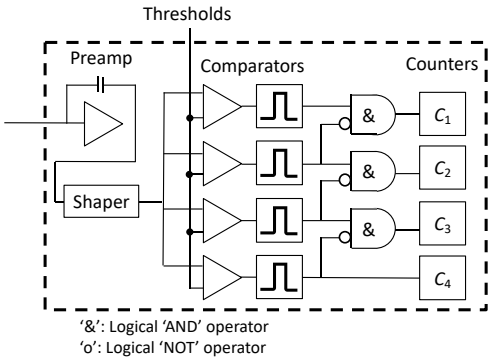


Fig. 1. A diagram for the pulse height analysis (PHA) with direct binning (DB) scheme for one pixel. A timer is associated with each threshold; every up-crossing event changes the status of the threshold from 'waiting' to 'counting,' and starts the timer. When the timer reaches a preset time *Tw*, the status of counter *k* and the inverse status of counter (*k*+1) go through AND operation and adds one count at counter C_k if the outcome is TRUE.

KT is with the Johns Hopkins University School of Medicine (Baltimore, Maryland, USA); SSH is with Mayo Clinic (Rochester, MN).

II. Methods

We outline TS and DB in Sec. II.A, statistical models in Sec. II.B., and the assessment schemes in Sec. II.C.

A. TS and DB

A diagram of PHA with DB scheme is shown in Fig. 1. The only difference from that of PHA with TS is the additional NOT and AND logical operators. Note that DB does NOT actively detect pulse peaks. Many active pulse peak detection schemes—such as taking a derivative of the pulse train and detecting its zeros—may be susceptible to electronic noise. Instead, DB uses up-crossing events as other PHAs and the status of an adjacent comparator to decide whether or not the corresponding counter should be incremented. It is a very simple modification from the PHA with TS.

When PP is severe and the pulse shape has a finite width, the baseline may be elevated to above zero when a new pulse arrives. This baseline elevation and its fluctuation over time may result in strange measurements with TS such as negative counts as shown in Fig. 2.

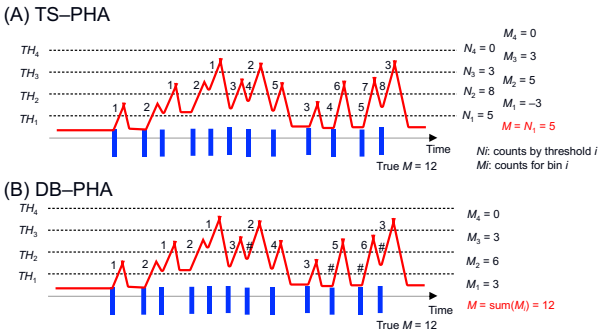


Fig. 2. The same pulse train generated by 12 incident photons (M=12) and processed by PHA with two different schemes, TS (A) and DB (B). The threshold counters of TS output counts (N_k) of 5, 8, 3, and 0 for 4 thresholds, from which the number of pulse peaks for four energy windows (M_k) are computed as –3, 5, 3, and 0, respectively. Notice a negative count (–3) at window 1. The total number of counts is the output of the lowest energy threshold (N_1) and it is 5, which is lower than true counts (M=12). The DB ignores up-crossing events marked by '#' because the above threshold is in 'counting state' when the timer reaches *Tw*. The four DB window outputs (M_k) are 3, 6, 3, 0, respectively. The total counts is a sum of all of the window outputs, and it is 12, which is the same as the true counts (M=12).

B. Statistical models

Let M be the total counts per reading. For both non-

paralyzable and paralyzable detection models, the expectation of M has been derived by Knoll [4] as

$$E(M) = M_{in} \times \frac{1}{1+a\tau} \quad \text{(non-paralyzable)} \quad (1)$$

$$E(M) = M_{in} \times exp(-a\tau), \text{ (paralyzable)} \quad (2)$$

where $a\tau$ is a relative incident count rate (dimensionless) and M_{in} is the number of incident photons per reading. The variance model has been derived by Yu and Fessler [5, 6] as

$$var(M) = M_{in} \times \frac{1}{(1+a\tau)^3}, \quad \text{(non-paralyzable)} \quad (3)$$

$$var(M) = \big(M_{in} \times exp(-a\tau)\big) \times \big(1 - 2a\tau\, exp(-a\tau)\big). \quad \text{(paralyzable)} \quad (4)$$

Let M_k be the output of energy window k per reading. The expectation of window data, $E(M_k)$, has been proposed by for various detection schemes—both non-paralyzable and paralyzable detectors by Taguchi, *et. al.* [7, 8], up-crossing detectors by Roessl, et. al., [9]—and we omit them due to limited space.

To our knowledge there is only one model for variance and covariance of M_k proposed by Wang, et. al., [10]. Wang made the following assumptions: (a) impulse pulse shape, (b) input pulses being not correlated, and (c) non-paralyzable detectors. A finite pulse shape such as Gaussian violates assumption (a); charge sharing violates (b); and paralyzable detectors or PHA DB violates (c). Based on these assumptions, the selection of window k of any pulse can be considered as a multinomial selection of k with a probability of $[b_1, b_1, \ldots, b_k, \ldots]$. Thus, the expectation of can be calculated as

$$E(M_k) = b_k E(M). \quad (5)$$

The variance of M_k depends on M, which is itself a random variable. Using the law of total variance, the variance is given as

$$var(M_k) = E\big(var(M_k|M)\big) + var\big(E(M_k|M)\big) = \cdots$$
$$= b_k E(M) + {b_k}^2\big(var(M) - E(M)\big). \quad (6)$$

Similarly, covariance is given as

$$cov(M_k, M_{k'}) = b_k b_{k'}\big(var(M) - E(M)\big). \quad (7)$$

C. Monte Carlo simulation

A Monte Carlo (MC) simulation was performed to compare two detection schemes: PHA with DB and PHA with TS; the time window Tw for DB was set at 10 ns, which is a half of pulse duration time (20 ns). We used a cadmium telluride PCD with a pixel size of (300 μm)^2, a thickness of 1.6 mm, 4×4 pixels, 4 energy thresholds at (20, 45, 70, and 95 keV), a charge cloud with a diameter of 36 μm, and an asymmetric Gaussian-like pulse shape. The x-ray spectrum was 140 kVp with some window filter and 10 cm of water. Tube current value were 1, 10, 50, 100, 200, 400, …, 2,000 mA and a tube current–time product (mAs) was fixed at 2×10^{-2} mAs/reading (e.g., 200 μs for 100 mA). A sum of 4×4 pixels were computed to produce super-pixel data for each reading. 2,200 noise realizations were used for each mA setting except for 1,600 noise realizations for 1 mA.

The mean, variance, and covariance of MC data, both total counts M and window data M_k, were calculated. A thousand bootstrapping samples were used to their standard deviations.

The Cramér–Rao lower bounds (CRLB) of the following three spectral tasks were computed by adding a thin additional attenuator of water, bone, and a K-edge material (tungsten) and repeated the above–described processes.

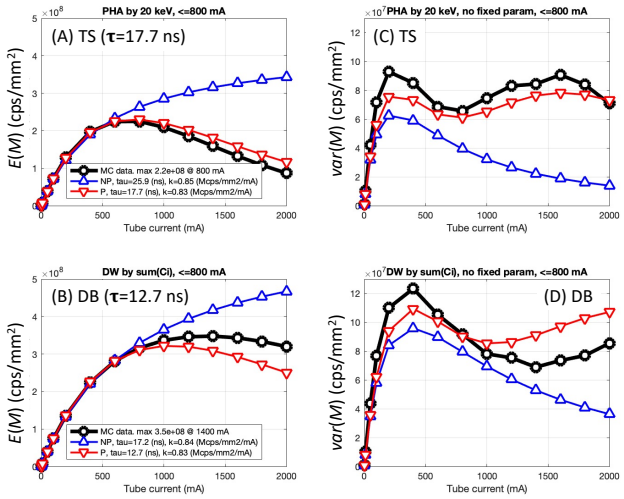


Fig. 3. Count rate curves (A,B) and variance of total counts (C,D). The data up to 800 mA were used to estimate deadtime τ for both paralyzable (P) detector and non-paralyzable (NP) detector, and they were 17.7 ns and 25.9 ns, respectively, for TS and 12.7 ns and 17.2 ns, respectively, for DB. The maximum count rate was 220×10^6 counts per second (cps) per mm^2 for TS and 350×10^6 cps/mm^2 for DB.

III. Results

A. Total counts M

$E(M)$: Overall, DB had a better count rate curve (CRC) than TS (Figs. 3A–3B). Both DB and TS had a better agreement with paralyzable detection model than with non-paralyzable detection model. The estimated detector deadtime for paralyzable detection model was 12.7 ns for DB and 17.7 ns for TS; DB had a 28% shorter deadtime. The maximum count rate was 350×10^6 counts per second (cps)/mm^2 for DB and 220×10^6 cps/mm^2 for TS; DB had 59% larger peak counts.

$var(M)$: Both DB and TS had a reasonably good agreement with paralyzable detection model, and they are better than those with non-paralyzable detection model (Figs. 3C–3D).

Plots of mean and variance of M showed that at 1 mA, the $var(M)$ of MC data were larger than that predicted by the paralyzable model by ~30% due to double-counting with charge sharing, and the differences were diminished as increasing the tube current (Figs. 4A–4B).

B. Window outputs M_k

$E(M_k)$: Probabilities of window k, b_k, were obtained from MC data and Wang model Eq. (5) was used to compute $E(M_k)$. Both DB and TS had an excellent agreement with the model (Figs. 4C–4D). Notice that windows 1 and 2 of TS had negative counts (Fig. 4C), and we believe that it was attributed to elevated baseline due to severe PP, as predicted by Fig. 2.

$var(M_k)$: DB had an outstanding agreement with the model except for window 1 with <500 mA (Fig. 4F). We believe that the discrepancy was attributed to the charge sharing, i.e., double-counting increasing the variance. In contrast, TS had

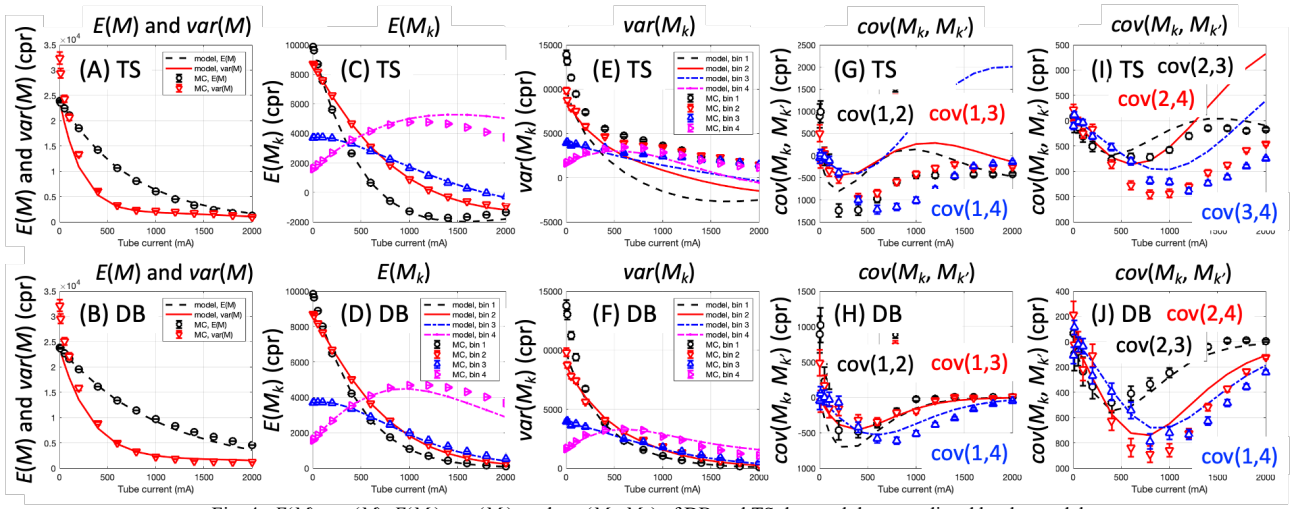


Fig. 4. $E(M)$, $var(M)$, $E(M_k)$, $var(M_k)$, and $cov(M_k, M_{k'})$ of DB and TS data and those predicted by the model

significantly higher variances than the model in general and the agreement was poor (Fig. 4E).

$cov(M_k, M_{k'})$: DB had an excellent agreement with the model (Figs. 4H, 4J); however, TS had a very poor agreement with the model and covariance values were negative and significantly lower (hence, they were larger in magnitude) (Figs. 4G, 4I).

The covariance between window data would be zero when all the data were Poisson distributed. The covariance was positive at lower count rates, as charge sharing was the dominant cause of non-zero covariance. The covariance was negative at higher count rates, as PP was the dominant cause. The break-even point seems to be 100–200 mA.

Overall, DB had a much better agreement with the model and variance and covariance values were lower than TS in general.

C. *CRLB of three spectral tasks*

Conventional CT imaging: Water thickness estimation was used as conventional CT imaging (Figs. 5A, 6A). DB had lower nCRLB values up to 1,100 mA, then higher values >1,100 mA. At 600 mA, DB had 10% lower nCRLB than TS (Fig. 6A).

The nCRLB with 1 threshold (window bin) was qualitatively similar to 4 windows, with differences between DB and TS significantly larger than the 4 windows case.

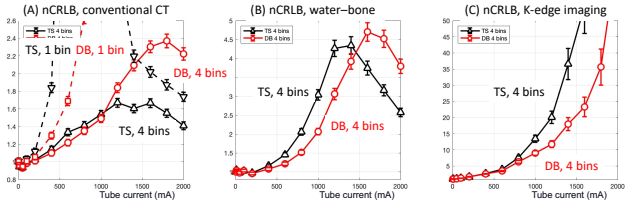


Fig. 5. The normalized CRLB (nCRLB) for 3 spectral tasks

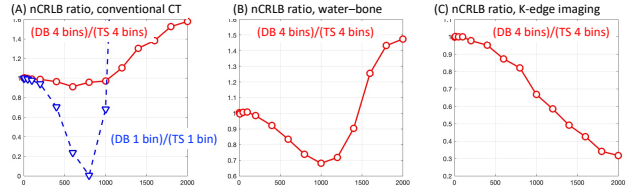


Fig. 6. The ratio of nCRLB for DB to that for TS for 3 spectral tasks

Water–bone material decomposition imaging: DB had lower nCRLB values up to 1,500 mA than TS (Fig. 5B) and at 1,000 mA DB was 30% better than TS (Fig. 6B).

K-edge imaging: DB had better nCRLB values throughout the tube current range investigated (Figs. 5C, 6C), although the nCRLB ratio appeared to be near minimum at 2,000 mA (Fig. 6C). At 2,000 mA, DB was 68% better than TS (Fig. 6C).

As the complexity of spectral tasks increased from conventional CT imaging to K-edge imaging, the advantage of DB over TS increased and so did the tube-current point at which the two schemes crossed over and had equivalent performance.

IV. Discussion

The proposed DB scheme was 28% faster than TS (if measured by deadtime) or 59% faster (if measured by the maximum count rate). Tube current values used in clinic would be up to 1,000 mA. Within this range and under the assumptions of our Monte Carlo model, DB was superior to TS with respect to variance and covariance of total counts and window outputs, normalized CRLBs of three spectral tasks, and the agreement with Wang's model on variance and covariance of window data. DB scheme is relatively simple because only timers and logical operators were added to TS. In contrast, an active peak detection scheme, which takes a derivative of pulse train and aims to find zero, requires additional pulse processing circuitry and may be sensitive to noise.

We found it very interesting that Wang's variance and covariance model agreed with DB data very well, even though any of the three assumptions (a)–(c) used in the model (see Sec. II.B) were satisfied by *neither* the MC simulation nor DB scheme. Having a good agreement with the model may be valuable when constructing a model-based algorithm with PHA DB data.

As shown in this study, TS scheme does not function as intended when PP is severe. There were negative $E(M_k)$, excessive $var(M_k)$, and negatively large $cov(M_k, M_{k'})$. The source of these problems can be understood to be is the

elevation and fluctuation of the baseline.

One may postulate, with no proof provided yet, that a use of baseline restoration (or baseline holder) might be able to decrease, although not eliminate, the negative effects observed with TS. It would also improve the spectral distortion with DB data as well. The results would strongly depend on a specific design of baseline restoration. The variance and covariance of M_k with baseline restoration has never been studied, and it will be an interesting topic of future work.

V. Conclusions

We have studied the PHA with DB scheme and found that DB had several advantages over the conventional scheme, TS. At low to moderate flux, DB could improve the resilience of PCDs to PP. For example, for a case where a 10 cm water object is imaged at 800 mA, our simulations predict that dose efficiency for conventional CT imaging, water–bone material decomposition imaging, and K-edge imaging is improved by 4%, 26%, and 18% respectively.

Acknowledgement

This study has been supported in part by an NIH research grant, R21 EB029739. We are grateful for Ms. Shalini Subramanian for discussion during the course of the study.

References

[1] K. Taguchi and J. S. Iwanczyk, "Vision 20/20: Single photon counting x-ray detectors in medical imaging," *Medical Physics,* vol. 40, no. 10, p. 100901, 2013. [Online]. Available: http://dx.doi.org/10.1118/1.4820371.

[2] K. Taguchi, R. Ballabriga, M. Campbell, and D. G. Darambara, "Photon Counting Detector Computed Tomography," *IEEE Transactions on Radiation and Plasma Medical Sciences,* vol. 6, no. 1, pp. 1-4, 2022, doi: 10.1109/TRPMS.2021.3133808.

[3] K. Taguchi, I. Blevis, and K. Iniewski, *Spectral, Photon Counting Computed Tomography* (Devices, Circuits, and Systems). New York: CRC Press. Taylor & Francis Group, 2020.

[4] G. F. Knoll, *Radiation detection and measurement*, 4th ed. New York: John F. Wiley and Sons, 2010.

[5] D. F. Yu and J. A. Fessler, "Mean and variance of single photon counting with deadtime," *Physics in Medicine and Biology,* vol. 45, no. 7, pp. 2043-2056, 2000. [Online]. Available: http://stacks.iop.org/0031-9155/45/i=7/a=324

[6] D. F. Yu and J. A. Fessler, "Mean and variance of coincidence counting with deadtime," *Nuclear Instruments and Methods in Physics Research Section A: Accelerators, Spectrometers, Detectors and Associated Equipment,* vol. 488, no. 1-2, pp. 362-374, 2002. [Online]. Available: http://www.sciencedirect.com/science/article/pii/S0168900202004606

[7] K. Taguchi, E. C. Frey, X. Wang, J. S. Iwanczyk, and W. C. Barber, "An analytical model of the effects of pulse pileup on the energy spectrum recorded by energy resolved photon counting x-ray detectors," *Medical Physics,* vol. 37, no. 8, pp. 3957-3969, 2010. [Online]. Available: http://link.aip.org/link/?MPH/37/3957/1

[8] K. Taguchi *et al.*, "Modeling the performance of a photon counting x-ray detector for CT: Energy response and pulse pileup effects," *Medical Physics,* vol. 38, no. 2, pp. 1089-1102, 2011. [Online]. Available: http://link.aip.org/link/?MPH/38/1089/1

[9] E. Roessl, H. Daerr, and R. Proksa, "A Fourier approach to pulse pile-up in photon-counting x-ray detectors," *Medical Physics,* vol. 43, no. 3, pp. 1295-1298, 2016, doi: doi:http://dx.doi.org/10.1118/1.4941743.

[10] A. S. Wang, D. Harrison, V. Lobastov, and J. E. Tkaczyk, "Pulse pileup statistics for energy discriminating photon counting x-ray detectors," *Medical Physics,* vol. 38, no. 7, pp. 4265-4275, 2011. [Online]. Available: http://dx.doi.org/10.1118/1.3592932

Co-clinical photon counting CT research for multi-contrast imaging

Cristian T. Badea, Darin P. Clark, Alex Allphin, Juan Carlos Ramirez-Giraldo, Prajwal Bhandari, Yvonne M. Mowery, Ketan B. Ghaghada

***Abstract*—Developing novel contrast agents for multi-energy photon-counting (PC)CT will require a clear translation pathway from preclinical validation to clinical applications. To begin this development, we have used a clinical PCCT scanner (Siemens *NAEOTOM Alpha*) to study the spectral separation of a few contrast elements (Iodine, I; Gadolinium, Gd; Hafnium, Hf; Tantalum, Ta; Bismuth, Bi; Calcium, Ca) with currently available scanning protocols (fixed: 120 kVp, 20 and 65 keV thresholds). We also explored the capabilities of clinical and preclinical PCCT to image mice with hind limb sarcomas injected with I- and Gd-containing nanoparticles (NP). Our results indicate that Ta or Hf are complementary to I or Gd, providing excellent spectral separation for future multi-agent studies. Based on preclinical PCCT with four energy thresholds, we also conclude that additional energy thresholds will benefit clinical PCCT. Furthermore, we demonstrate the role that multi-channel denoising and reconstruction algorithms will greatly benefit maximizing spatial and spectral resolution with clinical PCCT. Performing co-clinical research will facilitate the translation of novel imaging algorithms and NP contrast agents for PCCT.**

I. Introduction

Photon-counting detectors (PCD) can significantly improve CT contrast and enable quantitative material separation with a single CT scan. Because of these advantages, the maturation of PCCT technology promises to enhance routine CT imaging applications with high-fidelity spectral information. This enhancement has been demonstrated in large animal [1] and rabbit [2] studies using prototype clinical hardware. Our group has advanced preclinical PCCT by building prototype systems and demonstrating their value in cancer [3] and cardiac [4] studies in mice. The full potential of PCCT has not yet been realized, since it is dependent on the development and clinical translation of new contrast agents. Here, we investigate spectral separation of high Z element contrast materials using PCCT, and we demonstrate a co-clinical paradigm for PCCT research to translate preclinical validation experiments to clinical applications [5].

II. Methods

Conducting phantom experiments is crucial to the assessment of imaging performance. We next introduce the imaging systems and our multi-element phantom. We then present *ex vivo* imaging studies in mice.

A. Clinical PCCT

A phantom (Fig. 1) containing solutions of various contrast elements was prepared using a stock solution of 10 mg/mL elemental standard to prepare subsequent dilutions (Iodine, I; Gadolinium, Gd; Hafnium, Hf; Tantalum, Ta; Bismuth, Bi; Calcium, Ca). Dilutions were prepared in 2% nitric acid. For imaging studies, 50 mL centrifuge tubes were filled with solutions at concentrations of 10, 5, 2.5, 1.25, and 0.625 mg/mL. Our experiments used an FDA-approved clinical PCCT system (Siemens *NAEOTOM Alpha*) installed at Duke. The scans were performed at 120 kVp with a spiral pitch of 0.8 and used two fixed energy thresholds at 20 and 65 keV. The dose was 16 mGy. The phantom and mouse images were reconstructed with a dual-energy (DE) protocol using a Qr40 kernel and a voxel size of 0.29x0.29x0.4 mm, which yielded 70 keV virtual monoenergetic CT images (VMI) and two binned images (low energy, 20-65 keV; high energy, 65-120 keV). The reconstructed matrix size was 512x512. The reconstructed data was also decomposed on the scanner to create I maps overlaid on virtual non-contrast (VNC) images. The binned images allowed alternative material decompositions.

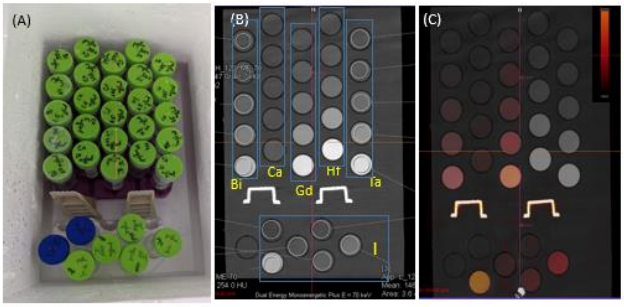


Fig. 1. The elements phantom: (A) a photo, (B) CT image corresponding to VMI at 70 keV and showing measurement circular ROIs, and (C) the iodine decomposition map (red) overlaid on top of the VNC image. Note that the I map also shows positive enhancement for Gd, Bi, and Ca; however, the Hf and Ta vials are clearly separated from I.

A second *ex vivo* mouse scan was performed to achieve higher spatial resolution (0.15x0.15x 0.2 mm, kernel Qr 98u). The radiation dose was increased to 64 mGy to reduce noise. The scanner reconstructs 70 keV VMI images, but does not currently reconstruct separate bin images or perform material

C. T. Badea, D. P. Clark, and A. Allphin are with the Quantitative Imaging and Analysis Lab, Dept. of Radiology while Y. M. Mowery is with the Dept. of Radiation Oncology at Duke University, Durham, NC 27710. P. Bhandari and K. B. Ghaghada are with the Baylor College of Medicine, and Texas Children's Hospital, Houston, Tx.. J. C. Ramirez-Giraldo is with Siemens Healthineers Contact: cristian.badea@duke.edu. This work was supported by the NIH (R01 CA196667, U24 CA220245, RF1 AG070149).

decomposition for this high resolution scanning mode. For both the high and low resolution clinical images we performed post-reconstruction denoising using our rank-sparse kernel regression algorithm [6].

B. PC Micro-CT

Our PC micro-CT system contains two imaging chains based on using both an energy-integrating detector and a PCD [3]. Each chain uses a G-297 X-ray tube (Varian Medical Systems, Palo Alto, CA) with a 0.3 mm focal spot size powered by an Epsilon high-frequency X-ray generator (EMD Technologies, Quebec, Canada). For this work, we have used only the imaging chain with a SANTIS 1604 CdTe-based PCD developed by Dectris, Ltd. (www.dectris.com). This detector has 150-µm pixel size and four energy thresholds. The source-to-detector distances were 831 and 680 mm, giving a magnification of approximately 1.2. To extend the PCD field-of-view along the z-axis and to reduce ring artifacts in our reconstructions, the subject was placed on a translational stage, and scans were performed using a helical trajectory. Our selection of kVp (120) and energy thresholds (20, 34, 50 and 65 keV) was based on the clinical scanner settings. The first and the last threshold (20 and 65 keV) were selected to match the clinical thresholds, while the other two were intended to match I (33.2 keV) and Gd (50 keV) K-edges. We used 200 ms exposures and 1 mA for each projection and 900 projections at each threshold. Based on ionization chamber measurements, the absorbed radiation dose was ~100 mGy. The projection data were air normalized, and the gaps between the tiles of the PCD were filled via interpolation.

Image Reconstruction: We performed an iterative reconstruction using the split Bregman method with the add-residual-back strategy [7] and rank-sparse kernel regression regularization (RSKR [6]), solving the following optimization problem:

$$\mathrm{X} = \underset{\mathrm{X}}{\arg\min} \frac{1}{2}\sum_e \|\mathrm{R}\mathrm{X}_e - \mathrm{Y}_e\|_2^2 + \lambda\|\mathrm{X}\|_{\mathrm{BTV}}. \qquad (1)$$

This algebraic reconstruction problem solves for the vectorized, reconstructed data, the columns of X, for each energy threshold simultaneously (indexed by *e*). The reconstruction for each threshold minimizes the reprojection error (R, system projection matrix) relative to the log-transformed projection data acquired at each threshold (the columns of Y). To reduce noise in the reconstructed results, this data fidelity term is minimized subject to the bilateral total variation (BTV) measured within and between energies via RSKR.

D. Material Decomposition

Extending the approach of Alvarez and Macovski [8], we performed post-reconstruction spectral decomposition both on clinical PCCT and PC micro-CT data. On the clinical DE PCCT data, we studied the following basis material decomposition: a) I and Ca, b) I and Gd and c) I and Ta. Using the PC micro-CT data we have performed a) I and Gd decomposition using 20 and 65 keV threshold images (unsubtracted) and b) photoelectric effect (PE), Compton scattering (CS), I and Gd decomposition using all 4 thresholds images. Vials of known concentrations of each material were used to construct a sensitivity matrix of enhancement per unit concentration (HU/mg/mL). Spectral decomposition was then performed by matrix inversion, followed by sub-space projection to prevent negative concentrations. After decomposition, the material maps were color coded and combined in ImageJ.

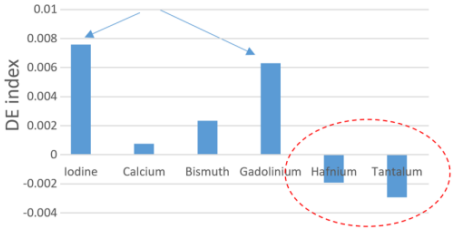


Fig. 2. A bar plot of the DE indices shows highest positive values for I and Gd and negative values for Hf and Ta.

C. Figures of merit

We have assessed our noise performance by measuring the standard deviation in water vials. The sensitivity of DE CT to particular contrast materials can be quantified by the DE index computed as follows:

$$DE\ index = \frac{XE1 - XE2}{XE1 + XE2 + 2000}, \qquad (2)$$

where $XE1, XE2$ are the measured CT numbers normalized by concentration for each element at the two energies.

Another figure of merit for material decomposition used here was the condition number of the sensitivity matrix (after unit normalization of each material's sensitivity vector). The condition number is the ratio of the largest and smallest singular values of a matrix, and it quantifies the potential for error amplification when performing matrix inversion (decomposition). A better spectral decomposition would correspond to a lower condition number.

D. Animal Experiments

For the tumor imaging study, we used a transplant model of soft tissue sarcoma that resembles human undifferentiated pleomorphic sarcoma. A sarcoma cell line was generated from an autochthonous soft tissue sarcoma (p53/MCA model) induced in C57BL/6 wild type mice by intramuscular injection of adenovirus expressing Cas9 endonuclease and sgRNA to *Trp53* gene (Adeno-sgp53-Cas9; Viraquest) followed by intramuscular injection of the carcinogen 3-methylcholenthrene (MCA)[9]. Transplant sarcomas were induced by intramuscular injection of 100,000 cells into the gastrocnemius muscle of C57BL/6 wild type mice Imaging studies were initiated when tumors were palpable (>6 mm diameter). Liposomal-based contrast agents containing iodine (Lip-I) and gadolinium (Lip-Gd) were fabricated similar to methods described previously[10]. Four mice with sarcomas were intravenously injected with Lip-Gd contrast agent (0.32 mg Gd/g body weight) by retro-orbital injection. Three days later (Day 3), the same animals were injected with Lip-I (1.32 mg I/g body weight) and shortly after they were euthanized. The mice were kept in formalin, and scanned with both clinical PCCT and PC micro-CT.

III. Results

A. Elements Phantom

Fig. 2 shows that I and Gd have the highest DE indices for

DE PCCT, and they can be separated best from Bone (Ca). Fig. 2 shows that I and Gd have the highest DE indices for DE PCCT, and they can be separated best from Bone (Ca). On the other hand, if another material could be used, Hf and Ta are best suited as they provide a negative DE index and therefore could be easily separated from either I or Gd. This can be explained by the positions of K-edges, i.e. for Ta (67 keV) and Hf (65 keV) relative to the two thresholds. Original reconstructions and RSKR denoising results are shown in Fig. 3. After RSKR denoising, the noise measured in the water vial decreased from 8.5 to 2.2 HU for the low energy threshold, E1, and from 10.6 to 2.2 HU for the high energy threshold, E2. Subtraction shows no structural information, thus proving the edge preserving performance. For example, denoising for Ta improved the detectable concentration (Rose criterion, CNR > 5) from 4.5 mg/mL to 1.25 mg/mL for E1 and from 6.5 mg/mL to 1.25 mg/mL for E2. The images were used for I/Ca decomposition (condition number: 11.8).

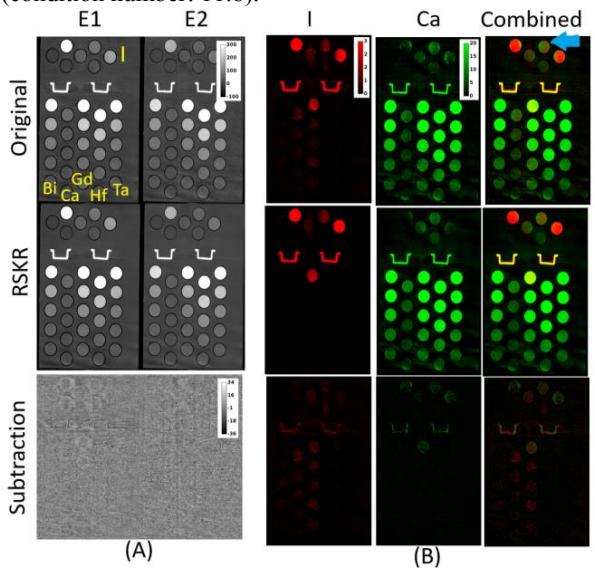


Fig. 3. (A) The original, RSKR denoised images and their subtractions for E1 and E2. (B) In-house material decomposition into I (red) and Ca (green) for original and denoised images. The combined color maps and the subtractions are also shown.

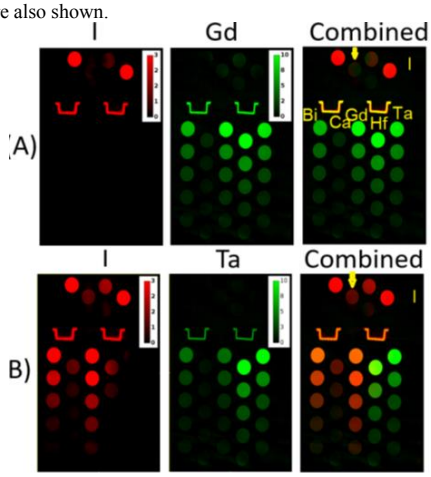


Fig. 4. (A) I/Gd and (B) I/Ta decompositions. Note the better separation of I in the I/Ta decomposition, as predicted by the lower condition number. Note the 1.25 mg/ml I vial (yellow arrow). Hf vials are also very well separated from I. The decompositions were performed on denoised images.

In Fig. 3, the I and Ca maps show levels of cross-contamination for vials of I with concentrations less than 5 mg/ml. Denoising improves the decomposition results; however, cross-contamination (blue arrow) between I and Ca still exists. Most of the other elements are mapped to the Ca basis. Decomposition images for I/Gd and I/Ta are shown in Fig.4. The condition numbers were 13.63 for I/Gd and 5.14 for I/Ta. As confirmed by Fig.4B, the I/Ta provides the best separation with no significant cross-contamination in the I vials and the best visualization of the low concentration vials.

B. Mouse scans

Fig. 5 presents the results for PC micro-CT in one mouse with a sarcoma. The iterative reconstructions with RSKR ensure similar noise levels in all energy bins (std ~ 20 HU). We show decompositions corresponding to two energy thresholds (20, 65 keV; I/Gd maps) and all four energy thresholds (I, PE, CS, Gd). The tumor appears to contain undetectable Gd levels; however, the vessels containing I are enhanced, and the overall quality of the decomposition is increased when using data corresponding to four energy thresholds.

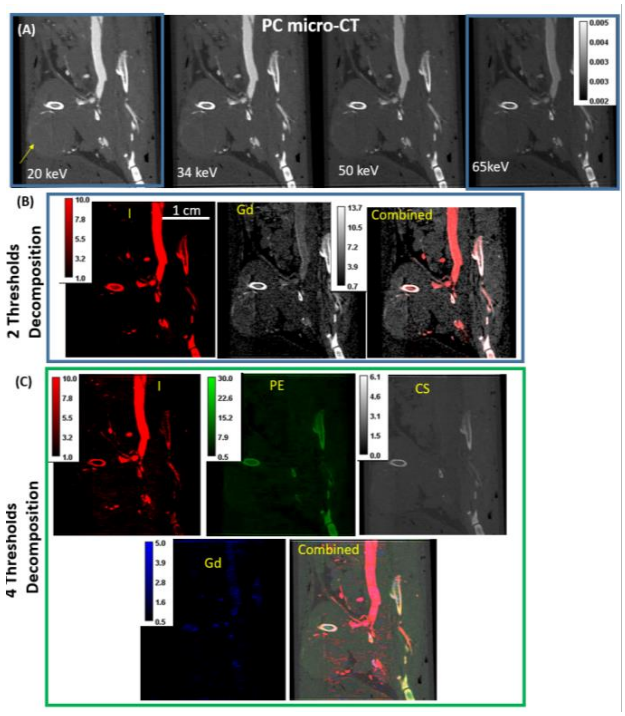


Fig. 5. (A) PC micro-CT images of a mouse with a sarcoma tumor reconstructed with 4 energy thresholds. A yellow arrow indicates the tumor. The vasculature is enhanced by the presence of contrast agents. (B) DE decomposition to I/Gd maps and their overlaid display. (C) The I/PE/CS/Gd decomposition using all 4 energy thresholds.

In Fig. 6 we show a comparison of clinical PCCT and PC micro-CT. The full mouse was scanned with clinical PCCT and the I/Gd decompostion confirms the expected uptake of Gd in the liver and spleen. Note that only the lower part of the body with the tumor was scanned with the PC micro-CT. When using two energy bins, both the clinical and DE PC micro-CT data have been decomposed into I/Gd maps. However, the tumor does not show Lip-Gd accumulation. This may be due to the

lower concentration of Gd injected compared to I. Lip-I is present in the vasculature. The cross-contamination between the I and Gd appears to be higher when using DE PC micro-CT data. The four thresholds PC micro-CT data separated into I/PE/CS/Gd shows higher quality and displays smaller vessels in the tumor that are not visible in the DE PC micro-CT rendering (see yellow arrow, Fig. 6).

Finally, in Fig. 7 we show a higher resolution scan of the mouse (voxel size of 0.15x0.15x 0.2 mm) scanned with the clinical PCCT system. The exisiting protocol provides only a single 70 keV VMI image for this setting. RSKR denoising reduces the noise standard deviation from 115 HU to 29 HU in soft tissue. Note how the resolution of the clinical PCCT approaches the level of resolution of the PC micro-CT scanner (compare with Fig. 5A).

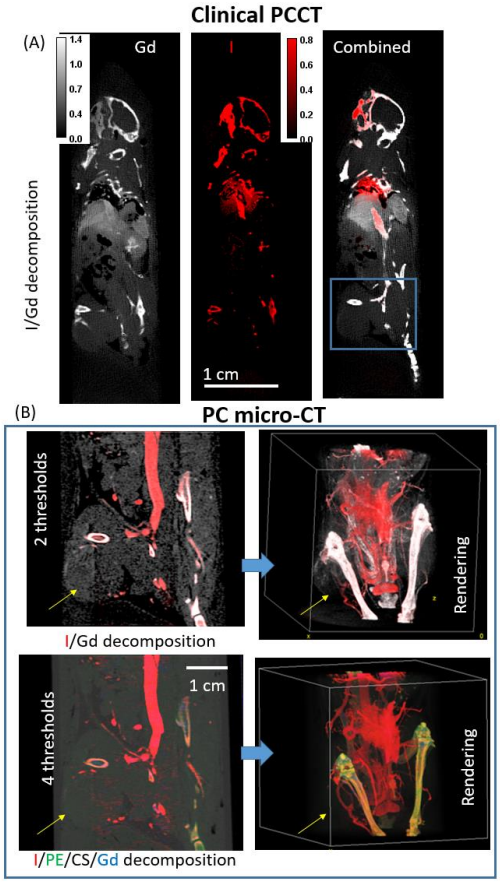


Fig. 6. A comparison of the clinical PCCT (A) and PC micro-CT (B) decompositions using 2 and 4 energy thresholds for the same mouse.

IV. Discussion and Conclusions

Our results indicate that Ta (Z=73, K-edge at 67 keV) and Hf (Z=72, K-edge at 65 keV) appear complementary to I and Gd, providing excellent candidates for multi-agent studies. Their superior performance is explained by their higher Z number and the proximity of their K-edges to the second threshold (65 keV) of the clinical PCCT scanner. Both Ta [11] and Hf [12] have been proposed as contrast agents for CT imaging. Although current spatial and spectral resolution is lower on the clinical PCCT compared to PC micro-CT, denoising and/or iterative reconstruction can enable spectral imaging at higher resolution and with more than two energies. By performing scans on both clinical and preclinical PCCT systems, we can bridge the translational gap for both imaging algorithms and the developments of new contrast agents including NPs that show promise in the field of cancer theranostics (combined therapy and diagnostics).

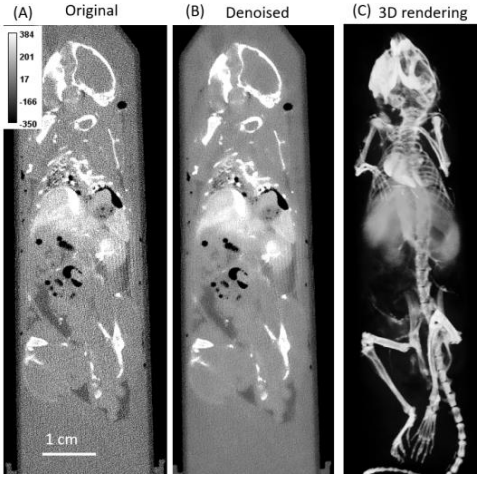


Fig. 7. (A) VMI from clinical PCCT. (B) Post-reconstruction denoising using RSKR. (C) A 3D rendering of the denoised set.

References

[1] R. Symons *et al.*, "Photon-counting CT for simultaneous imaging of multiple contrast agents in the abdomen: an in vivo study," *Medical physics,* vol. 44, no. 10, pp. 5120-5127, 2017.

[2] D. P. Cormode *et al.*, "Multicolor spectral photon-counting computed tomography: in vivo dual contrast imaging with a high count rate scanner," *Scientific reports,* vol. 7, no. 1, p. 4784, 2017.

[3] M. D. Holbrook, D. P. Clark, and C. T. Badea, "Dual source hybrid spectral micro-CT using an energy-integrating and a photon-counting detector," *Physics in medicine and biology,* vol. 65, no. 20, p. 205012, Oct 21 2020, doi: 10.1088/1361-6560/aba8b2.

[4] D. P. Clark, M. Holbrook, C. L. Lee, and C. T. Badea, "Photon-counting cine-cardiac CT in the mouse," *PLoS One,* vol. 14, no. 9, p. e0218417, 2019, doi: 10.1371/journal.pone.0218417.

[5] J. R. Ashton, J. L. West, and C. T. Badea, "In vivo small animal micro-CT using nanoparticle contrast agents," *Frontiers in pharmacology,* vol. 6, p. 256, 2015 2015.

[6] D. P. Clark and C. T. Badea, "Hybrid spectral CT reconstruction," *PloS one,* vol. 12, no. 7, p. e0180324, 2017.

[7] H. Gao, H. Yu, S. Osher, and G. Wang, "Multi-energy CT based on a prior rank, intensity and sparsity model (PRISM)," *Inverse Probl,* vol. 27, no. 11, Nov 1 2011, doi: 10.1088/0266-5611/27/11/115012.

[8] R. E. Alvarez and A. Macovski, "Energy-selective reconstructions in X-ray computerized tomography," *Physics in medicine and biology,* vol. 21, no. 5, pp. 733-44, Sep 1976.

[9] A. J. Wisdom *et al.*, "Single cell analysis reveals distinct immune landscapes in transplant and primary sarcomas that determine response or resistance to immunotherapy," *Nat Commun,* vol. 11, no. 1, p. 6410, Dec 17 2020, doi: 10.1038/s41467-020-19917-0.

[10] S. Mukundan *et al.*, "A Nanoscale, Liposomal Contrast Agent for Preclincal MicroCT Imaging of the Mouse," *AJR,* vol. 186, pp. 300–307, 2006.

[11] J. W. Lambert *et al.*, "An Intravascular Tantalum Oxide-based CT Contrast Agent: Preclinical Evaluation Emulating Overweight and Obese Patient Size," *Radiology,* vol. 289, no. 1, pp. 103-110, Oct 2018, doi: 10.1148/radiol.2018172381.

[12] M. Berger *et al.*, "Hafnium-Based Contrast Agents for X-ray Computed Tomography," *Inorg Chem,* vol. 56, no. 10, pp. 5757-5761, May 15 2017, doi: 10.1021/acs.inorgchem.7b00359.

Reproducibility in dual energy CT: the impact of a projection domain material decomposition method

Viktor Haase, Frédéric Noo, Karl Stierstorfer, Andreas Maier, and Michael McNitt-Gray

***Abstract*—Reproducibility of CT numbers represents an ongoing challenge, especially in clinical applications where exams are performed to assess disease progression or response to therapy. Dual energy CT (DECT) offers an opportunity for improved image quantification, but also presents unique issues. This work presents an initial phantom study that investigates reproducibility over time, across different scanners, and with regard to positioning of the phantom. Both an image domain and a projection data domain material decomposition method are used to create mono-energetic images from DECT data. The scanned object is the ACR CT accreditation phantom. Images were evaluated for reproducibility both inside the phantom inserts of module A as well as in regions between the inserts, where image artifacts are frequently visualized. The results demonstrate that artifacts are worse for off-centered positions. They also demonstrate that the data-based material decomposition provides comparable HU numbers within the inserts of interest like the image-based method, but provides substantially less artifacts and less HU variability in regions surrounding the inserts across the different phantom positions.**

I. Introduction

Reproducibility of CT numbers represents an ongoing challenge, especially in clinical applications where exams are performed to assess disease progression or response to therapy. In these exams, comparison to a baseline or previous exam is critical. When these applications involve quantitative assessments (e.g., lung nodule density, emphysema scoring), the reproducibility of HU across scans becomes critical. DECT offers an opportunity for improved image quantification, but also presents unique issues. Previously, we explored a projection data-based method for material decomposition that uses an analytical energy response model to create mono-energetic images from DECT data [1]. The purpose here is to investigate how well this method performs in terms of artifacts and HU variability in a phantom study. The investigation addresses two potentially challenging conditions: 1) scan repetition on another scanner with a three-year time difference, 2) changes in the position and orientation of the phantom.

II. Background

The methods to create mono-energetic images from DECT data can be grouped into two categories: image-based and data-based material decomposition approaches. In this work, we are comparing both approaches by using a specific realization of each of them which is shortly explained in this section.

V. Haase, and K. Stierstorfer are with Siemens Healthcare GmbH, Germany. A. Maier is with the Department of Computer Science, Friedrich-Alexander-Universität Erlangen-Nürnberg, Germany as well as V. Haase. F. Noo is with the Department of Radiology and Imaging Sciences, University of Utah, Salt Lake City, USA. M. McNitt-Gray is with the Department of Radiology, David Geffen School of Medicine, University of California, Los Angeles, USA.

For the image-based approach, the material decomposition takes place in the image domain, after the high and low energy projection data sets are separately reconstructed. Here, we use an an image-based decomposition approach that follows the steps outlined in [2]. In short, two region of interests (ROIs) are selected, each of which only includes pixels from one of the two basis materials. For each kV setting, the mean attenuation value of both ROIs is computed, then their ratio is compared with a ground truth ratio to identify an effective energy for the scan. Once the effective energy is known for the high and low energy scans, the decomposition proceeds on a pixel-by-pixel basis by inversion of a 2×2 system of equations. The input groups the two pixel values observed from the high and low energy scans. The matrix assembles the linear attenuation coefficient (LAC) of each basis material at the effective energy of the high and low energy scans. The output gives the components along each basis material. Finally, the pixel value in the mono-energetic image is obtained by linearly combining the LAC of the two basis materials at the desired energy using the obtained components.

For the data-based approach, the material decomposition is performed directly from the projection data of the high and low energy scans. The outcome of the decomposition is the line integrals of the images whose pixel values are the components along each basis material. Image reconstruction is taking place after the decomposition and yields the components for each pixel. The final step to create the mono-energetic image is the same as in the image-based approach. In [1], we proposed a data-based approach that uses an analytical energy response model for the CT system. In this work, we use the same method as it has shown robust results using real CT data.

III. Experimental Setup

A. Baseline Experiments

We use module A of the ACR CT accreditation phantom (model 464, Gammex-RMI, USA). Module A is a cylinder of water-equivalent material with a diameter of $20\,\text{cm}$ and a length of $4\,\text{cm}$. It contains five cylindrical inserts representing the X-ray attenuation behaviour of bone, polyethylene, acrylic, air, and water. The chemical composition characteristics of each material were provided by the phantom's manufacturer. In addition, the module shows horizontal ramps for slice thickness evaluation and four $1\,\text{mm}$ diameter steel beads to center the module in the z-direction. In 2018, the phantom was scanned using a state-of-the-art CT system (SOMATOM Definition AS+, Siemens Healthineers). All scans consisted of 2304 projections at a fixed bed position using an in-plane flying focal spot. The detector coverage was defined by 736 channels and 60 rows with a row width of $0.6\,\text{mm}$. The

dual energy scans were acquired sequentially, i.e., the high and low energy scans were performed consecutively, with the X-ray tube set to 120 kV and 300 mAs, and to 80 kV and 500 mAs, respectively. The phantom was placed so that the center of module A closely matches the center of the scan field of view (FOV). The scans were repeated ten times and the average of the repeated scans was taken to reduce noise in the data for easier image analysis.

The collected data was processed with both image-based and data-based material decomposition methods explained in Section II. The decomposition used the water-equivalent and bone-equivalent materials that are used to create the water and bone inserts in the ACR phantom. An offline filtered backprojection implementation of the vendor's own reconstruction software was used for image reconstruction. A semi-smooth kernel (D40s) that is recommended for quantitative applications was systematically applied.

B. Experiments on Reproducibility

New scans were taken three years later, in 2021, using the exact same phantom with a different scanner (SOMATOM Definition Flash, Siemens Healthineers). The scan geometry was the same except for the bowtie filter material and the beam collimation, which was changed to 32 rows with a 1.2 mm row width, so that the energy response model used in the data-based decomposition needed to be adjusted. The phantom was scanned using three different positionings. The first one was the same as that used for the baseline scan, i.e., the phantom is centered in the scan FOV. The second one included a vertical shift of 4.8 cm up. The last one included a counterclockwise rotation of 45° together with a vertical shift of 4.8 cm.

All projection data was processed in the same way as described previously to create mono-energetic images with the image- and the data-based approaches. For the non-centered positionings of the phantom, a shifted grid of pixels was used.

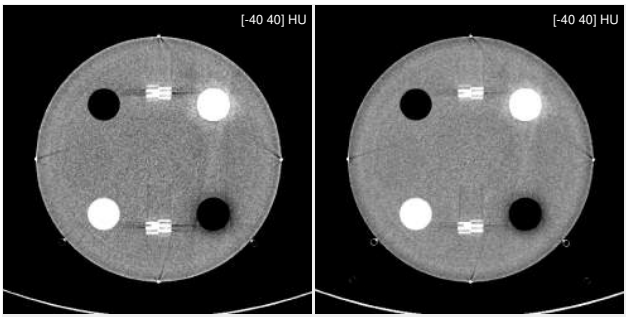
IV. Results

A. Baseline Results

Fig. 1 shows the reconstructions of the high and low energy scans used as baseline. Fig. 2 presents the mono-energetic images at 70 keV, as obtained with the image- and data-based decompositions. This energy was selected because it results in the lowest amount of noise in the images. The image-based decomposition leads to streak artifacts originating from the metal beads and to a bright artifact between the bone and the air insert, which is not present in the data-based result. To convey the quality of the implemented material decomposition approaches, Table I compares the mean attenuation value within the inserts in the mono-energetic images of Fig. 2 against the corresponding ideal attenuation values.

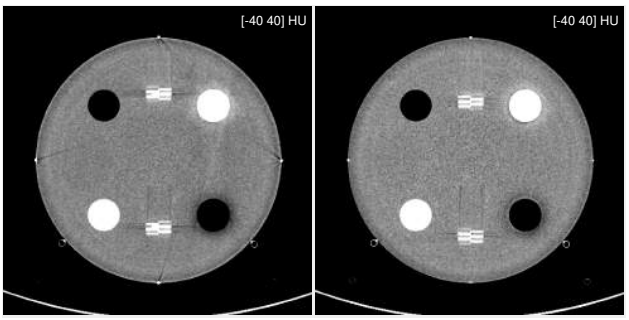
B. Results on Reproducibility

Reconstructions of the new high and low energy scans in the centered, off-centered, and off-centered and rotated positionings are displayed in Fig. 3 as an initial guide to interpret the mono-energetic reproducibility results, which are presented in Fig. 4. For the centered phantom, the images in Fig. 4 have an appearance similar to those from the baseline scans. When the phantom is moved out of the centered position, strong image artifacts around the bone insert can be observed in the image-based approach. These artifacts appear either absent or strongly reduced for the data-based approach.



(a) 80 kV, centered (b) 120 kV, centered

Fig. 1. Reconstructed images of the baseline high and low energy scans.



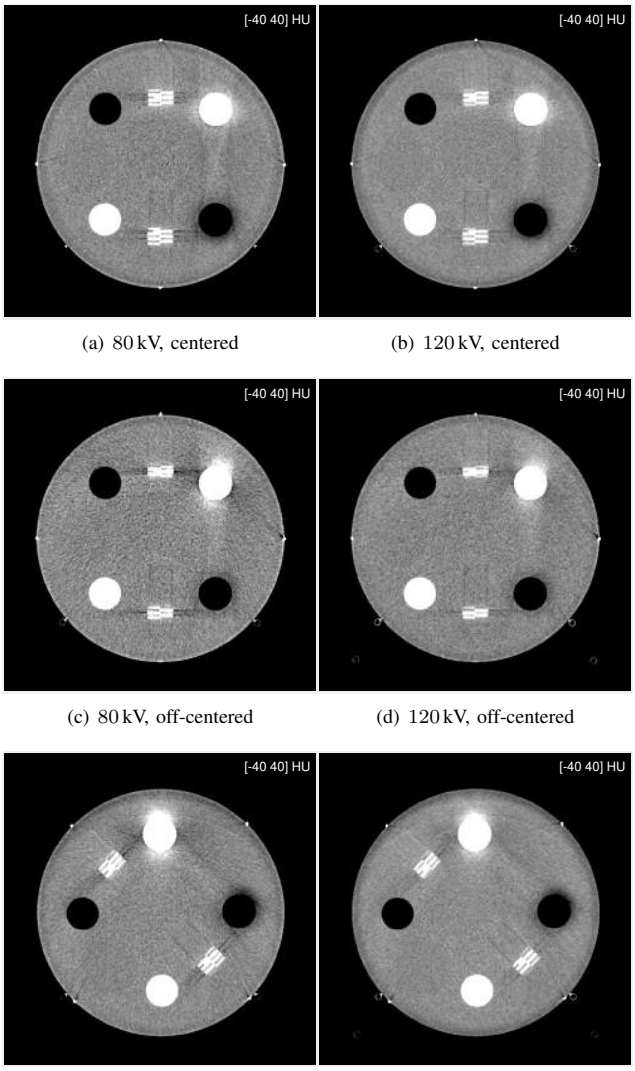
(a) Image-based decomposition (b) Data-based decomposition

Fig. 2. Mono-energetic images at 70 keV as computed from the baseline data, using (a) image-based and (b) data-based material decomposition.

For quantitative evaluation of the images in Fig. 4, Fig. 5 to Fig. 7 provide profile plots for differences with the baseline results. The mono-energetic images were registered to geometrically match the baseline results; the registration used cubic interpolation. Fig. 5 shows a circular profile around the bone insert, which was created by using 360 small ROIs around the bone insert, each with a radius of 5 pixels and placed at the same distance from the center of the bone insert. The differences in mean values for these ROIs are plotted starting at the 12 o'clock position going clockwise.

TABLE I
Mean linear attenuation values within the phantom's inserts in cm^{-1}, as obtained at 70keV using the baseline scan. Reference values given by NIST [3].

# ROI: Material	Reference	Image-based	Data-based
1: Bone	0.3696	0.3735 ± 0.0000	0.3655 ± 0.0001
2: Polyethylene	0.1718	0.1756 ± 0.0000	0.1771 ± 0.0000
3: Acrylic	0.2172	0.2173 ± 0.0000	0.2184 ± 0.0000
4: Air	0.0002	0.0038 ± 0.0000	0.0035 ± 0.0000
5: Water	0.1916	0.1935 ± 0.0000	0.1935 ± 0.0000

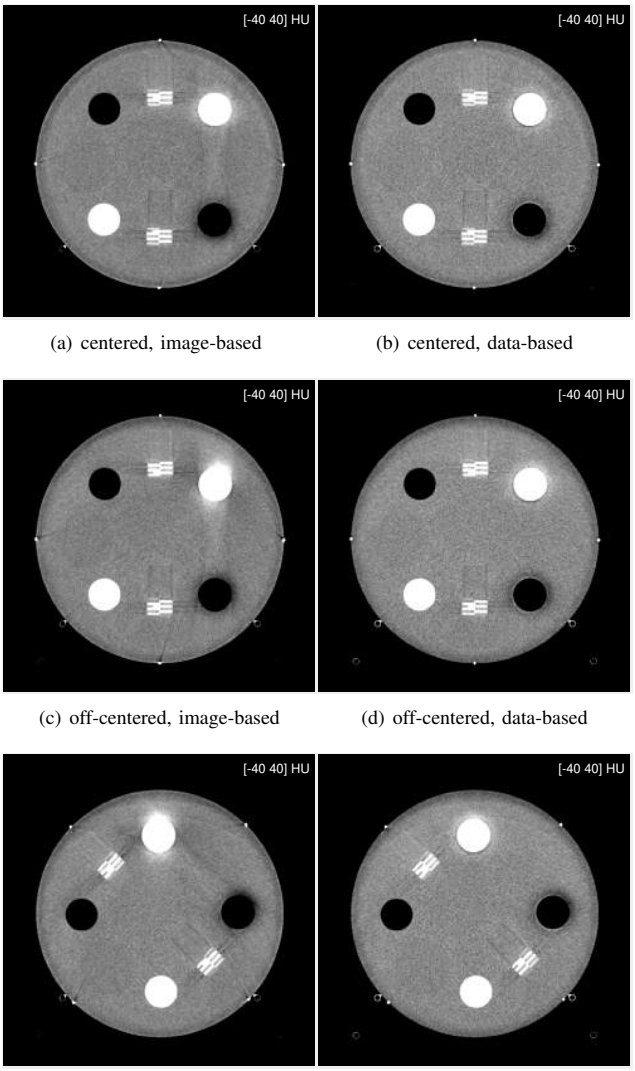


(a) 80 kV, centered (b) 120 kV, centered

(c) 80 kV, off-centered (d) 120 kV, off-centered

(e) 80 kV, off-centered + rotated (f) 120 kV, off-centered + rotated

Fig. 3. Reconstructed images for the new, high and low energy scans for the centered, off-centered, and off-centered and rotated positionings of the phantom.



(a) centered, image-based (b) centered, data-based

(c) off-centered, image-based (d) off-centered, data-based

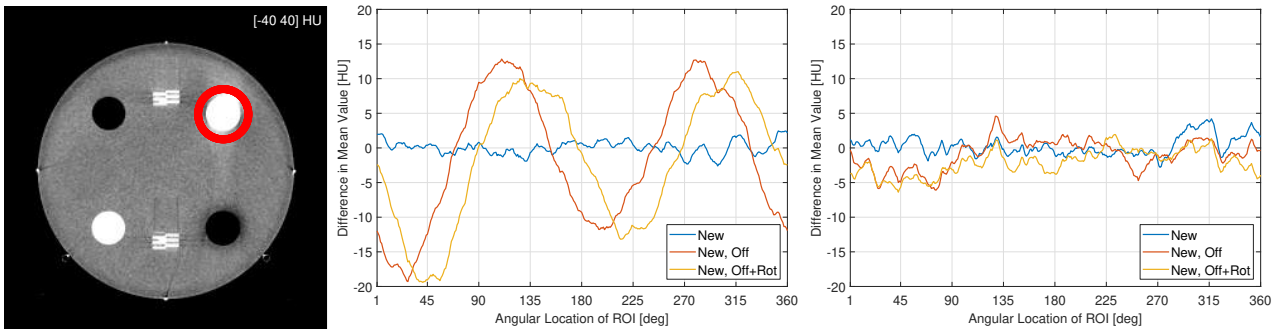
(e) off-centered + rotated, image-based (f) off-centered + rotated, data-based

Fig. 4. Mono-energetic images at 70 keV from the new scans: (left) image-based and (right) data-based material decomposition. Top to bottom: centered, off-centered, and off-centered and rotated positionings of the phantom.

The plots reveal more important deviations for the image-based approach than the data-based approach. For example, the image-based approach yields differences of -20 to 13 HU for the off-centered positioning of the phantom that only are of -6 to 5 HU when the decomposition is performed in the data domain. A second circular plot was created around the air insert in a similar manner as around the bone. Fig 6 shows that errors around the air insert are always in an acceptable range of -5 to 5 HU. It also shows that the results for the image-based approach are slightly better than for the data-based approach: for the off-centered positioning of the phantom, the curve for the data-based approach has a standard deviation of 2.03 HU while that for the image-based approach is 1.21 HU. A third profile plot is presented in Fig. 7, showing a thick horizontal profile, located below the bone insert within the main body of the phantom. Again, higher deviations from the baseline result can be observed when the image-based decomposition is used. We have also compared the mean values within the five inserts. The results (not shown here) convey that both image- and the data-based decomposition yields similar accuracy, with all absolute differences less than 5.0 HU.

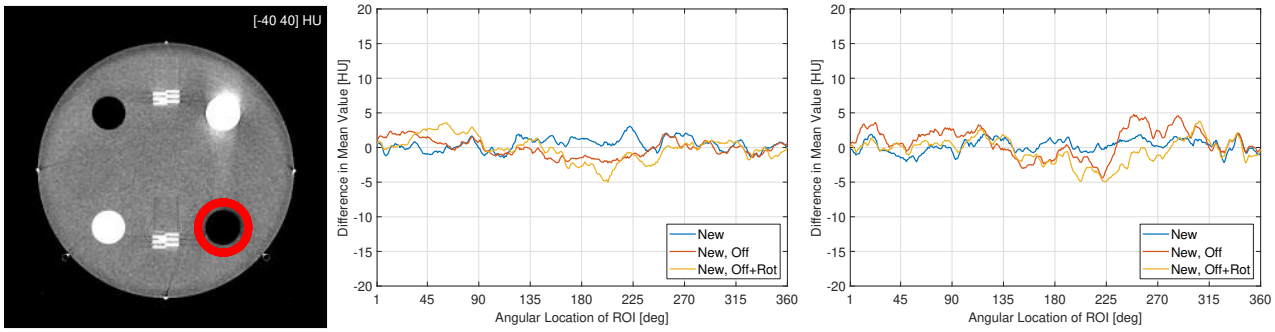
V. Discussion and Conclusion

We presented an initial phantom study investigating artifacts and reproducibility in HU that result from applying a method we previously presented for DECT with material decomposition in the data domain. The investigation included changes over time and scanner, as well changes in phantom positioning. This study demonstrates that our data-based decomposition shows promise as it globally produces more robust results than an image-based material decomposition. Future direction involves extending the method to helical scanning and extending



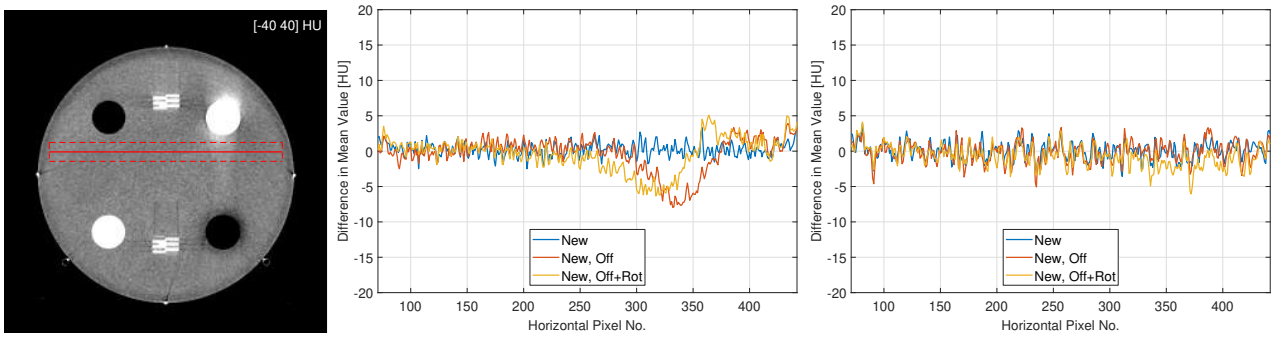
(a) Location of circular profile (b) Image-based decomposition (c) Data-based decomposition

Fig. 5. Quantitative analysis using a circular profile plot around the bone insert. The plots show the difference in mean ROI values relative to the baseline result. The position of the ROIs is highlighted in red in the left image. The angular location of the ROIs starts at the 12 o'clock location and moves clockwise.



(a) Location of circular profile (b) Image-based decomposition (c) Data-based decomposition

Fig. 6. Quantitative analysis using a circular profile plot around the air insert. The plots show the difference in mean ROI values relative to the baseline result. The position of the ROIs is highlighted in red in the left image. The angular location of the ROIs starts at the 12 o'clock location and moves clockwise.



(a) Location of horizontal profile (b) Image-based decomposition (c) Data-based decomposition

Fig. 7. Quantitative analysis using a thick profile that averages values over 31 neighboring horizontal lines within the phantom. In the left image, the central line of the profile is marked with a solid line and the vertical range with dashed lines. The plots show the difference relative to the baseline results.

the investigation to other challenges affecting reproducibility such as physiologic motion and imaging with contrast agent.

Acknowledgment

This project was partly supported by Siemens Healthcare GmbH and partly by the National Heart Lung and Blood Institute of the National Institutes of Health under R01HL153964-01. The concepts and information presented in this paper are based on research and are not commercially available. Its contents are solely the responsibility of the authors and do not represent the official views of the NIH.

References

[1] V. Haase, K. Hahn, H. Schöndube, K. Stierstorfer, A. Maier, and F. Noo, "Image reconstruction from real dual energy CT data using an analytical energy response model," in *2021 IEEE Nuclear Science Symposium and Medical Imaging Conference (NSS/MIC)*. IEEE, 2021.

[2] L. Ren, T. Allmendinger, A. Halaweish, B. Schmidt, T. Flohr, C. H. McCollough, and L. Yu, "Energy-integrating-detector multi-energy CT: Implementation and a phantom study," *Medical Physics*, 2021.

[3] NIST, "National Institute of Standards and Technology: X-ray table," https://www.nist.gov/pml/x-ray-mass-attenuation-coefficients, last accessed: 2022-01-18.

Dual-source photon-counting CT: consistency in spectral results at different acquisition modes and heart rates

Leening P. Liu[1,2], Nadav Shapira[2], Pooyan Sahbaee[3], Harold I. Litt[2], Marcus Y. Chen[4], Peter B. Noël[2]

***Abstract*— Cardiac CT is a useful tool for cardiovascular diagnostics that offers different acquisition modes, each with its advantages. The development of direct converting detector technology has resulted in the clinical translation of dual-source photon-counting CT. This takes advantage of the improved image quality at high heart rates from dual-source CT while making available spectral results for more precise material characterization and quantification. To evaluate the stability of spectral results among different acquisition modes and heart rates, a cardiac motion phantom with a rod mimicking a 50% coronary stenosis was scanned with a dual-source photon-counting CT in three different acquisition modes (retrospective dual-source spiral, prospective dual-source step-and-shoot, dual-source flash spiral) and at different heart rates (60, 80, 100 bpm). Dice scores of stenosed regions relative to a static scan, eccentricity of non-stenosed regions, full width half max, and normalized area under the curve of line profiles were calculated for iodine density maps, and virtual mono-energetic images at 40 and 70 keV. Dice scores and eccentricity were consistent and not significantly affected by acquisition mode or heart rate for spectral results. Full width half max and normalized area under the curve similarly illustrated minor differences between acquisition modes and heart rates. The consistency in these metrics demonstrate preserved image structure and allows for the use of spectral results with high confidence. Dual-source photon-counting CT will enable cardiovascular diagnostics with better material characterization and differentiation.**

***Index Terms*— Photon-counting CT, Cardiac CT, Spectral CT**

I. Introduction

CARDIAC computed tomography (CT) plays an important role in evaluation of heart disease [1]–[3]. These scans acquire high temporal resolution images in one of two ways: spiral, i.e. helical mode, or step-and-shoot. In the spiral/helical mode, images are reconstructed based on a retrospective selection of the phase of the cardiac cycle. This presents more control in time and image quality but comes at the cost of higher radiation exposure [4]. In comparison, step-and-shoot acquisition is a prospectively triggered process that results in reduced radiation doses [5], [6]. However, it has limited use at higher heart rates due to the inability to accurately select the phase of the cardiac cycle with minimal motion [7].

With the development of newer technologies, cardiac CT has further improved with dual-source CT. Dual-source CT adds an additional detector and x-ray tube at almost 90 degrees offset from the other detector and x-ray tube [8]. While a total of 180 degrees and fan angle of rotation are required to acquire images with a single source CT, dual source reduces the required rotation to 90 degrees. This effectively doubles the temporal resolution [8], which reduces motion artifacts to decrease the influence of heart rate on image quality [9], [10]. A third mode utilizes a high pitch dual source helical acquisition; this results in reduced radiation dose compared to spiral/helical mode [11], [12].

Another technology development, spectral CT, improves upon quantification, material characterization, and material differentiation by acquiring attenuation maps from two or more distinct x-ray spectra and generating spectral results, such as iodine density maps and virtual mono-energetic images (VMI) [13], [14]. The newest realization of spectral CT, photon-counting CT (PCCT) [15], in combination with dual-source CT not only has all three acquisition modes for cardiac CT but also allows for the acquisition of spectral results during cardiac CT [16], which was not previously available in cardiac CT. The addition of spectral results to cardiac CT presents an opportunity to advance characterization of heart disease with improved quantification and material differentiation but may be affected by the different acquisition modes and limited by heart rate.

II. Methods

To evaluate the effect of acquisition mode and heart rate on coronary material quantifications using different spectral results, a cardiac motion phantom with a three-dimensional motion simulator was utilized (QRM-Sim4D-Cardio, Quality Assurance in Radiology and Medicine, Möehrendorf, Germany). The phantom included rods submerged in a water tank. Analysis was performed on a rod that mimics a coronary artery filled with contrast and a fibro-fatty plaque that fills half the lumen, resulting in 50% stenosis. Programmed motion combined with a synthetic ECG signal simulated heart rates of 60, 80, and 100 beats per minutes (bpm).

This study was funded by the National Institutes of Health (R01EB030494) and Siemens Healthineers.

L.P. Liu is with the Department of Bioengineering and Department of Radiology, Perelman School of Medicine, Philadelphia, PA, USA. N. Shapira, H.I. Litt, and P.B. Noël are with the Department of Radiology, Perelman School of Medicine, Philadelphia, PA, USA. P. Sahbaee is with Siemens Medical Solutions, Malvern, PA, USA. M.Y. Chen is with the National Heart, Lung, and Blood Institute, National Institutes of Health, Bethesda, MD, USA. (e-mail: leening@seas.upenn.edu).

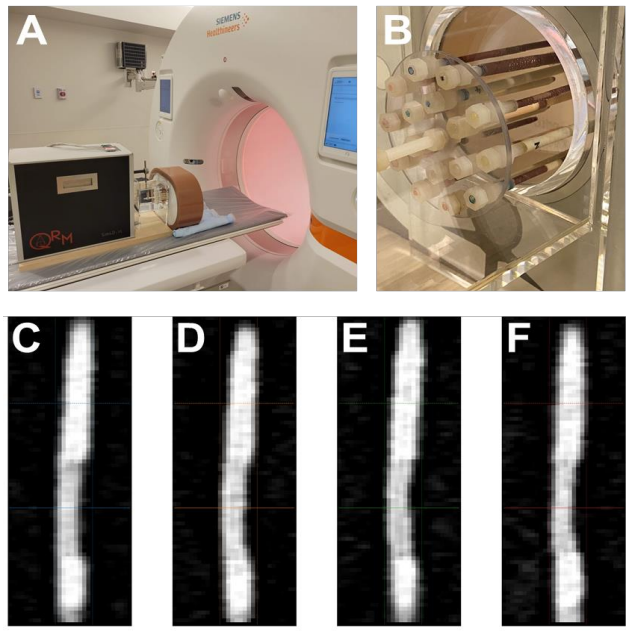


Fig. 1. Experimental setup. (A) A cardiac motion phantom coupled with a 3D motion simulator was scanned with a dual-source photon-counting CT. It contained a variety of artificial coronary arteries containing stenoses of different materials and extents (B). 70 keV VMI images (C-F) illustrate a fibro-fat stenosis within the rod scanned at different heart rates (0, 60, 80, 100 bpm).

The phantom was imaged on a first-generation dual-source PCCT (NAEOTOM Alpha, Siemens Healthineers, Erlangen, Germany) (**Figure 1**). The scanner offers three cardiac acquisition modes: retrospective dual source helical (DS spiral), prospective dual source step-and-shoot (DS sequence), and prospective high pitch dual source helical (flash spiral). In all modes, a standard clinical protocol (**Table 1**) was utilized to obtain images at 120 kVp for each heart rate: 60, 80, 100 bpm. In addition to dynamic scans, a static scan (motion-free) was acquired with a standard clinical protocol. Spectral results, including iodine density maps and VMI at 40 and 70 keV, were generated for each scan.

TABLE I
ACQUISITION PARAMETERS FOR CARDIAC MOTION PHANTOM

	Static	DS spiral	DS sequence	Flash spiral
Tube voltage	120 kVp			
Exposure time [s]	0.25	1.388	0.34	0.078
Spiral pitch factor	1	0.23	N/A	3.2
Exposure [mAs]	250	62	51	48
$CTDI_{vol}$ [mGy]	10	8	8	6
Collimation width	0/4 / 57.6 mm (single/total)			
Slice thickness	0.4 mm			
Convolution kernel	Bv36f			
Field of view	200 mm			
Matrix size	512 x 512			
Pixel spacing	0.390625 mm			

To determine consistency of lumen extent in images with different heart rates and acquisition modes, axial slices of non-stenosed and stenosed regions were extracted from scans. Adaptive thresholding with a Gaussian filter was applied to axial slices to generate a mask of the stenosed region. Using the mask from the static scan as reference, Dice similarity coefficients were calculated for 5 consecutive slices from each acquisition mode and heart rate combination. Dice scores were represented in a scatter plot as mean ± standard deviation. The eccentricity of non-stenosed regions of the rod were also examined. After adaptive thresholding of the non-stenosed region, an ellipse was fit to the contour (**Figure 2**). Eccentricity was calculated as

$$c = \sqrt{1 - \left(\frac{a}{b}\right)^2} \quad (1)$$

where a is the semi-minor axis and b is the semi-major axis. Values were represented in a scatter plot as mean ± standard deviation.

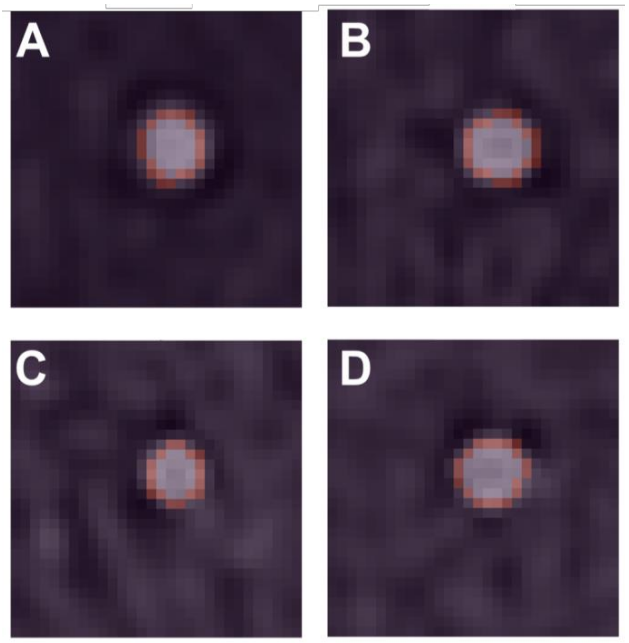


Fig. 2. Fitted ellipses of non-stenosed lumen. Ellipses (pink) matched rods on axial VMI 70 keV images for 0 (A), 60 (B), 80 (C), and 100 (D) bpm. Size and shape of the lumen were similar between the different heart rates with some slight distortion in shape that was reflected in eccentricity.

Line profiles were extracted from coronal slices at the stenosed region to further assess the consistency between acquisition modes and heart rates. A threshold was applied to the line profile to isolate the peak associated with the rod from the background signal. The value of this threshold was calculated as the average + standard deviation of the first and last 5 pixels of the line profile. The full width half max (FWHM) from this peak was calculated to describe the spread in points. Additionally, the normalized area under the curve (AUC) for the peak was determined by evaluating the AUC and normalizing by the maximum value of the peak. Values for both FWHM and normalized AUC were averaged across 5 slices and reported as mean ± standard deviation. For all four metrics, a 2-way ANOVA was implemented to evaluate the effect of acquisition mode and heart rate (60, 80, 100 bpm). A p-value of less than 0.05 was considered significant.

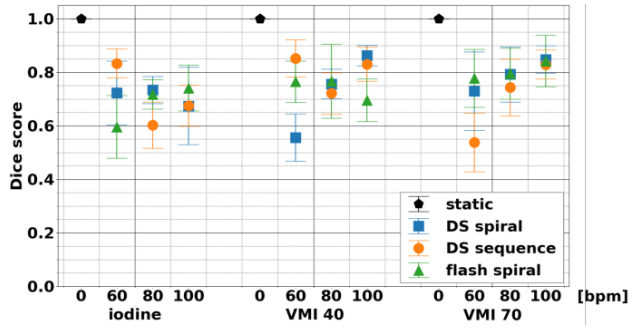


Fig. 3. Dice similarity coefficient of stenosed areas of coronary phantoms at different spectral results, acquisition modes, and heart rates. High similarity between Dice scores of different heart rates and acquisition modes demonstrates consistency.

III. Results

Dice scores (**Figure 3**) were similar between 60, 80, and 100 bpm for each of the acquisition modes and spectral results. The maximum average difference in Dice score between the three heart rates was 0.11, 0.20, and 0.19 for iodine density maps, VMI 40 keV, and VMI 70 keV, respectively. Of note, Dice scores from high-pitch spiral scans deviated less than Dice scores from other acquisition modes. There was no significant effect of acquisition mode or heart rate on the Dice score for iodine density maps (0.79, 0.69) and VMI 40 keV (0.67, 0.10) but not for the Dice score for VMI 70 keV (0.039, 0.002). In general, the high level of similarity illustrated image consistency of stenosed regions of the coronary artery phantom.

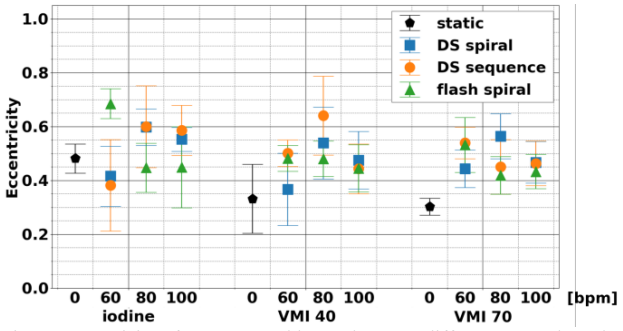


Fig. 4. Eccentricity of non-stenosed lumen between different spectral results, acquisition modes, and heart rates. Shape of non-stenosed lumen was not affected by acquisition modes or heart rates. The stability applied to all three spectral results examined: iodine, VMI 40 keV, and VMI 70 keV.

Eccentricity (**Figure 4**) was independent of heart rate and acquisition mode across the different spectral results. The shape of the non-stenosed area in dynamic scans demonstrated similar properties to those observed in static scans, with maximum magnitude of the average difference relative to the static scan of 0.05, 0.20, and 0.19 for iodine density maps, VMI 40 keV, and VMI 70 keV, respectively. These deviations correspond to small changes related to motion blurring. Even so, for the three different acquisition modes, eccentricity was not significantly different for different heart rates and acquisition modes for each of the three spectral results. This highlights that consistency

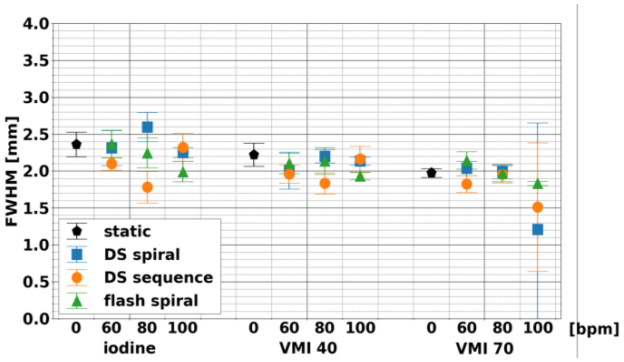


Fig. 5. Effect of acquisition modes and heart rates on full width half max of line profiles of coronary phantom. Between acquisition modes and heart rates, results were stable and did not vary for each spectral result.

was present at both non-stenosed and stenosed regions of the rod.

FWHM (**Figure 5**) at different acquisition modes and heart rates for iodine density maps, VMI 40 keV, and VMI 70 keV illustrated the stability in the sharpness of rod in comparison to its background. This stability was apparent with the maximum magnitude of the average difference relative to the static scan, which amounted to 0.30, 0.24, and 0.23 for iodine density maps, VMI 40 keV, and VMI 70 keV. Moreover, acquisition modes and heart rates did not significantly affect FWHM with the exception of acquisition mode for iodine density maps. Similarly, normalized AUC (Figure 6) was only significantly affected by acquisition mode or heart rate for iodine density maps, thus demonstrating stability between heart rates for each acquisition mode and most spectral results.

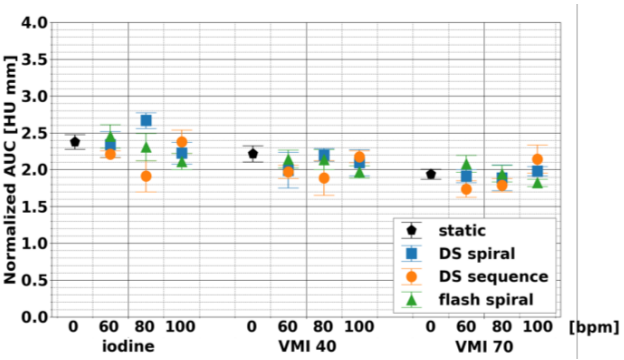


Figure 6. Comparison of normalized area under of the curve of line profiles of stenosed lumen in coronary artery phantom. Normalized area under of the curve varied little across heart rates and acquisition modes for each spectral result.

IV. Discussion

This systematic evaluation of the three acquisition modes for cardiac CT demonstrated stability at different heart rates and acquisition modes for iodine density maps, VMI 40 keV, and VMI 70 keV. This stability can be extended to other spectral results that were not examined here, and the minimal effect of acquisition mode and heart rate provides confidence in visualization and quantification of these spectral results. Moreover, this is the first-time spectral results are available for

a wide range of heart rates with dual-source CT. One of the main applications for these spectral results is examining and characterizing coronary artery plaque, which can be composed of different materials, i.e. calcium, fibro-fatty, or fatty tissue. As demonstrated in dual-energy CT, each spectral result plays a different role in material differentiation: iodine density maps isolate contrast from other materials, while virtual mono-energetic images at different keVs have specific advantages, such as increased contrast, reduced beam hardening, and metal artifact reduction. As a result, the stability of spectral results in dual-source PCCT augurs well for their utility in material quantification and characterization.

V. Conclusion

This initial study of dual-source photon-counting CT in cardiac mode illustrates potential for improved cardiovascular diagnostics. Clinical studies are an essential next step to establish PCCT in the clinical routine.

References

[1] S. Ulzheimer and W. A. Kalender, "Assessment of calcium scoring performance in cardiac computed tomography," *European radiology*, vol. 13, no. 3, pp. 484–497, Mar. 2003, doi: 10.1007/S00330-002-1746-Y.

[2] M. J. Budoff *et al.*, "Assessment of coronary artery disease by cardiac computed tomography: A scientific statement from the American Heart Association Committee on Cardiovascular Imaging and Intervention, Council on Cardiovascular Radiology and Intervention, and Committee on Cardiac Imaging, Council on Clinical Cardiology," *Circulation*, vol. 114, no. 16, pp. 1761–1791, Oct. 2006, doi: 10.1161/CIRCULATIONAHA.106.178458.

[3] G. L. Raff, M. J. Gallagher, W. W. O'Neill, and J. A. Goldstein, "Diagnostic Accuracy of Noninvasive Coronary Angiography Using 64-Slice Spiral Computed Tomography," *Journal of the American College of Cardiology*, vol. 46, no. 3, pp. 552–557, Aug. 2005, doi: 10.1016/J.JACC.2005.05.056.

[4] J. Hausleiter *et al.*, "Estimated radiation dose associated with cardiac CT angiography," *JAMA*, vol. 301, no. 5, pp. 500–507, Feb. 2009, doi: 10.1001/JAMA.2009.54.

[5] J. P. Earls *et al.*, "Prospectively Gated Transverse Coronary CT Angiography versus Retrospectively Gated Helical Technique: Improved Image Quality and Reduced Radiation Dose1," *https://doi.org/10.1148/radiol.2463070989*, vol. 246, no. 3, pp. 742–753, Mar. 2008, doi: 10.1148/RADIOL.2463070989.

[6] P. Stolzmann *et al.*, "Dual-source CT in step-and-shoot mode: Noninvasive coronary angiography with low radiation dose," *Radiology*, vol. 249, no. 1, pp. 71–80, Oct. 2008, doi: 10.1148/RADIOL.2483072032/ASSET/IMAGES/LARGE/R08SE22T04X.JPEG.

[7] J. Hsieh, J. Londt, M. Vass, J. Li, X. Tang, and D. Okerlund, "Step-and-shoot data acquisition and reconstruction for cardiac x-ray computed tomography," *Medical Physics*, vol. 33, no. 11, pp. 4236–4248, Nov. 2006, doi: 10.1118/1.2361078.

[8] T. G. Flohr Cynthia H McCollough Herbert Bruder Martin Petersilka Klaus Gruber Christoph Süβ Michael Grasruck Karl Stierstorfer Bernhard Krauss Rainer Raupach Andrew N Primak Axel Küttner Stefan Achenbach Christoph Becker Andreas Kopp Bernd M Ohnesorge *et al.*, "COMPUTER TOMOGRAPHY First performance evaluation of a dual-source CT (DSCT) system," *Eur Radiol*, vol. 16, pp. 256–268, 2006, doi: 10.1007/s00330-005-2919-2.

[9] S. Achenbach *et al.*, "Contrast-enhanced coronary artery visualization by dual-source computed tomography—Initial experience," *European Journal of Radiology*, vol. 57, no. 3, pp. 331–335, Mar. 2006, doi: 10.1016/J.EJRAD.2005.12.017.

[10] U. Ropers *et al.*, "Influence of Heart Rate on the Diagnostic Accuracy of Dual-Source Computed Tomography Coronary Angiography," *Journal of the American College of Cardiology*, vol. 50, no. 25, pp. 2393–2398, Dec. 2007, doi: 10.1016/J.JACC.2007.09.017.

[11] F. Morsbach *et al.*, "Performance of turbo high-pitch dual-source CT for coronary CT angiography: First ex vivo and patient experience," *European Radiology*, vol. 24, no. 8, pp. 1889–1895, May 2014, doi: 10.1007/S00330-014-3209-7/FIGURES/5.

[12] S. Achenbach *et al.*, "High-pitch spiral acquisition: A new scan mode for coronary CT angiography," *Journal of Cardiovascular Computed Tomography*, vol. 3, no. 2, pp. 117–121, Mar. 2009, doi: 10.1016/J.JCCT.2009.02.008.

[13] D. Caruso *et al.*, "Can dual-energy computed tomography improve visualization of hypoenhancing liver lesions in portal venous phase? Assessment of advanced image-based virtual monoenergetic images," *Clinical Imaging*, vol. 41, pp. 118–124, Jan. 2017, doi: 10.1016/J.CLINIMAG.2016.10.015.

[14] S. Ehn *et al.*, "Assessment of quantification accuracy and image quality of a full-body dual-layer spectral CT system," *Journal of Applied Clinical Medical Physics*, vol. 19, no. 1, pp. 204–217, 2018, doi: 10.1002/acm2.12243.

[15] L. P. Liu, N. Shapira, P. Sahbaee, M. Schnall, H. I. Litt, and P. B. Noël, "First-generation clinical dual-source photon-counting CT: ultra-low dose quantitative spectral imaging," *medRxiv*, p. 2021.11.17.21266432, Dec. 2021, doi: 10.1101/2021.11.17.21266432.

[16] N. R. van der Werf *et al.*, "Dose Reduction in Coronary Artery Calcium Scoring Using Mono-Energetic Images from Reduced Tube Voltage Dual-Source Photon-Counting CT Data: A Dynamic Phantom Study," *Diagnostics 2021, Vol. 11, Page 2192*, vol. 11, no. 12, p. 2192, Nov. 2021, doi: 10.3390/DIAGNOSTICS11122192.

Deep Scatter Estimation for Coarse Anti-Scatter Grids as used in Photon-Counting CT

Julien Erath, Jan Magonov, Joscha Maier, Eric Fournié, Martin Petersilka, Karl Stierstorfer, and Marc Kachelrieß

***Abstract*—Due to the smaller detector pixels in photon-counting CT, coarse anti-scatter grids are used. This may lead to high frequencies in the scattered radiation and therefore moiré artifacts in the reconstructed images can occur. It has been shown that deep convolutional neural networks are very effective to correct scatter artifacts in clinical CT. In this work we present an adapted version of the deep scatter estimation (DSE) to correct for the high frequency artifacts effectively. With the use of DSE the mean absolute error of the scatter artifacts is reduced from about 8 HU to under 1 HU. At the same time the moire artifacts can be prevented and additional post-processing in the image can be avoided.**

I. Introduction

Recently the U.S. Food and Drug Administration (FDA) has cleared the world's first photon-counting computed tomography (CT) scanner, the Siemens NAEOTOM Alpha. The FDA described it at the first major new technology for computed tomography imaging in nearly a decade. Since its introduction, photon counting detectors have shown a great potential to improve the spectral and spatial resolution of CT imaging [1]. The spectral information can be used for several different image techniques, for example virtual monochromatic imaging or iodine overlays. These possibilities can improve the clinical diagnostic of CT. To avoid the pulse pile-up effect and to improve the spatial resolution, the detector of the new photon-counting CT has very small detector pixels [2]–[4]. This can be useful to visualize, for example, small pulmonary vessels [5] or to show details of the inner ear [6].

Scattered radiation and the resulting image artifacts are well-known in clinical CT. Generally, there are two strategies to minimize the effect of scatter, first to reduce the measured scattered radiation with special hardware like anti-scatter grids (ASG) and second to correct the scatter in the pre-processing of the image reconstruction. To achieve the best possible image quality these two techniques should be combined.

The use of the smaller detector pixels influences the effect of the scattered radiation in the detectors. For energy integrating detector each pixel is surrounded by the lamella of the ASG. Since with the photon counting CT the detector pixels are much smaller, several pixels are combined between the lamellae of the ASG (left side in figure 1). This results in the fact that, dependent on the location and the angle of the incoming scattered photon, the scatter intensity can vary between the adjacent pixels in one ASG block. On the right side of figure 1 it can be seen that, depending on the angle of incidence, some photons will be absorbed by the anti-scatter grid and other will be measured in the detector. For energy-integrating detectors scattered intensities are usually low-frequency and only slowly change across the detector. These additional high frequencies in the scattered radiation can lead to ring artifacts in the reconstructed CT-images, as shown in figure 2.

Julien Erath and Jan Magonov are with Siemens Healthineers (Forchheim, Germany), the German Cancer Research Center (DKFZ) (Heidelberg, Germany) and with the Heidelberg University (Heidelberg, Germany). Eric Fournié, Dr. Martin Petersilka and Dr. Karl Stierstorfer are with Siemens Healthineers (Forchheim, Germany). Dr. Joscha Maier and Prof. Dr. Marc Kachelrieß are with the German Cancer Research Center (Heidelberg, Germany) and with the Heidelberg University (Germany). Corresponding author: Julien Erath (julien.erath@dkfz.de)

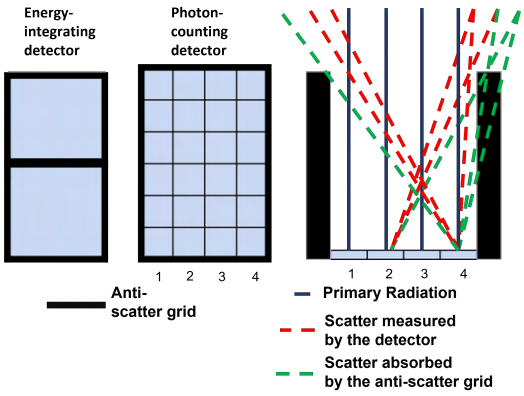


Fig. 1: In energy integrating detectors each pixel is surrounded by the lammela of the anti-scatter grid. In a photon counting detector there are several pixels between the lamellae. Depending on the direction of the incoming scatter photon, it will be absorbed by the anti-scatter grid. Thus two neighboring pixels will receive a different scatter intensity.

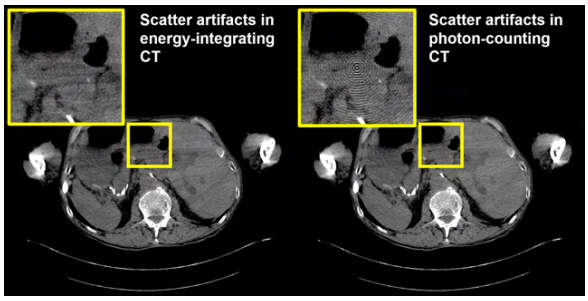


Fig. 2: Illustration of the moiré effect due to the use of coarse anti-scatter grids in photon-counting CT.

We have recently shown the efficient use of deep convolutional neural networks (CNN) to correct for scatter artifacts

in clinical CT with the deep scatter estimation (DSE) [7]–[9]. Once trained, the CNN's can lead to a very robust and fast scatter estimation. In this work, we propose the use of neural networks to correct for the artifacts resulting from coarse anti-scatter grids. Based on our previous work, we have modified the architecture of our neural network to estimate the new high frequencies in the occurring scattered radiation.

II. Methods and Material

A. Data generation

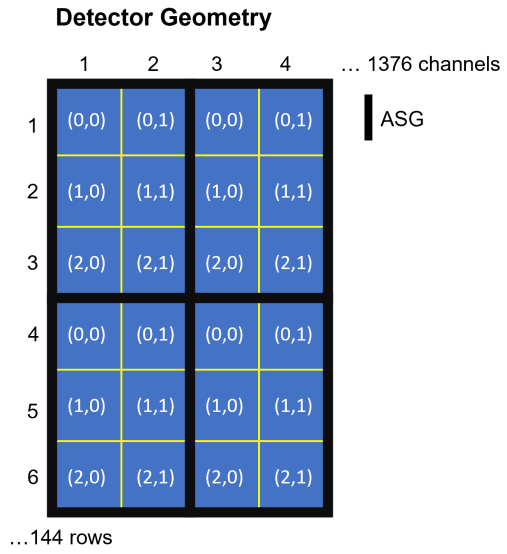


Fig. 3: Simulated detector geometry. There are 6 different pixels location within the anti-scatter grid which will measure a different intensity of scattered radiation. The total detector size 1376×144 pixels, consisting of 688×48 of the six-pixel-blocks.

To obtain the scatter data for the neural network, we have adapted our in-house Monte-Carlo simulation [10] to the geometry of the photon counting scanner NAEOTOM Alpha (Siemens Healthineers). In contrast to an energy-integrating detector where each pixel is surrounded by the lamella of the ASG, a coarse anti-scatter grid is used where 2×3 pixels are located between the lamellae of the ASG. The total detector size is 1376×144 pixels. As shown in figure 3, there are six different pixel locations under the coarse anti-scatter grid. If only one location is considered, the scattered radiation is further low-frequency. To consider this, post-smoothing needs to smooth only across same pixel locations and must not smooth between neighboring pixels. The distribution of the attenuation coefficient for the simulation of the CT scans is obtained from clinical full body CT exams. The patients are assumed to consist of four different materials (air, adipose tissue, soft tissue, and bone). One of these materials is assigned to each voxel based on CT value thresholds. In total 16 patients at 14 z-positions are Monte Carlo simulated every 10°. Twelve patients are used for training, two patients are used as validation data during the training and 2 patients are used as test data to evaluate the robustness of the networks. The corresponding primary intensities are obtained with a polychromatic forward projection. Overall 8064 paired scatter and primary data pairs are obtained for training and evaluation of the networks.

B. Deep scatter estimation

DSE is trained to predict scatter based on the acquired projection data as input [7]. In general DSE can be trained with measured or simulated scatter intensities [11]. Here the Monte Carlo-simulated scatter data will be used as the training and validation data. To adapt the network architecture to a coarse ASG the network receives the six different pixel locations as separate input channels (figure 4). The network is trained to predict six different output channels of the pixel locations, which then can be merged to obtain the scatter estimation for the whole detector. Each input and output channel has the size of 688×48 pixels.

C. Reconstruction

To evaluate the performance of the algorithm the scatter-corrected projections are reconstructed with the extended parallel back-projection (EPBP) algorithm [12]. Image quality is evaluated for the reconstructed images with the mean absolute error (MAE) compared to the ground truth. In the calculation of the MAE, air is excluded and only the patient area is considered.

III. Results

A clear moiré effect is visible in the reconstructed images without scatter correction. The average MAE in the uncorrected images is 8.4 HU. With the proposed algorithm the ring artifacts that appear can be significantly reduced and the average MAE is decreased to an average of 0.6 HU. In the figures 5 and 6 two examples of the scatter correction are shown. In the enlarged area, it is visible that the moiré effect is corrected. The difference images also show that the DSE algorithm corrects the scatter artifacts very well.

IV. Discussion

We propose a deep learning-based method to correct for scatter artifacts in photon-counting CT. In this work we demonstrate the potential of DSE to correct the artifacts that may occur due to the smaller detector pixel size and the coarse anti-scatter grid. The neural network is able to leverage information about the different pixel locations and shows an accurate scatter estimation. The results presented are based on simulated data, but the algorithm can also be applied to measurements. In our following step we plan to use the spectral information of the photon counting detectors to further improve the scatter corrections.

Acknowledgment

The Monte Carlo software RayConStruct-MC was provided by RayConStruct® GmbH, Nürnberg, Germany.

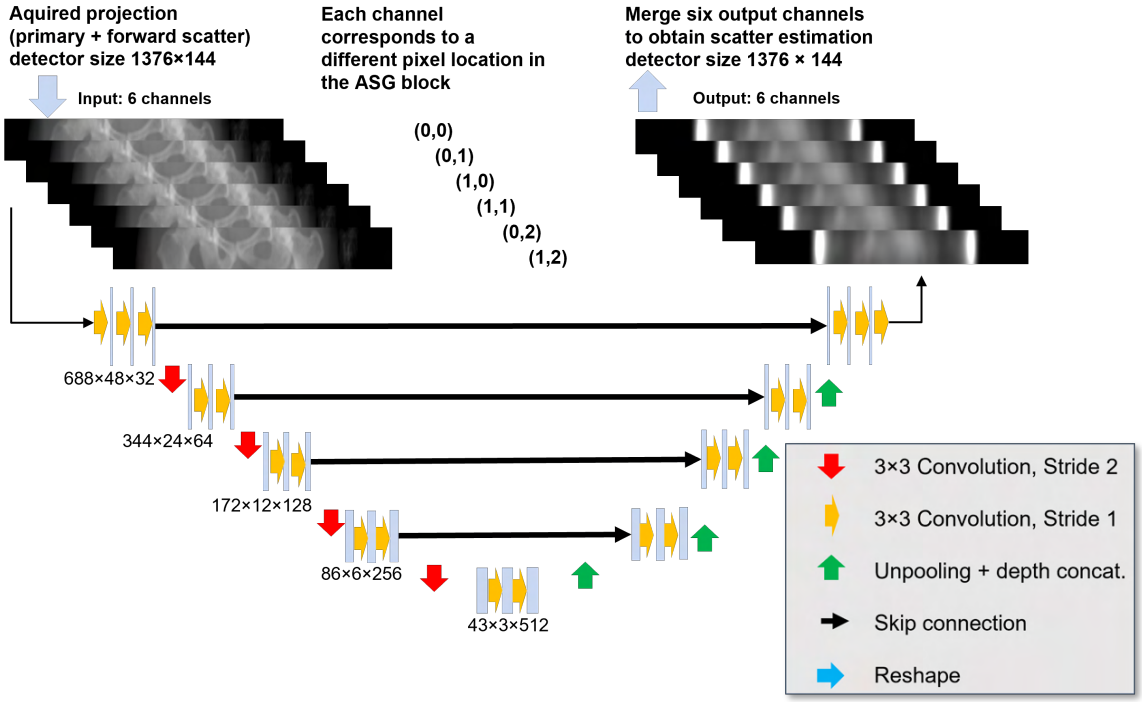


Fig. 4: The neural network has six different input and output channels corresponding to the different pixel location in the ASG.

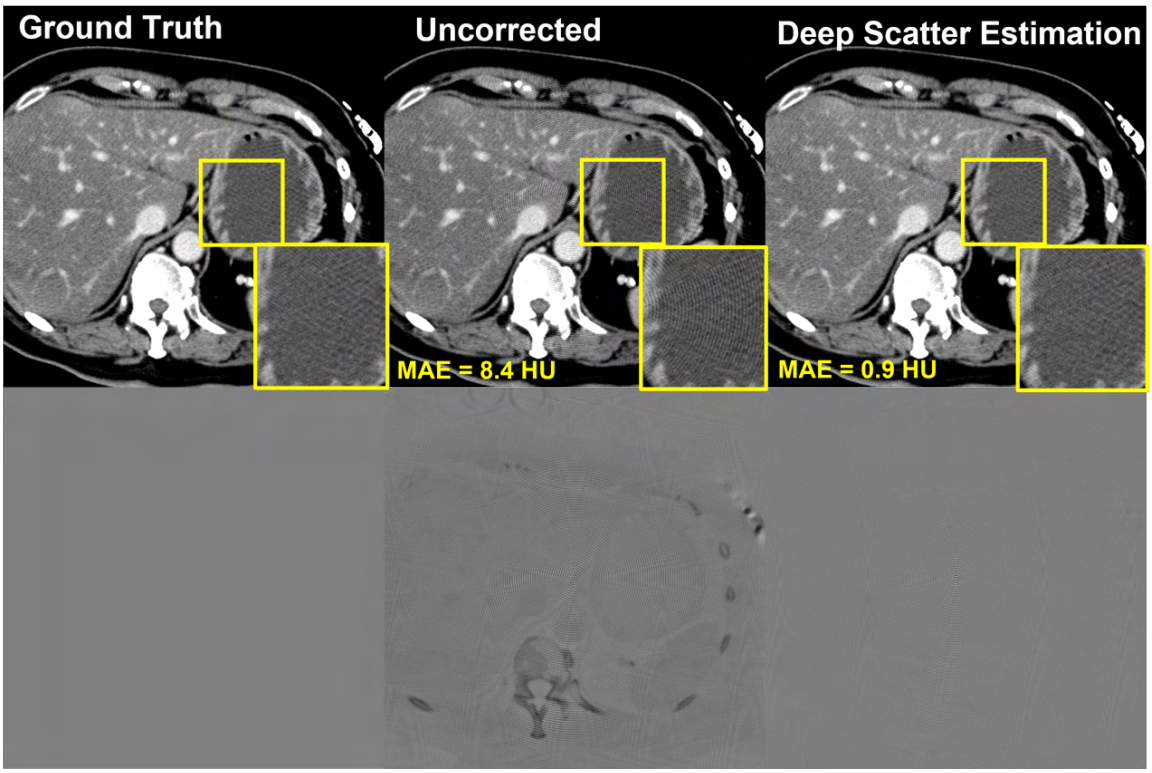


Fig. 5: First example of DSE for coarse ASGs. With the proposed neural network the MAE gets reduced from 8.4 HU to 0.9 HU. In the enlarged area it can be seen that the moiré artifacts gets corrected. C = 0 HU, W = 400 HU. Difference to GT: C = 0 HU, W = 50 HU.

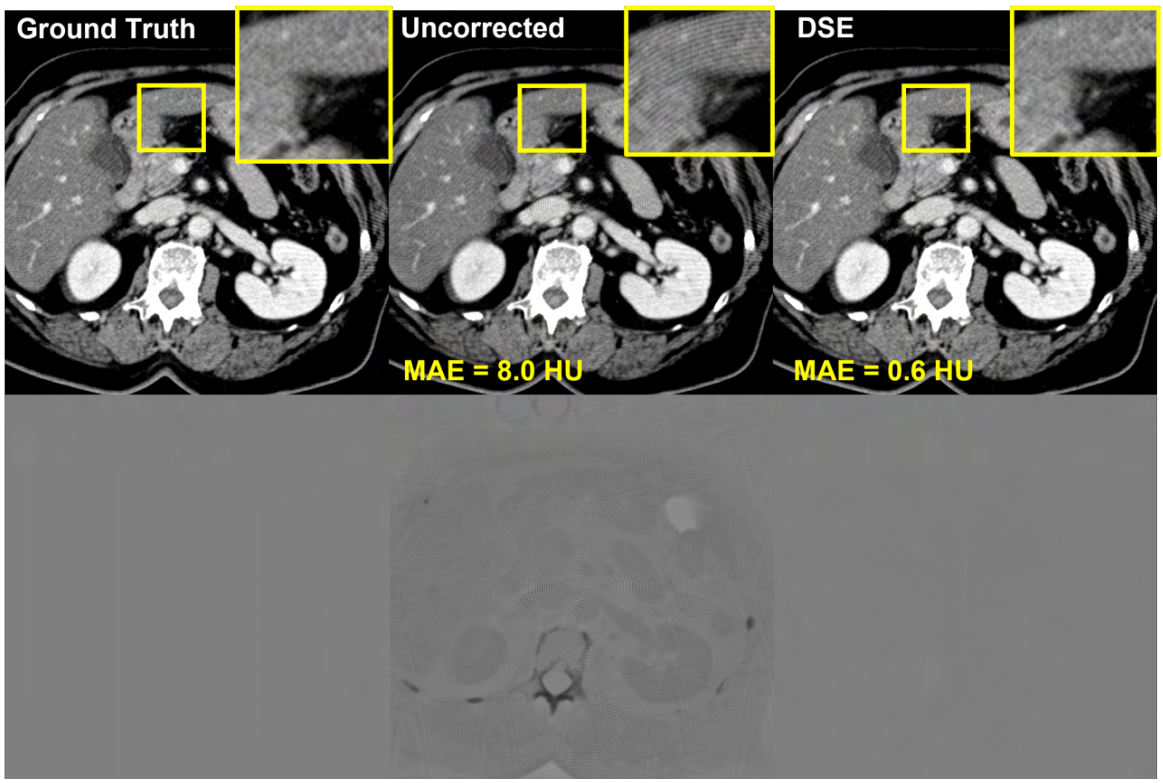


Fig. 6: Second example of DSE for coarse ASGs. Here the algorithm reduces the MAE from 8.0 HU to 0.6 HU. In the difference image, it is visible that the networks corrects both the scatter and moiré artifacts. C = 0 HU, W = 400 HU. Difference to GT: C = 0 HU, W = 50 HU.

References

[1] M. M. Lell and M. Kachelrieß, "Recent and upcoming technological developments in computed tomography: high speed, low dose, deep learning, multienergy," *Invest. Radiol.*, vol. 55, no. 1, pp. 8–19, 2020.

[2] K. Taguchi, E. C. Frey, X. Wang, J. S. Iwanczyk, and W. C. Barber, "An analytical model of the effects of pulse pileup on the energy spectrum recorded by energy resolved photon counting x–ray detectors," *Med. Phys.*, vol. 37, no. 8, pp. 3957–3969, Aug. 2010.

[3] K. Taguchi, M. Zhang, E. C. Frey, and X. Wang, "Modeling the performance of a photon counting x–ray detector for CT: Energy response and pulse pileup effects," *Med. Phys.*, vol. 38, no. 2, pp. 1089–1102, Feb. 2011.

[4] L. Klein, S. Dorn, C. Amato, S. Heinze, M. Uhrig, H.-P. Schlemmer, M. Kachelrieß, and S. Sawall, "Effects of detector sampling on noise reduction in clinical photon–counting whole–body computed tomography," *Investigative Radiology*, vol. 55, no. 2, pp. 111–119, Feb. 2020.

[5] S. Leng, K. Rajendran, H. Gong, W. Zhou, A. F. Halaweish, A. Henning, S. Kappler, M. Baer, J. G. Fletcher, and C. H. McCollough, "150-μm spatial resolution using photon-counting detector computed tomography technology: technical performance and first patient images," *Invest. Radiol.*, vol. 53, no. 11, pp. 655–662, 2018.

[6] A. Pourmorteza, R. Symons, A. Henning, S. Ulzheimer, and D. A. Bluemke, "Dose efficiency of quarter-millimeter photon-counting computed tomography: first-in-human results," *Invest. Radiol.*, vol. 53, no. 6, pp. 365–372, 2018.

[7] J. Maier, S. Sawall, M. Knaup, and M. Kachelrieß, "Deep scatter estimation (DSE): Accurate real-time scatter estimation for x-ray CT using a deep convolutional neural network," *Journal of Nondestructive Evaluation*, vol. 37, no. 3, p. 57, Jul. 2018.

[8] J. Maier, E. Eulig, T. Vöth, M. Knaup, J. Kuntz, S. Sawall, and M. Kachelrieß, "Real-time scatter estimation for medical CT using the deep scatter estimation: Method and robustness analysis with respect to different anatomies, dose levels, tube voltages, and data truncation," *Med. Phys.*, vol. 46, no. 1, pp. 238–249, Nov. 2019.

[9] J. Erath, T. Vöth, J. Maier, E. Fournié, M. Petersilka, K. Stierstorfer, and M. Kachelrieß, "Deep learning-based forward and cross-scatter correction in dual-source CT," *Medical Physics*, vol. 48, no. 9, pp. 4824–4842, 2021.

[10] M. Baer and M. Kachelrieß, "Hybrid scatter correction for CT imaging," *Phys. Med. Biol.*, vol. 57, pp. 6849–6867, 2012.

[11] J. Erath, J. Maier, E. Fournie, T. Vöth, K. Stierstorfer, and M. Kachelrieß, "Monte Carlo–free deep scatter estimation (DSE) for x-ray CT and CBCT." *Program of the 105th Scientific Assembly and Annual Meeting of the RSNA*, pp. SSE24–06, Dec. 2019.

[12] M. Kachelrieß, M. Knaup, and W. A. Kalender, "Extended parallel backprojection for standard 3D and phase–correlated 4D axial and spiral cone–beam CT with arbitrary pitch and 100% dose usage," *Med. Phys.*, vol. 31, no. 6, pp. 1623–1641, Jun. 2004.

Cross-Domain Metal Segmentation for CBCT Metal Artifact Reduction

Maximilian Rohleder[1,2], Tristan M. Gottschalk[1,2], Andreas Maier[1] and Bjoern W. Kreher[2]

***Abstract*—Metallic objects in the volume of a CBCT system can cause various artifacts after image reconstruction such as bright and dark streaks, local distortions of CT values and shadowing. In the intraoperative setting, this drastically reduces clinical value and can harden decision making. Most existing approaches to reduce such artifacts rely on a threshold-based metal segmentation in the reconstruction domain, which is prone to failure; especially in cases with extreme artifacts. Faulty metal masks impair these inpainting-based MAR methods and at times even worsen image quality by introducing secondary artifacts. In this work, a novel neural network topology is proposed to segment metallic objects in CBCT reconstruction domain by leveraging information of the given raw projection images. A reconstruction operator is embedded into this architecture, which enables the model to yield projection and reconstruction domain information during end-to-end training. This cross-domain approach is compared to the self-configuring segmentation method "nnUNet", which predicts the three dimensional metal masks directly from the artifact corrupted reconstruction. To provide a baseline, a segmentation using a global dice-optimal threshold is determined. Segmentation results on simulated data confirmed by 5-fold cross validation show that the cross-domain network yields a mean dice coefficient of 0.87 $\pm$ 0.05 at a prediction time of 4s per volume. The reference method achieves 0.86 $\pm$ 0.03 in 43s, whereas the optimal similarity using a threshold averages to 0.45 $\pm$ 0.22.**

***Index Terms*—Known Operator Integration, Metal Artifacts, Cone-Beam CT, Segmentation.**

I. Introduction

A. Motivation

MOBILE C-Arm devices are an integral part of modern surgical procedures. In addition to 2D X-Ray projection images used for guidance, modern systems allow to accurately verify the placement of tools and implants via 3D imaging. One major limitation of this modality are metal artifacts. Originating from simplifications in the reconstruction model and inconsistencies in the measured data, these image artifacts appear as bright and dark streaks, local distortions of CT values, and shadowing. As they emerge especially around metal objects, they obstruct the implant placement verification and thus drastically decrease the diagnostic value during the surgical intervention.

B. Existing MAR Approaches

Most modern Metal Artifact Reduction (MAR) methods reduce metal artifacts by inpainting the metal traces in projection domain. These metal traces are obtained by forward projection of the volumetric segmentation of metal objects. Inpainting in this context refers to the process of substituting pixel values to remove their contribution to the reconstruction image. Over the last decades, multiple approaches for inpainting have been proposed starting with simple linear or polynomial interpolation [1], [2], frequency domain interpolation [3], using wavelets [4] or with the help of machine learning [5]. Regardless of how elaborated the inpainting approach is designed, a faulty estimation of the metal mask can sabotage the effectiveness of said approaches, reduce the level of detail around the corrected metal, remove relevant anatomy, and even introduce new artifacts [6].

C. Related work

To enhance the mask quality, a shape-model based estimation of the metallic objects has been proposed [7]. A known object's outline is registered to a coarse volumetric metal segmentation to refine its shape. However, this approach is not generally applicable, as prior knowledge about the shape of the metal is usually not available.

The segmentation of metal in image domain is prone to error mainly because of the metal artifacts. The alterations of CT values around the depicted metal prohibit a purely value-based approach such as the traditional global thresholding. By including structural information, Convolutional Neural Networks (CNN) have proven to be successful in many medical segmentation scenarios.

Recent advances in known operator integration have facilitated the end-to-end training of cross-domain architectures. A derivable backprojection operator can be embedded into the model and used with gradient backpropagation for supervised learning [8]. The observation that metal artifacts originate during the domain transformation suggests that neural network architectures can benefit from projection domain information.

Multi-domain approaches have successfully been applied in Metal Artifact Avoidance (MAA) to estimate metallic objects from very few given projections [9]. However, the objective is fundamentally different compared to MAR. For MAA, a rough distribution of metal is sufficient to adapt the trajectory, whereas high detail masks are desirable for MAR. Furthermore, the system matrix based reconstruction from [9] cannot be applied to MAR segmentation due to the larger number of input projections.

A dual domain network topology has also been demonstrated for direct, learned MAR [10]. The authors report significant improvements over other single domain MAR approaches.

[1] Pattern Recognition Lab, Department of Computer Science, Friedrich-Alexander-University, Erlangen-Nuremberg, Germany
[2]Siemens Healthineers AG, Forchheim, Germany
Email: Maxi.Rohleder@fau.de

D. Planned Contributions

In this work, a novel cross-domain segmentation network which yields both raw rotational and reconstruction domain information is presented. It is compared to a CNN approach applied to artifact-corrupted reconstruction images and a simple threshold-based method.

II. Methods

A. Data

The cross-domain architecture proposed in this work requires training data, where the input consists of X-Ray projection images and the target metal mask is a volume-shaped binary array. It is crucial that the simulated projection images exhibit all physical effects which are relevant for the formation of metal artifacts. To model said effects, a simulation framework was derived from the DeepDRR method described in [11]. However, the provided pre-trained weights for material segmentation and scatter estimation could not be used as they did not generalize well on the raw data used in this project.

A total of 11 cone-beam CT volumes were acquired from 5 different specimens from a human spine cadaver study of the lumbar and thoracic region using a Siemens Cios Spin System.[1] To realistically model the shapes of metal objects, a library of 3D models of surgical screws, plates, k-wires, and towers are available from Nuvasive, San Diego, California. A set of metal objects is realistically positioned relative to the skeletal structures in the base volume using the 3D modelling suite Blender.[2] The position and orientation of each object is stored and considered during the simulation.

1) Simulation Process: First, each base volume is segmented into three materials $M \in \{air, tissue, bone\}$ using an empirically defined threshold. Then, projection images of the three material volumes and all metal objects are generated. The resulting material path-length projections are weighted with their spectrum-dependant attenuation coefficients according to the polychromatic Beer-Lambert law. To approximate the influence of scatter on the artifact formation, a constant background signal is added to the images. This simulation pipeline produces stacks of 200 X-Ray projection images of shape (488×488) over the angular range of 200 degrees which, after reconstruction, exhibit the desired metal artifacts. To generate the training labels, a cube-shaped binary array with side-length 256 is created from the metal objects depicted in one sample. With an isotropic voxel-size of 0.626, the binary volume resembles the standard volume size of the Siemens Cios Spin C-Arm System. Because of GPU memory limitations, the training data for this project has half the resolution and number of projection images compared to measured raw data.

2) Data Augmentation: In order to increase the dataset size, four random rotations around the z-axis are applied to each sample as an offline data augmentation step. This rotation is virtually applied prior to projection by simply appending to the projection matrices. Using this method, a total of 44 samples are generated from the 11 independent tool configurations.

[1]The work follows appropriate ethical standards in conducting research and writing the manuscript, following all applicable laws and regulations regarding treatment of animals or human subjects, or cadavers of both kind. All data acquisitions were done in consultation with the Institutional Review Board of the University Hospital of Erlangen, Germany.

[2]Open Source 3D Creation Suite, https://www.blender.org/

B. Optimal Threshold-Based Method

As current product-grade MAR methods utilize a threshold-based segmentation, it serves as a baseline here. To obtain a best-case estimate of such a global thresholding method, the optimal threshold is calculated for each sample. As an optimality criterion, the dice similarity coefficient (DSC) is used. Equation 1 shows the dice similarity defined on the binarized volume X and the binary label Y.

$$DSC(X, Y) = \frac{2|X \cap Y|}{|X| + |Y|} \tag{1}$$

A threshold is defined as optimal when the binarization it generates maximises the DSC metric. For each simulated CT volume, an optimal threshold is determined by testing 1000 values over the range of its histogram values.

C. Image Domain CNN Method

To obtain a standardized segmentation benchmark, the self-configuring framework *nnUNet* is used [12]. nnUNet automatically configures all relevant parameters of the popular architecture *U-Net* and adapts it to the so-called dataset fingerprint. The user can choose between a 2D, a 3D high-resolution, and a 3D low-resolution U-Net layout.

For this project, the 3D low-resolution (*3d-lowres*) U-Net model was chosen. The training of the other models was omitted, as the cross-domain method produces lower resolution masks and the results are compared on this lower resolution anyway. This model contains 6 mio. trainable parameters and is applied patch-wise on cube-shaped sub-volumes of side-length 128. The model is trained on CT volumes reconstructed from the simulated projection data. To predict volumes of side-length 256, the model is evaluated 27 times as the patches are strided by half a patch-size.

The performance of the *3d-lowres* architecture was evaluated using 5-fold cross validation. The simulated X-Ray images were reconstructed using filtered backprojection to serve as training input to this method. The data splitting, preprocessing and evaluation using the DSC metric was handled by the framework.

D. Cross-Domain Architecture

Inspired by the success of the 3D U-Net architecture, an encoder-decoder layout with skip-connections is used [13]. To enable end-to-end training with input and labels from different domains, a derivable backprojector is integrated into the network. This operator is implemented as the *PyroNN* filtered backprojection layer with a non-trainable Ram-Lak filter [8].

As seen in figure 1, this domain transform happens during the skip-connection step, such that the encoder is applied to projection domain and the decoder to reconstruction domain. The shapes of the tensors are indicated next to each stage of the

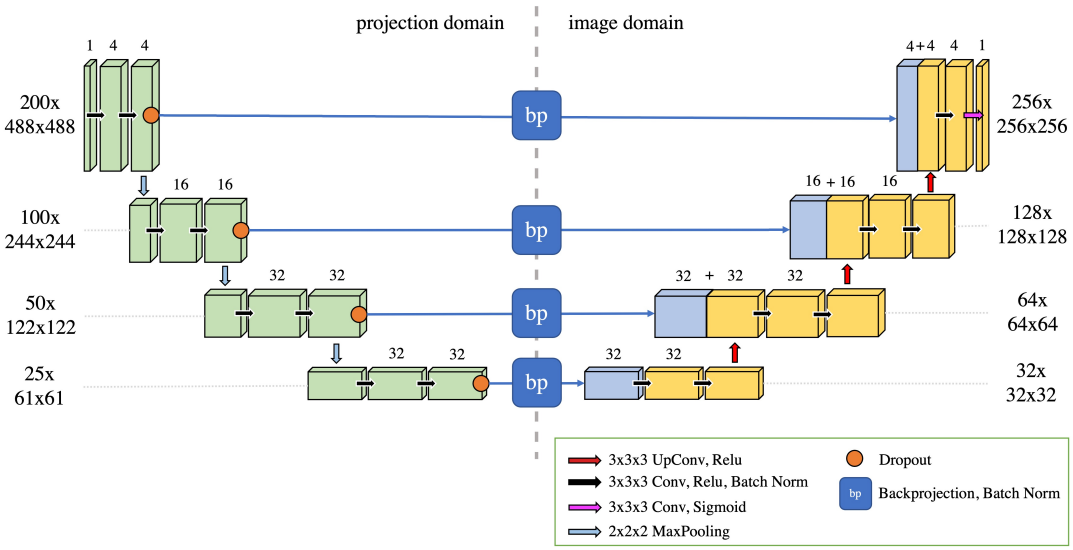


Fig. 1. **Cross-domain network architecture** with an encoder in projection domain and a decoder in image/reconstruction domain. The backprojection operators embedded on the skip-connections reconstruct volumes from the input projection domain feature maps and stack them along the channel axis (blue).

architecture, whilst the number of computed feature-maps is shown above each block symbolizing a layer's output. At each stage of the network, the stack of projections is downsampled both in the number of projections and their resolution.

To facilitate the reconstruction of differently sized volumes, the projection matrices and the resolution configurations are adapted accordingly. The convolutional layers with kernels of side length three use zero-padding to generate similarly sized feature maps. At the skip-connections, these feature maps are reconstructed individually and the resulting volume-shaped activations are stacked along the channel axis.

E. Cross-Domain Training Procedure

1) Preprocessing: The simulated intensity images are first converted to line integrals. A cross-validation split is defined with nine samples in the training set and the remaining two samples in the test set of each of the five splits. After splitting, the previously described offline data augmentation strategy is applied to boost the set sizes to $36/8$ (train/test).

During training, different noise levels are augmented. An additive, intensity dependent noise is drawn from a Poisson distribution. Furthermore, a convolutional noise model is used to imitate the detector characteristics and low-dose effects [14].

2) Training: The training samples are fed to the network in a batch-size of one due to GPU memory constraints. To compensate the resulting stochastic gradients, the Adam optimizer is applied with the standard parametrizations of the first and second order moments and an initial learning rate of 10^{-3}. Furthermore, the learning rate is reduced by a factor of 10 if the training loss plateaus for longer than 10 epochs. The training is terminated after the learning rate is reduced for the third time. As a loss function, the dice similarity defined in equation 1 is used.

III. Results

A. Quantitative

The cross-validation results and the evaluation of the optimal threshold-based method are shown in Table I. Our method achieves a dice similarity of 0.87 ± 0.05 averaged over the different data splits.

The reference method applied in reconstruction domain evaluates to a DSC of 0.86 ± 0.03 across all cross-validation runs. Note, that the data splits are not identical between our method and the reference image domain CNN.

The best possible segmentation using a per-sample threshold was evaluated for the entire simulated dataset. Overall, this method achieves a DSC metric of 0.45 ± 0.21.

TABLE I
Results of Model Evaluation (Dice Similarity)

	Ours	CNN	Threshold
split 0	0,8131	0,8089	-
split 1	0,8145	**0,8875**	-
split 2	0,9138	0,8574	-
split 3	0,8921	0,8699	-
split 4	**0,9140**	0,8613	-
mean±std	0,87±0,05	0,86±0,03	0,45±0,21

Apart from the segmentation performance, other attributes of the compared methods are displayed in Table II.

TABLE II
Additional Method Attributes

	Ours	CNN	Threshold
Inference Time	4s	45s	<1s
Inference GPU Memory	9.6Gb	2.1Gb	-
Training Duration	15h	97h	-
Training GPU Memory	23.2Gb	6Gb	-
#Params	200k	6mio.	1

The inference time of the reference image domain CNN method is about eleven times longer than our method. The

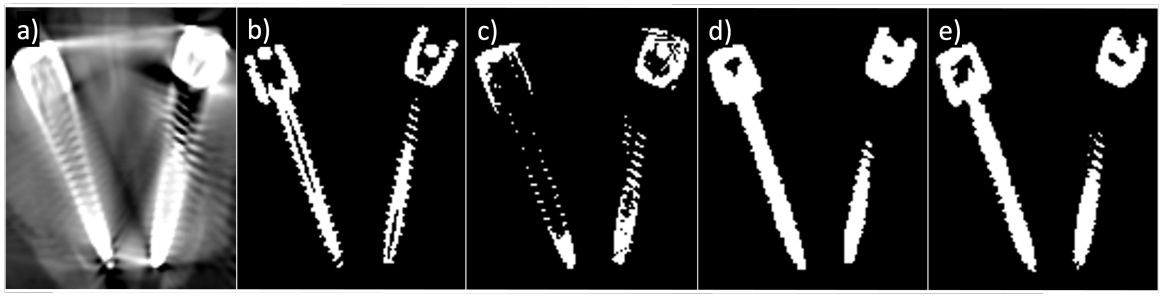


Fig. 2. **Cropped slice from a volumetric metal segmentation.** a) shows the corresponding artifact-corrupted reconstruction, b) is the ground truth, c) depicts the best-possible threshold-based segmentation, d) shows the result of the reference image domain CNN, and e) shows our methods prediction.

auto-configuring framework *nnUNet* trains the models for about 97h which is more than six times longer than our methods, which are trained 15h on average. The cross-domain network requires about four times more GPU memory during training and inference.

B. Qualitative

To intuitively compare the methods performance, an unseen sample is segmented using the three compared methods (Figure 2). From the resulting volumetric metal masks, a slice is selected and cropped to a region of interest showing two screws with tulips fixated to connecting rods.

For reference, the corresponding reconstruction is added. As the screws' main axes align with the acquisition trajectory, heavy artefacts are visible.

The thresholding method c) greatly underestimates the ground truth mask b). The image domain CNN and the cross-domain segmentation network produce similar looking masks, which, compared to the label, yield a dice similarity of 0.88 and 0.90 respectively.

IV. Conclusion

Summarizing the quantitative and qualitative results, it becomes evident that both learned methods largely improve the segmentation compared to the purely value-based thresholding baseline. Visually, there is no distinct difference between the cross-domain and single-domain network's predicted masks. However, the newly presented cross-domain method achieves this similar performance using 10% of trainable parameters and 15% of training time.

With the application in the surgical suite in mind, the cross-domain network offers the clear advantage of a faster inference time. This is largely attributable to the patch-wise application of the image domain CNN. On the downside, the novel network architecture has increased GPU memory requirements which might not be readily available on systems in the operating room due to financial cost.

Future work should focus on evaluating the effectiveness of the proposed approaches on measured data. Should this be successful, the impact of the improved metal masks on inpainting-based MAR methods needs to be investigated. Furthermore, the memory footprint of the cross-domain method can be reduced by revising the implementation of the backprojection operator.

References

[1] W. Kalender, R. Hebel, and J. Ebersberger, "Reduction of CT artifacts caused by metallic implants," *Radiology*, vol. 164, no. 2, pp. 576–577, 1987.

[2] A. H. Mahnken, R. Raupach, J. E. Wildberger, B. Jung, N. Heussen, T. G. Flohr, R. W. Günther, and S. Schaller, "A New Algorithm for Metal Artifact Reduction in Computed Tomography," *Investigative Radiology*, vol. 38, no. 12, pp. 769–775, 12 2003.

[3] B. Kratz and T. M. Buzug, "Metal artifact reduction in computed tomography using nonequispaced fourier transform," *IEEE Nuclear Science Symposium Conference Record*, pp. 2720–2723, 2009.

[4] A. Mehranian, M. R. Ay, A. Rahmim, and H. Zaidi, "X-ray CT metal artifact reduction using wavelet domain L0 sparse regularization," *IEEE Transactions on Medical Imaging*, vol. 32, no. 9, pp. 1707–1722, 2013.

[5] T. M. Gottschalk, B. W. Kreher, H. Kunze, and A. Maier, "Deep Learning Based Metal Inpainting in the Projection Domain: Initial Results," *Lecture Notes in Computer Science*, vol. 11905 LNCS, pp. 125–136, 2019.

[6] M. Stille, B. Kratz, J. Müller, N. Maass, I. Schasiepen, M. Elter, I. Weyers, and T. M. Buzug, "Influence of metal segmentation on the quality of metal artifact reduction methods," in *Medical Imaging 2013: Physics of Medical Imaging*, vol. 8668. SPIE, 3 2013, p. 86683C.

[7] A. Uneri, X. Zhang, T. Yi, J. W. Stayman, P. A. Helm, G. M. Osgood, N. Theodore, and J. H. Siewerdsen, "Known-component metal artifact reduction (KC-MAR) for cone-beam CT," *Physics in Medicine and Biology*, 2019.

[8] C. Syben, M. Michen, B. Stimpel, S. Seitz, S. Ploner, and A. K. Maier, "Technical Note: PYRONN: Python reconstruction operators in neural networks," *Medical Physics*, vol. 46, no. 11, pp. 5110–5115, 11 2019.

[9] P. Wu, N. Sheth, A. Sisniega, A. Uneri, R. Han, R. Vijayan, P. Vagdargi, B. Kreher, H. Kunze, G. Kleinszig, S. Vogt, S.-F. Lo, N. Theodore, and J. H. Siewerdsen, "Method for metal artifact avoidance in C-Arm cone-beam CT," in *Medical Imaging 2020: Physics of Medical Imaging*, H. Bosmans and G.-H. Chen, Eds., vol. 11312. SPIE, 3 2020, p. 78.

[10] W. A. Lin, H. Liao, C. Peng, X. Sun, J. Zhang, J. Luo, R. Chellappa, and S. K. Zhou, "DuDoNet: Dual domain network for CT metal artifact reduction," *Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition*, vol. 2019-June, pp. 10 504–10 513, 6 2019.

[11] M. Unberath, J. N. Zaech, S. C. Lee, B. Bier, J. Fotouhi, M. Armand, and N. Navab, "DeepDRR - A Catalyst for Machine Learning in Fluoroscopy-Guided Procedures," *Lecture Notes in Computer Science*, vol. 11073 LNCS, pp. 98–106, 2018.

[12] F. Isensee, P. F. Jaeger, S. A. A. Kohl, J. Petersen, and K. H. Maier-Hein, "nnU-Net: a self-configuring method for deep learning-based biomedical image segmentation," *Nature Methods 2020 18:2*, vol. 18, no. 2, pp. 203–211, 12 2020.

[13] . Çiçek, A. Abdulkadir, S. S. Lienkamp, T. Brox, and O. Ronneberger, "3D U-net: Learning dense volumetric segmentation from sparse annotation," in *Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics)*, vol. 9901 LNCS, 2016.

[14] A. S. Wang, J. W. Stayman, Y. Otake, S. Vogt, G. Kleinszig, A. J. Khanna, G. L. Gallia, and J. H. Siewerdsen, "Low-dose preview for patient-specific, task-specific technique selection in cone-beam CT," *Medical Physics*, vol. 41, no. 7, 2014.

Sparsier2Sparse: Weakly-supervised learning for streak artifact reduction with unpaired sparse-view CT data

Seongjun Kim, Byeongjoon Kim, and Jongduk Baek

***Abstract*—Sparse-view computed tomography (CT) becomes a major concern in the medical imaging field due to its reduced X-ray radiation dose. Recently, various convolutional neural network (CNN)-based approaches have been proposed, requiring the pairs of full and sparse-view CT images for network training. However, these paired data acquisition is impractical or difficult in clinical practice. To handle this problem, we propose the weakly-supervised learning for streak artifact reduction with unpaired sparse-view CT data. For CNN training dataset, we generate the pairs of input and target images from the given sparse-view CT data. Then, we iteratively apply the trained network to given sparse-view CT images and acquire the prior images. As the success factor of our novel framework, we estimate the original streak artifacts in the given sparse-view CT images from the prior images and subtract the estimated streak artifacts from the given sparse-view CT images. As a result, the proposed method has the best performance of lesion detection compared to the other methods.**

***Index Terms*—Computed tomography, Convolutional neural network, Weakly-supervised learning, Sparse-view CT.**

I. Introduction

COMPUTED tomography (CT) has been widely used for disease diagnosis in the medical imaging field due to its detailed observations of anatomical structures. However, the X-ray radiation exposure can increase a potential risk of cancers to the patients. To reduce the X-ray radiation dose, decreasing the number of projection views during CT scan, called sparse-view CT, can be implemented, whereas the severe streak artifacts are generated in the reconstructed images.

To overcome this, several iterative reconstruction (IR) methods, which iterativley optimize the both CT data fidelity and the image regularization terms, have been developed for sparse-view CT. For the prior of image regularization term, total variation [1], [2] is often utilized. However, IR calculates the forward and backprojection during optimization, requiring a huge amount of computational cost. Moreover, it is not easy to balance the parameters of fidelity and regularization terms for various imaging tasks. Therefore, the usage of IR methods is limited in medical imaging field, where the real time applications are required.

Recently, deep learning-based approaches have shown promising results in sparse-view CT reconstruction. Inspired by convolutional neural networks (CNNs), the image-domain method [3]–[5] trains the spatial information of streak artifacts distribution and reduces the streak artifacts effectively, whereas it also often reduces the signals overlapped by streak artifacts together. To overcome this limitation, the hybrid-domain approach [6]–[8], which contains both image and sinogram domain, not only learns the image prior but also utilizes the information of measured projection data during network training. From this, the hybrid-domain method improves the performance of streak artifact reduction while preserving the edge sharpness better than when only the image-domain is used.

Although the above CNN-based methods showed good performance, they are fully-supervised approaches that require the paired full and sparse-view CT images, where the anatomical structures are identical. In practice, it is not feasible to acquire the dataset of those pairs from the patients due to the principle of ALARA (as low as reasonably achievable). In this reason, to acquire the full and sparse-view CT image pairs, most of methods generate the sparse-view CT images by a computer simulation using the full view CT data. However, if only the sparse-view CT data are given, it is impossible to produce full view CT images that are anatomically identical. This is because image reconstruction in sparse-view CT is an undetermined inverse problem that has no general solution [9].

To tackle this problem, we propose a weakly-supervised learning for streak artifact reduction when the only unpaired sparse-view CT data are given. We acquire the pairs of data for CNN training from the given sparse-view CT data and train the network. We then apply trained network to the given sparse-view CT images iteratively. For the success of our framework, we perform streak estimation model for signal preservation.

II. Methods

Figure 1 shows the schematic diagram of the proposed method for reducing streak artifacts with unpaired sparse-view CT data.

This research was supported by the Bio and Medical Technology Development Program of the National Research Foundation (NRF) funded by the Ministry of Science and ICT (NRF2019R1A2C2084936 and 2020R1A4A1016619) and the Korea Medical Device Development Fund grant funded by the Korea government (the Ministry of Science and ICT, the Ministry of Trade, Industry and Energy, the Ministry of Health and Welfare, Republic of Korea, the Ministry of Food and Drug Safety) (202011A03).

Seongjun Kim, Byeongjoon Kim, and Jongduk Baek are with School of Integrated Technology and Yonsei Institute of Convergence Technology, Yonsei University, Incheon 21983, South Korea (e-mail: 2sjkim@gmail.com, bjkim2006@naver.com, jongdukbaek@yonsei.ac.kr)

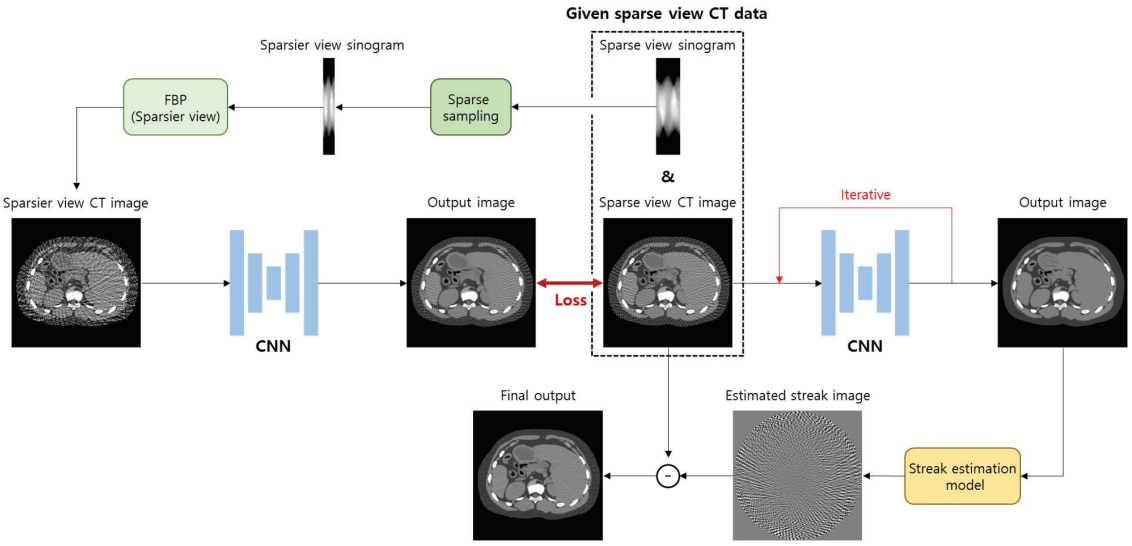


Fig. 1. The overall schematic of the proposed method. FP: forward projection; FBP: filtered back projection

A. Weakly-supervised learning

For CNN training to reduce streak artifacts, it is required to set the images with strong and weak streak artifacts as input and target, respectively. Since the only unpaired sparse-view CT data are given in our scenario, we should acquire pairs of the input and target from given sparse-view CT data. For this, we generate an image that has stronger streak artifacts than the original streak artifacts in given sparse-view CT image, which is denoted as a sparsier-view CT image. We then regard the sparsier and sparse-view CT images as the input and target for CNN training, respectively.

Here, the method of sparsier-view CT image generation is described in the following procedures. To make streak artifacts stronger, it can be achieved by decreasing the number of projection views because the strength of the streak artifacts is dependent on the number of projection views. Therefore, we down-sampled the given sparse-view sinogram by half and acquire a sparsier-view sinogram. Then, we reconstruct the sparsier-view CT image from the sparsier-view sinogram with the filtered backprojection (FBP) algorithm [10]. Note that we set down-sampling ratio as 2 to ease CNN training because the spasier view CT image will be significantly corrupted when the down-sampling ratio is bigger than 2.

Let $\hat{x}$ and x be the sparsier-view CT image and sparse-view CT image, respectively. Since the projection data of $\hat{x}$ is extracted and down-sampled from that of x, the streak artifacts in $\hat{x}$ will be amplified while keeping the directionality of the original streak artifacts in x. Therefore, we set a pair of $\hat{x}$ and x as CNN training dataset. We then train CNN with the mean squared error (MSE). The MSE is defined as

$$L_{MSE} = \mathbb{E}\left[\|f(\hat{x};\theta) - x\|^2\right] \tag{1}$$

where $f(\cdot;\theta)$ and θ denote the network operator and network parameters, respectively.

B. Iterative streak artifact reduction

Since the network trains to reduce streak artifacts by predicting the original streak artifacts in x from the amplified streak artifacts in $\hat{x}$, we iteratively apply the trained network to a given x to reduce the streak artifacts. In the first iteration, we set x as the network input and acquire the network output $f(x;\theta)$. Since a single iteration is not sufficient to reduce streak artifacts in x, more than two iterations are conducted. Note that the output of the previous iteration was set as the input of the current iteration. We perform this processing until the streak artifacts in x are reduced to the level we want. Note that the number of iterations can be flexibly controlled and it was set to a maximum of five in this work. We will notate y as the resulting image of the last iteration in this step.

C. Streak artifact estimation

Applying the trained network operator several times works to reduce streak artifacts rather than preserve the anatomical structure. Therefore, the streak artifacts are significantly reduced in y, but over-smoothing and blurring can occur, resulting in low performance of signal detection. In this reason, it is inadequate to utilize y itself as the final result. To preserve the fine details that may have been blurred, we estimate the original streak artifacts in x from y.

For streak estimation model, we first reconstruct full and sparse-view images from y by sequentially applying forward projection and FBP algorithm with full and sparse-view projection data, respectively. These full view image and sparse-view image are denoted as y_{full} and y_{sparse}, respectively. The estimated streak image is generated by subtracting y_{full} from y_{sparse}. We then acquire the final output by subtracting the estimated streak image from x.

D. Training configurations

We used the U-net structure [11] because it is well-known for effectively extracting the widely distributed streak artifacts

in the reconstructed images. The network was optimized by Adam optimizer [12] with default parameters and the learning rate 1e-4. To increase the datasets for CNN training, we implemented data augmentation that randomly flips and rotates the cropped images, where the images of 256 by 256 pixels are cropped from the original images of 512 by 512 pixels. We trained the network until 100 epochs and set batch size as 4.

III. Datasets and Experiments

A. XCAT dataset

We used abdomen and thorax regions of XCAT simulation dataset in our work. A total of 9690 slices was extracted from the 28 XCAT phantoms. We generated XCAT images in fan-beam geometry system and the simulation parameters are summarized in Table I. During image generation, the forward projection was implemented using Siddon's ray-driven algorithm [13] and the FBP algorithm was conducted with a Ram-Lak filter. For CNN dataset, we used 25 phantoms (8720 slices) and 3 phantoms (970 slices) for training and test set, respectively. The different details between sparse and full view CT images are explained in the following section.

TABLE I
Fan geometry simulation parameters

Parameters	Values
Source to iso-center distance	$50cm$
Source to detector distance	$50cm$
Detector pixel size	512 x 1
Detector cell size	$0.2cm^2$ x $0.1cm^2$
Reconstructed matrix size	512 x 512
Reconstructed pixel size	$0.0628cm^2$ x $0.0628cm^2$

B. Data generation

We generated sparse and full view CT images with 128 and 512 projection views, respectively, and each projection angle was equally spaced over 360 degrees. For Poisson noise, we set the number of incident X-ray photons as 10^6 for full view CT images. However, since the number of projection views in the sparse-view CT images is a quarter of full view CT images, the noise level of sparse-view CT images will be four times higher than that of full view CT images when the number of incident X-ray photons are the same. To make the noise level equal, we set the number of incident X-ray photons as 4×10^6 for sparse-view CT images. Note that the only unpaired sparse-view CT images are given for CNN training and the full view CT images are only used in testing phase.

C. Comparison methods

We additionally conducted a simple linear interpolation method and a fully-supervised learning [11] for comparison. The linear interpolation method estimated the missing view data of sparse-view sinogram by applying linear interpolation in view direction. For fully-supervised learning, we assumed that the paired of sparse and full view CT images were given for input and target of CNN training, respectively. The training configurations were the same as that we used in the proposed method.

IV. Results

Figure 2 shows the resulting images on XCAT dataset. The linear interpolation image showed less streak artifacts than 128 view FBP image through the estimation of missing view data, whereas the secondary artifacts such as edge distortion were produced due to the interpolation errors. The fully-supervised method reduced the streak artifacts and noise significantly. However, it can be seen that the lesions indicated by red and yellow arrows in ROIs of the fully-supervised method were blurred, leading to poor performance of visibility of those signals. In contrast, the proposed method produced the best visual similarity of ROIs to those of 512 view FBP image while preserving the edge sharpness. For the proposed method with iteration 1, its difference image had more edge errors than the results of other iterations. However, these errors disappear as more iterations are conducted.

Table II summarizes the average and standard deviation of normalized root MSE (NRMSE) and structural similarity index (SSIM) [14] on XCAT testset. We observed that the fully-supervised method showed the best scores of NRMSE and SSIM. For the proposed method, iteration 2 had the lowest NRMSE and highest SSIM scores. Although the proposed method with iteration 2 showed slightly worse scores than the fully-supervised method, visual inspection confirms that the proposed method produces better image texture without sacrificing the detectability of lesions.

To examine the effect of the streak estimation model in the proposed method, we did not apply the streak estimation model after the iterative streak artifact reduction procedure and the results are shown in figure 3. As the number of iterations increases, although the streak artifacts are gradually reduced, it can be seen that the images were getting blurred. For the ROIs of more than iteration 2, the signals were lost and the edge was also over smoothed. Although the results of iteration 1 preserved the signal indicated by red arrow, the streak artifacts were not reduced enough. From this, we chose to use iteration 2 to effectively reduce streak artifacts while preserving edge and signal shapes.

V. Conclusion

In this work, we proposed the weakly-supervised learning for streak artifact reduction with unpaired sparse-view CT data. From the results, the proposed method achieved the best performance in preserving lesions while reducing streak artifacts compared to the other results. By overcoming the difficulty of acquiring the pairs of sparse and full view CT images in practice, we expect that our novel framework can be utilized successfully in medical imaging field.

References

[1] J. C. Park, B. Song, J. S. Kim, S. H. Park, H. K. Kim, Z. Liu, T. S. Suh, and W. Y. Song, "Fast compressed sensing-based cbct reconstruction using barzilai-borwein formulation for application to on-line igrt," *Medical physics*, vol. 39, no. 3, pp. 1207–1217, 2012.

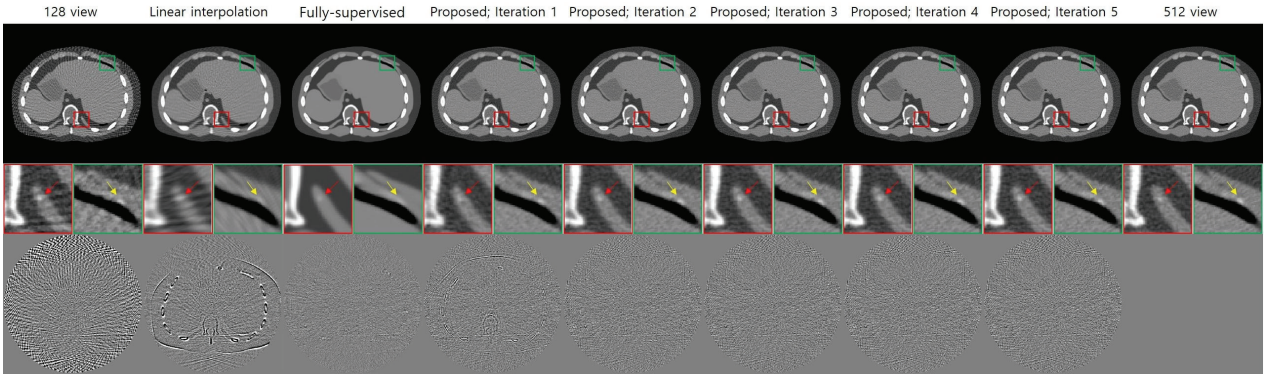


Fig. 2. The reconstructed images with XCAT dataset. The display window of field of view (FOV) and zoomed region of interest (ROI) images is (-240,320) in HU. The display window of difference images is (-100,100) in HU.

TABLE II
THE QUANTITATIVE EVALUATIONS OF XCAT DATASET

Metrics	128 view	Linear interpolation	Fully-supervised	Proposed				
				Iteration 1	Iteration 2	Iteration 3	Iteration 4	Iteration 5
NRMSE	0.0268 ± 0.0042	0.0157 ± 0.0036	**0.0064 ± 0.0013**	0.0122 ± 0.0015	**0.0102 ± 0.0012**	0.0104 ± 0.0012	0.0108 ± 0.0012	0.0111 ± 0.0012
SSIM	0.6838 ± 0.0484	0.8988 ± 0.0340	**0.9659 ± 0.0133**	0.9004 ± 0.0135	**0.9101 ± 0.0115**	0.9019 ± 0.0117	0.8938 ± 0.0119	0.8870 ± 0.0121

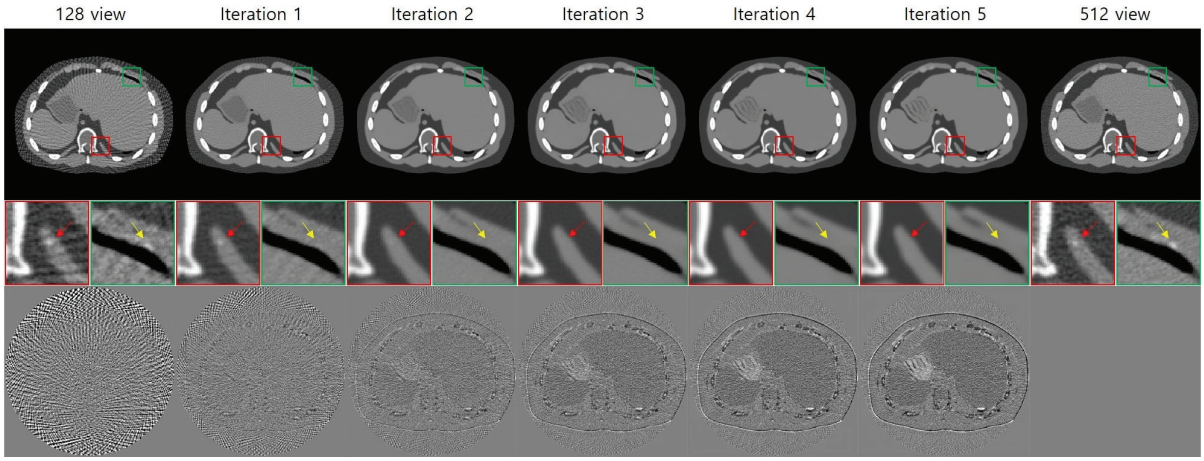


Fig. 3. Prior images after conducting different number of iterations in the proposed method. The display window of FOV and ROI images is (-240,320) in HU. The display window of difference images is (-100,100) in HU.

[2] E. Y. Sidky and X. Pan, "Image reconstruction in circular cone-beam computed tomography by constrained, total-variation minimization," *Physics in Medicine & Biology*, vol. 53, no. 17, p. 4777, 2008.

[3] Z. Zhang, X. Liang, X. Dong, Y. Xie, and G. Cao, "A sparse-view ct reconstruction method based on combination of densenet and deconvolution," *IEEE transactions on medical imaging*, vol. 37, no. 6, pp. 1407–1417, 2018.

[4] Y. Han and J. C. Ye, "Framing u-net via deep convolutional framelets: Application to sparse-view ct," *IEEE transactions on medical imaging*, vol. 37, no. 6, pp. 1418–1429, 2018.

[5] K. H. Jin, M. T. McCann, E. Froustey, and M. Unser, "Deep convolutional neural network for inverse problems in imaging," *IEEE Transactions on Image Processing*, vol. 26, no. 9, pp. 4509–4522, 2017.

[6] K. Liang, H. Yang, and Y. Xing, "Comparison of projection domain, image domain, and comprehensive deep learning for sparse-view x-ray ct image reconstruction," *arXiv preprint arXiv:1804.04289*, 2018.

[7] A. Zheng, H. Gao, L. Zhang, and Y. Xing, "A dual-domain deep learning-based reconstruction method for fully 3d sparse data helical ct," *Physics in Medicine & Biology*, vol. 65, no. 24, p. 245030, 2020.

[8] D. Lee, S. Choi, and H.-J. Kim, "High quality imaging from sparsely sampled computed tomography data with deep learning and wavelet transform in various domains," *Medical physics*, vol. 46, no. 1, pp. 104–115, 2019.

[9] W. Wu, D. Hu, C. Niu, H. Yu, V. Vardhanabhuti, and G. Wang, "Drone: Dual-domain residual-based optimization network for sparse-view ct reconstruction," *IEEE Transactions on Medical Imaging*, 2021.

[10] J. Hsieh, *Computed tomography: principles, design, artifacts, and recent advances*. SPIE press, 2003, vol. 114.

[11] B. Kim, M. Han, H. Shim, and J. Baek, "A performance comparison of convolutional neural network-based image denoising methods: The effect of loss functions on low-dose ct images," *Medical physics*, vol. 46, no. 9, pp. 3906–3923, 2019.

[12] D. P. Kingma and J. Ba, "Adam: A method for stochastic optimization," *arXiv preprint arXiv:1412.6980*, 2014.

[13] R. L. Siddon, "Fast calculation of the exact radiological path for a three-dimensional ct array," *Medical physics*, vol. 12, no. 2, pp. 252–255, 1985.

[14] Z. Wang, A. C. Bovik, H. R. Sheikh, and E. P. Simoncelli, "Image quality assessment: from error visibility to structural similarity," *IEEE transactions on image processing*, vol. 13, no. 4, pp. 600–612, 2004.

Dual Domain Closed-loop Learning for Sparse-view CT Reconstruction

Yi Guo, Yongbo Wang, Manman Zhu, Dong Zeng, Zhaoying Bian, Xi Tao and Jianhua Ma

***Abstract*—Sparse view sampling is one of the effective ways to reduce radiation dose in the CT imaging. However, artifacts and noise in sparse-view filtered back projection reconstructed CT images are obvious that should be removed effectively to maintain diagnostic accuracy. In this paper, we propose a novel sparse-view CT reconstruction framework, which integrates the projection-to-image and image-to-projection mappings to build a dual domain closed-loop learning network. For simplicity, the proposed framework is termed as closed-loop learning reconstruction network (CLRrcon). Specifically, the primal mapping (i.e., projection-to-image mapping) contains a projection domain network, a backward projection module, and an image domain network. The dual mapping (i.e., image-to-projection mapping) contains an image domain network and a forward projection module. All modules are trained simultaneously during the network training stage, and only the first mapping is used in the network inference stage. It should be noted that both the inference time and hardware requirements do not increase compared with traditional hybrid domain networks. Experiments on low-dose CT data demonstrate the proposed CLRecon model can obtain promising reconstruction results in terms of edge preservation, texture recovery, and reconstruction accuracy in the sparse-view CT reconstruction task.**

***Index Terms*—Computed tomography, closed-loop learning, image reconstruction, sparse view.**

I. Introduction

COMPUTED Tomography (CT) has been widely used in modern medical diagnosis and treatment due to its fast imaging speed, high resolution, etc. However, patients will receive lots of radiation dose during the CT examination, which is becoming a concern. Sparse-view scan can effectively reduce the radiation dose, but with the decrease of the scan views, the image quality will degrade when using the traditional filtered back projection (FBP) algorithm.

Numerous model-based iterative algorithms (MBIR) have been proposed for sparse-view CT reconstruction in the past decade. With correct prior assumptions [1], [2], the iterative algorithm can obtain high quality images. However, prior information is often manually selected, which cannot achieve desired results when it is not completely consistent with the actual collected projection data. In addition, the iterative algorithm requires repeated forward and backward projection until the desired image is obtained, which is time-consuming and requires a lot of computing resources.

This work was supported in part by the NSFC under Grant U21A6005 and Grant U1708261, the National Key R&D Program of China under Grant No. 2020YFA0712200. *(Corresponding author: Jianhua Ma.)*

Y. Guo, Y. Wang, M. Zhu, D. Zeng, Z. Bian, X. Tao and J. Ma are with the School of Biomedical Engineering, Southern Medical University, Guangdong, China; and the Guangdong Artificial Intelligence and Digital Economy Laboratory (Guangzhou), Guangdong, China (e-mail: jhma@smu.edu.cn).

In recent years, deep learning has achieved great success in the CT reconstruction task. The deep learning based reconstruction methods can be divided into the following three categories. The first category trains a network mapping from low-dose data to normal-dose data in the image domain or projection domain [3]–[5]. Some scholars attempt combine image domain network with projection domain network to form a hybrid model [6], [7]. The second category expands the iterative reconstruction algorithms into a network [8]–[10]. The third category builds a projection-to-image reconstruction network [11]–[13]. Through sufficient training, the network can directly reconstruct the image without artifacts from sparse-view projection data. In addition, Tao et al. [14] proposed to learn in the view-by-view backprojection tensor (VVBP-Tensor), and the experiment found that the results of this learning framework was significantly improved. At present, the application of deep learning in CT reconstruction mainly completes the learning from projection to image. However, this process lacks the constraint of comparing the difference between the calculated projection by projecting the reconstructed image and the measured projection like iterative algorithms.

Different from traditional supervised learning or semi-supervised learning, dual learning forms a closed-loop system by creating a dual problem of the primary problem. The primary problem and dual problem can mutually promote learning each other through this closed-loop system, so as to obtain better learning. Dual learning has been a great success in natural image processing such as image super-resolution [15] and raindrop removal [16]. For CT reconstruction, the advantages of traditional iterative reconstruction algorithms in low-dose CT reconstruction are mainly reflected in the use of forward and backward projection operators to update the target image by error feedback. This process implies a constraint that the estimated projection obtained by projecting reconstructed image should be consistent with the measured projection. Inspired by these works, we propose a closed-loop learning reconstruction model (CLRecon) for sparse-view CT reconstruction. In the proposed CLRecon, the original problem is learning the mapping from measured projection to image, and the dual problem is learning the mapping from image to measured projection. The mapping from image to projection can effectively constrain the mapping from projection to image to learn in the right direction, and thus helping to improve the quality of reconstruction.

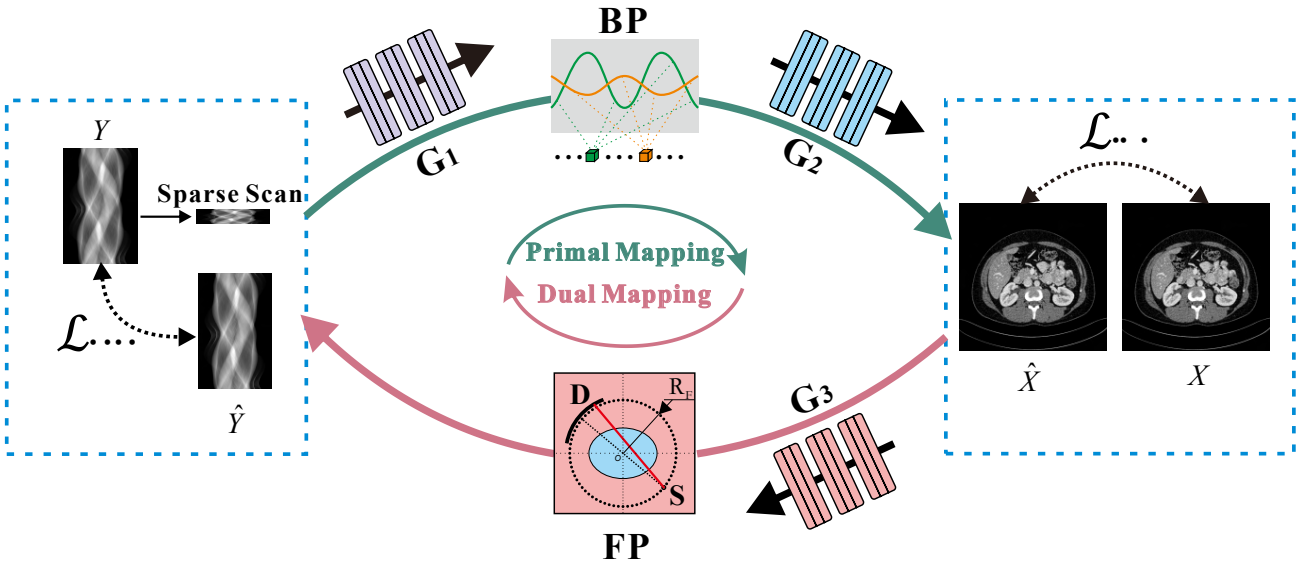


Fig. 1. The proposed closed-loop learning reconstruction framework.

II. Methods

A. Overview

Fig. 1 depicts the overview of our proposed closed-loop learning reconstruction framework. It consists of two learning mappings. The primal mapping is used to learn the transformation from sparse-view projection to reconstructed image, and the second one learns the dual mapping from reconstructed image to projection. The primal mapping includes a projection domain network, a gradient returnable backward projection module which is the implementation of FBP and an image domain network. The dual learning mapping consists of an image domain network and a gradient returnable forward projection module.

B. Network Architecture

We build the backward projection module by implementing FBP algorithm based on PyTorch deep learning library [17], which can retain gradient during forward propagation. Considering the sparsity of system matrix, we construct it by sparse matrix. This greatly reduces memory requirements and makes it possible to project the reconstructed image when training the neywork. In our proposed closed-loop learning sparse-view CT reconstruction framework, sub-network G_1, G_2, G_3 can be any network. In our next experiment, we chose FBPConvNet [18] as the base network and we remove the batch normalization layer as is shown in Fig. 2.

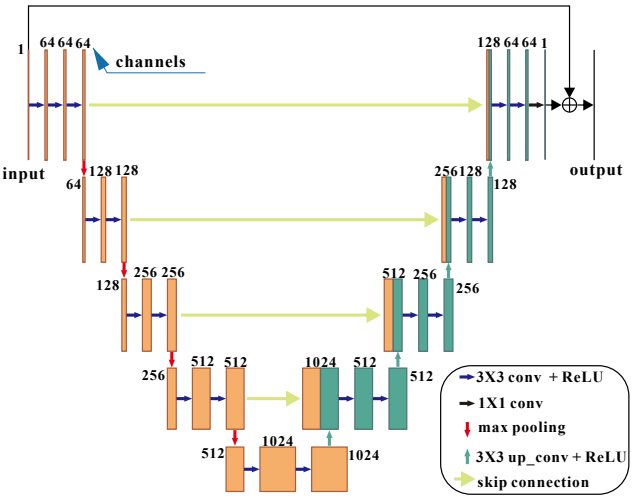


Fig. 2. The architecture of the base network.

C. Loss Functions

We adopt the simple mean square error (MSE) loss to train the network. To make network training is constrained by both label image and measured projection, we add the loss in both image domain and projection domain. For image domain loss, we can formulate it as follows:

$$\mathcal{L}_{img} = \frac{1}{MN}\sum_{i=1}^{M}\sum_{j=1}^{N}\|\hat{X}_{ij} - X_{ij}\|_2^2, \tag{1}$$

where $\hat{X}$ is the output of the network and X is the label image. M is the number of samples in a batch and N is the number of pixels in sample. The projection domain loss can be formulated as follows:

$$\mathcal{L}_{proj} = \frac{1}{MP}\sum_{i=1}^{M}\sum_{j=1}^{P}\|\hat{Y}_{ij} - Y_{ij}\|_2^2, \tag{2}$$

where $\hat{Y}$ is the output of the network and Y is the measured projection. P is the number of pixels in sample. If we have

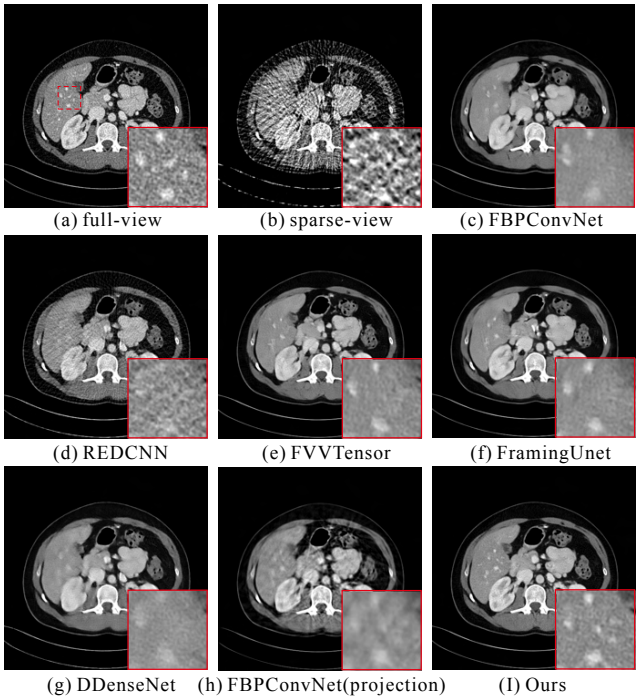


Fig. 3. Reconstruction results of different networks with a view number of 72. Display window: [-160, 240] HU.

full-view projection, Y can be full-view projection for better performance. The full objective function contains the image domain loss and the projection domain loss as follow:

$$\mathcal{L}_{mse} = \lambda \mathcal{L}_{img} + (1-\lambda)\mathcal{L}_{proj}, \tag{3}$$

where λ is a parameter that balances image loss and the projection loss, and in this work, it is set to be 0.8.

III. Experiments

A. Data

We used AAPM Low Dose CT Grand Challenge datasets for evaluation which consists of routine dose CT and the corresponding simulated LDCT data [19]. We used 10976 slices from 34 patients for training the network and 1095 slices from 6 patients for validation and testing. We projected the images to obtain the simulation fan-beam projection data. The geometry parameters projection were set as: projection view number of 1152, detector bin number of 736, image pixel space of $0.6934mm \times 0.6934mm$, and detector bin width of 1.2858 mm. In our experiment, we extracted 72 views with equal angle distribution to simulate the sparse-view scan.

B. Implementation Details

The framework was implemented in Python based on PyTorch deep learning library [17]. All reconstruction images have a size of 512 $\times$ 512 and the sinograms are with a size of $n \times 1152$, where n is the projection views. The Adam optimizer [20] was used to optimize the whole framework with the parameters $(\beta_1, \beta_2) =$ (0.9 0.999). The learning rate drop linearly from 10^{-3} to 10^{-5}. We trained the network on NVIDIA GeForce RTX 3090 GPUs.

TABLE I
QUANTITATIVE COMPARISON OF DIFFERENT MODELS.

Model	RMSE	SSIM
FBPConvNet	4.6328	0.8710
REDCNN	4.9033	0.8417
FVVTensor	4.2655	0.8827
FramingUnet	4.4261	0.8782
DDenseNet	4.4699	0.8770
FBPConvNet (projection)	5.0693	0.8603
Ours	**3.7063**	**0.9004**

TABLE II
QUANTITATIVE COMPARISON OF DIFFERENT COMBINATIONS OF MODULES.

Model	RMSE	SSIM
G_2	4.1230	0.8858
$G_1 + G_2$	3.9299	0.8937
$G_1 + G_2$ + FP	3.7685	0.8984
$G_1 + G_2 + G_2$ + FP	**3.7063**	**0.9004**

IV. Results

A. Experimental Results on Mayo Data

1) Qualitative analysis: We compared our method with the recent deep-learning-based methods, including FBPConvNet [18], FramingUNet [21], REDCNN [22], DDenseNet [23], FVVtensor [14]. FBPConvNet is closer to the network we used, but we did not use the batch normalization layer like it did. FramingUNet, REDCNN and DDenseNet optimized the network structure for better performance. FVVtensor learns in VVBP-Tensor and obtains good results. We also applied the FBPConvNet in projection domain to repair sparse views projection directly.

Fig. 3 shows the visual comparisons of our method and other methods on the reconstructed images with 72 views. We show the full-views image, sparse-view image and results of different models. To better compare the results of different networks, the ROI in the image is enlarged at the bottom of the image. It can be observed that our network can achieve excellent results in edge preservation and artifact removal.

2) Quantitative comparisons with state-of-the-art methods: Table I shows the quantitative comparison results of our method and other methods. It's the average of all the test slices. We can observe that our model achieves lower root mean square error (RMSE) and higher structured similarity index (SSIM) than other methods, which can prove that our network can obtain better reconstruction quality under this sparse-view degradation levels.

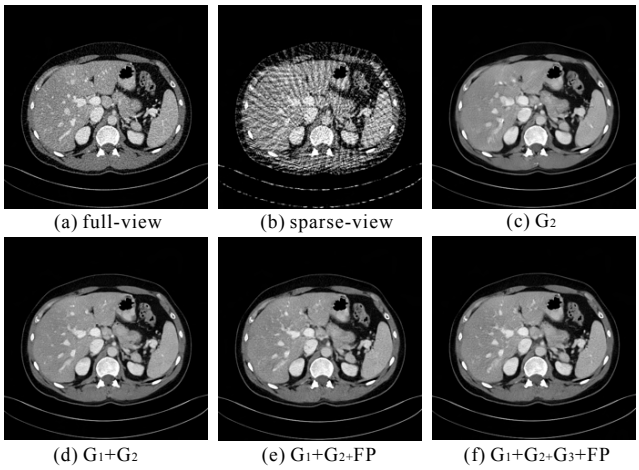


Fig. 4. Reconstruction results of different combinations of modules. Display window: [-160, 240]HU.

B. Ablation Study

In our proposed CLRecon, we add a image-to-projection mapping to the primal mapping (i.e., projection-to-image mapping) and form a closed-loop learning system. To show the effectiveness of this procedure, we compared the reconstruction results of different combinations of G_1, G_2, G_3 and forward projection module (FP). Fig. 4 shows the visual comparisons of different combinations of modules. We can observe that the image quality can be improved after adding G_3 and FP. Table II shows the quantitative comparison results of different experiments. We can see that it can achieve lower RMSE and SSIM with forward projection module and G_3. The network complexity did not increase because image-to-projection mapping was not used during the test stage. The improvement of reconstructed image quality lies in the improvement of network learning strategy rather than the increase of network depth. It proves that the proposed closed-loop learning can improve the reconstruction quality in sparse-view CT reconstruction.

V. Conclusion

We have presented a closed-loop learning reconstruction model (CLRecon) for sparse-view CT reconstruction in this paper. Specifically, the primal mapping is used to learn the transformation from sparse-view projection to reconstructed image, and the dual mapping learns from reconstructed image to projection. Our experiment shows that the addition of these two modules can improve the quality of reconstructed image. Since this mapping is only used in network training, the network parameters are not increased. The improvement of network performance lies in the change of learning strategy.

References

[1] E. Y. Sidky, X. J. P. i. M. Pan, and Biology, "Image reconstruction in circular cone-beam computed tomography by constrained, total-variation minimization," *Physics in Medicine & Biology*, vol. 53, no. 17, p. 4777, 2008.

[2] K. Kim, G. E. Fakhri, and Q. J. M. P. Li, "Low-dose CT reconstruction using spatially encoded nonlocal penalty," *Medical Physics*, vol. 44, no. 10, 2017.

[3] Z. Zhang, X. Liang, X. Dong, Y. Xie, and G. J. I. t. o. m. i. Cao, "A sparse-view CT reconstruction method based on combination of DenseNet and deconvolution," *IEEE Transactions on Medical Imaging*, vol. 37, no. 6, pp. 1407-1417, 2018.

[4] H. Lee, J. Lee, H. Kim, B. Cho, S. J. I. T. o. R. Cho, and P. M. Sciences, "Deep-neural-network-based sinogram synthesis for sparse-view CT image reconstruction," *IEEE Transactions on Radiation & Plasma Medical Sciences*, vol. 3, no. 2, pp. 109-119, 2018.

[5] H. Shan et al., "Competitive performance of a modularized deep neural network compared to commercial algorithms for low-dose CT image reconstruction," *Nature Machine Intelligence*, vol. 1, no. 6, pp. 269-276, 2019.

[6] A. Zheng, H. Gao, L. Zhang, Y. J. P. i. M. Xing, and Biology, "A dual-domain deep learning-based reconstruction method for fully 3D sparse data helical CT," *Physics in Medicine & Biology*, vol. 65, no. 24, p. 245030, 2020.

[7] Q. Zhang et al., "Artifact removal using a hybrid-domain convolutional neural network for limited-angle computed tomography imaging," *Physics in Medicine & Biology*, vol. 65, no. 15, p. 155010, 2020.

[8] H. Chen et al., "Learned experts' assessment-based reconstruction network (learn) for sparse-data ct,", 2017.

[9] H. Zhang, B. Liu, H. Yu, and B. J. I. T. o. M. I. Dong, "MetaInv-Net: Meta Inversion Network for Sparse View CT Image Reconstruction,", vol. 40, no. 2, pp. 621-634, 2020.

[10] A. Hauptmann, J. Adler, S. Arridge, and O. J. I. t. o. c. i. ktem, "Multi-scale learned iterative reconstruction,", vol. 6, pp. 843-856, 2020.

[11] B. Zhu, J. Z. Liu, S. F. Cauley, B. R. Rosen, and M. S. J. N. Rosen, "Image reconstruction by domain-transform manifold learning," *Nature*, vol. 555, no. 7697, pp. 487-492, 2018.

[12] J. He, Y. Wang, and J. J. I. t. o. m. i. Ma, "Radon inversion via deep learning," *IEEE Transactions on Medical Imaging*, vol. 39, no. 6, pp. 2076-2087, 2020.

[13] L. Shen, W. Zhao, and L. J. N. b. e. Xing, "Patient-specific reconstruction of volumetric computed tomography images from a single projection view via deep learning," *Nature Biomedical Engineering*, vol. 3, no. 11, pp. 880-888, 2019.

[14] X. Tao, Y. Wang, L. Lin, Z. Hong, and J. J. I. T. o. M. I. Ma, "Learning to reconstruct CT images from the VVBP-tensor," *IEEE Transactions on Medical Imaging*, 2021.

[15] Y. Guo et al., "Closed-loop matters: Dual regression networks for single image super-resolution," *IEEE/CVF Conference on Computer Vision and Pattern Recognition*, 2020, pp. 5407-5416.

[16] Y. Ye, Y. Chang, H. Zhou, and L. Yan, "Closing the Loop: Joint Rain Generation and Removal via Disentangled Image Translation," *IEEE/CVF Conference on Computer Vision and Pattern Recognition*, 2021, pp. 2053-2062.

[17] A. Paszke et al., "Pytorch: An imperative style, high-performance deep learning library,", vol. 32, pp. 8026-8037, 2019.

[18] G. Ongie, A. Jalal, C. A. Metzler, R. G. Baraniuk, A. G. Dimakis, and R. J. I. J. o. S. A. i. I. T. Willett, "Deep learning techniques for inverse problems in imaging," *IEEE Journal on Selected Areas in Information Theory*, vol. 1, no. 1, pp. 39-56, 2020.

[19] Low dose ct grand challenge. Available: https://www.aapm.org/grandchallenge/lowdosect/.

[20] D. P. Kingma and J. J. a. p. a. Ba, "Adam: A method for stochastic optimization," *Computer Science*, 2014.

[21] Y. Han and J. C. J. I. T. o. M. I. Ye, "Framing U-Net via Deep Convolutional Framelets: Application to Sparse-View CT," *IEEE Transactions on Medical Imaging*, p. 1418, 2018.

[22] C. Hu et al., "Low-Dose CT with a Residual Encoder-Decoder Convolutional Neural Network (RED-CNN)," *IEEE Transactions on Medical Imaging*, vol. 36, no. 99, pp. 2524-2535, 2017.

[23] Z. Zhang, X. Liang, D. Xu, Y. Xie, and G. J. I. T. o. M. I. Cao, "A Sparse-View CT Reconstruction Method Based on Combination of DenseNet and Deconvolution," *IEEE Transactions on Medical Imagin*, vol. 37, no. 6, pp. 1-1, 2018.

Hybrid Reconstruction Using Shearlets and Deep Learning for Sparse X-Ray Computed Tomography

Andi Braimllari, Theodor Cheslerean-Boghiu, Tobias Lasser

***Abstract*—In sparse X-ray Computed Tomography, the radiation dose to the patient is lowered by measuring less projections views compared to a standard protocol. In this work we investigate a hybrid approach combining shearlet representation with deep learning for reconstruction of sparse-view X-ray computed tomography. The proposed method is hybrid in that it reconstructs the parts that can provably be retrieved by utilizing a model-based approach, and it in-paints the parts that provably cannot through a learning-based approach. In doing so, we attempt to benefit from the best aspects of model- and learning-based methods. We demonstrate first promising results on publicly available data.**

***Index Terms*—sparse-view X-ray computed tomography, shearlets, deep learning.**

I. Introduction

X-RAY Computed Tomography (CT) is an essential technique that provides deep insight of a patient or an object of interest in a non-invasive manner. The forward model can be formulated as

$$Rf = y + \eta \tag{1}$$

where R represents the X-ray transform, f the quantity to be reconstructed (the absorption coefficients), while y represents the X-ray measurements and η some noise.

An important aspect of medical X-ray CT is the radiation exposure of the patients. One technique to lower the radiation dose is by lowering the number of projection views [14]. This is referred to as sparse-view or sparse X-Ray CT. The reconstructions of such sparse measurements tend to feature streak artifacts near edges tangent to the acquired X-rays [14]. With increasing sparsity and thus the increasing lack of measurement data, the severity of these streak artifacts also increases. Hence a reconstruction approach alleviating the impact of these streak artifacts is highly desirable.

The visibility principle [4] tells us in essence that the visible part of an object is comprised of the set of edges tangent to the acquired X-rays, and the invisible part is comprised of the edges non-tangent to these X-rays. Moreover, which edges can or cannot be reconstructed is dependent on the acquisition geometry, therefore known before acquisition.

Following [1], we leverage this principle by employing shearlets to resolve the wavefront set of such a signal. This enables us to properly reconstruct the visible edges using ℓ_1-regularization and to in-paint the invisible ones using deep learning. Using only model-based methods, by the visibility principle, we cannot retrieve the invisible information. On the other hand, using only deep learning on such an ill-posed problem, we might get satisfactory results up to a point, but we will not be able to certainly assert as to how much the original signal has changed [1].

Therefore, our proposed approach, which we will term **SDLX** for the remainder of this work, is a hybrid method that benefits from the best aspects of both model-based and learning-based approaches for sparse-view X-ray CT reconstruction. The SDLX method is very closely related to [1], which was developed for limited-angle X-ray CT. In the following we will present the SDLX method in detail, along with first results on a publicly available data set.

A. Braimllari was with and T. Cheslerean-Boghiu, T. Lasser are with Computational Imaging and Inverse Problems, Department of Informatics, School of Computation, Information, and Technology, Technical University of Munich, Germany, and with Munich Institute of Biomedical Engineering, Technical University of Munich, Germany.

II. Methods

A. Shearlets

Shearlets are a mathematical concept building on top of existing wavelet-theory components with distinct advantages. They represent a multi-scale framework that provides optimally sparse approximations of multivariate, anisotropic data. The approximation rate of shearlets is of $O(N^{-2})$, comparatively better than the $O(N^{-1})$ of wavelets. Shearlets are constructed by applying three operations, translation, dilation, and shearing, see [5] for details. They are applied to a single generating function ψ, resulting in a shearlet system

$$\psi_{a,s,t} = |\det M_{as}|^{1/2} \psi M_{as}(-t) \tag{2}$$

Here, $a \in \mathbb{R}_+$ dictates the dilation matrix A_a, $s \in \mathbb{R}$ dictates the shearing matrix S_s, while $t \in \mathbb{R}^2$ represents the translations. The composite matrix M_{as} is then defined as $M_{as} = A_a^{-1} S_s^{-1}$. The **continuous shearlet transform** is then

$$SH(f) = \langle f, \psi_{a,s,t} \rangle \tag{3}$$

For the **discerete shearlet transform**, we sample the parameter space $\mathbb{R}_+ \times \mathbb{R} \times \mathbb{R}^2$ at discrete points. This defines the **regular discrete shearlet system** as

$$SH(\psi) = \{\psi_{j,k,m} 2^{3j/4} \psi(S_k A_{2^j}), \forall (j,k,m) \in \mathbb{Z} \times \mathbb{Z} \times \mathbb{Z}^2\} \tag{4}$$

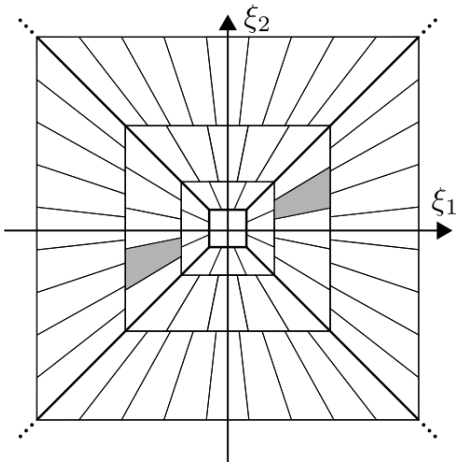
The **discrete shearlet transform** is defined similarly to the continuous scenario.

Given the directional bias exhibited by regular shearlets [8], we will instead be using the cone-adapted shearlets as they provide a remedy to it. For a visual representation of

the tiling that is generated in the Fourier domain, see Fig. 1. Additionally, we want to explicitly specify and emphasize, that the **cone-adapted discrete shearlet systems** as mentioned above, under mild assumptions, form a Parseval frame [9]. Based on this fact and the above statements, we know that

$$f = SH^T(SH(f)) \tag{5}$$

This equation represents a powerful reconstruction formula of the **discrete shearlet transform**, which is essential to our hybrid approach.



(a) f

Fig. 1: Frequency tiling of the **cone-adapted shearlet system**. By Afg genzel - Own work, CC BY-SA 3.0, https://commons.wikimedia.org/w/index.php?curid=27761187

By far the most important property of the shearlet systems to the hybrid approach is their ability to resolve the wavefront set of the signals at hand [1]. This allows us to differentiate the visible and invisible boundaries, and is accomplished by distinguishing different decay rates of the **shearlet transform**.

It is worth pointing out that we utilize classical shearlet systems, which are band-limited (compact support in the frequency domain [5]). Our implementation of such a **discrete cone-adapted band-limited shearlet transform** is based on [10], to which we refer the reader to for further details. We also utilize the **shearlet transform** available in [11] as an intermediary operation after running the ℓ_1-regularization.

B. Sparse Regularization with ADMM

Sparse regularization attempts to leverage the assumption that the output of a problem can be described by a fewer number of inputs, or put differently that for every output there exists a sparsifying representation system [1]. More specifically for low-dose CT, it has been shown that such methods enable more accurate reconstructions given very few tomographic measurements [1]. Therefore, such a paradigm is of interest to us for tackling sparse-view X-ray CT.

Alternating Direction Method of Multipliers (ADMM) is a general algorithm that works quite well in splitting the minimization of the sparsity-promoting ℓ_1-regularization term and the data fidelity term. Details on **ADMM** can be found in [12].

Based on (1), we are now able to construct and utilize shearlet-based sparse regularization. Explicitly expressing the reconstruction problem built so far, we write

$$\operatorname{argmin}_{f\geq 0} \frac{1}{2}\|Rf + y\| + \|SH(f)\|_{1,w} \tag{6}$$

in which $w \in \mathbb{R}_{\geq 0}^{W\times H\times L}$ represents the weights to the regularization term. This allows us to split the wavefront set of the signal into the visible and invisible parts, as described in the visibility principle. In this equation we have also explicitly specified a constraint for non-negative solutions, as it leads to better reconstructions [1].

We solve this minimization problem using **ADMM**.

C. Recovering the Invisible using Deep Learning

Deep Learning is one of the most influential paradigms of the last few decades with impressive results in a plethora of fields. In the recent years, considerable attempts have also been made towards the field of medical imaging as well. Many of the current model architectures and techniques pre-process the measurements or post-process the reconstructions, which can produce impressive results. However, it is not always immediately obvious that data fidelity was kept. In a medical setting this is not something that can be easily overlooked, as accuracy is a crucial.

In the hybrid approach that we are working on, the influence of deep learning is kept to a minimum. It is only used for inpainting missing information that can provably not be retrieved through classical model-based approaches. The architecture that we are using is **PhantomNet**, as proposed in [1].

PhantomNet is a fully-convolutional neural network based on one of the most prevalent architectures, **U-Net**. Different from **U-Net**, it is also a multi-channel input and multi-channel output network, based on the fact that it operates on the phase space and works with shearlet coefficients. More specifically, it takes a signal of shape (L, W, H) (e.g. $(61, 512, 512)$ and outputs a signal of the same shape. Here, W and H respectively represent the width and height of the image, while L dictates the number of layers of the shearlet coefficients. We refer the reader to [1] for full details on the architecture.

*D. The **SDLX** method*

Using the separate components of the hybrid approach outlined above, we now summarize all the steps that make up the **SDLX** method.

1) **Retrieve the visible coefficients**
   Compute ℓ_1-regularized solutions of the following problem
   $$g \in \operatorname{argmin}_{f\geq 0} \frac{1}{2}\|Rf + y\| + \|SH(f)\|_{1,w} \tag{7}$$
   by utilizing **ADMM** (or any other appropriate solver), which retrieves the visible coefficients based on the provided measurements. The input to this step are the sparse-view measurements y, while the output g is a reconstruction with sparse-view artifacts.

2) **Estimate the invisible coefficients**
We apply the **shearlet transform** to all of the images g generated above, which maps them from (W, H) to (L, W, H). The **PhantomNet** uses these shearlet coefficients as input, and it outputs objects of the same shape, which are the in-painted shearlet coefficients. After training **PhantomNet** ($\mathbf{PN}$), we use this model to estimate the invisible coefficients. If its weights are well adjusted, the following approximation should hold to a satisfactory threshold,

$$\mathbf{PN}(SH(g)) \approx SH(f)_{inv} \tag{8}$$

3) **Combine the visible and invisible coefficients**
Up until here we have the retrieved visible coefficients and a decent-enough estimation of the invisible coefficients (output of $\mathbf{PN}$). We sum them together and bring the entire output back to the spatial-domain through the **inverse shearlet transform**

$$f_{SDLX} = SH^T(SH(g)_{vis} + PN(SH(g))) \tag{9}$$

Here, f_{SDLX} is our end-result (of shape (W, H)), which contains the reconstruction of the sparse-view measurements along with the in-painting of the missing information.

The run-time of the proposed **SDLX** method is dominated by ADMM solving the ℓ_1-regularized problem. The implementations of our proposed algorithm are available at [3], [2].

III. Experiments and Results

The dataset used here is the one provided by the Mayo Clinic for the AAPM Low-Dose CT Grand Challenge [15]. It contains human abdomen scans with width and height of 512. We chose the low-dose scans, with pixel intensities in $[0 : 255]$.

For training of the PhantomNet we selected 10 patients, with the IDs of L004, L006, L014, L019, L033, L049, L056, L057, L058, and L064, which comprised a total number of 1525 scans. For testing we chose another patient, with an ID of L071.

To generate training pairs for the **PhantomNet**, we first simulate sparse, 64 projection view sinograms (over an arc of 360 degrees) of the 1525 training images using [2], adding 1% Gaussian noise. Then we compute ℓ_1-regularized reconstructions of those sinograms using ADMM as in (7), with 10 iterations of ADMM and 5 inner iterations of the conjugate gradient method on the normal equation. We manually selected the parameters of ADMM as $\rho_1 = 1/2$, $\rho_2 = 1$ (as in [1]), and $w = 0.001$. Afterwards, we apply the **forward shearlet transform** from [11].

We train the **PhantomNet** for 100 epochs in single-batches (e.g. one $(61, 512, 512)$ object as a batch) on a learning rate of $7e-5$ and weight decay of $1e-7$. The chosen optimizer is Adam. The loss function is the mean squared error loss from torch.nn.MSELoss.

For testing, we use the data from the patient with ID of L071, and simulate sparse-view sinograms with 64 projection views as above, adding 1% Gaussian noise. We execute the full **SDLX** method as in subsection II-D, using the same ADMM parameters as for training the phantom net. In the last step, we sum together the visible coefficients from the ℓ_1-regularization and the estimated invisible coefficients that the trained **PhantomNet** predicts, and apply the **inverse shearlet transform** to it.

TABLE I: Metrics of the 64-view reconstruction results on patient L071 compared to the ground truth. The lower the **RE**, the better. The higher the **PSNR**, **SSIM**, **HaarPSI**, the better.

	Metrics			
Method	RE	PSNR	SSIM	HaarPSI
f_{CG}	0.073	21.664	0.221	0.339
f_{ADMM}	0.061	22.409	0.246	0.352
f_{SDLX}	**0.026**	**26.001**	**0.271**	**0.626**

An example result f_{SDLX} of our proposed **SDLX** method is shown in Fig. 2 for one of the slices of patient ID L071, which was not seen during training. We also compare with the ground truth and reconstructions using the same ADMM as in the first step of SDLX, f_{ADMM}, and the result of an unregularized CG reconstruction using 10 iterations, f_{CG}.

It is apparent that this hybrid method is capable of in-painting the missing singularities for sparse-view CT. **SDLX** outperforms the other methods, a claim also supported by the metrics, as displayed in Table I. The metrics utilized are the Relative Error (RE), Peak Signal-to-Noise Ratio (PSNR), Structural Similarity Index Measure (SSIM), and Haar Wavelet-Based Perceptual Similarity Index (HaarPSI).

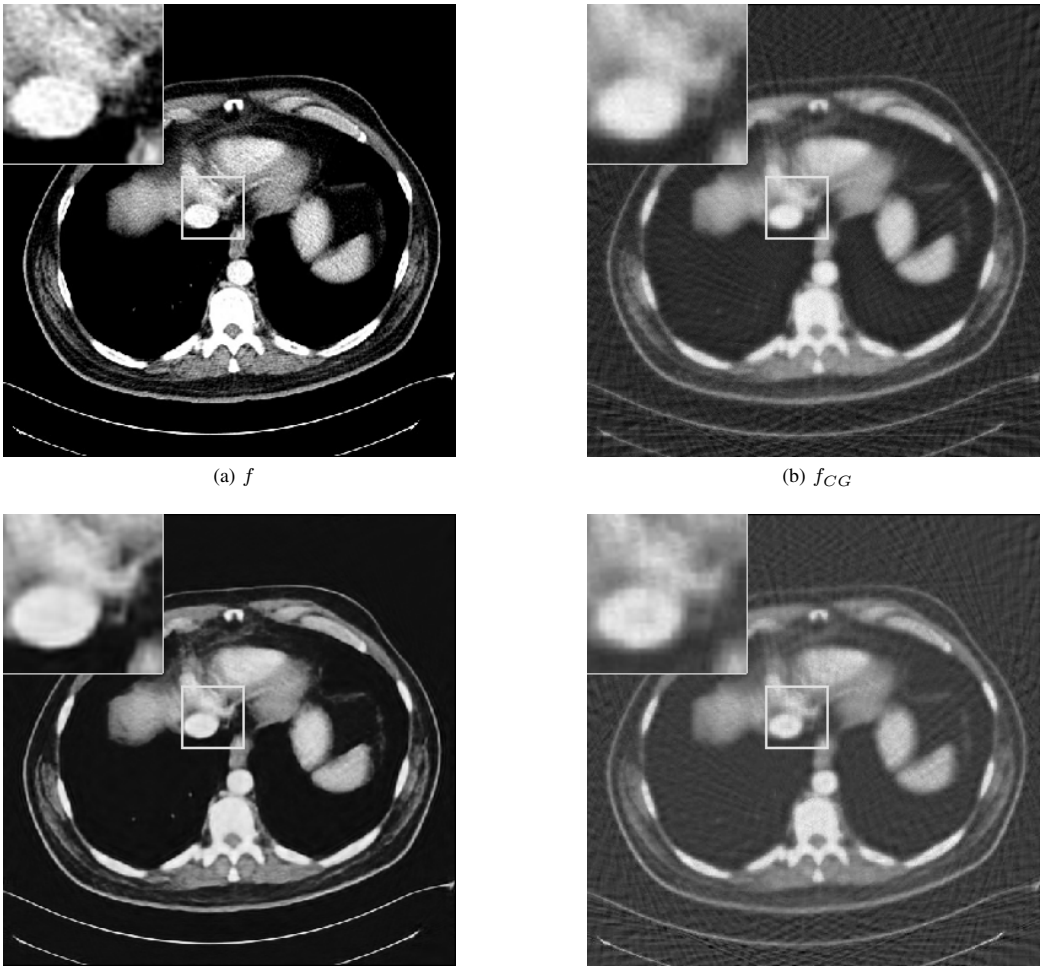
We also ran the same testing experiment without adding noise to the simulated sinograms. The trained model (on noisy simulated data) performed just as well on the unseen data, which serves as an indicator that **SDLX** is fairly robust towards noise.

IV. Discussion and Conclusion

The results from the experiment indicate that the hybrid approach works for sparse-view CT. However, a certain smoothing effect is also visible in the results. First experiments (not shown here) indicate that further tuning of the training of PhantomNet might negate this effect.

One aspect of the SDLX method that might introduce unexplained features is the deep learning step. Fortunately, this element is utilized here in a relatively controlled manner, given that it only handles the inference of the invisible coefficients. Further tuning of the hyper-parameters might be beneficial, as might be the study of more advanced models, such as transformers, instead of the U-net.

In summary, **SDLX** works because shearlets are capable of resolving the wavefront sets of the signals we are dealing with, and these decomposed coefficients adhere to certain rules which we can then learn. Adapting the work for limited-angle X-ray CT in [1], our first experiments for sparse-view X-ray CT on a publicly available data set show promising results.



(a) f (b) f_{CG}

(c) f_{SDLX} (d) f_{ADMM}

Fig. 2: 64-view reconstruction results on patient L071. The ground truth is f, while f_{SDLX} is the output of our proposed SDLX algorithm. For comparison, we also show reconstruction of the first step of the SDLX algorithm (f_{ADMM}) as well as an unregularized iterative CG reconstruction f_{CG}. The pixel intensities lie in $[0:255]$. For further details see section III.

ACKNOWLEDGMENT

We would like to express our gratitude to David Frank, a maintainer of the *elsa* library [2], for his code reviews.

REFERENCES

[1] T. A. Bubba and G. Kutyniok and M. Lassas and M. März and W. Samek and S. Siltanen and V. Srinivasan, *Learning The Invisible: A Hybrid Deep Learning-Shearlet Framework for Limited Angle Computed Tomography*. 2018.

[2] T. Lasser and M. Hornung and D. Frank, *elsa - an elegant framework for tomographic reconstruction*, 15th International Meeting on Fully Three-Dimensional Image Reconstruction in Radiology and Nuclear Medicine, 11072. International Society for Optics and Photonics, SPIE, Samuel Matej and Scott D. Metzler, 2019, 570 – 573, 10.1117/12.2534833, https://doi.org/10.1117/12.2534833.

[3] A. Braimllari, *SDLX, https://github.com/AndiBraimllari/SDLX*. 2021.

[4] E. T. Quinto, *Artifacts and Visible Singularities in Limited Data X-Ray Tomography*. 2017.

[5] G. Kutyniok and D. Labate, *Introduction to Shearlets*. 2012.

[6] E. J. Candès and D. L. Donoho, *New Tight Frames of Curvelets and Optimal Representations of Objects with C2 Singularities*. 2002.

[7] E. Candès and L. Demanet and D. Donoho and L. Ying, *Fast Discrete Curvelet Transforms*. 2006.

[8] K. Guo and G. Kutyniok and D. Labate, *Sparse Multidimensional Representations using Anisotropic Dilation and Shear Operators*. 2005.

[9] O. Christensen, *An Introduction to Frames and Riesz Bases*. 2003.

[10] S. Häuser and G. Steidl, *Fast Finite Shearlet Transform: a tutorial*, 3rd ed. 2014.

[11] M. Ronchetti and F. Voigtlaender, *Adaptive transform for manifold-valued data, https://github.com/dedale-fet/alpha-transform*. 2016.

[12] S. Boyd and N. Parikh and E. C. B. Peleato and J. Eckstein, *Distributed Optimization and Statistical Learning via the Alternating Direction Method of Multipliers*. 2011.

[13] O. Ronneberger and P. Fischer and T. Brox, *U-Net: Convolutional Networks for Biomedical Image Segmentation*. 2015.

[14] Y. Han and J. C. Ye, *Framing U-Net via Deep Convolutional Framelets: Application to Sparse-view CT*. 2018.

[15] C. McCollough, *TU-FG-207A-04: Overview of the Low Dose CT Grand Challenge*. 2016.

Photon Starvation Artifact Reduction by Shift-Variant Processing

Gengsheng L. Zeng[1,2]
[1]Department of Computer Science, Utah Valley University, Orem, UT 84058, USA
[2]Department of Radiology and Imaging Sciences, University of Utah, Salt Lake City, UT 84108 USA

e-mail: larry.zeng@uvu.edu

This work was supported in part by the National Institutes of Health (NIH) under Grant R15EB024283.

ABSTRACT The x-ray computed tomography (CT) images with low dose are noisy and may contain photon starvation artifacts. The artifacts are location and direction dependent. Therefore, the common shift-invariant denoising filters do not work well. The state-of-the-art methods to process the low-dose CT images are image reconstruction based; they require the raw projection data. In many situations, the raw CT projections are not accessible. This paper suggests a method to denoise the low-dose CT image using the pseudo projections generated by the application of a forward projector on the low-dose CT image. The feasibility of the proposed method is demonstrated by real clinical data.

INDEX TERMS Image processing, Image reconstruction, Biomedical imaging, Computed Tomography, Filters

I. INTRODUCTION

An immediate negative effect of using a low dose in CT imaging is that the images become noisy. The conventional denoising methods are based on the shift-invariant assumption. They can be implemented either in the spatial-domain as convolution methods or in the Fourier-domain as multiplication methods.

Shift-invariant filters can also be nonlinear. The nonlinear filters may outperform the linear filters in terms of sharp edge preservation.

Convolutional neural network (CNN) based methods can be very effective in removing noise from the images provided a large amount of noisy/noiseless image pairs are available to train the neural network.

This paper presents an effective nonlinear shift-variant procedure that does not need any image pairs to train. This proposed procedure blends the concepts of linear filtering, shift-variant filtering, and tomography. The feasibility and effectiveness of the proposed procedure are illustrated by its application to real clinical data.

II. METHODS

In this paper, we assume that the image x is already somehow reconstructed, for example, by the analytical filtered backprojection (FBP) algorithm. The image is noisy and contains photon starvation artifacts. The original measured projections are NOT available anymore.

A. The proposed algorithm

The proposed artifact reduction algorithm is introduced as follows.

Step 1. For a given image x_{old}, generate simulated pseudo projections as

$$p_i = a_i^T x_{old} \tag{1}$$

for all i.
Step 2. Select a threshold value T.
Step 3. Loop through all projections p_i.
If $p_i < T$, do nothing.
If $p_i \geq T$, replace p_i by its filtered version using a one-dimensional moving-average filter along the detector direction.
Step 4. Apply the filtered backprojection (FBP) algorithm to the processed pseudo projections, to obtain the final image x_{new}.

The threshold value T is a user-selected parameter, and we used T as the 75% of the maximum projection value in our study in this paper.

We now explain what motivates this algorithm. We do not choose any shift-invariant filters, because the artifacts are location and direction dependent. Since the state-of-the-art denoising algorithms are image reconstruction based, we choose an image reconstruction-based algorithm.

Our biggest obstacle is that we do not have an access of the original measurements in the projection domain. We only have a noisy reconstruction x_{old}. The simulated pseudo forward projection $a_i^T x_{old}$ is not the same as the originally measured projection.

The original projections due to noise are inconsistent. The inconsistency carries the noise information. The inconsistency information is lost in the forward projection $a_i^T x_{old}$. The objective function is already at its minimum with the pseudo projections because $p_i = a_i^T x_{old}$. Therefore, the strategy of selecting a set of weights to minimize the objective function does not help.

Realizing that the re-projected pseudo measurements do not carry the same information and do not have the same values as the original raw measurements, our novel strategy of this paper is to use the transmission data noise model to estimate the noise variance in the re-projected pseudo measurements. The FBP algorithm is selected to reconstruct the final image, because it is fast and easy to implement.

B. Image evaluation

The most common way to determine the effectiveness of artifact removal algorithms is by visual inspection or human observer studies. A quantitative evaluation metric adopted in this paper is the Sum Square Difference (SSD), defined as

$$SSD = \frac{\sum_{i,j}[X_{gold}(i,j) - X(i,j)]^2}{\sqrt{\sum_{i,j}[X_{gold}(i,j)]^2 \sum_{i,j}[X(i,j)]^2}}, \qquad (2)$$

where X_{gold} is the gold standard image, which is the FBP reconstruct from the regular-dose projections, and X is another image to compare with. The SSD essentially is the normalized distance between two images X_{gold} and X.

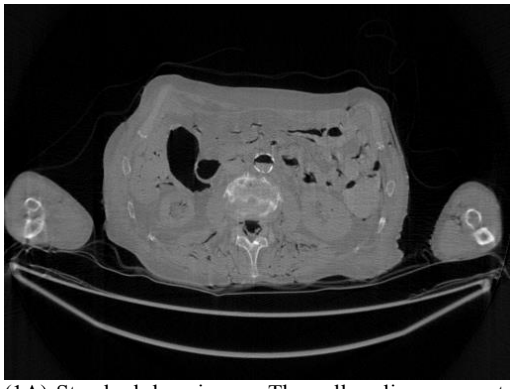
III. RESULTS

In this section, the methods are labeled with A – G. We point out that methods A and F use the 'unavailable' projections. In Figs. 1, the following labels are used for the images: (A) the gold standard image FBP reconstructed from the regular-dose x-ray projections; (B) the raw FBP reconstruction image reconstructed from the measured low-dose x-ray data; (C) the processed image using the proposed algorithm in the paper using the pseudo data; (D) the image is FBP reconstructed with a linear Hanning filter applied to the pseudo data; (E) the image is FBP reconstructed with a nonlinear bilateral filter applied to the pseudo data; (F) the image is post processing result of image from (B) with a BM3D filter in the image domain; (G) almost the same as (C) except that the 'unavailable' low-dose x-ray data is used instead of the pseudo data.

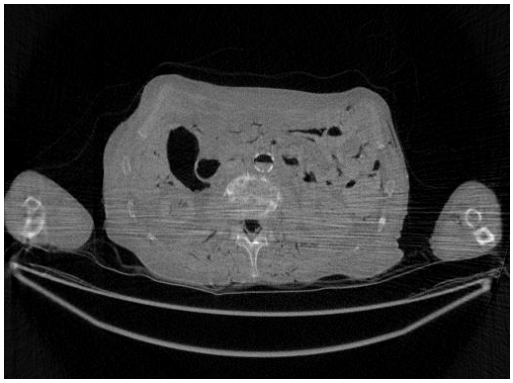
The numerical results of the Sum Square Difference (SSD) values are listed in Table 1. The SSD is a non-negative quantity, the smaller value the better. The ideal SSD value is 0. In all these cases, the proposed method gives the smallest SSD values, indicating the best performance.

The projection-domain images (also known as the sinograms) are displayed in Fig. 2. The images are (a) the 'unavailable' regular-dose projections, (b) the difference between the raw 'unavailable' low-dose projections and the 'unavailable' regular-dose projections, (c) the difference between the pseudo forward projections from the low-dose FBP reconstruction and the 'unavailable' low-dose projections, and (d) the difference between the processed version of the pseudo forward projections from the low-dose FBP reconstruction and the unprocessed version, respectively. It is observed from Fig. 2d that the proposed method only alters a very small portion of the projections.

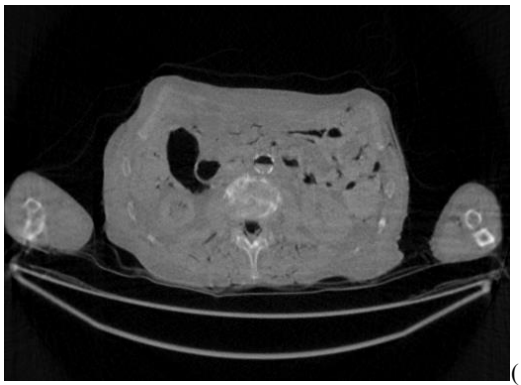
The proposed algorithm is effective in reducing the streaking artifacts and keeping the image resolution. As a comparison, the images produced by a linear Hanning filter, a nonlinear bilateral filter, or a BM3D filter are unable to to keep small details while the streaking artifacts are still severe.



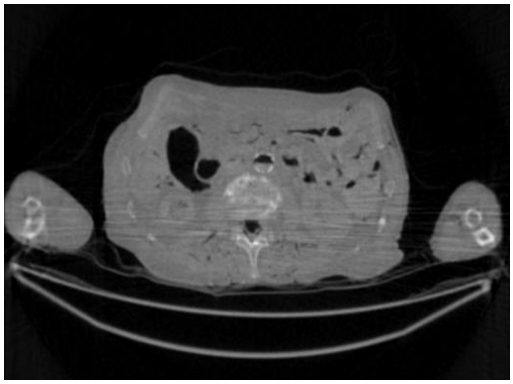
(1A) Standard dose image. The yellow line segment indicates the path that the line profiles are taken along in Fig. 14.



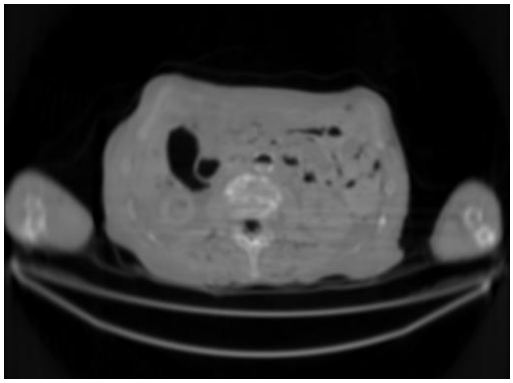
(1B) Low-dose image



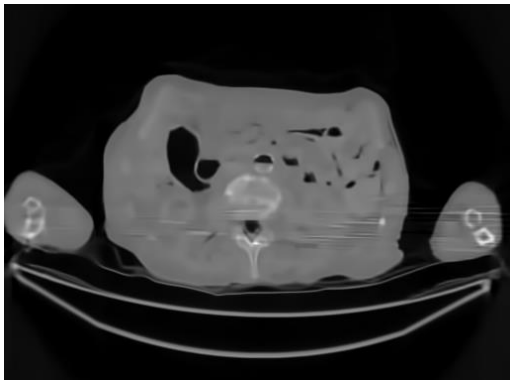
(1C) Low-dose image processed by proposed algorithm



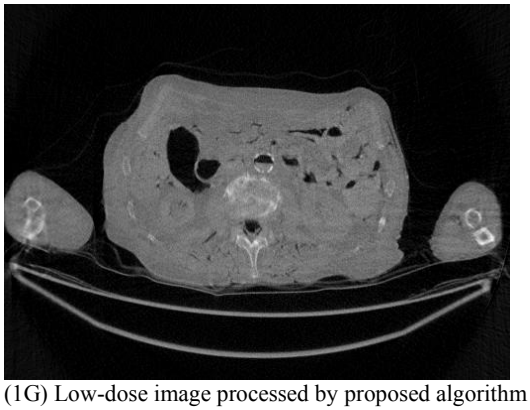
(1D) Low-dose image processed by FBP-Hann



(1E) Low-dose image processed by a bilateral filter



(1F) Low-dose image processed by a BM3D filter

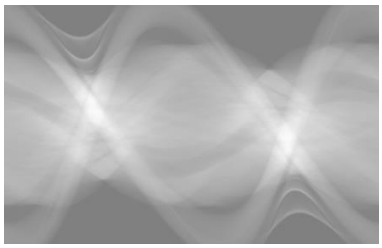


(1G) Low-dose image processed by proposed algorithm using the 'unavailable' measured projections

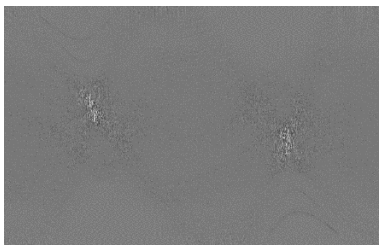
Figure 1. Processed and unprocessed images. The standard-dose image in (1A) is the gold standard. The image with the proposed method (1C) gives the best result among all images using the low-dose raw image (1B).

Table 1. Full width at half maximum value comparison

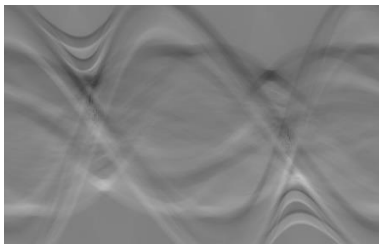
Method	FWHM (pixels)	Severe Artifacts?
A. Regular-dose FBP	2.54	No
B. Low-dose FBP, using the 'unavailable' low-dose measurements	2.34	yes
C. Low-dose FBP using proposed method	2.96	No
D. Low-dose FBP using linear Hann filter	3.76	Yes
E. Low-dose FBP using nonlinear bilateral filter	7.00	Yes
F. Measured Low-dose using BM3D filter	2.55	Yes
G. Low-dose FBP using proposed method, but using the 'unavailable' low-dose measurements	2.34	No
H. Low-dose FBP, using the pseudo data	2.98	No



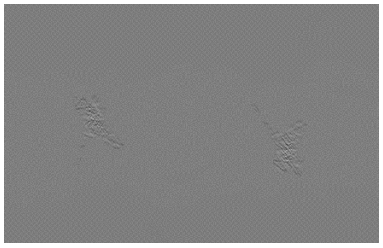
a. The sinogram for the regular-dose projections



b. The difference between the low-dose projections and the regular-dose projections



c. The difference between the low-dose projections and the forward projections of the FBP reconstruction from the low-dose projections



d. The difference between the forward projections of the FBP reconstruction from the low-dose projections and the processed projections by using the proposed algorithm

Figure 2. Sinogram domain images.

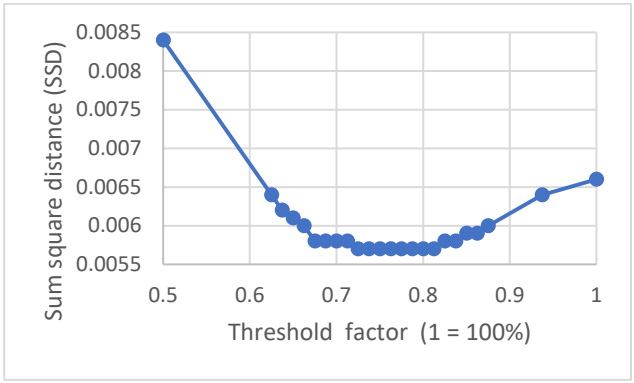


Figure 3. A sensitivity study of the SSD with respect to the threshold value *T*.

IV. DISCUSSION

When a noisy reconstructed image is available while the original projection measurements are no longer available, the pseudo re-projected line integrals are not helpful to reduce noise if a conventional iterative image reconstruction algorithm is to be used. The conventional iterative image reconstruction algorithms work in the principle of reducing the data fidelity term. By using the pseudo re-projected line integral data, this data fidelity term is already at its minimum value, which is zero.

One way of denoising is to stop the iterations early. This approach is equivalent to lowpass filtering, which is almost shift invariant. As we demonstrated in the Results section, shift-invariant denoising smooths the images but still cannot reduce the streaking artifacts.

A filter is referred to as shift-*in*variant if the filter operation is the same everywhere. In our proposed filter, the filter operation is only applied to a small amount of selected pseudo projections. Therefore, our proposed filter is shift variant.

Our proposed algorithm is NOT an iterative image reconstruction algorithm; it is an analytic FBP algorithm with a nonlinear pre-filter. In the FBP algorithm, a ramp filter (which is a high-pass filter) must be used to cancel the bachprojection blurring. The purpose the low-pass filter is to reduce the noise in the image. The application of a low-pass filter is optional in FBP, only when image denoising is necessary. The main goal of this paper is photon-starvation artifact reduction, we do not apply a linear low-pass filter in the FBP algorithm. In the proposed algorithm, there is a threshold value *T*; any pseudo projection data value that is less than this threshold value is not affected. The majority of the pseudo projections are less than this threshold. Thus, the image resolution degradation is kept to its minimum.

The proposed algorithm contains a user-determined hyper parameter *T*. This hyper parameter *T* is determined by trial and error. In fact, parameter *T* is not very sensitive. As shown in Fig. 3, the SSD vs *T* curve has a flat valley, which means that a wide range of the parameter *T* can give the optimal solution.

Three noise-reducing filters have been used to compare with the proposed shift variant filter in the task of photon starvation artifact reduction. Those three noise-reducing filters do not perform well for this task. If the filters are adjusted to remove the artifacts, many image details are removed as the price to pay. The message of our paper is that the noise reduction task is different from the artifact reduction task. For artifact reduction, where to filter (or equivalently, where not to filter) is far more important than what filter to use. Once the region to be filtered is identified, many filters are effective as long as the filters have enough smoothing power. We choose the simplest linear moving-average filter with a large enough kernel size. Other noise reduction filters such as bilateral and BM3D filters will work just fine when applied only in the specified region. The critical point is that we do not apply the lowpass filter to the entire image or the entire sinogram.

V. CONCLUSIONS

We have developed an effective method to reduce the photon starvation streaking artifacts in low-dose x-ray CT images. The proposed method is shift-variant; it only applies lowpass filtration for some pre-determined measurement values in the sinogram domain.

We assume that the raw, low-dose measurements are not available, and the noisy reconstruction is available. A set of pseudo re-projections are generated from noisy reconstruction. A threshold value *T* is selected by the user using a trial-and-error method. The pre-determined measurements are selected if the pseudo measurement value is greater than *T*. The pre-selected pseudo measurements are filtered in the sinogram domain by a simple moving-average lowpass filter along the detector direction. The FBP algorithm is performed to generate a final image using the selectively filtered pseudo measurements.

ACKNOWLEDGMENT

The author would like to thank Raoul M. S. Joemai of Leiden University Medical Center for providing raw projection data of the cadaver CT scans.

Data-driven Metal Artifact Correction in Computed Tomography using conditional Generative Adversarial Networks

Nele Blum[1], *University of Lübeck[1], Germany,* Thorsten M. Buzug[1,2] and Maik Stille[1,2], *Fraunhofer IMTE[2], Germany,*

***Abstract*—Metal objects in the field of view cause artifacts in the image, which manifest as dark and bright streaks and degrade the diagnostic value of the image. Standard approaches for metal artifact reduction are often unable to correct these artifacts sufficiently or introduce new artifacts. We propose a new data-based method to reduce metal artifacts in CT images applying conditional Generative Adversarial Networks to the corrupted data. A generator network is applied directly to the corrupted projections by the metal objects to learn the corrected sinogram data. Further, two discriminator networks are used to evaluate the image quality of the enhanced data from the generator. The method was initially developed based on a supervised approach. However, there is usually no ground truth for actual clinical data without artifacts, which is needed to train the networks. Therefore, the method was further improved to train an unsupervised network, i.e., without the ground truth. In addition the input data, the neighboring slices and the stochastic components of the image are included using the latent space representation of the data. The results show that the trained generator network can reasonably replace the missing projection data and reduce the artifacts in the reconstructed image.**

I. Introduction

ARTIFACTS are particularly apparent in computed tomography (CT) when high-density objects like metal implants or surgical instruments are present in the field of view. Various metal artifact reduction (MAR) methods have been proposed since the first publications on MAR [1]. Projection completion methods are often used due to their simplicity and fast application. They treat the data affected by metal objects as missing image information and replace them with synthetic data, usually obtained by interpolation [2]. A significant drawback of this method, especially in inhomogeneous image regions, is the loss of information in the metal trace. An alternative to this approach are iterative methods [3], yet they have the disadvantage of high computation times, especially for large 3D images. Recently, neuronal networks have been used to correct corrupted data [4], [5]. To train the networks effectively, the definition of a suitable loss function is crucial. Still, loss functions are often designed to optimize specific, quantifiable image parameters, even if a selection of image properties to describe good image quality is usually hard to define. However, instead of specifying parameters for a good image quality, it is possible to use an additional network. Besides the so-called generator network, which generates the improved projection data, a second network can be trained, the discriminator, to distinguish the actual projections from the synthetic data.

The first results using this method have already been published and showed the potential of the new method on the reduction of artifacts in the image compared to conventional approaches [6]. One major drawback of the published method, however, is the use of ground truth data to train the networks.

Prajot et al. propose a model for image inpainting, which allows learning the distribution of reconstructed images in a completely unsupervised setting by integrating the stochastic component that introduces an explicit dependency between this component and the generated output in the learning process [7]. We further developed this approach and incorporated it into our previous work. Different from Prajot's approach, it can be assumed that additional information is available in the neighborhood of the corrupted metal projections. This information can be used in the training process to get more realistic predictions from the generator network.

N.Blum is with the Institute of Medical Engineering, University of Lübeck, 23556 Germany, e-mail: (see blum@imt.uni-luebeck.de).

T.M.Buzug and M.Stille are with Fraunhofer Research Institution for Individualized and Cell-Based Medical Engineering IMTE, Lübeck, 23556 Germany and the Institute of Medical Engineering, University of Lübeck, 23556 Germany

II. Methods

To train the networks, a combination of different loss functions is used. In the objective function

$$G^* = \arg\min_{G,E}\max_{D1,D2} L^{\text{GAN}}(G, D1) + L^{\text{GAN}}(G, D2) + \lambda_z L^z(G, E) + \lambda_y L^y(G, D1, E) \quad (1)$$

several networks interact with each other. For the training process, the loss function is minimized by the generator G and the encoder E and maximized by the discriminators $D1$ and $D2$. No access to the metal-free projection data is usually available for the networks training. This is especially the case if the networks are to be trained on actual clinical data. Hence, in the first part of the loss function, the output data of the generator network are not used directly. Instead, a binary mask of the metal trace is first applied to the generated data using a function F, allowing the output data to be compared with the original input data. This is realized by applying an adversarial loss function

$$L_{\text{GAN}}(G, D1) = \mathbb{E}_y[\log(D1(y))] + \mathbb{E}_{y,m,z}[\log(1 - D1(F[G(y, z) + y, m]))] \quad (2)$$

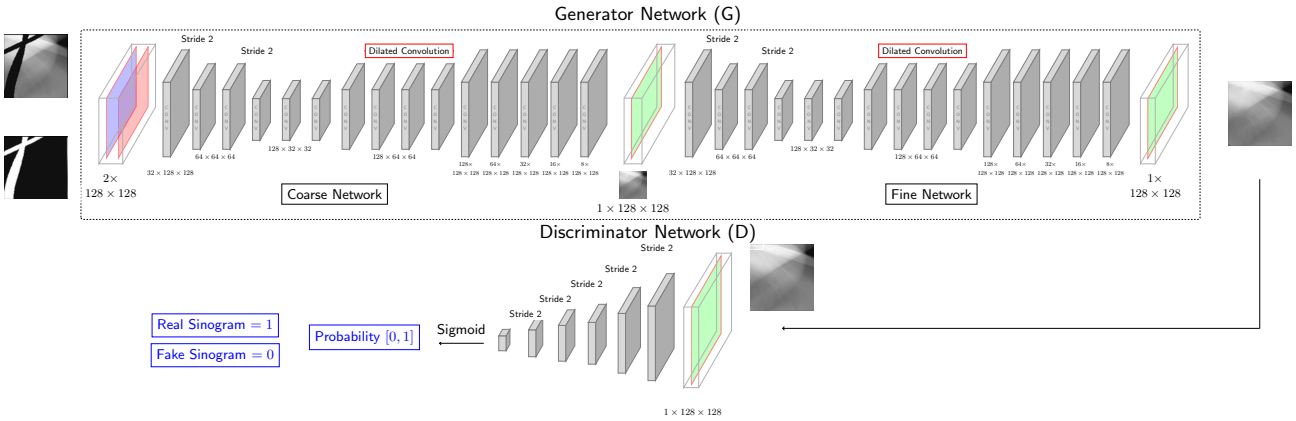


Fig. 1. Schematic representation of the used network architectures.

in which the generator and the first discriminator $D1$ compete against each other. Here, both networks are optimized based on a Min-Max game over the expectation values $\mathbb{E}$ of the input data y and the masked output data of the generator.
In addition to the incomplete or corrupt sinogram sections, there are also sections without the influence of a metal object. By searching for slices in the image in front of and behind the metal trace, these sections can be used as a reference for another discriminator network $D2$. In the second part of the loss function, again an adversarial loss is used in which the generator and the second discriminator compete against each other. The loss

$$L_{\text{GAN}}(G, D2) = \mathbb{E}_{x^*}[\log(D2(x^*))] + \mathbb{E}_{y,m,z,x^*}[\log(1 - D2(x^*, G(y, z)))] \quad (3)$$

is calculated now over the expectation values $\mathbb{E}$ of the input data y and the reference data x^* from the neighbor slices to integrate the information from the uncorrupted projection data into the learning process. In both parts of the objective function, the generator receives the incomplete sinogram data and the corresponding metal trace as an input. Using a two-stage architecture consisting of a coarse and a fine network [8], the generator generates an output image that is used as an input to the discriminator networks. Each discriminator network, a binary classification network, assigns a probability between $[0, 1]$, indicating whether the data are real or artificially generated compared to the original input data or the data from the neighbor slices, respectively. The structures of the generator and discriminator network architectures are represented in Figure 1. To stabilize the training process, the loss is extended by two complementary losses, which help to enforce the dependency on the stochastic component weighted by the parameter λ. First, an encoding z-loss is used to force the generator to use information from the stochastic component z from the latent representation of the given data and the masked output data of the generator network by adding

$$L^z(G, E) = \mathbb{E}_{y,m,z} \left\| z - E(F[G(y, z) + y, m]) \right\|^2 \quad (4)$$

to the objective function. However, as shown in [9] and [10], this loss exhibits a "stenography behavior" and is not sufficient alone to stabilize the network training.

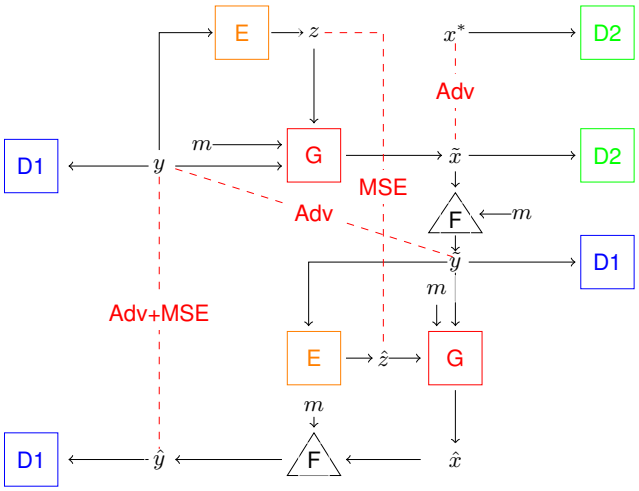


Fig. 2. Schematic representation of the used network interactions. The input data y are given together with a binary mask m and the latent representation into the generator G. The output data $\tilde{x}$ are then used as an input to the discriminator $D1$, by applying as masked function F, and the discriminator $D2$. Additionally, the masked output data $\tilde{y}$ are given again into the generator together with the latent representation $\tilde{z}$ and the binary mask. The output $\hat{y}$ of the second pass is given into $D1$, and an adversarial loss is calculated together with the MSE-error to the original input data. Further, both latent vectors z and $\tilde{z}$ are used to calculate an MSE-loss to update the generator and encoder network.

Therefore, an additional loss function, the encoding y-loss, is used. Here, the output data of the generator $\tilde{y}$ are not used directly but are given after the application of the function F, again into the generator. The masked output data of the second pass are then compared again with the original input data y by calculating

$$\begin{aligned} L^y(G, D1, E) = {} & \mathbb{E}_y[\log(D1(y))] \\ & + \mathbb{E}_{\tilde{y},\tilde{z}}[\log(1 - D(F[G(\tilde{y}, \tilde{z}) + \tilde{y}, m]))] \\ & + \mathbb{E}_{y,\tilde{z}} \left\| y - F[G(\tilde{y}, \tilde{z}) + \tilde{y}, m] \right\|^2 \end{aligned} \quad (5)$$

using an MSE-loss to constrain $G(\tilde{y}, \tilde{z})$ to be close to y and let their distribution be similar via an adversarial loss. Figure

TABLE I
TRAINING PARAMETERS

Parameter	Value
number training data	183 000
number validation data	31 000
batch size	32
number epochs	250
learning rate	$1 \cdot 10^{-5}$

2 shows the interaction between the four networks and the before described realization of the calculated loss functions for the network training.

III. Experiments

Simulated data from a software phantom [11] with varying parameter settings were used to train and test the networks. Objects with different shapes were inserted in 120 different data sets at random positions. Using simulations for 3D cone-beam CT, the metal affected projection data were generated. In the image domain, the metal objects were segmented and projected forward to obtain the metal trace, which is used to remove the metal corrupted data. The generated data were divided into training (90), validation (15), and test (15) data. To generate more training data and data with a higher variation, image sections of 128×128 were used instead of the full 3D sinograms. Training was performed using the training parameters listed in table I.

IV. Results

After successfully training the networks, the generator network can be applied to the test data. As shown in Figure 3, the network is able to reconstruct the missing data from the input data consisting of the data with the missing metal trace and the binary mask. Compared to the ground truth, hardly any differences can be seen. In the left part of Figure 4, an example slice from a completed 3D sinogram from the trained generator network is shown compared to the original input data above. Thereby, the artificially generated data by the generator network can barely be distinguished from the real surrounding data, making it difficult to detect the previous existing metal trace. The right side of the figure shows the resulting reconstruction obtained by an FDK reconstruction compared to the reconstruction from the original metal inserted data. Compared to the original data, it can be shown that the presence of artifacts could be reduced significantly by applying the generator network.
In Figure 5, the results of the generator network are shown compared to a linear interpolation for cylindrical hollow objects, which the network had not seen yet in the training process. It can be shown that the network, in contrast to the interpolation, is able to reconstruct the inner structures of these kinds of objects. The mean squared error (MSE) between the reconstructed images and the ground truth was calculated on the test data set for validation. As a result, the network was able to learn most of the missing structures, but some streak artifacts remain visible in the reconstructed image (Figure

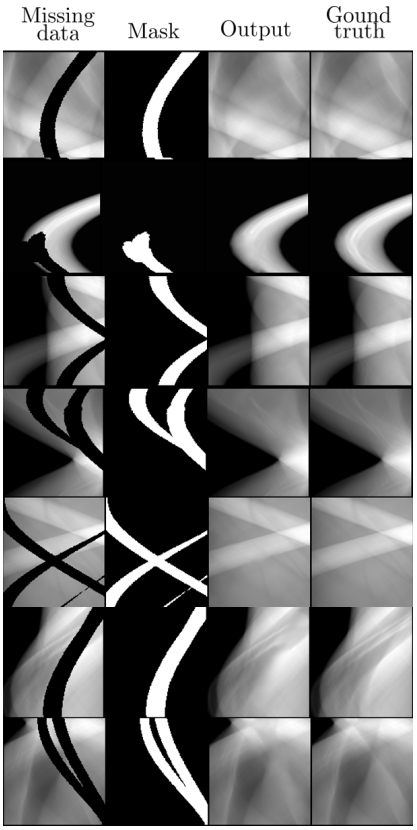


Fig. 3. Example data from the test data set completed by the generator network.

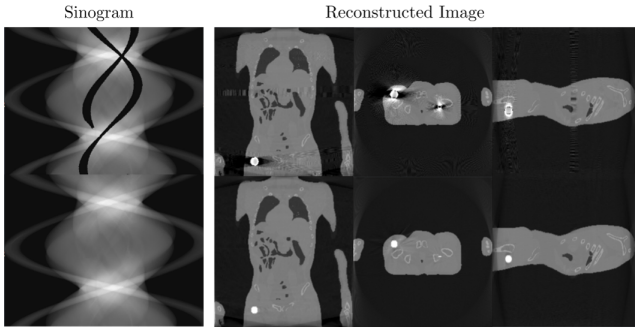


Fig. 4. Data before (top) and after application of the network (below).

6). Compared to linear interpolation, the average MSE value overall test images is recognizably lower with $7.7{*}10^{-9}$ versus $1.0{*}10^{-8}$. As well compared to the normalized metal artifact reduction (NMAR), the average MSE value is slightly smaller with $7.7 * 10^{-9}$ versus $7.9 * 10^{-9}$.

V. Discussion

The developed method shows the first promising results on the simulated test data set. However, a detailed study of the influence of different training parameters is still missing. In particular, the weighting of the individual components of the

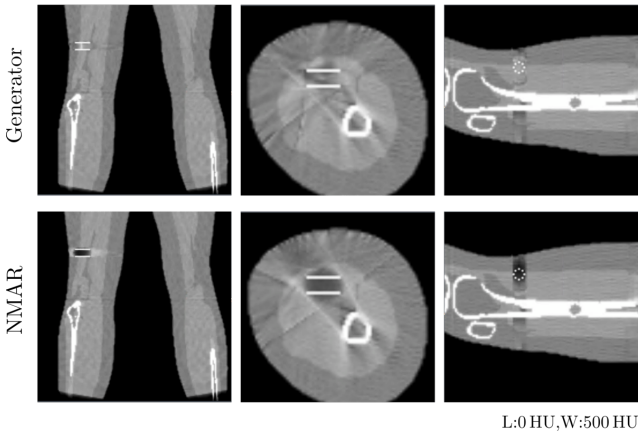


Fig. 5. Generator results (top) compared to normalized metal artifact reduction (below) for a cylindrical hollow object.

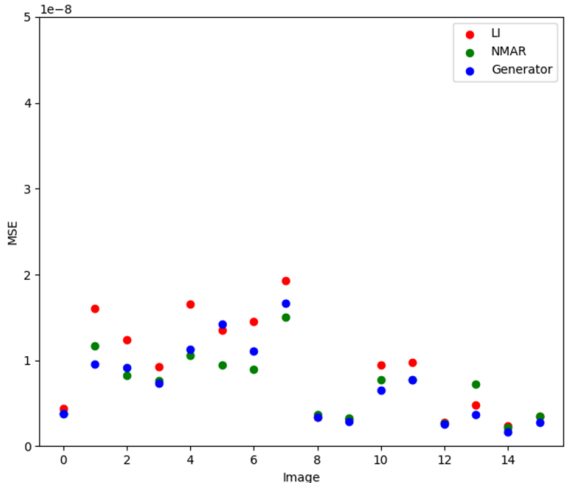


Fig. 6. Calculated MSE values between the reconstruction results of the generator and the ground truth compared to NMAR and linear interpolation.

loss function could further improve the training results.

One of the next steps is the integration of actual clinical data into the training process of the networks. Here, both the influence of different acquisition geometries on the network training, as well as the used resolution should be investigated. Increased or decreased resolution could lead to the necessity of adjusting the patch size for training the networks. Also, a validation method must be developed that is applicable to actual clinical data where, unlike for the simulated data, the ground truth is not available.

Further, the corrupted data should be integrated into the training process, as there might be more information about the image structures near the metal object available. One way to realize this is to replace the binary mask as an input of the network with the original data of the metal trace. Here, for example, a weighting depending on the position would also be conceivable, which could be integrated into the network convolution. Besides the integration of the original data, it is also possible to take advantage of the fact that 3D data are available so that the input data used as well as the networks can be adapted to 3D.

VI. Conclusion

The results demonstrate that the generator is able to replace the missing image information in the sinogram and reduce a vast number of artifacts in the reconstructed image. Furthermore, by developing a method to train the networks without using any ground truth, the networks can be applied to real clinical data in the next step without the need for major modifications. In the future, the corrupted projection data should be used as an input of the generator network instead of using the binary mask of the metal trace. The network should be able to use this information from the metal trace to avoid introducing false image structures.

Acknowledgment

This research was partially supported by TANDEM the competence center for medical technology supported by the European Union and the State of Schleswig-Holstein (grant no. 122-09-024) and TOMEDEX (Federal Ministry of Education and Research under grant number BMBF 13GW0371C).

References

[1] G. H. Glover and N. J. Pelc, "An algorithm for the reduction of metal clip artifacts in CT reconstructions," *Medical Physics*, vol. 8, no. 6, pp. 799–807, 1981. [Online]. Available: https://aapm.onlinelibrary.wiley.com/doi/abs/10.1118/1.595032

[2] W. J. H. Veldkamp, R. M. S. Joemai, A. J. van der Molen, and J. Geleijns, "Development and validation of segmentation and interpolation techniques in sinograms for metal artifact suppression in CT," *Medical Physics*, vol. 37, no. 2, pp. 620 – 628, 2010. [Online]. Available: https://doi.org/10.1118/1.3276777

[3] M. Stille and T. M. Buzug, "Augmented likelihood image reconstruction with non-local prior image regularization," in *Proc. 4th Intl. Mtg. on image formation in X-ray CT*, 2016, pp. 145–8.

[4] M. U. Ghani and W. Karl, "Fast Enhanced CT Metal Artifact Reduction using Data Domain Deep Learning," *IEEE Transactions on Computational Imaging*, vol. PP, pp. 1–1, 2019.

[5] C. Peng, B. Li, M. Li, H. Wang, Z. Zhao, B. Qiu, and D. Z. Chen, "An irregular metal trace inpainting network for x-ray CT metal artifact reduction," *Medical Physics*, vol. 47, no. 9, pp. 4087–4100, 2020. [Online]. Available: https://aapm.onlinelibrary.wiley.com/doi/abs/10.1002/mp.14295

[6] N. Blum, T. Buzug, and M. Stille, "Projection Domain Metal Artifact Reduction in Computed Tomography using Conditional Generative Adversarial Networks," *MIDL*, 2021.

[7] A. Pajot, E. de Bézenac, and P. Gallinari, "Unsupervised adversarial image inpainting," *CoRR*, vol. abs/1912.12164, 2019. [Online]. Available: http://arxiv.org/abs/1912.12164

[8] J. Yu, Z. Lin, J. Yang, X. Shen, X. Lu, and T. S. Huang, "Generative Image Inpainting with Contextual Attention," *arXiv preprint arXiv:1801.07892*, 2018.

[9] A. Almahairi, S. Rajeswar, A. Sordoni, P. Bachman, and A. C. Courville, "Augmented cyclegan: Learning many-to-many mappings from unpaired data," *CoRR*, vol. abs/1802.10151, 2018. [Online]. Available: http://arxiv.org/abs/1802.10151

[10] C. Chu, A. Zhmoginov, and M. Sandler, "Cyclegan, a master of steganography," 2017. [Online]. Available: https://arxiv.org/abs/1712.02950

[11] W. Segars, M. Mahesh, T. Beck, E. Frey, and B. Tsui, "Realistic CT simulation using the 4D XCAT phantom," *Medical Physics*, vol. 35, pp. 3800–3808, 2008. [Online]. Available: https://aapm.onlinelibrary.wiley.com/doi/abs/10.1118/1.2955743

CT-Value Conservation based Spatial Transformer Network for Cardiac Motion Correction

Xuan Xu, Peng Wang, Liyi Zhao, Guotao Quan*

***Abstract*—Artifact correction is a great challenge in cardiac imaging. During the correction of coronary tissue with motion-induced artifacts, the spatial distribution of CT value not only shifts according to the motion vector field (MVF), but also shifts according to the volume change rate of the local voxels. However, the traditional interpolation method does not conserve the CT value during motion compensation. A new sample interpolation algorithm is developed based on the constraint of conservation of CT value before and after image deformation. This algorithm is modified on the existing interpolation algorithms and can be embedded into neural networks with deterministic back propagation. Comparative experimental results illustrate that the method can not only correct motion-induced artifacts, but also ensure the conservation of CT value in the region of interest(ROI) area, so as to obtain corrected images with clinically recognized CT value. Both effectiveness and efficiency are proved in forward motion correction process and backward training steps in deep learning. Simultaneously, the visualized motion vector field transparentizes the correction process, making this method more interpretable than the existing image-based end-to-end deep learning method.**

***Index Terms*—Interpolation, Cardiac Motion Correction, Convolutional neural networks**

I. Introduction

RECENT study shows that cardiovascular disease is still the largest worldwide. Coronary Computed tomography angiography(CCTA) is a crucial technology to diagnose coronary heart disease as a simple, fast, noninvasive and safe imaging method. However, the beating characteristics, especially patients with high heart rates, introduce motion artifacts to the reconstruction, which significantly decreases the quality and confidence of the image and potentially limits the evaluation of coronary arteries or even makes misinterpretion. Existing technologies try to suppress artifacts from both hardware and software. Limited by the physical and mechanical properties of CT equipment, even the small incremental gain of the frame rotation time needs to make great efforts in engineering design.

Some methods try to improve image quality during reconstruction [1] [2].Rohkohl et al [3]initially proposed a Metric-based correction method and later improved and extended. Some registration-based also have shown good performance in compensating for strong motion artifacts.A classical non-rigid registration algorithm [4] uses the motion vector field estimated by bidirectional labeled point matching (BLPM) algorithm to perform 3D warping on a series of partial reconstructions. This algorithm uses thin plate spline interpolation algorithm (TPS) for interpolation. TPS, as a very robust spatial data interpolation method, was introduced by Duchon et al. [5] into geometric design, which is commonly used for non-rigid registration. Since the structure of TPS is differentiable, Spatial Transformer Networks (STN) [6] applies it in the network to achieve spatial alignment of feature maps.

Deep learning based cardiac motion correction method, as a particular case of image deblurring, usually follows two common ways: using deep network to estimate the motion vector field, and then combined with the traditional warp algorithm to deform. Another way is to learn from image to image, that is, the trained neural network can output the corrected image directly. Methods proposed by S. Jun [7] and N. Fu [8] have successfully proved that CNN has the ability to generate and learn coronary motion patterns. [9] [10] has successfully applied STN for end-to-end training. The obtained images can be well registered in shape, but due to the limitations of traditional interpolation in value conservation, the accuracy of the CT value needs to be investigated. Based on the principle of CT imaging, the overall integral value of the reconstruction image is not related to the states of motion of object in the fixed Field of View. However, various system biases may be introduced in the reconstruction process, resulting in differences between motion and static reconstruction results.

Separating raw data to generate multiple partial angle reconstructions and applying different MVFs with affine transformation is one way to eliminate the interpolation issue, but it requires more detailed and exact motion patterns for each subset. In order to solve the interpolation issue and meanwhile avoid increasing the complexity of correction process, the conservation integration constraint interpolation method is designed. This paper takes into account the proportional coefficient between the integral value and the area of deformation grid. This new deformation interpolation method is based on the existing interpolation method and can be embedded into the classical spatial transform network for back propagation.

II. MATERIALS AND METHODS

The outline of the proposed coronary correction pipeline is shown in Fig.1. Firstly, the front Network is designed to output three sets of deformation parameters θs.Then the deformation parameters of each pixel in the whole image are obtained by interpolation of θs, and are used to warp the coronary ROI. Simultaneously, calculate the deformation coefficient of the

Corresponding author: Guotao Quan* is with the Shanghai United Imaging Healthcare Co., Ltd email: guotao.quan@united-imaging.com

Xuan Xu is with the Shanghaitech University

Peng Wang and Liyi Zhao is with the Shanghai United Imaging Healthcare Co., Ltd,2258 Chengbei Rd, Jiading District, Shanghai China

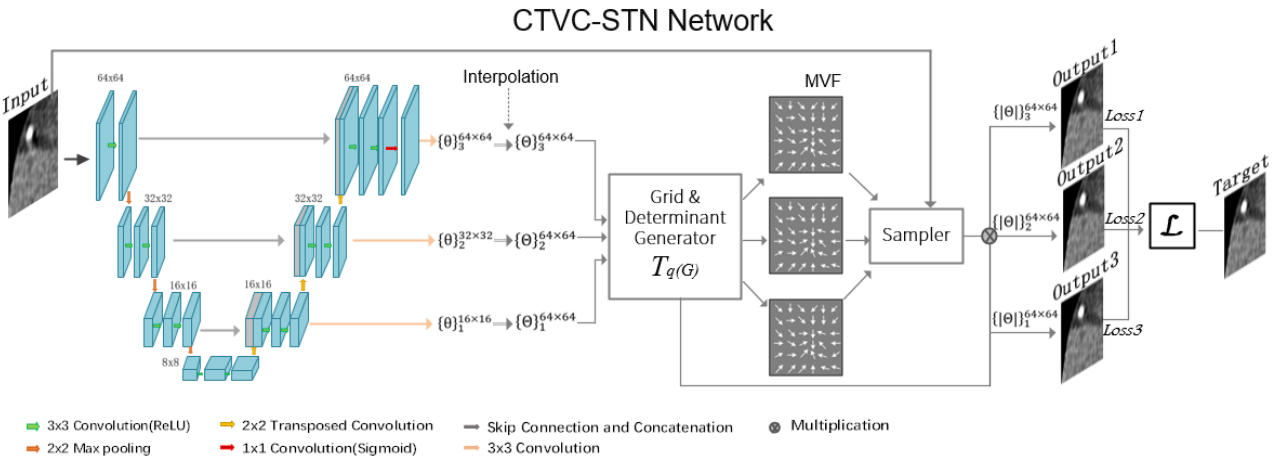


Fig. 1: Patch-based coronary correction pipeline of the proposed interpolation method. The front network receives an input which can be central axial, sagittal, or coronal slices of a motion-corrupted volumetric ROI and generate deformed parameters $\theta_1, \theta_2, \cdots, \theta_k$. CTVC-STN network feed with $\theta_1, \theta_2, \cdots, \theta_k$ to interpolate to get $\Theta_1, \Theta_2, \cdots, \Theta_n$ $(k \leq n)$ and sampling to output an corrected image.

area of the deformed grid. Then the deformation coefficient is multiplied by each corresponding pixel to obtain the final image which is the closest to the ground truth, not only in shape but also in CT value. The whole process can be back-propagated in the convolutional Network.

A. Deformation parameters estimation

Inspired by literature [11], a changed 2D-UNet network with deep supervision [12] is selected as the front network to generate deformation parameters $\theta_1, \theta_2, \cdots, \theta_k (k \geq 1)$ in decoder path, and it can be substituted by other suitable network structures. Since convolutional layers of different depths have different receptive fields, different from Unet [13], this network simultaneously outputs the learned deformation parameters from features extracted at different scales. For the specific coronary artery correction task in this paper, the features of three different scales are selected to estimating deformation parameters at the same time and calculate loss respectively, and finally the total loss is calculated by the combination of the three losses. This configuration can not only increase the stability of the network during training, but also support pruning the network during testing, which can increase the testing speed while ensuring the correction accuracy, thereby reducing the amount of network parameters within a controllable range.

B. CT Value Conservation Network based on Spatial Transformer (CTVC-STN)

In order to realize CT value conservation while ensuring the deformation, CTVC-STN network is proposed by improving on the basis of Spatial transformer Network as STN has shown some deficiencies in the end-to-end training of motion correction. First, STN introduces full connection layers to output an affine matrix θ with the size of 2×3 which increases the difficulty of training and limits that STN can only be used for small-size features.Secondly, STN uses conventional affine transformation and interpolation function to warp, in which the interpolation function can be bilinear interpolation, bicubic interpolation and thin plate spline interpolation. However, for coronary images with artifacts caused by different motion patterns, the simulated data as Fig.2 show that, ideally, the sum of CT value of the stationary and motion reconstructions are conserved. Therefore, this paper designs a deformation interpolation network CTVC-STN to keep the CT value conserved.

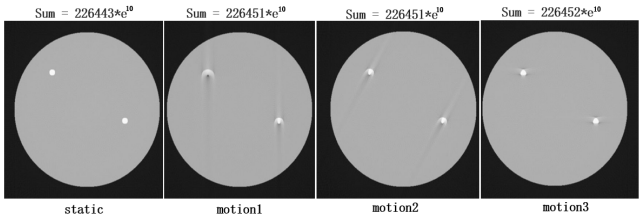


Fig. 2: One simulated case of the stationary reconstruction and motion reconstruction.The first image is the result of stationary reconstruction, and the rest three motioned images are reconstructed from different gantry starting positions and reconstruction angles. The CT value sums of the four images are approximatly equal. The CT value difference among them comes from system resolution error.

As illustrated in Fig.1, The Grid & Determinant Generator feeds with deformation parameters θs = $\theta_1, \theta_2, \cdots, \theta_k$,which are generated by front network. According to $\theta_1, \theta_2, \cdots, \theta_k$, input image is averagely discretized into k pixels. For the presupposed conditions, $\theta_1, \theta_2, \cdots, \theta_k$ are the accurate deformation parameters corresponding to k pixels, so the coordinates of these points can be directly obtained. Furtherly, combined with sampling and bilinear interpolation, the coordinates and pixel values of all points in the initial deformed image are also obtained. Backward mapping is used for sampling, that is, the pixel value of each point of the deformed image is traversed to find the corresponding coordinates on the original image, and then the surrounding pixels are used for simple interpolation. This method avoids holes generated during forward mapping. To realize CT value conservation of an image before and after deformation, ideally, when one pixel p of deformed image is contributed by q pixel grids on the original image where $q > 1$

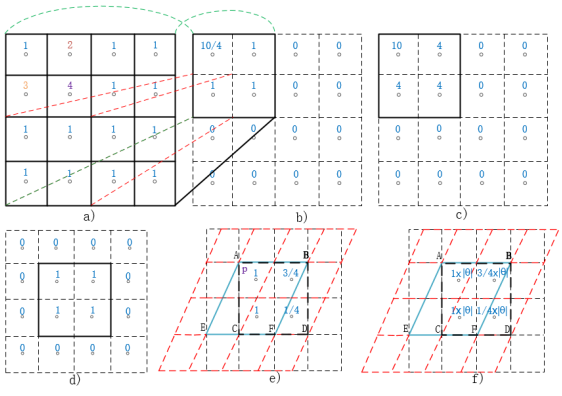


Fig. 3: A sample interpolation example. a) initial image of 4x4. b) result of bilinear interpolation, 2x2. c) result of proposed interpolation method with CT value-sum conservation. d)initial image with 2x2 focused ROI. e) the result of bilinear with deformation parameters θ. f)the result of proposed method. b),c) is still in the focused ROI, while e),f) not, so CT value conservation can be obtained in b),c), but invalid in e),f).

,as shown in Fig.3 e), $q > 2$, for these q pixels, calculating the ratio of covered area of each pixel grid of initial image. Then, the sum of the pixel values of these q grids multiplied by the corresponding ratio should be the exact pixel value of that pixel. An alternative solution is used here to alleviate the situation that computational complexity increase as initial image size becomes larger. To our knowledge, in the two-dimensional space, the geometric mathematical significance of the determinant of the matrix represents the directed area surrounded by two vectors [14]. Based on this mathematical theory, the area of each pixel in the affine transformation grid can be obtained.

A sample interpolation example is shown in Fig.3 a)-c). Geometrically, consider the pixels of the image as squares rather than points and the pixel value as the center points of the input's corner pixels. Fig.3 a) is the initial image with a size of 4x4, Fig.3 b) is the result of bilinear interpolation and sampling with scaling factor of 1/2. As it illustrated in Fig.3 b), the summation of all pixel value is non-conservation. According to the affine matrix of each pixel of this transformation and its determinant are as follows:

$$\theta = \begin{pmatrix} 2 & 0 & \delta x \\ 0 & 2 & \delta y \end{pmatrix}, \|\theta\| = 4 \tag{1}$$

Therefore, the value summation of the image obtained by multiplying Fig.3 b) by $\|\theta\|$ is equal to the initial image, as shown in Fig.3 c).

The proposed method maintains CT value conservation meets the following constraints: First, if the deformed image exceeds the size of the deformed grid, the boundary pixels will be lost in the sampling process,which are illustrated in Fig.3 e),f). Second, at the ideal limit resolution, even very exaggerated deformations will become very smooth. Therefore, the method can realize the conservation of CT value under ideal conditions. However, due to the difficulty of implementation and computational complexity, very precise grids are not used during implementation, which will cause errors. However, the following experiments show that the method can control the error within the clinically acceptable range, as shown in Fig.5.

III. RESULTS

A. Image acquisition

The combination network above is performed using supervised end-to-end training strategy. To train this model, the ground truth used in the training process are anonymous motion-artifacts-free cases with United Imaging Healthcare(UIH) uCT ATLAS devices. Referring to the forward model for simulating cardiac motion method proposed in related literature [15] [16] and our knowledge of cardiac beating patterns, artificial motion vector fields is generated to simulate all kinds artifacts. The artificial blurred data are input data for training. 9600 samples of 2D coronary patches with the size of 64x64 based on the above artifact simulation methods were generated. The samples were divided into 80% training data and 20% validation data. Test data set involves real motion blurred cases to examine the effectiveness of network and simulation method.

B. Neural Network Training

The training was performed on an NVIDIA TITAN RTX for 1000 epochs using an Adam optimizer with 0.1 decay, The batch size is 16 and the loss function is Mean Square Error(MSE).

C. Evaluation

Real data from several clinical patients with severe artifacts were used to test the trained network model.

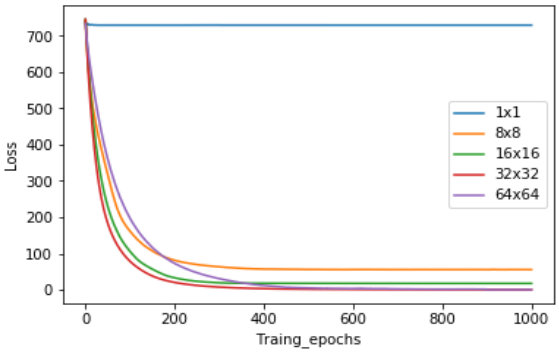


Fig. 4: The trend of different numbers of deformation parameters on the loss convergence as the training epoch increase. The loss is Mean Square Error.

The number of deformation parameters directly affects the accuracy of artifact correction. Fig.4 shows the trend of different deformation parameters on the loss convergence as the training epoch increases. It can be seen that since coronary motion is a relatively complex non-rigid deformation, it is impossible to correct the deformation of the entire image with a single parameter, so the loss is maintained at a relatively high level. With the increase of deformation parameters, the network can gradually learn complex motion deformation, and its number is positively correlated with the correction result.

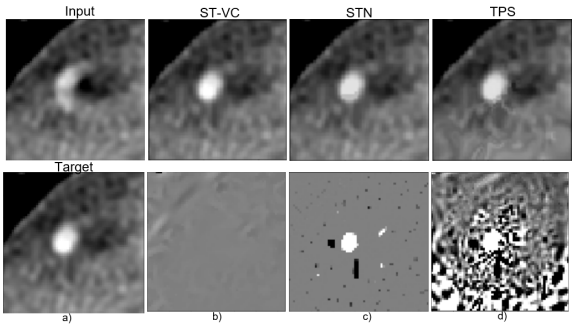


Fig. 5: The x-y plane of one image patch is visualized before and after different methods. Among these, CTVC-STN shows best correction both in shape and CT value. The CT value sum of Input is 3914156, CTVC-STN: 3914182, STN-biliear: 3909930, STN-TPS: 3908637, and Target: 3914188. CTVC-STN: can obtain an aproximate value coservation with a value error ratio of 1e-6. Figure b),c),d) show the differences between ground truth and the Results of STN-bilinear c), STN-TPS d) , our method b).

When the number of θ generated by the network is the same as the number of pixels in the input image, it is equivalent that each pixel has its own specific displacement vector, and a more accurate shape correction can be achieved under this configuration.

When Mean Square Error (MSE) and structural similarity index measure (SSIM) were used as the loss function, the above corrected CT value will be slightly deviated. The network needs to add the directed area of the deformation vector for further numerical correction. As shown in Fig.5, TPS and STN failed to maintain the image CT value conservation before and after correction, while CTVC-STN can achieve approximate conservation of CT value within the range of loss not exceeding 1e-6. Real cases were also tested to demonstrate the effectiveness of this method as shown in Fig.6, which shows that the designed network has a good performance in correcting drastic artifacts of coronary images in three planes.

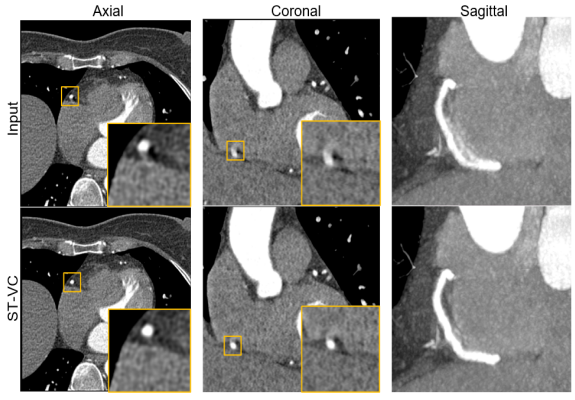


Fig. 6: One clinical case to show that the proposed pipeline is robust in related artifacts in three planes images of patient's coronary .

IV. Conclusion

The key contribution of this work is the solution that provides individual deformation of each pixel of the image, and can maintain the approximate conservation of the CT value in deformation. This method is a supplement to the existing interpolation algorithm and can be used in the network to support back propagation. A novel framework for motion correction of CCTA were experimented to verify the validity of this method. Compared with the existing interpolation method such as bilinear interpolation and TPS, it can get images with more accurate CT value and lower MSE. For the further works, we would focus on improving the network capacity that allows a full field of view image as input and realize self-attention to the regions of coronary or anywhere artifacts appear. The second improvement would take place to extend this method into a 3D network to achieve direct 3D coronary volume correction.

References

[1] U. Van Stevendaal, J. Von Berg, C. Lorenz, and M. Grass, "A motion-compensated scheme for helical cone-beam reconstruction in cardiac ct angiography," *Medical Physics*, vol. 35, no. 7Part1, pp. 3239–3251, 2008.

[2] A. A. Isola, M. Grass, and W. J. Niessen, "Fully automatic nonrigid registration-based local motion estimation for motion-corrected iterative cardiac ct reconstruction."

[3] C. Rohkohl, H. Bruder, K. Stierstorfer, and T. Flohr, "Improving best-phase image quality in cardiac ct by motion correction with mam optimization," *Medical Physics*, vol. 40, no. 3, 2013.

[4] R. Bhagalia, J. D. Pack, J. V. Miller, and M. Iatrou, "Nonrigid registration-based coronary artery motion correction for cardiac computed tomography," *Medical physics (Lancaster)*, vol. 39, no. 7, pp. 4245–4254, 2012.

[5] J. Duchon, "Splines minimizing rotation-invariant semi-norms in sobolev spaces," in *Constructive Theory of Functions of Several Variables*, ser. Lecture Notes in Mathematics. Berlin, Heidelberg: Springer Berlin Heidelberg, 2006, pp. 85–100.

[6] M. Jaderberg, K. Simonyan, A. Zisserman, and K. Kavukcuoglu, "Spatial transformer networks," *CoRR*, vol. abs/1506.02025, 2015. [Online]. Available: http://arxiv.org/abs/1506.02025

[7] S. Jung, S. Lee, B. Jeon, Y. Jang, and H. J. Chang, "Deep learning cross-phase style transfer for motion artifact correction in coronary computed tomography angiography," *IEEE Access*, vol. PP, no. 99, pp. 1–1, 2020.

[8] N. Fuin, A. Bustin, T. Küstner, I. Oksuz, and C. Prieto, "A multi-scale variational neural network for accelerating motion-compensated whole-heart 3d coronary mr angiography," *Magnetic Resonance Imaging*, vol. 70, 2020.

[9] N. Fish, R. Zhang, L. Perry, D. Cohen-Or, and C. Barnes, "Image morphing with perceptual constraints and stn alignment," 2020.

[10] I. Yoo, D. Hildebrand, W. F. Tobin, W. Lee, and W. K. Jeong, "ssemnet: Serial-section electron microscopy image registration using a spatial transformer network with learned features," 2017.

[11] Z. Zhou, M. Siddiquee, N. Tajbakhsh, and J. Liang, "Unet++: Redesigning skip connections to exploit multiscale features in image segmentation," *IEEE Transactions on Medical Imaging*, vol. 39, no. 6, pp. 1856–1867, 2020.

[12] C.-Y. L. Ğ01FSaining Xie Ğ01FPatrick W. Gallagher, "deep1y supervised nets."

[13] O. Ronneberger, P. Fischer, and T. Brox, "U-net: Convolutional networks for biomedical image segmentation," in *International Conference on Medical Image Computing and Computer-Assisted Intervention*, 2015.

[14] B. Kolman and A. Shapiro, "Matrices and determinants," *Algebra for College Students (Revised and Expanded Edition)*, pp. 419–440, 1982.

[15] "Deep-learning-based ct motion artifact recognition in coronary arteries," in *SPIE Medical Imaging Conference*.

[16] J. Maier, S. Lebedev, J. Erath, E. Eulig, S. Sawall, E. Fournié, K. Stierstorfer, M. Lell, and M. Kachelrie, "Deep learning-based coronary artery motion estimation and compensation for short-scan cardiac ct," *Medical Physics*.

Exploiting voxel-sparsity for bone imaging with sparse-view cone-beam computed tomography

Emil Y. Sidky, Holly L. Stewart, Christopher E. Kawcak, C. Wayne MacIlwraith, Martine C. Duff, and Xiaochuan Pan

***Abstract*—An optimization-based image reconstruction framework is developed specifically for bone imaging. This framework exploits voxel-sparsity by use of ℓ_1-norm image regularization and it enables image reconstruction from sparse-view cone-beam computed tomography (CBCT) acquisition. The effectiveness of the voxel-sparsity regularization is enhanced by using a blurred image representation. Ramp-filtering is included in the data discrepancy term and it has the effect of acting as a preconditioner, reducing the necessary number of iterations. The bone image reconstruction framework is demonstrated on CBCT data taken from an equine metacarpal condyle specimen.**

***Index Terms*—Voxel sparsity, CBCT image reconstruction, sparse-view, and equine imaging**

I. Introduction

THE majority of sparsity-exploiting image reconstruction techniques for sparse-view cone-beam computed tomography (CBCT) have employed total-variation (TV) as a sparsity regularizer because gradient sparsity is an effective prior for X-ray based imaging [1], [2]. Prior to this work, it was suggested that voxel-sparsity itself could be useful for application with few-view CBCT for imaging sparse structures such as blood vessel trees in CT angiography [3]. When applying sparsity-regularization to bone imaging, deciding between gradient and voxel sparsity is not obvious. Bone tissue occupies more volume than truly sparse structures such as blood vessels but less volume than the soft tissue. Bones also have high-contrast fine trabecular structure. Porosity will increase the number of non-zeros in a gradient-magnitude image but it will improve voxel sparsity, thus favoring the latter. Further complicating this decision is that some of the fine bone structures may not be resolvable by the CBCT system. Even though the underlying bone tissue may be porous on a scale of ≈ 10 microns or less, it will appear uniform at standard CBCT resolution, ≈ 200 microns, thus favoring the use of gradient sparsity.

In a prior study, we employed a micro-CT scan of an equine limb sample to generate a high-resolution computer phantom for studying sparsity regularization for image reconstruction in sparse-view CBCT [4]. In that work it was found that use of voxel sparsity regularization was more effective at reducing the streaks due to view-angle undersampling in CBCT. A model was developed that expresses the reconstructed volume in terms of a blurred voxel-sparse image.

E. Y. Sidky and X. Pan are with the Department of Radiology, University of Chicago, Chicago, IL, 60637 USA.

H. L. Stewart, C. E. Kawcak, and C. W. MacIlwraith are with the Department of Clinical Sciences, College of Veterinary Medicine and Biomedical Sciences, Colorado State University, Fort Collins, CO, 80532 USA.

M. C. Duff is with the Savannah River National Laboratory, Aiken, SC, 29808 USA

Fig. 1. FBP reconstructed slice images from the full CBCT dataset consisting of 720 views. The shown slices are for orthogonal planes that intersect at the lucency indicated by the red arrows. The image on the right corresponds to a trans-axial slice. The grayscale window is [0.02,0.32] cm^{-1}.

In this work, we apply this model to an actual CBCT scan of a horse limb sample. A full 720-view CBCT dataset is collected and sparse-view configurations are generated by subsampling this dataset. In Sec. II, we review the optimization-based model used for image reconstruction in sparse-view CBCT of bone tissue. In Sec. III, reconstruction slice images are shown demonstrated the parameter dependences of the algorithm. We conclude this abstract in Sec. IV.

II. The optimization-based framework for CBCT bone imaging

The proposed CBCT bone image reconstruction framework is based on the following optimization problem

$$\min_f \frac{1}{2}(XB_wf-g)^\top R(XB_wf-g) \text{ such that } \|f\|_1 \leq \gamma, \quad (1)$$

where the CBCT data and image volume are represented by an m-dimensional vector g and n-dimensional vector f, respectively; the $m\times n$ matrix X denotes CBCT projection; the $n\times n$ matrix B_w is Gaussian blurring using a width parameter w; the $m\times m$ matrix R is block-diagonal, where each block performs ramp filtering on each projection view; and γ is the constraint parameter on the volume image ℓ_1-norm. The ramp filter R has two purposes: (1) it accelerates convergence as demonstrated in Ref. [5], and (2) it can reduce artifacts due to low-frequency inconsistencies in the projection in a manner similar to the approach in Ref. [6]. The actual reconstructed volume is represented by B_wf and voxel sparsity is enforced on the underlying image f; in this way images with smooth transitions between tissue types can be represented with a high-degree of voxel sparsity [7].

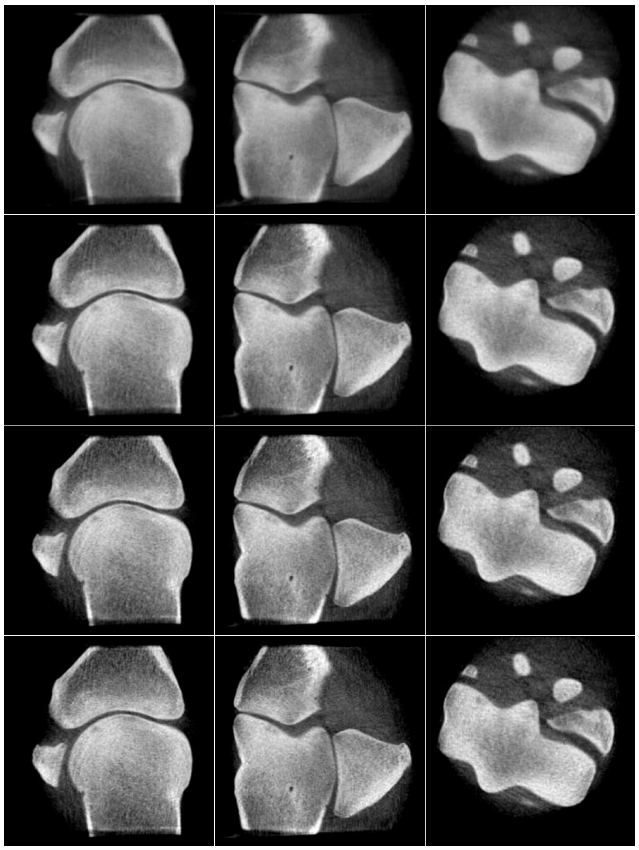


Fig. 2. Progression of the volume images with iteration number for reconstruction from a 64-view dataset spanning a 192 degree arc. The algorithm settings are $\gamma = 5.0 \times 10^6$, 0.0166 cm^{-1}/voxel, and $w = 0.2085$ mm, 1 voxel width. The shown slice images correspond to the same slices in Fig. 1. The rows correspond to the reconstructed volume at iteration 10 (1st row), 20 (2nd row), 50 (3rd row), and 100 (4th row). The grayscale window is [0.02,0.32] cm^{-1}.

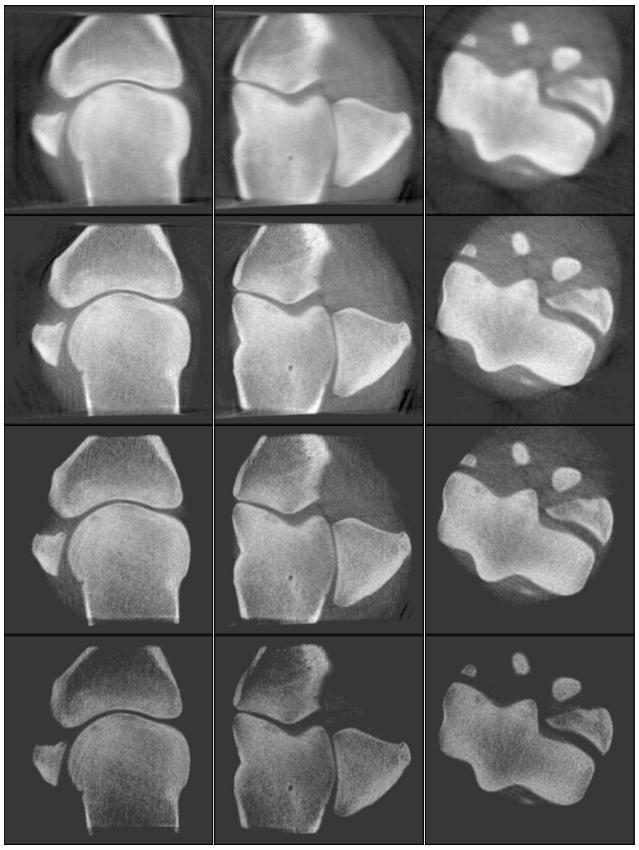


Fig. 3. Impact of varying γ on the volume images for image reconstruction from a 64-view dataset spanning a 192 degree arc. The other algorithm parameters, the iteration number and blur kernel width, are set to 50 iterations and $w = 0.2085$ mm, 1 voxel width, respectively. The γ constraint values are set to 8.0×10^6 (1st row), 6.0×10^6 (2nd row), 4.0×10^6 (3rd row), and 2.0×10^6 (4th row). The grayscale window is set wider, at [-0.1,0.35] cm^{-1}, than that of the other figures in order to appreciate the numbers of non-zero voxels.

To solve Eq. (1), we employ the primal-dual (PD) algorithm developed by Chambolle and Pock [8], [9]. To facilitate our PD implementation, we absorb the filtering operation into a combined system matrix and Eq. (1) is modified to

$$\min_f \frac{1}{2}\|M_w f - Sg\|_2^2 \text{ such that } \|f\|_1 \leq \gamma,$$

where

$$M_w = SXB_w, \text{ and } R = S^2.$$

The matrix S is the square root of the the ramp filter matrix, which can be computed readily in the frequency domain where the ramp filtering matrix is diagonal. The resulting algorithm has three control parameters: the iteration number, the ℓ_1-norm constraint parameter γ, and the voxel blur width w. We demonstrate the impact of all of these parameters on image reconstruction from sparse-view CBCT data of an equine limb specimen.

III. Results

We apply the bone image reconstruction algorithm to a CBCT data set acquired on an Epica Pegaso veterinary CT scanner. Three sparse-view configurations are investigated with a short-scan arc of 192 degrees; we consider 128, 64, and 32 views evenly spaced over the short-scan arc which corresponds to an angular spacing of 1.5, 3, and 6 degrees, respectively. All of the sparse-view CBCT datasets are subsampled from a full 720-view circular scan, and accordingly the modeled exposure decreases with the projection view number. The detector size is 1088x896 detector pixels, where each pixel is $(0.278\text{mm})^2$ in size. Image volumes are reconstructed onto a 768x768x512 voxel grid using cubic voxels with a width of 0.2085 mm. A reference volume is reconstructed by use of filtered back-projection (FBP) applied to the full 720-view dataset and shown in Fig. 1. The red arrows in the figure demonstrate subchondral bone lysis typical of a palmar osteochondral lesion in racehorses that are of clinical significance.

Varying the iteration number

In the first set of results for using the bone sparsity image reconstruction algorithm, we subsample the full 720-view dataset to a 64-view dataset over a 192 degree scanning arc.

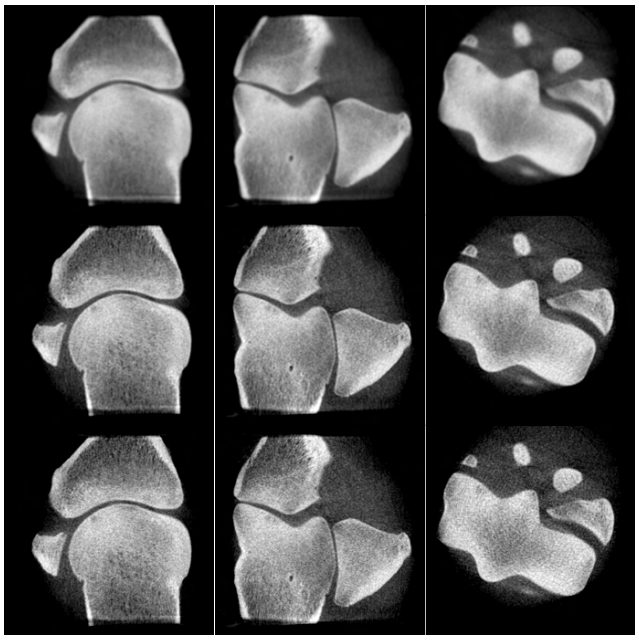


Fig. 4. Slice images of reconstructed volumes from the 128-view dataset using $w = 2 \times 0.2085$ mm (Top row) $w = 0.2085$ mm (Middle row), and $w = 0.0$ mm (Bottom row) at 50 iterations and $\gamma = 5.0 \times 10^6$. The grayscale window is [0.02,0.32] cm^{-1}.

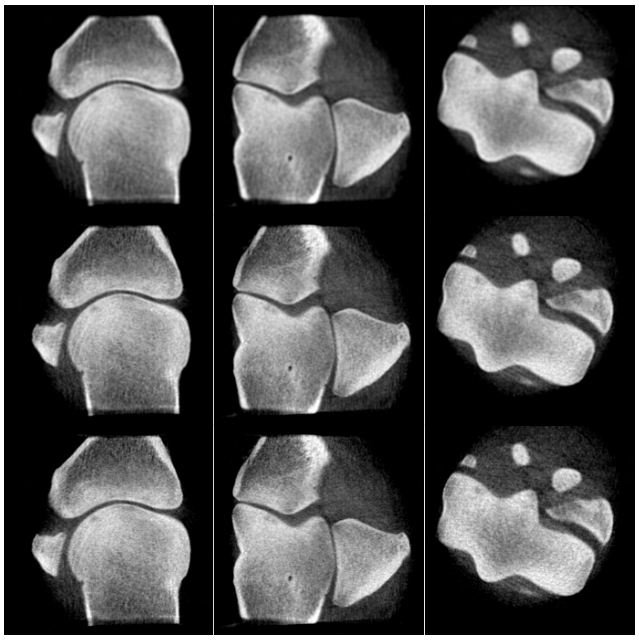


Fig. 5. Slice images of reconstructed volumes from the 64-view dataset using $w = 2 \times 0.2085$ mm (Top row) $w = 0.2085$ mm (Middle row), and $w = 0.0$ mm (Bottom row) at 50 iterations and $\gamma = 5.0 \times 10^6$. The grayscale window is [0.02,0.32] cm^{-1}.

Setting $w = 1$ and $\gamma = 5.0 \times 10^6$, slice images of the reconstructed volume are shown at 10, 20, and 50 iterations in Fig. 2. Through this progression of images, there is a clear trend of improving spatial resolution with iteration number up to the 50th iteration. Beyond this, the volume iterates change very little as demonstrated by the fact that there is little difference between the images at 50 and 100 iterations. For the remainder of the results shown in this abstract, we set the iteration number to 50.

In comparing the bottom row of images, at 50 iterations, to the FBP reference in Fig. 1 there is a noticeable loss in image quality, as might be expected since the view number is reduced by more than a factor of 10; however, features of clinical interest, such as the lucency in the bone tissue, are still clearly visible.

Varying the volume ℓ_1-norm constraint parameter γ

In the second parametric study, we investigate the impact of varying γ. Using the 64-view dataset, and setting the iteration number to 50 and $w = 0.2085$ mm, Fig. 3 shows reconstructed slice images for four values of γ, decreasing from top to bottom. The maximum value of $\gamma = 8.0 \times 10^6$ is chosen because at this value the ℓ_1-norm constraint is just barely active; with no ℓ_1-norm constraint the PD iterations achieves a volume with an ℓ_1-norm of 8.17×10^6. As by design, the images become more voxel sparse as γ decreases; at the largest value of γ, all of the soft-tissue is visible along with artifacts outside of the support of the sample. As γ decreases the non-zero voxel values outside of the object support begin to disappear. Driving to lower γ, the voxel values all decrease in magnitude and the soft tissue voxels also begin to disappear. At the lowest shown value of γ, only voxels containing bone tissue remain albeit at a reduced amplitude compared with the FBP reference of Fig. 1.

A less intuitive result of tightening the ℓ_1-norm constraint on the volume, i.e. reducing γ, is that the image resolution appears to improve as γ decreases. This trend runs counter to most other forms of regularization, where increasing the regularization strength tends to decrease spatial resolution. The marked increase in spatial resolution with reduced γ is a direct result of enforcing voxel sparsity on high-contrast porous bone tissue. Furthermore, ℓ_1 regularization does not penalize the difference between neighboring voxels, as many regularizers do. The apparent spatial resolution of the bone tissue improves mainly in going from $\gamma = 8.0 \times 10^6$ to $\gamma = 6.0 \times 10^6$. Decreasing γ further mainly impacts the bone tissue gray level. For the final set of results, this constraint is set to $\gamma = 5.0 \times 10^6$, a value that does not compromise spatial resolution and yet does not eliminate voxels at or near the bone tissue.

Varying w and scan configuration

For the final set of results, we consider three scan configurations using the 192 degree scanning arc with different angular sampling intervals as dictated by selecting 128, 64, and 32 evenly space projection view angles and shown in Figs. 4, 5, and 6, respectively. For each scan configuration image reconstruction is performed for w set to 0.417 mm (two voxel widths), 0.2085 mm (one voxel width) and zero mm for fixed iteration number and γ. We combine the w-dependence together with the display of images from different scan configurations, because the w-parameter is intended as

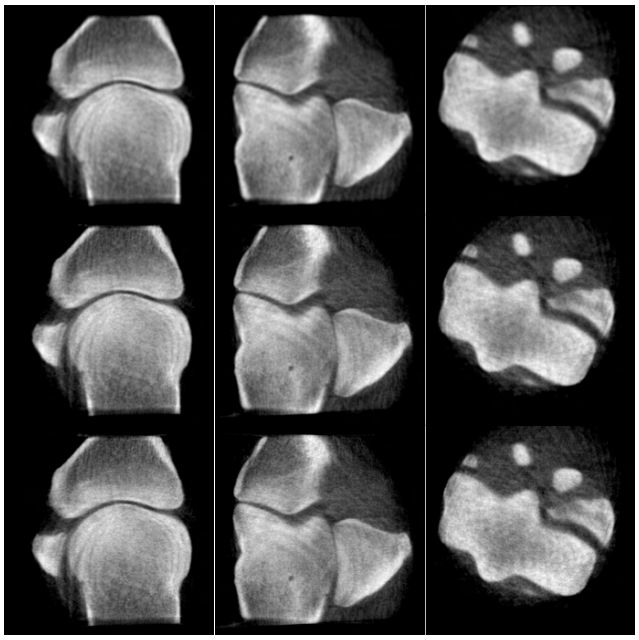


Fig. 6. Slice images of reconstructed volumes from the 32-view dataset using $w = 2 \times 0.2085$ mm (Top row) $w = 0.2085$ mm (Middle row), and $w = 0.0$ mm (Bottom row) at 50 iterations and $\gamma = 5.0 \times 10^6$. The grayscale window is [0.02,0.32] cm^{-1}.

the main control parameter on the image quality. For all configurations increasing w yields smoother images for the resulting reconstructed volume. Fine-tuning w will depend on the desired image quality metric that should be optimized. For the present discussion, we focus on $w = 1.0$, the middle row of these figures.

The 128-view scan results shown in Fig. 4 have decent image quality for the bone tissue with minimal streak artifacts and reasonably high spatial resolution. The lucent bone defect is still clearly visible in the $w = 1$ series of images. This scan also represents a six-fold reduction in the number of projections from the original CBCT dataset. Shown in Fig. 5 are the resulting images for a 64-view scan, a twelve-fold reduction in the number of views. The resolution is somewhat degraded in comparison with the 128-view results and mild streak artifacts overlapping the bone tissue become apparent, although the bone lucency is still visible. In the final set of results in Fig. 6, the view sampling is reduced by another factor of two, and the dataset contains only 32 views. For these images, the streak artifacts begin to seriously degrade the image quality and spatial resolution is markedly worse than the results for the 64-view configuration. Even so, some bone features are still visible and there may be clinical utility for this scan configuration.

IV. Conclusion

We have demonstrated an optimization-based framework for sparse-view CBCT image reconstruction designed for bone imaging using real CBCT data taken of an equine limb sample. The model involves representing the image as a blurred voxel-sparse image. The use of an ℓ_1-norm constraint on the reconstructed volume encourages voxel sparsity, which is useful for bone image because bone tissue takes up less volume than the surrounding soft tissue and it can have porous trabecular structure that is resolvable by the CBCT system. No explicit roughness regularizations is used in the framework other than the regularization that is afforded by the blurring kernel of the object model. Additional roughness regularization may not be necessary due to the high contrast of bone tissue in X-ray imaging. The results of applying the proposed bone imaging framework to sparse-view data show that greatly reduced view-angle sampling has the potential to yield clinically useful images and possibly enable new CBCT scan configurations for equine limb imaging. The development of new scan configurations may provide an opportunity to develop screening techniques useful to injury prevention in horses, potentially translating into other species.

Acknowledgment

This work was supported in part by the Grayson-Jockey Club Research Foundation and NIH Grant Nos. R01-EB026282 and R01-EB023968. MCD is also supported by Savannah River National Laboratory, which is operated by Battelle Savannah River Alliance for the U.S. Department of Energy under Contract No. 89303321CEM000080. The contents of this article are solely the responsibility of the authors and do not necessarily represent the official views of the National Institutes of Health.

References

[1] E. J. Candès, J. Romberg, and T. Tao, "Robust uncertainty principles: Exact signal reconstruction from highly incomplete frequency information," *IEEE Transactions on information theory*, vol. 52, no. 2, pp. 489–509, 2006.

[2] E. Y. Sidky and X. Pan, "Image reconstruction in circular cone-beam computed tomography by constrained, total-variation minimization," *Phys. Med. Biol.*, vol. 53, pp. 4777, 2008.

[3] M. Li, H. Yang, and H. Kudo, "An accurate iterative reconstruction algorithm for sparse objects: application to 3D blood vessel reconstruction from a limited number of projections," *Phys. Med. Biol.*, vol. 47, pp. 2599–2609, 2002.

[4] E. Y. Sidky, H. L. Stewart, C. E. Kawcak, C. W. McIlwraith, M. C. Duff, and X. Pan, "Bone sparsity model for computed tomography image reconstruction," in *15th International Meeting on Fully Three-Dimensional Image Reconstruction in Radiology and Nuclear Medicine*, 2019, vol. 11072, p. 110721U.

[5] T. Wang, H. Kudo, F. Yamazaki, and H. Liu, "A fast regularized iterative algorithm for fan-beam CT reconstruction," *Phys. Med. Biol.*, vol. 64, pp. 145006, 2019.

[6] E. Y. Sidky, D. N. Kraemer, E. G. Roth, C. Ullberg, I. S. Reiser, and X. Pan, "Analysis of iterative region-of-interest image reconstruction for x-ray computed tomography," *J. Med. Imag.*, vol. 1, pp. 031007, 2014.

[7] P. A. Wolf, J. S. Jørgensen, T. G. Schmidt, and E. Y. Sidky, "Few-view single photon emission computed tomography (SPECT) reconstruction based on a blurred piecewise constant object model," *Phys. Med. Biol.*, vol. 58, pp. 5629–5652, 2013.

[8] A. Chambolle and T. Pock, "A first-order primal-dual algorithm for convex problems with applications to imaging," *J. Math. Imag. Vis.*, vol. 40, pp. 120–145, 2011.

[9] E. Y. Sidky, J. H. Jørgensen, and X. Pan, "Convex optimization problem prototyping for image reconstruction in computed tomography with the Chambolle–Pock algorithm," *Phys. Med. Biol.*, vol. 57, pp. 3065–3091, 2012.

Estimation of Contrast Agent Concentration from Pulsed-Mode Projections to Time Contrast-Enhanced CT Scans

Isabelle M. Heukensfeldt Jansen, Eri Haneda, Bernhard Claus, Jed Pack,
Albert Hsiao, Elliot McVeigh, and Bruno De Man

***Abstract*— Cardiac CT exams are some of the most complex CT exams due to the need to carefully time the scan to capture the heart during a quiescent cardiac phase and when the intravenous contrast bolus is at its peak concentration in the left and/or right heart. We are interested in developing a robust and autonomous cardiac CT exam, using deep learning approaches to extract contrast and cardiac phase timing directly from projections. In this paper, we present a new approach to estimate contrast bolus timing directly from a sparse set of CT projections.**

We present a deep learning approach to estimate contrast agent concentration in left and right sides of the heart directly from a set of projections. We use a virtual imaging framework to generate training and test data, derived from real patient datasets. We finally combine this with a simple analytical approach to decide on the start of the cardiac CT exam.

***Index Terms*—bolus tracking, deep learning, enhancement prediction, scan timing**

I. Introduction

CARDIAC CT exams such as Coronary CT Angiography (CCTA) are some of the most complex CT exams due to the need to carefully time the scan to capture the heart during the quiescent cardiac phase (when the heart is relatively still) and when the contrast bolus in the heart chambers is at its peak concentration to achieve good contrast enhancement. The overall exam duration and the complexity of performing these exams (combined with limited reimbursement levels) have limited patient access to cardiac CT to academic hospitals and specialized cardiac imaging centers. Timing the CT scan to coincide with the peak contrast concentration can be done using a separate 'timing bolus' aquisition or with 'bolus tracking'. Both approaches have pros and cons and require highly trained operators to achieve consistent bolus enhancement.

Research reported in this publication was supported by the NIH/NHLBI grant R01HL153250. The content is solely the responsibility of the authors and does not necessarily represent the official views of the NIH.

Isabelle Heukensfeldt Jansen (isabelle.jansen@ge.com), Eri Haneda, Bernhard Claus, Jed Pack, and Bruno De Man are with GE Research, Niskayuna, NY. Albert Hsiao and Elliot McVeigh are with University of San Diego, La Jolla, CA.

Our overall project goal is to develop a smart cardiac CT scanner that autonomously determines the optimal scan time interval without ECG, traditional bolus tracking or timing bolus, using real-time deep learning analytics of sparsely pulsed projections [1,2]. Recent advances in X-ray tube technology have provided the ability to acquire pulsed-mode projections (PMPs) or only a few projections per gantry rotation. While not sufficient to perform image reconstruction, these PMPs can be utilized along with deep learning to predict the contrast agent concentration in specific compartments of the heart. Here we present results of a deep learning approach to estimate contrast agent concentration in left and right sides of the heart and to determine the start of the CCTA scan.

II. Background

A. Standard-of-care CCTA Protocol

In CCTA exams, an intravenous (IV) power injector and an ECG monitor are connected to the patient. A scout scan is performed for patient positioning, followed by a low-dose CT scan to determine a region-of-interest for tracking the contrast bolus. In a first standard-of-care protocol, a timing bolus (or test bolus) is administered as a 'trial run' [3], involving a separate injection of a small amount of contrast agent (10–20 ml) followed by a saline flush. After injection, a predetermined number of low-dose bolus timing scans are performed with narrow collimation and with a relatively long inter-scan delay (to minimize radiation dose). The enhancement is measured in a region of interest (ROI) to determine time-to-peak enhancement, which is used to compute the delay time between the administration of the main bolus and the start of diagnostic CCTA scan with wide collimation.

A second standard-of-care protocol uses real-time bolus tracking: the timing bolus is omitted, and the time-of-peak-enhancement is predicted in real-time. When a pre-defined threshold is reached, a diagnostic delay (e.g., 7-sec) is added to give a breath-hold command, open up the collimation, and re-position the patient, before starting the diagnostic CCTA scan [3]. The time-to-peak contrast enhancement depends on many factors, including target vessel, patient anatomy and cardiac output. The enhancement is measured in an ROI, and because of the necessary diagnostic delay, the threshold

(usually 100–150 HU) is far below the peak enhancement (250–600 HU) [1-3]. As a result, a longer bolus (and hence larger contrast volume) is used for robust scanning, i.e.: to avoid missing the peak.

B. *Proposed CCTA protocol*

Our proposed CCTA protocol uses IV contrast administration and a scout scan for patient positioning but eliminates the ECG as well as the additional CT scans to track the bolus concentration in ROIs. The patient and collimator are immediately positioned in the right location for the actual CCTA scan. After injection of the main bolus, a X-ray tube is pulsed ON and OFF, using a very low duty cycle for low radiation dose. For example, 1, 2, 4, 8, or 16 300-µs pulses may be performed per 0.28 sec rotation, resulting as many CT projections as there were pulses. Deep learning networks analyze these projections in real-time and determine a running estimate of the current contrast agent concentration as well the current cardiac phase (% R-to-R-peak). A separate timing algorithm uses these estimates to decide on the time for the breath-hold command and the start of the actual CCTA scan. This approach promises to eliminate or greatly reduce the diagnostic delay since the patient and collimator are already in the right position, so the peak enhancement may robustly be identified even for a smaller bolus and a narrower plateau. The PMPs minimize the additional radiation dose since no full-rotation scans are performed. The approach also leads to robust automation relying on deep learning algorithms to time the CCTA exam.

III. Materials And Methods

We briefly summarize the virtual imaging framework used to generate training data (section III.A), we then describe the convolutional neural network (CNN) to predict the contrast level present in a PMP (section III.B), and finally we derive a simple scan timing algorithm (section III.C).

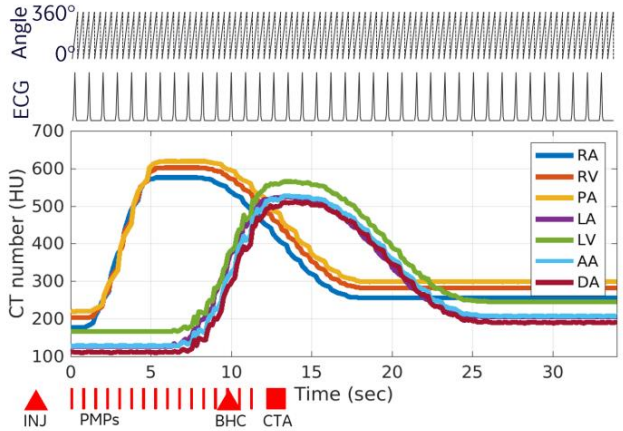


Fig. 1: Example of a CT exam instance as a function of time. From top to bottom: gantry rotation angle, ECG signal, average CT number in the 7 cardiac compartments, and the timing of injection (INJ), pulsed-mode projections, (PMPs), breath-hold command (BHC), and CCTA scan. The gantry rotation angle, ECG signal, and PMPs are shown for illustration purpose. The spacing is not exact.

A. *Training data generation*

We created a virtual imaging framework for creating cardiac CT projections at any combinations of view angles, cardiac phase, and bolus contrast timing as presented in detail in [4]. In summary, we did this by developing five-dimensional cardiac CT models from multi-phase clinical cardiac (cine) CT scans by segmenting the heart compartments and identifying a blood flow propagation map in each compartment. To model contrast dynamics at multiple bolus time points from datasets that were acquired at (approximately) a single bolus time point, we segmented the cardiac compartments, we parametrized the voxels inside those compartments based on their location along the flow direction, and then incremented the voxel values to model different bolus distributions based on location and time point. We then defined multiple instantiations of CT exams based on specific timing of cardiac cycle, contrast bolus, and CT scan and generated virtual CT projection data. The model contained segmentations for the right atrium (RA), right ventricle (RV), pulmonary artery (PA), left atrium (LA), left ventricle (LV), ascending aorta (AA), and descending aorta (DA). Figure 1 shows a specific CT exam instance as a function of time. The top row shows the CT gantry rotation angle. The second row shows the patient ECG signal. The 7 colored curves show CT number averaged over each of the 7 cardiac compartments. For each curve, we can clearly observe a rising edge, a plateau, and a more gradual decay. The large delay between compartments in the right left sides of the heart is due to the pulmonary circulation. We did not simulate dispersion for the left side of heart in this experiment. The bottom row shows the time of the injection, the start of the pulsed-mode projections (PMPs), the breath-hold command, and the actual CCTA scan.

In this work, a number of sample instantiations were defined

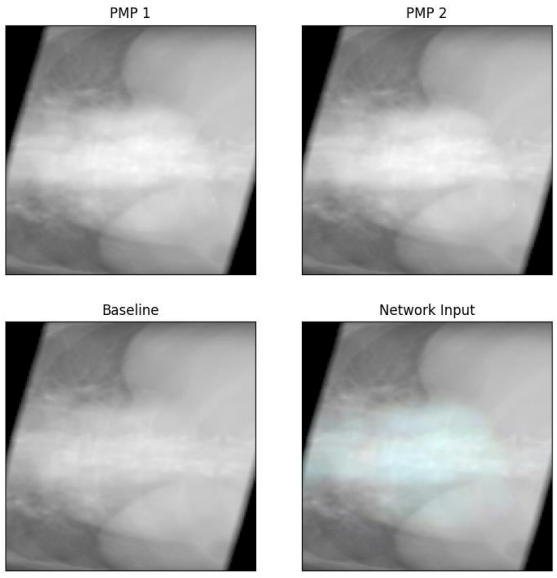


Fig. 2: A sample input image using 2 PMPs. In this configuration, the presence of a reference baseline created with pre-bolus images means the stacked layers show the presence of the bolus as color on a greyscale image.

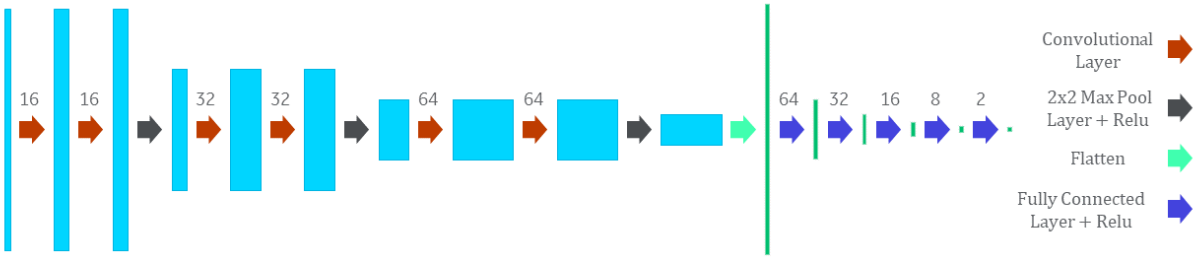


Fig. 3: Structure of the neural network. Where relevant, the number of features or nodes are shown.

by selecting and combining different patients, heart rates and contrast dynamics. For each sample, a series of virtual CT projection data was generated along two view angles, anterior-posterior (AP) and posterior-anterior (PA). For each series, a baseline reference image was defined as the average of all the PMPs from the matching view angle that had been estimated as not yet having any enhanced contrast from the bolus. Network input images were then created from grouping 5 PMPs (representing the 5 "latest" PMPs in a real-time scenario) together in a sliding window and stacking the PMPs with the baseline image matching the view angle of the last PMP in the series for a 128x128x6 input image. A sample input comprised of a reference image and 2 PMPs is shown in Figure 2.

Training, validation, and testing datasets were created to be independent of each other by separating the simulated dataset such that all the images from an individual cardiac model were assigned to only one of the sets. The training, validation, and testing sets comprised of 32, 4, and 4 models respectively, for a total of 300 simulated series of 250 PMPs each, or 61,254 possible samples in the training set when each series was broken down into sets of 5 PMPs to generate input images. Each sample was labeled with two volume-weighted averages, one of the contrast in the RA-RV chambers (right half) and one of the contrast in the LA-LV chambers (left half).

B. Neural Network Architecture and Training

The network comprised of a CNN followed by fully-connected layers to perform regression on the input images. In the CNN, pairs of convolutional layers were followed by batch normilization, 2x2 max pooling, and a relu activation function. This was repeated three times, using first a pair of convolutional layers with 16 features each, then 32 and 64 features each for the second and third repetition. The output of the CNN was then flattened and fed through fully connected layers followed by a relu activation function. Fig. 3 shows the network architecture. The network output was a pair of values representing the contrast level in the right and left halves.

The network was trained using a MSE loss function and ran for 500 epochs using an Adam optimizer. Each epoch consisted of 1,000 random samples from the training set images.

C. CCTA Scan Timing

Each series in the validation and testing set was broken down into a sequence of images representing a sliding window over the entire series. Each image was evaluated by the network to create a time series of the contrast levels for both the right and left chambers. Each time series was fed into an algorithm that looped over the series. For each PMP, all the predictions up to that PMP were evaluated to determine if peak contrast levels had been reached without any further information, to simulate a real-time scenario where peak contrast levels may not be known.

A simple timing algorithm was created to determine a trigger point based on thresholds defined on the predicted enhancement.

IV. Results

Network training converged to a R-squared value of 0.99 for the training set and 0.91 for the validation set. Fig. 4 shows the results of network training when the labels are taken as individual datapoints, for the training dataset (left) and validation dataset (right). The broader distribution of the validation dataset compared to that of the training dataset is a classic symptom of overfitting the data. Fig. 5 shows representative results of several series in the validation set when the predictions are plotted in a time series for each set. In Fig 5a, the predictions match the ground-truth labels closely. Figs. 5b-c show the two most common modes of error. In Fig. 5b, the bolus curve in the left heart is scaled to a lower value than the ground truth labels. In Fig 5c, the right heart bolus curve is confounded with the bolus curve in the left heart and has an additional bump in the distribution when the left-heart contrast is high. While the scaling error in Fig 5b was observed for both right heart and left heart estimates, the error shown in Fig 5c was exclusively seen as the right-heart

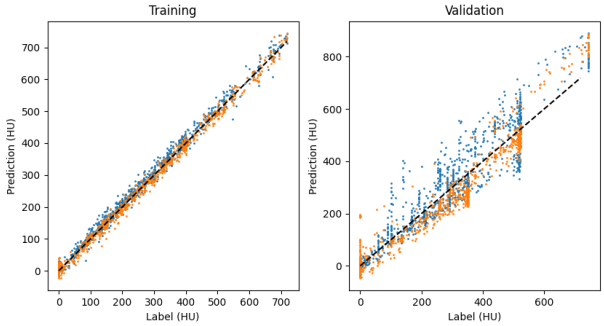


Fig. 4: Prediction vs Labeled contrast levels in the training and validation sets. For each plot, orange and blue represent the right and left chambers respectively.

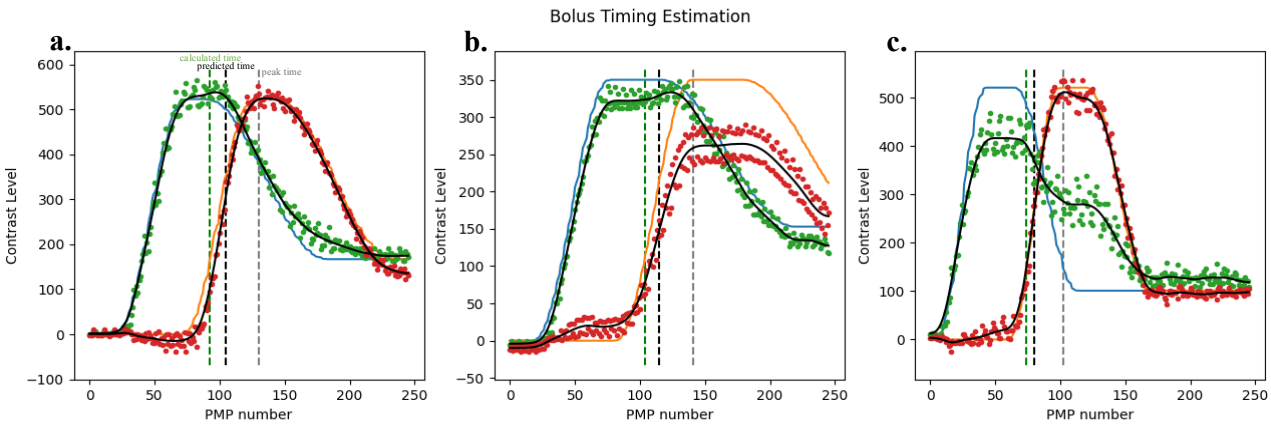


Fig. 5: Sample validation outputs of the network. Blue represents the right chamber enhancement with corresponding green predictions, while orange represents the left chamber enhancement with corresponding red predictions. Predictions are smoothed using a gaussian filter (resulting in black curves). The green and black dashed lines represent the desired and the trigger points based on a fixed threshold, while the grey dashed line shows the start of the peak enhancement in the left side of the heart. a.) The prediction closely matches the ground-truth; b.) the prediction in one chamber is off by a scaling factor; c.) the prediction in the right chambers is confounded with the left chambers, resulting in an addition bump after the bolus in the right chambers has reached the converged bolus level.

predictions gaining a secondary bump when the left heart contrast is bright. In both cases, the overall shape of the bolus rise time is preserved and is hence suitable for timing optimization algorithms.

The green and black dashed lines represent the desired and the trigger points based on a fixed threshold, while the grey dashed line shows the start of the peak enhancement in the left side of the heart. The RMS error was 11 PMPs in the testing set when compared to the same threshold applied to the ground-truth data, which corresponds with an RMS error of 1.6 seconds. More work is needed to optimize the trigger points and derive the time for the breath-hold command and for the start of the CCTA scan.

V. Conclusion

A neural network provides a powerful tool for analyzing the sparse projection data acquired with PMPs. For the purpose of scan timing, the rise and decay times are more critical than the precise levels of the contrast enhancement, which will guide us in designing improved networks.

This study used a simple threshold-based timing algorithm, in line with similar algorithms used in more traditional bolus tracking. A more sophisticated timing algorithm might derive trigger points for the breath hold command approximately 3-4 seconds in advance, then continuing to monitor the bolus curve to begin the scan once the actual peak is reached. By monitoring PMPs instead of an ROI, the equipment delays created by patient repositioning and opening the collimation is negated. This work shows promising results for the use of PMPs in optimizing timing for CCTA scans.

REFERENCES

[1] R. L. Hallett en D. Fleischmann, “Tools of the trade for CTA: MDCT scanners and contrast medium injection protocols”, Techniques in vascular and interventional radiology, vol 9, no 4, bll 134–142, 2006.

[2] S. Oda et al., “Low contrast and radiation dose coronary CT angiography using a 320-row system and a refined contrast injection and timing method”, Journal of cardiovascular computed tomography, vol 9, no 1, bll 19–27, 2015.

[3] J.-E. Scholtz en B. Ghoshhajra, “Advances in cardiac CT contrast injection and acquisition protocols”, Cardiovascular diagnosis and therapy, vol 7, no 5, bl 439, 2017.

[4] E. Haneda et al., “A five-dimensional cardiac CT model for generating virtual CT projections for user-defined bolus dynamics and ECG profiles”, in *The 7th International Conference on Image Formation in X-Ray Computed Tomography, Baltimore, MD, USA, June 12-16, 2022.* (submitted).

Time Separation Technique Using Prior Knowledge for Dynamic Liver Perfusion Imaging

Hana Haseljć, Vojtěch Kulvait, Robert Frysch, Fatima Saad, Bennet Hensen, Frank Wacker, Inga Brüsch, Thomas Werncke, and Georg Rose

***Abstract*—The perfusion imaging using C-arm CT could be used intraoperatively for liver cancer treatment planning and evaluation. To deal with undersampled data due to slow C-arm CT rotation and pause between the rotations, we applied model-based reconstruction methods. Recent works using the time separation technique with an analytical basis function set have led to a significant improvement in the quality of C-arm CT perfusion maps. In this work we apply the time separation technique with a prior knowledge basis function set extracted using singular value decomposition from CT perfusion reconstructions. On C-arm CT liver perfusion scan simulated based on the real CT liver perfusion scan we show that the bases extracted from only two CT perfusion scans are capable of modeling the C-arm CT data correctly.**

***Index Terms*—dynamic perfusion imaging, C-arm CT, prior knowledge, singular value decomposition**

I. Introduction

PERFUSION CT imaging is an important step in liver cancer therapy planning, see [1]. The C-arm CT perfusion imaging could interventionally assist in the cancer treatments by offering the possibility to evaluate the success of performed embolization or in ablation planning. The so far investigated protocols for estimation of parenchymal blood flow, see [2]–[4], would not be sufficient to capture the dynamic perfusion, so ten sweep perfusion protocol was suggested in [5]. However, the problem of undersampled data due to limited number of projections per rotation and pause times between in between remain. In [6] it was shown that the model-based reconstruction by applying Time separation technique (TST) with analytical basis function set could solve these problems. The possibility to use dedicated, so called prior knowledge, basis function set formed from singular vectors extracted by applying singular value decomposition (SVD) to CT perfusion reconstruction data was studied in [7], [8] on brain perfusion.

In this paper we extract the prior knowledge basis function set from liver CT perfusion reconstruction data. We use this basis function set for TST to reconstruct the simulated C-arm CT perfusion scan of animal liver. We compare perfusion maps of TST using prior knowledge basis constructed from two and three CT scans, perfusion maps of TST using analytical bases and perfusion maps of straightforward reconstruction to CT perfusion maps.

This work was partly funded by the European Structural and Investment Funds (International Graduate School MEMoRIAL, project no. ZS/2016/08/80646) and the German Federal Ministry of Education and Research within the Research Campus STIMULATE, grant no. 13GW0473A and 13GW0473B).

Hana Haseljić is with Institute for Medical Engineering and Research Campus STIMULATE, Otto-von-Guericke University, Magdeburg, Germany (e-mail: hana.haseljic@ovgu.de)

Vojtěch Kulvait is with Institute of Materials Physics, Helmholtz-Zentrum hereon, Geesthacht, Germany.

Robert Frysch, Fatima Saad and Georg Rose are with Institute for Medical Engineering and Research Campus STIMULATE, Otto-von-Guericke University, Magdeburg, Germany.

Bennet Hensen, Frank Wacker and Thomas Werncke are with Institute of Diagnostic and Interventional Radiology, Hannover Medical School, Hannover, Germany.

Inga Brüsch is with Institute for Laboratory Animal Science, Hannover Medical School, Hannover, Germany

II. Materials and Methods

A. Animal Experiments

The CT perfusion scans of three domestic pigs were acquired using SOMATOM Force CT after embolization. The iodinated contrast material used was Imeron 300. The right hepatic artery was embolized with tantalum-based embolization material (Onyx) and coils. The duration and contrast material injection details are given in Table I.

TABLE I
Perfusion Scans Duration and Details of Contrast Material Injection

Scan	Duration [s]	Dose [ml]	Flow rate [ml/s]	Flow duration [s]	Volume [ml]
1	65.997	14.0	3.0	7.0	20.0
2	56.997	10.43	2.8	5.07	14.9
3	41.998	10.43	2.9	5.1	14.9

B. Simulation of C-arm CT perfusion scan

We simulate the C-arm CT data by reprojecting the CT volume as in [9]. We reproject the Scan 3, see Table I with C-arm CT projector according the acquisition protocol of the experimental C-arm CT perfusion scans of the liver used in [5], [6]. The total scan time is divided in ten runs covering the ten rotations of 200° with 248 views with pause time $412.44\,\mathrm{s}$ between every two consecutive runs.

C. Prior Knowledge Extraction

Two prior knowledge basis function sets are extracted from CT perfusion scans. First is extracted from first two CT scans from Table I and the second one from all three by applying SVD on time attenuation curves (TAC) of voxels inside the organ regions as in [7]. All bones, catheters if visible

and surrounding organs were excluded. The CT scans were reconstructed using syngo CT VA50A software. The time-resolved volumes of each scan were interpolated across the time interval of shortest scan, see [9].

For the brain reconstruction it is recommended to use either three or five basis functions since more could cause instabilities in projections, see [10], [11]. In [6] five trigonometric functions formed the analytical basis function set. Based on the singular values, how well they fit the AIF and the Pearson correlation of perfusion maps with CT ground truth perfusion maps, we decide how many bases will form our basis function set, see Figures 1 and 2.

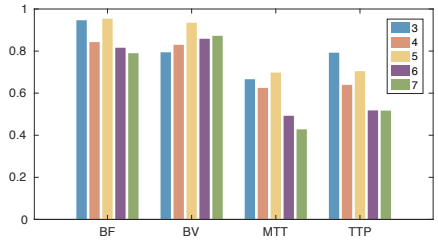


Fig. 1. Pearson correlation coefficient with respect to CT ground truth perfusion maps reconstructed using basis function sets of different sizes.

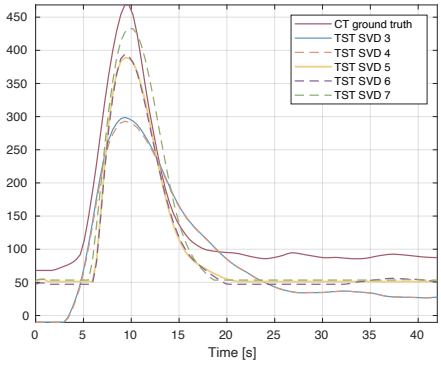


Fig. 2. Arterial input function fitted using basis function sets of different size.

By the means of the SVD our bases are orthonormal and therefore a suitable basis function set for TST. Our basis functions sets are shown in Fig. 3 and 4. The *Function 0* is a constant function and therefore left out from the graphs.

D. Time Separation Technique

The time separation technique, see [9], [10], is a model-based reconstruction where the time attenuation development of every voxel is modeled as a linear combination of mutually orthonormal bases (1).

$$\mathcal{B} = \{\Psi_1, \ldots, \Psi_N\}. \tag{1}$$

The same way the pixels in projections can be modeled so the reconstruction problem $Ax = p$ becomes

$$A\sum_{i=1}^{N} w_{v,i}\psi_i(t) = \sum_{j=1}^{N} w_{p,j}\psi_j(t) \quad t \in \mathcal{I}. \tag{2}$$

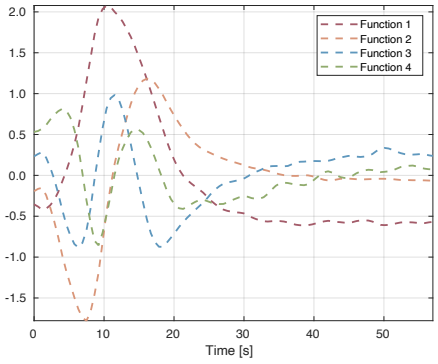


Fig. 3. Extracted prior knowledge basis function set from two animals on the scan duration interval.

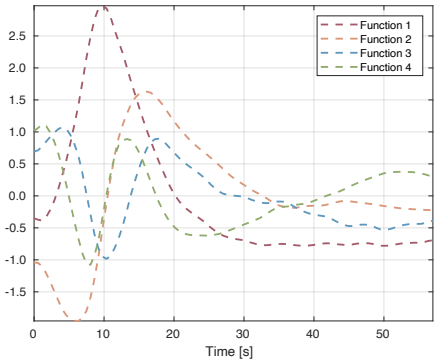


Fig. 4. Extracted prior knowledge basis function set from three animals on the scan duration interval.

where $\mathcal{I}$ is a vector of time points at which the volume was scanned. With scalar product of both sides with bases from $\mathcal{B}$ due to orthonormality of the bases, the reconstruction problem is reduced to N static reconstruction problems, see [9], [10], unlike the straightforward reconstruction where each rotation is reconstructed separately.

The reconstruction is done using algebraic reconstruction developed and implemented within [8], [12]. The voxel size is the same as in the CT scan $(x_v, y_v, z_v) = (0.7305, 0.7305, 1.5)$.

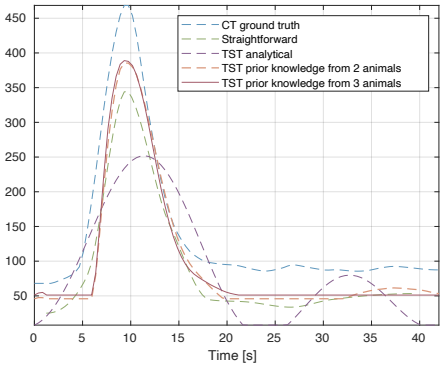


Fig. 5. Arterial input function of CT ground truth, straightforward simulated C-arm CT and using analytical and prior knowledge basis sets.

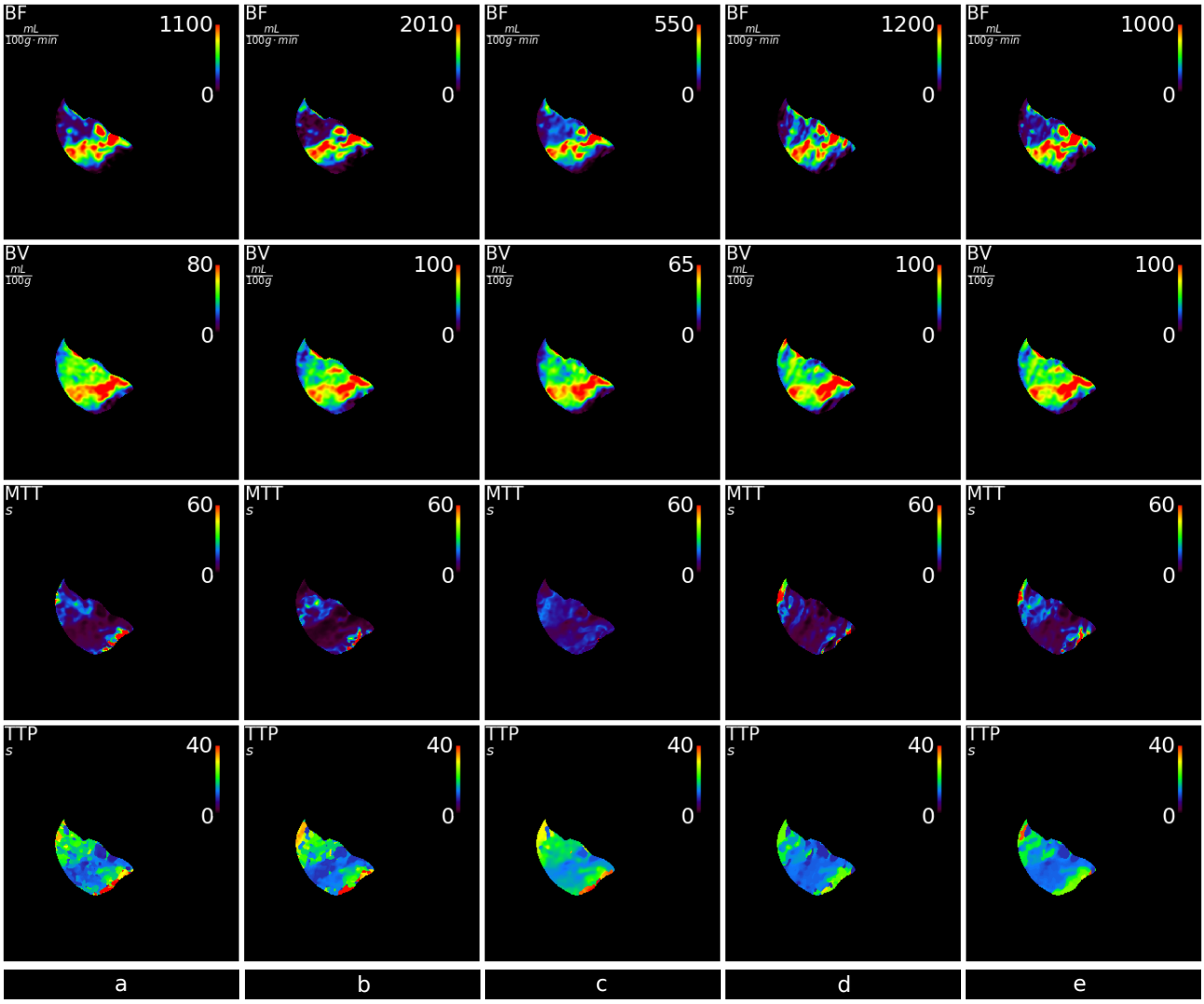


Fig. 6. Perfusion maps for scan 3. a: ground truth CT, b: simulated C-arm CT scan - straightforward reconstruction, c: simulated C-arm CT scan - TST using five trigonometric basis, d: simulated C-arm CT scan - TST using prior knowledge basis extracted from two animals, e: simulated C-arm CT scan - TST using prior knowledge basis extracted from three animals.

E. Perfusion Parameters Estimation

We calculate the perfusion parameters, blood flow (BF), blood volume (BV), mean transit time (MTT) and time to peak (TTP), using deconvolution technique, see [13]. The time attenuation profile of every voxel is described as a convolution of AIF with the residual function. To determine the residual function the pseudoinverse with Tikhonov regularization is applied to Eq. (3).

$$\mathrm{tac(t)} = \mathrm{aif(t)} * f_r(\mathrm{t}). \tag{3}$$

For the calculation of the perfusion parameters, see Eq. (4), except for the TTP, it is important to select arterial input function (AIF) properly. This is not the case for the TTP which is estimated as the time from the beginning of the acquisition of the maximum value of the voxel's attenuation. The selected AIF for location $(x, y, z) = (194, 257, 65)$ is shown in Fig. 5.

$$\begin{aligned} BF &= \max f_r(t), \quad BV = \sum_{i=1}^{n} f_r(\mathrm{i}), \\ MTT &= \frac{BV}{BF}, \quad TTP = \arg\max_t f_r(t). \end{aligned} \tag{4}$$

To avoid the possible instabilities of the BF we estimate the BF only based on the first $5\,\mathrm{s}$ of the acquisition as advised in [9]. The perfusion maps are calculated using [14].

III. Results and Discussion

We have simulated the C-arm CT perfusion scan and reconstructed it using model-based approach TST with prior knowledge basis function set as explained in previous section.

First, we compare two prior knowledge basis function sets. In the basis function set generated from two animals we can observe the noise especially after the $15\,\mathrm{s}$. In the basis

function set extracted on three animals these instabilities are less pronounced, see Fig. 4. However, both are fitting the AIF better than in straightforward approach and when fitted with trigonometric bases, see Fig. 5. It can be also observed that the *Function 1* from both basis function sets corresponds the AIF profile as the strongest pronounced signal. Note that all the values in AIF are lower due to introduced Gauss blur of $3.5\,\mathrm{px}$.

TABLE II
PEARSON CORRELATION COEFFICIENT AND NRMSE IN RESPECT TO CT GROUND TRUTH PERFUSION MAPS

Scan	BF	BV	MTT	TTP
Straightforward	0.9578	0.9528	0.7009	0.7741
	0.1528	0.1101	0.2003	0.1725
TST	0.9336	0.8837	0.4526	0.7921
analytic basis	0.1551	0.1404	0.3257	0.1886
TST	0.8898	0.8481	0.4779	0.5425
SVD 2 animals	0.1786	0.1698	0.2728	0.2131
TST	0.9537	0.9344	0.6976	0.7045
SVD 3 animals	0.1392	0.1242	0.2213	0.1823

The perfusion maps are shown in Fig. 6. The Pearson correlation coefficients and Normalized Root Mean Square Error (NRMSE) between the CT perfusion maps and simulated C-arm CT perfusion scans reconstructed by different means are given in Table II.

In all perfusion maps it can be observed that the qualitative information is preserved. In perfusion images we use the *asist* color map, see [15]. The lowest value in perfusion maps is dark blue. Hence, the dark blue area in BF and BV represents the embolized area. We can observe BF relative overestimation in TST with trigonometric basis and TST with prior knowledge extracted from two animals compared to ground truth. A relative underestimation is observed in straightforward approach.

From the correlation coefficients, see Table II, the perfusion maps generated using prior knowledge extracted from three animals are more similar to ground truth than the perfusion maps using prior knowledge extracted from two animals. When compared to perfusion maps using analytical bases, the prior knowledge using three animals is better in all considering the normalized RMSE. The TST using prior knowledge generated with three animals is the second best. The reason could be that the simulated CT scan is generated using CT and therefore better in terms of noise and sampling frequency compared to measurements of usual CBCT systems. Due to limited number of available CT perfusion scans of animal liver the same scan was used to simulate the C-arm CT perfusion data and to extract the prior knowledge. This, as well as the more scans involved in prior knowledge extraction, could have affected the better fitting of an AIF and better perfusion maps.

IV. CONCLUSION

From the results we see that the model-based reconstruction by applying TST with prior knowledge bases is comparable to CT perfusion maps and it is better than the straightforward reconstruction for BF. We have also shown that it is possible to extract the bases using only two CT scans and to get perfusion maps comparable to CT ground truth. However, more scans should be used to extract the prior knowledge. In the future we will focus on studying if these results can be confirmed using real C-Arm CT datasets.

REFERENCES

[1] H. Ogul, M. Kantarci, B. Genc, B. Pirimoglu, N. Cullu, Y. Kizrak, O. Yilmaz, and N. Karabulut, "Perfusion CT imaging of the liver: review of clinical applications," *Diagnostic and Interventional Radiology*, vol. 20, no. 5, pp. 379–389, aug 2014.

[2] R. Syha, S. Gatidis, G. Grözinger, U. Grosse, M. Maurer, L. Zender, M. Horger, K. Nikolaou, and D. Ketelsen, "C-arm computed tomography and volume perfusion computed tomography (VPCT)-based assessment of blood volume changes in hepatocellular carcinoma in prediction of midterm tumor response to transarterial chemoembolization: a single center retrospective trial," *Cancer Imaging*, vol. 16, no. 1, Sep. 2016. [Online]. Available: https://doi.org/10.1186/s40644-016-0088-y

[3] K. Mueller, R. Fahrig, M. Manhart, Y. Deuerling-Zheng, J. Rosenberg, T. Moore, A. Ganguly, and N. Kothary, "Reproducibility of parenchymal blood volume measurements using an angiographic c-arm CT system," *Academic Radiology*, vol. 23, no. 11, pp. 1441–1445, Nov. 2016. [Online]. Available: https://doi.org/10.1016/j.acra.2016.08.001

[4] R. L. O'Donohoe, R. G. Kavanagh, A. M. Cahalane, D. D. Houlihan, J. W. McCann, and E. R. Ryan, "C-arm cone-beam CT parenchymal blood volume imaging for transarterial chemoembolization of hepatocellular carcinoma: implications for treatment planning and response," *European Radiology Experimental*, vol. 3, no. 1, May 2019. [Online]. Available: https://doi.org/10.1186/s41747-019-0099-0

[5] S. Datta, K. Müller, T. Moore, L. Molvin, S. Gehrisch, J. Rosenberg, Y. Saenz, M. Manhart, Y. Deuerling-Zheng, N. Kothary, and R. Fahrig, "Dynamic measurement of arterial liver perfusion with an interventional c-arm system," *Investigative Radiology*, vol. 52, no. 8, pp. 456–461, Aug. 2017. [Online]. Available: https://doi.org/10.1097/rli.0000000000000368

[6] H. Haseljić, V. Kulvait, R. Frysch, B. Hensen, F. Wacker, G. Rose, and T. Werncke, "The Application of Time Separation Technique to Enhance C-arm CT Dynamic Liver Perfusion Imaging," in *16th International Meeting on Fully Three-Dimensional Image Reconstruction in Radiology and Nuclear Medicine*, 2021.

[7] C. Eckel, S. Bannasch, R. Frysch, and G. Rose, "A compact and accurate set of basis functions for model-based reconstructions," *Current Directions in Biomedical Engineering*, vol. 4, no. 1, pp. 323–326, Sep. 2018. [Online]. Available: https://doi.org/10.1515/cdbme-2018-0078

[8] V. Kulvait and G. Rose, "Software Implementation of the Krylov Methods Based Reconstruction for the 3D Cone Beam CT Operator," in *16th International Meeting on Fully Three-Dimensional Image Reconstruction in Radiology and Nuclear Medicine*, 2021.

[9] V. Kulvait, P. Hoelter, R. Frysch, H. Haseljić, A. Doerfler, and G. Rose, "Time separation technique with the basis of trigonometric functions as an efficient method for flat detector CT brain perfusion imaging," *Submitted for consideration to Medical physics*, 2021, preprint at https://arxiv.org/abs/2110.09438.

[10] S. Bannasch, R. Frysch, T. Pfeiffer, G. Warnecke, and G. Rose, "Time separation technique: Accurate solution for 4D C-arm-CT perfusion imaging using a temporal decomposition model," *Medical Physics*, vol. 45, no. 3, pp. 1080–1092, feb 2018.

[11] V. Kulvait, P. Hoelter, and G. Doerfler, Arnd Rose, "Noise and dose reduction in CT brain perfusion acquisition by projecting time attenuation curves onto lower dimensional spaces," in *SPIE Medical Imaging 2022*, 2021, accepted contribution.

[12] V. Kulvait, "Kct cbct a software package for algebraic ct reconstruction," https://github.com/kulvait/KCT_cbct, 2021.

[13] A. Fieselmann, M. Kowarschik, A. Ganguly, J. Hornegger, and R. Fahrig, "Deconvolution-based CT and MR brain perfusion measurement: Theoretical model revisited and practical implementation details," *International Journal of Biomedical Imaging*, vol. 2011, pp. 1–20, 2011.

[14] V. Kulvait, "Kpct perfviz a software package for creating perfusion maps," https://github.com/kulvait/KPCT_perfviz, 2021.

[15] K. Kudo, "Acute Stroke Imaging Standardization Group Japan recommended standard LUT (a-LUT) for perfusion color maps." http://asist.umin.jp/data-e.shtml, 2005, accessed April 9, 2021.

A hybrid neural network combining explicit priors for Low-dose CT reconstruction

Xiangli Jin, Yinghui Zhang, Ran An, Hongwei Li*

***Abstract*—Low-dose CT reconstructions suffer from severe noise and artifacts. Many methods have been proposed to increase the ratio between image quality and radiation dose through incorporating various priors. By learning priors in labelled data-set, neural network methods have achieved great success for this purpose . In CT applications, however, paired training data-sets are rarely available or difficult to obtain. Recently, unsupervised learning has attracted a lot of attention. Along this line, Noise2Inverse, an unsupervised neural network architecture, has shown the possibility of applying unsupervised learning for low-dose CT reconstructions. When the training data-sets (unlabelled) are not large enough or the training is insufficient, however, Noise2Inverse might perform not well. Another important issue is that network methods might suffer from intrinsic instability. In this regard, we propose to hybrid neural networks, especially the Noise2Inverse architecture, with traditional optimization models such that hand-crafted priors come into play as a remedy. Numerical experiments show that the proposed architecture improves Noise2Inverse in terms of both quality measures PSNR and SSIM, especially in the case of inadequate training.**

***Index Terms*—Low-dose CT, deep learning, image denoising, TV regularization, primal-dual**

I. Introduction

TO avoid exposing patients to more X-ray radiation, it makes sense to reduce the X-ray dose. However, compared with conventional CT, low-dose CT usually introduce serious noise and artifacts. In recent years, low-dose CT (LDCT) reconstruction has been a major challenge in medical CT applications. In general, LDCT reconstruction methods can be roughly divided into three categories. The first category is the filtering method, which directly performs filtering and smoothing on the projection data or the noisy reconstructions. Popular methods include NLM [3] and BM3D [4], which explore the structural similarity prior within a single image. The defect of these methods is that it can not distinguish well image structures and artifacts. The second category is the model-based optimization approach. By combining image priors into the objective function, noise and artifacts are removed in the reconstruction process. However, hand-crafted priors are often not accurate enough and might introduce negative effects like blurring . In addition, the optimization model usually need iterative solvers which are very time-consuming. The third category is deep learning methods like U-Net [9] and DnCNN [10]. Based on the powerful fitting ability of neural networks, priors or patterns concealed in "big data" could be extracted and utilized. Neural networks have demonstrated to be superior to traditional methods if adequate labelled training data are available. However, for CT applications, high quality training data are usually difficult to acquire.

Unsupervised learning methods for image denoising gain popularity in recent years. The X2Y series algorithms have attracted much attention. The Noise2noise [8] method explains the MSE loss in Bayesian framework by which noisy images are allowed to be used as reference images, when certain independence conditions are satisfied. Later on, by introducing the idea of blind spot, Noise2Self [2] and Noise2Void [7] were proposed which ruled out the need for independent multiple snapshots of the same scene. It has been demonstrated that these unsupervised methods could achieve competitive denoising results compared to the supervised ones.

Based on the idea of Noise2Void, Noise2Inverse was proposed in [5] for CT image reconstruction. Basically, it consists of two procedures. The first one is to prepare the training data-sets. The projection data are divided into non-overlapping groups which are then used to reconstruct noisy images of the same "scene"(object). Since the projections are independent of each other, the noise in these images should follow the same distribution and be independent of each other. The second procedure is to train a denoising convolutional neural network (CNN) with the constructed training data in the first procedure. Noise2Inverse has shown promising results in LDCT denoising.

In practice, the training data sets might be quite limited available, such that under-fitting or over-fitting occurs. Another important issue is the stability of neural networks. It has been demonstrated that neural networks could suffer from intrinsic instability issue [1]. In our experiments, when the training data sets were relatively small or insufficient training were performed, i.e. the number of training epochs was set to be relatively small, the effectiveness of Noise2Inverse would be compromised.

To further improve the quality of LDCT reconstructions, we propose to combine the convolutional neural networks (CNN) with traditional optimization models. So, the proposed hybrid neural network (Hybrid NN) architecture consists of two blocks: the denoising CNN block and a training-parameter free optimization-based denoising block. Especially, in this paper, the second block is mapped from the primal-dual [6] algorithm for total-variation (TV) denoising, which encodes the piecewise-constant prior of the ideal image. Since the hybrid neural network builds in hand-crafted priors, its stability

All authors are with the School of Mathematical Sciences, Capital Normal University, Beijing, 100048, China and are also with Beijing Advanced Innovation Center for Imaging Technology, Capital Normal University, Beijing, 100048, China.

E-mail: hongwei.li91@cnu.edu.cn

should have been improved.

The main contributions of this paper are two-fold.

- A hybrid neural network that blends neural networks and traditional optimization models is proposed. The hybrid model could leverage both the advantages of the prior concealed in data (through neural network learning) and hand-crafted priors (through explicit regularization).
- A special blending mechanism is devised which allows manually adjusting the "weights" of the neural network block and the optimization block, such that the designed architecture could better fit the size of the training datasets.

The remainder of this paper is organized as follows. The hybrid neural network is introduced and described in Section II. In Section III, experiments are carried out to validate the proposed hybrid neural network. We conclude the paper in Section IV.

II. Method

As stated before, the proposed hybrid NN consists of two blocks: CNN denoiser and optimization-based denoiser. To leverage the power of unsupervised learning, we borrow the idea of Noise2Inverse to prepare the training data-sets. This is illustrated in Fig. 1. The backbone of the proposed architecture, i.e. the hybrid NN, is illustrated in Fig. 2

A. Noise2Inverse for LDCT

The key idea of the Noise2Inverse method is that it partitions the projection data into non-overlapping groups of equal size, each of them are then reconstructed into noisy images. Since there are no explicit correlations between the projection data groups, the reconstructed images could be thought of being independent observations, so the assumption for unsupervised training is satisfied. With training data-sets prepared, conventional convolutional neural networks like U-Net [9] and DnCNN [10] could be utilized to achieve denoising.

B. Traditional optimization model

Let o denote the output of the first block. In the forward pass of the proposed hybrid NN, o is fed to the second block, i.e. an optimization model for further processing. Let A denote the desired output of the optimization model, then by adding TV regularizer and non-negative constraint, the optimization model can be written as

$$A = \arg\min_{A\geq 0} \left\{ \frac{1}{2}\|o - A\|_2^2 + \lambda \|\nabla A\|_1 \right\}. \tag{1}$$

A popular algorithm for solving the above model is the primal-dual Chambolle-Pock method. Given $A^0 = 0$, $\xi^t = 0$:

$$\begin{cases} \xi^{t+1} = \xi^t + \tau\lambda\nabla A^t \\ \eta^{t+1} = \mathrm{P_B}\left(\xi^{t+1}\right) \\ A^{t+1} = \frac{A^t + \tau\left(o + \lambda div\left(\eta^{t+1}\right)\right)}{1+\tau} \\ A^{t+1} = \max\left(0, A^{t+1}\right) \end{cases}$$

where t is the iteration number, τ is the time step, and λ is the regularization parameter which controls the denoising strength. $\mathrm{P_B}$ is an element-wise projection operator onto l_2 Ball:

$$\mathrm{P_B}(y) = \begin{cases} y, & \text{if } \|y\|_2 \leq 1 \\ \frac{y}{\|y\|_2}, & \text{if } \|y\|_2 > 1 \end{cases}.$$

When performing backpropagation, one needs to calculate the gradients regarding to the variables A^t. Deep learning frameworks like Pytorch provide tools for automatic gradients calculation. However, in our tests, automatic calculations were rather slow. So, we use the python package cupy to wrap up cuda kernels to serve our need. The required gradients can be derived as below. Let $o_{t+1} = o_t = \cdots o_1$, then we have

$$\begin{cases} dA_{t+1} = \mathcal{R}\left(\frac{1}{1+\tau} + \frac{\tau^2\lambda^2}{1+\tau}\operatorname{div}\mathcal{M}_{t+1}\nabla\right) dA_t \\ \qquad\quad + \mathcal{R}\frac{\tau\lambda}{1+\tau}\operatorname{div}\mathcal{M}_{t+1} d\xi_t + \mathcal{R}\frac{\tau}{1+\tau} do_t \\ d\xi_{t+1} = d\xi_t + \tau\lambda\nabla dA_t \\ do_{t+1} = do_t \end{cases} \tag{2}$$

where, $\mathcal{R}$ denotes the derivative for the non-negative constraint, div is the divergence operation, and $\mathcal{M}_{t+1}$ represents the derivative of the projection operator $\mathrm{P_B}$ in the $(t+1)$th iteration. The matrix form reads

$$\begin{pmatrix} dA_{t+1} \\ d\xi_{t+1} \\ do_{t+1} \end{pmatrix} = V_{t+1} \begin{pmatrix} dA_t \\ d\xi_t \\ do_t \end{pmatrix} \tag{3}$$

with

$$V_{t+1} = \mathcal{N} \begin{pmatrix} \frac{1}{1+\tau} + \frac{\tau^2\lambda^2}{1+\tau}\operatorname{div}\mathcal{M}_{t+1}\nabla & \frac{\tau\lambda}{1+\tau}\operatorname{div}\mathcal{M}_{t+1} & \frac{\tau}{1+\tau} \\ \tau\lambda\nabla & 1 & 0 \\ 0 & 0 & 1 \end{pmatrix},$$

where

$$\mathcal{N} = \begin{pmatrix} \mathcal{R} & & \\ & 1 & \\ & & 1 \end{pmatrix}.$$

Let L denote the loss function, then

$$\begin{aligned} dL &= \mathrm{Tr}\left[\left(\frac{\partial L}{\partial o}\right)^T do\right] = \mathrm{Tr}\left[\left(\frac{\partial L}{\partial A_{t+1}}\right)^T dA_{t+1}\right] \\ &= \mathrm{Tr}\left[\begin{pmatrix} \frac{\partial L}{\partial A_{t+1}} \\ 0 \\ 0 \end{pmatrix}^T V_{t+1} \begin{pmatrix} dA_t \\ d\xi_t \\ do_t \end{pmatrix}\right] \\ &= \mathrm{Tr}\left[\begin{pmatrix} \frac{\partial L}{\partial A_{t+1}} \\ 0 \\ 0 \end{pmatrix}^T V_{t+1}\cdot V_t \cdots V_2 \begin{pmatrix} \mathcal{R}\frac{\tau}{1+\tau} \\ 0 \\ 1 \end{pmatrix} do\right] \end{aligned}$$

where, $\mathrm{Tr}(Q)$ computes the trace of matrix Q.

Finally, the required gradient can be extracted as

$$\frac{\partial L}{\partial o} = \begin{pmatrix} \mathcal{R}\frac{\tau}{1+\tau} \\ 0 \\ 1 \end{pmatrix}^T V_2^T \cdots V_t^T \cdot V_{t+1}^T \begin{pmatrix} \frac{\partial L}{\partial A_{t+1}} \\ 0 \\ 0 \end{pmatrix}.$$

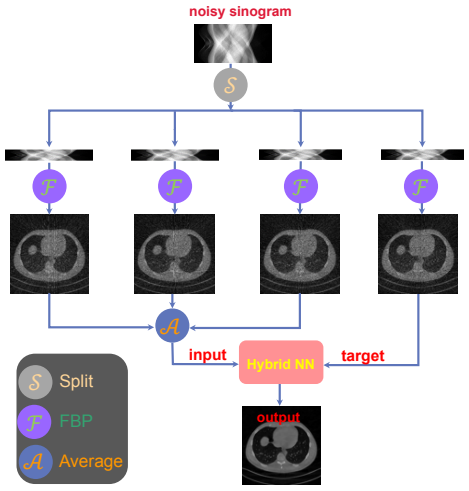


Fig. 1. The flow chart of hybrid NN for LDCT. The input noisy sinogram is split into four groups. Then, FBP is applied to reconstruct 4 noisy images from which the training data-sets are constructed.

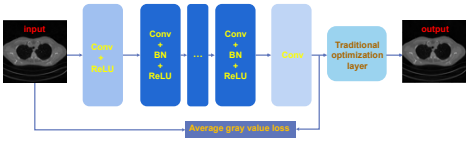


Fig. 2. Illustration of hybrid NN.

C. Loss function

Clearly, there exists an interaction between the two designed blocks, and the hyper-parameters, e.g. the number of primal-dual iterations, the depth of the CNN block, the weighting parameter λ, etc. would affect the networks' performance. Particularly, the two blocks actually run for the same goal, i.e. act as denoisers, the good performance of the second block, i.e. the primal-dual algorithm, might inhibit the performance of the CNN block. To combat such a negative effect, we add a skip connection from the CNN block to the final output of the network, which requires that the CNN block should preserve the average gray values of its input images. This is achieved by adding an additional term to the loss function:

$$loss_{agv} = \|\bar{G}_{input} - \bar{G}_{CNN_{output}}\|^2$$

where $\bar{G}_{(I)}$ means the average gray value of image I. The final loss function is

$$loss_{total} = MSE(target, output) + \alpha \cdot loss_{agv},$$

where MSE means the mean squared error loss, and α is a positive scalar parameter

III. Experiments And Results

A. Datasets and settings

To verify the effectiveness of the proposed hybrid NN, a real clinical dataset was used, which was for "LUNA (LUng Nodule Analysis) 16−ISBI 2016 Challenge" (https://luna16.grand-challenge.org/download/). The dataset contains 1308 thoracic volumes from 1010 people, including 244,527 image slices size of 512×512. We randomly selected 7 patients, 2270 slices, from which 200 slices were randomly chosen as the training dataset and 10 slices as the testing dataset. The projection data are acquired with a virtual CT system equipped with a parallel beam source and a linear detector consisting of 720 cells. The projection data are acquired for 360 projection views uniformly distributed in the angular range $[0, \pi]$. To simulate low-dose radiation, Poisson noise with incident intensity I_0 is added to the raw data as follows

$$p_{noisy} = -ln\left(\frac{I_0 \times e^{-p}}{I_0}\right), \quad (4)$$

where p and p_{noisy} denote the noise-free and noisy sinogram data, respectively. In our tests, we set $I_0 = 3 \times 10^4$. As Fig.1 shows, we split the sinogram into four non-overlapping groups in the way that each group consists of projection angles uniformly distributed in $[0, \pi]$.

The proposed hybrid NN is implemented with the PyTorch framework. All the experiments are executed on a a single graphic processing NVIDIA card RTX 2080Ti with 11GB video memory. The gradient $\frac{\partial L}{\partial o}$ is computed by cuda kernels wraped by cupy https://github.com/cupy/cupy.

B. Hyperparameters selection

For the proposed hybrid NN, the hyperparameters include the number of layers of the CNN, the number of primal-dual iterations, the scalar parameters τ, λ for the optimization layer and α for the loss function. Considering the convergence requirement of the primal-dual iterations, We set $\tau = \frac{1}{2\lambda}$. In our tests, the CNN has 12 layers, and α is set in a trial and error manner.

C. The effect of the explicit prior on the performance of CNN

To verify whether the second block, i.e. the explicit regularization has an influence on the performance of the CNN block, the following experiments were carried out: training the hybrid NN with increasing number of the primal-dual iterations t =10, 30 and 50, then testing the first block, i.e. CNN along with the learned weights. The results shown in Fig.3 indicate that, with increasing t, the performance of the CNN block becomes poorer and poorer. This coincides with our intuition. By increasing t, the performance of the optimization model (second block) is improved, so the importance of the CNN block gets weaken.

As Section II-C has stated, to preserve the performance of the CNN block, the contrast-preserving loss $loss_{agv}$ is added to the loss function. The experimental results shown in Fig.4 suggest that, to some extent, the $loss_{agv}$ is indeed helpful to bring back the performance of the CNN block.

D. Comparison with the Noise2Inverse model

To verify the superiority of the proposed hybrid NN, experiments against the Noise2Inverse model are performed. For both methods, the training runs for 150 epochs. Even though there are many perspectives for comparison, in this experiment, we check how the two methods behave against noise

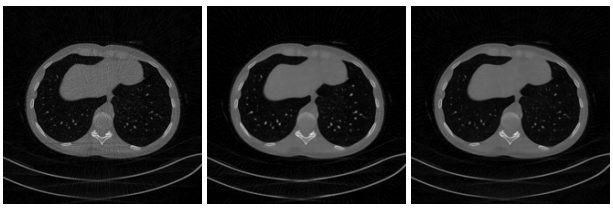
(a) $t = 10$ (b) $t = 30$ (c) $t = 50$

Fig. 3. Testing results of the CNN block, with iteration number $t = 10, 30, 50$ for the primal-dual algorithm, respectively.

(a) Reference (b) without $loss_{agv}$ (c) with $loss_{agv}$

Fig. 4. Results without and with $loss_{agv}$, $t = 50$. The upper right corners show the zoomed-ins for the red-framed regions.

change. Two tests with different noise levels, i.e. $I_0 = 3 \times 10^4$ and $I_0 = 1 \times 10^4$ are performed, and the results are shown in Fig.5. For the low level noise case ($I_0 = 3 \times 10^4$), the two methods achieve similar quality. As Fig.5 (b) and (c) show there are little visual differences. The quantitative PSNR indices are 32.522 and 33.820, and the SSIM values are 0.735 and 0.799, respectively, which might weakly indicate the advantages of the proposed hybrid NN. When checking the results with higher noise level ($I_0 = 1 \times 10^4$), however, the advantages of the proposed hybrid NN become apparent. As shown in Fig.5(e) and (f), there are remaining noise within the result of Noise2Inverse, while hybrid NN still achieve high quality denoising, similar to the low level noise case.



(a) FBP, $I_0 = 3 \times 10^4$ (b) Noise2Inverse (c) Hybrid NN

(d) FBP, $I_0 = 1 \times 10^4$ (e) Noise2Inverse (f) Hybrid NN

Fig. 5. Comparison with Noise2Inverse. The two rows show the results with noise levels corresponding to $I_0 = 3 \times 10^4$ and $I_0 = 1 \times 10^4$, respectively.

IV. Conclusion

Deep learning based methods achieve state-of-the-art results for low-dose CT reconstructions. The required training datasets, however, might not be available for real applications. Another important issue is that neural networks usually suffer from intrinsic instability. In this paper, we propose a hybrid NN aiming to leverage the power of both learning based methods and conventional optimization based methods. By adding an average contrast preserving loss, the two blocks, i.e. CNN and optimization algorithm, could work in harmony such that the hybrid NN performs better than any of them alone.

Even though the hybrid NN in this paper consists of a CNN block and a TV denoising algorithm, the basic idea actually allows for any possible combinations of a neural network architecture and a optimization model.

Acknowledgment

Thanks for the support of the National Natural Science Foundation of China (NSFC) (61971292, 61827809 and 61871275).

References

[1] Vegard Antun, Francesco Renna, Clarice Poon, Ben Adcock, and Anders C Hansen. On instabilities of deep learning in image reconstruction and the potential costs of ai. *Proceedings of the National Academy of Sciences*, 117(48):30088–30095, 2020.

[2] Joshua Batson and Loic Royer. Noise2self: Blind denoising by self-supervision. In *International Conference on Machine Learning*, pages 524–533. PMLR, 2019.

[3] Antoni Buades, Bartomeu Coll, and J-M Morel. A non-local algorithm for image denoising. In *2005 IEEE Computer Society Conference on Computer Vision and Pattern Recognition (CVPR'05)*, volume 2, pages 60–65. IEEE, 2005.

[4] Kostadin Dabov, Alessandro Foi, Vladimir Katkovnik, and Karen Egiazarian. Image denoising by sparse 3-d transform-domain collaborative filtering. *IEEE Transactions on image processing*, 16(8):2080–2095, 2007.

[5] Allard Adriaan Hendriksen, Daniël Maria Pelt, and K Joost Batenburg. Noise2inverse: Self-supervised deep convolutional denoising for tomography. *IEEE Transactions on Computational Imaging*, 6:1320–1335, 2020.

[6] Fan Jia, Jun Liu, and Xue-Cheng Tai. A regularized convolutional neural network for semantic image segmentation. *Analysis and Applications*, 19(01):147–165, 2021.

[7] Alexander Krull, Tim-Oliver Buchholz, and Florian Jug. Noise2void - learning denoising from single noisy images. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, June 2019.

[8] Jaakko Lehtinen, Jacob Munkberg, Jon Hasselgren, Samuli Laine, Tero Karras, Miika Aittala, and Timo Aila. Noise2noise: Learning image restoration without clean data. In *International Conference on Machine Learning*, pages 2965–2974. PMLR, 2018.

[9] O. Ronneberger, P.Fischer, and T. Brox. U-net: Convolutional networks for biomedical image segmentation. In *Medical Image Computing and Computer-Assisted Intervention (MICCAI)*, volume 9351 of *LNCS*, pages 234–241. Springer, 2015. (available on arXiv:1505.04597 [cs.CV]).

[10] Kai Zhang, Wangmeng Zuo, Yunjin Chen, Deyu Meng, and Lei Zhang. Beyond a gaussian denoiser: Residual learning of deep cnn for image denoising. *IEEE transactions on image processing*, 26(7):3142–3155, 2017.

High Resolution Cerebral Perfusion Deconvolution via Mixture of Gaussian Model based on Noise Properties

Sui Li, Zhaoying Bian, Dong Zeng, and Jianhua Ma

***Abstract*—Cerebral perfusion computed tomography (CPCT) imaging provides a rapid and accurate noninvasive measurements of the acute stroke by generating hemodynamic parameter maps with a qualitative and quantitative way. However, due to it performs a multiple consecutive scanning protocol at one area of the head, the radiation exposure is relatively higher than a routine protocol. And lowering radiation dose in CPCT protocol would increase the amount of noise and hence influence hemodynamic parameters for patients with acute stroke. Some advanced methods have been proposed and show a great potential in noise suppression for low-dose CPCT imaging. And most of them assume that the embedded noise obeys an independent and identically distribution (i.i.d), but the noise may be more complicated in practical scenarios. In this work, we first analysze the noise properties in low-dose CPCT images. And then present a novel perfusion deconvolution method with a self-relative structure similarity information and a mixture of Gaussians (MoG) noise model (named SR-MoG) to accurately estimate the hemodynamic parameters directly at the low radiation exposure. Experiments implemented on digital brain perfusion phantom verify that the presented SR-MoG method can achieve promising gains over the existing deconvolution approaches.**

***Index Terms*—Cerebral perfusion CT, mixture of Gaussian model, noise properties, deconvolution.**

I. Introduction

CEREBRAL stroke is an acute cerebrovascular disease, it has a high incidence rate, high mortality rate, and high disability rate [1]. Cerebral perfusion computed tomography (CPCT) is an imaging modality that provides detailed information about blood flow to the brain in a rapid, noninvasive, and quantitative way for cerebral stroke diagnosis. Specifically, the CPCT measures the temporal changes for each voxel through a series of repeating CT scanning, which is used to improve detection of acute infarction and help assess the degree of collateral circulation by providing hemodynamic parameters maps (HPMs), i.e., cerebral blood flow (CBF), cerebral blood volume (CBV), and mean transit time (MTT). [2]. However, CPCT performs repeated scans on the same region, so it is associated with high radiation dose. And lowering radiation dose in CPCT imaging would increase the amount of noise and hence influence HPMs' accuracy.

To reduce radiation dose in CPCT imaging while maintaining HPMs' accuracy, various algorithms have been proposed. Generally, related algorithms can be grouped into two categories. The first category is to reduce noise in the reconstructed CPCT images at low-dose measurements and to estimate HPMs via standard deconvolution process [3-5], e.g., image-based restoration algorithms, statistical iterative reconstruction algorithms. However, since noise distribution characteristics are not considered accurately, these methods might suffer from resolution loss, thereby degrading the HPMs' accuracy. The second category is to directly calculate HPMs via design strong prior information to stable the residue function with an iterative deconvolution procedure [6-7]. These models can improve the HPMs' accuracy in the case of low-dose measurement. However, most of models are based on the assumption that the embedded noise in the CPCT images is an independent and identically distribution, and this would deviate from the practical scenarios wherein the noise distributions among different frames could be significantly different from each other.

Therefore, supposing the noise in CPCT imaging is complex and does not obey an i.i.d, in this work, we present a novel framework to estimate HPMs by characterizing the noise distribution properly. Specifically, a non-independent and identically distributed mixture of Gaussians (MoG) noise model is utilized to modeling the residue function in the perfusion deconvolution procedure. In addition, we introduce the self-relative structure similarity information of CPCT images as a prior term, which can be easily integrated into the MoG model to further improve the HPMs' results. Finally, the fast iterative shrinkage-thresholding algorithm (FISTA) [8] is developed to solve this model. The experimental results of the digital brain perfusion phantom simulation demonstrate the presented algorithm can greatly achieve promising gains over the existing deconvolution approaches.

S. Li, Z. Bian, and D. Zeng are with the School of Biomedical Engineering, Southern Medical University, Guangzhou 510515, China, and also with Guangzhou Key Laboratory of Medical Radiation Imaging and Detection Technology, Southern Medical University, Guangzhou 510515, China.

*J. Ma is the corresponding author with the School of Biomedical Engineering, Southern Medical University, Guangzhou 510515, China, and also with Guangzhou Key Laboratory of Medical Radiation Imaging and Detection Technology, Southern Medical University, Guangzhou 510515, China. (Email: jhma@smu.edu.cn).

II. Methodology

A. The noise properties for low-dose CPCT image

Fig. 1 shows the normal-dose CPCT images, the low-dose CPCT images, and the corresponding difference maps at three different frames (the 1st, 12th, and 20th frames, respectively). As depicted in Fig. 1(d), it shows the corresponding histograms of the noise maps in Fig. 1(c). It can be observed that the peak of histogram at the 12th frame is lower than those at the 1th and 20th frames. The possible reason is that the 12th frame contains more contrast than the other frames, and it would suffer much more noise-induced artifacts due to the contrast agent. Fig. 2 further illustrates the different types of noise at different areas at a single frame. It can be observed that different areas contains different noise distribution at a single frame. Therefore, it can be concluded that the noise distribution in low-dose CPCT images is complex, and in this work, we exploit the mixture of Gaussian (MoG) to model the noise E of low-dose CPCT image as follows:

$$P_E(E|\Pi,\Sigma)=\sum_{k=1}^{K}\pi_k\mathcal{N}(E|0,\Sigma_k)\,, \tag{1}$$

where $\pi_k \geq 0$ is the mixing coefficient with $\sum_k \pi_k = 1$. And $\mathcal{N}(\cdot\,|0,\Sigma_k)$ represents a Gaussian distribution with zero mean and variance Σ_k.

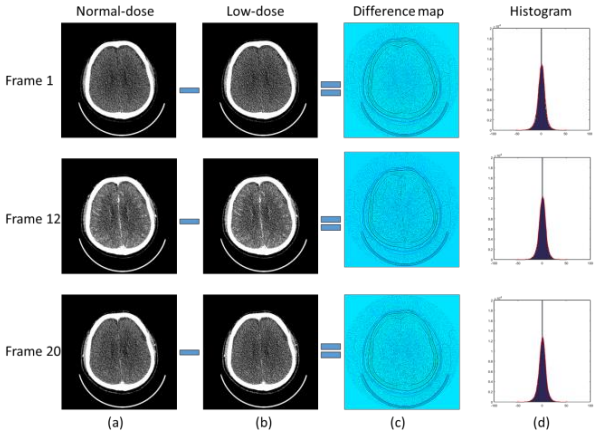


Fig. 1. Installation of (a) the normal-dose CPCT images at frame 1, 12 and 20, (b) the low-dose CPCT images at frame 1, 12 and 20, (c) the difference images from (a) and (b), and (d) the histogram map of the difference images from (c)..

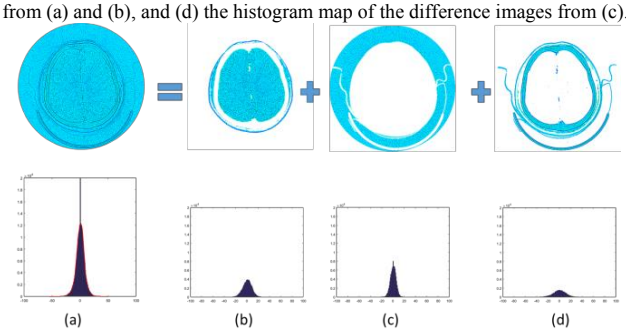


Fig. 2. (a) Illustration of the noise at the 12th CPCT frame and its histogram. (b)-(d) Illustration of three different Gaussian components and their histograms, obtained by MoG model with three Gaussian components of the noise in (a).

B. SR-MoG deconvolution model

The CPCT images can be considered as the combination of the background component with temporal similarity and the foreground component with spatial contiguity and noise [9]. Meanwhile, the concentration maps are obtained by subtracting the baseline image wherein the baseline image equals the average of early CPCT frames before the contrast agent enters the blood vessel actually. Moreover, the residue function data reflects the relative concentration of contrast agent in the blood vessels over time. Therefore, it can be concluded that the residue function can also be considered as the combination of the background component and foreground component as well. Motivated by these observations, the residue function data X consists of background component U, foreground component F, and embedded noise component E, which can be modeled as follows:

$$X = U + F + E \tag{2}$$

According to Eq. (2) and the noise model in Eq. (1), we can obtain following likelihood distribution for U and F:

$$\begin{aligned} P_X(X|U,F,\Pi,\Sigma) &= P_E(X-U-F|\Pi,\Sigma) \\ &= \sum_{k\&=1}^{K}\pi_k\mathcal{N}(X-U-F|0,\Sigma_k) \end{aligned} \tag{3}$$

To estimate the underlying variable U, we introduce a we introduce self-relative information B to provide prior information for U, , then, we provide following prior distribution for U:

$$P_U(U) = \mathcal{N}(U|B,\sigma_1) \tag{4}$$

where σ_1 denotes the variance of the inevitable noise in B.

Moreover, since the foreground component F is variant and sparse over time, it is rational to introduce a sparse prior distribution $P_F(F)$ on it. This prior can be express as a regularization term when estimating F. By combining the priors and the likelihood in Eq. (4), it can be then expressed as following posteriori distribution:

$$\begin{aligned} P(U,F,\Pi,\Sigma|X) &\propto P_X(X|U,F,\Pi,\Sigma)\times P_U(U)\times P_F(F) \\ &= \sum_{k}\pi_k\mathcal{N}(X-F-U|0,\Sigma_k) \\ &\times \mathcal{N}(U|B,\sigma_1)\times P_F(F) \end{aligned} \tag{5}$$

Therefore, by integrating the perfusion deconvolution procedure with the likelihood model as aforementioned, the perfusion deconvolution model can be rewritten as:

$$\begin{aligned} \hat{X} = \arg\min & \frac{1}{2}\|AX-C\|_2^2 \\ & + \log\sum_{k=1}^{K}\pi_k\mathcal{N}(X-U-F|0,\Sigma_k) \\ & + \log N(U|B,\sigma_1) + \beta\log P_F(F) \\ & s.t.\, X = U+F+E \end{aligned} \tag{6}$$

where the $\log P_F(F) = \|D(F)\|_F^2$ and the operator $D(\cdot)$ is the first order difference matrix. Then we can estimate the underlying background and foreground, as well as other

involved parameters. By incorporating the self-relative structural similarity information and MoG noise distribution latent in CPCT images into a perfusion deconvolution model, we can define the self-relative structural-information-induced mixture of Gaussian (SR-MoG) model.

III. EXPERIMENTS

A. Data

To validate and evaluate the performance of the presented SR-MoG perfusion deconvolution model at low-dose cases, a digital brain perfusion phantom package [10] was used in this work. The specific digital brain perfusion phantom consists of infarct core and ischemic penumbra in the white and gray matters, as well as the healthy tissue with 40 temporal frames at size 256 × 256. Fig. 3 shows the phantom and the corresponding time-density curves of penumbral, strike core, tissue and vessel. Additionally, to simulate low-dose CPCT images under different dose levels, the simulate strategy is based on the previous study [11], and two dose levels were simulated with 50% and 25% of normal-dose.

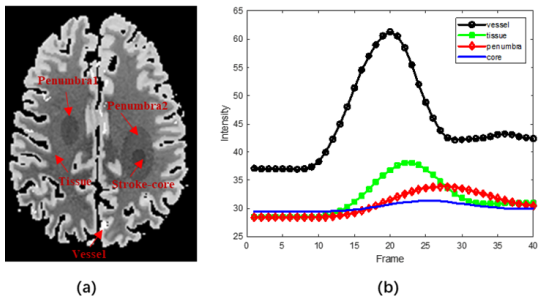


Fig. 3 Illustration of digital CPCT phantom (a) and the corresponding TDCs (b). The display window is [10, 65] HU.

B. Comparison methods

To validate and evaluate the performance of the presented SR-MoG method, five deconvolution methods were compared against: block-circulant truncated SVD (bSVD) [12], Tikhonov regularization [13], the tensor total variation regularized approach (TTV) [6], the structure total variation regularized approach (STV) [7] and the SR-MoG with a single Gaussian noise component (SG). Extensive experiments with different parameter settings were conducted for all competing methods, and the appropriate parameters are selected based on the optimal visual inspection.

IV. RESULT

Fig. 4 shows the estimated HPMs, i.e., CBF, CBV, and MTT generated by different algorithms of digital brain perfusion phantom with 50% dose level. The reference HPMs obtained by the digital brain perfusion phantom package directly, thus it can be as the "golden standard" in this comparison. The remaining HPMs all estimated from the FBP reconstructed low-dose CPCT images by the competing deconvolution methods for comparison. It can be observed that several noise is evident in the traditional deconvolution methods, i.e., bSVD, Tikhonov, especially in healthy region of brain. And for the TTV and STV algorithms, the results shows more improvement by visual inspection, while the detailed structure of the stroke-core region is still corrupted by the noise-induced artifacts, as shown by the red arrows at CBF maps in Fig. 4. Moreover, the results from SG and SR-MoG are remarkable over that from TTV and STV in terms of noise reduction and structure information preservation. However, by careful visual inspection, the values of SG at penumbra region, indicated by yellow arrows in Fig. 4, are lighter than that in reference, and results of SR-MoG are closer than other deconvolution approach.

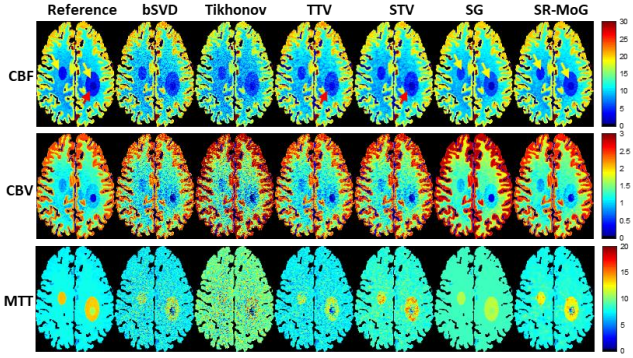


Fig. 4 Perfusion maps of the digital brain perfusion phantom estimated by bSVD, Tikhonov, TTV, STV, SG and SR-MoG approaches, respectively, under the 50% dose. CBF in unit of mL/100 g/min, CBF in mL/100 g, MTT in sec.

Fig. 5 illustrates the performance of all competing methods on digital phantom with 25% dose, respectively. With the radiation dose further reducing, all compared methods fail to remove all noise, especially at 25% dose. While the presented SR-MoG performed better than other compared methods in terms of noise suppression, structure preservation and value consistency, especially for the penumbra and stroke-core regions.

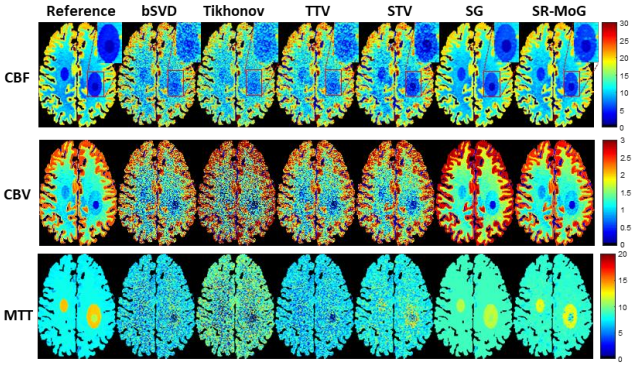


Fig. 5 Perfusion maps of the digital brain perfusion phantom estimated by bSVD, Tikhonov, TTV, STV, SG and SR-MoG approaches, respectively, under the 25% dose. CBF in unit of mL/100 g/min, CBF in mL/100 g, MTT in sec.

To further demonstrate the performance of the presented SR-MoG method, RMSE and SSIM measurements are utilized and corresponding comparison results of the different methods at 50% dose and 25% dose, and the result is listed in Table. 1. It can be seen that the presented SR-MoG method outperforms the other competing methods by a large margins in terms of the RMSE and SSIM measurements at the two dose levels. Thus results demonstrate the presented SR-MoG deconvolution algorithm achieve significant ability compared with other algorithm for noise suppression and structure consistent.

Table 1. RMSE and SSIM for HPMs (CBF, CBV, MTT) estimated using the different competing methods at 25% and 50% of normal dose. Bold font indicates the best performance.

		CBF		CBV		MTT	
Methods	Dose	RMSE	SSIM	RMSE	SSIM	RMSE	SSIM
bSVD	25%	2.9152	0.9123	0.4537	0.9955	2.4328	0.9213
Tikhonov		4.6676	0.9143	1.8534	0.9790	3.4193	0.8510
TTV		2.8414	0.9236	0.4288	0.9960	2.3473	0.9301
STV		2.9210	0.9502	0.5705	0.9944	1.7319	0.9564
SG		2.0843	0.9978	0.5974	0.9946	2.9620	0.9321
SR-MoG		**1.9456**	**0.9963**	**0.1405**	**0.9990**	**1.6187**	**0.9718**
bSVD	50%	1.4844	0.9741	0.2280	0.9988	1.3656	0.9730
Tikhonov		4.1308	0.9690	1.7753	0.9849	2.6510	0.9158
TTV		1.2916	0.9828	0.2190	0.9990	1.3018	0.9765
STV		1.7399	0.9868	0.4508	0.9969	1.2121	0.9793
SG		1.0834	0.9978	0.5960	0.9947	11.1994	0.8917
SR-MoG		**0.7962**	**0.9988**	**0.1616**	**0.9990**	**0.8843**	**0.9889**

V. Conclusion

In this work, we analyze the noise properties in low-dose CPCT images and then we find that the noise latent in CPCT images can be modeled as MoG distribution. Meanwhile, based on the analysis, we present a novel CPCT deconvolution method using the self-relative structure similarity information and MoG distribution latent in CPCT images to estimate the HPMs directly. Specifically, the self-relative structure similarity information is obtained through pre-processed low-dose CPCT images. Digital brain perfusion phantom simulation is used to evaluate the deconvolution performance of the presented SR-MoG method qualitatively and quantitatively.

Acknowledgment

This work was supported in part by the NSFC under Grant U21A6005 and Grant U1708261.

References

[1] B. C. V. Campbell, D. A. De Silva, M. R. Macleod, S. B. Coutts, L. H. Schwamm, S. M. Davis, and G. A. Donnan, "Ischaemic stroke," Nature Reviews Disease Primers, vol. 5, no. 1, p. 70, 2019.

[2] A. Konstas, G. Goldmakher, T.-Y. Lee, and M. Lev, "Theoretic basis and technical implementations of CT perfusion in acute ischemic stroke, part 2: Technical implementations," American Journal of Neuroradiology, vol. 30, no. 5, pp. 885–892, 2009.

[3] S. Li, D. Zeng, J. Peng, Z. Bian, H. Zhang, Q. Xie, Y. Wang, Y. Liao, S. Zhang, J. Huang, D. Meng, Z. Xu, and J. Ma, "An efficient iterative cerebral perfusion CT reconstruction via low-rank tensor decomposition with spatial–temporal total variation regularization," IEEE Transactions on Medical Imaging, vol. 38, pp. 360–370, 2019.

[4] S. Li, D. Zeng, Z. Bian, D. Li, M. Zhu, J. Huang, and J. Ma, "Learning non-local perfusion textures for high-quality computed tomography perfusion imaging," Physics in Medicine & Biology, vol. 66, no. 11, p. 115007, 2021.

[5] J. Ma, H. Zhang, Y. Gao, J. Huang, Z. Liang, Q. Feng, and W. Chen, "Iterative image reconstruction for cerebral perfusion CT using a precontrast scan induced edge-preserving prior," Phys. Med. Biol., vol. 57, pp. 7519-7542, 2012.

[6] R. Fang, S. Zhang, T. Chen, and P. Sanelli, "Robust low-dose CT perfusion deconvolution via tensor total-variation regularization," IEEE Transactions on Medical Imaging, vol. 34, 2015.

[7] D. Zeng, X. Zhang, Z. Bian, J. Huang, H. Zhang, L. Lu, W. Lyu, J. Zhang, Q. Feng, W. Chen, and J. Ma, "Cerebral perfusion computed tomography deconvolution via structure tensor total variation regularization." Medical Physics, vol. 43, pp. 2091-2107, 2016.

[8] A. Beck and M. Teboulle, "A fast iterative shrinkage-thresholding algorithm for linear inverse problems," SIAM journal on imaging sciences, vol. 2, no. 1, pp. 183–202, 2009.

[9] D. Zeng, Q. Xie, W. Cao, J. Lin, H. Zhang, S. Zhang, J. Huang, Z. Bian, D. Meng, Z. Xu, Z. Liang, W. Chen, and J. Ma, "Low-dose dynamic cerebral perfusion computed tomography reconstruction via kroneckerbasis- representation tensor sparsity regularization," IEEE Transactions on Medical Imaging, vol. 36, pp. 2546–2556, 2017.

[10] https://lme.tf.fau.de/dataset/digital-brain-perfusion-phantom

[11] J. Ma, Z. Liang, Y. Fan, Y. Liu, J. Huang, W. Chen, and H. Lu, "Variance analysis of x-ray CT sinograms in the presence of electronic noise background," Medical Physics, vol. 39, no. 7Part1, pp. 4051–4065, 2012.

[12] O. Wu, L. Østergaard, R. M. Weisskoff, T. Benner, B. R. Rosen, and A. G. Sorensen, "Tracer arrival timing-insensitive technique for estimating flow in MR perfusion-weighted imaging using singular value decomposition with a block-circulant deconvolution matrix.," Magnetic Resonance in Medicine, vol. 50, pp. 164–74, Jul 2003.

[13] T. S. Koh, X. Y. Wu, L. H. Cheong, and C. C. T. Lim, "Assessment of perfusion by dynamic contrast-enhanced imaging using a deconvolution approach based on regression and singular value decomposition," IEEE Transactions on Medical Imaging, vol. 23, no. 12, pp. 1532–1542, 2004.

Simulating Arbitrary Dose Levels and Independent Noise Image Pairs from a Single CT Scan

Sen Wang, Adam Wang

***Abstract*—Deep learning-based image denoising and reconstruction methods have shown promising results for low-dose CT. When high-quality reference images are not available for training the network, researchers found a powerful and effective counterpart called Noise2Noise, which trains the neural network using paired data with independent noise. However, it is uncommon to have paired CT scans with independent noise (e.g., from two scans). In this paper, a method is proposed to generate such paired data for potential usage in deep learning training by simultaneously simulating a low-dose image at arbitrary dose level and an image with independent noise from a single CT scan. Their independence is investigated both analytically and numerically. In our numerical study, a Shepp-Logan phantom was utilized in MATLAB to generate the ground-truth, normal-dose, and low-dose images for reference. Noise images were obtained for analysis by subtracting the ground-truth from the noisy images, including the normal-dose/low-dose images and the paired products of our proposed method. Our numerical study matches the analytical results very well, showing that the paired images are not correlated. Under an additional assumption that they form a bivariate normal distribution, they are also independent. The proposed method can produce a series of paired images at arbitrary dose level given one CT scan, which provides a powerful new method to enrich the diversity of low-dose data for deep learning.**

***Index Terms* — low-dose CT, synthetic CT, neural network**

I. Introduction

Low-dose computed tomography (CT) is one of the most direct and effective ways to reduce the radiation dose to patients. However, a trade-off between image quality and patient dose always exists. Many efforts have been put into this area to find better ways of balancing the trade-off. Deep learning methods are one of the most recent and promising developments in reducing noise in CT imaging. When high-quality images are accessible for training, neural networks trained either in image domain [1], [2] or during the reconstruction process [3], [4] showed promising performance.

On the other hand, there are several works attempting to tackle the problem without the presence of high-quality images by exploiting the Noise2Noise pipeline[5]. Wu *et al* [6] showed that a denoising network with Noise2Noise training is equivalent to training with clean labels (high-quality images) when a few conditions are satisfied. One of the four conditions is that the network should have paired noisy data with zero-mean, independent noise. For Noise2Noise application in CT imaging, it is crucial to find such paired data. While such data could be acquired with two scans of the same patient, this exposes the patient to additional dose and will have misregistration artifacts. Pairing simulated low-dose images with the original normal-dose images does not satisfy this condition since some of the noise in the simulated low-dose image comes from the normal-dose image, so the two images do not have independent noise. In one Noise2Noise approach, Wu *et al* [7] constructed the independent image pairs via random projection splitting. Yuan *et al* [8] proposed a Noise2Noise based denoising method named 'Half2Half'. In their training pair construction, binomial selection was applied to the projection data, splitting it into two pseudo half-dose scans.

For the aforementioned methods, the dose allocation is fixed, and both of them split dose evenly to the paired images. In this paper, we propose a method to simulate arbitrary dose levels and independent noise from an existing CT scan. Paired images can be generated at any desired dose reduction level from a single CT scan, which provides more diversity in training data given the same normal-dose CT scans.

II. Methods

For simplicity, the normal-dose projection domain measurements (raw data) P_{ND} can be modeled as the sum of a Poisson and Gaussian random variable [8]:

$$P_{ND} \sim \text{Poisson}(\lambda) + \text{Gaussian}(0, \sigma_e^2), \qquad (1)$$

where λ is the mean counts and σ_e is the standard deviation of electronic noise. If we denote the photon counts from the source as I_0 and the object pathlength as l, the mean counts can be

Sen Wang is with the Department of Radiology, Stanford University, Stanford, CA USA 94305. (Email: senwang@stanford.edu).

Adam Wang is with the Departments of Radiology and Electrical Engineering, Stanford University, Stanford, CA USA 94305. (Email: adamwang@stanford.edu)

formulated as

$$\lambda = I_0 \exp(-l). \quad (2)$$

Hence, the expectation and variance of P_{ND} can be given by equations (3) and (4):

$$E[P_{ND}] = \lambda, \quad (3)$$
$$\mathrm{Var}(P_{ND}) = \lambda + \sigma_e^2. \quad (4)$$

For a specified dose level d $(0 < d < 1)$, according to equation (1) the measured counts $P_{LD}^{(d)} \sim \mathrm{Poisson}(d\lambda) + \mathrm{Gaussian}(0, \sigma_e^2)$. Similarly, the expectation and variance of $P_{LD}^{(d)}$ are $E[P_{LD}^{(d)}] = d\lambda$ and $\mathrm{Var}(P_{LD}^{(d)}) = d\lambda + \sigma_e^2$, respectively.

For the low-dose simulation process, we want to emulate the behavior of $P_{LD}^{(d)}$ at dose level d from normal-dose scan P_{ND}. For conventional low-dose simulation, this is a well-known process to insert noise in the projection data. Detailed steps are listed in Table 1.

TABLE 1. CONVENTIONAL NOISE INSERTION

Step	Operation
1	Generate $Q \sim \mathrm{Gaussian}(0, \lambda)$ and $E \sim \mathrm{Gaussian}(0, \sigma_e^2)$.
2	Let $P_{sLD}^{(d)} = d(P_{ND} + a(Q + bE))$, where $a = \sqrt{\frac{1}{d} - 1}$, $b = \sqrt{\frac{1}{d} + 1}$. Then $E[P_{sLD}^{(d)}] = d\lambda$, $\mathrm{Var}(P_{sLD}^{(d)}) = d^2((\lambda + \sigma_e^2) + a^2(\lambda + b^2\sigma_e^2)) = d\lambda + \sigma_e^2$.
3	$P_{LD}^{(d)}$ and $P_{sLD}^{(d)}$ are independent and identically distributed random variables.

The conventional noise insertion adds additional quantum noise Q and electronic noise E, which are scaled by a factor depending on the dose level. In practice, when generating Q, we use P_{ND} as a surrogate for variance λ since the true λ is unknown from a single realization. The result is defined as:

$$P_{sLD}^{(d)} = d(P_{ND} + a(Q + bE)), \quad (5)$$

which can be viewed as a synthetic projection acquired at dose level d as it shares the identical probability distribution (noise properties) as $P_{LD}^{(d)}$. While this enables a simulated low-dose image, we still need a paired zero-mean, independent noise realization for Noise2Noise training. To this end, we define $P_{IN}^{(d)}$ as:

$$P_{IN}^{(d)} = d\left(P_{ND} - \frac{1}{a}\left(Q + \frac{1}{b}E\right)\right). \quad (6)$$

As is shown in the Appendix, we prove that

$$E[P_{sLD}^{(d)} P_{IN}^{(d)}] = d^2\lambda^2 = E[P_{sLD}^{(d)}]E[P_{IN}^{(d)}], \quad (7)$$

which means that they are uncorrelated. On the assumption that $(P_{sLD}^{(d)}, P_{IN}^{(d)})$ form a bivariate normal distribution, they are independent if they are uncorrelated. This is a reasonable assumption for any modest number of photon counts, where the Poisson distribution is approximately Gaussian, and the other noise (electronic, added noise Q, E) are all Gaussian. Importantly, both $P_{sLD}^{(d)}$ and $P_{IN}^{(d)}$ use the same noise realizations of Q and E, but they are scaled inversely and subtracted in $P_{IN}^{(d)}$ as compared to $P_{sLD}^{(d)}$, which leads to the uncorrelated property.

We thus have $P_{IN}^{(d)}$, which has zero-mean, independent noise of $P_{sLD}^{(d)}$. In the Noise2Noise conditions, independent noisy image pairs are required. The proposed method can generate such paired data $(P_{sLD}^{(d)}, P_{IN}^{(d)})$, where $P_{sLD}^{(d)}$ simulates data acquired at arbitrary dose level d. Note that $P_{IN}^{(d)}$ does not correspond to any specific dose level, but rather is designed to satisfy the Noise2Noise conditions. This enables a diversity of dose levels, which may be beneficial to training CT denoising networks.

III. NUMERICAL SIMULATION

To validate the proposed method, numerical simulation was carried out in MATLAB with the built-in Shepp-Logan phantom and Radon projection method for a monoenergetic source.

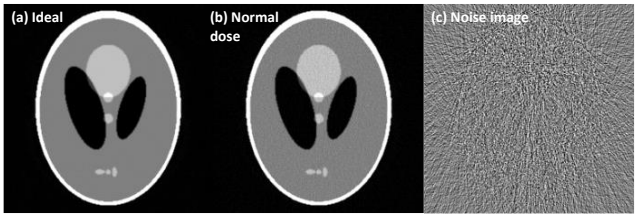


Fig. 1. Reconstructed image and noise image of Shepp-Logan phantom. (a) ideal image, noiseless ground-truth. (b) reconstructed image using normal dose, flux of source I_0= 5e4 photons, σ_e= 5 counts. (c) noise image, subtracting (a) from (b). The display window for (a) and (b) is [0, 0.4] cm^{-1}. The display window for (c) is [-0.02, 0.02] cm^{-1}.

In the simulation, the x-ray source flux was set to 5e4 photons per ray and σ_e= 5 counts, which is referred to as normal dose for the remainder of the paper.

The projections were reconstructed with filtered backprojection (FBP), and the images are illustrated in Fig. 1. We include the ideal image with noiseless projections [using equation (2)] in Fig. 1(a), which is the ground-truth image. Fig. 1(b) is the reconstructed image under normal dose [using equation (1)]. By subtracting Fig. 1(a) from Fig. 1(b), we obtain the noise image as shown in Fig. 1(c).

From the normal dose projections, it is possible to synthesize projections at a specific dose level following the conventional noise insertion steps in Table 1. We can also simulate a real low-dose scan acquired at the same dose level. Fig. 2 displays the results of both reconstructed and noise images for normal dose, real low-dose ($R_{LD}^{(0.3)}$, the reconstructed image from $P_{LD}^{(0.3)}$), and synthetic low-dose ($R_{sLD}^{(0.3)}$, the reconstructed image from $P_{sLD}^{(0.3)}$), respectively, for dose level d=0.3, or 30% of the normal dose. Standard deviations of the phantom region are labeled on the noise images, where we find good correspondence between the synthetic low-dose image and the real low-dose image, as well as the increased noise in the low-dose images compared

with the normal dose.

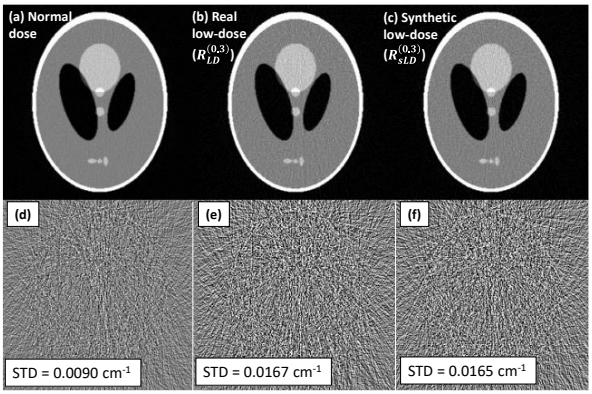


Fig. 2. Noise insertion results. (a), (b) and (c) are reconstructed and noise images for normal dose, real low-dose, and synthetic low-dose, respectively. Noise images are determined by subtracting the ground-truth image in Fig. 1(a). The display window for the first row (a-c) is [0, 0.4] cm^{-1} and is [-0.02, 0.02] cm^{-1} for the second row (d-f).

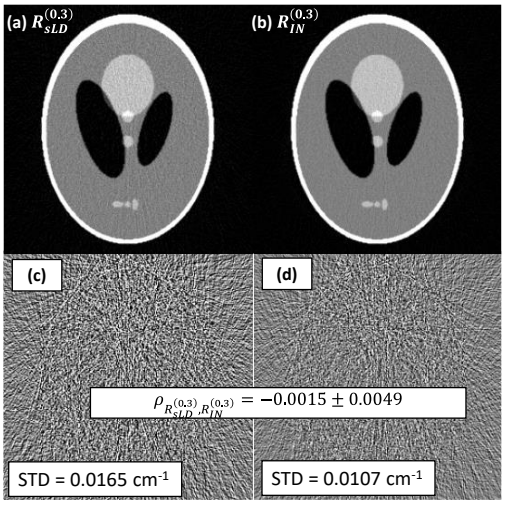


Fig. 3. Independent noise images at 30% dose level. Image $R_{sLD}^{(0.3)}$ is at simulated low dose d=0.3, and $R_{IN}^{(0.3)}$ has noise that is independent of $R_{sLD}^{(0.3)}$ ($\rho_{R_{sLD}^{(0.3)},R_{IN}^{(0.3)}} = -0.0015$), although it has a different noise magnitude. The display window is [0, 0.4] cm^{-1} for the first row and [-0.02, 0.02] cm^{-1} for the second row.

Equations (5) and (6) guide us in the generation of independent noise images from normal dose images. Fig. 3 shows a realization of the ($R_{sLD}^{(0.3)}$, $R_{IN}^{(0.3)}$) pair at 30% dose level. Correlation $\rho_{R_{sLD}^{(0.3)},R_{IN}^{(0.3)}}$ between the noise images of $R_{sLD}^{(0.3)}$ and $R_{IN}^{(0.3)}$ was calculated across all pixels in the phantom and across 10 realizations and was found to be near zero, which supports the independence we desire. As expected when $d < 0.5$, the noise magnitude in $R_{IN}^{(d)}$ is lower than that in $R_{sLD}^{(d)}$ since more noise is added into $R_{sLD}^{(d)}$ than is added to $R_{IN}^{(d)}$ according to the inverse scaling of the added quantum and electronic noise.

For other dose levels, the processing can be easily repeated, which forms the curves in Fig. 4. The horizontal axis denotes the relative dose levels from 5% to 95% of normal dose. The vertical axis is the noise in image domain. The blue squares are the real low-dose images $R_{LD}^{(d)}$ at corresponding dose levels. The red dots are noise levels of synthetic low-dose images $R_{sLD}^{(d)}$ from the normal dose image R_{ND}. Again, they fit the blue squares very well at all dose levels. The orange dots are noise levels of images $R_{IN}^{(d)}$ with independent noise from the synthetic low-dose images $R_{sLD}^{(d)}$. At lower dose, the independent $R_{IN}^{(d)}$ image tends to have lower noise level, showing different noise behaviors to real or synthetic low-dose images.

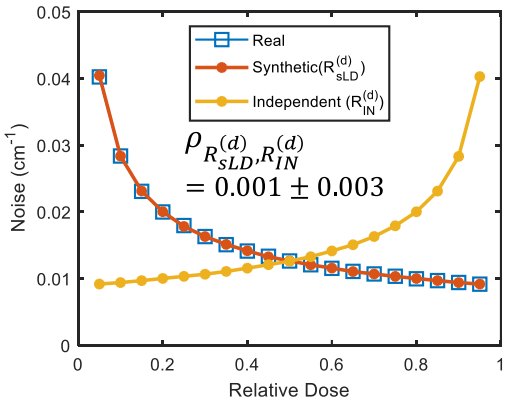


Fig. 4. Noise at different dose levels. The average correlation between images $R_{sLD}^{(d)}$ and $R_{IN}^{(d)}$ across all dose levels is $\rho_{R_{sLD}^{(d)},R_{IN}^{(d)}} = 0.001$.

In general, the independent noise image $R_{IN}^{(d)}$ does not correspond to a specific dose level, even though the noise levels appear approximately symmetric to that of the low-dose images about $d = 0.5$. For example, the noise level in $R_{IN}^{(d)}$ at 80% dose level is generally not equal to that in $R_{sLD}^{(d)}$ at 20%. However, for the special case of no electronic noise ($\sigma_e = 0$), it can be shown that this is the case, and the $R_{IN}^{(d)}$ image represents a dose level of $1 - d$.

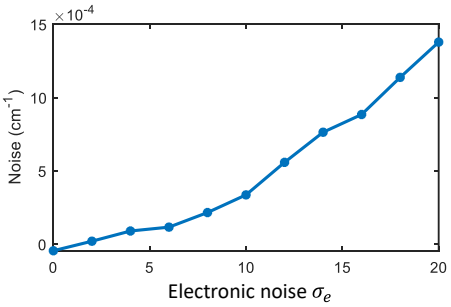


Fig. 5. Noise level difference between $R_{sLD}^{(0.5)}$ and $R_{IN}^{(0.5)}$ at 50% dose level

We demonstrate this assertion with a simple test. For the 50% dose level, we plot the difference in noise between the $R_{sLD}^{(0.5)}$ and $R_{IN}^{(0.5)}$ images (Fig. 5). When there is no electronic noise ($\sigma_e = 0$), the noise levels are indeed identical, but for $\sigma_e > 0$, more electronic noise is added to the synthetic 50% dose image $R_{sLD}^{(0.5)}$ than the independent noise image $R_{IN}^{(0.5)}$. Therefore, in general we are not splitting dose or creating another low-dose

image. Instead, we have created an additional image with independent noise, which satisfies the Noise2Noise conditions.

In Fig. 6, the correlations between ($R_{sLD}^{(d)}$, $R_{IN}^{(d)}$) and ($R_{sLD}^{(d)}$, R_{ND}) are plotted in blue and red. As expected, the correlations between ($R_{sLD}^{(d)}$, $R_{IN}^{(d)}$) at different dose levels are close to 0. On the contrary, the correlations between ($R_{sLD}^{(d)}$, R_{ND}) increases with higher dose (red curve) since more of the noise in the synthetic low-dose image comes from the original normal-dose image. At lower dose, the increased amount of inserted noise leads to lower correlations with the original image.

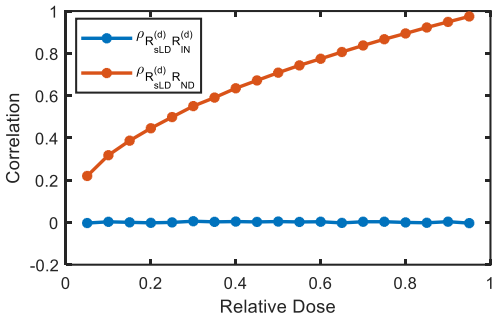


Fig. 6. Correlation at different relative dose levels

We also list the correlation coefficients under different electronic noise (σ_e) levels in Table 2. The correlation between synthetic image $R_{sLD}^{(d)}$ and independent noise image $R_{IN}^{(d)}$ is generally near zero, which agrees with the theoretical analysis in equation (7).

TABLE 2. CORRELATION COEFFICIENTS

σ_e	$\rho_{R_{sLD}^{(d)},R_{IN}^{(d)}}$	
	Mean	STD
0	0.0003	0.0023
5	0.0011	0.0022
10	0.0014	0.0023
20	0.0008	0.0028

IV. DISCUSSION AND CONCLUSION

In this paper, a simulation tool was demonstrated for simultaneously synthesizing low-dose images at arbitrary dose level and independent noisy images. The method extends the conventional noise insertion procedure and creates a byproduct image with independent noise along with the low-dose image at a specific dose level. Correlation between the synthetic and independent noise images was investigated both analytically and numerically, which verified that they are uncorrelated. Thus, they are independent under the assumption that they form a bivariate normal distribution.

For now, we only carried out preliminary validation with a simple simulation in MATLAB. Future work will extend these concepts to a more accurate forward projection model with polychromatic spectrum and non-ideal detector response (energy integrating or photon counting). Also, we are using a linear FBP reconstruction algorithm so that projection domain analysis can be transferred directly to the image domain (although this does include a non-linear log step). Iterative reconstruction methods may violate our linearity assumptions in the image domain, even if the projection domain noise properties hold. Another challenge might be the accuracy of our noise models in severely attenuated areas with photon starvation, such as behind metal. Lastly, we plan to demonstrate the utility of our independent noise simulation on CT denoising networks by fully leveraging the Noise2Noise principle. Our belief is that training with a wide range of simulated dose levels paired with independent noise will outperform other training methods like Half2Half or pairing simulated low dose images with normal dose images.

APPENDIX

In this section, we prove that ($P_{sLD}^{(d)}$, $P_{IN}^{(d)}$) are uncorrelated ($E[P_{sLD}^{(d)}P_{IN}^{(d)}] = E[P_{sLD}^{(d)}]E[P_{IN}^{(d)}]$). Given the definitions of $P_{sLD}^{(d)}$ and $P_{IN}^{(d)}$, it is straightforward to show

$$E[P_{sLD}^{(d)}]E[P_{IN}^{(d)}] = (d\lambda)(d\lambda) = d^2\lambda^2. \tag{8}$$

On the other side:

$$\begin{aligned} E\left[P_{sLD}^{(d)}P_{IN}^{(d)}\right] &= d^2E[P_{ND}^2 + P_{ND}\cdot a(Q+bE) - P_{ND}\cdot\frac{1}{a}\left(Q+\frac{1}{b}E\right) \\ &\quad -(Q+bE)\left(Q+\frac{1}{b}E\right)] \\ &= d^2\left(E\left[P_{ND}^2\right] - E\left[Q^2\right] - E\left[E^2\right]\right) \\ &= d^2\left(\mathrm{Var}(P_{ND}) + E^2\left[P_{ND}\right] - E\left[Q^2\right] - E\left[E^2\right]\right) \\ &= d^2\left(\lambda + \sigma_e^2 + \lambda^2 - \lambda - \sigma_e^2\right) \\ &= d^2\lambda^2 = E\left[P_{sLD}^{(d)}\right]E\left[P_{IN}^{(d)}\right] \end{aligned} \tag{9}$$

where the independence of added noise Q and E from the measured data P_{ND} gives us $E[P_{ND}Q] = 0, E[P_{ND}E] = 0$.

REFERENCES

[1] H. Chen *et al.*, "Low-dose CT with a residual encoder-decoder convolutional neural network," *IEEE transactions on medical imaging*, vol. 36, no. 12, pp. 2524–2535, 2017.

[2] E. Kang, J. Min, and J. C. Ye, "A deep convolutional neural network using directional wavelets for low-dose X-ray CT reconstruction," *Medical physics*, vol. 44, no. 10, pp. e360–e375, 2017.

[3] I. Y. Chun, Z. Huang, H. Lim, and J. Fessler, "Momentum-Net: Fast and convergent iterative neural network for inverse problems," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 2020.

[4] D. Wu, K. Kim, G. El Fakhri, and Q. Li, "Iterative low-dose CT reconstruction with priors trained by artificial neural network," *IEEE transactions on medical imaging*, vol. 36, no. 12, pp. 2479–2486, 2017.

[5] J. Lehtinen *et al.*, "Noise2noise: Learning image restoration without clean data," *arXiv preprint arXiv:1803.04189*, 2018.

[6] D. Wu, K. Gong, K. Kim, X. Li, and Q. Li, "Consensus neural network for medical imaging denoising with only noisy training samples," in *International Conference on Medical Image Computing and Computer-Assisted Intervention*, 2019, pp. 741–749.

[7] D. Wu, K. Kim, and Q. Li, "Low-dose CT reconstruction with Noise2Noise network and testing-time fine-tuning," *Medical Physics*, vol. 48, no. 12, pp. 7657–7672, 2021.

[8] N. Yuan, J. Zhou, and J. Qi, "Half2Half: deep neural network based CT image denoising without independent reference data," *Physics in Medicine & Biology*, vol. 65, no. 21, p. 215020, 2020.

Dark-Field Imaging on a Clinical CT System: Sample Data Processing and Reconstruction

Jakob Haeusele, Clemens Schmid, Manuel Viermetz, Nikolai Gustschin, Tobias Lasser, Frank Bergner, Thomas Koehler, Franz Pfeiffer

***Abstract*—Grating-based phase-contrast and dark-field X-ray imaging is a promising technology for improving the diagnosis and imaging capabilities of breast cancer and lung diseases. While traditional X-ray techniques only consider the attenuation coefficient, phase-contrast and dark-field imaging are also capable of measuring the refractive index decrement and the so-called linear diffusion coefficient, a measure of a sample's small-angle scattering strength. Consequently, the technique provides additional information about the micro-structure of a sample. While it is already possible to perform human chest dark-field radiography, it is assumed that its diagnostic value increases when performed in a tomographic setup. The thereby acquired three-dimensional mappings of the three modalities yield detailed information about morphological changes without being obscured by overlaying structures.**

This work presents the sample data processing and reconstruction pipeline of the first human-sized clinical dark-field CT system. In this novel setting we require a processing concept which is (1) compatible with continuous rotation, (2) can compensate for perturbances induced by system vibrations, and (3) still enables short processing and reconstruction times. An advanced sliding window approach was chosen for the sample data extraction to meet requirements (1) and (3). Furthermore, we present the corrective measures that have to be applied in the employed processing and reconstruction algorithms to mitigate the effects of vibrations and deformations of the interferometer gratings. The developed techniques are shown to successfully reduce the emergence of artefacts in the reconstructed images.

***Index Terms*—x-rays, computed tomography, image reconstruction, signal analysis, parameter estimation**

I. Introduction

X-RAY imaging is an invaluable technique in medical imaging that enables fast measurements of a sample's attenuation coefficient. However, it is limited to attenuation based contrast and cannot exploit additional information of the X-ray wavefront like its phase and small-angle scattering. Grating-based X-ray dark-field imaging is capable of measuring all three effects by introducing a Talbot-Lau interferometer in the beam path [1], [2].

Due to the added diffraction gratings the X-ray wavefront is modulated to create a reference pattern on the detector. As the sample interacts with the incident wavefront the observed pattern is distorted: Porous materials induce small-angle scattering that smears out the observed interference pattern. A proper analysis of this pattern thus yields additional information about the sample's microstructure.

The so-called dark-field signal measured by this procedure has proven a promising new tool in the diagnosis of lung diseases as it can show micro-structural changes in the lung parenchyma [3], [4].

As a first step in transferring this technology to clinical usage a prototype human dark-field scanner is already in use in the university hospital München rechts der Isar. In a first clinical study it could show the potential of dark-field imaging for the diagnosis of COPD and COVID-19 [5].

However, this system is only capable of measuring chest radiographs. For an unobstructed 3D-view of a patients lung a tomographic setup is needed, as demonstrated for living mice [6]. Therefore, in a next step we installed a Talbot-Lau interferometer in a clinical CT. The design of this first human sized dark-field CT prototype is presented in [7].

Previous lab-based dark-field CT systems were designed to meet the stability requirements of the interferometric method. In contrast, our goal was to fit a Talbot-Lau interferometer into an existing clinical CT system with minimal hardware changes. This implies, in particular, a higher level of vibrations than desired, which we address by more complex data processing. Furthermore, the continuous rotation of the gantry prohibits the use of conventional phase demodulation techniques since the detector is positioned differently in each shot.

This work presents the second part of the data processing pipeline of the setup – the processing of sample scans covering the detectors full field of view. The prerequisite reference processing of air scans is discussed in [8]. A schematic overview of the full processing pipeline is depicted in Fig. 1.

II. Model

The basic forward model of grating-based phase-contrast and dark-field X-ray imaging is given by [1]

$$y = T \cdot I \cdot \left[1 + D \cdot V \cdot \cos\left(\Phi + \phi\right)\right], \tag{1}$$

where the state of the interferometer is described by the parameters I, V and ϕ which denote the mean intensity,

J. Haeusele, C. Schmid, M. Viermetz, N. Gustschin, and F. Pfeiffer are with the Chair of Biomedical Physics, Department of Physics, Technical University of Munich, 85748 Garching, Germany. (email: jakob.haeusele@tum.de)

J. Haeusele, C. Schmid, M. Viermetz, N. Gustschin, T. Lasser, and F. Pfeiffer are with the Munich Institute of Biomedical Engineering, Technical University of Munich, 85748 Garching, Germany.

T. Lasser is with Computational Imaging and Inverse Problems, Department of Informatics, Technical University of Munich, 85748 Garching, Germany.

F. Bergner and T. Koehler are with Philips Research, 22335 Hamburg, Germany.

T. Koehler and F. Pfeiffer are with the Institute for Advanced Study, Technical University of Munich, 85748 Garching, Germany.

F. Pfeiffer is with the Department of Diagnostic and Interventional Radiology, School of Medicine and Klinikum rechts der Isar, Technical University of Munich, 81675 München, Germany

the fringe amplitude (visibility), and the interferometer phase, respectively. Once a sample is introduced it attenuates, small-angle scatters, and phase-shifts the incident wave which alters the measured fringe pattern. These changes can be modeled by three sample parameters: the sample transmission T, the dark-field D, and the (differential) phase Φ.

For a data acquisition scheme utilizing a continuously rotating gantry and in the presence of vibrations and other fluctuations, all of the introduced parameters generally depend on the individual detector pixel x and time t.

For the presented dark-field CT prototype the effects of the observed fluctuations are described by linear combinations of per-pixel correction arrays R that describe the impact of the main modes of grating deformations on the measured fringe pattern and their respective per-shot strength coefficients c. These perturbations alter the flatfields I_0, V_0, and ϕ_0 of the three interferometer parameters as follows:

$$I(x,t) = I_0(x) \cdot \left(1 + \sum_{n=1}^{N_i} c_{i,n}(t) R_{i,n}(x)\right) \tag{2}$$

$$V(x,t) = V_0(x) \cdot \left(1 + \sum_{n=1}^{N_v} c_{v,n}(t) R_{v,n}(x)\right) \tag{3}$$

$$\phi(x,t) = \phi_0(x) + \sum_{n=1}^{N_p} c_{p,n}(t) R_{p,n}(x). \tag{4}$$

While the flatfields I_0, V_0, and ϕ_0 and the correction arrays R_i, R_v, and R_p are determined during the reference processing of an air scan (see Fig. 1a, b), the fluctuation strength coefficients of the sample scan c_i, c_v, and c_p have to be determined during sample processing. Subsequently, the sample parameters T, D and Φ can be extracted.

III. Fluctuation Estimation

The observed fluctuations described by c_i, c_v, and c_p are caused by periodic processes happening during the measurement. These include, most importantly, the gantry rotation, the change of the X-ray focal spot due to induced magnetic fields, and vibrations caused by motors on the gantry. While the effects of the gantry movement lead to a reproducible low frequency change in the coefficients, the other effects result in high frequency oscillations. It is therefore possible to extract the low frequency components of the coefficients $\mathrm{LF}\left[c_n^{\mathrm{ref}}(t)\right]$ from the reference scan by low-pass filtering the reference coefficients. Only high frequency effects are estimated separately during sample processing.

By analyzing multiple reference scans it was found that the high frequency components of the fluctuation coefficients can be approximated well by a sum of a few sinusoidals. These oscillate at characteristic frequencies of the setup which can be assigned to different components on the gantry. The most important frequencies are two dominant frequencies from the X-ray tube asynchronous motor which drives the anode rotation: the stator frequency is 189 Hz and the rotor frequency approximately 176 Hz. Moreover, vibrations induced by the cooling unit at 117 Hz are included in our model.

It was further found that the amplitudes of these oscillations stay constant from scan to scan and can therefore also be extracted from the respective reference coefficient. Additionally, the phase relationship of the oscillations between different coefficients stays constant. These two facts can be illustrated nicely by using pairs of coefficients to plot Lissajous curves which can then be compared between scans. An example is given in Fig. 1f: Here it can be seen that all values of the intensity fluctuation coefficients lie on a well defined Lissajous ellipse. This ellipse stays constant between scans, which indicates a constant amplitude and phase relation of the coefficients oscillations.

Based on this analysis each coefficient c_n is expressed as

$$c_n(t) = \sum_{k=1}^{N_f} A_{n,k}^{\mathrm{ref}} \sin(2\pi f_k t + \phi_{n,k}^{\mathrm{ref}} + \delta_k) + \mathrm{LF}\left[c_n^{\mathrm{ref}}(t)\right], \tag{5}$$

where $A_{n,k}^{\mathrm{ref}}$ and $\phi_{n,k}^{\mathrm{ref}}$ are the reference amplitude and oscillation phase of the respective reference coefficient c_n^{ref} and frequency f_k which were retrieved during processing of an air scan. Employing this model has the advantage that only the global phase offsets of each frequency δ_k have to be found for the sample scan. They are extracted via Fourier-analysis and subsequent fitting of initial guesses of the total flux and interferometer phase. These are acquired by conventional processing algorithms. Then, all coefficients can be synthesized from parameters extracted during the reference processing.

IV. Signal Extraction and Reconstruction

Once the fluctuation strengths have been estimated, per-shot references can be generated by applying equations (2), (3), and (4). Next, the measured interference pattern has to be demodulated to retrieve the three sample channels T, D, and Φ. This is achieved by first applying a change of variables to equation (1) which is defined by

$$A = T, \quad B = TD\cos(\Phi), \quad C = TD\sin(\Phi) \tag{6}$$

$$M_A = I, \quad M_B = IV\cos(\phi), \quad M_C = -IV\sin(\phi). \tag{7}$$

The forward model is now linear in the new set of sample parameters A, B, and C:

$$y = AM_A + BM_B + CM_C. \tag{8}$$

Demodulation is performed by defining patches of pixels on the detector and consecutive shots in the measured sinogram y_{meas}. For each of these patches a least squares optimization is performed to minimize a cost function $\mathcal{C}_{\mathrm{patch}}$ and find the sample parameters. The cost function is defined as

$$\mathcal{C}_{\mathrm{patch}} = \sum_{x,t \in \mathrm{patch}} \frac{\left[y_{\mathrm{meas}}(x,t) - y_{\mathrm{patch}}(x,t)\right]^2}{w(x,t)}, \tag{9}$$

with weights w that can be used to introduce a distance weighting in the patch to weight pixels and shots less towards the borders of the patch.

To account for a continuously rotating gantry and spatial changes of the sample parameters, in each patch they are

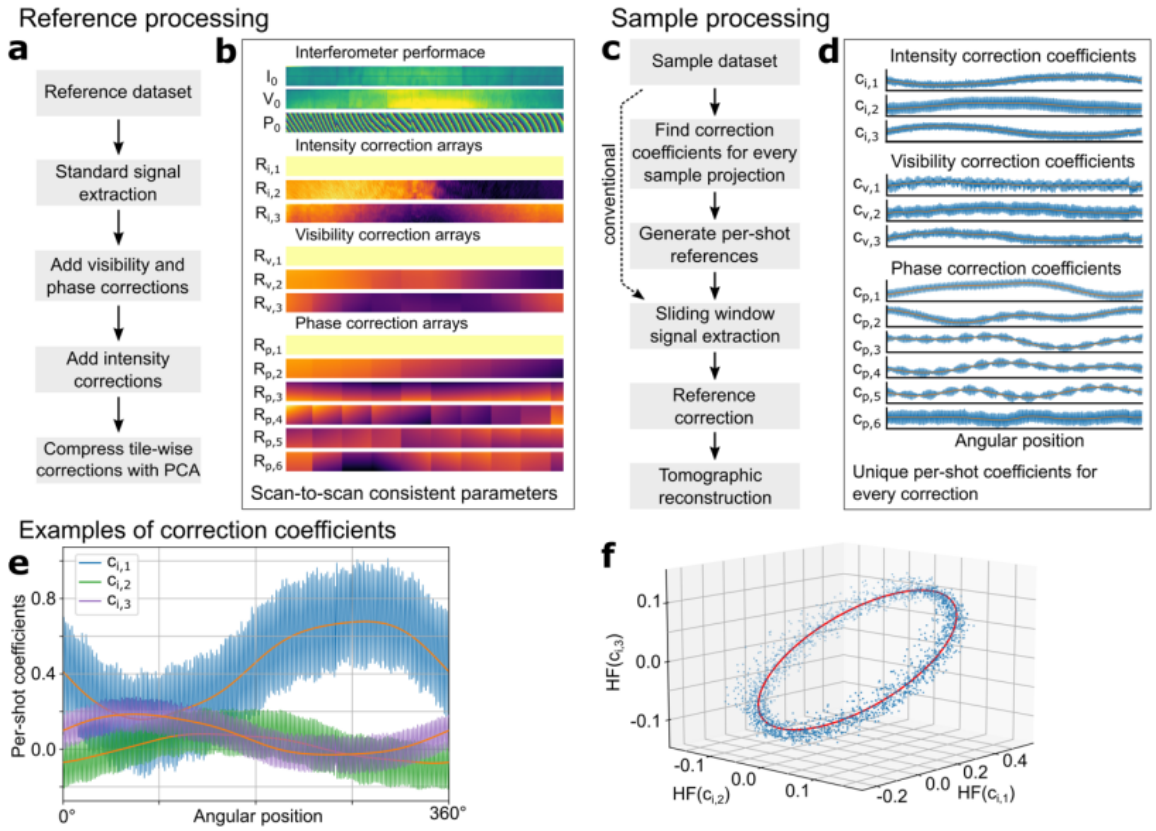


Fig. 1. Processing pipelines for reference and sample scans. **a,** Reference processing pipeline to extract scan-to-scan persistent system characteristics from an air scan. Here we introduce local intensity fluctuation corrections and information compression using principal component analysis (PCA). **b,** Interferometer performance and correction array results from reference processing. **c,** Sample processing pipeline, based on sliding window processing. As the high frequency oscillations differ from the reference scan, we use an optimization step to identify the optimal linear combination of the correction arrays to estimate the current sample free fringe parameters and thus to suppress vibration artefacts. **d,** Results of the correction coefficient optimization using prior knowledge from the reference scan. **e,** Coefficients of the three intensity correction arrays. The angular position dependent drift (orange) is scan-to-scan consistent and is utilized as prior knowledge in the correction optimization step. **f,** Three-dimensional scatter plot of the three intensity correction coefficients after subtraction of the low-frequency component shown in e. A correlation can be observed which also is used during the optimization step as prior knowledge. Figure and caption adapted from [7].

represented as linear combinations of known basis functions and their a priori unknown coefficients:

$$A = \sum_m a_m \alpha_m(x,t), \;\; B = \sum_n b_n \beta_n(x,t), \;\; C = \sum_k g_k \gamma_k(x,t) \tag{10}$$

For simplicity the basis functions α, β, and γ were chosen to be polynomials of x and t. Consequently, the model function y_{patch} assumed in each patch is given by:

$$\begin{aligned} y_{\text{patch}}(x,t) = &\sum_m a_m \alpha_m(x,t) M_A(x,t) + \\ &+ \sum_n b_n \beta_n(x,t) M_B(x,t) + \\ &+ \sum_k g_k \gamma_k(x,t) M_C(x,t). \end{aligned} \tag{11}$$

This model is linear in the basis function coefficients. Therefore, the least squares optimization is a linear regression and possesses an analytical solution that yields the optimal coefficients:

$$a_m^{\text{opt}}, b_n^{\text{opt}}, g_k^{\text{opt}} = \underset{a_m, b_n, g_k}{\operatorname{argmin}} \mathcal{C}_{\text{patch}}. \tag{12}$$

The full demodulation of the measured sinogram is achieved by iterating over all pixels and at each step processing the patch with the current pixel positioned in its center. The sample parameters at the current position are then given by evaluating equations (10) at the center of the patch using the least squares optimization results.

After performing this patch-wise sliding window phase retrieval the variables are transformed back using:

$$T = A, \quad D = \frac{\sqrt{B^2 + C^2}}{A}, \quad \Phi = \arctan2(B, C). \tag{13}$$

This yields the three sinograms $T(x,t)$, $D(x,t)$ and $\Phi(x,t)$. After transforming T and D into line integrals by applying the negative logarithm function they can be reconstructed using a filtered backprojection. For the dark-field a weighting has to be performed during reconstruction to account for sensitivity

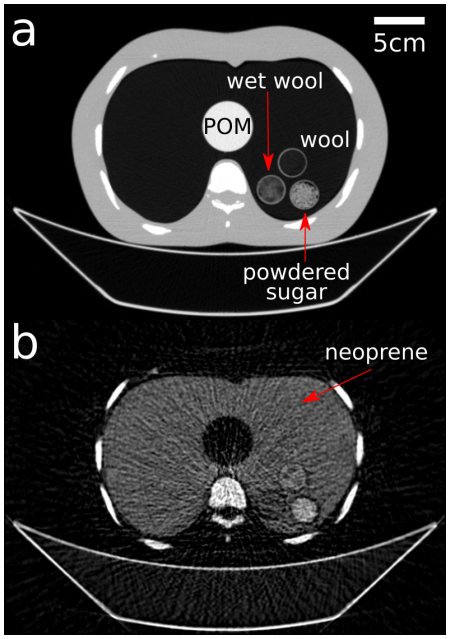


Fig. 2. Processing results of a measurement of a human thorax phantom. Sub-figure **a** shows the reconstructed attenuation channel (-1000, 160) [HU] while sub-figure **b** depicts the dark-field. It can be observed that dry wool, powdered sugar, and the large porous neoprene insert simulating the lung show a strong dark-field signal, while the dense POM cylinder vanishes fully.

differences depending on the sample's position within the interferometer [9]. Lastly, post processing steps are applied to enhance the image quality and correct for dark-field beam hardening effects.

V. Results

The results of applying the processing pipeline to a measurement of a human thorax phantom are depicted in Fig. 2. The phantom was filled with a neoprene insert to simulate the scattering behavior of lung tissue. Additionally, three Falcon tubes filled with dry wool, wet wool, and powdered sugar as well as a POM cylinder were placed inside the neoprene insert.

The attenuation channel reconstruction appears artefact free. In the dark-field, there are minor streak artefacts, and noise is corrupting the image quality slightly. However, this could be mitigated by employing further post-processing and filter algorithms.

It can be observed that the neoprene insert, dry wool, powdered sugar and bones, show a strong dark-field signal, while the POM cylinder and the soft tissue surrounding the ribs vanish fully in the dark-field reconstruction. This demonstrates the system's capability to differentiate between porous and dense materials. Moreover, it shows the possibility of analyzing the scattering strengths of the different materials quantitatively.

VI. Conclusion

This work presented the sample processing pipeline of the first clinical dark-field CT scanner. The two main processing challenges in translating grating-based dark-field imaging to a clinical CT system are addressed and solutions are proposed.

The first challenge lies in describing the interferometer state accurately for each shot in spite of vibrations. This was addressed by analyzing and adapting measurements of an empty reference scan to the sample measurement.

The second challenge lies in the demodulation of the signal. A continuous tomographic data acquisition shows the sample under a different angle for each projection. This leads to movement artefacts when trying to demodulate the signal with a conventional sliding window approach. To mitigate the effects of sample movement, an advanced patch-wise sliding window algorithm is presented that estimates the sample movement locally using continuous basis functions. Since the proposed model leads to a linear optimization problem, demodulation can be performed fast, while only minor artefacts related to noise are present in the final reconstruction.

Acknowledgment

The authors wish to thank Julia Herzen, Maximilian von Teuffenbach, and Amanda Pleier for their help and support. This work was carried out with the support of the Karlsruhe Nano Micro Facility (KNMF, www.kit.edu/knmf), a Helmholtz Research Infrastructure at Karlsruhe Institute of Technology (KIT). We acknowledge the support of the TUM Institute for Advanced Study, funded by the German Excellence Initiative, the European Research Council (ERC, H2020, AdG 695045) and Philips GmbH Market DACH.

References

[1] F. Pfeiffer *et al.*, "Phase retrieval and differential phase-contrast imaging with low-brilliance x-ray sources," *Nature Physics*, pp. 258–261, 2006.

[2] T. Weitkamp *et al.*, "Tomography with grating interferometers at low-brilliance sources," International Society for Optics and Photonics, SPIE, 2006, pp. 249 –258.

[3] K. Hellbach *et al.*, "In vivo dark-field radiography for early diagnosis and staging of pulmonary emphysema," *Invest Radiol*, 2015.

[4] A. Yaroshenko *et al.*, "Improved in vivo assessment of pulmonary fibrosis in mice using x-ray dark-field radiography," *Scientific reports*, p. 17 492, 2015.

[5] K. Willer *et al.*, "X-ray dark-field chest imaging for detection and quantification of emphysema in patients with chronic obstructive pulmonary disease: A diagnostic accuracy study," *The Lancet Digital Health*, e733–e744, 2021.

[6] A. Velroyen *et al.*, "Grating-based x-ray dark-field computed tomography of living mice," *EBioMedicine*, pp. 1500–1506, 2015.

[7] M. Viermetz *et al.*, "Darkfield computed tomography reaches the human scale," *PNAS*, in press.

[8] C. Schmid *et al.*, "Modeling vibrations of a Talbot-Lau interferometer on a continuously rotating clinical CT gantry," submitted.

[9] U. van Stevendaal *et al.*, "Reconstruction method incorporating the object-position dependence of visibility loss in dark-field imaging," International Society for Optics and Photonics, SPIE, 2013, pp. 260 –270.

S^{2}MS: Self-Supervised Learning Driven Multi-Spectral CT Image Enhancement

Chaoyang Zhang, Shaojie Chang, Ti Bai, and Xi Chen

***Abstract*—Photon counting spectral CT (PCCT) can produce reconstructed attenuation maps in different energy channels, reflecting energy properties of the scanned object. Due to the limited photon numbers and the non-ideal detector response of each energy channel, the reconstructed images usually contain much noise. With the development of Deep Learning (DL) technique, different kinds of DL-based models have been proposed for noise reduction. However, most of the models require clean data set as the training labels, which are not always available in medical imaging field. Inspiring by the similarities of each channel's reconstructed image, we proposed a self-supervised learning based PCCT image enhancement framework via multi-spectral channels (S^{2}MS). In S^{2}MS framework, both the input and output labels are noisy images. Specifically, one single channel image was used as output while images of other single channels and channel-sum image were used as input to train the network, which can fully use the spectral data information without extra cost. The simulation results based on the AAPM Low-dose CT Challenge database showed that the proposed S^{2}MS model can suppress the noise and preserve details more effectively in comparison with the traditional DL models, which has potential to improve the image quality of PCCT in clinical applications.**

***Index Terms*—Spectral CT, denoising, Noise2Noise, deep learning**

I. Introduction

PHOTON counting spectral CT (PCCT) can separately collect the incident photons in different energy bins, which has high energy resolution and can generate more accurate material decomposition [1], [2]. Nevertheless, with the increase number of energy bins, counting rate is limited in each individual channel, which results in a relatively low signal-to-noise ratio (SNR). Moreover, there are complicated noises caused by non-ideal response of detector, such as fluorescence x-ray effects, K-escape, charging sharing, and pulse pileups [3]. Noise in the reconstructed CT images will seriously affect diagnosis of doctors.

To reduce noise in CT images, recent deep learning (DL) technique has been widely developed in the field of CT image denoising and shows the potential in applications. Yang et al. used the generative adversarial network (GAN) with Wasserstein distance and perceptual similarity to reduce the noise in CT images [4]. Lv et al. proposed an PCCT image denoising method via fully convolutional pyramid residual network, which suppresses noise in each single energy channel image [5].

However, traditional DL methods require high-quality clean images as training labels to achieve high performance, which are difficult to obtain especially in medical imaging field. To solve this problem, Lehtinen et al. introduced the Noise2Noise model (N2N), where the network was trained to map one noisy realization to another noisy realization [6]. In addition, photon counting spectral CT (PCCT) provides an opportunity to produce reconstructed attenuation maps in different energy channels, which reflect energy properties of the scanned object. Using the similarity of images in different energy bins, we proposed a self-supervised learning driven PCCT image denoising method via multi-spectral channels based on the N2N network model (S^{2}MS). In our framework, the input are images from multi-channels and a channel-sum image, while the output is image of one single channel. Both input and output are noisy PCCT images. Compared with N2N, our proposed S^{2}MS fully used all the reconstructed images in different energy bins at the same time, rather than processing images in each channel separately. All the simulated experiments were carried out and the results show the S^{2}MC model is effective and accurate in noise reduction and detail preservation.

II. Materials and Methods

A. Basic Principle of PCCT

Compared with the traditional energy integration detector, photon counting detector (PCD) can separately count out the number of photons in each energy channel, which can be used to obtain the projection in different energy bins. Fig. 1 shows an example of PCCT images in four energy channels. The similarity between images in different energy bins can be used for noise reduction.

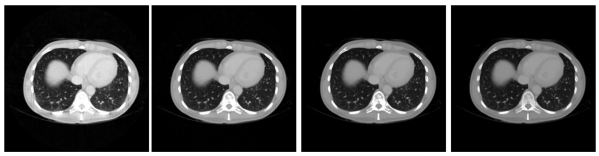


Fig. 1. An example of reconstructed PCCT images in different energy bins. The display window is [0,0.4] cm^{-1}.

C. Zhang and X. Chen are with the School of Information and Communication Engineering, Xi'an Jiaotong University, Xi'an, Shanxi 710049, China (e-mail: xi_chen@mail.xjtu.edu.cn)

S. Chang is with the Department of Radiology, Stony Brook University, Stony Brook, NY 11794, USA (e-mail: shaojiechang01@gmail.com)

T. Bai is with the Department of Radiation Oncology, University of Texas Southwestern Medical Centre, Dallas, TX 75390, USA

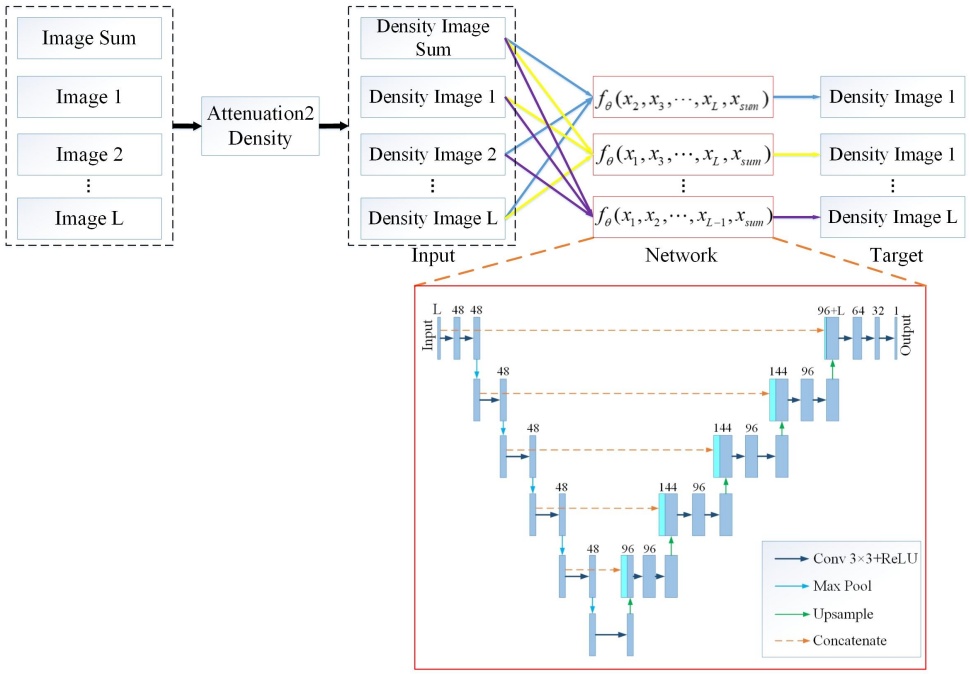


Fig. 2. The noise2noise network-based PCCT image denoising framework with self-supervised learning via multi-spectral channels (S²MS). The attenuation images were divided by the mass attenuation coefficient, converted into density images.

B. Deep-Learning based Denoising

In deep learning based denoising method, the input of the network is generally regarded as the following:

$$x_i = y_i + n_i \tag{1}$$

where x_i is the corrupted input, y_i is the clean target and n_i denotes the corresponding noise. According to the type of training target, DL-based denoising can be divided into the following two types:

1) Supervised Learning-N2C

Traditionally, supervised learning is always used for deep-learning based CT image denosing, which means training a regression model with pairs $(\widehat{x_i}, y_i)$ and minimizing the function:

$$\theta^* = \arg\min_{\theta} \frac{1}{N} \sum_i \left\| f_\theta(\widehat{x_i}) - y_i \right\| \tag{2}$$

where $f_\theta(\widehat{x_i})$ is the denoising convolutional neural network (CNN), θ is weight, N is the total number of training samples. In this paper, the image denoising based on supervised learning is referred as Noise2Clean (N2C).

2) Self-Supervised Learning-N2N

Opposite to Noise2Clean, Noise2Noise (N2N) is a self-supervised learning framework where input and target are both corrupted. It can be expressed as:

$$\theta^* = \arg\min_{\theta} \frac{1}{N} \sum_i \left\| f_\theta(y_i + n_{i1}) - (y_i + n_{i2}) \right\| \tag{3}$$

where n_{i1} and n_{i2} are two independently noise realizations. It has been assumed that the Noise2Noise training is equivalent to Noise2Clean training under certain mild conditions [6], [7]:

1. $N \to \infty$;
2. Conditional expection $E\{n_{i2} \mid y_i\} = 0$;
3. n_{i1} and n_{i2} are independent;
4. $\forall i,\ \left| f_\theta(y_i + n_i) \right| < \infty$.

Since filter in the convolutional neural network is shift-invariant, different parts of the image may be served as multiple training samples. Even if the size of training data is small, the actual number of training samples is large enough to satisfy condition 1. After reconstructing, the noise in image domain is zero-mean and independent in different energy channels [8], which means condition 2 and condition 3 are both satisfied in our method. Condition 4 can be easily satisfied by choosing the meaningful parameters of the network.

C. Noise2Noise Network for PCCT Image Denoising

By using the similarities of reconstructed images in different energy bins, we proposed a noise2noise network-based PCCT image denoising framework based on self-supervised learning via multi-spectral channels (S²MS). The basic process of our framework is shown in Fig. 2.

In S²MS, L-1 reconstructed images in single channel and a channel-sum image (linear attenuation map) were divided by the mass attenuation coefficient of each channel (Attenuation 2Density), converted into density images before training, which were used as the input of L channels. Then, the left single channel image was also converted into density image as the target. The S²MS network can be described as:

$$\theta^* = \arg\min_{\theta} \frac{1}{N} \sum_i \left\| f_\theta(y_{i1} + n_{i1}, \cdots, y_{i(E-1)} + n_{i(E-1)}, y_{isum} + n_{isum}) - (y_{iE} + n_{iE}) \right\| \tag{4}$$

where $f_\theta(\widehat{x_i})$ is the denoising convolutional neural network (CNN) with L inputs, y_{isum} is the clean channel-sum image, n_{isum} is the noise in channel-sum image, $y_i = \{y_{i1}, y_{i2}, \cdots, y_{iL}\}$ are the clean reconstructed images in different single energy channel and $n_i = \{n_{i1}, n_{i2}, \cdots, n_{iL}\}$ are the corresponding noise.

After training, S^{2}MS can denoise the PCCT image in single channel. The denoised density image was multiplied by mass attenuation coefficient (Density2Attenuation), converted into a linear attenuation image (Fig. 3).

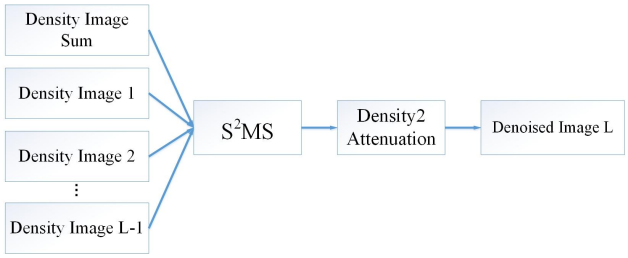


Fig. 3. The denoised process of the trained S^{2}MS network.

D. Experiment Setup

1) Dataset Establishment

In this study, CT images from the 2016 Low-dose CT Grand Challenge dataset [9] were used to simulate the PCCT images. 1000 slices of 7 patients were randomly divided into training dataset, validation dataset and test dataset according to the ratio of 8:1:1. An equal spatial fan-beam geometry was assumed to simulate the projecting process. The distance from the source to the system origin was 142 cm, the distance from the source to the detector was 180 cm, and there were 512 detector elements with the width of 0.1 cm per element. A total of 512 projections were acquired in an angular range of 360 degrees. The projection data were collected in four different energy bins 30-45keV (channel 1), 45-60keV (channel 2), 60-80keV (channel 3), and 80-100keV (channel 4). In each energy channel, 1000 PCCT images were acquired which had 512*512 pixels. Poisson noise was introduced in the simulation process. Totally 1×10^5 photons emitted along each x-ray path and the number of photons per energy channel was proportional to the normalized spectrum of each channel. Finally, the PCCT images can be reconstructed by FBP algorithm. Before training, the reconstructed images divided by the mass attenuation coefficient to convert into the density images which were used as the input and target of S^{2}MS.

2) Network Implementation

U-Net architecture in [6] was used in our study (Fig. 2). The encoder-decoder network includes a shrinking multi-scale decomposition path and a symmetric expansion path, with skip connections on each layer. Adam optimizer was used with a learning rate of 0.0003. The loss function was designed based on Mean Squared Error (MSE):

$$MSE(x,y) = \frac{1}{mn}\sum_{i=1}^{m}\sum_{j=1}^{n}\left\|x(i,j) - y(i,j)\right\|^2 \qquad (7)$$

where x is input and y is target of the network.

The training was performed on a server with Intel Xeon Silver 4214 CPU and GeForce RTX 3090 24G GPU. The network was coded in Pytorch1.9.1 using Ubuntu20.04.

3) Comparison Study

To evaluate the performance of our proposed method, the Noise2Clean (N2C) and traditional Noise2Noise (N2N) were used for comparison. In N2C network, PCCT image reconstructed from the projections without noise was used as the target. In N2N network, two independent projections were simulated and the corresponding reconstructed images were used as input and output, respectively. Since our study focused on a denoising method rather than the network structure, the U-Net architecture in Fig. 2 was also used for N2C and N2N.

4) Evaluation Metrics

In our study, structure similarity (SSIM) and Root Mean Squared Error (RMSE) were used as the evaluation metrics. SSIM measures the structural similarity by comparing both the mean value and distribution relevance between denoised image and reference, and RMSE measures the L2-norm error between the estimated image and the ground truth.

III. Results

Denoised images generated by our proposed method S^{2}MS, N2C and N2C were shown in Fig. 4. Our proposed method can effectively reduce noise in each PCCT channel image. Especially in channel 3 and channel 4, our proposed method is able to retain richer structural information while suppressing most noise. Four regions of interest (ROIs) were selected (red rectangles) to show the detail preservation performance (Fig.5). In comparison with other methods, the proposed S^{2}MS can remove more noise while preserving more details.

We selected a fixed size (200*200) ROI (blue rectangle in Fig. 4) and calculated the SSIM and RMSE (Table I). Our proposed S^{2}MS achieved the highest SSIM and the lowest RMSE in each energy channel, which indicated that S^{2}MS made a better performance on image denoising for PCCT in comparison with N2N and N2C.

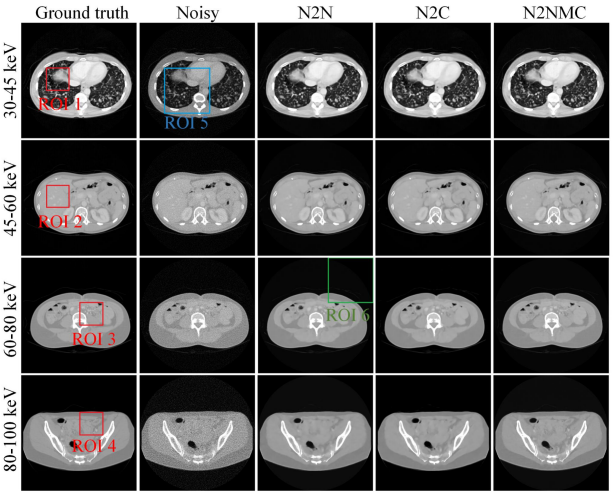


Fig. 4. Four example reconstructed slices in four energy channels (30-45keV, 45-60keV, 60-80keV, 80-100keV). The display windows for linear attenuation from the top to the bottom rows are [0,0.4] cm^{-1}, [0,0.4] cm^{-1}, [0,0.35] cm^{-1}, and [0,0.35] cm^{-1}, respectively.

TABLE I
DENOISING RESULTS OF DIFFERENT METHODS ON TEST DATASET

Energy	Method	SSIM	RMSE
30-45 keV	N2N	0.9754	0.0131
	N2C	0.9765	0.0130
	S²MS	**0.9800**	**0.0117**
45-60 keV	N2N	0.9738	0.0087
	N2C	0.9750	0.0085
	S²MS	**0.9853**	**0.0061**
60-80 keV	N2N	0.9697	0.0072
	N2C	0.9719	0.0070
	S²MS	**0.9825**	**0.0048**
80-100 keV	N2N	0.9442	0.0081
	N2C	0.9479	0.0077
	S²MS	**0.9737**	**0.0046**

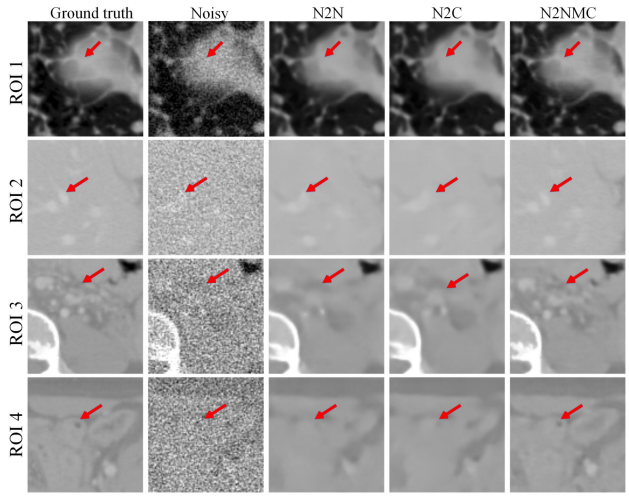


Fig. 5. Details of reconstructed images in Fig.3. The first and second rows are PCCT images in channel 3 (60-80 keV), the third and fourth rows are PCCT images in channel 4 (80-100 keV), and the display window for all images is [0,0.35] cm^{-1}.

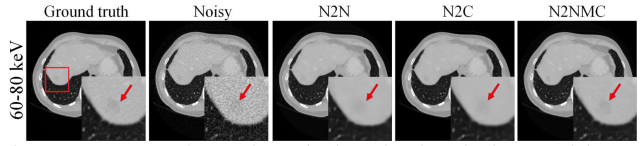


Fig. 6. A reconstructed PCCT image in channel 3, the noisy image and the outputs of different network. The shading in the red rectangle is the region of lesion which is magnified and shown. The display window is [0,0.35] cm^{-1}.

To further validate the performance of the proposed method in PCCT clinical application, CT images with lesion is shown in Fig. 6. Denoised reconstructed images in channel 3 (60-80 keV) of different methods are illustrated and the lesion region (red arrow) is magnified. The lesion is hardly observed in the noisy image while it can be clearly in the reconstructed image by S²MS. The lesion area was blurred in the reconstructed images denoised by other methods. This result indicated that our method has potential in clinical application.

IV. DISCUSSIONS AND CONCLUSION

In particular, in the output of N2N and S²MS, a gray circular shadow can hardly be seen in the air area (Fig. 7). There is no anatomic structure in this area, which makes the noise distribution and background signal in this part are totally different from those in human part. Therefore, the output of the network may get wrong values in non-human regions. The shadow is hard to see and has little influence in diagnosis.

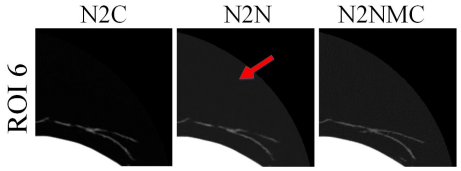


Fig. 7. The magnified image of ROI (green rectangle in Fig. 3). The display window is [0,0.1] cm^{-1}.

In conclusion, we have proposed a Noise2Noise-based PCCT image denoising framework via multi-spectral channels (S²MS). In this study, noisy PCCT images were used as both the input and the output to train the network. To make full use of the spectral data in L channels, the reconstructed images in L-1 single channels and channel-sum image were used as the input and the left single channel image was used as the output. Compared with the traditional DL denoising method, simulation results show that the proposed method can obtain a reconstructed image with high quality: noise is reduced remarkably and detail features is well remained. No clean image needed makes the proposed S²MS has potential in practical application. In the future work, our S²MS will be regarded as a priori information to be combined with the material decomposition framework and the experimental data will be used to test the network.

REFERENCES

[1] S. Leng, L. Yu, J. G. Fletcher, C. A. Mistretta, and C. H. McCollough, "Noise reduction in spectral CT: Reducing dose and breaking the trade-off between image noise and energy bin selection," *Med. Phys*., vol. 38, no. 9, pp. 4946-4957, Sep. 2011.

[2] X. Wu, P. He, Z. Long, X. Guo, M. Chen, X. Ren, P. Chen, L. Deng, K. An, P. Li, B. Wei, and P. Feng, "Multi-material decomposition of spectral CT images via fully convolutional DenseNets," *J. X-Ray Sci. Technol.*, vol. 27, no. 3, pp. 461–471, Jul. 2019.

[3] K. Taguchi, C. Polster, O. Lee, K. Stierstorfer, S. Kappler, "Spatio-energy cross talk in photon counting detectors: Detector model and correlated poisson data generator," *Med. Phys.*, vol. 43, no.12, pp. 6386-6404, Dec. 2016.

[4] Q. Yang, P. Yan, Y. Zhang, H. Yu, Y. Shi, X. Mou, M. K. Kalra, Y. Zhang, L. Sun, and G. Wang, "Low-dose CT image denoising using a generative adversarial network with Wasserstein distance and perceptual loss," *IEEE Trans. Med. Imag.*, vol. 37, no. 6, pp. 1348-1357, Jun. 2018.

[5] X. Lv, X. Ren, P. He, M. Zhou, Z. Long, X. Guo, C. Fan, B. Wei, and P. Feng, "Image denoising and ring artifacts removal for spectral CT via deep neural network," *IEEE Access*, vol. 8, pp. 225594-225601, Dec. 2020.

[6] J. Lehtinen, J. Munkberg, J. Hasselgren, S. Laine, T. Karras, M. Aittala, and T. Aila, "Noise2Noise: Learning image restoration without clean data," in *Proc. Int. Conf. Mach. Learn.*, 2018, pp. 2965-2974.

[7] D. Wu, K. Gong, K. Kim, X. Li, and Q. Li. "Consensus neural network for medical imaging denoising with only noisy training samples," in *Int. Conf. MICCAI*, 2019, pp. 741-749.

[8] W. Fang, D. Wu, K. Kim, M. K Kalra, R. Singh, L. Li, and Q. Li, "Iterative material decomposition for spectral CT using self-supervised Noise2Noise prior," *Phys. Med. Biol.*, vol. 66, no. 15, pp. 5013-5030, Jul. 2021.

[9] C. McCollough. "TU-FG-207A-04: overview of the low dose CT grand challenge," *Med. Phys.*, vol. 43, no.6, pp. 3759-3760, Jun. 2016.

Virtual Non-Metal Network for Metal Artifact Reduction in the Sinogram Domain

Da-in Choi, Taejin Kwon, Jaehong Hwang, Joon Il Hwang, Yeonkyoung Choi and Seungryong Cho

***Abstract*—Often, the artifacts caused by high-density objects degrade the quality of the image with streaks and information loss in CT imaging. In recent years, machine learning has proven itself a powerful tool to resolve some of the challenges faced in reducing metal artifacts. In this work, a novel method of metal artifact reduction (MAR) without metal segmentation by using a CNN network is proposed. The approach focuses on removing the need for the sensitive metal segmentation step to improve robustness and aims to tackle beam hardening directly in the sinogram domain. In the proposed method, we trained the network with sinogram pairs that include metal objects and those that include virtual non-metal (VNM) replacement objects. A VNM object is designed to be less dense than metal but more dense than soft tissue. The novelty of this method lies in the sinogram-to-sinogram training without the need for metal segmentation by replacing the metal object to a virtual non-metal object in the sinogram to reduce beam hardening and successfully compensate for the information loss.**

***Index Terms*—Computed tomography (CT), convolutional neural network (CNN), metal artifact reduction (MAR), deep learning**

I. Introduction

High density materials cause degradation of the CT image quality via factors such as beam-hardening, photon starvation and scatter. Metal artifact is the overarching term referring to the resulting artifacts observed as many streaks, loss of image information, structural deformation and more [1-2]. For decades, researchers devised multitude of methods to tackle metal artifacts. The most common and analytic MAR methods are sinogram interpolation based, such as linear MAR and Normalized MAR (NMAR). These methods aim to replace the metal trace in the sinogram with neighboring information. In doing so, the essential step for these methods is metal segmentation [3-5]. However, for cases with severe beam hardening and photon starvation, segmenting the metal accurately is a great challenge.

The prowess of CNN shines in medical imaging when it comes to segmentation and solving complex problems by finding patterns and training features [6-8]. In recent years, new methods that incorporate CNN for metal segmentation to aid in interpolation have been proposed. For instance, CNN-MAR proposed by Zhang et al. train the network to synthesize an artifact reduced image based on the non-corrected, beam hardening corrected (BHC) and linear interpolated images [9]. The CNN result is used as the prior image in the NMAR process. The DuDoNet proposed by Lin et al. takes advantage of the CNN to enhance the sinogram while retaining geometric consistency and to improve the reconstructed image based on the linear interpolated images [10]. While both methods improved image quality, for metal artifacts of different shapes and sizes, the result image is sensitivity to metal segmentation and to soft tissue smoothing and deformation.

Since metal artifact and prior image generation are sensitive, more approaches that do not require metal segmentation are published. However, most utilize reconstructed images for training. As a result, metal artifacts that are not fully taken care of in the MAR processed and labeled data used for training remain in the test result. Thus, the quality of the image domain training is limited by the MAR preprocessing implemented to the training dataset. To overcome this limitation, Park et al. trained U-net with metal-corrupted sinogram and metal artifact corrected sinogram with metal mask for hip prosthesis replaced by air [11].

In this work, we propose to train a sinogram-to-sinogram CNN to tackle the effect of beam hardening without the need for metal tracing. The common tactic to replace the metal trace via interpolation of the neighboring pixels or to replace the metal object with air is not desirable in that the metal trace not only contain the metal but also soft tissue. We propose to reduce metal artifact by tackling the beam hardening effect in the sinogram region. This method aims to replace the metal sinogram regions with objects with significantly less density than metal. The resulting sinogram, when reconstructed will restore soft tissue information in the shading artifact region and reduce streak artifacts. The feasibility of the proposed method is tested by a simulation study. The image quality of the result is evaluated qualitatively and quantitatively against the conventional NMAR algorithm.

II. Materials and Methods

A. Dataset preparation

Attaining a large set of projection data of clinical volumes was found difficult. Thus, this study was performed with a

This research was supported in part by the Ministry of Science and ICT (MSIT, Korea) & the Ministry of Trade, Industry and Energy (MOTIE, Korea) [Project Number: 20014921].

D. Choi, T. Kwon, J. Hwang, J. Hwang, Y. Choi and S. Cho are with the Medical Imaging and Radiotherapy Lab, Department of Nuclear and Quantum Engineering, Korea Advanced Institute of Science and Technology, Daejeon, South Korea (contact: drchoi@kaist.ac.kr). S. Cho is in correspondence of this paper. (e-mail: scho@kaist.ac.kr).

simulation data set. To obtain a more realistic simulation data, a material map is generated by applying thresholds to the human body CT image without metal implants attained from NIH Clinical Center [12]. The thresholds are set to divide the object into bone, soft tissue, and fat. Then, elliptical objects of random size, number and position were added to the material map. The position of the elliptical objects was limited to the body. The elliptical objects were then given a material value for either titanium for metal and carbon for VNM substitute. Finally, the projection data was generated using a polynomial energy forward projection simulator with the parameters shown in Table 1. An example of the material map and the simulator results are shown in Fig. 1.

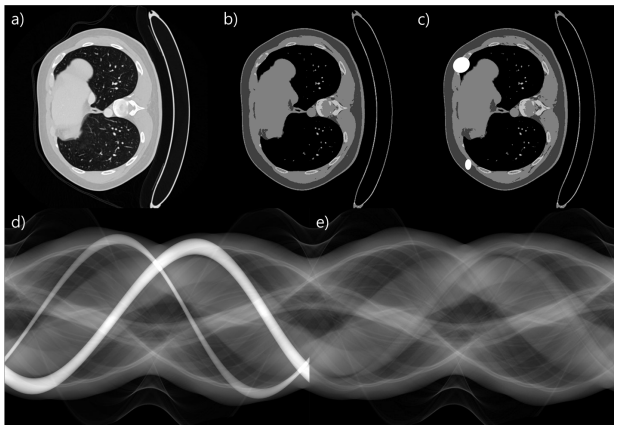


Fig. 1. a) CT image, b) Material map, c) material map with metal, d) sinogram simulated with metal, e) sinogram simulated with VNM-material

TABLE I
FAN-BEAM PARAMETERS FOR SIMULATED DATA

Parameters	Values
Views per rotation	720
Detector pixel number	512 x 1
Detector pixel pitch	0.8 mm
Distance of source to detector	1300 mm
Distance of source to object	900 mm
X-ray source	120 kVp

B. U-Net Training

The U-Net architecture implemented for this experiment is similar to the original U-Net published by Ronneberger et al. and is shown in Fig. 2 [13]. The last output channel is adjusted to be a single channel since the goal of this network is not to segment but produce a new sinogram image. The training input is the metal-inclusive sinogram set generated in the simulator and the label is the VNM-inclusive sinogram set generated in the simulator. From the total of 2351 data pairs, 80% was used for training and 10% was used for validation and test. In order to increase the variety in the dataset, data augmentation was performed by shifting the sinograms randomly for each epoch. The batch size was 4. The network was optimized with ADAM optimizer and the MSE loss function. The network was trained for 400 epochs with learning rate of 10^{-4}, coming to a convergence.

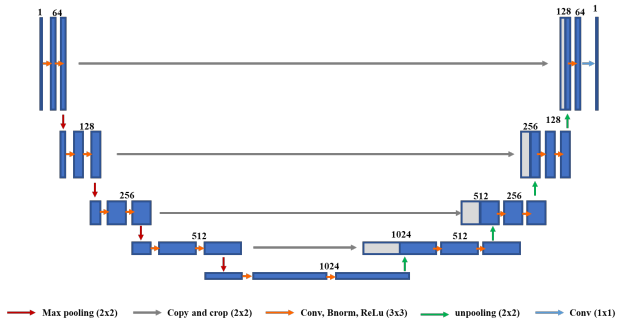


Fig. 2. U-Net structure with the input as a metal-inclusive sinogram and the output image as the VNM-inclusive sinogram pair.

III. RESULTS

For testing, 235 sinogram pairs were synthesized in the same manner as the training data using the simulator. The metal-inclusive sinograms were loaded to the trained network and the resulting output sinograms were reconstructed with an FBP algorithm. The simulated VNM-inclusive sinograms were set as the reference. Fig. 3 shows an example of the test cases. A metal masked image that was extracted with thresholding during NMAR were added to the reconstructed images of the proposed method in order to compare the results qualitatively.

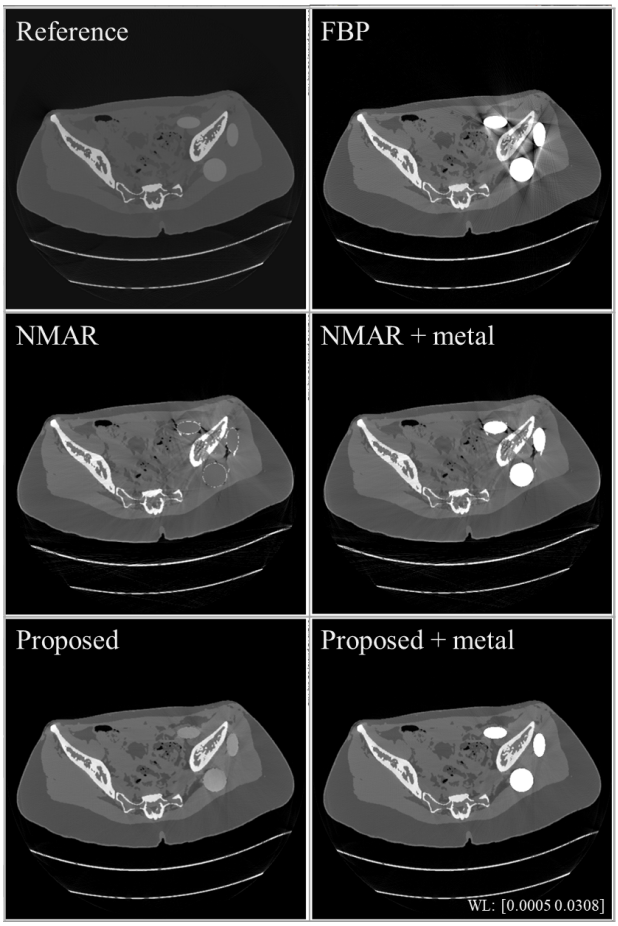


Fig. 3. Image reconstructed from the sinograms of the simulator (Reference and FBP), NMAR with and without metal, and proposed method with and without metal.

Root mean square error (RMSE) and structured similarity index (SSIM) values in the four ROI were calculated to assess the effectiveness of the MAR algorithm in not only removing metal artifacts but also in retaining the anatomical structures [14]. The four ROIs were selected to assess the performance of the NMAR algorithm and the proposed method pertaining to different metal artifacts. ROI 1 aims to evaluate the reduction of dark streaks and the recovery of the soft tissue. ROI 2 focuses on the bone structure near a large metal object. While ROI 3 does not show noticeable metal artifacts, streak artifacts degraded the contrast and the structure of the soft tissue. ROI 4 has streak artifacts on boneless soft tissue. ROI 5 includes all pixels except the metal mask. The ROIs were selected such that the metal replacements are not included but the neighboring pixels are. Table II shows the result of the quantitative calculation performed on the different ROIs marked and displayed on Fig. 4.

IV. Discussion

Depending on the ROI, the difference is significant. The proposed method excels in recovery of the soft tissue and bone information even when there is harsh beam hardening since no prior metal mask is required and the sinograms are substituted not only in the metal mask region but throughout. This can be observed in ROI 1 and 2. In ROI 1, the dark region marked with the red arrow is the neighboring region of a large metal object. Since NMAR tries to recover the lost information with the prior image, the bone information is not recovered fully. This may be due to the fact that the dark region was masked as air instead of tissue in the prior image due to the low pixel value. The proposed method does not require a well-defined prior image and the beam hardening effects are corrected in the sinogram domain. As a result, the marked region is

Similarly, in ROI 2, the proposed method successfully retrieves the bone and the neighboring soft tissue information that are almost lost in the FDK result (red arrow). The success of the proposed method in retrieving the soft tissue information even for those pixels with values close to air poses speculation on the simplicity of the soft tissue model and possible overfitting. However, being that the sinograms for the training set and the test set were simulated from material maps of different patient sets with varying anatomical details, it is unlikely. This can be further investigated by testing clinical data with more complex soft tissue and bone structures to the same network.

The SSIM values of FBP and NMAR are similar for ROI 2 and 4 while that of the proposed method is significantly better. These two ROIs are contaminated with dark and white streaks and have a great potential for creating misleading NMAR priors. For the case of ROI 4, the streaks are prevalent in the soft tissue even after NMAR. Furthermore, beam hardening artifact at the perimeter of the metal objects were not corrected properly and the remaining streaks are observed. The proposed method not only reduces streaks better but also improves the visibility and details of the soft tissue structure.

As for ROI 3, the streaks were reduced well by both NMAR and the proposed method. Yet, the reconstructed NMAR lacks contrast and suffers from soft tissue deformation in multitude of areas. An example of structural degradation is noted by a red arrow. This is quantitatively reflected in the low SSIM value for NMAR compared to the proposed method: 0.7694 compared to 0.9353.

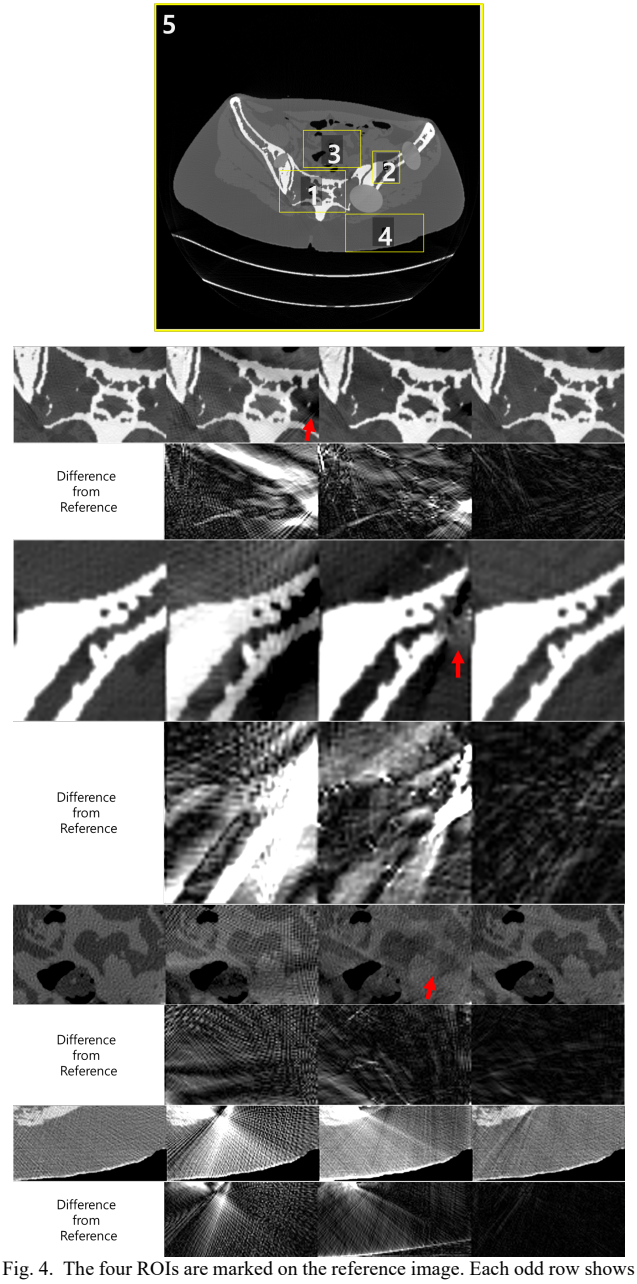


Fig. 4. The four ROIs are marked on the reference image. Each odd row shows the according ROI. The window level of each ROI is [0.001, 0.035], [0.01, 0.025], [0.01, 0.025], and [0.01, 0.015]. Every even row shows the difference image from the reference and the window level is [0 0.005]. The window levels are adjusted to observe the artifacts and their reduction with ease. The columns show the enlarged images at the ROI in the order of reference, FBP, NMAR and proposed method result.

Finally, to compare the effectiveness of the MAR algorithm in general for all areas, ROI 5 was evaluated. The RMSE of all

three results are negligible since large part of the RMSE computes air region which are similar for all cases. However, the performance of MAR on the overall image can be noted significantly from the improved SSIM value. Overall, the proposed method is successful in reducing the metal artifact in the sinogram domain. The quality of the reconstructed image improves as the beam hardening artifacts and streak artifacts are corrected. Significantly, unlike methods that require metal segmentation or depend on the quality of the prior images, the proposed method has merit in working without the need for metal segmentation. Furthermore, the proposed method replaces the metal-inclusive sinogram with a virtual non-metal sinogram. In doing so, the discontinuity of the sinogram is reduced compared to air replacement and streak artifact improvement.

TABLE II
QUANTITATIVE ANALYSIS OF THE ROI

ROI	Parameters	RMSE	SSIM
1	FBP	2.514E-3	0.8289
	NMAR	1.747E-3	0.8655
	Proposed	5.058E-4	**0.9521**
2	FBP	4.574E-3	0.6053
	NMAR	3.093E-3	0.6529
	Proposed	6.471E-4	**0.8612**
3	FBP	1.230E-3	0.5928
	NMAR	8.113E-4	0.7694
	Proposed	3.600E-4	**0.9353**
4	FBP	1.532E-3	0.4031
	NMAR	1.476E-3	0.5753
	Proposed	2.265E-4	**0.8705**
5	FBP	2.6283E-6	0.6368
	NMAR	2.9660E-6	0.7062
	Proposed	4.7452E-7	**0.9556**

V. CONCLUSION

Overall, the propose method outperforms FBP and NMAR in reducing metal artifacts and appropriately constructing the soft tissue and bone information. Through this simulation study, we have successfully demonstrated the feasibility of the proposed method. The result of the preliminary simulation study shows promise for further exploration of the method. We are interested in testing the method with datasets with greater complexity. It could be done by increasing the number of materials and types of metals used to create the sinogram pairs. Additionally, the shape of the metal objects used for the feasibility study were simply elliptical. Metal objects with sharp corners and complex shapes can be added to assimilate screws and needles. Finally, noise and photon starvation effect can be added to the data simulator to test the algorithm under exacerbated metal artifact conditions.

REFERENCES

[1] B. De Man, J. Nuyts, P. Dupont, G. Marchal, and P. Suetens, “Metal streak artifacts in x-ray computed tomography: A simulation study,” IEEE Transactions on Nuclear Science, vol. 46, no. 3, pp. 691–696, 1999.

[2] M. L. Kataoka, M. G. Hochman, E. K. Rodriguez, P.-J. P. Lin, S. Kubo, and V. D. Raptopolous, “A review of factors that affect artifact from metallic hardware on multi-row detector computed tomography.,” *Curr. Probl. Diagn. Radiol.*, vol. 39, no. 4, pp. 125– 36, 2010.

[3] E. Meyer, F. Bergner, R. Raupach, T. Flohr, and M. Kachelrieß, “Normalized metal artifact reduction (NMAR) in computed tomography,” *Medical Physics*, vol. 37, no. 10, pp. 5482–5493, 2010.

[4] V. Ruth, D. Kolditz, C. Steiding, and W. A. Kalender, “Metal artifact reduction in X-ray computed tomography using computer-aided design data of implants as prior information,” *Invest. Radiol.*, vol. 52, no. 6, pp. 349–359, 2017.

[5] J. W. Stayman, Y. Otake, J. L. Prince, A. J. Khanna, and J. H. Siewerdsen, “Model-based tomographic reconstruction of objects containing known components.,” *IEEE Trans. Med. Imaging*, vol. 31, no. 10, pp. 1837–48, Oct. 2012

[6] L. Zhu, Y. Han, L. Li, X. Xi, M. Zhu and B. Yan, "Metal Artifact Reduction for X-Ray Computed Tomography Using U-Net in Image Domain," in IEEE Access, vol. 7, pp. 98743-98754, 2019, doi: 10.1109/ACCESS.2019.2930302.

[7] J. Lee, J. Gu and J. C. Ye, " Unsupervised CT metal artifact learning using attention-guided β-CycleGAN ", *IEEE Trans. Med. Imag.*, vol. 40, no. 12, pp. 3932-3944, Dec. 2021.

[8] D. F. Bauer, C. Ulrich, T. Russ, A.-K. Golla, L. R. Schad, and F. G. Zöllner, “End-to-End Deep Learning CT Image Reconstruction for Metal Artifact Reduction,” *Applied Sciences*, vol. 12, no. 1, p. 404, Dec. 2021.

[9] Y. Zhang and H. Yu, "Convolutional Neural Network Based Metal Artifact Reduction in X-Ray Computed Tomography," in *IEEE Transactions on Medical Imaging*, vol. 37, no. 6, pp. 1370-1381, June 2018.

[10] W. A. Lin, H. Liao, C. Peng, X. Sun, J. Zhang, J. Luo, R. Chellappa, and S.K. Zhou, "Dudonet: Dual domain network for ct metal artifact reduction," in *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, 2019, pp. 10512-10521.

[11] H. S. Park, S. M. Lee, H. P. Kim and J. K. Seo, “CT sinogram-consistency learning for metal-induced beam hardening correction”, 2017, [online] Available: https://arxiv.org/abs/1708.00607.

[12] K. Yan, X. Wang, L. Lu, R. M. Summers, "DeepLesion: Automated Mining of Large-Scale Lesion Annotations and Universal Lesion Detection with Deep Learning" in *Journal of Medical Imaging*, 2018.

[13] O. Ronneberger, P. Fischer, and T. Brox, “U-Net: Convolutional networks for biomedical image segmentation.” in *Medical Image Computing and Computer–Assisted Interventions (MICCAI)*, volume abs/1505.04597 of LNCS, pages 234–241. Springer, 2015.

[14] Zhou Wang, A. C. Bovik, H. R. Sheikh and E. P. Simoncelli, "Image quality assessment: from error visibility to structural similarity," in *IEEE Transactions on Image Processing*, vol. 13, no. 4, pp. 600-612, April 2004.

Attenuation Image Guided Effective Atom Number Image Calculation Using Image-domain Neural Network for MeV Dual-energy Cargo CT Imaging

Wei Fang, Liang Li, *IEEE Senior Member*

***Abstract*–Traditional inspection technology for cargo or container in customs and harbours is MeV X-ray radiography. The biggest limitation of radiography imaging is the overlapping problem. While MeV dual energy CT can provide cross-section image, which is free of the overlapping problem. Besides, the recorded dual-energy projection data provides the ability for material decomposition. Electron density image and effective atom number image can be further calculated from the material decomposition coefficients. However, the quality of effective atom number image can be very poor and the behind reasons are multifaceted. First of all, MeV dual-energy CT material decomposition is much more difficult than keV dual-energy CT since the mass attenuation coefficients of different materials are very close in MeV energy range. Besides, metals in the cargo or container may cause strong beam hardening artefacts, further degrading the image quality of material decomposition results. Last but not least, the way of calculation effective atom number image also tends to bring noise. In this paper, we proposed a deep learning framework for effective atom number image calculation. The network was input with low and high-energy reconstructions and the effective atom number image that was directly calculated using the derived formula and output with the estimated effective atom number image. The simulation results show the effectiveness of the proposed deep learning framework.**

I. Introduction

THE routine security screening of cargos and containers in airports, stations and harbors is very important for the protection of homeland security. The container is large in size and the contents of these cargos are very complicated. It will be a very labour-intensive work if performing manual inspection. Currently, the dominant technique for performing this kind of security screening is X-ray radiography. The biggest limitation for X-ray radiography technology is the overlapping problem. If many objects are stacked along the X-ray, the projection information of different objects will be stacked and it will be difficult to discriminate one object from another by judging from the generated radiography image. CT imaging can provide the attenuation coefficients map of a cross-section without object overlapping. MeV dual-energy CT can be recognized as the next-generation tool for performing routine screening of cargos and containers [1].

MeV dual-energy CT can also provide the ability for material discrimination. In MeV material decomposition, the quality of the generated atom number image is usually worse than the reconstructed low and high-energy attenuation images. However, the low and high-energy attenuation images and the atom number image share similar image structures. The attenuation reconstructions can be used to improve the effective atom number image within the deep learning framework.

In this paper, we proposed a deep learning framework for calculating the effective atom number image for MeV container CT imaging. The deep neural network model has three channels as input. The first two channels are the low and high attenuation images. The third channel is the effective atom number image that calculated by directly using the derived formula. The output of the network is the network estimated effective atom number image. The network was trained with simulated MeV dual-energy container CT dataset. The simulation was performed using the shape of XCAT [2] model but filled with materials that often exist in container and cargo. The model was validated on simulation data. The results and quantitative analysis showed the effectiveness of the proposed method.

II. Physical Model

Compared to traditional single-energy CT, dual-energy CT can provide the capability of material decomposition. The reason why dual-energy CT can perform material decomposition is that the attenuation of materials decays as energy rises and different materials have different decay curves. To be more specific, the decay curves are decided by two aspects of factors. The first aspect is material-related properties, such as mass density, atomic number and mass number. The second aspect is the energy of incident photons. The feasibility that these two aspects of factors can be decoupled is the foundation for material decomposition. Because the objective of material decomposition is to estimate the first aspect of factors, which are material-related properties. In keV clinical dual-energy CT, decomposition coefficient images are calculated to indicate bone-like and soft tissue-like structures. While in MeV container CT, the decomposition coefficient image is more like a medium-result and the electon

Manuscript received January 24, 2022. This work was funded by grant from NNSFC 12027811.

Wei Fang and Liang Li are with the Department of Engineering Physics, Tsinghua University, Beijing, 100084, China & Key Laboratory of Particle & Radiation Imaging (Tsinghua University), Ministry of Education, 100084, China.

Corresponding author: lliang@tsinghua.edu.cn

density image and the effective atom number image are more of physical meanings in security inspection.

The ability of material for attenuating X-ray is mostly based on three physical effects, which are photoelectric effect, compton scattering effect and pair production effect. These three effects can all be decomposed into material-related and energy related factor. In MeV energy range, The photoelectric effect caused attenuation is very small and can be neglected. So currently we only consider compton scattering and pair production as two main effects. For compton scattering part of attenuation, it can be decomposed as the product of material-related factor and energy-related factor as

$$\mu_{cs} \approx 2\rho\frac{Z}{A}\cdot\frac{1}{2}N_A\sigma_{cs} = a_2 \cdot f_2, \tag{1}$$

where ρ is the mass density, Z is the atomic number and A is the mass number. N_A represents the avogadro constant. σ_{cs} represents the single electron compton scattering cross section, which is a function of energy E and can be depicted by using Klein-Nishima formula. Similarly, the pair production caused part of attenuation can be represented as

$$\mu_{pp} = 2\rho\frac{Z^2}{A}\cdot\frac{1}{2}K_{pp}N_A E = a_3 \cdot f_3, \tag{2}$$

where K_{pp} is a constant irrelevant to material or energy.

Material basis and effect basis are just two types of basis for attenuation representation in material decomposition. By listing the equality of the two types of representation, we can figure out that they are totally equivalent. Formula (3) and formula (4) show that coefficients of double effect basis can be represented with the coefficients of double material basis,

$$a_2 = 2\rho\frac{Z_{eff}}{A} = b_1 \cdot 2\rho_1\frac{Z_1}{A_1} + b_2 \cdot 2\rho_2\frac{Z_2}{A_2} \tag{3}$$

$$a_3 = 2\rho\frac{Z_{eff}^2}{A} = b_1 \cdot 2\rho_1\frac{Z_1^2}{A_1} + b_2 \cdot 2\rho_2\frac{Z_2^2}{A_2}, \tag{4}$$

where ρ_i represents the mass density of the basis material i, Z_i and A_i respectively represent the atomic number and mass number of material i. The objective of material basis decomposition is to get the material-specific coefficient b_1 and b_2. In this work, we use post-processing as material decomposition method. After we get the decomposition coefficients b_1 and b_2, the electron density ρ_e and effective atomic number Z_{eff} can be calculated by using the following formulas. The definition of electron density is just the coefficient of compton scattering. It can be represented with basis material coefficients as

$$\rho_e = a_2 = 2\left(b_1\rho_1\frac{Z_1}{A_1} + b_2\rho_2\frac{Z_2}{A_2}\right). \tag{5}$$

The effective atomic number can be calculated by dividing the coefficient of pair production a_3 with the coefficient of compton scattering a_2 [3].

$$Z_{eff} = \frac{a_3}{a_2} = \left[b_1\rho_1\frac{(Z_1)^2}{A_1} + b_2\rho_2\frac{(Z_2)^2}{A_2}\right]\Big/\rho_e, \tag{6}$$

The effective atom number image is prone to have noise because there is a ρ_e at the denominator in the formula (6). This will cause strong noise in area that electron density is extremely low or air area where theoretically electron density is zero but practically is a extremely small value due to noise or discretization error. In practice, the effective atom number image is extremely noisy and very sensitive to the beam hardening artefacts.

III. The Experimental Settings

To cope with the strong noise and artefacts in effective atom number image, we proposed a deep learning framework for the calculation of effective atom number image. The network was input with three channels. The first two channels are low and high energy attenuation reconstructions. The third channel is the directly calculated effective atom number image using formula (6). The low and high-energy attenuation images are related to the effective atom number and have good quality of image structure details, which can be used to guide the restoration process of the effective atom number image in the information flow of neural networks. Similar ideas have already existed in some traditional image denoising algorithms, such as guided image filtering [4] and HYPR framework [5]. Here we try to extend this idea in the framework of deep learning. The diagram of the proposed method was illustrated in Figure 1.

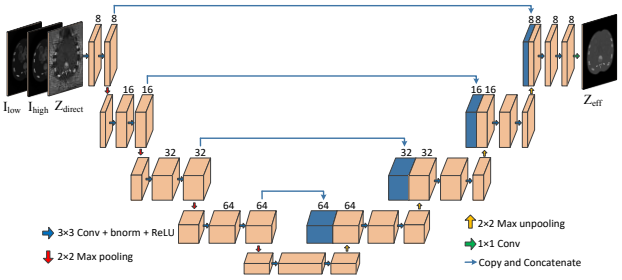


Fig. 1. The schematic plot of the method.

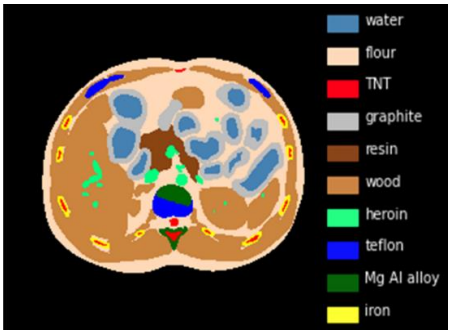


Fig. 2. One phantom image sample used for simulation.

For the training data, we have built an in-house simulation tool for MeV dual-energy CT imaging. The simulation uses Monte Carlo generated 6/9 MeV spectra. The phantom is with the shape of XCAT model but is filled with materials that

frequently appear in security CT inspection, such as iron, carbon, wood and Teflon. The total photon counts for simulation is 10^7. For the scanning protocol, the angle of views is 5701. The number of detector elements is 3701. The size of the reconstructed image is 256×256. One sample phantom image is shown in Figure 2.

There are 300 phantom images used for simulation. For the simulation process, the input is the spectra and phantom and the output is the dual-energy projections. Since we know the material that was practically used for simulation, we can calculate the real effective atom number image as label. Of the 300 phantom images, 240 of them were used for training, 30 of them were used for validation and 30 of them were used for validation.

IV. The Simulation Results

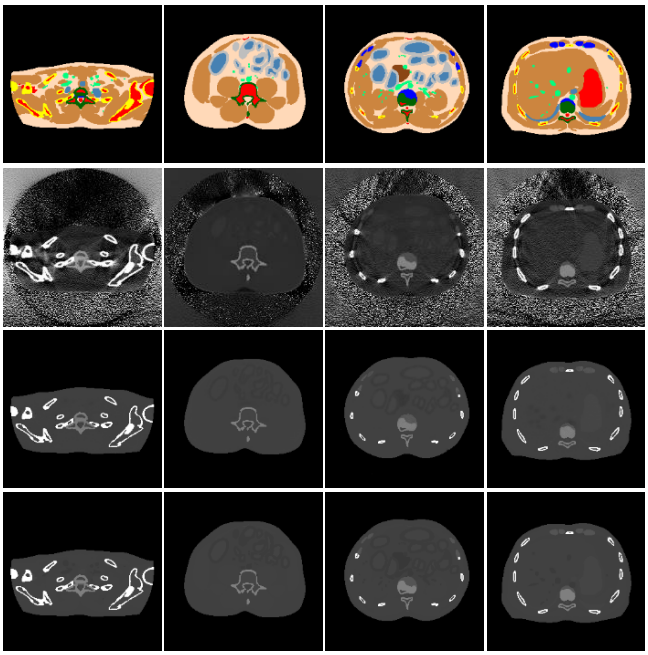


Fig. 3. The first row shows the phantom images. The second row is the directly calculated effective atomic number image using formula (9). The third row is the network output. The fourth row is the ground truth.

The results of simulation dataset on the testing set were shown in Figure 3. The first row shows the corresponding phantom images and the legend is the same with that in Figure 2. The second row shows the directly calculated effective atom number image using formula (6). We can see that the directly calculated image is severely corrupted by beam hardening artefacts. The noise in the air area is also very strong because there is a ρ_e term at the denominator of formula (6). The third row shows the effective atom number images that produced by using the proposed deep learning framework. We can figure out that the proposed framework can significantly reduce beam hardening artefacts and avoid the noise in air area well. The result produced by proposed deep learning framework does not come with no flaws. Some low-contrast structure details were not fully restored compared to the ground truth images shown in the fourth row. Quantitative analysis was performed on the simulation results. Figure 4 and Figure 5 respectively show the calculated Root of Mean Square Error (RMSE) and Structural Similarity (SSIM) [6] of the simulation results on different materials.

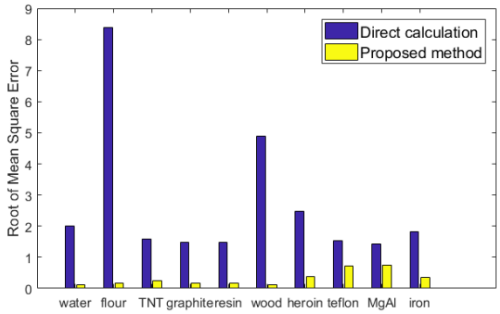


Fig. 4. RMSE of the simulation result on different materials.

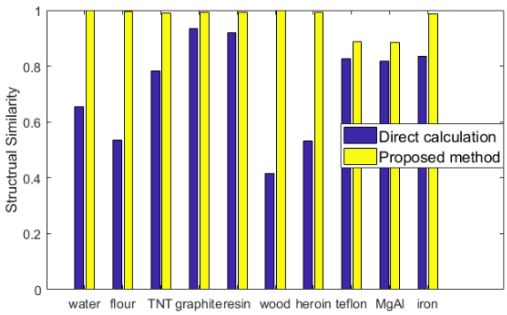


Fig. 5. SSIM of the simulation result on different materials.

We also evaluate the proposed method on manually designed phantoms, which are more like real cases in actual container CT imaging. Figure 6 and Figure 7 show the truck and car phantom.

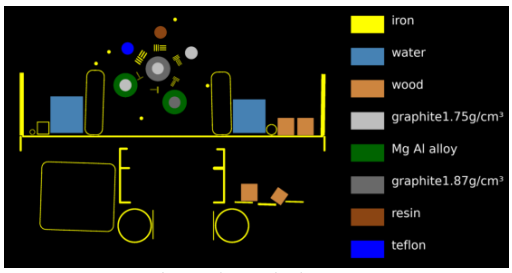


Fig. 6. The truck phantom.

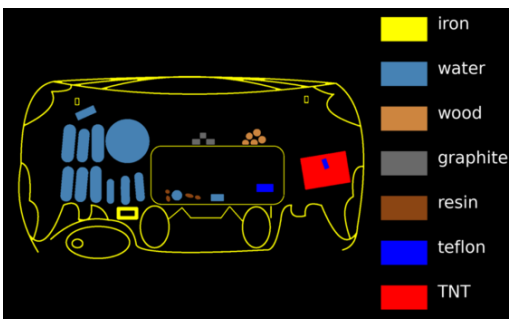


Fig. 7. The car phantom.

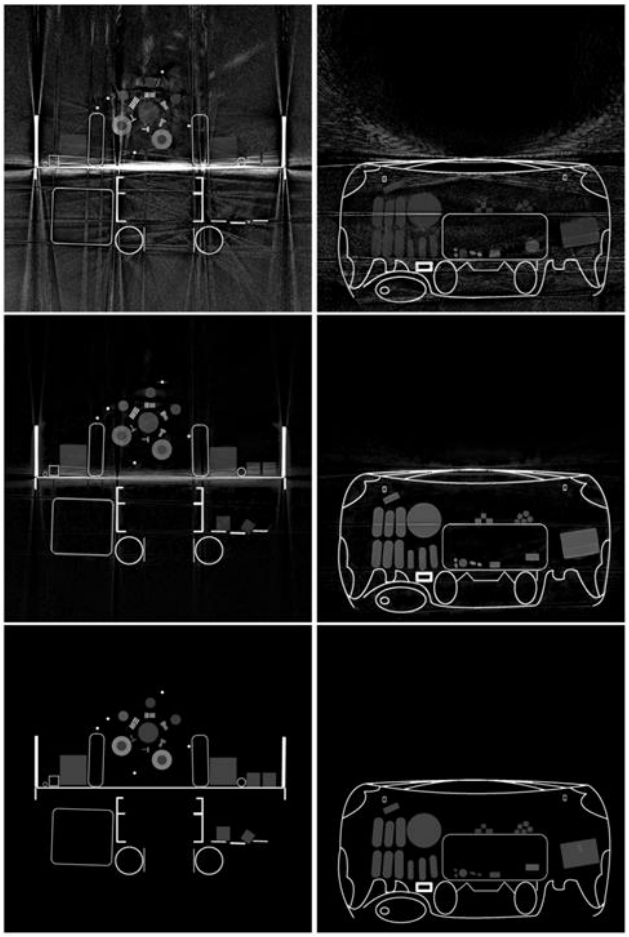


Fig. 8. The results on the manually designed phantoms. The first row shows the directly calculated effective atom number images. The second row shows network produced images. The third row shows the ground truth images.

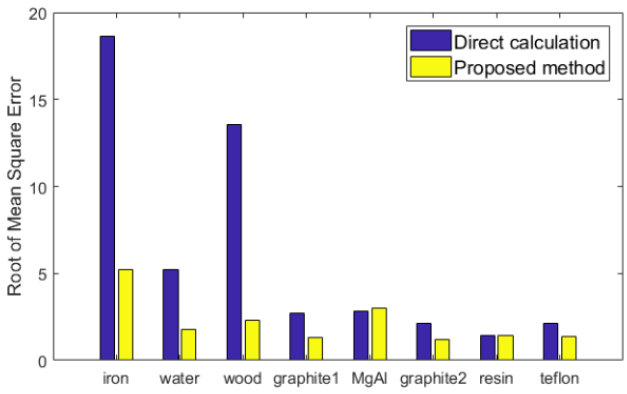


Fig. 9. RMSE of the results on the manually designed phantoms.

Figure 8 shows the simulation results on manually designed phantom images. We can see the strong beam hardening artefacts and severe noise in air area in the directly calculated effective atom number images in the first row. The second row shows the results produced by the proposed neural network. We can see that the beam hardening artefacts have been significantly reduced and the noise was suppressed well in network produced results. The third row shows the ground truth images. Figure 9 and Figure 10 show the RMSE and SSIM of the results on the manually designed phantoms.

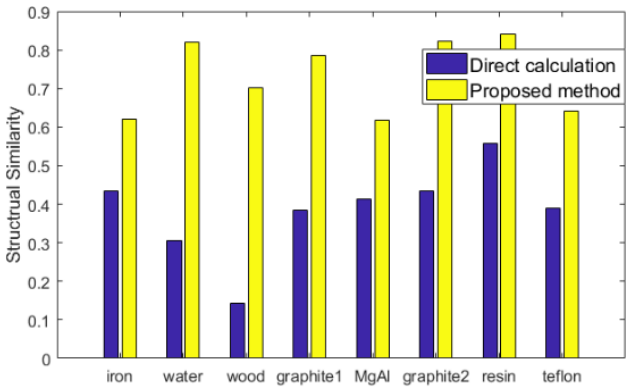


Fig. 10. SSIM of the results on the manually designed phantoms.

Due to the space limit, the experimental results on actual commercial MeV dual-energy CT system were not reported in this abstract.

V. Conclusion

In this paper, we proposed a deep learning framework for calculating the effective atom number image. The network is a modified U-net and was input with three channels and output with the estimated effective atom number image. The input three channels include low and high-energy attenuation reconstruction images and directly calculated effective atom number image. The network uses the guidance from low and high-energy reconstruction to improve the quality of effective atom number image. A MeV dual-energy CT simulation was performed using XCAT phantom's shape but filled with materials that frequently exist in container CT imaging. The simulated dataset was used as training dataset. The proposed method was validated on the simulation dataset, including the XCAT model and two manually designed phantoms. The quantitative analysis indicates the effectiveness of the method.

References

[1] L. Li, T. Zhao, and Z. Chen, "First Dual MeV energy x-ray CT for container inspection: design, algorithm, and preliminary experimental results," *IEEE Access,* vol. 6, pp. 45534-45542, 2018.

[2] W. P. Segars, G. Sturgeon, S. Mendonca, J. Grimes, and B. M. Tsui, "4D XCAT phantom for multimodality imaging research," *Medical physics,* vol. 37, pp. 4902-4915, 2010.

[3] Y. Xing, L. Zhang, X. Duan, J. Cheng, and Z. Chen, "A reconstruction method for dual high-energy CT with MeV X-rays," *IEEE Transactions on Nuclear science,* vol. 58, pp. 537-546, 2011.

[4] K. He, J. Sun, and X. Tang, "Guided image filtering," *IEEE transactions on pattern analysis and machine intelligence,* vol. 35, pp. 1397-1409, 2012.

[5] C. A. Mistretta, O. Wieben, J. Velikina, W. Block, J. Perry, Y. Wu, *et al.*, "Highly constrained backprojection for time-resolved MRI," *Magnetic Resonance in Medicine: An Official Journal of the International Society for Magnetic Resonance in Medicine,* vol. 55, pp. 30-40, 2006.

[6] Z. Wang, A. C. Bovik, H. R. Sheikh, and E. P. Simoncelli, "Image quality assessment: from error visibility to structural similarity," *IEEE transactions on image processing,* vol. 13, pp. 600-612, 2004.

Residual W-shape Network (ResWnet) for Dual-energy Cone-beam CT Imaging

Xiao Jiang, Hehe Cui, Zihao Liu, Lei Zhu and Yidong Yang

***Abstract*—Deep learning has achieved great success in many medical imaging tasks without explicit solutions. In this work, learning method was applied to dual-energy cone-beam CT imaging. We proposed a Residual W-shape Network (ResWnet). ResWnet consists of three modules: scatter correction module $\mathcal{S}$, material decomposition module $\mathcal{M}$, decomposition denoising module $\mathcal{D}$. Both $\mathcal{S}$ and $\mathcal{D}$ use ResUnet architecture, and this light-weight model fuses multi-level features, achieving satisfied performance with a small number of parameters. $\mathcal{S}$ acts on dual-energy attenuation projections to reduce the scatter contaminations, and $\mathcal{D}$ acts on material composition projections to suppress the noise. $\mathcal{M}$ links the modules $\mathcal{S}$ and $\mathcal{D}$, and is used for domain transform from attenuation projections to material projections. This process could be approximated by polynomials with pre-calibrated parameters, that is, $\mathcal{M}$ is a known operator in proposed network with no trainable parameters. This helps to reduce model parameters and improve the performance with small training dataset. Using public head CT dataset, we simulated dual-energy cone-beam CT projections and material projections. Proposed ResWnet was trained, validated and tested on this simulated dataset, verifying its effectiveness in projection-domain scatter correction and low-noise decomposition.**

***Index Terms*—Cone-beam CT, Dual Energy CT, Resnet, Deep Learning**

I. Introduction

CONVENTIONAL CT measures the spatial distribution of x-ray linear attenuation coefficient (LAC) [1]. Dual-energy CT (DECT) [2], which scans object with two different x-ray spectrums, extends the measurement to the energy dimension, and the quantitative information provided by DECT facilitates various new applications, including but not limited to electron density/stopping power calculation [3], synthesis of monochromatic images [4], virtual-non-enhanced images [5]. After several decades development, DECT has become a powerful tool in clinical diagnosis [6]. Another widely used CT is cone-beam CT (CBCT) [7]. Taking advantages of high spatial resolution, large volume coverage and open structure, CBCT provides flexible image guidance in image-guided radiotherapy [8] and image-guided intervention [9], and the flexible geometry is also well suitable for some dedicated clinical tasks, such as breast CT, extremity CT and dental CT [10].

Recently, some groups investigated the feasibility of dual-energy cone-beam CT [11-13] that combines the advantages of DECT and CBCT. Our group implemented a rotation filter [14] configuration to acquired dual-energy data within single rotation. We further proposed a joint bilateral filtering-based algorithm to suppress the image streaks and amplified decomposition noise [15]. However, photon scatter, a major issue in CBCT imaging [16], was not taken into consideration in our previous research. The scatter contamination could severely degrade the imaging accuracy, hampering quantitative dual-energy imaging. Moreover, according to the dual-energy imaging theory [17], image-domain decomposition cannot provide accurate material composition, as well as cannot eliminate beam-hardening effects. Polynomial fitting-based projection-domain decomposition [18] tackles this issue but it is sensitive to projection noise [19].

This work aims to perform projection-domain scatter correction and material decomposition. To this end, we designed a Residual W-shape Network (ResWnet), which consists of two cascade ResUnets. In order to reduce the model parameters and achieve satisfied performance using small training dataset, a known decomposition operator was used to link the two ResUnet.

II. Methodology

A. Dual-Energy Cone-Beam Projection Model

The polyenergetic forward model for dual-energy cone-beam projections is written as:

$$\boldsymbol{p}_{l/h} = \boldsymbol{I}_{l/h} \int \omega_{l/h}(E) \exp\left(-\int \boldsymbol{\mu}(\boldsymbol{x}, E)\mathrm{d}l\right)\mathrm{d}E \qquad (1)$$

where l/h means physical quantities under low/high incident x-ray spectrum, $\boldsymbol{p}$ represents transmission photons, ω represents spectrum, $\boldsymbol{\mu}$ stands for LAC. Due to low scanning dose and large volume coverage, CBCT projections suffer high noise level ($\boldsymbol{n}_{l/h}$) and severe scatter contamination ($\boldsymbol{s}_{l/h}$). Consequently, the actual measurements are modeled as:

This work was supported by the Fundamental Research Funds for the Central Universities (Grant No. WK2030000037), Anhui Provincial-level S&T Megaprojects (Grant No. BJ2030480006), National Natural Science Foundation of China (Grant No. 81671681), the Ministry of Science and Technology of China Key Research and Development Projects (Grant No. 2016YFC0101400) and by the Fundamental Research Funds for the Central Universities (Grant No. WK2030040089)."

Xiao Jiang, Hehe Cui, Zihao Liu, Lei Zhu and Yidong Yang are with Department of Engineering and Applied Physics, University of Science and Technology of China, Hefei, Anhui, 230026 China (e-mail: jx2015, hehecui, zihaoliu, leizhusg, ydyang@mail.ustc.edu.cn.

Corresponding authors: Lei Zhu and Yidong Yang.

$$\boldsymbol{p}_{l/h}^{ns} = \boldsymbol{p}_{l/h} + \boldsymbol{n}_{l/h} + \boldsymbol{s}_{l/h} \tag{2}$$

The superscript n, s represent noise and scatter, respectively. After log normalization, the line integrals with and with scatter and noise are:

$$\boldsymbol{q}_{l/h} = -\log\left(\int \omega_{l/h}(E)\exp\left(-\int \boldsymbol{\mu}(\boldsymbol{x},E)\mathrm{d}l\right)\mathrm{d}E\right) \tag{3}$$

$$\boldsymbol{q}_{l/h}^{ns} = -\log\left(\int \omega_{l/h}(E)\exp\left(-\int \boldsymbol{\mu}(\boldsymbol{x},E)\mathrm{d}l\right)\mathrm{d}E + \frac{\boldsymbol{n}_{l/h}+\boldsymbol{s}_{l/h}}{\boldsymbol{I}_{l/h}}\right) \tag{4}$$

B. Projection-Domain Material Decomposition

According to the dual-energy CT theory [17], the LAC could be decomposed as:

$$\boldsymbol{\mu}(\boldsymbol{x},E) = \boldsymbol{\rho}_1(\boldsymbol{x})\phi_1(E) + \boldsymbol{\rho}_2(\boldsymbol{x})\phi_2(E) \tag{5}$$

$\phi_{1/2}$ is energy-dependent basis function, which could be interpreted as mass attenuation coefficient of two basis materials, then $\boldsymbol{\rho}_{1/2}$ is the density of the basis material correspondingly. Plug Eq. (5) into Eq. (3), we obtain:

$$\boldsymbol{q}_{l/h} = -\log\left(\int \omega_{l/h}(E)\exp(-\phi_1(E)\boldsymbol{q}_1 - \phi_2(E)\boldsymbol{q}_2)\,\mathrm{d}E\right) \tag{6}$$

where $\boldsymbol{q}_1 = \int \boldsymbol{\rho}_1(\boldsymbol{x})\mathrm{d}l$, $\boldsymbol{q}_2 = \int \boldsymbol{\rho}_2(\boldsymbol{x})\mathrm{d}l$. Given incident spectrum $\omega_{l/h}$ and two basis functions $\phi_{1/2}$, the material decomposition is an inverse problem which recovers $\boldsymbol{q}_{1/2}$ from $\boldsymbol{q}_{l/h}$. Since no explicit expression for this inversion, A commonly used analytical decomposition method is polynomial approximation [18]:

$$\boldsymbol{q}_1 = \sum_{i,j=0}^{N} \alpha_{i,j}\,\boldsymbol{q}_l^i\boldsymbol{q}_h^j, \qquad \boldsymbol{q}_2 = \sum_{i,j=0}^{N} \beta_{i,j}\,\boldsymbol{q}_l^i\boldsymbol{q}_h^j \tag{7}$$

where $\alpha_{i,j}$, $\beta_{i,j}$ are pre-calibrated coefficients using calibration phantom. Although this method could achieve high accuracy via increasing the polynomial order N, previous research revealed that this decomposition is sensitive to projection noise [19].

C. Residual W-shape Network

Figure 1 presents the ResWnet Architecture, which consists of scatter correction module, material decomposition module and decomposition denoising module. For simplicity, projections represent log-normalized projections hereafter.

1) Scatter Correction Module ($\mathcal{S}$: $\boldsymbol{q}_{l/h}^{ns} \mapsto \boldsymbol{q}_{l/h}$)

Scatter correction module aims to remove the noise and scatter signal in projections. In this work, scatter correction module $\mathcal{S}$ employs a Residual U-shape network (ResUnet) [20]. The paired $\boldsymbol{q}_l^{ns}, \boldsymbol{q}_h^{ns}$ are concatenated as two-channel input, which then passes six encoder blocks and five decoder blocks successively. To avoid gradient vanishing and exploding [21] in very deep networks, each encoder and decoder block adopts residual architecture [22] with a shortcut connection from input to output. Considering that $\boldsymbol{q}_{l/h}^{ns}$ and $\boldsymbol{q}_{l/h}$ share the same structure with only numerical difference, a global shortcut connection is applied to directly add raw projections to the output projections. Different from the conventional Unet, the last two encoder blocks in $\mathcal{S}$ keep the same number of features to reduce model parameters without compromising the performance.

2) Material Decomposition Module ($\mathcal{M}$: $\boldsymbol{q}_{l/h} \mapsto \boldsymbol{q}_{1/2}$):

Material decomposition module $\mathcal{M}$ aims to recover the material composition projections $\boldsymbol{q}_{1/2}$ from the dual-energy projections $\boldsymbol{q}_{l/h}$. As discussed in Sec.II.B, this inversion could be approximated by polynomials functions, and the parameters $\alpha_{i,j}$, $\beta_{i,j}$ could be determined by pre-calibration. Thus, there is no need to train a sub network for decomposition. In proposed network, $\mathcal{M}$ acts as depicted in Eqs. (7). N was set to 4 and there is no trainable parameter in this module.

3) Decomposition Denoising Module ($\mathcal{D}$: $\boldsymbol{q}_{1/2}^{s} \mapsto \boldsymbol{q}_{1/2}$):

Since the module $\mathcal{S}$ cannot reduce the noise level to zero and

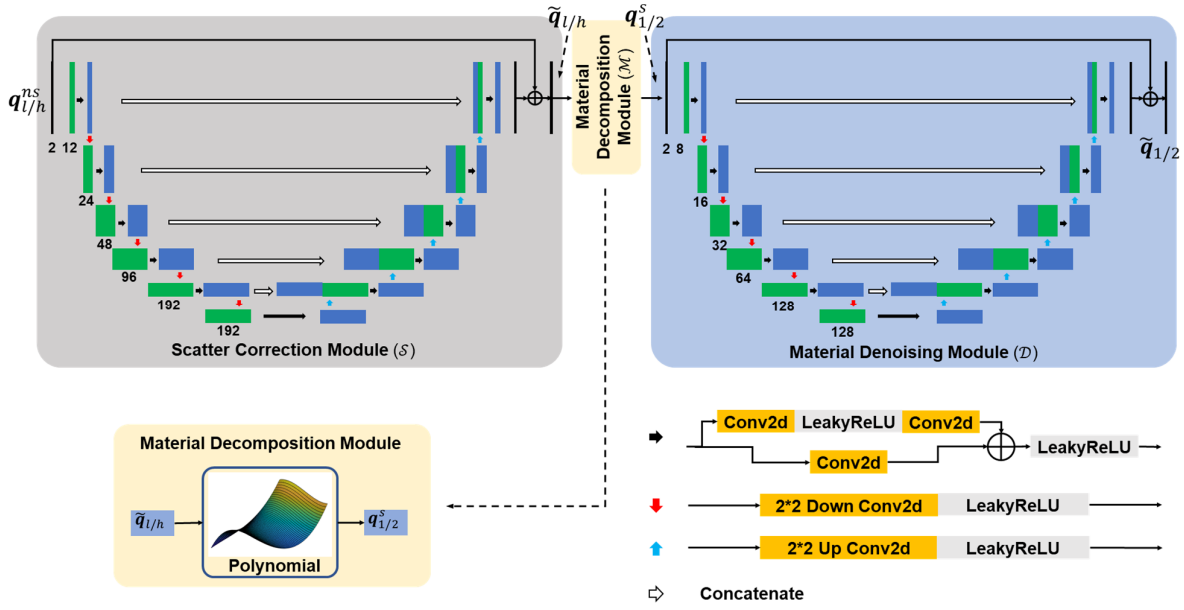


Fig. 1. ResWnet architecture. Gray, blue and yellow boxes represent scatter correction, material decomposition, decomposition denoising module, respectively.

the polynomial-based decomposition is sensitive to noise, the output of $\mathcal{M}$ would be noisy. Decomposition denoising module $\mathcal{D}$ was used for further suppress the noise. Same as the scatter correction module, $\mathcal{D}$ also employs the ResUnet architecture.

D. Data Generation and network Training

To obtain the training, validation and testing datasets, we first download 22 head CT scans from public dataset in https://wiki.cancerimagingarchive.net/pages/viewpage.action?pageId=39879146. These CT images were decomposed into four tissues: fat, muscle, 200mg/cc bone and 800mg/cc bone using multi thresholds segmentation. We performed polyenergetic forward projection of each volume using two spectrums generated by Spektr [23]. These are the label projections $\boldsymbol{q}_{l/h}$ for scatter correction. The input projections $\boldsymbol{q}_{l/h}^{ns}$ were then obtained via adding scatter and noise signals generated by Monte Carlo simulation. Muscle and 800mg/cc bone were selected as basis materials in this work, and fat and 200mg/cc bone were decomposed onto these two basis. The label projections of material decomposition $\boldsymbol{q}_{1/2}$ were generated by forward projecting the muscle and 800mg/cc composition images. Each head produced 100 projections, and the projection angles equally distributed between 0 and 2π.

During the model training, network $\mathcal{S}$ were firstly optimized by:

$$\mathcal{S}^* = \underset{\mathcal{S}}{\operatorname{argmin}}\, w_1 \left\| \mathcal{S}\left(\boldsymbol{q}_{l/h}^{ns}\right) - \boldsymbol{q}_{l/h} \right\|_1 + w_2 \left\| \mathcal{M}\left(\mathcal{S}\left(\boldsymbol{q}_{l/h}^{ns}\right)\right) - \mathcal{M}\left(\boldsymbol{q}_{l/h}\right) \right\|_1 \quad (8)$$

The second term was added because we hope $\mathcal{S}$ could not only reduce the scatter signal, but also produce a projection noise distribution that minimizes the decomposition noise. In this work, w_1, w_2 were set to 0.9 and 0.1, respectively. Using the trained network $\mathcal{S}^*$, the network $\mathcal{D}$ was finally trained by:

$$\mathcal{D}^* = \underset{\mathcal{D}}{\operatorname{argmin}} \left\| \mathcal{D}\left(\mathcal{M}\left(\mathcal{S}^*\left(\boldsymbol{q}_{l/h}^{ns}\right)\right)\right) - \boldsymbol{q}_{1/2} \right\|_1 \quad (9)$$

Parameters of both $\mathcal{D}$ and $\mathcal{S}$ were optimized by Adam optimizer with an initial learning rate of 0.0004 which decay 8% after each epoch. Batch size was set to 4 and training stopped after 100 epochs.

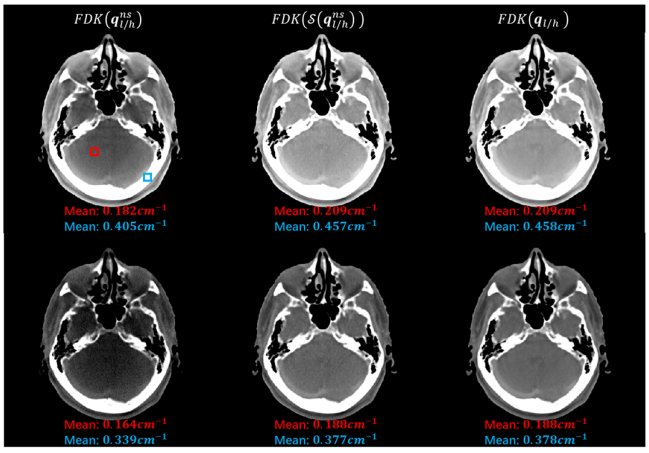


Fig. 2. Dual-energy CT images. Top and bottom rows are low-energy and high-energy CT images, respectively. Mean values of two boxed areas are listed below each image. Display window: [0.15,0.25]cm^{-1}

III. Results

Corrected projections and decomposed material projections are obtained after the simulated cone-beam projections pass the first and the second ResUnet, respectively. Tomographic images are reconstructed via conventional FDK algorithm [24].

Figure 1 displays the dual-energy CT images. As in the left column, scatter contamination leads to obvious shading artifacts on soft tissue. The bone tissue, although clearly visualized, has a numerical error more than 10%. Proposed network successfully removes the image shading, with preservation of small bones and details in intracranial soft tissues. On both soft tissues and bone tissues, proposed network reduced the error to less than 0.5%.

Figure 2 displays the material composition images. As discussed above, polynomial-fitting based decomposition in the left column suffers terrible noise, especially on the soft tissue images. Without significantly compromising the spatial resolution, proposed network suppresses the noise of bone composition and muscle composition images by 32.7% and 65.6%, respectively. Subtle details around nasal cavity are faithfully recovered by proposed method as well.

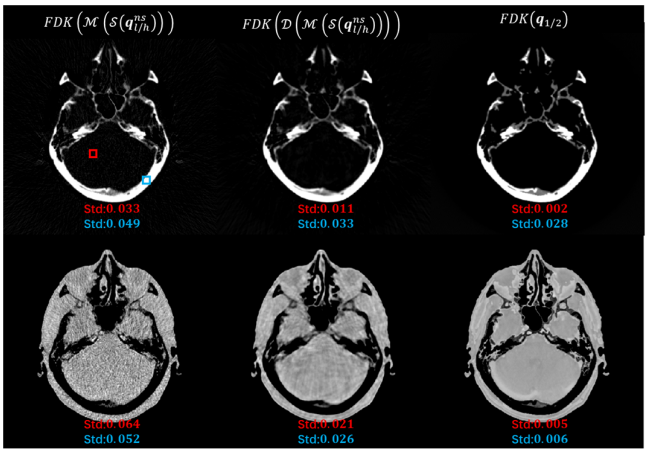


Fig. 3. Material composition images. Top and bottom rows are bone and muscle images, respectively. Std values of two boxed areas are listed below each image. Display window: bone: [0,0.8], muscle: [0.6 1.2]

Conclusion And Discussion

In this work, we proposed a ResWnet for dual-energy cone-beam CT imaging. Three modules were designed for scatter correction, material decomposition and decomposition denoising, respectively. To reduce the model complexity and optimize the performance using small training dataset, both two trainable modules perform transform between the same domain, and the domain transform from attenuation projection to material composition projection is achieved by a known operator, i.e., polynomial-based decomposition. Simulation study preliminarily demonstrated the performance of proposed methods on scatter correction and low-noise decomposition.

Furthermore, some details lost in the material images, especially on the soft tissue images, sharp bone edges were also blurred to some extent. More complex model is needed to improve the performance. However, current study only used 2000 projections for training and validation, which is not capable of training more complex models. More cone beam projections will be simulated in the next step. Another issue is that current simulated projections only account for the primary signal, photon scatter and noise., and other factors such as off focus and detector glare are not considered. More accurate forward projector is needed to simulate real projections, then trained model could be applied to physical measurements acquired in the real system.

REFERENCES

[1] J. Hsieh, *Computed tomography: principles, design, artifacts, and recent advances*. SPIE press, 2003.

[2] A. Graser, T. R. Johnson, H. Chandarana, and M. Macari, "Dual energy CT: preliminary observations and potential clinical applications in the abdomen," *European radiology,* vol. 19, no. 1, p. 13, 2009.

[3] L. I. R. Garcia, J. F. P. Azorin, and J. F. Almansa, "A new method to measure electron density and effective atomic number using dual-energy CT images," *Physics in Medicine & Biology,* vol. 61, no. 1, p. 265, 2015.

[4] L. Yu, S. Leng, and C. H. McCollough, "Dual-energy CT–based monochromatic imaging," *American journal of Roentgenology,* vol. 199, no. 5_supplement, pp. S9-S15, 2012.

[5] L.-J. Zhang *et al.*, "Liver virtual non-enhanced CT with dual-source, dual-energy CT: a preliminary study," *European radiology,* vol. 20, no. 9, pp. 2257-2264, 2010.

[6] M.-J. Kang, C. M. Park, C.-H. Lee, J. M. Goo, and H. J. Lee, "Dual-energy CT: clinical applications in various pulmonary diseases," *Radiographics,* vol. 30, no. 3, pp. 685-698, 2010.

[7] D. Jaffray and J. Siewerdsen, "Cone-beam computed tomography with a flat-panel imager: initial performance characterization," *Medical physics,* vol. 27, no. 6, pp. 1311-1323, 2000.

[8] X. Liang, Y. Jiang, and T. Niu, "Quantitative cone-beam CT imaging in radiotherapy: Parallel computation and comprehensive evaluation on the TrueBeam system," *IEEE Access,* pp. 2169-3536, 2019.

[9] M. Maybody, C. Stevenson, and S. B. Solomon, "Overview of navigation systems in image-guided interventions," *Techniques in vascular and interventional radiology,* vol. 16, no. 3, pp. 136-143, 2013.

[10] R. Fahrig, D. A. Jaffray, I. Sechopoulos, and J. W. Stayman, "Flat-panel conebeam CT in the clinic: history and current state," *Journal of Medical Imaging,* vol. 8, no. 5, p. 052115, 2021.

[11] R. Cassetta *et al.*, "Fast-switching dual energy cone beam computed tomography using the on-board imager of a commercial linear accelerator," *Physics in Medicine & Biology,* vol. 65, no. 1, p. 015013, 2020.

[12] L. E. Schyns *et al.*, "Optimizing dual energy cone beam CT protocols for preclinical imaging and radiation research," *The British Journal of Radiology,* vol. 90, no. 1069, p. 20160480, 2017.

[13] L. Shi *et al.*, "Characterization and potential applications of a dual-layer flat-panel detector," *Medical physics,* vol. 47, no. 8, pp. 3332-3343, 2020.

[14] C. Fang, G. Xu, and L. Zhu, "Single scan dual energy cone beam CT using a rotating filter," in *Medical Imaging 2020: Physics of Medical Imaging*, 2020, vol. 11312, p. 113123S: International Society for Optics and Photonics.

[15] X. Jiang, C. Fang, P. Hu, H. Cui, L. Zhu, and Y. Yang, "Fast and effective single-scan dual-energy cone-beam CT reconstruction and decomposition denoising based on dual-energy vectorization," *Medical Physics,* vol. 48, no. 9, pp. 4843-4856, 2021.

[16] L. Zhu, Y. Xie, J. Wang, and L. Xing, "Scatter correction for cone-beam CT in radiation therapy," *Medical physics,* vol. 36, no. 6Part1, pp. 2258-2268, 2009.

[17] R. E. Alvarez and A. Macovski, "Energy-selective reconstructions in x-ray computerised tomography," *Physics in Medicine & Biology,* vol. 21, no. 5, p. 733, 1976.

[18] P. Stenner, T. Berkus, and M. Kachelriess, "Empirical dual energy calibration (EDEC) for cone-beam computed tomography," *Medical physics,* vol. 34, no. 9, pp. 3630-3641, 2007.

[19] M. Petrongolo, X. Dong, and L. Zhu, "A general framework of noise suppression in material decomposition for dual-energy CT," *Medical physics,* vol. 42, no. 8, pp. 4848-4862, 2015.

[20] Z. Liu and H. Yuan, "An Res-Unet Method for Pulmonary Artery Segmentation of CT Images," *Journal of Physics: Conference Series,* vol. 1924, no. 1, p. 012018 (6pp), 2021.

[21] S. Al-Abri, T. Lin, M. Tao, and F. Zhang, "A Derivative-Free Optimization Method With Application to Functions With Exploding and Vanishing Gradients," *IEEE Control Systems Letters,* vol. PP, no. 99, pp. 1-1, 2020.

[22] K. He, X. Zhang, S. Ren, and J. Sun, "Deep residual learning for image recognition," in *Proceedings of the IEEE conference on computer vision and pattern recognition*, 2016, pp. 770-778.

[23] J. Punnoose, J. Xu, A. Sisniega, W. Zbijewski, and J. Siewerdsen, "spektr 3.0—A computational tool for x-ray spectrum modeling and analysis," *Medical physics,* vol. 43, no. 8Part1, pp. 4711-4717, 2016.

[24] L. A. Feldkamp, L. C. Davis, and J. W. Kress, "Practical cone-beam algorithm," *Journal of the Optical Society of America A,* vol. 1, no. 6, pp. 612-619, 1984.

Dark-Field Imaging on a Clinical CT System: Modelling of Interferometer Vibrations

Clemens Schmid, Manuel Viermetz, Nikolai Gustschin, Jakob Haeusele, Tobias Lasser, Thomas Koehler, Franz Pfeiffer

***Abstract*—X-ray computed tomography (CT) is an invaluable imaging technique for non-invasive medical diagnosis. However, for soft tissue in the human body the inherent small difference in attenuation limits its significance. Grating-based X-ray phase-contrast is a relatively novel imaging method which detects additional interaction mechanisms between photons and matter, namely refraction and small-angle scattering, to generate additional images with different contrast. The experimental setup involves a Talbot-Lau interferometer whose susceptibility to mechanical vibrations hindered acquisition schemes suitable for clinical routine in the past. We present a processing pipeline to identify spatially and temporally variable fluctuations occurring in the first interferometer installed on a continuously rotating clinical CT gantry. The correlations of the vibrations in the modular grating setup are exploited to identify a small number of relevant vibration modes, allowing for an artifact-free reconstruction of a sample.**

I. Introduction

GRATING-based X-ray differential phase-contrast [1] uses the Talbot effect to retrieve additional information about the sample from the X-ray wavefront. Besides the conventional attenuation coefficient, the refractive index decrement and ultra-small-angle scattering as the linear diffusion coefficient [2] can be obtained. This is achieved by generating a periodic interference pattern by a modulation grating G_1, creating self-images at specific distances. The "intensity" is the local mean of the pattern, the relative magnitude of the modulation is called "visibility", and the position of the pattern is the "phase". These quantities are altered by the presence of a sample, where attenuation leads to an overall intensity reduction of the pattern, refraction shifts its lateral position and coherent small-angle scattering (diffusion) reduces the visibility. As the interference pattern is usually too small to be resolved directly, an analyzer grating G_2 is placed in front of the detector to sub-sample the wavefront [3]. One of the gratings is moved in small increments to obtain the convolution of the G_2 modulation with the interference pattern at multiple positions. From these data points the three signals transmission, dark-field contrast, and differential phase shift can be retrieved. This procedure is called "phase stepping". The method was developed with highly coherent synchrotron radiation and brought to laboratory setups by including a third grating G_0 in the interferometer [4]. Placed between G_1 and a conventional X-ray source, it transforms the latter into many narrow slit sources which are mutually incoherent but produce individual G_1 interference patterns adding up constructively at the detector plane. The combination of G_0, G_1, and G_2 gratings is called a Talbot-Lau interferometer. Fig. 1 shows a sketch of the experimental setup of such an interferometer with inverse geometry [5], in which the sample is placed between G_1 and G_2.

The first Talbot-Lau interferometer mounted in a continuously rotating clinical gantry is presented in [6]. It is a modified, commercial CT platform with 80 cm bore size, operated in a standard clinical scan protocol and with sufficient field-of-view for imaging a human. The sampling of the stepping curve relies on vibrations intrinsic to the system, meaning no explicit phase stepping is performed. These vibrations also cause fluctuations of the interference pattern (intensity, visibility, and fringe phase) independently from the sample properties.

For artefact-free reconstruction, it is required to separate changes of the interference pattern caused by vibrations from those caused by the sample. The goal of the presented work is to model the vibration-induced changes in such a way that this separation is possible.

II. Methods

The experimental setup of interest in this work is a commercial clinical gantry platform (Brilliance iCT, Philips) which has been retrofitted with gratings to enable human CT scans giving dark-field contrast. A schematic of the interferometer and the expected modes of vibrations is shown in Fig. 1. The system design is presented in [6]. The gantry is operated in a continuously rotating manner and the acquisition utilizes the vibrations intrinsic to the system which generate sufficient sampling of the stepping curve to perform phase retrieval.

The general approach for separating the changes of the interference pattern induced by vibrations from those induced by the sample is to derive a parametric model from an air scan, which describes the interference pattern generated by the gratings and its variability. The goal is to minimize the number of parameters related to the actual vibration state, as they need to be determined for each subsequent sample scan due to their limited reproducibility. The process is set up in two steps: In the first step, polynomial variations for phase and visibility

C. Schmid, M. Viermetz, N. Gustschin, J. Haeusele, and F. Pfeiffer are with the Chair of Biomedical Physics, Department of Physics, Technical University of Munich, 85748 Garching, Germany. (email: clem.schmid@tum.de)

C. Schmid, M. Viermetz, N. Gustschin, J. Haeusele, T. Lasser, and F. Pfeiffer are with the Munich Institute of Biomedical Engineering, Technical University of Munich, 85748 Garching, Germany.

T. Lasser is with Computational Imaging and Inverse Problems, Department of Informatics, Technical University of Munich, 85748 Garching, Germany.

T. Koehler is with Philips Research, 22335 Hamburg, Germany.

T. Koehler and F. Pfeiffer are with the Institute for Advanced Study, Technical University of Munich, 85748 Garching, Germany.

F. Pfeiffer is with the Department of Diagnostic and Interventional Radiology, School of Medicine and Klinikum rechts der Isar, Technical University of Munich, 81675 Munich, Germany

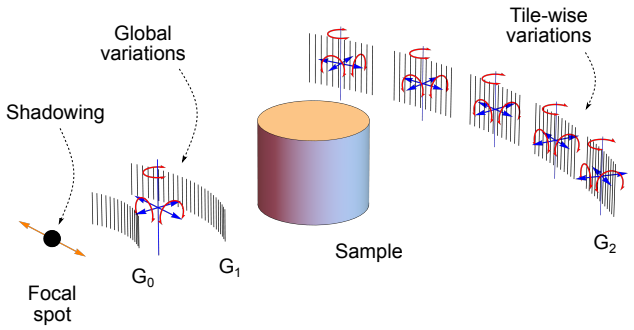


Fig. 1. Schematic depiction of the setup. It is an inverse Talbot-Lau interferometer with the sample placed between G_1 and G_2. G_0 and G_1 are close to the X-ray source and consist of a single grating, respectively. The G_2 consists of multiple smaller gratings which are tiled to cover the whole detector. The G_0-G_1 combination vibration creates phase fluctuations globally on the detector. The individual G_2 tile vibration creates additional tile-wise variations. The focal spot movement creates global intensity and visibility fluctuations due to shadowing. The interference pattern amplitude is further reduced by grating movement during an exposure.

are assumed. In a second step, the number of parameters is reduced using a principal component analysis."

A. Forward model

The canonical forward model for the measured intensity using a stepped Talbot-Lau interferometer is [4]

$$y_{pt}^{\text{simple}} = I_p \left(1 + V_p \cos\left(\phi_p + \gamma_t\right)\right) , \tag{1}$$

with the expected intensity y_{pt}^{simple} in detector pixel $p \in \{1, \dots, P\}$ at stepping position index $t \in \{1, \dots, T\}$, flat-field intensity I_p, flat-field visibility V_p, flat-field phase ϕ_p, and the global phase shift γ_t induced by moving one of the gratings perpendicular to the grating bars. The flat-fields (I_p, V_p, ϕ_p) are intrinsic to the interferometer setup and usually assumed to be constant during a scan and between scans. In laboratory setups it is assumed that γ_t is known and exactly the same for air scan and sample scan.

As indicated in Fig. 1 we expect various vibrations which will lead to a pixel- and time-dependent change of the intensity, visibility, and phase of the interference pattern. The forward model (1) is extended to

$$y_{pt} = I_p \left(1 + I_{pt}^{\text{vib}}\right) \left[1 + V_p \left(1 + V_{pt}^{\text{vib}}\right) \cos\left(\phi_p + \phi_{pt}^{\text{vib}}\right)\right] , \tag{2}$$

with I_{pt}^{vib}, V_{pt}^{vib}, and ϕ_{pt}^{vib} representing spatial and temporal fluctuations over p and t.

B. Phase fluctuations

As indicated in Fig. 1 we assume translations and rotations of the G_0 grating, G_1 grating, the G_2 carrier, and the individual G_2 tiles. We initially treat each G_2 tile as an independent interferometer, so p refers to a pixel behind one G_2 tile.

According to [7], the resulting changes of the phase can be accurately modeled by low-order two-dimensional polynomials over the detector. We define a two-dimensional polynomial term $\boldsymbol{\mathcal{P}}_{ij} = (\mathcal{P}_{ijp})$ of order i along the width and order j along the height of the interferometer in pixel p as $\mathcal{P}_{ijp} = w(p)^i h(p)^j$. $w(p)$ is the w coordinate along the width and $h(p)$ the h coordinate along the height of the interferometer in pixel p. Each term of the polynomial is multiplied with a coefficient γ_{ijt} to give ϕ_{pt}^{vib}

$$\phi_{pt}^{\text{vib}} = \sum_{i,j} \gamma_{ijt} \mathcal{P}_{ijp} \, . \tag{3}$$

A maximum order of one is chosen along the grating bars (in j here) and two perpendicular to them (in i here) [7].

C. Visibility fluctuations

According to [8] change in total phase (i.e. the terms inside the cosine) during an exposure leads to a visibility drop proportional to the first time-derivative of the total phase. Besides movement of the gratings, changes of the X-ray focal spot location can also cause visibility fluctuations as the gratings are bent to a cylindrical surface and carefully focused onto the intended source position. It is assumed that both can be approximated by two-dimensional polynomials. We formulate a general model for the visibility fluctuation

$$V_{pt}^{\text{vib}} = \sum_{i,j} \beta_{ijt} \mathcal{P}_{ijp} \, , \tag{4}$$

with the maximum polynomial order in i and j doubled in comparison to (3), such that the first time-derivative of the phase vibrations can be modeled.

D. Principal vibration components

Instead of modeling each G_2 tile independently, it is desirable to approximate the fluctuations with a small number of dominant shared modes to reduce the amount of free parameters per exposure.

We use principal component analysis (PCA) to reduce the number of parameters. Let $\boldsymbol{X} \in \mathbb{R}^{P \times T}$ be a data matrix with P variables in the rows and T observations in the columns. The singular value decomposition on $\boldsymbol{X}$ is given as $\boldsymbol{X} = \boldsymbol{U}\boldsymbol{\Sigma}\boldsymbol{V}^{\mathrm{T}}$, with the orthogonal matrix $\boldsymbol{U} \in \mathbb{R}^{P \times P}$, the diagonal matrix $\boldsymbol{\Sigma} \in \mathbb{R}^{P \times T}$, and the orthogonal matrix $\boldsymbol{V} \in \mathbb{R}^{T \times T}$. $\boldsymbol{\Sigma}$ contains the singular values of $\boldsymbol{X}$ on its diagonal which are defined to be in descending order. The rows of $\boldsymbol{\Sigma}\boldsymbol{V}^{\mathrm{T}}$ are the "principal components" of $\boldsymbol{X}$ and the columns of $\boldsymbol{U}$ are the "principal directions" of $\boldsymbol{X}$, i.e. the magnitude of each principal component per observation. Let the function PCA be defined as acting on a matrix $\boldsymbol{X}$ and returning $\boldsymbol{\Sigma}\boldsymbol{V}^{\mathrm{T}}$ and $\boldsymbol{U}^{\mathrm{T}}$: $\mathrm{PCA}(\boldsymbol{X}) \to (\boldsymbol{\Sigma}\boldsymbol{V}^{\mathrm{T}}, \boldsymbol{U}^{\mathrm{T}})$.

To apply PCA on the combined fluctuations from all G_2 tiles, their terms V_{pt}^{vib} and ϕ_{pt}^{vib} are concatenated along the width of the interferometer and the index p is changed from locally on a G_2 tile to globally on the detector. With this definition of PCA, the joint principal components $\boldsymbol{\mathcal{B}} = (\mathcal{B}_{kp})$ and $\boldsymbol{\mathcal{C}} = (\mathcal{C}_{kp})$ (both $\in \mathbb{R}^{P \times T}$) over all G_2 tiles of the visibility and phase fluctuations are determined via

$$\mathrm{PCA}(V_{pt}^{\text{vib}}) \to (\mathcal{B}_{kp}, \beta_{kt}^{\star}) \, ; \; \mathrm{PCA}(\phi_{pt}^{\text{vib}}) \to (\mathcal{C}_{kp}, \gamma_{kt}^{\star}) \, . \tag{5}$$

The coefficients $\beta_{kt}^{\star}$ and $\gamma_{kt}^{\star}$ correspond only to the magnitude of principal component k at exposure t, not to the original polynomial model.

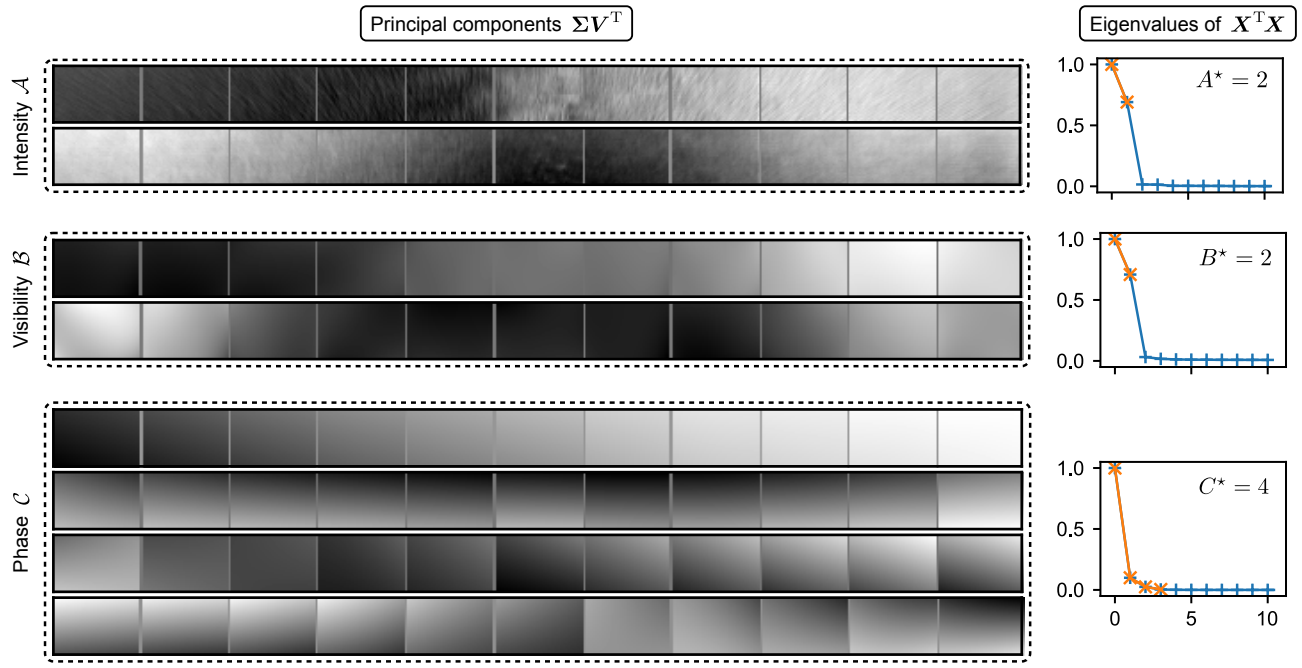


Fig. 2. The most dominant principal components of the spatial fluctuations in intensity, visibility, and phase (left). All images are scaled to [-1, 1]. Gray vertical lines correspond to G_2 tile boundaries. The size of each principal component is equal to the number of detector columns times rows. The number of principal components used for processing is determined by the scree plots of the respective (normalized) eigenvalues of $\boldsymbol{X}^{\mathrm{T}}\boldsymbol{X}$ obtained by the PCA (right). Two dominant components in intensity and visibility are identified based on the first "knee" in the scree plots. For the phase channel there is no distinct knee and a number of four principal components are chosen based on the impact on the reconstruction of the refractive index decrement. The vibrations in visibility and phase are a combination of global and tile-wise characteristics.

The fluctuations can now be approximated with a reduced number $B^\star$ and $C^\star$ of modes:

$$V_{pt}^{\mathrm{vib}} \approx \sum_{k=1}^{B^\star} \beta_{kt}^\star \mathcal{B}_{kp}\,;\;\; \phi_{pt}^{\mathrm{vib}} \approx \sum_{k=1}^{C^\star} \gamma_{kt}^\star \mathcal{C}_{kp}\,. \tag{6}$$

The dominant modes of the intensity fluctuations $\boldsymbol{\mathcal{A}} = (\mathcal{A}_{kp}) \in \mathbb{R}^{P\times T}$ are determined by applying PCA on the normalized residuum

$$I_{pt}^{\mathrm{vib}} = \frac{\hat{y}_{pt}}{y_{pt}} - 1\,;\;\; \mathrm{PCA}(I_{pt}^{\mathrm{vib}}) \to (\mathcal{A}_{kp}, \alpha_{kt}^\star)\,, \tag{7}$$

with the measured values $\hat{y}_{pt}$ and the predicted signal y_{pt} including only visibility and phase fluctuations. Again we approximate I_{pt}^{vib} with a reduced number $A^\star$ of modes:

$$I_{pt}^{\mathrm{vib}} \approx \sum_{k=1}^{A^\star} \alpha_{kt}^\star \mathcal{A}_{kp}\,. \tag{8}$$

III. Results

We show the results from processing an air scan with the proposed PCA method in Fig. 2. The number of eigenvalues of $\boldsymbol{X}^{\mathrm{T}}\boldsymbol{X}$ before the first "knee" in the scree plots of the principal components in intensity and visibility are used as $A^\star$ and $B^\star$, i.e. the number of vibration modes to keep for processing a sample scan. Because of the lack of a distinct knee in the phase channel, the number of modes $C^\star$ are increased until there is no noticeable difference in the reconstructed refractive index decrement. We obtain $A^\star = B^\star = 2$ and $C^\star = 4$. Especially the phase vibrations show tile-wise behavior, while substrate structure is visible in the intensity. The vertical gray lines correspond to the gaps between G_2 tiles.

The main goal of the vibration modeling is to facilitate artefact-free reconstruction. This requires a simultaneous estimation of vibration parameters and sample parameters, which is in general a difficult task and described in a different paper [9]. Here, we disentangle these two estimations by the following approach for demonstrating the accuracy of the model: The selected sample is small so that only the central part of the detector is covered. This allows us to fit the vibration parameters robustly by a least-squares fit to the outer parts of the detector for each exposure. Subsequently, object parameters are estimated using a sliding-window phase-retrieval [10].

The vibration modes in Fig. 2 are used on the scan of an object consisting of a Polyoxymethylene (POM) cylinder of 5 cm diameter and six falcon tubes with 3 cm diameter each, filled with wool at three different levels of dampness, chocolate chips, water, and neoprene, respectively.

One scan consists of 2400 exposures over a full 360° gantry rotation which takes 1 s. The X-ray tube is operated at 80 kVp and 550 mA. A subsequent air scan is used with the proposed reference processing method to extract the PCA vibration model shown in Fig. 2.

The results are compared with a simplified pipeline in which fluctuations in all channels are only modeled with polynomials $\mathcal{P}_{ijp}$ up to second order along width and height of the whole detector. Also in intensity, polynomials $\boldsymbol{\mathcal{P}}$ replace the dominant

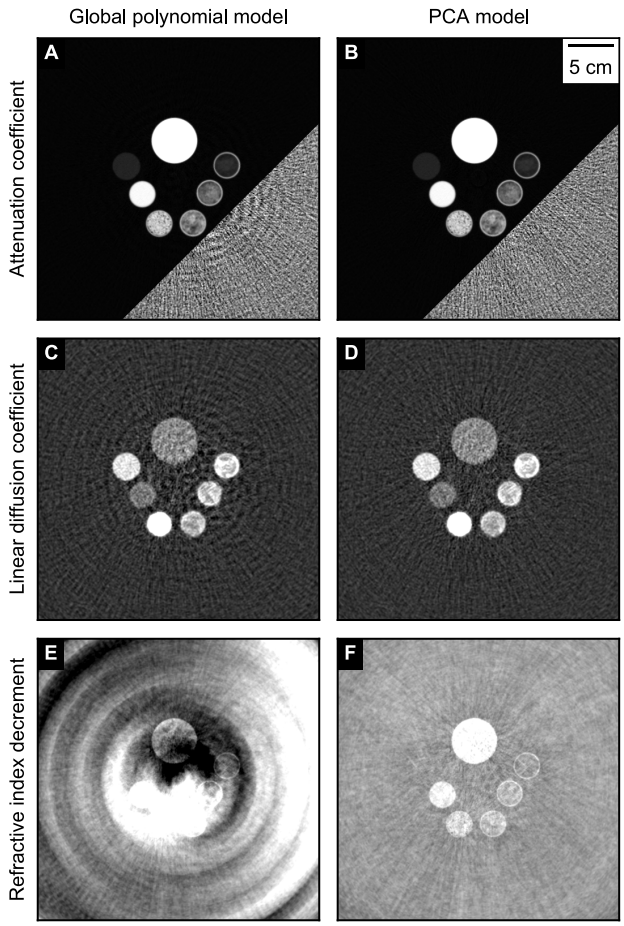


Fig. 3. Reconstruction of a phantom in attenuation coefficient (top), linear diffusion coefficient (middle), and refractive index decrement (bottom) with a global polynomial fluctuation model (left) in all channels and principal vibration components obtained via the proposed pipeline (right). All channels are free of artifacts when processed with the PCA model as shown in (B), (D) and (F). The materials of the cylinders are (clock-wise starting with the large cylinder) Polyoxymethylene (POM), dry wool, moist wool, soaked wool, chocolate chips, water, and neoprene. No beam-hardening correction is applied which leads to POM and water showing a non-zero diffusion coefficient. The windowing is (A), (B): [-0.05, 2.2] $\times$ 10^{-1} cm^{-1}; (C), (D): [-2, 8] $\times$ 10^{-2} cm^{-1}; (E), (F): [-4, 2] $\times$ 10^{-1} cm^{-1}. Small window in (A), (B): [-3, 3] $\times$ 10^{-3} cm^{-1}.

components $\mathcal{A}$ from PCA. This simplified pipeline does not handle tile-wise vibrations nor the intricate details of the intensity fluctuations shown in Fig. 2.

We show the central slices of the resulting volumes in attenuation, diffusion coefficient [11], and refractive index decrement [12] in Fig. 3. The reconstruction is performed via filtered back-projection [13]. The simplified pipeline using global polynomials suffers from artifacts in all channels. The attenuation and linear diffusion coefficient show roughly circular fringe artifacts. The refractive index decrement is dominated by concentric rings and a bright/dark blob structure in the center. None of these artifacts appear in the volumes reconstructed with vibrations from PCA.

IV. Conclusion

We proposed a processing scheme to identify and correct for vibrations of a Talbot-Lau interferometer mounted inside a rotating clinical CT gantry. The tile-wise vibrations are coupled by applying principal component analysis (PCA) and only the first few dominant components are used for processing a sample scan. In the intensity channel dominant fluctuation components are identified by PCA on the normalized residual. The resulting vibration model has few parameters per exposure for the intricate fluctuations, still allowing for an artifact-free reconstruction of a sample. A comparison with a vibration model using global polynomial fluctuations in all channels shows that they are not sufficient to capture the system's dynamics and lead to artifacts in the reconstruction of a sample.

Acknowledgment

The authors wish to thank Pascal Meyer, Jürgen Mohr, Frank Bergner, Maximilian von Teuffenbach, and Amanda Pleier for their support. This work was carried out with the support of the Karlsruhe Nano Micro Facility (KNMF, www.kit.edu/knmf), a Helmholtz Research Infrastructure at Karlsruhe Institute of Technology (KIT). We acknowledge the support of the TUM Institute for Advanced Study, funded by the German Excellence Initiative, the European Research Council (ERC, H2020, AdG 695045) and Philips GmbH Market DACH.

References

[1] A. Lohmann and D. Silva, "An interferometer based on the Talbot effect," *Opt. Commun.*, pp. 413–415, 1971.
[2] F. Pfeiffer *et al.*, "Hard-X-ray dark-field imaging using a grating interferometer," *Nat. Mater.*, pp. 134–137, 2008.
[3] A. Momose *et al.*, "Demonstration of X-ray Talbot interferometry," *Jpn. J. Appl. Phys., Part 2*, pp. L866–L868, 2003.
[4] F. Pfeiffer *et al.*, "Phase retrieval and differential phase-contrast imaging with low-brilliance X-ray sources," *Nat. Phys.*, pp. 258–261, 2006.
[5] T. Donath *et al.*, "Inverse geometry for grating-based x-ray phase-contrast imaging," *Journal of Applied Physics*, p. 054 703, 2009.
[6] M. Viermetz *et al.*, "Dark-field computed tomography reaches the human scale," PNAS, accepted.
[7] J. Dittmann, A. Balles, and S. Zabler, "Optimization based evaluation of grating interferometric phase stepping series and analysis of mechanical setup instabilities," *J. Imaging*, 77, 2018.
[8] F. Horn *et al.*, "Implementation of a Talbot-Lau interferometer in a clinical-like c-arm setup: A feasibility study," *Scientific Reports*, 2018.
[9] J. Haeusele *et al.*, "Dark-field for human CT – Sample data processing at the first clinical prototype," submitted to 2022 CT Meeting.
[10] I. Zanette *et al.*, "Trimodal low-dose x-ray tomography," *Proceedings of the National Academy of Sciences*, pp. 10 199–10 204, 2012.
[11] M. Bech *et al.*, "Quantitative x-ray dark-field computed tomography," *Physics in Medicine & Biology*, p. 5529, 2010.
[12] F. Pfeiffer *et al.*, "Hard x-ray phase tomography with low-brilliance sources," *Phys. Rev. Lett.*, p. 108 105, 2007.
[13] U. van Stevendaal *et al.*, "Reconstruction method for object-position dependent visibility loss in dark-field imaging," in *Medical Imaging 2013: Physics of Medical Imaging*, International Society for Optics and Photonics, 2013.

Fully Utilizing Contrast Enhancement on Lung Tissue as a Novel Basis Material for Lung Nodule Characterization by Multi-energy CT

Shaojie Chang, Yongfeng Gao, Marc J. Pomeroy, Ti Bai, Hao Zhang, and Zhengrong Liang

***Abstract*— Based on well-established X-ray physics in computed tomography (CT) imaging, the spectral responses of different materials contained in lesions are different, which brings richer contrast information at various energy bins. Hence, obtaining the material decomposition of different tissue types and exploring its spectral information for lesion diagnosis becomes extremely valuable. The lungs are housed within the torso and consist of three natural materials, i.e., soft tissue, bone, and lung tissue. To benefit the lung nodule differentiation, this study innovatively proposed to use lung tissue as one basis material along with soft tissue and bone. This set of basis materials will yield a more accurate composition analysis of lung nodules and benefit the following differentiation. Moreover, a corresponding machine learning (ML)-based computer-aided diagnosis framework for lung nodule classification is also proposed and used for evaluation. Experimental results show the advantages of the virtual monoenergetic images (VMIs) generated with lung tissue material over the VMIs without lung tissue and conventional CT images in differentiating the malignancy from benign lung nodules. The gain of 9.63% in area under the receiver operating characteristic curve (AUC) scores indicated that the energy-enhanced tissue features from lung tissue have a great potential to improve lung nodule diagnosis performance.**

***Index Terms*—Multi-energy CT reconstruction, Computer-aided diagnosis, Machine learning, Malignant and benign differentiation**

I. Introduction

In the conventional single energy CT, different tissue types can be represented by the linear attenuation coefficients (LACs). LACs are not unique for any given material and depend on the photon energies interacting with the material and the mass density of the material. Therefore, two different tissues may share similar intensity values, making it challenging to perform material decomposition based on the segmentation with the intensity values alone. Compared to the conventional single energy CT, additional measurements with a second or more energy (called multi-energy) allow the identification of two or three materials. Multi-energy CT also enables multiple materials (composition) analysis within the same region of interest, e.g., lesion. Based on well-established X-ray physics in CT imaging, the spectral responses of different materials contained in lesions are different, which brings richer contrast information at different energy bins. Hence, obtaining the material decomposition of different tissue types and exploring its spectral information for lesion diagnosis becomes extremely valuable.

The lungs as the primary organs of the respiratory system are essential for humans to breathe. Lung cancer is the leading cause of cancer death in the US. According to the report from World Health Organization, there are around 2.20 million deaths in 2020. CT has been recommended as an advanced non-invasive tool for cancer screening in the early stage, which provides fully three-dimensional (3D) information for volumetric-based lesion detection. Torso CT images consist of three natural materials, which are lung, bone, and soft tissue. To benefit the lung nodule differentiation, this study innovatively proposed to use lung tissue as one basis material along with soft tissue and bone. This set of three basis materials will yield a more accurate composition analysis of lung nodule and benefit the following differentiation.

In dual-energy CT (DECT), the scanned object can be decomposed into two basis materials (e.g., bone and soft tissue) with two datasets, one at high and the other at low energy [1]. To analyze more than two materials, one straightforward strategy is to obtain an additional measurement at one more energy, which we called triple energy CT (TECT). And it could provide three decomposed materials reconstructed from the three datasets. In the previous application, when using triple spectrums to decompose three materials, two materials are natural and the third is usually injected contrast agent (e.g., bone, soft tissue, and iodine) [2]. In this study, we use triple spectrums to decompose all-natural tissues: lung tissue, soft tissue, and bone. In addition, we also explore using two spectrums to decompose three materials, which can further lower the radiation dose. This paper will describe the material decomposition with the novel set of basis materials in detail. Moreover, a machine learning (ML)-based computer-aided diagnosis (CADx) framework for lung nodule classification is

This work was partially supported by the NIH/NCI grant #CA206171.

Shaojie Chang and Yongfeng Gao are with the Department of Radiology, Stony Brook University, Stony Brook, NY 11794, USA

Marc J. Pomeroy, and Zhengrong Liang are with the Departments of Radiology and Biomedical Engineering, Stony Brook University, Stony Brook, NY 11794, USA (e-mail: Jerome.Liang@SUNYSB.EDU).

Ti Bai is with the Department of Radiation Oncology, University of Texas Southwestern Medical Centre, Dallas, TX 75390, USA

Hao Zhang is with the Department of Medical Physics, Memorial Sloan Kettering Cancer Center, NY 10065, USA

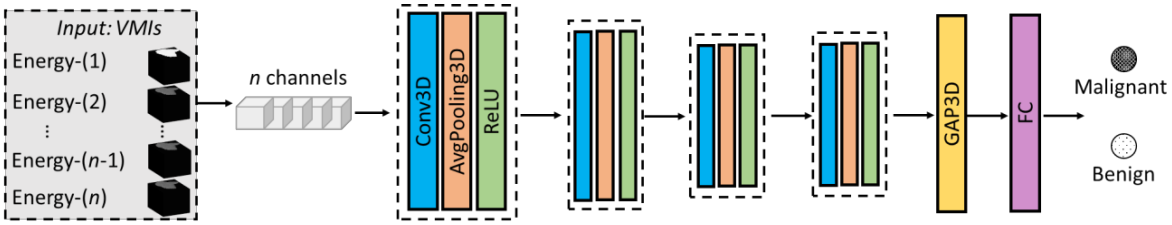


Fig. 2: The 3D CNN architecture of CADx for lung nodule diagnosis.

also proposed, which is based on the data-driven deep learning-based convolutional neural network (DL-CNN) method.

The remainder of this paper is organized as follows. Section II will describe the material decomposition method and the proposed computer-aided diagnosis framework. Section III presents the experiment design and results. Discussion and conclusions are drawn in Sections IV and V.

II. Methods

A. Material Decomposition

Considering the material composition, a linear attenuation coefficient (LAC) can be represented by R types of basis materials, e.g. soft tissue, bone, and so on. The LAC function $\mu_j(\varepsilon)$ at the j^{th} pixel of the image is decomposed as[3]:

$$\mu_j(\varepsilon) = \sum_{r=1}^{R} m_r(\varepsilon)\rho_{rj}, \quad (1)$$

where ρ_{rj} denotes the density of material r at the j^{th} pixel, $m_r(\varepsilon)$ represents the mass attenuation coefficient of material r at energy ε.

By the use of DECT, two basis materials of bone and soft tissue are usually considered as the basis materials in the material composition. From the decomposed basis materials of bone and soft tissue, a series of virtual monoenergetic images (VMIs) can be generated. As lungs become a major clinical concern, where the lung tissue has a distinct feature from the soft tissue, it is desirable to consider the lung tissue as a basis material in addition to the bone and soft tissue. This is the major motivation of this study. The material decomposition can be performed directly in the sinogram data domain or the image domain after the images are reconstructed from the sinogram data, where the image reconstruction is usually performed by a linear operator, such as FBP.

1) Image-domain material decomposition

Regarding the material decomposition in the image domain, Eq. (1) can be rewritten as,

$$\mu_j(\varepsilon) = \sum_{r=1}^{R} m_r(\varepsilon)\rho_{rj} = \sum_{r=1}^{R} m_r(\varepsilon)\rho_r f_{rj} = \sum_{r=1}^{R} \mu_r(\varepsilon) f_{rj}, \quad (2)$$

where ρ_r is the density of the material r and $\mu_r(\varepsilon)$ denotes linear attenuation coefficient of material r at energy ε. Notation f_{rj} is a unitless tissue fraction that quantifies the contribution of material r to attenuation in pixel j, which needs to be solved by the image domain decomposition methods. Theoretically, more unknowns need more equations to find the solution. For a conventional single energy spectral CT image, we only have one FBP CT image. For DECT, we have two FBP images from each of the two energy spectral data, respectively. From the two FBP images, we can obtain two basis material images. For TECT, e.g. 60kVp, 100kVp, and 140 kVp, we have three FBP images correspondingly.

By setting $R = 3$ for Eq. (2), we have three linear equations for three basis materials with the triple energy CT FBP images μ^{Eq} (Tube voltage = $E_q, q = 1,2,3$) as follows [4],

$$\mu_j^{E_1} = \mu_1(E_1^{eff})f_{1j} + \mu_2(E_1^{eff})f_{2j} + \mu_3(E_1^{eff})f_{3j}, \quad (3a)$$

$$\mu_j^{E_2} = \mu_1(E_2^{eff})f_{1j} + \mu_2(E_2^{eff})f_{2j} + \mu_3(E_2^{eff})f_{3j}, \quad (3b)$$

$$\mu_j^{E_3} = \mu_1(E_3^{eff})f_{1j} + \mu_2(E_3^{eff})f_{2j} + \mu_3(E_3^{eff})f_{3j}, \quad (3c)$$

where E_q^{eff} denotes the effective energy of the corresponding X-ray spectrum at each selected tube voltage E_q. $\mu_1(E_q^{eff})$, $\mu_2(E_q^{eff})$, and $\mu_3(E_q^{eff})$ describes the linear attenuation coefficient for basis material soft tissue, bone, and lung tissue at each effective energy, respectively.

Hence, the fractions f of each material can be solved by minimizing the mean squared difference between the calculated attenuation coefficients with (3a)-(3c) and the reconstructed attenuation coefficients at each energy as follows,

$$\emptyset(f) = \sum_{q=1}^{3}\sum_{j=1}^{J}(\mu_j^{E_q} - \sum_{r=1}^{3}\mu_r(E_q^{eff})f_{rj})^2, \quad (4)$$

Similar to [4], a grid search method can be utilized to optimize the objective function (4) to obtain the decomposed three materials.

B. Virtual Monoenergetic CT images Generation

From the three basis material images obtained by the image-domain or pre-log data domain, a series of VMIs at selected n energies are generated with the corresponding tissue mass attenuation coefficients as follows. For example, we can choose $n = 10$ with the energy values of 5, 8, 10, 12, 15, 20, 25, 30, 35, 40 and 45 keV, where the HU values in this energy range of 5-45 kVp have the maximum differences among the tissues to generate the tissue contrast features for CADx. Hence, the virtual monoenergetic CT images or VMIs in the range can be expressed by,

$$\mu_j^{virtual}(\varepsilon) = \sum_r \mu_r(\varepsilon) f_{rj}, \quad (5)$$

where $\varepsilon = [5,8,10,15,20,25,30,35,40,45]$, $\mu_r(\varepsilon)$ represents the linear attenuation coefficient of r tissue type.

C. Machine Learning (ML)-based CADx

1) Deep learning (DL)-based CNN model:

For the DL-based model, a 3D CNN architecture is designed, which has a multi-channel input with each energy image, as shown in Fig. 2. We first use four convolutional layers to extract the features, which are then pooled together with a global average pooling (GAP3D) layer such that the final features have a global receptive field. Then, we use a fully connected (FC) layer-based classifier to distinguish the malignant and benign lesions. Specifically, each convolutional layer consists of three operators: 3D convolution (Conv3D), 3D average pooling (Avgpooling3D), and rectified linear unit (ReLU). We adopt

the binary cross-entropy loss to train the model. The details of the CNN model are listed in Table I.

TABLE I: DETAILS OF NETWORK DESIGN

Layer	Type	Channels	Kernel Size	Padding	Stride	Activation
1	Conv3D	32	(7,7,7)	(1,1,1)	(1,1,1)	ReLU
2	Avgpooling3D	-	(2,2,2)	-	(2,2,2)	-
3	Conv3D	64	(5,5,5)	(1,1,1)	(1,1,1)	ReLU
4	Avgpooling3D	-	(2,2,2)	-	(2,2,2)	-
5	Conv3D	128	(3,3,3)	(1,1,1)	(1,1,1)	ReLU
6	Avgpooling3D	-	(2,2,2)	-	(2,2,2)	-
7	Conv3D	128	(3,3,3)	(1,1,1)	(1,1,1)	ReLU
8	Avgpooling3D	-	(2,2,2)	-	(2,2,2)	-
9	GAP3D	-	-	-	-	-
10	FC	128	-	-	-	-

Based on the above model, the final classification result will show the lung nodule diagnosis performance with the explored lung tissue material spectral information in VMIs.

III. EXPERIMENTS AND RESULTS

A. Datasets

114 patients were scheduled for CT-guided lung nodule needle biopsy with X-ray exposure of clinical dose at 120 kVp, 100 mAs in Stony Brook University Hospital, USA. With the pathological report, a total of 114 lung nodules with 50 benign and 64 malignant were confirmed. Each CT scan covers a portion of the patient's entire chest volume, resulting in 100–200 image slices of 512 × 512 array size, and each image voxel is nearly cubic with an edge size of 1 mm.

B. Multi-energy CT Scan Simulation

In this study, we simulated the DECT and TECT scans with the above dataset. First, each slice image in the datasets was segmented as basis materials (soft tissue/bone for DECT, soft tissue/bone/lung tissue for TECT) with a simple threshold method and transferred the linear attenuation coefficient to density by dividing the corresponding material linear attenuation coefficient at 75 keV, which is the equivalent energy of 120 kVp X-ray spectrum. Hence, the material fractions images f_r were obtained. Next, a poly-energetic forward model was used to simulate the DECT scan at 80/140 kVp and TECT scan at 60/100/140 kVp by (19):

$$\bar{N}_{wi} = \sum_{\varepsilon} S_w(\varepsilon)\, exp\left(-\sum_j A_{ij} \sum_{r=1}^{R} \mu_r(\varepsilon) f_{rj}\right). \quad (6)$$

An equidistant fan-beam geometry was assumed for a DECT scanner. The distance from the X-ray source to the isocenter is 570 mm and the distance from the source to the detector (SDD) is 1040 mm. 1160 projections from a full angle were acquired with 672 detector elements with a width of 1.4 mm per element. The X-ray spectrum at 60-, 80-, 100- and 140-kVp were generated by the SpekCalc software with 3mm Al filtration. Then, the multi-energy CT scan simulation was done. In this paper, all the reconstructed basis material fraction image slices have an array size of 512 × 512 with each pixel in the image slices covering an area of 1 mm × 1 mm. The value of β in the cost function (14) was empirically set to be 0.5 for the pre-log reconstruction.

C. CNN Training Implementation

For the input to the CNN-based implementation, we first converted each *n*-energy data with the resolution of 64×64×64 voxels. And these converted energy volumetric images were fed into the multi-channel 3D CNN as shown in Fig. 4 for training. And the target is the results from the pathological reports of the malignant and benign lesions. The k-fold (k=5) cross-validation was implemented to test the robustness and avoid data bias. The procedure is as follows. We firstly shuffled the dataset randomly and split it into 5 folds. For each fold, we randomly divided the dataset into training and testing datasets. And then we trained a model on the training dataset and evaluated it on the testing dataset. Finally, we retained the evaluation score for each fold and the average score was calculated. In this study, the CNN model was trained for 100 epochs with a learning rate of 0.001 and batch size of 8 using Adam optimizer [5].

D. Lung tissue material decomposition

Fig.4 shows the lung tissue material decomposition images from the TECT in the image domain. The result indicates that lung tissue can be reconstructed as one basis material with multi-energy CT data.

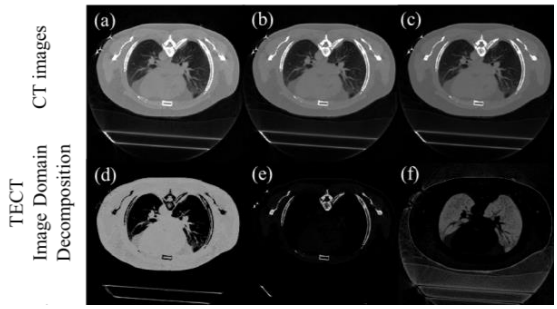


Fig. 4: Axis slice of CT images in one dataset: (a) 60 kVp, (b) 100 kVp, (c) 140kVp. Results of TECT image domain decomposition: (d) soft tissue, (e) bone, (f) lung tissue. CT images display window: [0,0.5] cm^{-1}, Tissue fraction images display window: [0,1.2].

E. Classification Performance

With the three decomposed material fraction images from the above, the contrast-enhanced 10-energy VMIs data are generated by (18) and incorporated into the CADx part of our proposed ML-based methods. Meanwhile, to further verify the significance of lung tissue as the basis material, the traditional decomposed materials (soft tissue/bone) from DECT are also adopted to generate the VMIs for ML-based CADx. The conventional single CT data is also applied to CADx for comparison. For simplicity, we define the material decomposition methods in the image domain of DECT as **DE-FBP**, the corresponding methods of TECT as **TE-FBP**. To illustrate the effectiveness and the generality of our ML-based CADx, the AUC of each dataset was calculated.

1) DL-based CNN model

TABLE II: MEAN AUC VALUES FOR DL-BASED CNN DIAGNOSIS MODEL

Data	METHOD	Lung tissue as basis material	AUC
Conventional CT (120kVp)	FBP	-	59.54
DECT images	DE-FBP	×	65.53
TECT images	**TE-FBP**	✓	**69.17**

For the DL-based CNN model, the conventional CT and the enhanced 10-energy VMI data w/ and w/o lung tissue basis material generated by different multi-energy data were incorporated into our 3D-CNN network, respectively. We

calculated the mean values of AUC scores, which are shown in Table II. We have the following observations. First, the CADx outcome from the conventional CT is consistent with the expectation that all the nodules are undifferentiable to be benign or malignant by human experts who recommended biopsy on these nodules. By the reports of their biopsies as the ground truth, the CADx with 59.54% AUC score performed slightly better than human experts, whose score would be random or 50%. Using the outcome from the conventional CT as the baseline, the CADx outcome from multi-energy CT data achieves higher mean AUC values than the conventional 120 kVp data, which verifies the effectiveness of the contrast enhancement brought from the VMIs. Second, by considering the lung tissue as a basis material, the results show that the mean AUC values improved 3.64% for lung nodule characterization in comparison with the image-domain method by multi-energy CT. The main reason why the VMIs with lung tissue perform well is that the lung tissue has a distinct feature from the soft tissue and brings richer information for DL-based diagnosis.

IV. Discussion

In this study, we proposed to use lung tissue as one basis material along with soft tissue and bone with the multi-energy CT material decomposition for lung nodule characterization. The proposed CADx framework with lung tissue basis material fully utilizes the contrast enhancement at each energy bin and improves the diagnosis performance. However, there are opportunities for further refinements. First, how the decomposed image quality affects the CADx performance would be a very interesting future research. Second, for the VMIs generation, we selected 10 energies as an example. Ideally, when the basis material fraction images are well reconstructed, an arbitrary number of VMIs could be generated. Hence, it remains an interesting topic to analyze the effect of the selected energy number on the CADx performance and optimize the range and number of energies in the range. Third, the feature quality of enhanced energy images varies among each other and not all of them are guaranteed to contain meaningful features to improve the CADx performance. Exploring the meaningful features from the VMIs and developing an adaptive learning model should also be further investigated as future work. Finally, clinical evaluations with more human lung data sets are needed to test the robustness of the proposed method.

V. Conclusions

In conclusion, this study proposed an end-to-end computer-aided diagnosis framework for lung nodule characterization by multi-energy CT, which fully utilizes contrast enhancement on lung tissue as a basis material in reconstruction. Experimental results demonstrated that the lung tissue is able to be decomposed by multi-energy CT data with either the image-domain. The results also indicated the advantages of the VMIs generated with lung tissue material over the VMIs without lung tissue and conventional CT images in ML-based lung nodule classification.

References

[1] S. Chang et al., "Spectrum estimation-guided iterative reconstruction algorithm for dual energy ct," IEEE transactions on medical imaging, vol. 39, no. 1, pp. 246-258, 2019.

[2] W. Zhao et al., "A unified material decomposition framework for quantitative dual‐and triple‐energy CT imaging," Medical physics, vol. 45, no. 7, pp. 2964-2977, 2018.

[3] M. Abella, C. Martínez, M. Desco, J. J. Vaquero, and J. A. Fessler, "Simplified Statistical Image Reconstruction for X-ray CT With Beam-Hardening Artifact Compensation," IEEE transactions on medical imaging, vol. 39, no. 1, pp. 111-118, 2019.

[4] P. Granton, S. Pollmann, N. Ford, M. Drangova, and D. Holdsworth, "Implementation of dual‐and triple‐energy cone‐beam micro‐CT for postreconstruction material decomposition," Medical physics, vol. 35, no. 11, pp. 5030-5042, 2008.

[5] D. P. Kingma and J. Ba, "Adam: A method for stochastic optimization," arXiv preprint arXiv:1412.6980, 2014.

Image Reconstruction in Phase-Contrast CT with Shortened Scans

Zheng Zhang[1], Buxin Chen[1], Dan Xia[1], Emil Y. Sidky[1], Mark Anastasio[3], and Xiaochuan Pan[1,2]
[1]*Department of Radiology, The University of Chicago, USA*
[2]*Department of Radiation and Cellular Oncology, The University of Chicago, USA*
[3]*Department of Bioengineering, The University of Illinois at Urbana-Champaign, USA*

***Abstract*—Phase-contrast CT (PCCT) is an emerging tool that has found numerous applications, including applications to preclinical imaging. There remains a need for reducing the imaging time in current PCCT. One approach to reducing imaging time is to reduce the scanning angular range in PCCT. However, accurate image reconstruction from data collected over a limited angular range (LAR) is challenging because it poses a problem of accurate inversion of the PCCT imaging model that can be highly ill-conditioned in LAR scans. In this work, we conduct an investigation of accurate image reconstruction through inverting the imaging model for LAR scanning configurations in propagation-based (PB) PCCT. We have developed a directional-total-variation (DTV) algorithm for image reconstruction from knowledge of the discrete X-ray transform (DXT) over a LAR for CT imaging. Observing the mathematical similarity between the DXT in CT and the imaging model in PB-PCCT, we develop and tailor the DTV algorithm for image reconstruction from LAR data in PB-PCCT. Results of our study show that the tailored DTV algorithm can yield image reconstruction with reduced LAR artifacts that can be observed otherwise in images reconstructed by use of the existing algorithm in PB-PCCT imaging. For a given LAR, it can be divided into sub arcs of LARs. We also investigate a scanning configuration with two orthogonal arcs of LARs separated by 90°, and observe that the two-orthogonal-arc scanning configuration may allow image reconstruction more accurately than does a single-arc scanning configuration even though the total angular ranges in both scanning configurations are identical. While boundary images can be reconstructed from data, we develop the DTV algorithm for reconstruction of the image, i.e., the refractive index distribution, instead of its boundary image from data in PB-PCCT. Once the image is obtained, the Laplacian operator can be applied to it for yielding its boundary image.**

***Index Terms*—limited-angular range (LAR), directional total variation, primal-dual algorithm, phase contrast CT (PCCT)**

I. Introduction

Physical quantities such as the refractive index within an object scanned can be estimated in phase-contrast CT (PCCT) for yielding unique contrast mechanisms differing from that in conventional CT. It is believed that PCCT may hold potential value for certain preclinical and clinical applications. However, a number of issues need to be addressed adequately before its wide adoption in preclinical and clinical applications. One of the practical constraints that PCCT faces is its considerably long imaging time. There exist studies on the minimization of the total imaging time by reducing the number of scanning views while maintaining the imaging time unchanged at each view in PCCT [1]–[3]. In this study, we investigate a different approach to reducing the imaging time in PCCT by lowering its scanning angular range to a limited angular range (LAR) that is less than the full angular range of π. A PCCT scan over a LAR can readily be implemented in PCCT without invoking any hardware changes simply by stopping X-ray illumination and data collection beyond the LAR prescribed. However, accurate image reconstruction from such LAR data is challenging because it poses a problem of accurate inversion of the PCCT imaging model that can be highly ill-conditioned under LAR conditions.

In this work, we conduct an investigation of accurate image reconstruction from LAR data by inverting the imaging model in propagation-based (PB) PCCT. We have developed a directional-total-variation (DTV) algorithm [4] for image reconstruction from knowledge of the discrete X-ray transform (DXT) over a LAR for CT imaging. Observing the mathematical similarity between the DXT in CT and the imaging model in PB-PCCT, we develop and tailor the DTV algorithm for image reconstruction from LAR data in PB-PCCT. For a given LAR, it can be divided into sub arcs of LARs. We thus also investigate a scanning configuration with two orthogonal arcs of LARs [5]–[7], which are of equal angular ranges but their centers are separated by 90°. For simplicity, we refer to a configuration with two orthogonal arcs of LARs as the two-orthogonal-arc configuration, whereas the configuration with a single arc of LAR as a single-arc configuration.

In many of the existing works in PB-PCCT, algorithms have been developed only for reconstruction of the image boundaries, as shown in Fig. 1b, instead of the image itself in Fig. 1a. Unlike these existing algorithms, the DTV algorithm developed in this work reconstructs the image, i.e., the refractive index distribution, from LAR data in PB-PCCT. From the reconstructed image, one can then readily obtain the boundary image.

II. Materials and Methods

A. Imaging model in PB-PCCT

In PB-PCCT, assuming the object's attenuation of X-ray is weak, the 2D imaging model can be expressed as [1], [8]

$$\mathbf{g} = \mathcal{H}\Delta\mathbf{f}, \tag{1}$$

where vector $\mathbf{g}$ of size M denotes model data in PB-PCCT; vector $\mathbf{f}$ of size N denotes a 2D discrete image of the refractive index within the imaged object; $\Delta = \frac{\partial^2}{\partial x^2} + \frac{\partial^2}{\partial y^2}$ is the discrete

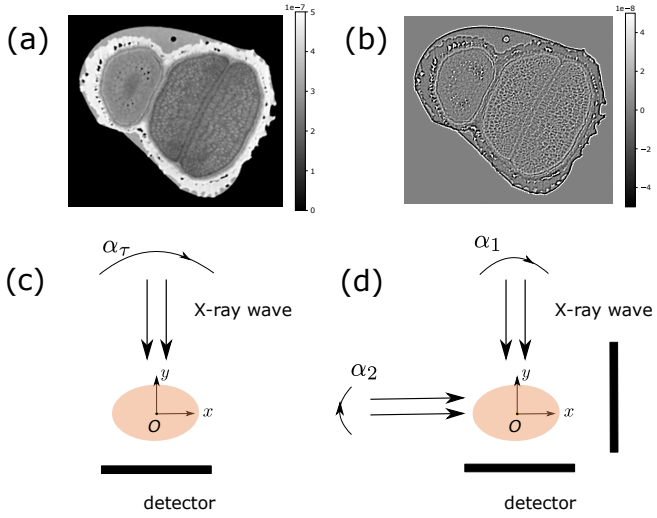


Fig. 1. (a) A numerical, realistic plant-seed phantom, i.e., truth image $\mathbf{f}^{[\text{truth}]}$, from which PB-PCCT data are generated. Display window: [0.0, 5×10^{-7}] arbitrary units (AU); (b) a boundary image generated by applying the discrete Laplacian operator to the plant-seed phantom in panel (a). Display window: [-5×10^{-8}, 5×10^{-8}] AU; (c) single-arc configuration with LAR of α_τ, and (d) two-orthogonal-arc configuration in which the two sub-arcs of LARs are symmetric, respectively, relative to the y- and x-axis, covering equal angular ranges α_1 and α_2, where $\alpha_\tau = \alpha_1 + \alpha_2$.

Laplacian operator on the image space; and system matrix $\mathcal{H}$ of size $M \times N$ represents a discrete Radon transform from image space to data space. In this work, element h_{ij} of $\mathcal{H}$ denotes the intersection between ray j and pixel i. For given $\mathcal{H}$ corresponding to a LAR scanning configuration discussed below, we use Eq. (1) to generate model data $\mathbf{g}$ from numerical, realistic plant-seed phantom $\mathbf{f}$ on a grid of 190×200 square pixels of size 0.12 mm in Fig. 1a [9].

B. Scan configuration

We consider a single-arc and a two-orthogonal-arc scanning configurations [7], as shown in Figs. 1c and 1d, with a parallel-beam projection geometry, and the discrete Radon transform is used as the imaging model in PB-PCCT. We also assume that the former covers LAR of α_τ and that the latter includes two sub-arcs of LARs of α_1 and α_2, symmetric relative, respectively, to the y- and x-axis. In the work, without loss of generality, we consider the case with $\alpha_\tau = \alpha_1 + \alpha_2$ and $\alpha_1 = \alpha_2$. Furthermore, the detector array consists of 512 bins of size 0.046 mm. We generate noiseless data from the plant-seed phantom (i) with the single-arc configurations of $\alpha_\tau = 60°$, $90°$, $120°$, and $150°$, with an angular interval of 0.5° per view, and (ii) with the two-orthogonal-arc configurations of $2\alpha_1 = 2\alpha_2 = 60°$, $90°$, $120°$, and $150°$. Based upon the noiseless data, we also generate noisy data which is equivalent to adding Gaussian noise to the intensity data [8], and the Gaussian noise has zero mean and standard deviation of 1.5% of the intensity for each detector bin.

C. Directional-TV reconstruction algorithm

Image reconstruction in PB-PCCT is equivalent to inverting the imaging model in Eq. (1). It is well-known that, while existing analytic algorithms can accurately reconstruct boundary images from data collected over FAR of $180°$, they cannot yield accurate boundary images nor images from data collected over LARs that are considerably lower than $180°$. Instead, we formulate the reconstruction problem from LAR data in PB-PCCT as an optimization problem and tailor the DTV algorithm to obtain images through solving the optimization problem.

We consider a convex optimization program containing a data-ℓ_2-norm fidelity and DTV constraints along the x- and y-directions, which is given by [4]

$$\mathbf{f}^\star = \underset{\mathbf{f}}{\text{argmin}} \Phi(\mathcal{H}\Delta\mathbf{f} - \mathbf{g}^{[\mathcal{M}]}) \quad \text{s.t.} \quad \Psi(\mathbf{f}) \tag{2}$$

where vector $\mathbf{g}^{[\mathcal{M}]}$ of size M denotes discrete measured data; $\Phi(\mathcal{H}\Delta\mathbf{f} - \mathbf{g}^{[\mathcal{M}]})$ denotes the ℓ_2-norm of the difference between $\mathcal{H}\Delta\mathbf{f}$ and measured data $\mathbf{g}^{[\mathcal{M}]}$; $\Psi(\mathbf{f})$ is the constraint term in the optimization program which includes three components: image DTV constraint along x-axis, image DTV constraint along y-axis, and image non-negativity constraint; the DTV constraint applies an upper bound on an image DTV, which is defined as ℓ_1-norm of a vector obtained by calculating two-point differences of $\mathbf{f}$ along either x- or y-axis; and the image non-negativity constraint is to enforce non-negative values for each pixel in the image. In many of existing works [1], [10], [11], algorithms were developed largely for inverting $\mathcal{H}$ in the imaging model, thus yielding the boundary image $\Delta\mathbf{f}$, instead of image $\mathbf{f}$ itself. In this work, we develop and tailor the DTV algorithm to solve Eq. (2) directly for obtaining image $\mathbf{f}$. Once image $\mathbf{f}$ is obtained, one can readily apply the Laplacian to it to obtain a boundary image, as shown in Fig. 1b. For references, we also reconstruct images by using the filtered-backprojection (FBP) algorithm [8], [12].

For evaluating image reconstructed, we first perform visual inspection of the images reconstructed. Additionally, we also perform a quantitative evaluation of image reconstructed from data over LAR by using the normalized root-mean-square-error (nRMSE) metric

$$\text{nRMSE}(\mathbf{f}^{[\text{recon}]}) = ||\mathbf{f}^{[\text{recon}]} - \mathbf{f}^{[\text{truth}]}||_2 / ||\mathbf{f}^{[\text{truth}]}||_2, \tag{3}$$

where $\mathbf{f}^{[\text{recon}]}$ and $\mathbf{f}^{[\text{truth}]}$ denote the reconstructed and truth images. Metric nRMSE provides measures of quantitative accuracy between the reconstructed and truth images. We note that $\text{nRMSE}(\mathbf{f}^{[\text{recon}]})$ approaches 0 as $\mathbf{f}^{[\text{recon}]}$ approaches $\mathbf{f}^{[\text{truth}]}$.

III. Results

A. Reconstruction from noiseless data

We first conducted image reconstruction, as shown in Fig. 2, by using the DTV and FBP algorithms from noiseless data with the two-orthogonal-arc and single-arc configurations over total angular ranges $\alpha_\tau = 60°$, $90°$, $120°$, and $150°$. In this case, measured data and model data are identical,

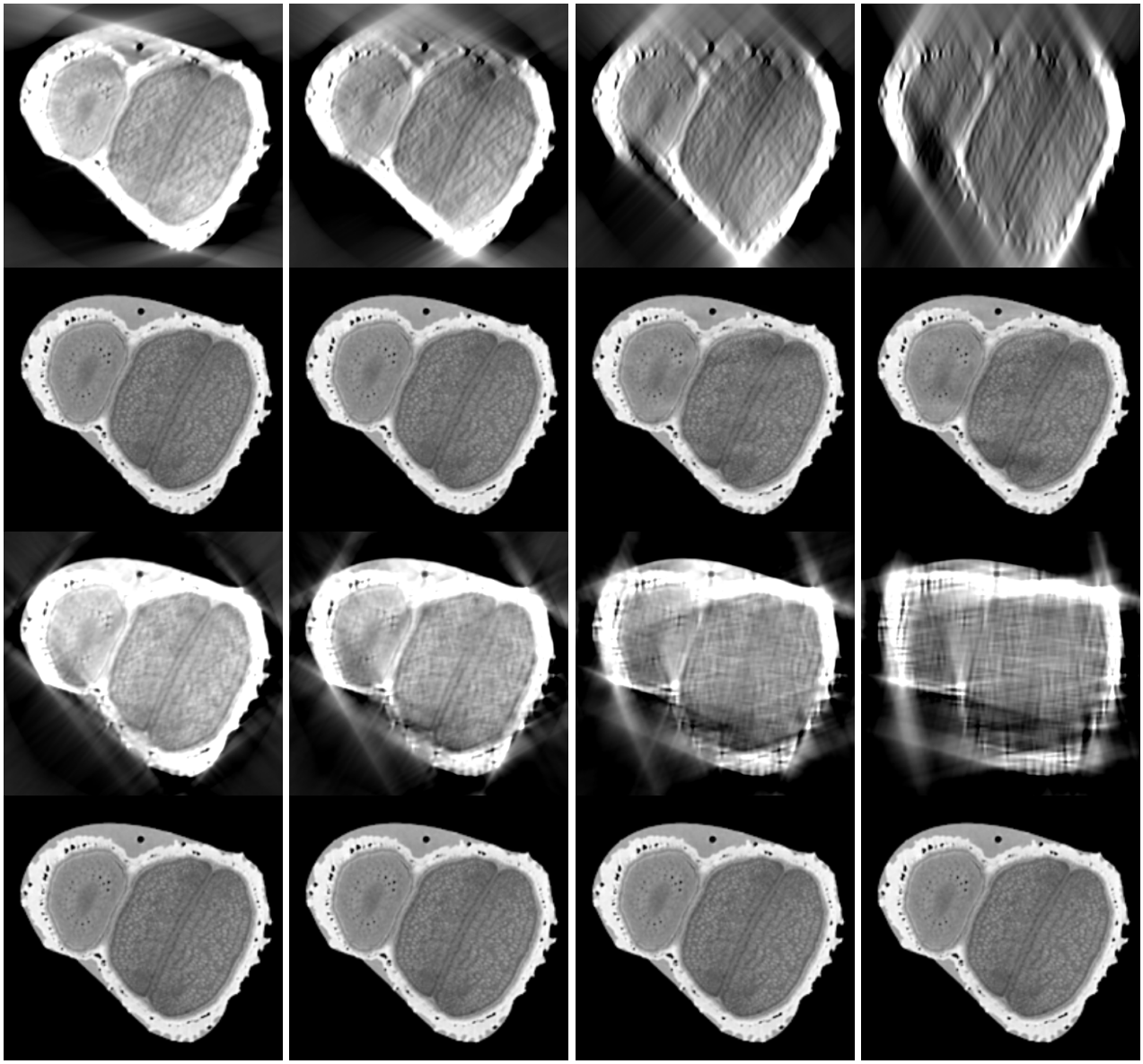


Fig. 2. Images reconstructed by use of the FBP (rows 1&3) and DTV (rows 2&4) algorithm from noiseless data for single-arc configurations (rows 1&2) of $\alpha_\tau = 150°$ (column 1), $120°$ (column 2), $90°$ (column 3), and $60°$ (column 4), respectively, and for two-orthogonal-arc configurations (rows 3&4) of $\alpha_1 = \alpha_2 = 75°$ (column 1), $60°$ (column 2), $45°$ (column 3), and $30°$ (column 4), respectively. Display window: [0.0, $5{\times}10^{-7}$] AU.

TABLE I
nRMSEs of images reconstructed from noiseless data generated with single-arc and two-orthogonal-arc configurations of different α_τ.

Configuration \ LAR	150°	120°	90°	60°
Single-arc	2.56×10^{-5}	3.85×10^{-4}	1.27×10^{-2}	2.71×10^{-2}
Two-orthogonal-arc	8.40×10^{-7}	1.63×10^{-5}	1.44×10^{-4}	2.21×10^{-3}

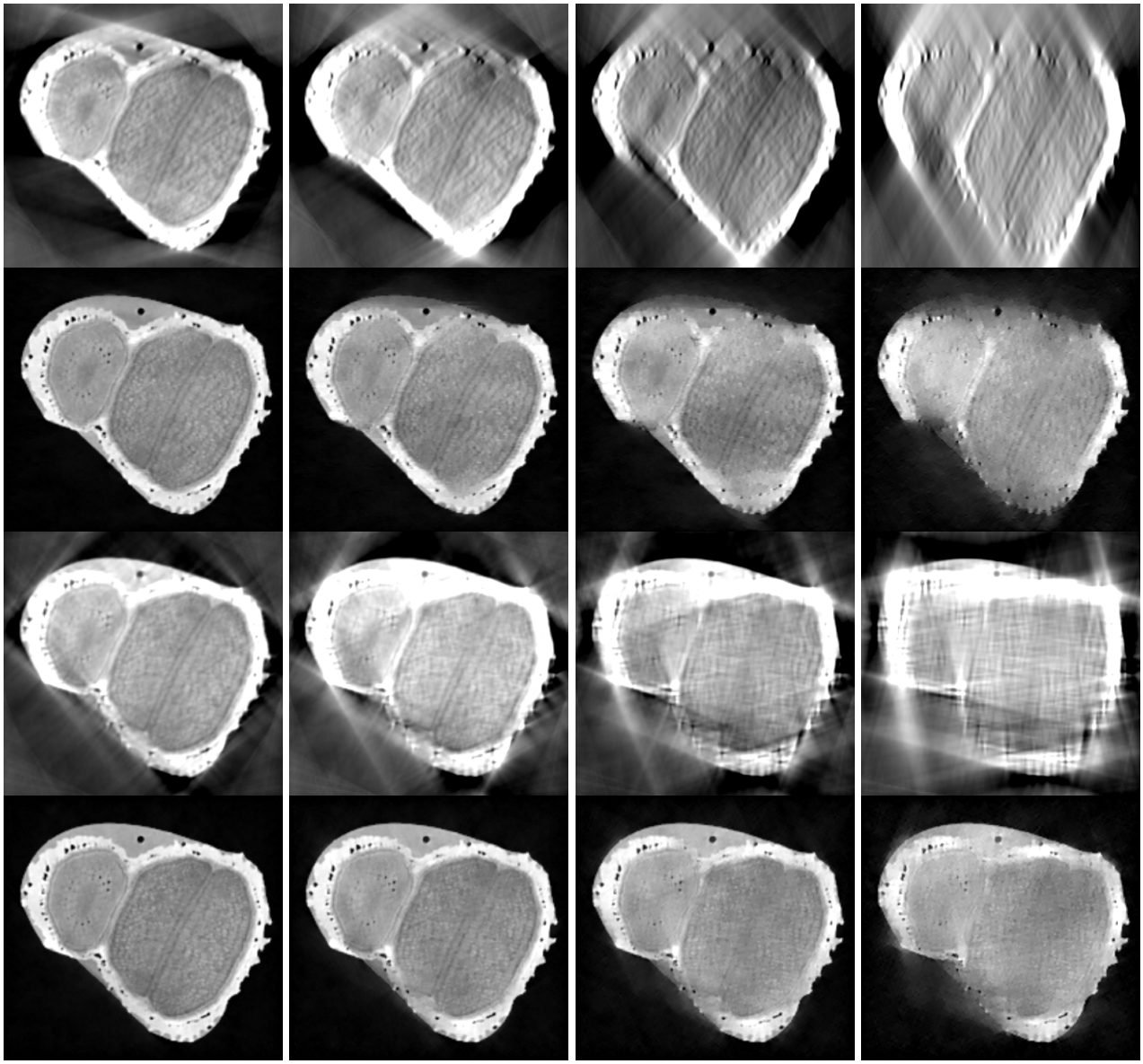


Fig. 3. Images reconstructed by use of the FBP (rows 1&3) and DTV (rows 2&4) algorithm from noisy data for single-arc configurations (rows 1&2) of $\alpha_\tau = 150^\circ$ (column 1), 120° (column 2), 90° (column 3), and 60° (column 4), respectively, and for two-orthogonal-arc configurations (rows 3&4) of $\alpha_1 = \alpha_2 = 75^\circ$ (column 1), 60° (column 2), 45° (column 3), and 30° (column 4), respectively. Display window: [0.05, 0.9].

TABLE II
nRMSEs of the DTV reconstructions from noisy data generated with single-arc and two-orthogonal-arc configurations of different α_τ.

Configuration \ LAR	150°	120°	90°	60°
Single-arc	0.134	0.187	0.269	0.392
Two-orthogonal-arc	0.118	0.167	0.220	0.283

i.e., $\mathbf{g}^{[\mathcal{M}]} = \mathbf{g} = \mathcal{H}\Delta\mathbf{f}$. We first observe that FBP reconstructions from all LAR data contain significant artifacts such as leakage and distortion. It can also be observed that all DTV reconstructions of the two-orthogonal-arc configurations visually resemble the truth plant-seed phantom. However, DTV reconstructions of the single-arc configurations show visible artifacts for $\alpha_\tau \leq 90^\circ$. We also calculated nRMSEs of the DTV reconstructions and list the corresponding values in Table I. It can be observed that the nRMSE increases as total-angular range α_τ decreases. In addition, the nRMSEs of DTV reconstructions for two-orthogonal-arc configurations are generally smaller than those of the single-arc configurations for a given α_τ.

B. Reconstruction from noisy data

We then conducted image reconstruction with the DTV and FBP algorithms from noisy data $\mathbf{g}^{[\mathcal{M}]}$, as described in Sec. II-B, by using the two-orthogonal-arc and the single-arc configurations covering total angular ranges $\alpha_\tau = 60^\circ$, 90°, 120°, and 150°, and show results in Fig. 3. Similarly to the noiseless-data study, there exist significant LAR artifacts in all FBP reconstructions. For the DTV reconstructions, in general, as the total angular coverage α_τ decreases, LAR artifacts increase in image reconstructions. We observe that DTV reconstructions with the two-orthogonal-arc configuration outperform those with the single-arc configuration in terms of artifacts reduction. This is understandable because the data model is less ill-conditioned for the former than for the latter, and thus the reconstruction of the former is less sensitive to data inconsistency, i.e., data noise, than that of the latter. In particular, DTV reconstructions with single-arc configurations considerably deteriorate for $\alpha_\tau \leq 90^\circ$, and some detailed structures cannot be identified. DTV reconstructions with two-orthogonal-arc configurations, however, can still reveal most of the structures for $\alpha_\tau = 60^\circ$. We also conduct a quantitative analysis of the DTV reconstructions by computing their nRMSEs relative to the truth phantom, and show results in Table II. We notice that, for DTV reconstructions from LAR data containing noise, the nRMSEs of reconstructions with the two-orthogonal-arc configuration are smaller than those of reconstructions with the single-arc configuration. Additionally, the nRMSE increases as the total angular range decreases.

IV. Discussion

In the work, we have investigated image reconstruction from LAR data in PB-PCCT by developing and tailoring the DTV algorithm that was developed previously for image reconstruction from LAR data in CT. We note that the DTV algorithm developed reconstructs the image, i.e., the refractive index distribution, and that the application of the Laplacian operator to the image subsequently yields its boundary image. We have studied and compared images reconstruction for single-arc and two-orthogonal-arc configurations of different LARs. Results of our study show that the DTV algorithm developed can reconstruct images, instead of their boundary images, directly from data in PB-PCCT, and that the DTV images from LAR data are with reduced LAR artifacts observed otherwise in images reconstructed with other algorithms. Furthermore, observations can also be made that the two-orthogonal-arc configurations can further improve reconstruction accuracy of the single-arc configurations with $\alpha_\tau = \alpha_1 + \alpha_2$, especially as α_τ is reduced. This work may provide insights into the design of LAR scanning configurations for potentially reducing imaging time in PCCT. In the future, it is of interest to investigate image reconstructions from LAR data with additional noise and also for two-orthogonal-arc configurations with $\alpha_1 \neq \alpha_2$. Moreover, the DTV algorithm can also be applied to other PCCT imaging techniques to enabling LAR imaging, such as analyzer-crystal-based PCCT and grating-based PCCT.

Acknowledgment

This work was supported in part by NIH Grant Nos. R01-EB026282, R01-EB023968, and 1R21CA263660-01A1. The contents of this article are solely the responsibility of the authors and do not necessarily represent the official views of the National Institutes of Health.

References

[1] E. Y. Sidky, M. A. Anastasio, and X. Pan, "Image reconstruction exploiting object sparsity in boundary-enhanced X-ray phase-contrast tomography," *Opt. Expr.*, vol. 18, no. 10, pp. 10 404–10 422, 2010.

[2] L. Wang, X. Li, M. Wu, L. Zhang, and S. Luo, "A sparse-projection computed tomography reconstruction method for in vivo application of in-line phase-contrast imaging," *Biomed. Engin. Onlin.*, vol. 12, no. 1, pp. 1–13, 2013.

[3] T. Gaass, G. Potdevin, M. Bech, P. Noël, M. Willner, A. Tapfer, F. Pfeiffer, and A. Haase, "Iterative reconstruction for few-view grating-based phase-contrast CTAn in vitro mouse model," *Europhy. Lett.*, vol. 102, no. 4, p. 48001, 2013.

[4] Z. Zhang, B. Chen, D. Xia, E. Y. Sidky, and X. Pan, "Directional-TV algorithm for image reconstruction from limited-angular-range data," *Medical Image Analysis*, p. 102030, 2021.

[5] B. Chen, Z. Zhang, D. Xia, E. Y. Sidky, and X. Pan, "Dual-energy CT imaging with limited-angular-range data," *Phys. Med. Biol.*, vol. 66, no. 18, p. 185020, 2021.

[6] ——, "Dual-energy CT imaging over non-overlapping, orthogonal arcs of limited-angular ranges," *J. X-Ray Sci. and Technol.*, vol. 29, no. 6, pp. 975–985, 2021.

[7] Z. Zhang, B. Chen, D. Xia, E. Y. Sidky, and X. Pan, "Image reconstruction from data over two orthogonal arcs of limited-angular ranges," *Med. Phys.*, pp. –In Press, 2022.

[8] A. V. Bronnikov, "Theory of quantitative phase-contrast computed tomography," *JOSA A*, vol. 19, no. 3, pp. 472–480, 2002.

[9] (2018) MITOS. [Online]. Available: https://www.mitos.de/en/gallery.html

[10] T. Köhler, B. Brendel, and E. Roessl, "Iterative reconstruction for differential phase contrast imaging using spherically symmetric basis functions," *Med. Phy.*, vol. 38, no. 8, pp. 4542–4545, 2011.

[11] M. Nilchian, C. Vonesch, P. Modregger, M. Stampanoni, and M. Unser, "Fast iterative reconstruction of differential phase contrast X-ray tomograms," *Opt. Expr.*, vol. 21, no. 5, pp. 5511–5528, 2013.

[12] T. E. Gureyev, G. R. Myers, Y. I. Nesterets, D. M. Paganin, K. M. Pavlov, and S. W. Wilkins, "Stability and locality of amplitude and phase contrast tomographies," in *Proc. SPIE Develop. in X-Ray Tomogr. V*, vol. 6318. SPIE, 2006, pp. 278–292.

Self-trained Deep Convolutional Neural Network for Noise Reduction in CT

Zhongxing Zhou, Akitoshi Inoue, Cynthia McCollough, and Lifeng Yu

***Abstract*—Supervised deep convolutional neural network (CNN)-based methods have been actively used in clinical CT to reduce image noise. The networks of these methods are typically trained using paired high- and low-quality data from a large number of patients and/or phantom images. This training process is tedious, and the network trained under a given condition may not be generalizable to patient images acquired and reconstructed at different conditions. In this paper, we propose a self-trained deep CNN (ST_CNN) method which does not rely on pre-existing training datasets. The training is accomplished using extensive data augmentation through projection domain and the inference is applied to the data itself. Preliminary evaluation on patient images demonstrated that the proposed method could achieve similar image quality in comparison with conventional deep CNN denoising methods pre-trained on external datasets.**

***Index Terms*— Computed tomography (CT), low-dose CT, deep learning, convolutional neural network (CNN), supervised training**

I. Introduction

In recent years, Deep convolutional neural network (CNN) has been one of the main driving forces for CT image denoising [1-7]. A majority of existing CNN-based denoising methods are supervised, which learn the mapping function between the low-quality image (e.g., low dose) and its high-quality (e.g., high dose) counterpart [1-5]. In order to have a CNN denoiser that generalizes well to new patient data, a large amount of low-/high-quality image pairs from a large number of patient and/or phantom images are needed to sufficiently cover the data distribution. However, this training process is costly, and the model trained from one dataset may not generalize well to another dataset acquired or reconstructed at a different condition. The inter-patient differences can also make it challenging to learn a model that generalizes well across patients.

To tackle this challenge, we propose a self-trained deep CNN (ST_CNN) method for noise reduction in CT which does not rely on pre-existing training datasets. This method trains the network directly using the data itself through extensive data augmentation (random rotation and noise addition) through projection domain and the inference is applied to the data itself. We demonstrated that this method could achieve similar performance as conventional deep CNN denoising methods trained on external datasets. There are three major potential benefits of this method. First, by removing the need of a large number pre-existing training dataset, it can be applied to any CT data, even if the data condition were not previously trained. Second, the self-training mechanism eliminates the generalizability issue that may occur for network models applied to datasets that are different from the training datasets. Third, the trained model can be applied to and finetuned for each individual patient if repeated CT exams are expected, which may maximize the benefit of image quality improvement and radiation dose reduction.

Corresponding author: Lifeng Yu

Zhongxing Zhou, Cynthia McCollough, and Lifeng Yu are with Department of Radiology, Mayo Clinic, Rochester, MN, 55905, USA (e-mail: zhou.zhongxing@mayo.edu; mccollough.cynthia@mayo.edu; yu.lifeng@mayo.edu).

II. Methods

The proposed ST_CNN method belongs to the family of image-domain supervised deep learning techniques, but there is a distinct difference in the training scheme from the existing approaches, as described in Figs. 1 and 2. The availability of sufficient amount of patient cases for training is one key factor contributing to the performance of the conventional supervised deep learning methods.

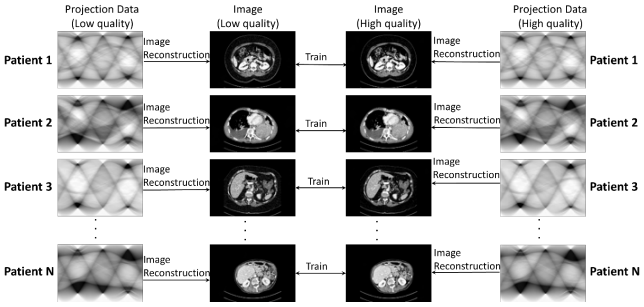


Fig. 1. Overview of the training scheme for conventional image-domain supervised deep learning techniques

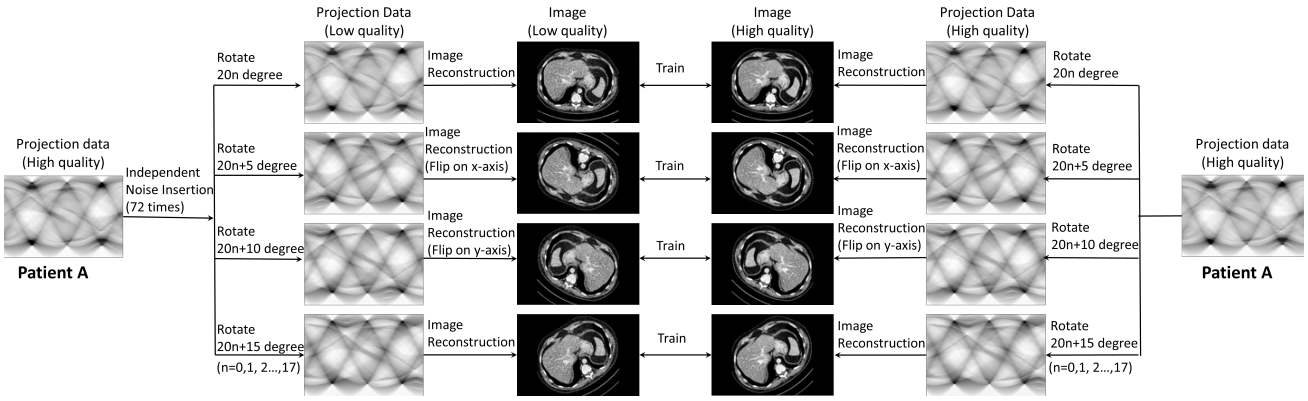


Fig. 2. Overview of the training scheme for the proposed ST_CNN method.

The proposed ST_CNN method is trained based on the data acquired from one single patient by generating a large amount of paired low-quality and high-quality images from the same patient (Figure 2).

The trained model is used to denoise the data acquired from the same patient. The training scheme is described as follows:

A. Low-quality image generation and augmentation:

1) Apply independent noise insertion multiple times (e.g., 72 times) on the original projection data of a specific patient to generate the corresponding low-quality projection data (e.g., 25% dose or 10% dose). The low dose levels can be randomized during the process of noise insertion.

2) Generate images with a large amount of different rotation angles. This cannot be accomplished by simply rotating the image itself as that will introduce errors through interpolation. Our approach is to apply the rotation angle directly on the projection data so that images are rotated at arbitrary angle without introducing additional errors. For example, we applied 72 different rotation angles (e.g., every 5 degree on 360 degree) to the low-quality projection data. After reconstruction, 72 sets of images with different rotation angles are obtained.

3) Divide the images into 4 groups: group1 (rotate 20*n degree, n=0, 1, …, 17); group 2 (rotate 20*n+5 degree); group 3 (rotate 20*n+10 degree); group 4 (rotate 20*n +15 degree). The images in group 2 are flipped on x-axis, and the images in group 3 are flipped on y-axis.

B. High-quality image generation and augmentation:

1) (Optional) Apply independent noise insertion multiple times (e.g., 72 times) on the original projection data. The amount of noise inserted is less than that to generate the low-quality images. One can also bypass this step so that the original dose level is directly used.

2) Apply rotation augmentation (e.g., every 5 degree on 360 degree) on the original projection data of the same patient to generate multiple high-quality projections.

3) Follow step 3) in section A to generate 4 groups of high-quality images.

C. Low-/high-quality image pairs generation:

Generate matched high- and low-quality patches with multiple slices (e.g., 64×64×7 voxels) from the reconstructed images in the first 3 groups for model training. The images in the 4th groups are used to generate the matched patches for model validation.

D. Model training:

The CNN denoising model can be based many of the popular network architecture. Here we employed a recently developed 2D residual-based CNN denoiser [3] for both ST_CNN and conventional deep CNN methods. The identical network architecture (Figure 3) was used for both methods so that any performance difference can be attributed to the different training methods. To optimize the performance of the CNN model, we used 7 adjacent CT slices as the channel input of the 2D residual CNN model [5]. The CNN inputs were first standardized (derived by subtracting the mean value and dividing by the standard deviation), and then subjected to initial layers that generated 128 feature maps using 2D convolutional layers. The feature maps were further processed by a series of 2D residual blocks, each of which consisted of repeated layers of 2D convolutional, batch normalization, and rectified linear unit activation. Then the output of residual blocks was projected back to a single-channel image by using a single convolutional layer with linear activation. This single-channel image was the estimated noise, which was further subtracted from the central input slice to get the final denoising result.

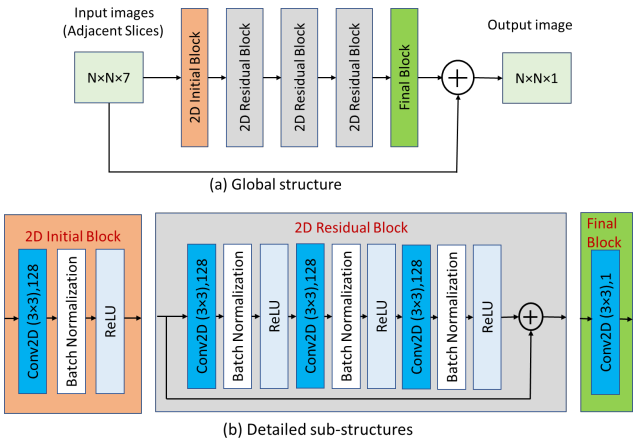


Figure 3. The architecture of residual-based 2D CNN denoiser. (a) Global structure of the network containing a 2D initial block, three 2D residual blocks, and a final block. (b) Details regarding the convolutional layers and transformations used within each block. Conv2D = two-dimensional convolutional layer, N = arbitrary image size, M = number of slices (1,3,7), ReLU = rectified linear units

III. Results

Figure 3 compares full-dose (FD) images reconstructed and denoised using 4 different methods: (a) filtered-backprojection (FBP), (b) iterative reconstruction (IR), (c) conventional CNN, and (d) ST_CNN. The images were from a patient case in the Mayo/AAPM Low-dose CT Grand Challenge data library (Case number: L291). In Figure 3 and the following figures of this article, "CNN" refers to the conventional residual CNN method [3]. The FBP and IR reconstructions used matched kernels of B30 and I30 at a strength setting of 3. The conventional CNN was trained and validated using FBP FD and FBP QD image pairs from a subset of totally 30 patient data (17 patients for training and 5 patients for validation). The residual network architecture was identical to that used in the ST_CNN. The trained conventional CNN model was applied to denoise the FBP FD images of the rest of the patient data (e.g., L291). The ST_CNN was trained and validated using augmented FBP 10% dose and FBP FD image pairs of a specific patient (e.g., L291), and then was applied to denoise the original FBP FD images of the same patient. The performance of two CNN models were assessed visually by an experienced radiologist. For the overall image quality evaluation, the radiologist ranked the ST_CNN method better than the conventional one because of more homogeneous liver parenchyma and better low-contrast lesion visibility (Arrows in the figure point to two subtle malignant liver tumors).

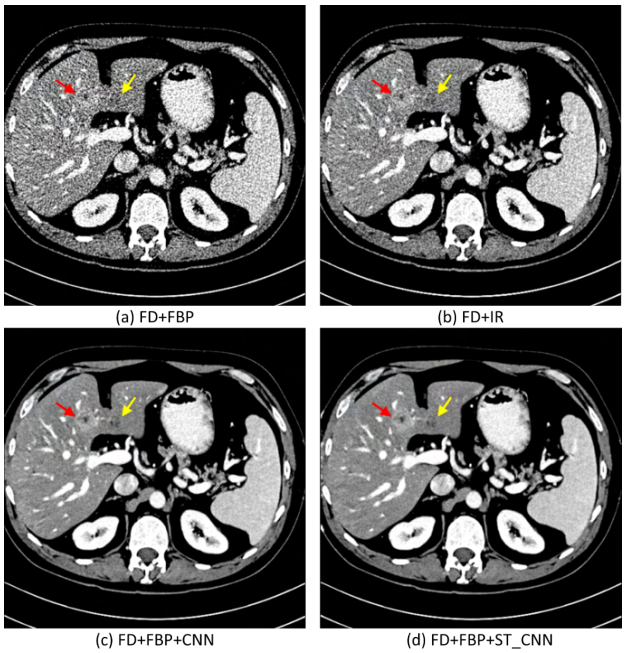


Figure 3. An example comparing the FBP and IR full dose (FD), conventional and self-trained CNN-denoised FBP FD images from a patient case in the Mayo/AAPM Low-dose CT Grand Challenge data library (Case number: L291). The arrows point to two subtle liver lesions. Slice thickness was 1 mm. To visualize the different appearance better, the display window was narrowed down to [60,200] HU.

To have a reference standard for quantitative evaluation of the performance, the self-trained CNN was trained and validated using augmented 10% dose (FBP) and QD (FBP) image pairs, and then applied to denoise the FBP QD images of the same patient. In this way, the original FD images can be used as the reference standard. The previously trained conventional CNN was used to denoise the same FBP QD images for CNN performance comparison. Figure 4 compares images reconstructed and denoised using 4 different conditions: (a) QD+FBP, (b) QD+IR, (c) QD+FBP+CNN, (d) QD+FBP+ST_CNN, and the two FD reconstructions were used as the reference standard: (e) FD+FBP and (f) FD+IR. Performance of the two CNN models were assessed visually by the same radiologist. In terms of low-contrast lesion visibility, conventional and self-trained CNN appeared to have a similar performance (Arrows in the figure point to two subtle malignant liver tumors). For the overall image quality evaluation, the radiologist ranked the self-trained CNN method better than the conventional one because of more homogeneous liver parenchyma and less false positive structures (A zoomed-in ROI in the liver parenchyma corresponding to the green box is shown in the bottom-right).

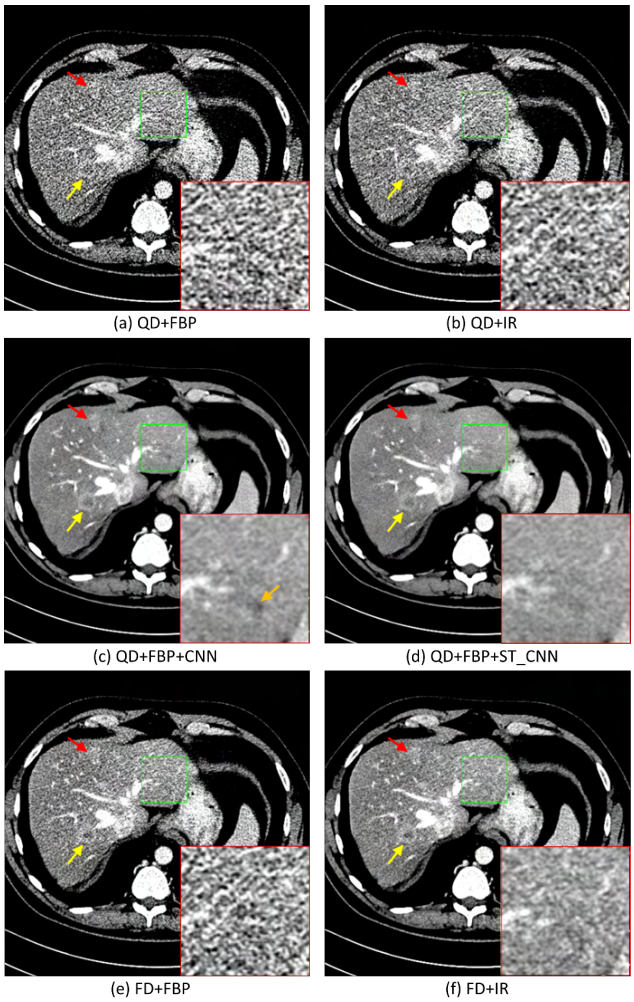


Figure 4. An example comparing the FBP and IR quarter dose (QD), FBP and IR full dose (FD), conventional and self-trained CNN-denoised QD images from a patient case in the Mayo/AAPM Low-dose CT Grand Challenge data library (Case number: L291). The arrows point to two subtle liver lesions. A zoomed-in ROI within the green box is shown in the bottom-right corner. Note the orange arrow on the conventional CNN-denoised image points to a false positive lesion that does not exist on the self-trained CNN-denoised image. Slice thickness was 1 mm. To visualize the different appearance better, the display window was narrowed down to [60,200] HU.

Using FD+FBP as the reference, the root mean square error (RMSE), peak signal-to-noise ratio (PSNR), and structural similarity (SSIM) were calculated for the conventional CNN and ST_CNN-denoised QD images (Table 1). The results provide clear evidence that the ST_CNN method has a performance similar to that of conventional deep CNN denoising methods without the need of a large number of training data.

TABLE I
QUANTITATIVE RESULTS (MEAN±SDS) ASSOCIATED WITH CONVENTIONAL AND SELF-TRAINED CNN METHODS FOR PATIENT CASE L291

	PSNR	RMSE	SSIM
Conventional CNN	41.9±2.1	19.8±3.8	0.95±0.02
Self-trained CNN	41.8±2.1	20.1±3.7	0.95±0.02

IV. CONCLUSION

We have designed a patient-specific self-trained CNN denoising method, aided by data augmentation through projection domain. Preliminary clinical evaluation demonstrated that the proposed method may achieve similar image quality in comparison with conventional deep CNN denoising methods pre-trained on a large number of patient cases. This new technique has the potential to overcome the generalizability issue of conventional training methods and to provide optimized noise reduction for each individual patient.

ACKNOWLEDGMENT

The authors acknowledge the computing facility (mForge) provided by Mayo Clinic for research computing. Dr. Zhou was supported by Mayo Radiology Research Fellowship.

REFERENCES

[1] H. Chen, Y. Zhang, W. Zhang, P. Liao, K. Li, J. Zhou, and G. Wang, "Low-dose CT via convolutional neural network," *Biomed. Opt. Express*, vol. 8, no. 2, pp. 679–694, 2017.

[2] H. Chen, Y. Zhang, MK Kalra, F. Ling, Y. Chen, P. Liao, J. Zhou, and G. Wang, "Low-dose CT with a residual encoder-decoder convolutional neural network," *IEEE Trans. Med. Imag.*, vol. 36, no. 12, pp. 2524–2535, 2017.

[3] Nathan R. Huber, Andrew D. Missert, Lifeng Yu, Shuai Leng, Cynthia H. McCollough, "Evaluating a Convolutional Neural Network Noise Reduction Method When Applied to CT Images Reconstructed Differently Than Training Data," *Journal of Computer Assisted Tomography*, vol. 45, no. 4, pp. 544-551, 2021.

[4] W. Yang, H. Zhang, J. Yang, J. Wu, X. Yin, Y. Chen, H. Shu, L. Luo, G. Coatrieux, and Z. Gui, "Improving low-dose CT image using residual convolutional network," *IEEE Access*, vol. 5, pp. 24698–24705, 2017.

[5] Zhongxing Zhou, Nathan R. Huber, Akitoshi Inoue, Cynthia H. McCollough, Lifeng Yu, "Residual-based convolutional-neural-network (CNN) for low-dose CT denoising: impact of multi-slice input," in SPIE Medical Imaging. 2022. SPIE.

[6] J. M. Wolterink, T. Leiner, M. A. Viergever and I. Isgum, "Generative adversarial networks for noise reduction in low-dose CT", *IEEE Trans. Med. Imag.*, vol. 36, no. 12, pp. 2536-2545, 2017.

[7] Q. Yang, P. Yan, Y. Zhang, H. Yu, Y. Shi, X. Mou, M. K. Kalra, and G. Wang, "Low dose CT image denoising using a generative adversarial network with Wasserstein distance and perceptual loss," *IEEE Trans. Med. Imag.*, vol. 37, no. 6, pp. 1348 – 1357, 2018

2D-3D motion registration of rigid objects within a soft tissue structure

Nargiza Djurabekova[1], Andrew Goldberg[2], David Hawkes[1], Guy Long[3], Felix Lucka[1,4] and Marta M. Betcke[1]

***Abstract*—The study of rigid body dynamics in soft tissue is an essential part of orthopaedic imaging. Our focus is the foot and ankle structure, which consists of 28 bones surrounded by a variety of soft tissue such as 30 muscles, numerous ligaments, bursae and nerves. The importance of understanding the involved dynamics is evident by the frequent need to find functional joint replacements to fight diseases such as hindfoot arthritis. We study a simplified problem by constructing a phantom with two "bones" submerged in silicone and a remotely controlled LEGO robot that moves one of the two bones.**

We perform motion registration by manipulating the bone segmented from an initial static scan and matching its digitally reconstructed radiographs to a sequence of scanned X-ray projections. The registration of the 3D volume to 2D projections requires the right combination of optimization and interpolation to achieve an optimal result. We test four different approaches with the help of the Flexible Algorithms for Image Registration (FAIR) toolbox.

***Index Terms*—2D3D registration, dynamic X-ray CT, cone beam CT, rigid motion**

I. Introduction

THIS project covers the approach to dynamic imaging of the foot and ankle through the lens of image registration, in particular volume to projection registration. We explore the parametric, bone-by-bone approach as we assume that the movement of the individual bones can be modelled through rigid transformation. Each bone is segmented out of the initial static scan and the volume containing one of the bones is matched to a sequence of fluoroscopic projections. We use the term fluoroscopy here to mean an "X-ray movie", i.e. photons shooting in quick succession to track the movement of the scanned object while the source, detector and the sample stage remain stationary. The registration is performed with the help of the Flexible Algorithms for Image Registration (FAIR) package [1] for MATLAB and to test this approach we use the phantom described in section I-A.

A. Data acquisition

The data for this project was acquired at the CWI research institute in Amsterdam, using the state-of-the-art X-ray scanner called FleXray. During the acquisition, the FleXray is completely sealed and can only scan small (up to 10 cm tall) objects, so we constructed a remotely controlled base and a simple phantom consisting of two "bones" made from gypsum plaster submerged in some silicone. Silicone was chosen to reproduce the effect of cartilage and other soft tissue between and surrounding the bones.

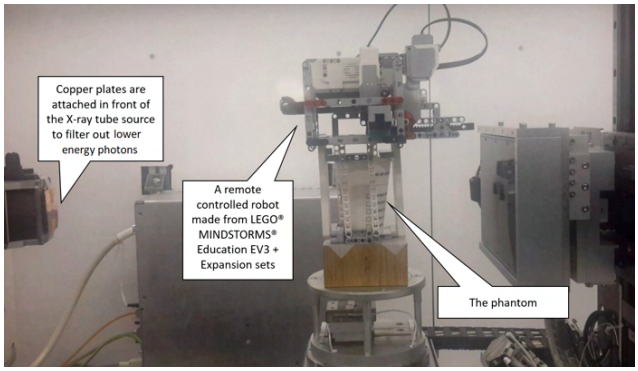


Fig. 1. The FleXray scanner set-up with the CWI phantom.

The data was collected at 78 kV with maximum power due to the high attenuation of gypsum and silicone. To remove lower energy photons, layers of thin copper plates were attached to the source as can be seen in fig. 1. Static frame-by-frame data of 40 individual positions was acquired as well as data from continuously shooting X-ray beams (i.e. fluoroscopic data) from a few selected angles.

B. Image registration

Image registration is an umbrella term for aligning two or more sets of images. It is often used in image processing for combining images to either extend the view, such as in panoramic images, or to enhance an existing view and reduce signal-to-noise ratio. The purpose of image registration is often to identify similar features in different images in order to acquire further information about the features. This has a very clear application in medical imaging, where scans are often performed multiple times during treatment and sometimes even using different devices (e.g. MR and X-ray systems). To find out the exact differences between the images, it is crucial to align the images correctly, i.e. perform registration.

To avoid the need for experts to identify correct features in order to perform landmark-based registration, we instead choose to rely on intensity based registration. Intensity based methods align images by comparing their intensity values and contrasts, and often require little to no preparation in terms of marker placement etc., although they can be quite time consuming during the registration process. An example of a fully

[1] Centre for Medical Image Computing (CMIC) at the Department of Computer Science, University College London (UCL)
[2] Royal National Orthopaedic Hospital, UCL
[3] CurveBeam Europe, Ltd
[4] Computational Imaging (CI) Group at the Centrum Wiskunde & Informatica (CWI)

intensity based method can be found in [2] where individual vertebrae in a 3D CT volume were robustly registered to the corresponding vertebrae in a 2D X-ray image. The registration was performed by optimising for one of the six degrees of freedom of the 3D bone at a time. We propose a method in which we use the FAIR toolbox to optimise for all six transformation parameters simultaneously.

C. Practical background

Image registration can occur when both images are in the same domain. That is, both are in $\Omega \subset \mathbb{R}^d$, where d is the common dimension, usually 2 or 3. So for some reference image $u_R : \Omega \to \mathbb{R}$, we try to transform a template image $u_T : \Omega \to \mathbb{R}$ using the transformation $\rho : \mathbb{R}^d \to \mathbb{R}^d$ in a way that minimizes the following objective function:

$$\arg\min_w F(w) = dist(u_T(\rho(w, x)), u_R(x)) + \mathcal{R}(w). \quad (1)$$

Here, $dist$ is some distance measure function and $\mathcal{R}$ is a regularizer, $x \in \Omega$ is the set of coordinates and w is the vector of parameters for the transformation ρ. In our case, the template is a 3D bone s segmented out of the whole image u_T and the reference is a projection (or a set of projections) g_R. Since the registration is to be performed between a 3D volume s and 2D measurement data g_R, one must first generate digitally reconstructed radiographs (DRRs) from the 3D volume, and then perform the registration between the DRR(s) and given X-ray measurement data. Let $s : \Omega \to \mathbb{R}$ be the template image and $g_R : \Omega' \to \mathbb{R}$ be a reference image. Domain Ω' is a set of one or more stacked X-ray projections from different angles, $\Omega' \subset \mathbb{R}^{d-1} \times [0, 2\pi)$. Let also A_θ be the X-ray cone-beam forward operator, where θ is the set of indices representing the present projection angles. Then for 2D3D registration, the eq. 1 becomes

$$\arg\min_w F(w) = dist(A_\theta(s(\rho(w, x))), g_R(x)) + \mathcal{R}(w).$$

As we are working with rigid bodies, our parameter vector w is going to consist of 6 elements corresponding to 6 degrees of freedom, 3 for rotation w_1, w_2, w_3 and 3 for translation w_4, w_5, w_6.

II. Numerical realization and results

RECALL that the objective function we want to minimize is

$$\arg\min_w F(w) = dist(A_\theta(s(\rho(w, x))), g_R(x)) + \mathcal{R}(w). \quad (2)$$

We set the distance function to be the squared ℓ^2 norm, also known as the sum of squared differences, which we refer to as ψ. So if the residual $r = A_\theta(s(\rho(w, x))) - g_R(x)$, then the $dist$ function is

$$dist(A_\theta(s(\rho(w, x))), g_R(x)) = \psi(r) = \frac{1}{2}||r||_2^2 = \frac{1}{2}||A_\theta(s(\rho(w, x))) - g_R(x)||_2^2.$$

In the discrete setting, where the image is defined on a grid of cells, $\psi(r)$ is multiplied by a scaling factor h which is defined as the cell width. This is particularly useful for the multiscale optimization, where different levels/scales correspond to different grid densities.

As a constraint on transformation parameters w, we use Tikhonov's regularization, i.e. given a reference set of parameters w^{ref},

$$\mathcal{R}(w) = \frac{1}{2}(w - w^{ref})^\intercal M(w - w^{ref}),$$

where M is a diagonal matrix of weights for the transformation parameters.

A. Pre-processing the data

For the two-dimensional reference images, one or two projection images are used for each of the 41 static scans (one initial position + 40 poses along a programmed motion path). The template image is obtained by reconstructing the static scan in initial position and segmenting it with K-means [3] into 3 separate masks - air, silicone (as well as lego) and gypsum, corresponding to air, soft tissue and bone. Noise from the reconstruction u_T is smoothed by a convolution with a Gaussian filter. The bone mask is split into two masks corresponding to the two shapes coded as the m_{tibia} (top) and the m_{talus} (bottom). These bones are then segmented out of the reconstruction

$$s_{bones} = (m_{tibia} + m_{talus}) \odot u_T$$

by means of element-wise multiplication ($\odot$) to obtain a background volume which can be projected with the forward operator A_θ into the data domain

$$g_{bg} = A_\theta(u_T - s_{bones}).$$

These background projections can then be removed from all following reference images to give the best chance to our intensity based registration algorithm,

$$g_{bones} = g - g_{bg}.$$

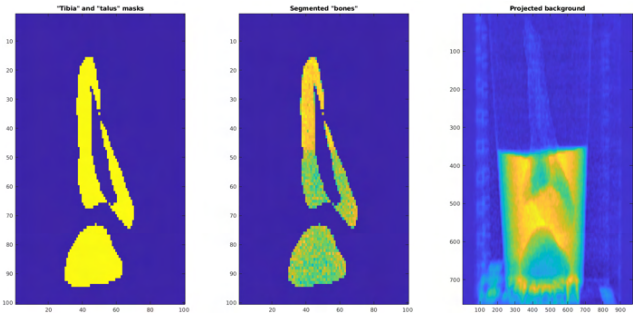


Fig. 2. Demonstration of data pre-processing. *Left:* A slice through the bone masks volume, top: m_{tibia}, bottom: m_{talus}; *Middle:* A slice through the segmented bones volume s_{bones} acquired by applying the bone masks to the full reconstruction of the initial pose; *Right:* A simulated projection of the "background", g_{bg}.

The masks, segmented bones and an example of a projected background for the static initial position can be found in fig. 2.

As the bones move further away from the initial position, the background extracted from it becomes a worse approximation to the true background due to decreasing overlap of the occlusion in the initial position with the occlusion in the present frame. However, as this only affects a small part of the background volume, there is still a benefit to subtracting the background.

B. Results

We run the registration problem on all 40 frames + the first initial position using 4 different strategies. The registration itself is performed using Gauss-Newton method and the template image s_1 is always the bone (in our case tibia, or s_{tibia}) in the initial position, while w_1^{ref} is set to $w_0 = [0, 0, 0, 0, 0, 0]^\intercal$.

Setup 1.
For each consecutive frame to be registered (iteration $k+1$), the GN method is warm started by setting w^{ref} to the parameters w_k obtained from the registration of the previous frame k

$$w_{k+1}^{ref} = w_k$$

Meanwhile the template image s for a consecutive frame registration is updated based on the previous iteration's output parameters, using rigid transformation T_{rigid} and spline interpolation $\mathcal{T}_{SI}$, so

$$s_{k+1} = \mathcal{T}_{SI}(T_{rigid}(s_k, w_{k+1})).$$

To simplify notation, we combine the rigid transformation and the interpolation into one transformation operator T. The updated s_{k+1} is used as a template image for the following iteration. Results for setup 1 are visualized in fig. 4(a).

Setup 2.
In this setup, we obtain the template image s_k (for $k > 1$) directly from s_1 via rigid transformation with parameters w_k, while the reference parameters w^{ref} for the optimization stay the same for all iterations. Then in every iteration we obtain global transformation parameters as opposed to setup 1, where the parameters are computed locally. For every iteration $k = 1, 2, ..., 40$,

$$s_{k+1} = T(s_1, w_{k+1}).$$

Note that w^{ref} and s_1 remain unchanged through all iterations. Results for setup 2 are visualized in fig. 4(b).

Setup 3.
As in setup 2, except that the registration is now performed using the coarse-to-fine multiscale approach for registration with the goal to register over larger distances. Results for setup 3 are visualized in fig. 4(c).

Setup 4.
In this setup we propose a scheme to aggregate the optimized parameters and use those along with the s_1. As can be seen in fig. 3 this scheme aggregates incremental transformations $w_1, ..., w_k$. In practice, this is performed stepwise using the `nested_parameters` function described in appendix A. This way, interpolation is applied only once to s_1 to move it into position of frame k. The variables for this setup are updated as follows:
For $k = 0$,

$$w_{k+1}^{ref} = w_1^{ref}$$
$$w_{k+1} = \text{GN}(w_{k+1}^{ref})$$
$$w_{k+1}^{nest} = w_{k+1}$$
$$s_{k+1} = T(s_1, w_{k+1}^{nest})$$

For $k = 1, 2, 3, ..., 40$

$$w_{k+1}^{ref} = w_k$$
$$w_{k+1} = \text{GN}(w_{k+1}^{ref})$$
$$w_{k+1}^{nest} = \texttt{nested\_parameters}(w_k^{nest}, w_{k+1})$$
$$s_{k+1} = T(s_1, w_{k+1}^{nest})$$

The updated template s is used as initialization for the following iterations. Results for setup 4 are visualized in fig. 4(d).

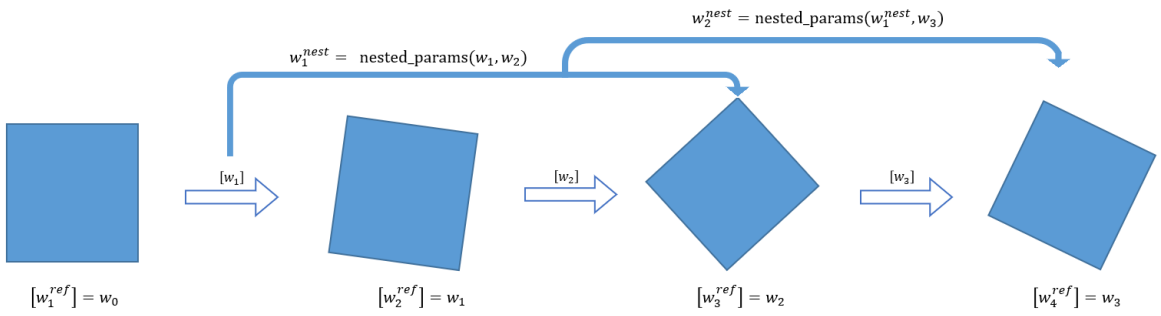


Fig. 3. Diagram of parameter aggregation via the `nested_parameters` function.

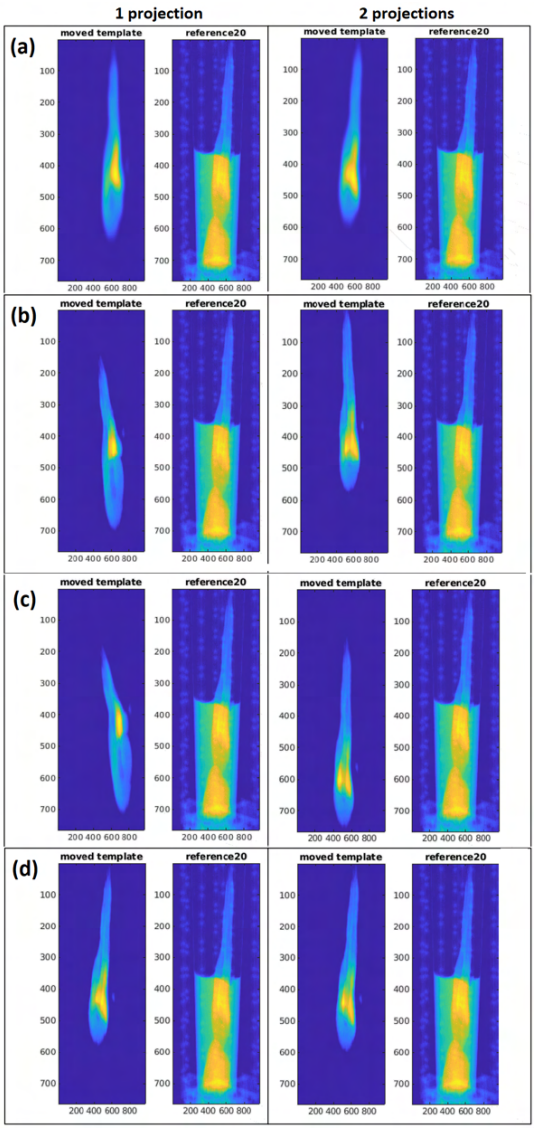


Fig. 4. Demonstration of the four different setups. Registration performed with 1 projection at angle $45°$ (Left) and with 2 projections at angles $0°$ and $90°$ (Right). Rows (a) to (d) correspond to registration results for frame 20 for setups 1 to 4 respectively.

III. Discussion

In this work we consider the 2D3D parametric image registration between fluoroscopic X-ray data and a segmented volume. The registration is performed with spline interpolation and Gauss-Newton optimization as implemented in the FAIR toolbox [1], adjusted for our 2D3D problem. We create an objective function eq. 2, which we optimize for rigid transformation parameters that would align the template bone with the reference projections. Gauss-Newton algorithm functions as the optimizer. The main difficulty arises with the initialization, so we test out four different setups.

The first setup involves initializing the optimization of a consecutive frame with a previous frame's transformed template and optimal parameters. This leads to very blurry results due to what amounts to recursive interpolation (see fig. 4(a)). To avoid this, the second setup initializes the rigid transformation with all parameters w set to 0 to directly compute the transformation from the initial position to the currently processed timestep. However, as the bone moves further away from its initial position, the results become consistently worse. And then as the bone moves back, closer to the initial position, the algorithm is able to recover and realign the bones (see fig. 4(b)). To try and make up for the second setup's inability to account for large displacements, we repeat its initialization in setup 3 but now the registration is performed with a coarse-to-fine grid approach. As can be seen from fig. 4(c), though, this multiscale approach does not lead to better results. This is likely due to the presence of background elements in the reference image(s). In particular, the presence of the other bone causes the template to be aligned with it, as can be most clearly seen in fig. 4(c) with 2 projections. Finally, for setup 4, we propose initializing the Gauss-Newton optimization with the previously optimised parameters and updating the template image from the initial position with aggregated parameters. These aggregated parameters come from nesting previous transformations together to achieve one single transformation (using `nested_parameters` function, appendix A). In fig. 4(d), we can see that this approach does not manage to perform correct registration with just one projection, since if it wrongly estimates in one frame, the mistake only gets amplified with each frame and the algorithm can no longer recover. The moving bone can even end up completely out of the visible region of interest in the final few frames. However, two orthogonally directed projections (at $0°$ and $90°$) constrain the problem enough to eliminate the ambiguity between motion and scaling, leading to promising results depicted in fig. 4(d) on the right, with the transformed template bone tracking the true motion quite accurately.

Acknowledgment

The authors would like to thank NVIDIA Corporation for the GeForce Titan Xp GPU used to create the reconstructions presented here.

References

[1] J. Modersitzki, *FAIR: Flexible Algorithms for Image Registration*. Philadelphia: SIAM, 2009.

[2] G. Penney, P. Batchelor, D. Hill, D. Hawkes, and J. Weese, "Validation of a two- to three-dimensional registration algorithm for aligning preoperative ct images and intraoperative fluroscopy images," *Medical physics*, vol. 28, pp. 1024–32, 07 2001.

[3] X. Jin and J. Han, *K-Means Clustering*. Boston, MA: Springer US, 2010, pp. 563–564. [Online]. Available: https://doi.org/10.1007/978-0-387-30164-8$_4$25

APPENDIX A
NESTED_PARAMETERS FUNCTION

The `nested_parameters` function is a function that aggregates previous registration parameters to move a rigid object from the initial position at $t = 0$ to all the positions in the following timeframes. The nesting or aggregation of parameters is derived as follows. Recall that rigid transformation ρ on x can be expressed as

$$\rho = R(x) + \tau_{ext} = R(x) + (I - R)c + \tau, \tag{3}$$

where $\tau_{ext} = [\tau^1_{ext}, \tau^2_{ext}, \tau^3_{ext}]^\intercal$ is the extended translation vector that contains the shift to the centre of the domain c, rotation Rc, and shift back, as well as the actual translation vector $\tau = [\tau^1, \tau^2, \tau^3]^\intercal$. This extended translation vector $\tau_{ext} = (I - R)c + \tau$ corresponds to the last three elements of the rigid transformation parameters vector w. The first three parameters of w are the angles of rotation α, β and γ (with respect to the axes x, y and z), which are used to calculate the corresponding rotation matrix R. We want to be able to perform the transformation 3 recursively to compute the effective rotation R_{nest} and translation τ_{nest} of two timeframes together. Let R_1 and τ_{ext} refer to the rotation and translation performed on the first timeframe and R_2 and τ'_{ext} on the second one. Then the effective transformation ρ' can be computed with the nested rotation matrix R_{nest} and nested translation vector τ_{nest} as follows

$$\rho' = R_{nest}(\rho) + \tau_{nest} = R_2(R_1(x) + \tau_{ext}) + \tau'_{ext}.$$

Computing the translation parameters of τ_{nest} is straightforward

$$\tau_{nest} = R_2(\tau_{ext}) + \tau'_{ext},$$

while finding the rotation angles a, b and c of R_{nest} is a little more involved. We know that as a 3D rotation matrix, R_{nest} has the form

$$R_{nest} = \begin{bmatrix} \cos a \cos b & \cos a \sin b \sin c - \sin a \cos c & \cos a \sin b \cos c + \sin a \sin c \\ \sin a \cos b & \sin a \sin b \sin c + \cos a \cos c & \sin a \sin b \cos c - \cos a \sin c \\ -\sin b & \cos b \sin c & \cos b \cos c \end{bmatrix}.$$

To solve for the rotation parameters (a, b, and c), we consider the following equation

$$\begin{aligned} R_{nest} &= R_2 R_1 \\ &= \begin{bmatrix} \cos\alpha\cos\beta & \cos\alpha\sin\beta\sin\gamma - \sin\alpha\cos\gamma & \cos\alpha\sin\beta\cos\gamma + \sin\alpha\sin\gamma \\ \sin\alpha\cos\beta & \sin\alpha\sin\beta\sin\gamma + \cos\alpha\cos\gamma & \sin\alpha\sin\beta\cos\gamma - \cos\alpha\sin\gamma \\ -\sin\beta & \cos\beta\sin\gamma & \cos\beta\cos\gamma \end{bmatrix} \times \\ &\begin{bmatrix} \cos\alpha'\cos\beta' & \cos\alpha'\sin\beta'\sin\gamma' - \sin\alpha'\cos\gamma' & \cos\alpha'\sin\beta'\cos\gamma' + \sin\alpha'\sin\gamma' \\ \sin\alpha'\cos\beta' & \sin\alpha'\sin\beta'\sin\gamma' + \cos\alpha'\cos\gamma' & \sin\alpha'\sin\beta'\cos\gamma' - \cos\alpha'\sin\gamma' \\ -\sin\beta' & \cos\beta'\sin\gamma' & \cos\beta'\cos\gamma' \end{bmatrix}. \end{aligned}$$

To make the calculation simpler to follow, we shall replace $\cos\alpha, \cos\beta$ and $\cos\gamma$ by $c(1)$, $c(2)$ and $c(3)$. For $\sin\alpha$, $\sin\beta$, $\sin\gamma$ we have $s(1)$, $s(2)$, $s(3)$ and analogously for the α', β', γ' parameters, their sine and cosine functions turn into $s'(1)$, $s'(2)$, $s'(3)$, and $c'(1), c'(2), c'(3)$. This gives us

$$\begin{aligned} R_{nest} = R_2 R_1 = &\begin{bmatrix} c(1)c(2) & c(1)s(2)s(3) - s(1)c(3) & c(1)s(2)c(3) + s(1)s(3) \\ s(1)c(2) & s(1)s(2)s(3) + c(1)c(3) & s(1)s(2)c(3) - c(1)s(3) \\ -s(2) & c(2)s(3) & c(2)c(3) \end{bmatrix} \\ &\times \begin{bmatrix} c'(1)c'(2) & c'(1)s'(2)s'(3) - s'(1)c'(3) & c'(1)s'(2)c'(3) + s'(1)s'(3) \\ s'(1)c'(2) & s'(1)s'(2)s'(3) + c'(1)c'(3) & s'(1)s'(2)c'(3) - c'(1)s'(3) \\ -s'(2) & c'(2)s'(3) & c'(2)c'(3) \end{bmatrix} \\ = &\begin{bmatrix} R_{11} & R_{21} & R_{31} \\ R_{12} & R_{22} & R_{32} \\ R_{13} & R_{23} & R_{33} \end{bmatrix}, \end{aligned}$$

where

$$\begin{aligned} R_{11} &= c(1)c(2)(c'(1)c'(2)) + (c(1)s(2)s(3) - s(1)c(3))(s'(1)c'(2) \\ &\quad - (c(1)s(2)c(3) + s(1)s(3))s'(2), \\ R_{21} &= c(1)c(2)(c'(1)s'(2)s'(3) - s'(1)c'(3)) + (c(1)s(2)s(3) - s(1)c(3))(s'(1)s'(2)s'(3) \\ &\quad + c'(1)c'(3)) + (c(1)s(2)c(3) + s(1)s(3))(c'(2)s'(3)), \\ R_{31} &= c(1)c(2)(c'(1)s'(2)c'(3) + s'(1)s'(3)) + (c(1)s(2)s(3) \\ &\quad - s(1)c(3))(s'(1)s'(2)c'(3) - c'(1)s'(3)) + (c(1)s(2)c(3) + s(1)s(3))(c'(2)c'(3)), \\ R_{12} &= s(1)c(2)(c'(1)c'(2)) + (s(1)s(2)s(3) + c(1)c(3))(s'(1)c'(2)) \\ &\quad - (s(1)s(2)c(3) - c(1)s(3))s'(2), \\ R_{22} &= s(1)c(2)(c'(1)s'(2)s'(3) - s'(1)c'(3)) + (s(1)s(2)s(3) \\ &\quad + c(1)c(3))(s'(1)s'(2)s'(3) + c'(1)c'(3)) + (s(1)s(2)c(3) - c(1)s(3))(c'(2)s'(3)), \\ R_{32} &= s(1)c(2)(c'(1)s'(2)c'(3) + s'(1)s'(3)) + (s(1)s(2)s(3) \\ &\quad + c(1)c(3))(s'(1)s'(2)c'(3) - c'(1)s'(3)) + (s(1)s(2)c(3) - c(1)s(3))(c'(2)c'(3)), \\ R_{13} &= -s(2)(c'(1)c'(2)) + (c(2)s(3))(s'(1)c'(2)) - (c(2)c(3))s'(2), \\ R_{23} &= -s(2)(c'(1)s'(2)s'(3) - s'(1)c'(3)) + (c(2)s(3))(s'(1)s'(2)s'(3) \\ &\quad + c'(1)c'(3)) + (c(2)c(3))(c'(2)s'(3)), \\ R_{33} &= -s(2)(c'(1)s'(2)c'(3) + s'(1)s'(3)) + (c(2)s(3))(s'(1)s'(2)c'(3) \\ &\quad - c'(1)s'(3)) + (c(2)c(3))(c'(2)c'(3)). \end{aligned}$$

Then the rotation parameters a, b, c are:

$$b = \arcsin(-R_{13}) = \arcsin(s(2)(c'(1)c'(2)) - (c(2)s(3))(s'(1)c'(2)) + (c(2)c(3))s'(2)),$$

$$\begin{aligned} a = \arcsin(R_{12}/\cos b) = \arcsin((s(1)c(2)(c'(1)c'(2)) + (s(1)s(2)s(3) \\ +c(1)c(3))(s'(1)c'(2)) - (s(1)s(2)c(3) - c(1)s(3))s'(2))/\cos b), \end{aligned}$$

$$\begin{aligned} c = \arcsin(R_{23}/\cos b) = \arcsin((-s(2)(c'(1)s'(2)s'(3) - s'(1)c'(3)) \\ + (c(2)s(3))(s'(1)s'(2)s'(3) + c'(1)c'(3)) + (c(2)c(3))(c'(2)s'(3)))/\cos b). \end{aligned}$$

So the whole nested transformation vector w_{nest} is then defined as

$$w_{nest} = [w^1_{nest}, w^2_{nest}, w^3_{nest}, w^4_{nest}, w^5_{nest}, w^6_{nest}]^\intercal = [a, b, c, \tau^1_{ext}, \tau^2_{ext}, \tau^3_{ext}]^\intercal,$$

and serves as the output to the `nested_parameters`(w, w') function, where w is the vector of transformation parameters of the first timeframe and w' of the second. The process can be repeated to compute the effective rotation and transformation for any number of steps, assuming w_{nest} is a good estimate of all the previous timeframes' transformations.

Gas Bubble Motion Artifact Reduction through Simultaneous Motion Estimation and Image Reconstruction

Kai Wang, Hua-Chieh Shao, You Zhang, Chunjoo Park, Steve Jiang, Jing Wang

***Abstract*--Motion of gas bubbles (gastrointestinal gas) in abdominal region can produce significant artifacts during on-board CBCT scanning, which adversely affects the imaging quality and limits the process of CBCT-based adaptive planning and image-guided radiotherapy. In this study, we tested the effectiveness of simultaneous motion estimation and image reconstruction technique (SMRIE) for improving CBCT image quality and HU accuracy of abdominal scan. The improved image quality in the simulation study demonstrated that SMRIE technique is promising for on-board CBCT gas bubble motion artifact reduction.**

***Index Terms*—CBCT image reconstruction, motion artifact reduction, motion estimation.**

I. Introduction

On-board cone-beam computed tomography (CBCT) is widely used in image-guided radiation therapy (IGRT) and shows the potential to aid in adaptive radiation therapy (ART) [1-4]. Its artifacts, range from those due to inherent limitation of imaging physics, such as scattering, beam hardening, to those caused by the long-time scanning process, such as patient respiratory motion, have been well studied [5-7]. Much work has been done on techniques to reduce these artifacts and improve the imaging quality. However, the artifacts produced by isolated aperiodic motions of small structures, such as gastrointestinal gas bubble, have been rarely described, and no available method has been discussed to correct them to our knowledge [2, 4]. As the gas bubble motion artifact is frequently seen in abdominal scan and could induce severe artifacts (as shown in Fig. 1), an artifact correction method designed for gas bubble motion is desired to improve the abdominal CBCT image quality and then aid in online patient setup and adaptive radiation therapy.

Serval studies have been conducted to characterize the imaging features of gas bubble motion artifacts through phantom and/or clinical experiment [6, 8, 9]. These studies revealed that data inconsistency of the high-contrast tissue-air boundary on different x-ray projections due to gastrointestinal peristalsis is the main reason for the streaky and/or hyperdense artifacts, and the shape of artifact can be affected by the size, speed, position and moving direction of gas bubble. Through viewing the patient scans, they found small intestine area was the most commonly affected region.

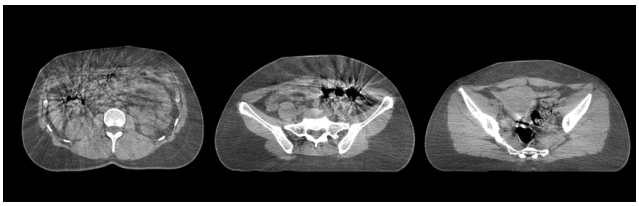


Fig. 1. On-board abdominal CBCT gas bubble artifacts in different regions of a same patient in a single scan (-300~300 HU).

Winklhoder *et al*. used rapid-kV switching (rs) dual-energy computed tomography (DECT) to reduced gastrointestinal peristalsis-related steak artifact and observed significant artifact reduction in both phantom and patient abdominal imaging [7]. However, the current on-board CBCT doesn't have energy discrimination ability, and the required time for scanning of on-board CBCT is much longer than diagnostic CT which could induce more gas bubble motion artifacts. Liu *et al.* proposed a cycle generative adversarial network (cycleGAN) for post-imaging processing of on-board abdominal CBCT and generate high-quality CBCT-based synthetic CT (sCT) for pancreatic adaptive radiotherapy [4]. Although much less gas bubble artifact can be seen in the generated sCT, the post-processing strategy and black-box property of deep learning might lead to fake abdominal structures in the regions affected by artifacts.

As the distribution and motion of gas bubbles in abdominal region are irregular in temporal domain, the widely used phase partition strategy for 4D CBCT motion artifact reduction cannot be readily applied for gas bubble artifact reduction. To overcome this limitation, we hypothesize that if the CBCT projection images were continuously divided into small groups, the gas bubble motion within each group can be neglected, an artifact reduced image can be reconstructed through estimation of the deformation vector fields (DVFs) between each group in image domain and motion-compensated image reconstruction.

This work was supported in part by the U.S. National Institutes of Health under Grant R01 EB027898.

K. Wang, H. Shao, Y. Zhang, C. Park, S. Jiang, and J. Wang are with the Department of Radiation Oncology, University of Texas Southwestern Medical Center, Dallas, TX 75390, USA (email: Kai.Wang@UTSouthwestern.edu, Hua-Chieh.shao@UTSouthwestern.edu, You.Zhang@UTSouthwestern.edu, Chunjoo.Park@UTSouthwestern.edu, Steve.Jiang@UTSouthwestern.edu and Jing.Wang@UTSouthwestern.edu).

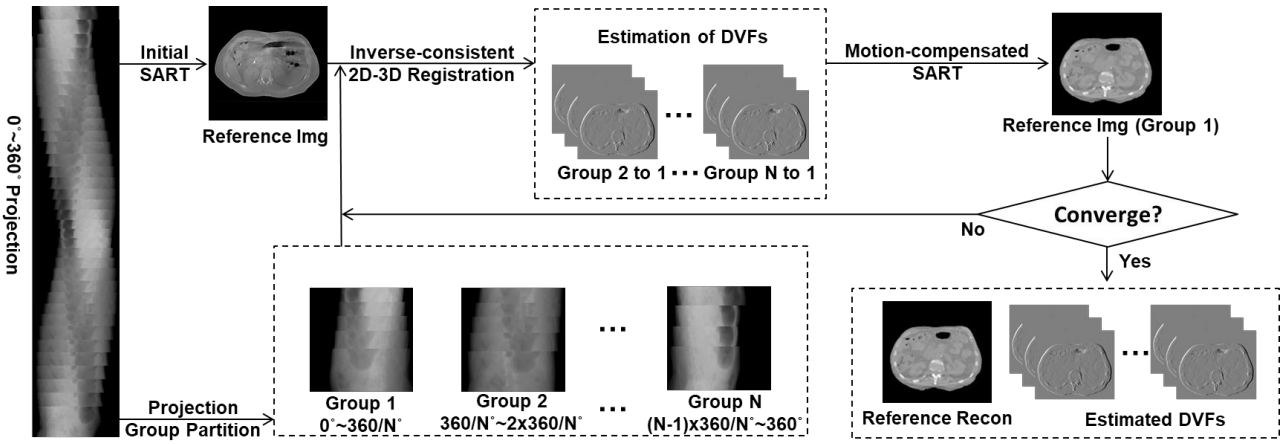


Fig. 2. Workflow of gas bubble artifact reduction using simultaneous motion estimation and image reconstruction (SMEIR) method. SART: simultaneous algebraic reconstruction technique; DVF: deformation vector field; N: number of projection groups.

Based on this hypothesis, in this work, we utilize the simultaneous motion estimation and image reconstruction (SMEIR) approach for abdominal CBCT image reconstruction and obtain the motion vectors simultaneously. The performance of the proposed strategy was evaluated with a simulation study.

II. Materials and Methods

A. The SMEIR algorithm

SMEIR algorithm has two main steps: motion estimation and motion-compensated reconstruction. Before SMEIR, we reconstructed an initial reference image using all the current projection images with simultaneous algebraic reconstruction (SART) method. Because of the existence of gas bubble motion during CBCT scan, the initial image would contain gas bubble motion artifacts, similar to images in Fig. 1. Different from the original SMEIR which was proposed for respiration artifact reduction, there is no motion phase information to guide motion phase partition for gas bubble motion. Instead, we evenly divided the projections into *N* projection groups following our hypothesis, each group has continued projection images covering 360°/*N* projection angle. We then applied the Demons registration algorithm to generate the initial inter-group DVFs between the reference group (group 1 in our experiment) and all other groups. 2D projection images from each projection group and 3D reference reconstruction image were used here for the corresponding DVF estimation. In the motion-compensated reconstruction step, we used all the projection images and the estimated DVFs to reconstruct a new reference CBCT by using the motion-compensated simultaneous algebraic reconstruction method (MC-SART). These two steps were iteratively used until the reconstruction process reached the converge condition. The overall workflow of SMEIR method for abdominal CBCT gas bubble motion artifact reduction is shown in Fig. 2, and the objective function is:

$$\widehat{\mu^1}, \hat{v} = \underset{\mu^1, v}{\operatorname{argmin}} \sum_{n=1}^{N} \|p^n - A\mu^1(x + v^{1\to n})\| + \beta\phi_1(\mu^1) + r\phi_2(v) \quad (1)$$

where μ^1 is the line attenuation coefficient of CBCT at group 1, $v^{1\to n}$ denotes the deformation matrix to transform CBCT from group 1 to group *n* (*n*=1, 2, …, *N*), p is the projection, β and r are regularization terms which control the balance between data fidelity, image total variation sparsity and DVF smoothness constrain. A is the CBCT projection system matrix, which was calculated through ray-tracing technique. Regularization function ϕ_1 is total variation in our study, and ϕ_2 measures free-form energy of the deformation fields which is defined as:

$$\phi_2(v) = \sum_{v\in R^3} \sum_{i=1}^{3} \sum_{j=1}^{3} \left(\frac{\partial v^i}{\partial x^j}\right)^2 \quad (2)$$

To be specific, in the 2D-3D registration motion estimation step, we iteratively update the following objective functions group by group:

$$\hat{v}^{n\to 1} = \underset{v^{n\to 1}}{\operatorname{argmin}} \|p^1 - A\mu^n(x + v^{n\to 1})\| + r\phi_2(v^{n\to 1}) \quad (3)$$

$$\hat{v}^{1\to n} = \underset{v^{1\to n}}{\operatorname{argmin}} \|p^n - A\mu^1(x + v^{1\to n})\| + r\phi_2(v^{1\to n}) \quad (4)$$

r is empirically set to 0.05 in our study. Nonlinear conjugate gradient algorithm was used to minimize these two functions alternatively, the DVFs updated in the first function were inverted and served as the initial for the second function. T_1=5 iterations were used in our experiment for motion estimation in each iteration of SMEIR.

In MC-SART step, we have the current reconstruction results for each group as:

$$\mu^n = \mu^1(x + v^{1\to n}) \quad (5)$$

and the reference group image can be updated through a modified SART algorithm, which is:

$$\mu_j^{1,new} = \mu_j^1 + \lambda \frac{\sum_{n,k} d_{jk}^{n\to 1} \sum_i [a_{ik} \frac{p_i - \sum_k a_{ik} u_k^n}{\sum_{k=1}^{J} a_{ik}}]}{\sum_{n,k} d_{jk}^{n\to 1} \sum_i a_{ik}} \quad (6)$$

where $d_{jk}^{n\to 1}$ denotes element of the DVF matrix that deforms image from group *n* to group 1. T_2=10 iterations were used in our study for motion compensated reconstruction in each iteration of SMEIR.

B. Materials

To simulate the irregular motion during on-board CBCT

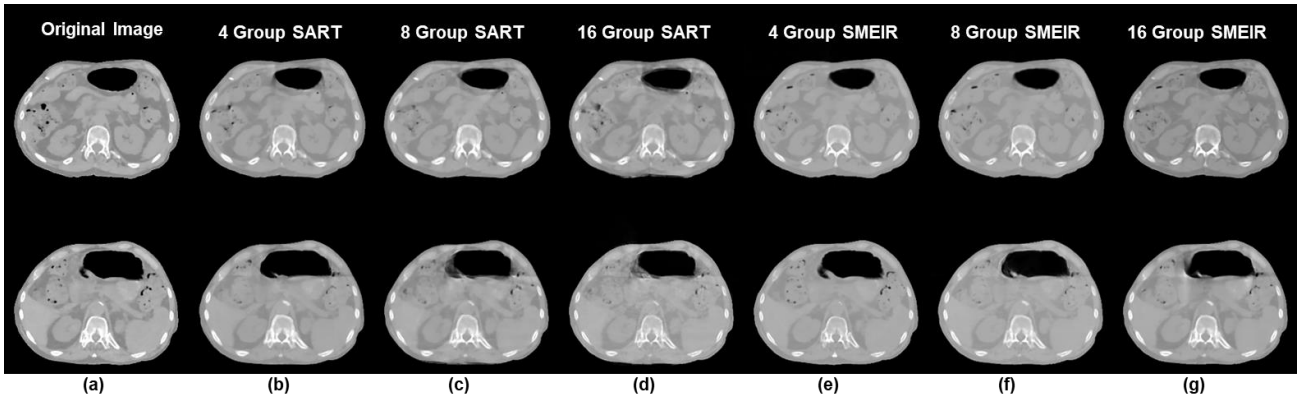


Fig. 3. Original group 1 images, simulated gas bubble artifacts in SART results, and reconstructed images with SMEIR. (a) are two slices of original group 1 image. (b)~(d) are the SART reconstruction of different motion group projections, simulated gas bubble motion artifacts can be seen around the boundary of the gas bubble in the images. (e)~(g) are the SMEIR results which reduced the artifact around the gas bubble boundaries.

scanning, we used two 3D abdominal CT images for a same patient acquired with 3 mins interval as the base images in this study. Dimension of the CT images are both 512×512×256, with voxel size of 0.125×0.125×0.192cm^3. The deformation vector field (DVF) between these two scans were calculated in the image domain and used as $v^{1\to N}$, and other N-2 (N=4, 8, 16) different DVFs were generated through randomly scaling $v^{1\to N}$in *x*, *y*, *z* directions by [-2, 2] to the simulate different DVFs. Then, the first CT image was deformed to *N-2* different CT images corresponding to the simulated DVFs. A *N* group 360° full-fan CBCT projection was then simulated through combining the projection images from different CT images, where projections from every continuous 360°/*N* beam angle are generated from the same CT image using the ray tracing technique. The total number of projection images are set as 360, detector size is 1280 × 318 with pixel size 0.0772 × 0.1344cm^2, distance from source to detector is 154cm, distance from source to isocenter is 100cm. To explore the effect of selection of group number *N*, *N*=4, 8, 16 were simulated in our experiment. For image reconstruction, image dimension of 256 × 256 × 128 with voxel size of 0.25×0.25×0.25cm^3 were used, and we ran 20 iteration of SMEIR for all the experiments.

III. Results

We showed the original group 1 image sample slices and reconstructed image with different methods and motion groups in Fig. 3. The real gas bubble artifacts in an on-board CBCT abdominal scan are shown in Fig. 1, while the simulated gas bubble artifacts in 4-, 8- and 16-groups simulated CBCT scan with SART reconstruction are shown in Fig. 3 (b)~(d). As the simulated motion mainly occurred on a single large air bubble (stomach) the pattern of artifacts in these images are more similar to the third image in Fig. 1. The corresponding reconstructed image with SMEIR is shown in Fig. 3 (e)~(g).

As the number of motion groups increase, in the SART reconstructed images, the visual quality of the gas bubble boundary became more blurred, and small streak artifact can be seen when *N*=16. Compared to the SART reconstruction results, the boundaries of reconstructed gas bubbles with SMEIR are well preserved and HU values around bubble regions are more accurate when compared with the original group 1 images, no obvious streak artifact can be seen.

IV. Discussions and Conclusion

In this study we modified SMEIR for gas bubble motion artifact reduction. From the simulation experiment, we can clear see the image quality improvement in the gas bubble boundary area by using SMEIR over SART.

There are several limitations of the current study. First, scattering is an important factor that influences CBCT imaging quality, for a more realistic simulation, we need to consider it in a future study. Second, given the inherent irregularity of gas bubble motion in both temporal domain and spatial domain, it's difficult to simulate the gas bubble motion. Our next step will be more focusing on artifact reduction using real clinical data. Third, dividing the projection data into projection groups equally is not an optimal strategy. The larger the *N* is, the smaller the motion within each projection group, and smaller angle coverage. Although big *N* can improve the data consistency within each projection group, small projection angle coverage might not provide enough information to update the DVF. Therefore, a data-driven projection group partition method is desired.

Deep learning-based methods have showed promising performance on both medical image reconstruction and post-processing tasks, and serval cycleGAN-style networks were proposed to synthesis high-quality images using CBCT for CBCT-guided adaptive radiotherapy [4, 10]. Although not designed for gas bubble motion artifact reduction, some of them performed well in the abdominal region [4]. One of our future works will be incorporating deep learning methods to the workflow of gas bubble motion artifact reduction and motion estimation.

In conclusion, we introduced SMRIE technique for gas bubble motion artifact reduction, the improved image quality of the gas bubble area in this study demonstrated that SMRIE technique is promising for on-board CBCT gas bubble motion artifact reduction.

V. REFERENCES

[1] G. X. Ding *et al.*, "A study on adaptive IMRT treatment planning using kV cone-beam CT," *Radiotherapy and Oncology,* vol. 85, no. 1, pp. 116-125, 2007.

[2] J. Wang and X. Gu, "Simultaneous motion estimation and image reconstruction (SMEIR) for 4D cone‐beam CT," *Medical physics,* vol. 40, no. 10, p. 101912, 2013.

[3] Y. Li *et al.*, "Dosimetric benefit of adaptive re-planning in pancreatic cancer stereotactic body radiotherapy," *Medical Dosimetry,* vol. 40, no. 4, pp. 318-324, 2015.

[4] Y. Liu *et al.*, "CBCT‐based synthetic CT generation using deep‐attention cycleGAN for pancreatic adaptive radiotherapy," *Medical physics,* vol. 47, no. 6, pp. 2472-2483, 2020.

[5] N. C. Estabrook, J. B. Corn, M. M. Ewing, H. R. Cardenes, and I. J. Das, "Dosimetric impact of gastrointestinal air column in radiation treatment of pancreatic cancer," *The British journal of radiology,* vol. 91, no. xxxx, p. 20170512, 2017.

[6] J. S. Lee, S. H. Kim, J. J. Kim, B. S. Kim, G. M. Choi, and D. R. Kim, "Gastrointestinal Air Motion Artifact Which Can Be Mistaken for Active Gastrointestinal Bleeding in Multidetector Computed Tomography: Phantom and Clinical Study," *Journal of computer assisted tomography,* vol. 44, no. 1, pp. 145-152, 2020.

[7] S. Winklhofer *et al.*, "Reduction of peristalsis-related gastrointestinal streak artifacts with dual-energy CT: a patient and phantom study," *Abdominal radiology,* vol. 41, no. 8, pp. 1456-1465, 2016.

[8] F. Liu, C. Cuevas, A. A. Moss, O. Kolokythas, T. J. Dubinsky, and P. E. Kinahan, "Gas bubble motion artifact in MDCT," *American Journal of Roentgenology,* vol. 190, no. 2, pp. 294-299, 2008.

[9] R. Clackdoyle and L. Desbat, "Data consistency conditions for truncated fanbeam and parallel projections," *Medical physics,* vol. 42, no. 2, pp. 831-845, 2015.

[10] X. Liang *et al.*, "Generating synthesized computed tomography (CT) from cone-beam computed tomography (CBCT) using CycleGAN for adaptive radiation therapy," *Physics in Medicine & Biology,* vol. 64, no. 12, p. 125002, 2019.

Comparing One-step and Two-step Scatter Correction and Density Reconstruction in X-ray CT

Alexander N. Sietsema, Michael T. McCann, Marc L. Klasky, and Saiprasad Ravishankar

***Abstract*—In this work, we compare one-step and two-step approaches for X-ray computed tomography (CT) scatter correction and density reconstruction. X-ray CT is an important imaging technique in medical and industrial applications. In many cases, the presence of scattered X-rays leads to loss of contrast and undesirable artifacts in reconstructed images. Many approaches to computationally removing scatter treat scatter correction as a preprocessing step that is followed by a reconstruction step. Treating scatter correction and reconstruction jointly as a single, more complicated optimization problem is less studied. It is not clear from the existing literature how these two approaches compare in terms of reconstruction accuracy. In this paper, we compare idealized versions of these two approaches with synthetic experiments. Our results show that the one-step approach can offer improved reconstructions over the two-step approach, although the gap between them is highly object-dependent.**

***Index Terms*—Computed tomography, computational imaging, density estimation, scatter correction, model-based iterative reconstruction.**

I. Introduction

THE presence of scattered X-rays presents a challenge for X-ray computed tomography (CT) imaging systems. For example, in the context of medical cone-beam CT, scatter causes a loss in soft-tissue contrast and artifacts such as cupping, streaks, bars, and shadows [1]. The same artifacts appear in nondestructive testing applications of X-ray CT, where they can interfere with subsequent quantification tasks [2]. There is a large body of work on preventing scatter using hardware and on correcting it using software; see [1], [3] for a review.

Hardware approaches to scatter correction include collimation (blocking unwanted X-rays at the source, thereby preventing them from contributing to scatter) or increasing the distance between the source and detector, which reduces the amount of scattered radiation reaching the detector. Among software approaches to scatter correction, a key distinction is whether scatter correction happens as a preprocessing step before CT reconstruction or jointly with CT reconstruction. In the former case, which we term *two-step* reconstruction, scatter is typically modeled as a function of the direct (i.e., not scattered) radiograph. This model is used to remove scatter from the measured data and estimate the direct radiograph, which is subsequently used for reconstruction. In the latter case, which we term *one-step* reconstruction, a model of scatter is included in a model-based CT reconstruction algorithm.

In our literature review, we found that the two-step formulation is the more common approach, including among recent work, e.g., [4], [5], [6], [7]. A two-step approach is also implicitly assumed in works where only scatter correction is considered, e.g., [8]. The popularity of the two-step approach may be because it usually involves solving two simple optimization problems (scatter correction and reconstruction) rather than a challenging joint problem.

The one-step approach appears less well-studied. The model-based reconstruction formulation in [9] includes a scatter term, but it is assumed to be known. The authors in [10] iteratively alternate between reconstruction and scatter estimation, but do not formulate a joint optimization problem. The review [1] describes a joint scatter correction and reconstruction approach, but only in general terms and without implementing it.

The main goal of this work is to compare one-step and two-step reconstruction approaches to find when, if ever, the added complexity of one-step reconstruction yields better results. To this end, we compare idealized one- and two-step algorithms on synthetic data in our studies. These experiments are intentionally simple: scatter is modelled using a convolution with a Gaussian kernel, noise is Gaussian, and the beam is monoenergetic; we believe this experimental setup includes many of the key features of real descattering and reconstruction while leaving out aspects that may obscure the difference between the one-step and two-step approaches, e.g., inaccuracies in the radiographic forward model and beam hardening.

In the following sections, we formulate the scatter correction and reconstruction problem, describe our one- and two-step algorithms, and present our experiments, results, and conclusions.

II. Scatter Correction and Reconstruction

We focus on a model of 2D monoenergetic X-ray tomography that includes scatter. Given a (vectorized) density profile $\boldsymbol{\rho} \in \mathbb{R}^{N_1 N_2}$, our model of the total transmission $\boldsymbol{t}$ is

$$\boldsymbol{t} = \boldsymbol{s} + \boldsymbol{d}, \quad \boldsymbol{s} = \boldsymbol{K}\boldsymbol{d}, \quad \boldsymbol{d} = \exp(-\xi \boldsymbol{A}\boldsymbol{\rho}), \tag{1}$$

A. Sietsema is with the Dept. of Mathematics, Michigan State University, East Lansing, MI 48824, USA (sietsem6@msu.edu).

S. Ravishankar is with the Dept. of Computational Mathematics, Science and Engineering and the Dept. of Biomedical Engineering, Michigan State University, East Lansing, MI 48824, USA (ravisha3@msu.edu).

M. McCann and M. Klasky are with the Theoretical Division, Los Alamos National Laboratory, Los Alamos, NM 87545, USA (mccann@lanl.gov, mklasky@lanl.gov).

This abstract has been accepted for poster presentation at the CT Meeting 2022.

where $\boldsymbol{s}$ is the scattered signal, $\boldsymbol{d}$ is the direct (i.e., scatter-free) signal, $\boldsymbol{A} \in \mathbb{R}^{M \times N_1 N_2}$ is the X-ray transform, ξ represents the mass attenuation coefficient for a given material, $\exp(\cdot)$ is applied element-wise, and $\boldsymbol{K}$ represents linear convolution by a kernel $\boldsymbol{k}$, used to approximate scatter.

The choice to model scatter as a kernel convolved with the direct is common in the scatter correction literature [11], [7], [12], [13], [14]. This provides a fast scatter model that is at least representative of models used in practice. As our main goal is to bring out the differences between the one-step and two-step formulations, we leave more complicated models for future investigation.

A. One-step vs two-step scatter correction and reconstruction

Two-step Approach: The two-step method involves first solving a scatter correction problem that is followed by solving a density reconstruction problem. We formulate the first step, i.e. scatter correction, as

$$\boldsymbol{d}^* = \underset{\boldsymbol{d}}{\operatorname{argmin}} \|\boldsymbol{t} - (\boldsymbol{K}\boldsymbol{d} + \boldsymbol{d})\|_2^2, \tag{2}$$

where we minimize an ℓ_2 fit between the measured transmission and the model for it (i.e., direct + scatter) to account for noisy data. This step would correct for scatter in all the measured (one or multiple) CT views.

Following scatter correction, we invert the nonlinear part of (1) with the elementwise operation, $\boldsymbol{\rho}_A^*[m,n] = -\log(\boldsymbol{d}^*[m,n])/\xi$. Note that we would need to set any values where $\boldsymbol{d}^*[m,n]$ was less than or equal to zero to zero, as the logarithm would be invalid there. We use the symbol $\boldsymbol{\rho}_A^*$ here because this quantity represents the areal density [7].

After performing this scatter correction, we solve the following optimization problem to reconstruct the underlying density:

$$\boldsymbol{\rho}^* = \underset{\boldsymbol{\rho}}{\operatorname{argmin}} \|\boldsymbol{\rho}_A^* - \boldsymbol{A}\boldsymbol{\rho}\|_2^2 + \alpha R(\boldsymbol{\rho}), \tag{3}$$

where R is a regularization functional and α is a nonnegative parameter. In essence, this approach first estimates the direct $\boldsymbol{d}^*$ by removing scatter from the transmission $\boldsymbol{t}$, and then takes that estimate and uses it to reconstruct an estimate of the object density $\boldsymbol{\rho}^*$.

The main advantage of the two-step method is its simplicity: both the first and second step are well-studied formulations of linear inverse problems that can be readily solved with standard algorithms. The scatter correction step involves linear least squares (with possible constraints) for which there are many good algorithms, e.g., a fixed-point method such as Jacobi iteration [15], the conjugate gradient method [15] or ADMM [16]. The reconstruction step can be solved efficiently using ADMM.

One-step Approach: One-step scatter correction and reconstruction involves jointly optimizing over the entire forward model rather than first optimizing for the direct radiograph. This amounts to solving the following optimization problem:

$$\boldsymbol{\rho}^* = \underset{\boldsymbol{\rho}}{\operatorname{argmin}} \|\boldsymbol{t} - (\boldsymbol{K} + \boldsymbol{I})\exp(-\xi\boldsymbol{A}\boldsymbol{\rho})\|_2^2 + \alpha R(\boldsymbol{\rho}), \tag{4}$$

where R is again a regularization functional. Intuitively, the one-step method has the benefit of not relying on the estimation of the direct: in the two-step method, our overall estimate of the density is limited by the estimate of $\boldsymbol{d}^*$ we obtain from solving (2). However, the one-step optimization could be more challenging depending on the complexity of the entire forward model.

B. Implementation of one-step and two-step methods

In our implementation of the one-step and two-step methods described above, we consider imaging a spherically symmetric, single-material object, parameterized by its radial profile $\boldsymbol{\rho} \in \mathbb{R}^N$. This simple model captures the important elements of X-ray reconstruction with scatter, while remaining fast to optimize; similar models find application in nondestructive testing [17] and have been used for developing scatter correction methods [7]. As a result, our operator $\boldsymbol{A}$ is the forward Abel transform and is followed by spinning the 1D signal into a 2D image on which the convolution (for scatter) is applied. For an example of data generated with this model, see Figure 1.

For regularization, (R in (3) and (4)) we use total variation on the profile, i.e., $R(\boldsymbol{\rho}) = \sum_{k=1}^{n} |\boldsymbol{\rho}[k] - \boldsymbol{\rho}[k-1]|$.

In both approaches, we solve the underlying optimization problems using the LBFGS algorithm in PyTorch [18]. Note that PyTorch automatically handles non-differentiability issues. Additionally, due to the poor conditioning of the Abel matrix $\boldsymbol{A}$, we used the separable quadratic surrogate (SQS) preconditioning as follows:

$$\boldsymbol{P} = \operatorname{diag}\left(\boldsymbol{A}^\top \boldsymbol{A}\mathbf{1}\right)^{-1}, \tag{5}$$

where $\mathbf{1} \in \mathbb{R}^N$ denotes a vector of ones. This preconditioner is used for performing the one-step optimization and the second step of the two-step optimization. We apply this preconditioning by premultiplying $\boldsymbol{\rho}$ by $\boldsymbol{P}$ prior to optimizing. For the one-step method, this then becomes the preconditioned optimization formulation

$$\boldsymbol{\rho}' = \underset{\boldsymbol{\rho}}{\operatorname{argmin}} \|\boldsymbol{t} - (\boldsymbol{K} + \boldsymbol{I})\exp(-\xi\boldsymbol{A}\boldsymbol{P}\boldsymbol{\rho})\|_2^2 + \alpha \sum_{k=1}^{n} |\boldsymbol{P}\boldsymbol{\rho}[k] - \boldsymbol{P}\boldsymbol{\rho}[k-1]|, \tag{6}$$

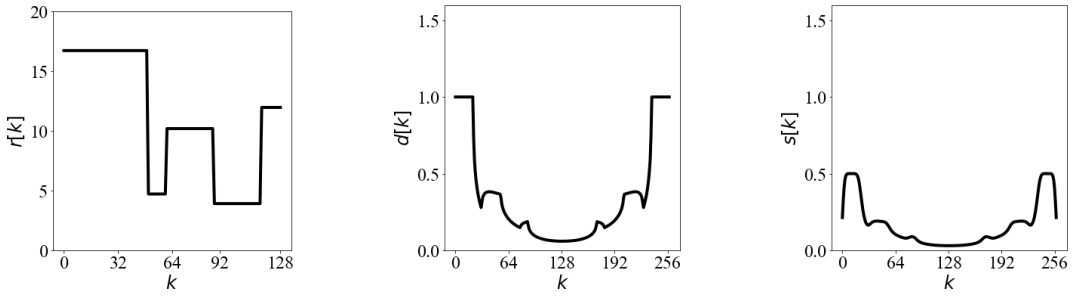
where the solution is recovered via $\boldsymbol{\rho}^* = \boldsymbol{P}\boldsymbol{\rho}'$; a similar formulation is used for the two-step problem.

III. EXPERIMENTS AND RESULTS

We compare the one-step and two-step algorithms by scatter correcting and reconstructing ten synthetically generated transmissions. In all experiments, we quantify performance by taking the root mean square error (RMSE) between the ground truth density and the reconstructed density for each algorithm.

A. Data generation

To generate test objects, we first randomly selected indices in 1D at which shells start and end, assuming a maximum radius of $N = 129$ pixels of the object. We converted



(a) Density (b) Direct (c) Scatter

Fig. 1: Density, direct, and scatter lineouts for profile 4 in our data set.

these indices into a piecewise-constant profile with steps at each shell boundary. Finally, we picked shell densities (in $\mathbb{R}$) uniformly in the range (0, 20), and assigned each shell a density. These 1D radial profiles were spun to create 2D images (representing slice of 3D volume) as part of the forward projection process [19]. The spinning process slightly cropped the images to avoid edge effects.

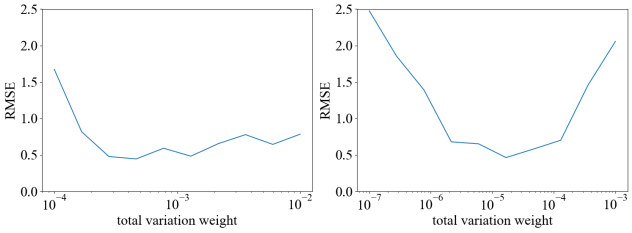
In order to generate transmissions, we applied the forward model in (1), choosing $\xi = 1 \times 10^{-3}$ and fixing a Gaussian scatter kernel. In our implementation, this kernel is a three-fold convolution with a 7×7 Gaussian blur with standard deviation 1.5 pixels. Spinning occurs after $\boldsymbol{A\rho}$ is computed. Finally, Gaussian noise is added in the end with $\mu = 0, \sigma = 3 \times 10^{-2}$, and negative values in the transmission are set to zero (non-physical). See Figure 1 for an example of the density, direct, and scatter profiles.

B. Implementation details

All experiments were performed in Python using simulated data, and optimization is performed using the PyTorch package. The one-step algorithm was run with a learning rate (step size of Wolfe line search) of 3×10^{-2} and a total variation weight of 1×10^{-3} with 20 iterations. The two-step algorithm used a learning rate of 1.0 and no total variation with 10 iterations for the first step, and a learning rate of 1×10^{-2} and a total variation weight of 7×10^{-4} with 20 iterations for the second step. All parameters were optimized by performing a grid search over possible combinations of learning rate and total variation parameters. See Figure 2 for the results of the grid search. The forward Abel transform was the Hansenlaw method [20] from the PyAbel Python package [19].

C. Results

Results for each of the ten profiles are summarized in Table I. Overall, the one-step algorithm outperforms the two-step algorithm, with a median RMS error of 1.548, compared to 2.575 for the latter. Two example reconstructions are shown in Figure 3. Qualitatively, the two-step method can be quite noisy, especially near the center of the reconstruction, despite the regularization. However, the two-step model may reconstruct thin shells better (Figure 3b), although this benefit is on the whole negligible.

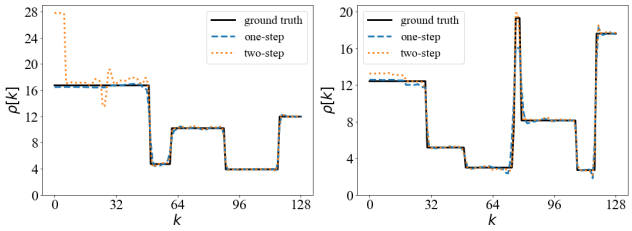


(a) One-step (b) Two-step

Fig. 2: Reconstruction accuracy (RMSE) as a function of the total variation weight (α in (3) and (4)). The optimal learning rate was tuned independently for each weight value. The minima achieved are used in later fitting.

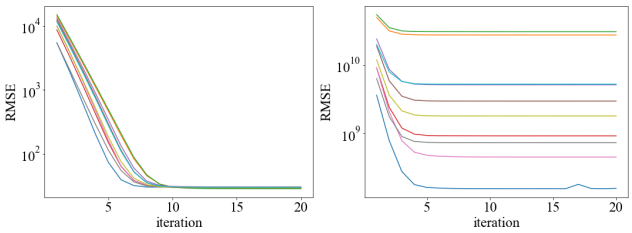
TABLE I: One-step vs. two-step reconstruction performance.

Profile	One-step RMSE	Two-step RMSE
1	**0.827**	1.445
2	**2.077**	9.242
3	**2.021**	9.552
4	**0.568**	2.468
5	**1.210**	2.683
6	**2.488**	3.071
7	0.853	**0.552**
8	**0.422**	0.623
9	**1.887**	2.397
10	**2.305**	3.491



(a) Profile 4 (b) Profile 7

Fig. 3: Comparison of one-step and two-step reconstruction results on two profiles.



(a) One-step (b) Two-step (second step only)

Fig. 4: Comparison of one-step and two-step (second step only) convergence rates for each of the ten profiles. Satisfactory convergence is demonstrated in all cases. Note that the two-step loss is calculated in the areal density space, so its values are much higher than the one-step method's.

Both methods tend to struggle with profiles with high-density (densities above 8) shells near the center. This is likely due to the conditioning effects of the Abel matrix, which the SQS preconditioning is only partially able to solve. Uniformly high-density profiles can also cause some problems in both methods (e.g., profiles 2 and 3). This may be an artifact of our noise model, as denser objects produce smaller transmissions, which results in a low signal-to-noise-ratio when noise with a constant variance is added.

Finally, we summarize convergence results, where we plot the data-fidelity terms in (3) and (4) over the LBFGS iterations. From Figure 4, we see that both algorithms converge quickly. We omitted the first step for the two-step method here since it converges nearly instantly and we plotted for only the one-step method and the second step of the two-step method.

IV. CONCLUSIONS

In this paper, we performed an empirical comparison of one-step and two-step X-ray CT scatter correction and reconstruction. Our experiments showed that the one-step approach can demonstrate significant improvements over the common two-step approach. While certainly not exhaustive, these experiments suggest that the added complexity of the one-step method may be worth it. Our future work in this area would extend these experiments to more complicated regimes, including using more complicated scatter estimation models (e.g., [7]), polyenergetic spectra, and multiple materials. The same comparisons could be also run on more realistic synthetic data (e.g., generated from particle transport simulations as in [7]) to validate the efficacy of one-step descattering in a setting where the scatter model used during reconstruction does not perfectly match the scatter in the data. Finally, we aim to run similar comparisons on real, experimental data.

REFERENCES

[1] E.-P. Rührnschopf and K. Klingenbeck, "A general framework and review of scatter correction methods in X-ray cone-beam computerized tomography. part 1: Scatter compensation approaches," *Med. Phys.*, vol. 38, no. 7, pp. 4296–4311, June 2011.

[2] J. J. Lifton, A. A. Malcolm, and J. W. McBride, "An experimental study on the influence of scatter and beam hardening in X-ray CT for dimensional metrology," *Meas. Sci. Technol.*, vol. 27, no. 1, pp. 015007, Dec. 2015.

[3] E.-P. Rührnschopf and K. Klingenbeck, "A general framework and review of scatter correction methods in cone beam CT. part 2: Scatter estimation approaches," *Med. Phys.*, vol. 38, no. 9, pp. 5186–5199, Aug. 2011.

[4] N. Bhatia, D. Tisseur, and J. M. Létang, "Convolution-based scatter correction using kernels combining measurements and monte carlo simulations," *J Xray Sci Technol.*, vol. 25, pp. 613–628, 2017.

[5] J. Maier, S. Sawall, M. Knaup, and M. Kachelrieß, "Deep scatter estimation (DSE): Accurate real-time scatter estimation for X-ray CT using a deep convolutional neural network," *J. Nondestruct. Eval.*, vol. 37, no. 3, July 2018.

[6] Y. Nomura, Q. Xu, H. Shirato, S. Shimizu, and L. Xing, "Projection-domain scatter correction for cone beam computed tomography using a residual convolutional neural network," *Med Phys*, vol. 46, no. 7, pp. 3142–3155, jun 2019.

[7] M. T. McCann, M. L. Klasky, J. L. Schei, and S. Ravishankar, "Local models for scatter estimation and descattering in polyenergetic X-ray tomography," *Opt. Express*, vol. 29, no. 18, pp. 29423–29438, 2021.

[8] D. Tisseur, N. Bhatia, N. Estre, L. Berge, D. Eck, and E. Payan, "Evaluation of a scattering correction method for high energy tomography," *EPJ Web Conf.*, vol. 170, pp. 06006, 2018.

[9] J. Nuyts, B. Man, J. A. Fessler, W. Zbijewski, and F. J. Beekman, "Modelling the physics in the iterative reconstruction for transmission computed tomography," *Phys. Med. Biol.*, vol. 58, no. 12, pp. R63–R96, June 2013.

[10] E. Mainegra-Hing and I. Kawrakow, "Fast monte carlo calculation of scatter corrections for CBCT images," *J. Phys.: Conf. Ser*, vol. 102, pp. 012017, Feb. 2008.

[11] M. Sun and J. M. Star-Lack, "Improved scatter correction using adaptive scatter kernel superposition," *Phys. Med. Biol*, vol. 55, no. 22, pp. 6695–6720, Oct. 2010.

[12] Roland E Suri, Gary Virshup, Luis Zurkirchen, and Wolfgang Kaissl, "Comparison of scatter correction methods for CBCT," in *SPIE Med. Imaging*. International Society for Optics and Photonics, 2006, vol. 6142, p. 614238.

[13] B Ohnesorge, T Flohr, and K Klingenbeck-Regn, "Efficient object scatter correction algorithm for third and fourth generation CT scanners," *Eur. Radiol*, vol. 9, no. 3, pp. 563–569, 1999.

[14] L Alan Love and Robert A Kruger, "Scatter estimation for a digital radiographic system using convolution filtering," *Med. Phys*, vol. 14, no. 2, pp. 178–185, 1987.

[15] J. R. Shewchuk, "An introduction to the conjugate gradient method without the agonizing pain," Tech. Rep., Carnegie Mellon University, 1994.

[16] S. Boyd, N. Parikh, E. Chu, B. Peleato, and J. Eckstein, "Distributed optimization and statistical learning via the alternating direction method of multipliers," *Found. Trends Mach. Learn.*, vol. 3, no. 1, pp. 1–122, Jan. 2011.

[17] T. J. Asaki, P. R. Campbell, R. Chartrand, C. E. Powell, K. R. Vixie, and B. E. Wohlberg, "Abel inversion using total variation regularization: applications," *Inverse Probl. Sci. Eng.*, vol. 14, no. 8, pp. 873–885, Dec. 2006.

[18] A. Paszke, S. Gross, F. Massa, A. Lerer, J. Bradbury, G. Chanan, T. Killeen, Z. Lin, N. Gimelshein, L. Antiga, A. Desmaison, A. Kopf, E. Yang, Z. DeVito, M. Raison, A. Tejani, S. Chilamkurthy, B. Steiner, L. Fang, J. Bai, and S. Chintala, "Pytorch: An imperative style, high-performance deep learning library," in *Advances in Neural Information Processing Systems*, H. Wallach, H. Larochelle, A. Beygelzimer, F. d'Alché-Buc, E. Fox, and R. Garnett, Eds. 2019, vol. 32, Curran Associates, Inc.

[19] S. Gibson, D. D. Hickstein, R. Yurchak, M. Ryazanov, D. Das, and G. Shih, "Pyabel/pyabel: v0.8.4," 2021.

[20] E. W. Hansen and P. Law, "Recursive methods for computing the Abel transform and its inverse," *J. Opt. Soc. Am. A: Opt. Image Sci. Vis.*, vol. 2, no. 4, pp. 510, Apr. 1985.

Material decomposition from unregistered dual kV data using the cOSSCIR algorithm

Benjamin M. Rizzo, Emil Y. Sidky, and Taly Gilat Schmidt

***Index Terms*—Dual and multi-energy imaging, Image reconstruction, Photon counting CT.**

I. Introduction

THE constrained one-step spectral CT Image Reconstruction method (cOSSCIR) has been developed to estimate basis material maps directly from spectral CT data [1], [7]. cOSSCIR includes modeling of polyenergetic x-ray transmission, which can prevent beam hardening artifacts. Also, constraints can be placed on the basis material maps to stabilize the material decomposition inversion. The cOSSCIR framework has been investigated for spectral data from photon counting detectors. Recently, preliminary work has applied cOSSCIR to dual energy CT using an integrating detector model, where the dual energy problem is modeled in the cOSSCIR framework as a two-window spectral CT problem with significant overlap between the acquisition spectra. Previous work assumed the spectral data are fully registered (i.e., all spectral measurements are collected for each ray), as is generally the case with current photon-counting detector technologies. However, clinical dual energy diagnostic CT or cone-beam CT data may be acquired using Rapid- or Slow-kV acquisitions, resulting in unregistered spectral CT data. Typically, unregistered dual kV acquisitions are reconstructed by a two-step approach that first reconstructs CT images from each spectra and then performs image-domain material decomposition. However, such approaches cannot take advantage of polyenergetic modeling and are susceptible to beam hardening artifacts. One advantage of one-step direct inversion material decomposition methods such as cOSSCIR is that polyenergetic transmission can be modeled while the spectral data need not be registered. This study investigates the application of cOSSCIR to unregistered, dual energy acquisitions. First, an inverse crime simulation using a pelvic phantom was performed to determine whether the cOSSCIR optimization algorithm converges for the case of rapid switching dual kV. Our results demonstrate that convergence is possible for the unregistered, dual energy, problem using cOSSCIR. We further demonstrate the approach on a preliminary experimental dataset.

This work was supported in part by NIH Grant Nos. ROI-R01EB023968

B. M. Rizzo is with the Department of Biomedical Engineering at Marquette University and the Medical College of Wisconsin, Milwaukee, WI, 53233 (e-mail: benjamin.rizzo@marquette.edu).

E. Y. Sidky is with the Department of Radiology at The University of Chicago, Chicago, IL, 60637 (e-mail: sidky@uchicago.edu).

T. G. Schmidt is with the Department of Biomedical Engineering at Marquette University and the Medical College of Wisconsin, Milwaukee, WI, 53233 (e-mail: tal.gilat-schmidt@marquette.edu).

II. Purpose

The goal of this study is to investigate the use of cOSSCIR on unregistered dual kV data, such as that acquired by rapid kV switching acquisition approaches. Specifically, this investigation will focus on inverse-crime studies and experimental reconstructions to establish the feasibility of using cOSSCIR on unregistered dual-kV data.

III. Methods

The cOSSCIR algorithm is designed to solve the concave-convex optimization problem [1], [2], [3],

$$x^*_{km} = \text{argmin}\Big\{ D(c_{w\ell}, \hat{c}_{w\ell}) \Big\} \; : \sum_m \|x_m\|_{TV} \leq \gamma,$$

where the objective is to find the optimal basis material images x^*_{km} that minimizes

$$D_{TPL}(c_{w\ell}, \hat{c}_{w\ell}) = \sum_{w\ell} [\hat{c}_{w\ell}(y) - c_{w\ell} \, \log(\hat{c}_{w\ell}(y))],$$

where D_{TPL} is the Transmission data discrepancy derived from a Poisson likelihood (TPL). The data model utilized by the method is

$$\hat{c}_{w\ell} = N_{w\ell} \sum_i S_{w\ell i} \exp(-\mu_{mi} P_{\ell k} x_{km}),$$

where $\hat{c}_{w\ell}$ represents the mean photon counts along a ray ℓ in energy window w. The variable $N_{w\ell}$ represents the number of counts along ray ℓ in energy window w in the absence of an object, and $S_{w\ell i}$ is the normalized x-ray energy spectrum. $P_{\ell k}$ is the x-ray projection operator. Finally, μ_{mi} represents the attenuation of material m at energy E_i. This one-step direct inversion [2], [5], [6] method for CT has been investigated for the photon counting problem. The method was found to be more stable than traditional two-step approaches as it utilizes all rays in the reconstruction, and places constraints on the basis images x_{km} directly. In the one-step approach, the spectral measurements need not be registered, and cOSSCIR may potentially be applied to dual energy CT. Dual energy CT, therefore, presents a unique opportunity to test cOSSCIR on the unregistered reconstruction problem. In the case of dual energy CT, the spectral model requires minimal modification to account for an energy integrating detector. The spectra $S_{w\ell i}$ may be replaced by

$$\hat{S}_{w\ell i} = \beta_{w\ell i} \, S_{w\ell i} \, E_i,$$

where $\beta_{w\ell}$ represents the detector gain along the transmission path ℓ in energy window w. The effect of unregistered data was investigated by simulating data from a pelvis phantom, consisting of a bone and water basis material maps, as shown in Fig. 1. Using a polyenergetic x-ray transmission

model [7], projections at 80 and 140 kV data were simulated using realistic x-ray spectra. One simulation generated fully registered noiseless data for spectra at 128 equally spaced views. A second simulation generated data at 256 equally spaced views, but with the 80 and 140 kV spectra alternating for each view angle. This second simulation models a rapid kV acquisition approach. Both sets of simulation data were

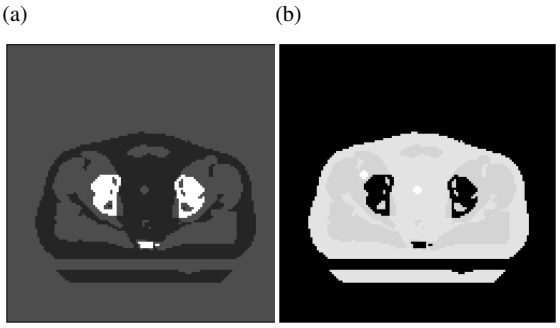


Fig. 1: Pelvis phantom; (a) The bone basis material map, and (b) the water basis material map.

reconstructed using the cOSSCIR framework, and the same x-ray transmission model was used to simulate the data. During the reconstructions, two error metrics were monitored with iteration; the Root-Mean Squared-Error (RMSE) between the phantom basis images and the basis map estimates, and the TPL between the measured data and the modeled data.

Additionally, Dual-kV data was obtained from an experimental micro-CT system using a high resolution flat panel detector (Varian PaxScan 2520 DX) and a micro-focal x-ray source (Hamamatsu 9181-02). A cylindrical phantom with PMMA, Polystyrene, and Teflon inserts was imaged. For each phantom, fully registered projections were acquired from 500 equally spaced view angles using 80 and 130 kV.

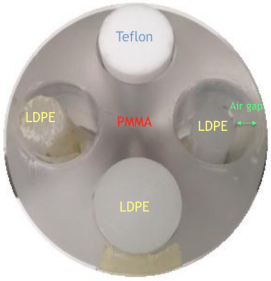


Fig. 2: Rod phantom consisting of PMMA, LDPE, and Teflon

A rapid-kV-switching acquisition was then simulated by reconstructing from only 250 views for each spectra, with the spectra alternating across angle. For each subsampled dataset, the 80 and 130 kV sinograms were offset by a single view angle. To further test the method, Slow-kV-switching data was subsampled from the registered data by alternating between the 80 and 130 kV sinograms every 10 consecutive views. Reconstructions were then performed using the cOSSCIR algorithm using aluminum and PMMA as the basis materials. For each basis material image, Regions-of-interest (ROIs) were placed over the LDPE, PMMA, and Teflon regions, and the mean value in each material's ROI was measured. The percent error was taken between the ROI measurements and the predicted ground truth material decomposition values obtained from the XCOM NIST database [4].

IV. Results

cOSSCIR was performed on the noiseless data to investigate the effect of unregistered data on the reconstructions, and the resulting material maps are shown in Fig. 3 for 8000 iterations. Rows (b) shows the results of the unregistered reconstructions. The top row (a) shows the results obtained from using fully registered 128 view data for comparison.

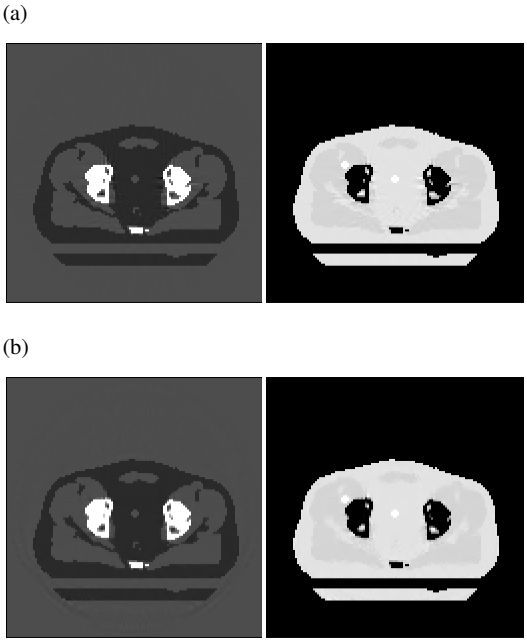


Fig. 3: Pelvis phantom reconstructions for (a) registered 128 views, (b) Rapid-kV 128 views/kV. The columns show the bone and water basis reconstructions, respectively.

In general, the basis images are recovered well for the unregistered 128 views/kV in Fig. 3(b), compared to the registered reconstruction using 128 views/kV in Fig. 3(a). The algorithm performance was further evaluated by monitoring the data error and basis image error with iteration, and are plotted in Fig. 4 for 128 views/kV.

As in the case of the reconstructions, the basis image error indicate good recovery after 8000 iterations. The RMSE between the phantom and basis images suggest the registered and unregistered 128 view/kV reconstructions are approaching the same solution. More iterations are required, in the case of the unregistered data, before the image RMSE approaches the same values as the registered reconstructions. Although the data error is decreasing with iteration in Fig. 4(a), in general, the TPL also indicates more iterations were required in the unregistered case, suggesting there is a penalty to using fewer projections per spectra compared to the fully registered

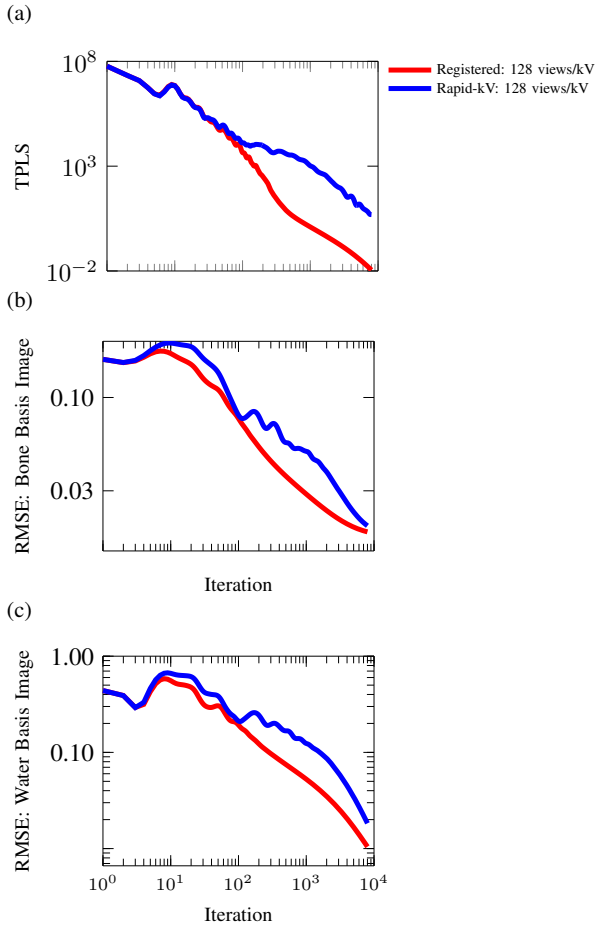


Fig. 4: (a) Data error, (b) RMSE for the bone basis image, (c) RMSE for the water basis image.

case. An interesting aspect of these unregistered inverse-crime studies was convergence to a solution was only observed using μ-preconditioning [1], where the material attenuation energy functions $\mu(E)$ were orthogonalized. μ-preconditioning was not required for convergence on the registered data, but was used in the results presented above.

Experimental data was also reconstructed as a preliminary demonstration cOSSCIR applied to dual-kV data. 500 view dual energy data were acquired, then unregistered rapid- and slow-kV data were simulated by applying masks to the fully registered data for a total of 256 views/kV. A 256 views/kV registered dataset was also constructed from the 500 view acquisition. Results for the 256 views/kV reconstructions are shown in Fig. 5 using aluminum and PMMA as the basis materials for 8000 iterations. μ -preconditioning was also used to perform all reconstructions. Compared to the registered results in Fig. 5, the recovered rapid- and slow-kV PMMA basis image and virtual mono-energetic images are quite similar. There are noticeable ray artifacts in the rapid- and slow-kV aluminum basis images. However, these appear to be largely in areas of the image where aluminum has limited contribution.

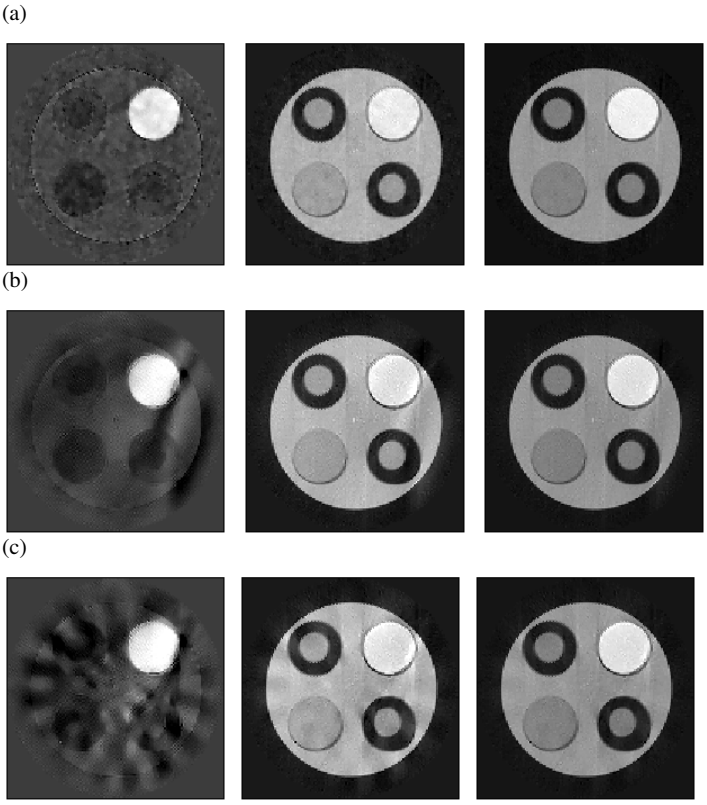


Fig. 5: Reconstructions using (a) Registered, (b) Rapid-kV, and (c) Slow-kV data. The columns display the aluminum, PMMA, and the 50 keV mono-energetic image.

Material quantification was performed on the LDPE, PMMA, and Teflon ROI's of the phantom. The percent errors between the recovered grayscale values and their ground truth for each basis are shown in Table I. The recovered material vectors are nearly identical for the registered and rapid-kV reconstructions, demonstrating that unregistered acquisition does not impact quantitative accuracy.

V. New And Breakthrough Work To Be Presented

This is the first demonstration and application of cOSSCIR on unregistered dual-kV data.

VI. Conclusions

Our preliminary results demonstrate the application of the one-step cOSSCIR algorithm to unregistered data. The inverse-crime studies demonstrate a challenging reconstruction prob-

TABLE I

	% Error	LDPE ROI	PMMA ROI	Teflon ROI
Registered	PMMA	1.11	0.80	1.76
	AL	-21.41	-	6.24
Rapid-kV	PMMA	0.99	0.20	0.58
	AL	-21.41	-	3.27
Slow-kV	PMMA	0.60	1.20	1.10
	AL	-2.92	-	0.45

lem, compared to reconstructing registered data. More iterations and orthogonalized material attenuation energy functions are required for basis image recovery. Our results suggest this may be due to using fewer view per spectra. Finally, the inverse-crime results were extended to experimental unregistered dual-kV data using a physical rod phantom. The results demonstrate that the cOSSCIR may also be applied to experimental, unregistered, dual-kV data. Furthermore, the unregistered data appears to have minimal impact on material basis quantification using 256 views/kV. Investigations into improving the basis image quality and material decomposition are currently underway. Work is underway to compare the cOSSCIR results to a two-step image-domain decomposition method.

REFERENCES

[1] R.F. Barber, and E.Y. Sidky, "MOCCA: Mirrored Convex/Concave Optimization for Nonconvex Composite Functions," Journal of Machine Learning Research, vol. 17, no. 144, pp. 1-51, 2016

[2] R.F. Barber, and E.Y. Sidky, T.G. Schmidt, "An algorithm for constrained one-step inversion of spectral CT data," Phys. Med. Biol., vol. 61, no. 10, pp. 3784-3818, 2016

[3] R.F. Barber, and E.Y. Sidky, "Convergence for nonconvex ADMM, with applications to CT imaging," arXiv: 2020/06. 2006.07278 [math]

[4] M.J. Berger, J.H. Hubbell, S.M. Seltzer, J. Chang, J.S. Coursey, R. Sukumar, D.S. Zucker, and K. Olsen (2010), XCOM: Photon Cross Section Database (version 1.5). [Online] Available: http://physics.nist.gov/xcom [2022, January 24]. National Institute of Standards and Technology, Gaithersburg, MD.

[5] C. Cai, T. Rodet, S. Legoupil, A. Mohammad-Djafari. "A full-spectral Bayesian reconstruction approach based on the material decomposition model applied in dual-energy computed tomography," Med Phys. vol. 40, no. 11. Nov. 2013

[6] I.A. Elbakri, J.A. Fessler. "Statistical Image Reconstruction for Polyenergetic X-Ray Computed Tomography," IEEE Trans. on Med. Imaging, vol. 21, no. 2, pp. 89-99, Feb. 2002

[7] T.G. Schmidt, R.F. Barber, and E.Y. Sidky, "A Spectral CT Method to Directly Estimate Basis Material Maps From Experimental Photon-Counting Data," IEEE Trans. Med. Imaging, vol. 36, no. 9, pp. 1808-18, Nov. 2017

[8] T.G. Schmidt, R.F. Barber, and E.Y. Sidky, "Spectral CT metal artifact reduction using weighted masking and a one step direct inversion reconstruction algorithm," Proceedings of the SPIE, Medical Imaging 2020, vol. 11312 Physics of Medical Imaging, 2020

PixelPrint: Three-dimensional printing of patient-specific soft tissue and bone phantoms for CT

Kai Mei, Michael Geagan, Nadav Shapira, Leening P. Liu, Pouyan Pasyar,
Grace J. Gang, J. Webster Stayman, and Peter B. Noël

Abstract— **Patient-based CT phantoms, with realistic image texture and densities, are essential tools for assessing and verifying CT performance in clinical practice. This study extends our previously presented 3D printing solution (PixelPrint) to patient-based phantoms with soft tissue and bone structures. To expand the Hounsfield Unit (HUs) range, we utilize a stone-based filament. Applying PixelPrint, we converted patient DICOM images directly into FDM printer instructions (G-code). Density was modeled as the ratio of filament to voxel volume to emulate attenuation profiles for each voxel, with the filament ratio controlled through continuous modification of the printing speed. Two different phantoms were designed to demonstrate the high reproducibility of our approach with micro-CT acquisitions, and to determine the mapping between filament line widths and HU values on a clinical CT system. Moreover, a third phantom based on a clinical cervical spine scan was manufactured and scanned with a clinical spectral CT scanner. CT image of the patient-based phantom closely resembles the original CT image both in texture and contrast levels. Measured differences between patient and phantom are around 10 HU for bone marrow voxels and around 150 HU for cortical bone. In addition, stone-based filament can accurately represent boney tissue structures across the different x-ray energies, as measured by spectral CT. This study demonstrates the feasibility of our 3D-printed patient-based phantoms to be extended to soft-tissue and bone structure while maintaining accurate organ geometry, image texture, and attenuation profiles for spectral CT.**

Index Terms—**Computed Tomography, 3D printing, Image Quality Phantoms, Quality Assurance**

I. Introduction

Anthropomorphic patient-based phantoms are essential tools in computed tomography (CT) research and clinical practice. Academic and clinical CT communities would benefit from a fast and inexpensive manufacturing method to produce patient-based phantoms compared to currently available commercial solutions.

Over the last decade, several approaches for 3-dimensional (3D) printing of tissue-mimicking phantoms have been proposed for validation and evaluation of CT imaging technology [1]-[3]. We recently introduced PixelPrint [4], which has illustrated the capability to generate patient-specific lung phantoms with accurate attenuation profiles and textures. This method directly translates DICOM image data into G-code, eliminating the need for slicing software utilized in common conventional 3D printing methods. By this, our method does not require segmentation and triangulation of surface geometry models and thus enables the generation of sophisticated patient-based phantoms.

In this study, we applied PixelPrint with a new type of filament (StoneFil) to enable 3D printing of soft tissue and bone structures. The filament, consisting of 50% gravimetric powdered stone filling, is 37% denser than regular PLA. First, we manufactured three identical phantoms to evaluate the reproducibility of our printing patterns with a micro-CT system. Next, we generated a calibration phantom to calibrate printed filament ratios with Hounsfield Units (HU). Finally, a patient-specific cervical phantom was generated and evaluated with a clinical spectral CT system. We demonstrated that PixelPrint can readily and reliably produce realistic patient-based phantoms for representing various anatomical structures.

II. Materials and Methods

A. Printer and filament

In this study we used a fused-filament 3D printer (Lulzbot TAZ 6 with M175 tool head, Fargo Additive Manufacturing Equipment 3D, LLC Fargo, ND, USA) and a 0.40 mm brass nozzle. StoneFil filament with a diameter of 1.75 mm (FormFutura, AM Nijmegen, the Netherlands) was extruded at a nozzle temperature of 200 °C. To generate different x-ray attenuations, printing speeds were varied between 6.0 to 30 mm/s, producing line widths from 0.2 to 1.0 mm. Printer head acceleration was set to 500 mm/s^2 and the jerk setting (or acceleration threshold) was kept at 8 mm/s.

K. Mei, M. Geagan, N. Shapira, L. Liu, P. Pasyar, and P. B. Noël are with Department of Radiology, Perelman School of Medicine, University of Pennsylvania, Philadelphia, PA, USA

L. Liu is with Department of Bioengineering, University of Pennsylvania, Philadelphia, PA, USA

G. J. Gang, and J. W. Stayman are with Department of Biomedical Engi neering, Johns Hopkins University, Baltimore, MD, USA.

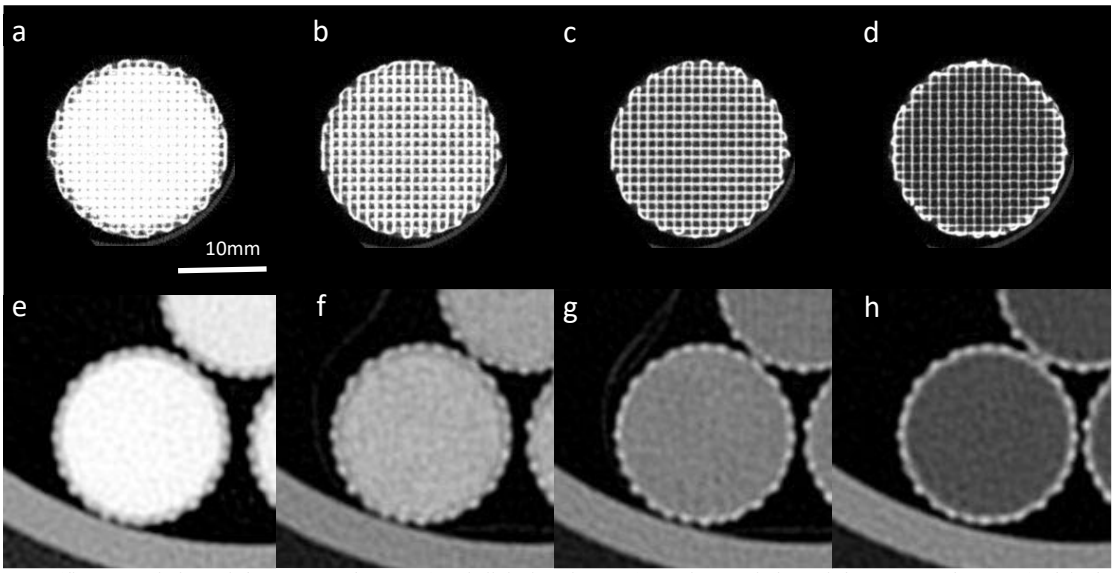


Fig. 1. Micro-CT (first row, window min/max = -1000/2500 HU) and clinical CT images (second row, window min/max = -1000/1000 HU) of the four sections of a micro-CT phantom, with infill ratios of 1.0, 0.7, 0.5 and 0.3 (from left to right).

B. PixelPrint

For each phantom in this study, G-code was automatically generated by PixelPrint through translating input images, i.e., DICOMs, into printer instructions. The generated G-code files contain instructions for the 3D printer to produce multiple 2D layers. Each printed layer consisted of an array of spaced parallel filament lines at a fixed spacing but of varying line widths, causing a partial volume effect that corresponds to the desired densities and resulting HU values in the final CT slice. Here, the infill ratio is defined as the amount of filament occupying a given unit volume, entailing printed line widths that are wider for high-density areas and narrower for low-density areas. Figure 1 illustrates the internal structure of 2D layers imaged with a micro-CT and the resulting partial volume effects in clinical CT scans. Further details regarding the PixelPrint technique can be found in our previous publication [5].

TABLE I
SCAN PROTOCOL OF THE PATIENT DATA AND PHANTOM SCAN

	Patient scan	Phantom scan
Scanner model	Siemens SOMATOM Definition Edge	Philips IQon Spectral CT
Tube voltage	120 kVp	120 kVp
Tube current	105 mA	105 mA
Exposure time	1.0 s	1.248 s
Spiral pitch factor	0.8	1.0
Exposure	131 mAs	131 mAs
$CTDI_{vol}$	8.8 mGy	9.9 mGy
Collimation width	0.6 / 38.4 mm	0.625 / 40.0 mm
Slice thickness	0.60 mm	0.67 mm
Convolution kernel	I26s\3	C
Field of view	99.75 x 99.75 mm^2	224 x 224 mm^2
Matrix size	228 x 228 $pixel^2$	512 x 512 $pixel^2$
Pixel spacing	0.4375 mm	0.4375 mm

Collimation width values are noted as single / total collimation width.

C. Phantoms

1) Micro-CT phantom

Three cylindrical phantoms were generated to evaluate the reproducibility of PixelPrint, utilizing both micro-CT and clinical CT imaging. Each phantom was printed as a small rod, with a diameter of 20 mm and a length of 60 mm, that consisted of four sections with different infill ratios (100%, 70%, 50%, and 30%). StoneFil filament lines were printed at a spacing of 1 mm, and the corresponding line widths were 1, 0.7, 0.5, and 0.3 mm, respectively. A thin outer layer was added to each phantom to support the structure (crucial for low density sections). The three phantoms were printed using the same G-code input and 3D printer.

2) Calibration phantom

A calibration phantom was designed to further evaluate the performance of PixelPrint when utilizing the StoneFil filament. The phantom is designed as a cylinder, with a diameter of 10 cm and a height of 1 cm, that includes seven equally divided pie-shaped sections. Each section was printed with a fixed line spacing of 0.5 mm but at variable filament line widths (0.2-0.5 mm), corresponding to different infill ratio (40-100%, with 10% intervals).

3) Cervical phantom

The Institutional Review Board (IRB) approved this retrospective study. A cervical phantom was created based on a patient-specific image volume (10 x 10 x 10 cm^3) that was acquired on a clinical CT scanner (Siemens SOMATOM Definition Edge, Siemens Healthcare GmbH, Erlangen, Germany) at 120 kVp with a standard diagnostic dose ($CTDI_{vol}$: 8.8 mGy). See Table 1 for detailed acquisition parameters for the patient and phantom scans.

The patient data consist of four cervical vertebrae (C4 to C7), including a clear view of trachea and esophagus. A circular region of interest with a diameter of 10 cm was cropped in the axial slices, forming a cylindrical phantom to fit in the bore of a QRM chest phantom (QRM GmbH, Möhrendorf, Germany).

TABLE II
SCAN PROTOCOLS FOR THE MICRO CT PHANTOM

	Micro CT scan	Clinical CT scan
Scanner model	MILabs U-CT	Philips IQon Spectral CT
Tube voltage	50 kVp	120 kVp
Tube current	0.21 mA	130 mA
Exposure time	54 s	1.923 s
Spiral pitch factor	Axial scan	0.39
Exposure	11.3 mAs	250 mAs
CTDIvol	69 mGy	16.4 mGy
Collimation width	-	0.625 / 40.0 mm
Slice thickness	0.08 mm	0.67 mm
Convolution kernel	-	YC
Field of view	22.16 x 22.16 mm^2	100 x 100 mm^2
Matrix size	277 x 277 $pixel^2$	512 x 512 $pixel^2$
Pixel spacing	0.080 mm	0.195 mm

Collimation width values are noted as single / total collimation width.

HUs were converted to infill ratios, based on the results from the calibration phantom, where the maximum HU the StoneFil filament can reach was 1000 HU. Further, the maximum infill ratio was 100% (0.5 mm line width), and the minimum was 40% (0.2 mm line width).

D. CT scan and image analysis

The three micro-CT phantoms were scanned on a commercial micro-CT (U-CT system, MILabs, CD Houten, the Netherlands) with a tube voltage of 50 kVp in three consecutive scans. In addition, the phantom was also scanned on a clinical CT system (IQon spectral CT, Philips Healthcare, the Netherlands) at 120 kVp with a high-resolution protocol and a small field-of-view. Additional acquisition parameters of the two scans are listed in Table 2.

Micro-CT images were exported from the scanner and reprocessed with a multi-planar reconstruction algorithm (MPR) in Horos (Horos Project, Annapolis, MD, USA). Mean and standard deviation HU values of the four sections in the sagittal slices were measured, and their linearity was assessed.

The calibration and cervical phantom were scanned inside the QRM chest phantom with a clinical CT system (IQon spectral CT, Philips Healthcare, the Netherlands). The protocol parameters approximately matched those of the original clinical examination of the patient (see Table 1).

For the calibration phantom, mean and standard deviation HU values of seven areas were measured. Square regions of interest (ROI) of 15 x 15 mm^2 were manually placed in each density region. For the cervical phantoms, images are exported and registered to the original patient images using the OpenCV Library (https://opencv.org). The dual-energy CT acquisition was utilized to measure HU at various virtual monoenergetic levels to quantity the spectral performance of our phantom.

III. RESULT

Underlying grid-like structure generated by PixelPrint are visible in micro-CT images however appear as constant regions in high-resolution clinical images due to the partial volume effect (Figure 1). In the micro-CT images, printed lines were

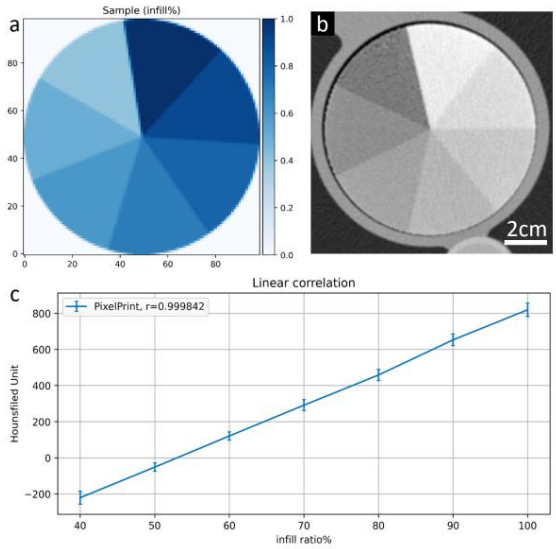


Fig. 2. Calibration phantom for Stonefil filament (a) Ground truth design of the phantom: a cylinder that is equally divided into areas with infill ratios of 40-100%, with 10% intervals. (b) Phantom images from a clinical CT scanner. Window max/min is -1000/1000 HU. (c) Mean HU values of the seven areas versus the corresponding infill ratios, with standard deviations indicated as error bars.

observed having equal spacings (1 mm) and a constant line width within each region. A layered structure with introduced offsets is distinctly visible in orthogonal views. When comparing the three identically manufactured phantoms, one can appreciate the high reproducibility that Pixelprint offers. In both micro- and clinical CT scans, an accurate linear relationship between infill ratios and mean HU was measured (r = 0.984 and 0.982).

For the calibration phantom (Figure 2), each of the seven regions had homogeneous intensities with excellent linearity. The highest infill ratio (100%) region has a mean of 819 HU and a standard deviation of 37 HU, while the lowest infill ratio (40%) has a mean of -220 HU and a standard deviation of 36 HU. A Pearson's correlation coefficient of 0.99 indicated a very high linear correlation between infill ratios and HUs.

Qualitatively, the CT image of the patient-based phantom closely resembles the original CT image both in texture and contrast levels, including various small bone and soft tissue structures. Mean virtual monoenergetic HU values of vertebra voxels from spectral CT reconstructions of the phantom scan show high corresponds to those of a 300 mg/ml calcium insert, with an RMSE of 138 HU after accounting for density differences. (Figure 3g). The intensity of bone marrow in the phantom (region 1 in Figure 3b) is about 10 HU different from the corresponding value in the original patient image. Limited by the density of the StoneFil filament, cortical bone voxels (regions 2-3 in Figure 3b) are about 150 HU less than expected.

IV. DISCUSSION

Continuing our previous work on producing patient-based lung phantoms, we demonstrated that PixelPrint combined with higher density filament is capable of creating soft tissue and

bone phantoms. We demonstrate PixelPrint's high level of reproducibility and robustness by examining the underlying grid-like structure with micro-CT scans, which cannot be resolved with high-resolution clinical CT acquisitions. In addition, we show a high correspondence in attenuation values for boney structures across the entire x-ray energy range.

Our study has limitations; the maximum Hounsfield unit created from StoneFil was approximately 819 HU in the patient-based cervical spine phantom. Further investigations are necessary to extend the dynamic range of PixelPrint, potentially by working with a dual-filament 3D printer.

V. Conclusion

The present study demonstrates the feasibility of our 3D-printed patient-based phantoms to be extended to soft-tissue and bone structure while maintaining accurate organ geometry, image texture, and spectral attenuation profiles.

Acknowledgment

The authors acknowledge support through the National Institutes of Health (R01EB030494).

References

[1]Mitsouras D, Liacouras PC, Wake N, Rybicki FJ. RadioGraphics Update: Medical 3D Printing for the Radiologist. RadioGraphics. 2020;40(4):E21-E23. doi:10.1148/rg.2020190217

[2]Okkalidis, Nikiforos. "A novel 3D printing method for accurate anatomy replication in patient-specific phantoms." Medical physics 45.10 (2018): 4600-4606.

[3]Okkalidis, Nikiforos, and George Marinakis. "Accurate replication of soft and bone tissues with 3D printing." Medical physics 47.5 (2020): 2206-2211.

[4]Mei K, Geagan M, Leonid Roshkovan, Harold I. Litt, Grace J. Gang, Nadav Shapira, J. Webster Stayman, and Peter B. Noël. "Three-dimensional printing of patient-specific lung phantoms for CT imaging: emulating lung tissue with accurate attenuation profiles and textures." Medical physics (2021).

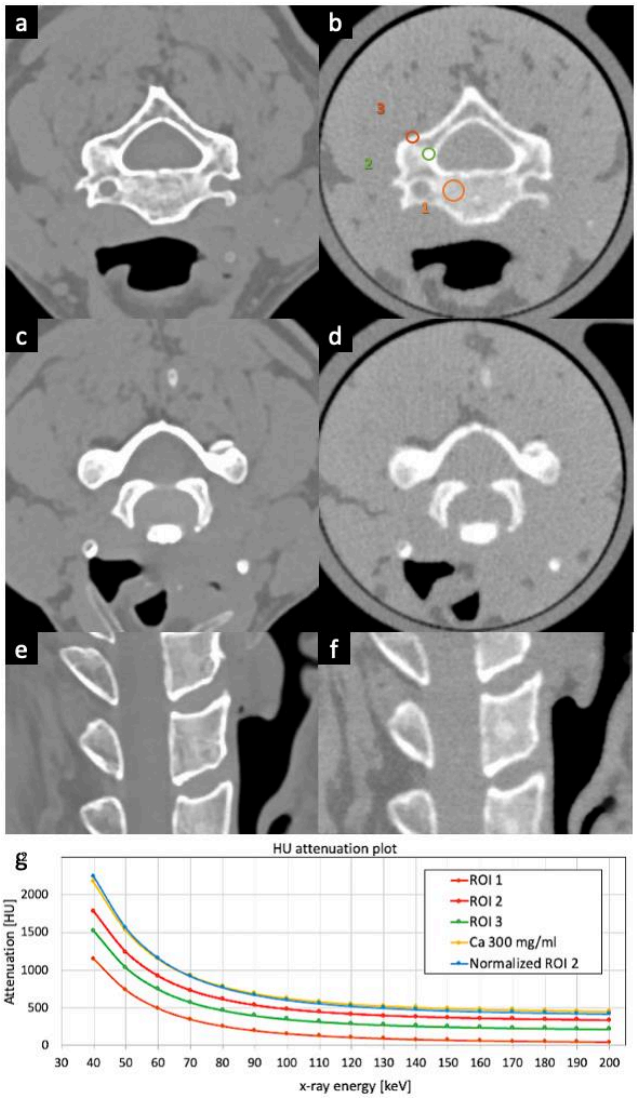


Fig. 3. Patient image and the PixelPrint cervical phantom. (a) (c) and (e) Original CT images utilized by PixelPrint to create the cervical phantom. (b) (d) and (f) CT images of the cervical phantom. All images have window level of 0 HU and width of 1200 HU. (g) Virtual monoenergetic HU measured with spectral CT at the denoted ROIs in (b) alongside reference values from a 300 mg/ml calcium insert (marked in yellow).

Practical Workflow for Arbitrary Non-circular Orbits for CT with Clinical Robotic C-arms

Yiqun Ma[1], Grace J. Gang[1], Tess Reynolds[2], Tina Ehtiati[3], Junyuan Li[1], Owen Dillon[2], Tom Russ[4], Wenying Wang[1], Clifford Weiss[1], Nicholas Theodore[1], Kelvin Hong[1], Ricky O'Brien[2], Jeffrey Siewerdsen[1], J. Webster Stayman[1]

***Abstract*—Non-circular orbits in cone-beam CT (CBCT) imaging are increasingly being studied for potential benefits in field-of-view, dose reduction, improved image quality, minimal interference in guided procedures, metal artifact reduction, and more. While modern imaging systems such as robotic C-arms are enabling more freedom in potential orbit designs, practical implementation on such clinical systems remains challenging due to obstacles in critical stages of the workflow, including orbit realization, geometric calibration, and reconstruction. In this work, we build upon previous successes in clinical implementation and address key challenges in the geometric calibration stage with a novel calibration method. The resulting workflow eliminates the need for prior patient scans or dedicated calibration phantoms, and can be conducted in clinically relevant processing times.**

I. Introduction

FOR decades, CT imaging has largely relied on standard circular and helical source-detector orbits for data acquisition. In recent years, new imaging systems (eg. robotic C-arms) have enabled the exploration of more advanced non-circular orbits for added benefits including increased field-of-view (FOV) size [1], improved image quality and/or dose reduction [2], weight-bearing extremity imaging [3], and metal artifact reduction [4].

Despite the increased interest, non-circular orbits that require more complex motion remain difficult to implement and research on clinical systems. There are several challenges — especially when many new orbits are desired including those that are customized to be patient- and/or task-specific. First, without access to sophisticated control systems, it is difficult to command the system to realize designed orbits. To date, we have relied on largely manual controls to achieve non-circular orbits on robotic C-arms [4]–[6]. Second, the manual element in the data acquisition leads to irreproducible scans. Additionally, the system geometry parameters recorded by the robot are not accurate enough to be used for 3D reconstructions due to system vibrations and gravity-induced strain on mechanical parts. Therefore, each scan requires online geometric calibration before reconstruction. Previously, we used a 3D-2D registration process for geometric calibration, in which we used a prior reconstruction as the registration target and then iteratively register each acquired projection to forward projections from the registration target [7]. However, this method requires a prior reconstruction and relatively long computation time, both of which may not be available in a clinical setting, and thus presents an obstacle to clinical research and translation.

In this work, we outline a workflow for implementing non-circular orbits on clinical systems, focusing on orbits previously demonstrated with inherent metal artifact reduction capabilities. This workflow includes a novel geometric calibration method based on fixed fiducials with arbitrary and *a priori* unknown placement, which overcomes the aforementioned challenges in clinical implementation.

II. Methods

We introduce a workflow for arbitrary-trajectory CT data acquisition in the following subsections. This includes details of acquisition on an experimental test bench and two different clinical robot C-arms, as well as a new online geometric calibration method and image reconstruction. Many clinical systems are limited in the degree of automatic orbit control. Some currently can only be driven manually while others allow increasing levels of automation. While fully manual techniques allow for research investigations, clinical translation will require additional manufacturer support.

A. Orbit Design and Implementation

Previously, we have investigated several types of non-circular orbits that can largely eliminate metal artifacts including *sinusoidal* and *multiple-arc* trajectories [6], [8]. In this work we consider simplified versions of two orbit types: *sawtooth orbits* and *double-circle-plus-arc (DCArc) orbits*. (See Figure 1). Both orbits were designed with a fixed isocenter, with LAO/RAO gantry rotation angle and CRAN/CAUD gantry tilt angle being the parameters in non-circular actuation. In the sawtooth orbit, the source oscillated between $\pm 20^\circ$ in tilt at a constant speed for two full cycles while rotating 360°. The DCArc orbit consisted of two tilted circular scans at $\pm 25^\circ$ plus an arc, where the gantry did not rotate while tilting from $+29^\circ$ to -28°. On clinical C-arms, the orbits were realized by manually driving using the bedside joystick in fluoroscopy mode (*Experiment B*), manually advancing through pre-programmed navigation points [6], or using a dedicated control system supplied by the manufacturer (*Experiment C*).

B. Geometric Calibration

As mentioned above, calibration of trajectories can require scan-specific estimation of the system geometry. Towards this end, we placed steel ball bearings (BBs) on the surface of the object as fiducials. The positions were unknown *a priori* but were presumed to remain rigidly aligned with respect to each other during the scan. The locations of BBs in projection images were extracted and used as inputs to a geometric estimation routine. Details of this procedure follow.

This work was supported, in part, by NIH grant R01EB027127 and, in part, by the Cancer Institute of New South Wales Fellowship 2021/ECF1293

[1]Johns Hopkins University, Baltimore, MD, 21205
[2]University of Sydney, Australia
[3]Siemens Healthineers
[4]Heidelberg University, Mannheim, Germany

1) BB Extraction: The BB locations were identified in each projection by performing a 2D correlation between the line integral images and a disc-shaped kernel of roughly the same size as the BBs. The highest correlations represent the approximate location of the centroids of the BBs. This location is further refined by computing the centroid of pixel values in a square region-of-interest (ROI) about the initial location estimate. Predicted BB locations based on linear interpolation across adjacent frames were used to identify individual BBs and maintain continuity. For scans where the BBs had very low contrast against background features such as metal and thick bones, the extraction was initialized by manually selecting the approximate centroid locations on the first image.

2) Geometry Optimization: Due to the BB's spherical symmetry, accurately backprojected rays of the *centroid* of a BB should intersect at an infinitesimal point in space. With an inaccurate geometry, these rays may not intersect. Thus, we can potentially estimate the true geometry by minimizing the mean Euclidean distance from the BB centers to corresponding backprojected rays, or, the *reprojection error* (RPE):

$$\mathbf{RPE}(\mathbf{p}, \Omega_n) = \frac{1}{K} \sum_{k=1}^{K} \mathbf{d}\left[\mathbf{p}_k, \mathbf{L}_k(\Omega_n)\right] \tag{1}$$

where n is the projection index; k is the BB index; Ω_n are parameters of the view-dependent system geometry; $\mathbf{p}$ contains all 3D BB center locations; $\mathbf{L}_k(\Omega_n)$ is the line equation of the ray backprojected from the k-th BB centroid on the n-th projection view using geometry Ω_n; $\mathbf{d}\left[\mathbf{p}, \mathbf{L}\right]$ is the Euclidean distance from point $\mathbf{p}$ to line $\mathbf{L}$.

We presume that the BB center locations are unknown (e.g. no prior scan or calibration) and also need to be estimated. We start with an approximate geometry (e.g. based on rough encoder positions or commanded location) as an initial guess. In this case, backprojected lines will likely not intersect at a point but instead a larger region that is generally close to the true BB location in 3D. For each pair of backprojected rays, we calculate the *nearest point* between these rays - the midpoint of the line segment orthogonal to and connecting both rays. We form a point cloud by computing the nearest points between all ray pairs, and the mean location of the point cloud is used as the estimated BB center: $\hat{\mathbf{p}}_k = \overline{\mathbf{P}}_N\left[\mathbf{L}_k(\Omega)\right]$ where $\overline{\mathbf{P}}_N\left[\mathbf{L}(\Omega)\right]$ is the mean of all nearest points formed from N rays backprojected using geometry Ω. We iteratively update $\hat{\mathbf{p}}$ and Ω by minimizing the RPE to jointly estimate BB point cloud centers and the system geometry:

$$\hat{\Omega}_{n,i} = \underset{\Omega}{\operatorname{argmin}}\, \mathbf{RPE}_n(\hat{\mathbf{p}}_{i-1}, \Omega) \tag{2}$$

where i is the iteration number. Note that the solutions of this objective will have fixed points where the point clouds have shrunk to infinitesimal size. It is possible that such a scenario is not the true geometry; however, such a geometry should provide accurate representations at those points. In this work, the optimization is performed using the MATLAB (MathWorks Inc.) function *fmincon*.

C. Reconstruction and Metal Artifact Reduction (MAR)

For image formation for the non-circular scan trajectories, we used a model-based iterative reconstruction algorithm.

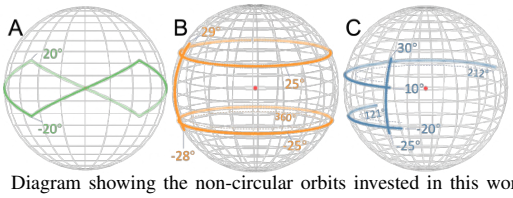


Fig. 1. Diagram showing the non-circular orbits invested in this work. Red dot at center of spheres is the isocenter. Colored dots on the sphere are source positions in the orbit. Important CRAN/CAUD tilt angles and arc lengths are marked on each plot. (A) sawtooth orbit. (B) double-circle-plus-arc (DCArc) orbit. (C) multi-arc orbit.

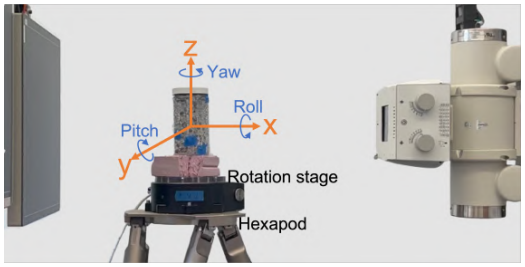


Fig. 2. Photo of bench setup and diagram of the 6 DoF motion space of the hexapod stage.

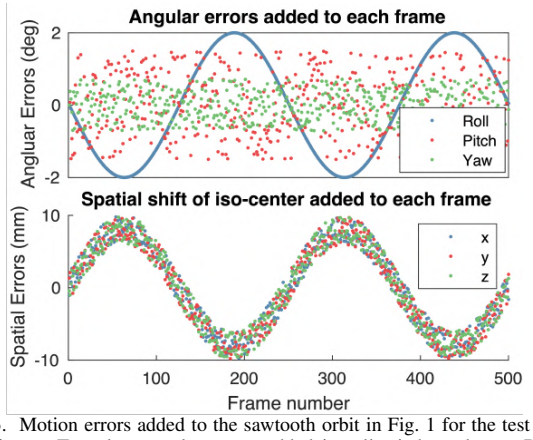


Fig. 3. Motion errors added to the sawtooth orbit in Fig. 1 for the test bench experiment. Top plot: angular errors added in roll, pitch, and yaw. Bottom plot: spatial shift of iso-center added in x, y, and z axes.

Specifically, we used a modified Penalized Weighted Least-Squares (PWLS) objective with a quadratic penalty. For all experiments, 50 iterations of a separable paraboloidal surrogates algorithm [9] were applied. A simple MAR algorithm [10] was implemented to eliminate streaks, etc. associated with the BBs or implants. In short, the metal regions in the projection images were segmented and those regions were filled using interpolated data. For implants, the metal volumes were added back into the metal-free reconstruction.

D. Experiment A: X-ray Test Bench

To investigate the online registration approach under controlled conditions, we performed a phantom study on a dedicated x-ray testbench.

1) Phantom Design: The test phantom (Figure 4A) consisted of a 3D-printed cervical spine placed in a plastic cylinder filled with plastic spheres of variable sizes for background clutter. Eight steel BBs of 3.17 mm diameter were affixed to the side of the container. BB placement followed a roughly spiral fashion to reduce the chance of overlapping BBs in projections; however, precise locations were unknown *a priori*.

2) Bench Experiment Implementation: The test bench includes an X-ray tube (Varex Rad-94), a flat-panel detector (Varex PaxScan 4343CB), a 6 degree-of-freedom (DoF) hexapod (PI H-900K Series) and rotation stage (PI PRS-200) (Figure 2), which enables precise emulation of arbitrary source-detector trajectories. In previous work [6], inaccuracies in positioning of a Siemens Artis Zeego were identified. We found systematic inaccuracies in the system encoded positions in LAO/RAO angle, CRAN/CAUD angle, isocenter, and a slight in-plane panel rotation around its normal axis, but also found that the C-arm was sufficiently rigid that we could assume a fixed source-detector distance and piercing point [6], [7]. In our bench studies, we emulated a C-arm system whose source remained fixed relative to the detector but had similar isocenter shifts and angular inaccuracies. These inaccuracies were added to an ideal sawtooth orbit (Figure 1A) and were realized by the hexapod using 3 DoF linear motions and 3 DoF rotations in roll, pitch, and yaw. Figure 3 illustrates the error in each DoF added to each frame. The error amplitudes in yaw (rotation) and pitch (tilt) were equal to the angular step size in each axis. A sinusoidal error pattern that was previously observed during data acquisitions (likely caused by gravity-induced sagging) was also added. The spatial shifts were generated with a sine wave of amplitude 8 mm plus random noise of ± 2 mm for an exaggerated level of uncertainty as compared with the clinical system. A normal circular scan was also acquired and each scan had 500 frames.

E. Experiment B: Siemens Artis Zeego Robotic C-arm

1) Phantom Design: To test and quantify the performance of the new calibration method on the Zeego C-arm, we used a cylindrical calibration phantom as proposed in [11], which contained 16 steel BBs, with the addition of a thin tungsten wire of 0.13 mm diameter suspended in the middle to probe in-plane image fidelity (Figure 4B).

2) Siemens Artis Zeego implementation: The DCArc orbit (Figure 1B) was performed manually with the bedside joystick controller, and projections were acquired in fluoroscopy mode. The orbit was acquired in three parts: two full tilted circles in LAO/RAO and one arc in CRAN/CAUD. For comparison, a standard circular scan was also acquired. The acquisition process required no modification to the system. The raw images were extracted with a dedicated software tool.

F. Experiment C: Siemens Artis Pheno Robotic C-arm

1) Phantom Design: To further test the proposed workflow, we imaged a torso phantom with Sawbones spine (Figure 4C). Four pedicle screws (Evolution Surgical, Sydney, Australia) were placed into three vertebrae (L2–4). Ten BBs of diameter 2 mm were taped to the surface of the torso phantom.

2) Siemens Artis Pheno implementation: We combined non-circular orbits with an upright weight-bearing setup — positioning that could provide more diagnostic information for spine diseases [1], [3], [12]. Since the phantom was sat upright on the bed, the C-arm's CRAN/CAUD tilt axis now functioned as the primary LAO/RAO gantry rotation. Due to motion range limitation, in order to achieve the 210° arc in the *multi-arc orbit* (Figure 1C), we manually rotated the phantom

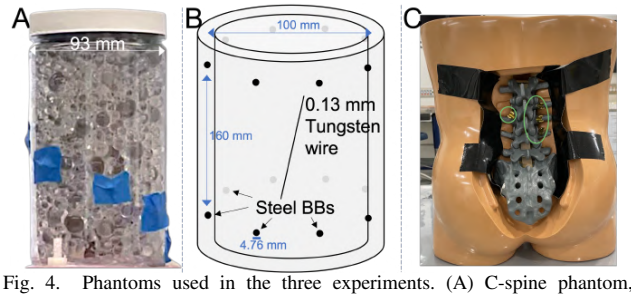


Fig. 4. Phantoms used in the three experiments. (A) C-spine phantom, showing the clear container, the plastic spheres inside, and steel BBs taped on the outside; the 3D-printed cervical spine is not visible due to the spheres. (B) Diagram of the cylindrical phantom adding a tungsten wire to a Cho calibration phantom with BBs. (C) Torso phantom with Sawbones spine, pedicle screws (green ovals), and steel BBs affixed with black tape.

approximately 90° between two identical 120° arc scans at a 10° tilt. Another 120° arc was acquired at a $-20°$ tilt, and a tilting arc between $-25°$ and 30° was acquired.

III. Results

A summary of results for all experiments is shown in Figures 5–7. Pre- and post-online calibration reconstructions are shown for each case. In all cases, the online calibration improves image quality. In experiment A/test bench (Figure 5) we see that the online BB calibration has similar image quality to both a well-calibrated circular scan as well as a calibration based on 2D-3D registration using a prior circular scan. In experiment B/Zeego (Figure 6), the central tungsten wire in the phantom is used to compute a FWHM estimate of the point-spread-function (PSF) which is comparable in both a circular and online BB-calibrated scan, whereas the pre-calibrated scan is too diffuse to obtain a FWHM estimate. Experiment C/Pheno (Figure 7) shows significantly improved visualization of anatomy with the online calibration. Moreover, previous results showing the ability to reduce metal artifacts with non-circular scans are evident.

The right side of each Figure 5-7 shows RPE distribution for the initial (pre-calibration) guess, and after 1, 2, and 3 iterations of the joint location/geometry estimation process. Each violin plot contains N RPE values from the N projections of the scan. The white circle at the middle of each violin represents the median RPE, the horizontal line denotes mean RPE, and the top right insert shows the RPE for the initial guess, which does not fit within y-axes of the main plots.

All three RPE distribution plots show that the joint estimation significantly improve accuracy after 1 iteration and appears largely converged after 2 iterations, with marginal improvements with a third iteration. Calibration performance results after 2 iterations are summarized in Table I. Figure 7 shows the biggest RPE improvement from the initial guess and also the highest computation time per frame, which is potentially caused by the worst accuracy of the initial guess. We observe an increasing total computation time with more frames, although other factors such as BB number and accuracy of initial guess may also be correlated. For reference, the 3D-2D registration method routinely took more than 15 seconds *per frame*. Note computation times were for a mid-range laptop using prototype code, whereas 3D-2D registrations were run on a workstation with a high-end GPU.

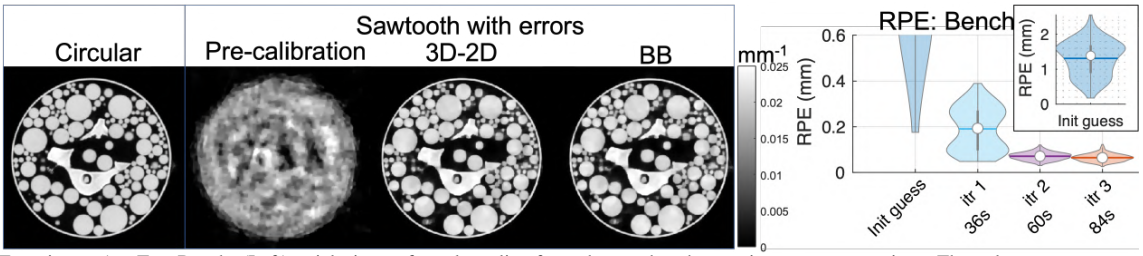


Fig. 5. Experiment A - Test Bench: (Left) axial views of a select slice from the test bench experiment reconstructions. The columns compare results pre- and post-calibration and between the 3D-2D registration and BB calibration methods. (Right) Summary of RPE as a function of iteration.

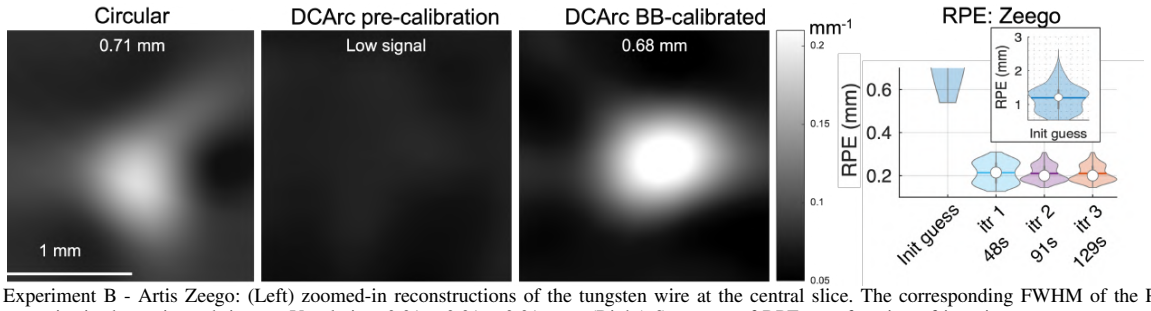


Fig. 6. Experiment B - Artis Zeego: (Left) zoomed-in reconstructions of the tungsten wire at the central slice. The corresponding FWHM of the PSF from the tungsten wire is shown in each image. Voxel size: 0.01 x 0.01 x 0.01 mm. (Right) Summary of RPE as a function of iteration.

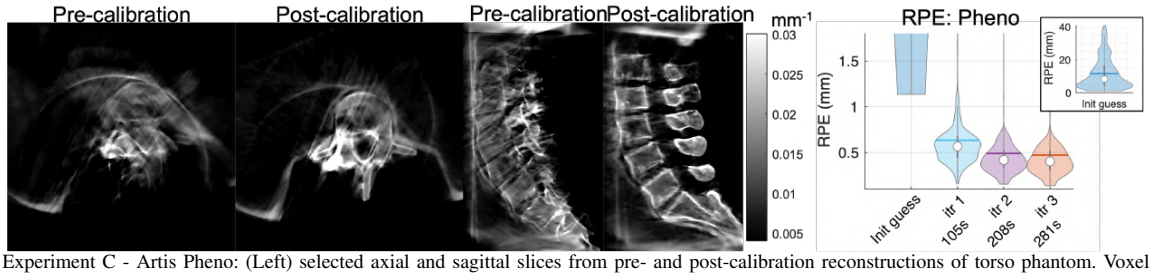


Fig. 7. Experiment C - Artis Pheno: (Left) selected axial and sagittal slices from pre- and post-calibration reconstructions of torso phantom. Voxel size: 0.5 x 0.5 x 0.5 mm. (Right) Summary of RPE as a function of iteration.

TABLE I
GEOMETRY ESTIMATION PERFORMANCE AFTER 2 ITERATIONS

	Exp. A Test Bench	Exp. B Siemens Zeego	Exp. C Siemens Pheno
Number of frames	500	656	857
Number of BBs	8	16	10
Elapsed time	60 s	91 s	208 s
Mean RPE	0.065 mm	0.210 mm	0.490 mm
Median RPE	0.065 mm	0.201 mm	0.418 mm
Std Dev RPE	0.0169 mm	0.0377 mm	0.370 mm
Time/frame	0.12 s	0.14 s	0.24 s
Time/frame/BB	0.015 s	0.009 s	0.024 s

IV. DISCUSSION AND CONCLUSION

This work establishes a practical workflow for non-circular CBCT scans on clinical robotic C-arms by overcoming several challenges in geometric calibration. The proposed geometric calibration method requires no prior scans, is fast to compute, and maintains comparable image quality to methods based on 2D-3D registration with prior images. This approach allows for scans with only approximately known geometries due to hardware limits in control and position, and for patient- and task-specific scans that vary between procedures and that cannot be individually pre-calibrated.

While we observe that previously investigated advantages like artifact reduction for metal implants can be realized with this approach, more detailed investigations are ongoing. Moreover, we expect that refinements in the BB extraction process and subsequent optimization can be improved. In particular, approaches to handle BBs obscured by metal implants and BBs coming in and out of the FOV, will further deliver a practical automatic workflow. Moreover, future work includes the development of strategies for BB placement based on task and anatomical site. Despite these current limitations, the proposed workflow is an important step in delivering fast online calibration without the need for prior images that will facilitate clinical translation.

REFERENCES

[1] Noo, F *et al.*, in *Medical Imaging 2018: Physics of Medical Imaging*, Chen, G.-H *et al.*, Eds. SPIE, Mar. 2018.
[2] Stayman, J. W *et al.*, *J. Med. Imaging (Bellingham)*, vol. 6, no. 2, p. 025002, Apr. 2019.
[3] Choi, J.-H *et al.*, *Med. Phys.*, vol. 41, no. 6, p. 061902, Jun. 2014.
[4] Gang, G. J *et al.*, *Conf. Proc. Int. Conf. Image Form. XRay Comput. Tomogr.*, vol. 2020, pp. 400–403, Aug. 2020.
[5] Ma, Y *et al.*, AAPM Virtual Meeting, Jul. 2021.
[6] ——, SPIE Medical Imaging, Feb. 2022.
[7] Ouadah, S *et al.*, *Phys. Med. Biol.*, vol. 61, no. 7, pp. 2613–2632, Apr. 2016.
[8] Gang, G. J *et al.*, *Proc. SPIE*, vol. 11312, Feb. 2020.
[9] Erdogan, H and Fessler, J. A, *Phys. Med. Biol.*, vol. 44, no. 11, pp. 2835–2851, Nov. 1999.
[10] Kalender, W. A *et al.*, *Radiology*, vol. 164, no. 2, pp. 576–577, Aug. 1987.
[11] Cho, Y *et al.*, *American Association of Physicists in Medicine (AAPM)*, Mar 2005.
[12] Zhao, C *et al.*, *Med. Phys.*, vol. 47, no. 8, pp. 3305–3320, Aug. 2020.

Rigid Motion Correction Based on Locally Linear Embedding for Helical CT Scans with Photon-counting Detectors

Mengzhou Li, Chiara Lowe, Anthony Butler, Phil Butler, and Ge Wang, *Fellow, IEEE*

***Abstract*—X-ray photon-counting detector (PCD) provides low noise, high resolution, high contrast and spectral responses which makes it popular in the CT imaging applications. Same as conventional CT scans with energy integrating detectors, patients' involuntary motions may blur the reconstruction and generate artifacts, and this issue is more prominent with PCD due to its high resolution and extended imaging time result from limited counting rate. In addition, PCDs often come with a significant amount of bad pixels and this makes the reconstruction hard to be handled satisfactorily with analytical methods, which rules out many state-of-the-art motion correction methods with fast analytical reconstruction cores. In this paper, we extend our previous locally linear embedding (LLE) based cone beam motion correction work to helical scanning geometry which is more commonly used due to the high cost of large-area PCDs. Specifically, besides the simple adaption of the LLE based parallel searching idea to helical scans with PCDs, we propose to use unreliable-volume masking to improve the motion estimation accuracy and to perform incremental type updating to majorly reduce the size of sampling grids for further acceleration. Our numerical experiment results demonstrate that using unreliable-volume masking significantly reduce the estimation errors around the two ends of the reconstruction volume, and a five times reduction of sampling grid size while maintaining a slightly better performance is observed by adopting the incremental updating strategy. The experimental results on physical photon-counting scans of patient wrist show significant resolution improvement and contrast enhancement after correction with our method which reveals subtle fine structures hidden by the motion artifacts.**

***Index Terms*—Rigid motion correction, Motion estimation, Locally linear embedding, Helical CT scan, Photon-counting detectors (PCDs).**

I. Introduction

MOTION induced artifact is a long standing problem in X-ray computed tomography which frequently happens in many clinical diagnostic scans. For instances, head CT and cardiac CT often suffer from two most representative types of motion artifacts, rigid motion artifacts and deformable motion artifacts, respectively. In this work we mainly focus on the rigid type motion compensation. The rigid motion of the object can be virtually taken as the transformation of the source and detector pair. Without proper correction for the motion, the reconstruction will get degraded by equivalent geometric error artifacts, typically like, blurring, double-edges, and even streaks. Compensations for rigid motion is fairly straightforward in an iterative reconstruction framework, which basically minimizes a loss function by alternatively updating image reconstruction and motion estimation till convergence [1]. However, this method is often slow due to the heavy computational cost from each iterative reconstruction. For acceleration, modified FDK [2] and weighted filtered backprojection (WFBP) with motion-accounting rebinning [3] have been recently developed to replace the iterative reconstruction in the loop. Along the other direction, a powerful parallel searching algorithm based on locally linear embedding (LLE) is proposed for cone beam CT [4] to replace the gradient-base optimization steps to majorly reduce the required number of iterations till convergence, resulting a significant speed boost with the help of GPU.

M. Li and G. Wang are with the Department of Biomedical Engineering, Rensselaer Polytechnic, Troy, NY, 12180 USA (wangg6@rpi.edu).

C. Lowe, A. Butler and P. Butler are with University of Canterbury and MARS Bioimaging Ltd, Christchurch, New Zealand.

Recent development of X-ray photon-counting detector begins to revolutionize the CT imaging field with the flag-event of the Siemens photon-counting CT getting approved by FDA. But currently the photon-counting detection technique is still not mature [5] which also brings additional challenges to the motion correction (MC) problem. First of all, due to complex manufacturing process, many PCDs in the market come with a significant amount of ineffective pixels as a result of manufacture cost controlling. In addition, large PCD array is often tiled from many small PCD chips. The tiling often results in gaps due to the insufficient room for electronics. Hence, the gaps together with the ineffective pixels form the bad pixels in the projection which are sometimes so big that can not be easily addressed with interpolations. As a result, the traditional analytical reconstruction is hard to be applied here. Second, due to the high cost of a PCD compared to a same-area flat panel detector, helical scan is often performed for axial FOV extension. These two issues make the problem unique and we start from the state-of-the-art LLE based cone beam MC method [4], further tailor the method for helical scan, and modify the motion updating step from a global absolute value searching manner to an incremental type for improved performance.

II. Methods

A. LLE based motion correction

The rigid MC for a moving object is equivalent to the compensation for the geometrical misalignment of the source and detector pair with a fixed object. The goal is to minimize the difference between the measured projections and the re-projected projections with estimated motions incorporated. More explicitly, it formulates,

$$\arg\min_{\boldsymbol{p},\boldsymbol{x}} \|\boldsymbol{b} - A(\boldsymbol{p})\boldsymbol{x}\|_2^2, \tag{1}$$

where $A(\boldsymbol{p})$ is the system matrix after incorporating the motion parameters $\boldsymbol{p}$, $\boldsymbol{x}$ and $\boldsymbol{b}$ denote the reconstruction image and the projection measurements respectively. The optimization is often performed by alternatively updating $\boldsymbol{p}$ and $\boldsymbol{x}$. In addition, since the motion parameters for one view is separable from those for others, the problem can be dealt as a set of sub-problems for each view in the motion estimation updating step,

$$\boldsymbol{p} = [\boldsymbol{p}_1, \cdots, \boldsymbol{p}_i, \cdots, \boldsymbol{p}_N]^T,\ \boldsymbol{p}_i = \arg\min_{\boldsymbol{p}_i} \|\boldsymbol{b}_i - A_i(\boldsymbol{p}_i)\boldsymbol{x}\|_2^2, \tag{2}$$

where $i = 1, 2, \cdots, N$ denotes the selection for the ith view out of N in total, and $\boldsymbol{p}_i$ is the motion parameters for the ith view in the form of $[t_x, t_y, t_z, \theta_x, \theta_y, \theta_z]^T$ characterized by six freedom parameters describing the translation and rotation along and around three axes.

Instead of using gradient based optimization technique, LLE utilizes a powerful parallel searching strategy by densely sampling a predefined parametric range. The basic idea is that if the sampling grid is sufficiently small, the true parameter vector should be so close to its K-nearest neighbors in the sampling grid and can be expressed as the a linear combination of them such that its corresponding projection measurement can also be represented with the linear

combination of the K-nearest reprojected projections associated with the neighboring parameter samples and with the same weights, i.e.,

$$\boldsymbol{p}_i^* = \sum_{k=1}^{K} w_k \boldsymbol{p}_i^{(k)}, \quad \boldsymbol{b}_i = \sum_{k=1}^{K} w_k A_i(\boldsymbol{p}_i^{(k)})\boldsymbol{x}, \tag{3}$$

where $\sum_{k=1}^{K} w_k = 1$, and $\boldsymbol{p}_i^{(k)}$ is one of the K nearest samples for $\boldsymbol{p}_i^*$ in the grid $(\boldsymbol{p}_i^{(1)}, \cdots, \boldsymbol{p}_i^{(j)}, \cdots, \boldsymbol{p}_i^{(S)})$ with S samples in total, and $\boldsymbol{p}_i^*$ denotes the ground truth motion parameter. This can understood as following, based on Taylor series expansion of A_i at $\boldsymbol{p}_i^*$,

$$A_i(\boldsymbol{p}_i) \approx A_i(\boldsymbol{p}_i^*) + (\boldsymbol{p}_i - \boldsymbol{p}_i^*)^T \frac{\partial A_i(\boldsymbol{p}_i^*)}{\partial \boldsymbol{p}_i}, \tag{4}$$

Hence,

$$\begin{aligned}\sum_{k=1}^{K} w_k A_i(\boldsymbol{p}_i^{(k)}) &= \sum_{k=1}^{K} w_k (A_i(\boldsymbol{p}_i^*) + (\boldsymbol{p}_i^{(k)} - \boldsymbol{p}_i^*)^T \frac{\partial A_i(\boldsymbol{p}_i^*)}{\partial \boldsymbol{p}_i}) \\ &= A_i(\boldsymbol{p}_i^*) + (\sum_{k=1}^{K} w_k (\boldsymbol{p}_i^{(k)} - \boldsymbol{p}_i^*)^T) \frac{\partial A_i(\boldsymbol{p}_i^*)}{\partial \boldsymbol{p}_i}\end{aligned}$$

If we have $\boldsymbol{p}_i^* = \sum_{k=1}^{K} w_k \boldsymbol{p}_i^{(k)}$, we can easily reach the conclusion,

$$\sum_{k=1}^{K} w_k A_i(\boldsymbol{p}_i^{(k)}) = A_i(\boldsymbol{p}_i^*), \ \boldsymbol{b}_i = \sum_{k=1}^{K} w_k A_i(\boldsymbol{p}_i^{(k)})\boldsymbol{x}. \tag{5}$$

As a close approximation, in practical algorithm we estimate the parameters from $\boldsymbol{p}_i$ one by one for each subproblem to save computational cost rather than sampling multidimensional grid, and all subproblems are updated in parallel for each of this iteration. Specifically, let the rth parameter $\boldsymbol{p}_{ir}$ in one subproblem to be updated. The key steps are

1) to generate a dense sample grid $\{\boldsymbol{p}_{ir}^{(s)} | s = 1, \cdots, S\}$ for $\boldsymbol{p}_{ir}$;
2) to calculate the reprojected projection grid $\{(\boldsymbol{b}_i^{(s)}, \boldsymbol{p}_{ir}^{(s)}) | s = 1, \cdots, S|\}$ by replacing the rth parameter of $\boldsymbol{p}_i$ with the sample grid above;
3) to find the K nearest neighbors $\{(\tilde{\boldsymbol{b}}_i^{(k)}, \tilde{\boldsymbol{p}}_{ir}^{(k)}) | k = 1, \cdots, K\}$ for $\boldsymbol{b}_i$ from the grid projections in terms of Euclidean distance;
4) to find the weights $\boldsymbol{w} = [w_1, \cdots, w_K]^T$ for the K neighbors to best fit the measurement by optimizing the following problem,

$$\boldsymbol{w} = \arg\min_{\boldsymbol{w}} \left\| \boldsymbol{b}_i - \sum_{k=1}^{K} w_k \tilde{\boldsymbol{b}}_i^{(k)} \right\| \text{ s.t. } \sum_{k=1}^{K} w_k = 1; \tag{6}$$

5) to update the estimation as $\boldsymbol{p}_{ir} = \sum_{k=1}^{K} w_k \tilde{\boldsymbol{p}}_{ir}^{(k)}$.

Note that the described updates for different views can be implemented in parallel with GPU, above these are the sequential iterations on r, and the whole process can be iterated a few times to reach convergence (we refer to [4] for more details).

B. Bad pixel masking

Most PCDs contain a nonnegligible amount of ineffective pixels, including initial real bad pixels from manufacturing, the tile gaps between chips, and developing dead/bad pixel due to aging/degradation over time. Those pixels come with unreliable responses and in a big number compared to the traditional EID detectors. One exemplary projection with a 14-chip PCD is shown in Fig. 1. Since the number is so big that simple interpolations cannot address the issue without introducing significant artifacts. Their unreliable responses will also influence the motion estimation accuracy.

Hence, our strategy to suppress the issue is to turn off those pixels with a mask and utilize iterative reconstruction methods to avoid their contribution to the reconstruction and MC. We adopted two simple criteria for unreliable pixels detection with the open beam projection

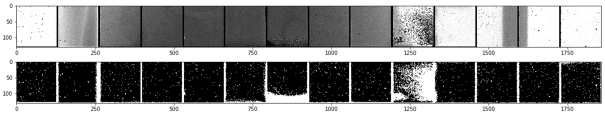


Fig. 1. An exemplary raw projection image (top) and corresponding bad pixel mask (bottom).

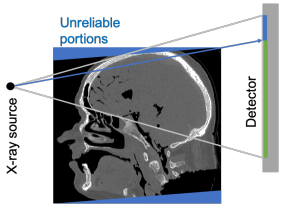


Fig. 2. Geometric illustration for the reliable volume masking. Any ray passing through the unreliable portions (marked blue on the volume) is taken as unreliable and resulting a mask indicating the areas in a projection (marked blue on the detector) which should be discarded during the data fidelity loss calculation for motion estimation.

data, (1) the temporal mean value of the pixel is a statistically outlier from the group of all pixels; (2) the temporal variance of the pixel is a statistically outlier from the group of all pixels. The mask for the unreliable pixels will be used for reconstruction and to exclude their contributions in the calculation of fidelity loss in Eq. 6.

C. Unreliable volume masking

Due to axial truncation, there are portions at two ends of a reconstruction volume to be unreliable due to data insufficiency, as illustrated in Fig. 2. Since these regions are only related to a few projections, the resulting reconstruction will be solely determined by these projections regardless of their geometrical misalignment, which will counter-balance the misalignment loss they contribute to the reliable portion, and finally degrade the reconstruction and influence the overall motion estimation for other projections as well. The original cone beam LLE MC method [4] does not consider this effect, however, we found this effects could make a significant difference if not dealt with properly.

To minimize the aforementioned effects, we utilize a mask to reject the contribution from the rays passing through the two unreliable portions to the data fidelity loss calculation in Eq. 6. The generation of the mask follows three key steps,

1) determine the unreliable portions and generate a volume mask with one indicating the unreliable portions;
2) forward project the volume mask to obtain the projection images;
3) take threshold on the projected images comparing to zero and generate the unreliable volume mask in the projection domain.

There are several different approaches to determine the unreliable portions. One could determine them according to the Tam-Danielsson window, or could investigate the reconstruction and manually select the slice range to balance the noise and image quality and reserve some margins. In this study, we follows the latter one.

D. Incremental updating

In our previous work [4], the algorithm directly estimates the motion vectors from the updated reconstruction in each main iteration, and a same large dense sampling grid is used for all iterations. However, the error of the motion estimation is expected to gradually decease as iteration goes, which is also what we observed in practice.

Global searching in a large grid may be a waste of computation in later iterations, hence, we propose a more efficient incremental type searching strategy, i.e., in each new iteration, we inherent both the intermediate motion estimation and the updated reconstruction from the last iteration, and perform further incremental refinement on the estimation rather than start a global absolute value search. Since we know the error is decreasing during iterations, we can gradually shrink the sampling space while maintaining the same number of samples to generate finer grid for improved searching accuracy. On the other hand, we can use a fixed small number of samples resulting an initial coarse sampling grid and gradually shrink the sampling range to obtain finer grid during iterations, hence, we are able to boost the searching efficiency without sacrificing significant performance.

E. Virtual bed removal

In most CT scans, the object is supported by a bed or holder which is considered static in reference to the scan geometry, for example, a couch in a medical CT or a sample holder in a micro-CT. In the reconstruction, those static beds preserve sharp edges and a relatively good image quality and are more robust to object motions. However, during the MC, those static items begin to move in reference to the object. In other words, if we perfectly compensate the motion for the object, the bed portion will get blurred and degraded. This will counterbalance the decrease of the loss function in Eq. 6 due to the improved motion estimation.

In order to avoid this effect, we virtually remove the bed from the projection data, and use this new projection data to perform motion correction. The steps to remove the bed are described as follows,

1) reconstruct the volume with the initial scan geometry;
2) segment the bed (i.e., all static portions) from the volume with some margin and generate a bed mask;
3) set voxels outside the bed mask to zero in the reconstruction;
4) forward project the obtained bed-only volume to get the bed-only projections;
5) subtract the bed-only projections from the original projection data to obtain the desired projections with bed removed for MC.

III. Results

A. Simulation experiment

The benefits of unreliable volume masking and incremental updating have been investigated through simulation. The additional head phantom from visible human project [6] is used as our sample, and the volume is first resampled to have isotropic voxels of size $0.5^3 mm^3$. The detector size is 64 rows by 640 columns with a pixel size of $1mm \times 1mm$, and the scanning pitch is 43.13 mm with 984 projections per rotation. To investigate the searching ability of the method, we do not add noise in our experiments. Two rotations are performed with motions added to six freedoms shown in Fig. 3, where Ti and Ri denote the translation and rotation along and around the ith axis, respectively. These amount of motions cause severe artifacts in the reconstruction as demonstrated in Fig. 3 (b).

During the MC, the image is reconstructed at a volume size of $240 \times 240 \times 100$ with a voxel size of $1^3 mm^3$ for motion estimation. Figure 4(a) shows the MC result using LLE based correction method. Those images are greatly improved from the original reconstruction. Despite a few artifacts, they appear close to the ground truth with perfect motion compensation shown in Fig. 3(b). Figure 4(b) plots the motion curves, estimated with the absolute type updating and the unreliable volume masking, against the ground truth along views. The estimated curves align well with the references especially with those of rotation motion demonstration the effectiveness of LLE based correction method.

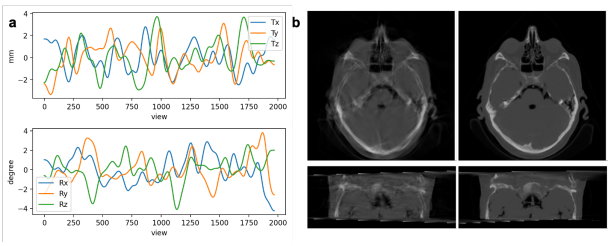


Fig. 3. The motion added along each freedom (a) and the axial and sagittal views of corresponding reconstructions without motion compensation and with perfect motion compensation (b), displayed in window [0, 0.8], unit of cm^{-1}.

The results of different strategy combinations of unreliable volume masking and incremental updating are compared in Figs.s 4(c) and 4(d), including: (absolute updating, masking, $N_s = 50$), (absolute updating, no masking, $N_s = 50$), (absolute updating, masking, $N_s = 10$), and (incremental updating, masking, $N_s = 10$), where N_s denotes the size of the sampling grid, and the masking is the short for the unreliable volume masking. Figure 4(c) shows the change of the root mean squared error (RMSE) of the forward projections of the reconstruction against the measurements during the iterative updates of the reconstruction. Note that even with the perfect motion estimation the RMSE does not reach zero due to the large-voxel reconstruction, and this RMSE is marked as a reference line in the figure. All four settings converge after 5 iterations and the incremental updating with masking ranks first even it only uses a five times smaller sampling grid compared to the second place, the setting with absolute updating, masking and $N_s = 50$. The full size sampling grid makes the absolute updating with masking perform better than a smaller sampling grid as anticipated, and the absolute updating without masking performs the worst although it uses a full size sampling grid. To compare the motion estimation accuracy among them, we calculate the virtual spatial positions of the source and detector and virtual normal vector of the detector assuming a static patient. The positional and angular errors against the ground truth are plotted in Fig. 4(d). The other three settings except the no masking one have closely overlapped curves, while the no masking setting demonstrates a a cupping effect as predicted in section II-C, i.e., significantly larger misalignment is observed at the two ends and extending towards the center compared to the other three. Those results demonstrate the effectiveness of reducing errors of motion estimation around two ends with the masking, and the efficiency boost with incremental type updating while maintaining good accuracy.

B. Physical data experiment

The scan of a patient wrist was performed on a MARS Spectral clinical scanner equipped with a PCD array tiled from 14 chips, and each chip with 128×128 pixels of $110 \times 110 \mu m^2$ size. The helical scan covers a FOV of 120mm in diameter and 100mm in length, with 373 projections per rotation and 5713 projections in total. The source is operated at 120 kVp, 310 μA with 0.25 mm Brass filtration. The source to detector distance and source to iso-center distance are 949 mm and 625 mm, respectively. We select one portion from the projection data consisting of 746 projections (2 rotations, projection volume around size $128 \times 1792 \times 746$) to test our MC method for demonstration. To minimize the noise influence, the channel with the lowest threshold (i.e., counts all photons with energy above 7keV) is used for MC. The reconstruction has an isotropic voxel size of $180^3 \mu m^3$, resulting a volume of size $746 \times 746 \times 123$. We set the number of samples along each freedom as 50, and the range for

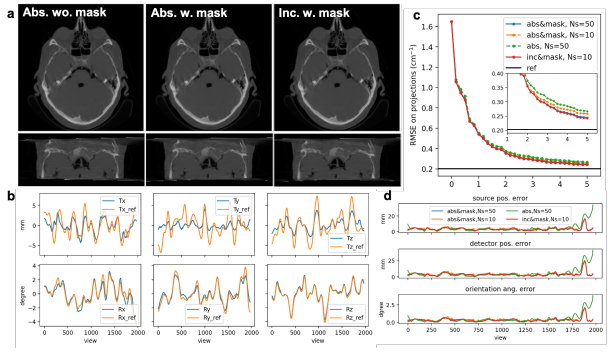


Fig. 4. Numerical simulation results of MC using four settings, including: (1) absolute updating, without the unreliable volume masking, $N_s = 50$ (abs. wo. mask); (2) absolute updating, with the masking, $N_s = 50$ (abs. w. mask); (3) absolute updating, with the masking, $N_s = 10$ (abs. w. mask); and (4) incremental updating, with the masking, $N_s = 10$ (inc. w. mask). (a) shows the axial and sagittal views of compensated reconstructions using settings 1, 2 and 4; (b) shows the estimated motion curve against the ground true with respect to the view, obtained with setting 2; (c) is the RMSE of the reprojections from the reconstruction against the measurements during the MC iterations; and (d) illustrates the positional error of the virtual source and detector positions and the angular error of the virtual detector orientation against the ground truth.

translations in x, y, z and the range for rotations around x, y, z are [-0.9 mm, 0.9 mm] and $[-2^\circ, 2^\circ]$ respectively. SART algorithm is used for reconstruction with 200 iterations. For the optimization along each freedom we iterate two times, and the complete sequential updating for all six freedoms (the main loop) is repeated for 5 times.

Figure 5 shows the results before and after MC. In Fig. 5 (b), image becomes sharper (see the structures in the red circle) and subtle structures are revealed from the blurry clouds after correction (see the region pointed by the read arrow). Similar results are observed in Fig. 5 (c), and the blurry double-edge phenomenon has been removed through correction as shown in the region pointed by the left arrow. In addition, fine structures are reveal inside the bony region pointed by the second arrow, showing improved resolution. Figure 5 (e) provides similar evidence for sharper edges and improved resolution. In Fig. 5 (d), weak structures are also enhanced as shown in the circles. Note that the missing chunk after reconstruction at the bottom right corner in Fig.s 5 (b) and (c) is the bed that has been virtually removed.

Figure 6 shows the change of RMSE of the reprojected projections of the reconstruction volume against the measured projections with bed removed during iterations. The corrections of six freedoms are performed sequentially as shown in the curve. The RMSE rapidly drops in the first two iterations and then converges gradually which demonstrate the effectiveness of our method. Usually three iterations are enough and no significant structure difference is observed between the result with three iterations and that with five iterations.

IV. Conclusion

In conclusion, we have presented a LLE-based motion correction method for helical photon-counting CT which uses iterative method as the reconstruction core and incorporates unreliable volume masking and incremental updating strategies. Accuracy improvement of motion estimation and speed boost due to the significantly reduced sampling grid size benefitting from the two strategies are observed in our numerical experiments. We also performed experiments on a real clinical human wrist data, and show that significant resolution and contrast enhancement are achieved, which reveals subtle fine structures hidden by artifacts, after correction with our method.

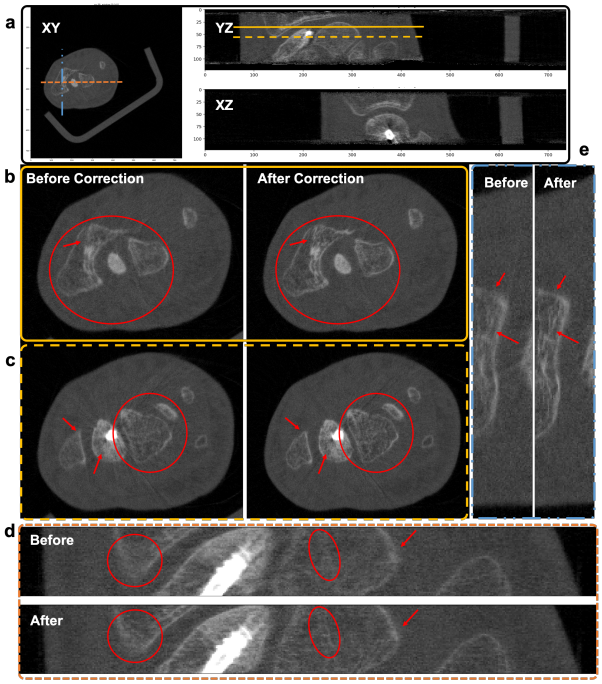


Fig. 5. Motion correction results for real patient wrist data. (a) The overview of the reconstruction volume before correction; (b) and (c) different axial slices before and after correction (the axial positions indicated in (a)); (d) and (e) are coronal view and sagittal view comparison (the cross-section postions marked in (a)). All images displayed in the window [0, 0.8], unit: cm^{-1}.

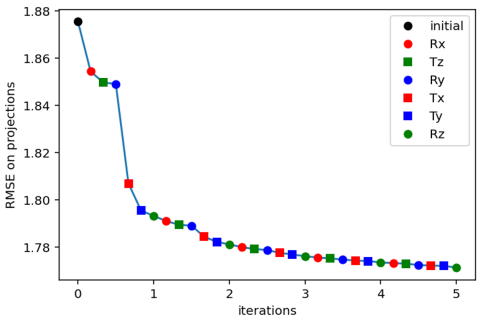


Fig. 6. The RMSE change on projection data during the correction iterations.

References

[1] T. Sun, J.-H. Kim, R. Fulton, and J. Nuyts, "An iterative projection-based motion estimation and compensation scheme for head x-ray ct," *Medical physics*, vol. 43, no. 10, pp. 5705–5716, 2016.

[2] J. Nuyts and R. Fulton, "Iterative fdk reconstruction for helical ct of the head with rigid motion compensation," in *The 6th Int. Conf. on Image Formation in X-Ray Computed Tomography*, 2020, pp. 248–251.

[3] S. Jang, S. Kim, M. Kim, K. Son, K.-Y. Lee, and J. B. Ra, "Head motion correction based on filtered backprojection in helical ct scanning," *IEEE transactions on medical imaging*, vol. 39, no. 5, pp. 1636–1645, 2019.

[4] M. Chen, P. He, P. Feng, B. Liu, Q. Yang, B. Wei, and G. Wang, "General rigid motion correction for computed tomography imaging based on locally linear embedding," *Optical Engineering*, vol. 57, no. 2, p. 023102, 2018.

[5] M. Li, D. S. Rundle, and G. Wang, "X-ray photon-counting data correction through deep learning," *arXiv preprint arXiv:2007.03119*, 2020.

[6] M. J. Ackerman, "The visible human project," *Proceedings of the IEEE*, vol. 86, no. 3, pp. 504–511, 1998.

Wednesday, June 15

Modeling and Assessment

Invited Talk on Deep Learning

Deep Learning Assessment

Wednesday Poster Session

Spectral and Polyenergetic CT Reconstruction

An Attempt of Directly Filtering the Sparse-View CT Images by BM3D

Larry Zeng[1,2]
[1]Department of Computer Science, Utah Valley University, Orem, UT 84058, USA
[2]Department of Radiology and Imaging Sciences, University of Utah, Salt Lake City, UT 84108 USA

e-mail: larry.zeng@uvu.edu

This work was supported in part by the National Institutes of Health (NIH) under Grant R15EB024283.

ABSTRACT The x-ray computed tomography (CT) images with sparse-view data acquisition contain severe angular aliasing artifacts. The common denoising filters do not work well. The state-of-the-art methods to process the sparse-view CT images are deep learning based; they require a large amount of training data pairs. This paper considers a situation where no training data sets are available. All we have is one sparse scan of a patient. This paper attempts to use a BM3D filter to reduce the artifacts by introducing an artifact power spectral density function, which is calculated with computer simulations. The results in this paper show that the proposed method is not effective enough for practice applications. However, some insights may lead us to further investigations.

INDEX TERMS Image processing, Image reconstruction, Biomedical imaging, Computed Tomography, Filters

I. INTRODUCTION

The motivation for using low-dose x-ray computed tomography (CT) is to reduce the patient radiation exposure [1-3]. Since x-ray radiation exposure may play a role in getting cancers, it is advised to reduce the x-ray exposure to an As-Low-As-Reasonably-Achievable (ALARA) level [4]. One way of low-dose imaging is the sparse view method, but sparse angular sampling frequently leads to characteristic streak artifacts. This under sampling situation is also a case of compressed sensing.

Many researchers attempted to solve this compressed sensing problem. One method is the iterative image reconstruction method that minimizes the total-variation (TV) norm or other measures of the image [5]-[11]. Most recently, research activities are mainly in the deep learning area [12]-[19]. It is fair to say that deep learning methods are dominating the current publications and conferences.

This paper investigates a nonlinear filter that is not deep learning based. Our filter is based on the BM3D denoising method, which was proposed by Dabov *et al*. [20][21]. The BM3D method uses block matching and aggregation strategies to obtain three-dimensional image blocks; its denoising uses Wiener filtering. The BM3D is currently the state-of-the-art in image denoising.

The BM3D method requires two inputs: the noisy image and the noise power spectral density image. The original purpose of BM3D is for random noise reduction. In our application of sparse-view tomography, our main concern is the angular aliasing streak artifacts. These artifact patterns are deterministic and object dependent. These artifacts are usually more pronounced than the random noise. The strategy of this paper is to treat the deterministic artifacts as random noise when calculating the 'noise' power spectral density function (image).

II. METHODS

A. 'Noise' power spectral density

For a given CT image, G, resulted from sparse-view projection measurements, its associated artifact power spectral density function, P, is difficult to obtain. This is because the true image, T, is not available.

In this paper, the artifact power spectral density function, P, is obtained by noiseless computer simulations, that simulate full-scan and sparse-scan projections of some random objects. A full scan has sufficient angular measurements. The reconstructed images from this full-scan data set are treated as (gold standard) true images, T_{simu}.

The artifact image, A, is the differences between the gold standard true image, T, and the given sparse-scan image, G_{simu}:

$$A = T_{simu} - G_{simu}. \quad (1)$$

In this paper, we use 1000 random simulated objects. Therefore, we have 1000 2D artifact images, A's.

Let B be the 2D Fourier transform of image A defined in (1). For each element in B, we calculate its norm square and

denote the resulting frequency-domain image be P. This resultant 2D image, P, has the same dimension as the image A, is real, and is nonnegative. Even in the noiseless cases, P is not zero due to the sparse-view streaking artifacts. In forming 1000 versions of P, no noise is added. Therefore, the image P is better referred to as the artifact spectral function (instead of the noise spectral function).

Let $\bar{P}$ be the average artifact power spectral density image from our 1000 artifact power spectral density images, P's. This averaged artifact power spectral density image $\bar{P}$ is used in the proposed algorithm.

B. The proposed algorithm

In the conventional BM3D algorithm, the noise is assumed to be stationary. The Wiener filter is used for denoising in the BM3D algorithm. The Wiener filter assumes stationary noise with a noise power spectral density function $\bar{P}$. However, the artifacts are not stationary. Strictly speaking, it is not proper to use our artifact power spectral density function in the BM3D algorithm. Despite of these concerns, we propose an *ad hoc* algorithm:

$$H = BM3D(G_{CT}, \bar{P}), \qquad (2)$$

where G_{CT} is a 2D given sparse-view CT image, $\bar{P}$ is the averaged artifact spectral density image, H is the processed output image, and *BM3D* is the conventional BM3D algorithm.

We must point out that in calculating $\bar{P}$, the sparse simulation G_{simu} in (1) must have the same imaging and sampling parameter as the situation when sparse-scan CT image, G_{CT}, is obtained. For example, if G_{CT} is reconstructed from a data set of 200 views and with a focal-point to axis-of-rotation of 600 mm, the P image must be obtained using 200 views and 600 mm as well for the sparse-view data.

C. Computer simulations

We generated 1000 noiseless random 256×256 phantoms, each of which had 2 random ellipses of random shapes, random locations, and random intensities. We generated 2 versions of projections for each computer-generated phantom: one with 60 views over 360° (sparse scan case); the other one with 180 views over 360° (full scan case). Images were reconstructed using the filtered backprojection (FBP) algorithm using all projections for both full scan and sparse scan cases. One averaged artifact power spectral density image, $\bar{P}$, was calculated from these 1000 phantoms.

We then generated a new random 256×256 phantom and generated a sparse scan with 60 views (test case). The FBP reconstruction, G_{CT}, was calculated from this new test case 60-view data. The proposed algorithm (2) was applied to this FBP image, G_{CT}, to obtain the final image, H.

D. Clinical data

Here we had one set of sparse-scan CT images for one patient. The set contained 512×512 2D images. The original projections were not available. We knew the imaging geometry. The number of views was 200 views over 360°. In order to use the proposed method to reduce the angular aliasing artifacts, we generated a new averaged artifact power spectral density image $\bar{P}$ with 1000 512×512 2D random computer simulated sparse/full image pairs.

III. RESULTS

A. Computer simulation results

Fig. 1 shows 2 (out of 1000) representative random phantoms. Their sparse-view versions using 60 views are shown in Fig. 2. Fig. 3 shows the average artifact power spectral density image by considering 1000 sparse/full pairs of the simulated images.

Two new random phantoms sparse-scan images are shown in Fig. 4. These new phantoms are NOT among the 1000 phantoms used in estimating the artifact power spectral density image, because the new ones contain 3 ellipses while the old ones contain 2 ellipses. The results of the proposed method are shown in Fig. 5.

B. Patient data results

Three patient image pairs are shown in Figs. 6, 7 and 8, respectively. The images are sparse-scan images without and with the proposed BM3D processing. Fig. 9 shows the $\bar{P}$ image for the patient study.

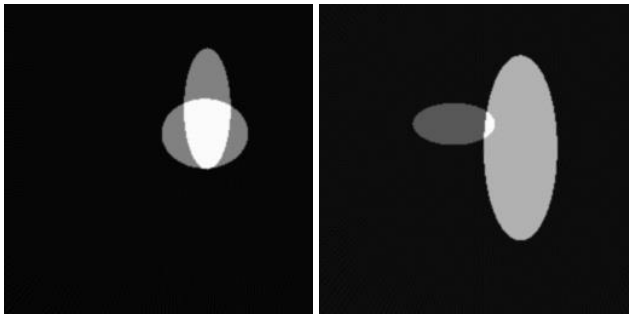


Figure 1. Computer simulated random full-scan images.

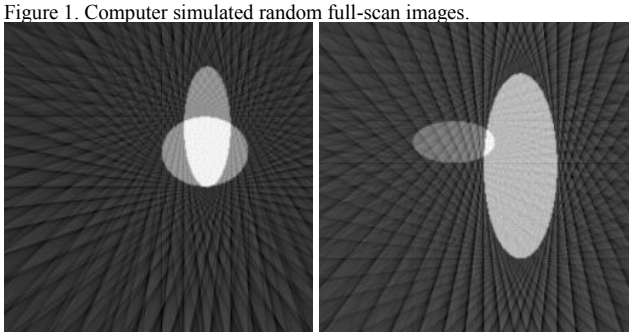


Figure 2. Computer simulated random sparse-scan images.

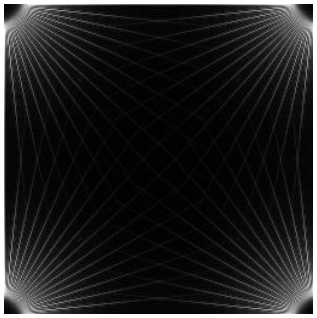


Figure 3. The averaged artifact power spectral density image for the computer simulation study.

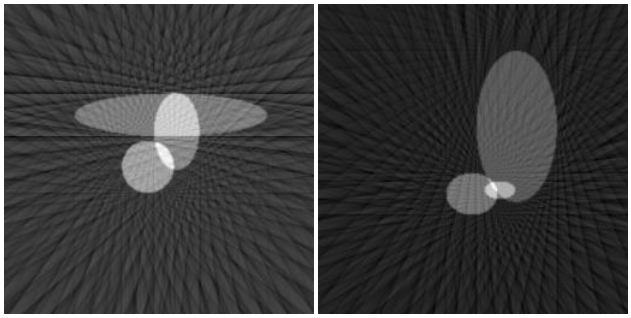


Figure 4. The test sparse-scan image.

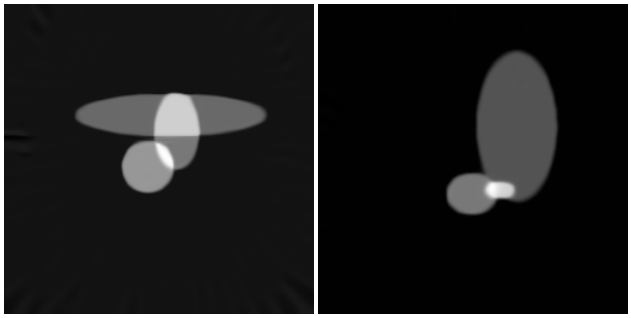


Figure 5. The test sparse-scan image processed by the proposed method.

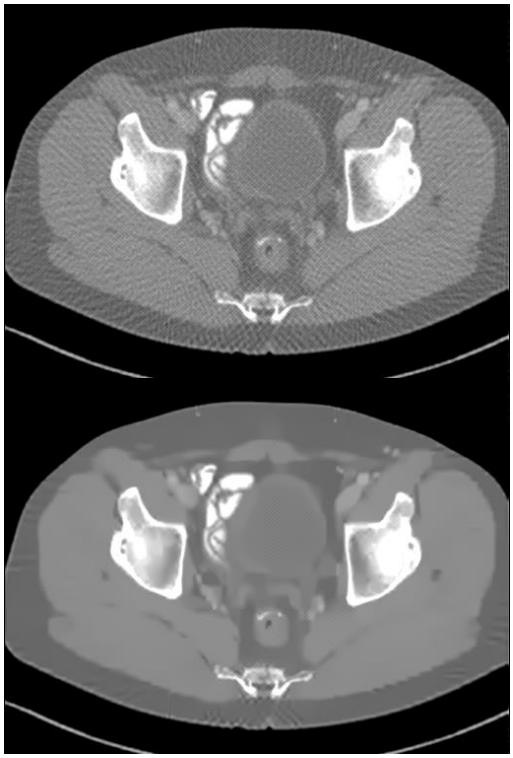


Figure 6. The sparse-scan patient image slice #160 before (upper) and after (lower) processing by the proposed method.

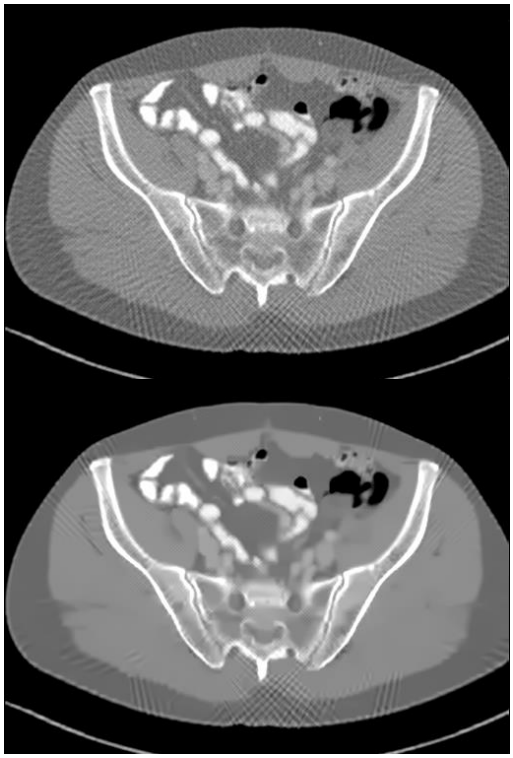


Figure 7. The sparse-scan patient image slice #120 before (upper) and after (lower) processing by the proposed method.

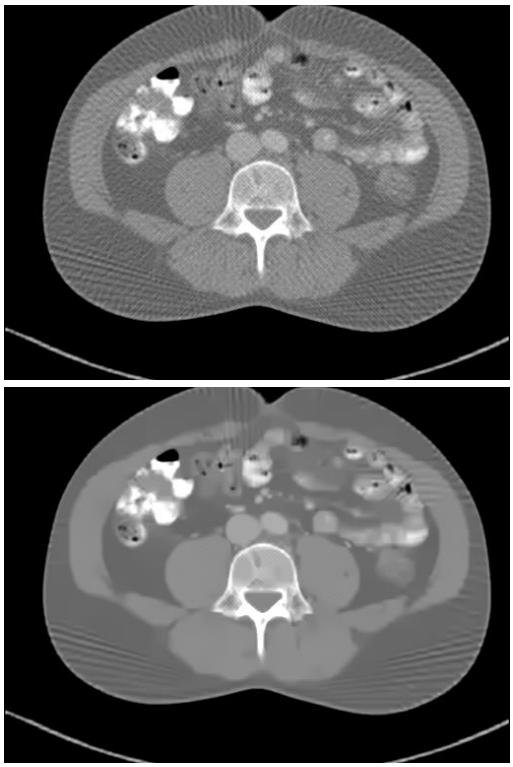


Figure 8. The sparse-scan patient image slice #100 before (upper) and after (lower) processing by the proposed method.

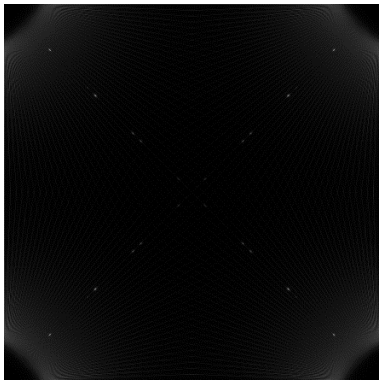


Figure 9. The averaged artifact power spectral density image for the patient CT image processing.

V. CONCLUSIONS

We have attempted a method to reduce the sparse-scan angular aliasing artifacts without using any patient training data. This method is a direct application of the BM3D filter by replacing the noise power spectral density function with the artifact power spectral density function.

The BM3D filter assumes stationary noise that is characterized by the noise power spectral density function. Noise and artifacts are never the same. Noise is random, while artifacts are somewhat deterministic. Artifacts are not stationary. Strictly speaking, the artifact power spectral density function does not exist because it is not stationary.

Our *ad hoc* method assumes the norm square of the Fourier transform of the error image as the artifact power spectral density function, which is calculated with computer simulations and depends on the imaging geometry only. Patient data is not used in finding the artifact power spectral density function.

Our results indicate that the proposed method is not effective enough for practical applications. The artifacts are still present, and the images are over-smoothed after processing. More work needs to be done. However, insights our from this study suggest that some features can be obtained my simulations when there is no real data available. Another thing we observe is that the Wiener filter is not an effective method to remove artifacts, and a better approach should be considered.

REFERENCES

[1] W. Yu, C. Wang, X. Nie, and D. Zeng, "Sparsity-induced dynamic guided filtering approach for sparse-view data toward low-dose x-ray computed tomography," *Phys. Med. Biol*., vol. 63, no. 23, 235016, 2018.

[2] M. Lell and M. Kachelrieß, "Recent and upcoming technological developments in computed tomography: High speed, low dose, deep learning, multienergy," *Investigative Radiology*, vol. 55, Issue 1, pp. 8-19, 2020. doi: 10.1097/RLI.0000000000000601

[3] M. Messerli, T. Kluckert, M. Knitel, *et al*, "Ultralow dose CT for pulmonary nodule detection with chest x-ray equivalent dose—a prospective intra-individual comparative study," *Eur Radiol*., vol. 27, pp. 3290–3299, 2017.

[4] A. W. K. Yeung, "The 'As Low As Reasonably Achievable' (ALARA) principle: a brief historical overview and a bibliometric analysis of the most cited publications," *Radioprotection*, vol. 54(2), pp. 103–109, 2019. https://doi.org/10.1051/radiopro/2019016

[5] E. Y. Sidky and X. Pan, "Image reconstruction in circular cone-beam computed tomography by constrained, total-variation minimization," *Phys. Med. Biol*., vol. 53(17), pp. 4777–4807, 2008. doi: 10.1088/0031-9155/53/17/021.

[6] S. Abbas, J. Min, and S. Cho, "Super-sparsely view-sampled cone-beam CT by incorporating prior data," *J. X-Ray Sci. Technol*., vol. 21(1), pp. 71–83, 2013.

[7] J. Huang, Y. Zhang, J. Ma, D. Zeng, Z. Bian, S. Niu, Q. Feng, Z. Liang, and W. Chen, "Iterative image reconstruction for sparse-view CT using normal-dose image induced total variation prior," *PLoS One*, vol. 8(11), e79709, 2013. doi: 10.1371/journal.pone.0079709.

[8] Z. Zheng, Y. Hu, A. Cai, W. Zhang, J. Li, B. Yan, and G. Hu, "Few-view computed tomography image reconstruction using mean curvature model with curvature smoothing and surface fitting," *IEEE Trans Nucl Sci*., vol. 66(2), pp. 585–596, 2019. doi: 10.1109/TNS.2018.2888948.

[9] G. A. Jones and P. Huthwaite, "Limited view X-ray tomography for dimensional measurements," *NDT & E Int*., vol. 93, pp. 98–109, 2018. doi: 10.1016/j.ndteint.2017.09.002.

[10] V. V. Vlasov, A. B. Konovalov, and S. V. Kolchugin, "Hybrid algorithm for few-views computed tomography of strongly absorbing media: algebraic reconstruction, TV-regularization, and adaptive segmentation," *J. Electron Imag*., vol. 27(4), 043006, 2018. doi: 10.1117/1.JEI.27.4.043006.

[11] C. de Molina, E. Serrano, J. Garcia-Blas, J. Carretero, M. Desco, and M. Abella, "GPU-accelerated iterative reconstruction for limited-data tomography in CBCT systems," *BMC Bioinformatics*, vol. 19, 171, 2018. doi: 10.1186/s12859-018-2169-3.

[12] Y. Han and J. C. Ye, "Framing U-net via deep convolutional framelets: application to sparse-view CT," *IEEE Transactions on Medical Imaging*, vol. 37, no. 6, pp. 1418-1429, June 2018, doi: 10.1109/TMI.2018.2823768.

[13] H. Lee, J. Lee, H. Kim, B. Cho and S. Cho, "Deep-neural-network-based sinogram synthesis for sparse-view CT image reconstruction," *IEEE Transactions on Radiation and Plasma Medical Sciences*, vol. 3, no. 2, pp. 109-119, March 2019, doi: 10.1109/TRPMS.2018.2867611.

[14] Z. Zhang, X. Liang, X. Dong, Y. Xie and G. Cao, "A sparse-view CT reconstruction method based on combination of DenseNet and deconvolution," *IEEE Transactions on Medical Imaging*, vol. 37, no. 6, pp. 1407-1417, June 2018, doi: 10.1109/TMI.2018.2823338.

[15] S. Xie, X. Zheng, Y. Chen, *et al*., "Artifact removal using improved GoogleNet for sparse-view CT reconstruction," *Sci. Rep*., vol. 8, 6700, 2018. https://doi.org/10.1038/s41598-018-25153-w

[16] W. Wu, D. Hu, C. Niu, H. Yu, V. Vardhanabhuti and G. Wang, "DRONE: Dual-domain residual-based optimization network for sparse-view CT reconstruction," *IEEE Transactions on Medical Imaging*, vol. 40, no. 11, pp. 3002-3014, Nov. 2021, doi: 10.1109/TMI.2021.3078067.

[17] C. Zhang, Y. Li, G-H Chen, "Accurate and robust sparse-view angle CT image reconstruction using deep learning and prior image constrained compressed sensing (DL-PICCS)." *Medical Physics*, vol. 48, issue10, pp. 5765-5781, October 2021. https://doi.org/10.1002/mp.15183

[18] J. Liu, Y. Sun, W. Gan, X. Xu, B. Wohlberg and U. S. Kamilov, "SGD-Net: Efficient model-based deep learning with theoretical guarantees," *IEEE Transactions on Computational Imaging*, vol. 7, pp. 598-610, 2021, doi: 10.1109/TCI.2021.3085534.

[19] J. Xiang, Y. Dong and Y. Yang, "FISTA-net: Learning a fast iterative shrinkage thresholding network for inverse problems in imaging," *IEEE Transactions on Medical Imaging*, vol. 40, no. 5, pp. 1329-1339, May 2021, doi: 10.1109/TMI.2021.3054167.

[20] K. Dabov, A. Foi, V. Katkovnik, and K. Egiazarian, "Image denoising by sparse 3-D transform-domain collaborative filtering," *IEEE Trans. Image Process*., vol. 16, pp. 2080–2095, 2007. doi: 10.1109/TIP.2007.901238.

[21] K. Dabov, A. Foi, V. Katkovnik, and K. O. Egiazarian, "Image restoration by sparse 3D transform-domain collaborative filtering," *Image Processing: Algorithms and Systems VI, International Society for Optics and Photonics*; San Jose, CA, USA, p. 681207, 2008.

Assessment of perceptual quality measures for multi-exposure radiography and tomography

Joaquim G. Sanctorum, Sam Van der Jeught, Sam Van Wassenbergh, Joris J. J. Dirckx

***Abstract*—Fusion of x-ray projection images obtained with different exposure levels is a promising technique to study objects with features beyond the dynamic range of the x-ray detector. Various multi-exposure fusion techniques are described in the literature, yet a direct comparison between these methods is not available. This was mainly due to the absence of objective quality measures dedicated to multi-exposure x-ray images and tomographic reconstructions, a problem remaining unsolved to this day. Therefore, we compare several fusion algorithms in terms of perceptual quality using recently reported quality measures based on structural similarity in this work. Moreover, we investigate whether these quality measures apply to tomographic slices as well. Our results indicate that the reliability of the quality measures is more convincing for fused projection images as opposed to reconstructed slices. Additionally, it is shown that fusion algorithms developed for optical photography are also suitable for multi-exposure x-ray image fusion to increase perceptual quality.**

***Index Terms*—dynamic range, multi-exposure fusion, structural similarity, quality measures**

I. Introduction

Due to the limited dynamic range of x-ray detectors, attenuation information may be incomplete in a single exposure image or projection series. In radiography, this problem can occur when the object is a heterogeneous mixture of materials with highly differing attenuation properties [1] or when the object has a wide range of thickness components [2]. In computed tomography (CT), since the object is scanned under many different angles, objects may have an aspect ratio that is too large for the dynamic range of the detector [3]. These issues cause under- and overexposure in the recorded projections, inhibiting correct image evaluation or subsequent tomographic reconstruction.

Apart from hardware modifications to increase the dynamic range of detectors, various methods have been proposed to increase the information contained in x-ray projections through multi-exposure fusion (MEF). In essence, projections are recorded with different exposure parameters to capture information in multiple attenuation ranges, which are then fused to gather the information in different attenuation ranges in the same (fused) projection image. Different exposure situations and fusion methods are created by altering the integration time [1], the tube current [3], [4], or the voltage [2], [5].

Although various acquisition and fusion schemes were reported, to the best of our knowledge, no direct comparison has been made between those methods. This is partly because the methods were developed in different fields and for different purposes (although the problem is general), but more importantly, there was no reported objective quality measure for MEF x-ray images.

For the fusion methods, it is vital to investigate which method represents best the fused information from the input images in the fused image. This work compares several fusion methods and evaluates their perceptual quality using MEF quality measures reported recently in the literature. Additionally, we investigate the possibility to use these measures to assess the perceptual quality of CT slices obtained with MEF projection data. The great strength of such a measure is that a subjective property, like perceptual quality, can be described with an objective measure, independent of the observer.

II. Materials and methods

A. Image acquisition

Radiographs of a preserved piglet specimen were acquired using a single source-detector pair of our stereoscopic x-ray set-up, the 3D²YMOX system [6], [7]. The piglet specimen was chosen for its high aspect ratio, resulting in a considerable difference in absorption between longitudinal and lateral transmission. During the acquisition, the voltage was set to 60 kVp. A low-exposure dataset was recorded using a tube current of 45 mA, and the tube current was put to 90 mA for a high-exposure dataset. For each dataset, 450 projections (2048 pixels × 2048 pixels, pixel size of 0.143 mm) were recorded during continuous rotation of the rotation stage (rotation period of 3 s, stationary source and detector) with a shutter speed of 0.5 ms. The distances from the x-ray source (SRD) and the detector (DRD) to the axis of rotation were 1025 mm and 278 mm, respectively. As the set-up is highly modular, the geometry of the set-up was calibrated using a phantom-based method [8], [9]. Prior to geometry calibration and reconstruction, the

Abstract was submitted 17 January 2022. This work was funded by a grant from the special research fund of the University of Antwerp (BOF-GOA 2016 33927)

J. G. Sanctorum, S. Van der Jeught, and J. J. J. Dirckx are with the Biophysics and Biomedical Physics research group (BIMEF), Department of Physics, University of Antwerp, 2020 Antwerp, Belgium (e-mail: joaquim.sanctorum@uantwerpen.be, sam.vanderjeught@uantwerpen.be, joris.dirckx@uantwerpen.be).

S. Van Wassenbergh is with the Functional Morphology (FunMorph) research group, Department of Biology, University of Antwerp, 2610 Wilrijk, Belgium (e-mail: sam.vanwassenbergh@uantwerpen.be).

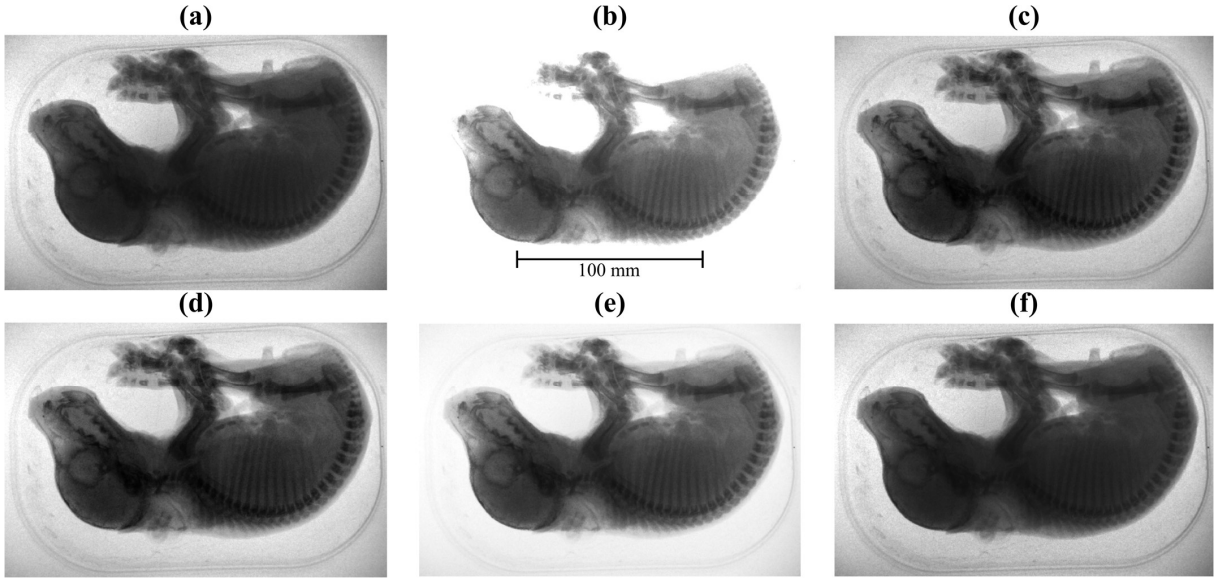


Fig. 1. The original and fused radiographs of the piglet specimen (lateral transmission). **(a)** Low exposure. **(b)** High exposure. **(c)** Mertens. **(d)** Paul. **(e)** Sisniega. **(f)** Krämer. In these projections, the axis of rotation is positioned in the center, from top to bottom. In each panel, the gray values range from 0 to 4095. The scale bar in the top center panel applies to all panels.

geometric distortion in the projections induced by the x-ray image intensifier was corrected using a method based on digital image correlation [10], [11].

B. Reconstruction

Tomographic reconstruction was carried out using the ASTRA-toolbox [12] with a Matlab interface (version 2021b). The reconstruction algorithm of choice was the simultaneous iterative reconstruction technique (SIRT), of which 150 iterations were performed with a non-negativity constraint. The voxel size was isotropic and measured 0.246 mm. As the imaging set-up was not calibrated for the attenuation coefficient of water, the gray values of the reconstructed volume were not scaled to Hounsfield units. Instead, to use the objective quality measures, the gray values were scaled between 0 and 4095, which is the gray value range of the x-ray detector.

C. Quality measures and assessment

Ma *et al*. [13] proposed an objective quality measure for MEF images in optical photography based on the structural similarity index measure (SSIM) [14], denoted as multi-exposure-fusion SSIM, or MEF-SSIM. In their work, the MEF-SSIM is calculated from the local contrast and structural components of corresponding image patches of the input sequence and the fused image. The calculation is done on multiple scales to incorporate both small-range and large-range luminance patterns and structures.

A perceptual quality measure dedicated to x-ray projections was proposed very recently by Qi *et al.* [15], which is also based on the SSIM. Their method calculates the SSIM from the fused maximum gradient amplitude map from the input sequence and the gradient map from the fused image, weighted with the contrast sensitivity function (CSF) to obtain a final measure compatible with the human visual system (HVS). The authors did not explicitly name their proposed method, but since it is based on the gradient amplitude map weighted with the CSF, we refer to their method as weighted-gradient-amplitude SSIM or WGA-SSIM. The mathematical framework of both measures is not included, as it is well documented in the original publications [13], [15].

Both of these measures will be used to assess the quality of exposure-fused x-ray projections and CT reconstructions obtained with different fusion methods. In the present work, we consider two fusion methods originating from optical photography proposed by Mertens *et* al. [16] and by Paul *et al.* [17], and two fusion methods dedicated to x-ray images recorded with different tube currents proposed by Krämer *et al*. [3] and by Sisniega *et al*. [4]. The fusion methods will be referred to by the corresponding name of their first author.

III. Results

A. Radiographs

High and low exposure sets of projections of the piglet specimen were recorded and fused with the fusion methods mentioned before. Examples of projections are shown in Fig. 1. Visually, from Fig. 1(a), there is little contrast between the cervical vertebrae and the surrounding soft tissue when a low exposure is used. The contrast is increased when a high exposure is used, at the expense of loss of soft tissue attenuation (for example, at the snout and tail). By fusion of these projections, the contrast in certain regions can be increased while retaining the soft tissue attenuation, as is shown in Figs. 1(c)-(e). The method of Krämer was included intentionally, as it produces nearly the same projection as the low-exposure one, but with a higher signal-to-noise ratio (SNR), as was reported

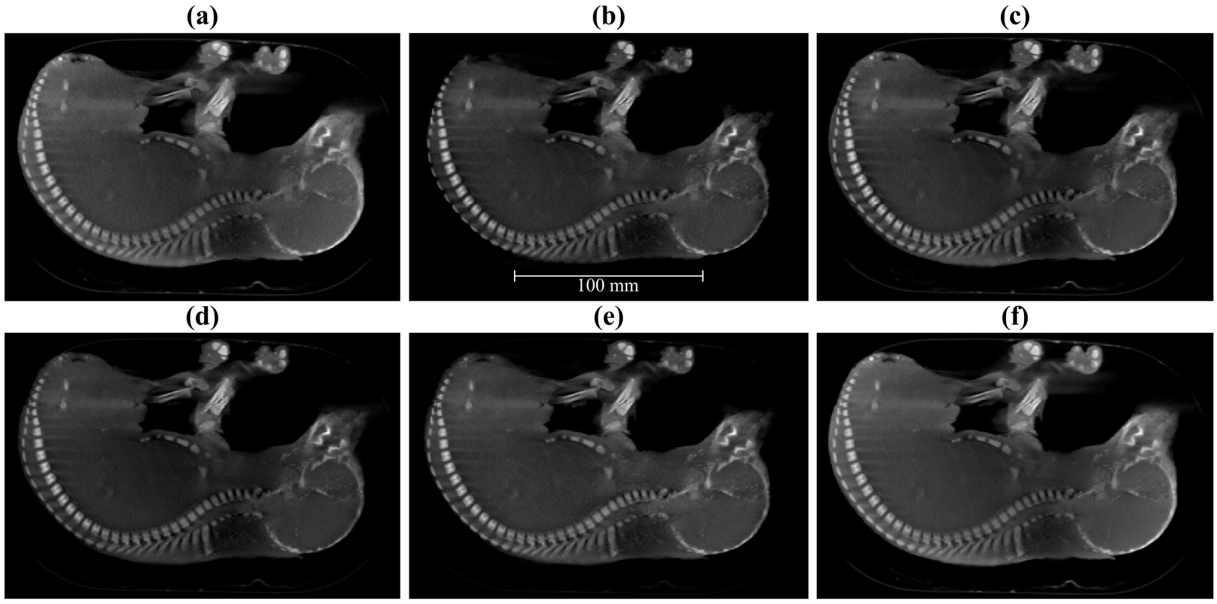


Fig. 2. Reconstructed slices of the piglet specimen. **(a)** Low exposure. **(b)** High exposure. **(c)** Mertens. **(d)** Paul. **(e)** Sisniega. **(f)** Krämer.

by the authors and observed in our results.

The fusion methods all produce different results, and the main question is which one has the best perceptual quality. To answer this question, the MEF-SSIM and WGA-SSIM scores of the 450 fused images in the projection series were calculated, and the average values with corresponding standard deviation are shown in Table I. The MEF-SSIM measure favors the method of Mertens, whereas the WGA-SSIM measure favors the method of Paul, as a higher score indicates a higher quality (with a maximum of 1).

TABLE 1
QUALITY MEASURES FOR FUSED PROJECTIONS AND SLICES

	Mertens	Paul	Sisniega	Krämer
$MEF\text{-}SSIM_P$	**0.988±0.004**	0.986±0.002	0.897±0.003	0.96±0.02
$WGA\text{-}SSIM_P$	0.986±0.002	**0.9903±0.0007**	0.984±0.003	0.981±0.003
$MEF\text{-}SSIM_S$	**0.93±0.06**	0.91±0.06	0.90±0.06	0.92±0.06
$WGA\text{-}SSIM_S$	**0.986±0.008**	0.981±0.006	0.979±0.005	0.98±0.01
CNR_{CV}	1.65	1.94	**2.08**	1.73
CNR_{ST}	3.03	3.51	**3.53**	3.39

In the names of the SSIM quality measures, subscripts P and S denote projection and slice, respectively. The subscripts CV and ST stand for cervical vertebrae and sternum for the CNR values. The highest reported values are indicated in bold.

B. Reconstructed slices

In addition to the fused projection series, a fused flat-field image was obtained with each fusion method from low-exposure and high-exposure flat-field images, subsequently used for flat-field and log correction prior to tomographic reconstruction. Examples of reconstructed slices are shown in Fig 2. In contrast to the radiographs, the difference between the slices is less apparent. Visually, the high-exposure reconstructed slice yields a better contrast between bones and soft tissue (for example, at the cervical vertebrae), at the expense of loss of signal in the soft tissue (missing tail and parts of the trotters, top center) and even some bones in the tail are not reconstructed. The MEF-SSIM and WGA-SSIM scores of the reconstructed slices are shown in Table I. From these measures, the method of Mertens appears to yield the best results.

As the HVS is sensitive to contrast, we also present contrast-to-noise (CNR) values between soft tissue and bone in one of the cervical vertebrae and the sternum, found in Table I. It is shown that the method of Sisniega yields the greatest CNR in both locations. The CNR values were calculated as [4]:

$$CNR = \frac{|\mu_{bone} - \mu_{tissue}|}{\sqrt{\sigma_{bone}^2 + \sigma_{tissue}^2}}, \quad (1)$$

in which μ and σ are the mean pixel gray value and corresponding standard deviation in an 8 pixels × 8 pixels image region.

IV. DISCUSSION

Table I shows that the MEF-SSIM measure indicates that the projections fused with the method of Mertens yield the best perceptual result. However, the error bar overlaps with that of the method of Paul. Visually, both projections are indeed very similar (Fig. 1). The method of Paul is favored by the WGA-SSIM measure, which is partly in agreement with the MEF-SSIM measure.

It is quite interesting that both quality measures indicate that the fusion methods originating from optical photography outperform those dedicated to x-ray imaging, as these optical methods were not considered for x-ray imaging before, to the best of our knowledge. Visually, the result obtained using the method of Sisniega is most dissimilar compared to the others, which is well represented by its lower MEF-SSIM score, possibly caused by the low contrast between the background and the soft tissue. On the other hand, all WGA-SSIM scores are very similar (all error bars overlap, except for the method of

Paul), while the fused projections are visually not. The relative differences between the MEF-SSIM scores are more representative of the perceptual differences of the projections than the WGA-SSIM scores.

In the case of the slices, the quality scores indicate that the method of Mertens yields the best perceptual results. However, it is important to notice that all error bars overlap for both quality measures. Visually, the method of Sisniega provides good contrast between soft tissue and bone (for example, in the cervical vertebrae), which is also represented by the CNR values. Moreover, the tail is reconstructed and most of the soft tissue is visible. The highest quality scores for the method of Mertens are not supported by the CNR values and actual visual perception. The soft tissue near the cervical vertebrae is quite dark, as is the case when the method of Paul is used, which is probably an amplified cupping artifact due to beam-hardening. In comparison, the gray values of the soft tissue are more uniform in the result obtained with the method of Sisniega.

These results suggest that the proposed quality measures are not reliable for reconstructed slices, yet the perceptual quality scores of the fused projections are very reasonable and in agreement with the visual quality. Moreover, the results imply that high perceptual quality in the fused projections does not guarantee high perceptual quality in the reconstructed slices. This indicates that quality measures dedicated to tomographic reconstructions are needed to validate the quality of multi-exposure methods for tomographic purposes.

V. CONCLUSIONS

The results demonstrate that the quality measures are reliable for selecting the fusion method that yields the highest perceptual quality in fused projections and that optical fusion methods are suitable for MEF x-ray imaging as well. Yet, the same is not valid for reconstructed slices. As reconstructed slices have other demands than projection data, dedicated quality measures are needed. In future research, we plan to investigate the possibilities for a new quality measure suitable for reconstructed slices obtained from fused projection data.

REFERENCES

[1] M. A. Haidekker, L. D. Morrison, A. Sharma, and E. Burke, ‘Enhanced dynamic range x-ray imaging’, *Computers in Biology and Medicine*, vol. 82, pp. 40–48, Mar. 2017, doi: 10.1016/j.compbiomed.2017.01.014.

[2] L. Bin, H. Yan, P. Jinxiao, and C. Ping, ‘Multi-energy image sequence fusion based on variable energy X-ray imaging’, *Journal of X-Ray Science and Technology*, no. 2, pp. 241–251, 2014, doi: 10.3233/XST-140422.

[3] P. Krämer and A. Weckenmann, ‘Multi-energy image stack fusion in computed tomography’, *Measurement Science and Technology*, vol. 21, no. 4, p. 045105, Apr. 2010, doi: 10.1088/0957-0233/21/4/045105.

[4] A. Sisniega, M. Abella, M. Desco, and J. J. Vaquero, ‘Dual-exposure technique for extending the dynamic range of x-ray flat panel detectors’, *Physics in Medicine and Biology*, vol. 59, no. 2, pp. 421–439, Jan. 2014, doi: 10.1088/0031-9155/59/2/421.

[5] P. Chen, S. Yang, Y. Han, J. Pan, and Y. Li, ‘High-dynamic-range X-ray CT imaging method based on energy self-adaptation between scanning angles’, *OSA Continuum*, vol. 3, no. 2, p. 253, Feb. 2020, doi: 10.1364/OSAC.380634.

[6] J. G. Sanctorum *et al.*, ‘Methods for characterization and optimisation of measuring performance of stereoscopic x-ray systems with image intensifiers’, *Meas. Sci. Technol.*, vol. 30, no. 10, p. 105701, Aug. 2019, doi: 10.1088/1361-6501/ab23e7.

[7] ‘3D2YMOX | 3-Dimensional DYnamic MOrphology using X-rays (3D2YMOX) | University of Antwerp’. https://www.uantwerpen.be/en/research-groups/3d2ymox/ (accessed Jan. 03, 2022).

[8] V. Nguyen *et al.*, ‘A low-cost geometry calibration procedure for a modular cone-beam X-ray CT system’, *Nondestructive Testing and Evaluation*, pp. 1–14, Jun. 2020, doi: 10.1080/10589759.2020.1774580.

[9] V. Nguyen, J. G. Sanctorum, S. Van Wassenbergh, J. J. J. Dirckx, J. Sijbers, and J. De Beenhouwer, ‘Geometry Calibration of a Modular Stereo Cone-Beam X-ray CT System’, *Journal of Imaging*, vol. 7, no. 3, Art. no. 3, Mar. 2021, doi: 10.3390/jimaging7030054.

[10] J. G. Sanctorum, S. Van Wassenbergh, P. Aerts, and J. J. J. Dirckx, ‘Technical Note: Correction of geometric x-ray image intensifier distortion based on digital image correlation’, *Medical Physics*, vol. 47, no. 2, pp. 597–603, 2020, doi: 10.1002/mp.13944.

[11] J. G. Sanctorum, S. Van Wassenbergh, V. T. H. Nguyen, J. De Beenhouwer, J. Sijbers, and J. J. J. Dirckx, ‘Projection-angle-dependent distortion correction in high-speed image-intensifier-based x-ray computed tomography’, *Measurement Science and Technology*, Aug. 2020, doi: 10.1088/1361-6501/abb33e.

[12] W. van Aarle *et al.*, ‘Fast and flexible X-ray tomography using the ASTRA toolbox’, *Opt. Express, OE*, vol. 24, no. 22, pp. 25129–25147, Oct. 2016, doi: 10.1364/OE.24.025129.

[13] K. Ma, K. Zeng, and Z. Wang, ‘Perceptual Quality Assessment for Multi-Exposure Image Fusion’, *IEEE Transactions on Image Processing*, vol. 24, no. 11, pp. 3345–3356, Nov. 2015, doi: 10.1109/TIP.2015.2442920.

[14] Z. Wang, A. C. Bovik, H. R. Sheikh, and E. P. Simoncelli, ‘Image Quality Assessment: From Error Visibility to Structural Similarity’, *IEEE Trans. on Image Process.*, vol. 13, no. 4, pp. 600–612, Apr. 2004, doi: 10.1109/TIP.2003.819861.

[15] Y. Qi, Z. Yang, and L. Kang, ‘Multi-exposure X-ray image fusion quality evaluation based on CSF and gradient amplitude similarity’, *XST*, vol. 29, no. 4, pp. 697–709, Jul. 2021, doi: 10.3233/XST-210871.

[16] T. Mertens, J. Kautz, and F. V. Reeth, ‘Exposure Fusion’, in *15th Pacific Conference on Computer Graphics and Applications (PG’07)*, Oct. 2007, pp. 382–390. doi: 10.1109/PG.2007.17.

[17] S. Paul, I. S. Sevcenco, and P. Agathoklis, ‘Multi-Exposure and Multi-Focus Image Fusion in Gradient Domain’, *J CIRCUIT SYST COMP*, vol. 25, no. 10, p. 1650123, Oct. 2016, doi: 10.1142/S0218126616501231.

Geometric calibration of seven degree of freedom Robotic Sample Holder for X-ray CT

Erdal Pekel, Florian Schaff, Martin Dierolf, Franz Pfeiffer, and Tobias Lasser

***Abstract*—We present a geometric calibration method for integrating a seven degrees of freedom robotic arm as a sample holder within an existing laboratory X-ray computed tomography setup. We aim to provide a flexible sample holder that is able to execute non-standard and task-specific trajectories for complex samples. The calibration is necessary to identify the accurate pose of the sample which deviates from the expected pose due to inaccurate placement of the robotic arm. The robotic arm is integrated with a unified software package that allows for path planning, collision detection, geometric calibration and reconstruction of the sample. With our software the user is able to command the robotic arm to execute arbitrary trajectories for a given sample in a safe manner and output its reconstruction to the user. We present experimental results with a circular trajectory where the robotic sample holder achieves identical visual quality compared to a conventional sample holder.**

I. Introduction

In this work we introduce a geometric calibration method for using a flexible robotic arm with seven degrees of freedom as a sample holder within a laboratory X-ray Computed Tomography (CT) setup. The calibration mechanism is required as a result of the insufficient placement accuracy of the robotic arm. A purpose-built sample holder with an embedded geometric structure is used to calibrate the position and orientation of the sample for later use in the reconstruction step.

The robotic arm adds flexibility to the setup as a sample holder by enabling arbitrary rotation and placement of the sample. This allows non-standard trajectories that are not restricted in their sequence, such as conventional circular or helical trajectories. In addition, the robotic sample holder can avoid occlusions on the projections that would normally be introduced by limitations of static setups where the sample is inherently mounted to non-moving parts (e.g. mounted on a plate). In the following we present our work on the integration of a robotic arm with seven degrees of freedom within a lab X-ray CT setup together with a suitable calibration mechanism.

The system can easily execute specific trajectories that can overcome the limitations of fixed trajectories which will enable imaging modalities that require non-standard acqusition sequences in the future, such as Anisotropic X-ray Dark-field Tomography, a novel imaging technique that allows the extraction of X-ray scattering and phase contrast information by employing grating interferometers [1], [2]. The robotic sample holder will enable arbitrary rotations covering the full sphere and hence expose the 3D structures of the target object by measuring the full dark-field contrast from all possible angles.

E. Pekel and T. Lasser are with Computational Imaging and Inverse Problems, Department of Informatics, School of Computation, Information, and Technology, and Munich Institute of Biomedical Engineering, Technical University of Munich, Germany.

Florian Schaff, Martin Dierolf and Franz Pfeiffer are with the Chair of Biomedical Physics, Department of Physics, School of Natural Sciences, and Munich Institute of Biomedical Engineering, Technical University of Munich, Germany.

Franz Pfeiffer is also with the Department of Diagnostic and Interventional Radiology, School of Medicine, and Klinikum rechts der Isar, Technical University of Munich, Germany.

II. Robotic Sample Holder

In this section the methods for operating the robotic arm as a sample holder in a lab X-ray CT setup are discussed in detail. After introducing the hardware components of the system more specific parts like the sample holder, calibration and reconstruction are described.

A. Hardware Setup

The hardware components of the system are displayed in fig. 1 and 2. The main difference to a conventional X-ray CT setup is the seven degrees of freedom robotic arm *Panda* from the manufacturer FRANKA EMIKA [3]. It has a maximum reach of 855 mm and a repeatability of 0.1 mm when repeatedly moved from a specific starting pose to a goal pose. It has two fingers that can move on a fixed axis and grasp objects. The maximum allowed payload is 3 kg. The robotic arm and the depth cameras are connected directly to a computer while the detector is accessible through a network interface. The robotic arm can be turned off in case of emergency from outside of the safety hutch with a power switch (see fig. 2).

Two *Intel Realsense D435* depth camera capture the movements of the robot and provide 3D information about the surroundings as a point cloud. The cameras are connected directly to the workstation and they are used for the collision detection mechanism.

The robotic arm is mounted on a table inside a safety hutch for X-ray CT which houses the X-ray source and the detector (see fig. 1). The detector has a maximum resolution of 2880x2880 and is connected to a different workstation on the network which provides a network interface for triggering image capturing.

B. Sample Holder

The sample holder is a critical component of the system as it allows the robotic arm to grasp samples of arbitrary shape and is a fundamental part of the calibration process where the position and orientation of the sample is identified. The 3D models of the sample holder and the rail component are visualized in figure 1. The sample holder consists of two parts. The bottom part is where the robot's fingers can grasp the

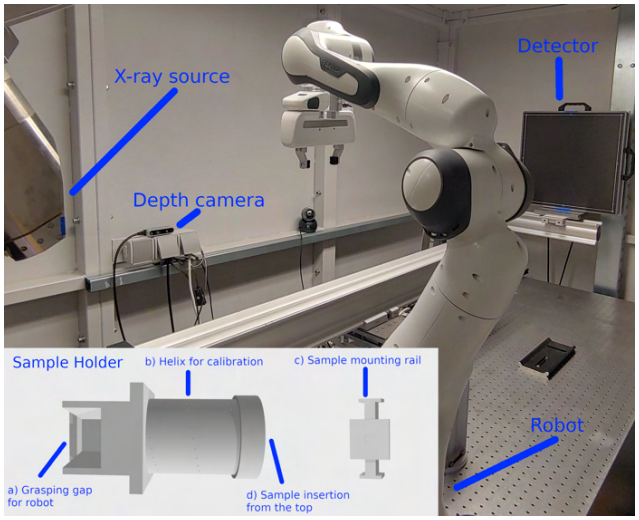


Fig. 1. Hardware setup. The robotic arm is mounted on a table with the source and the detector inside a safety hutch. The source to robot distance is 40 cm and robot to detector distance is 176 cm. The sample holder consists of two physical parts. The first part is grasped by the robotic arm and houses the calibration structure. The second part is where the sample is mounted and slides into the base part.

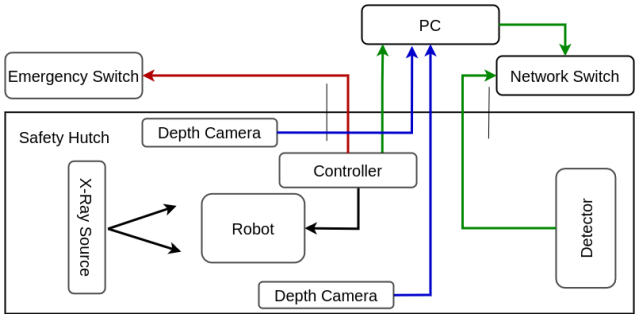


Fig. 2. Hardware schematics. Two depth cameras monitor the movement of the robot and send a stop signal to the robot controller when the executed trajectory interferes with obstacles. The robotic arm can be stopped by a manual power switch that was routed to the operator table outside of the hutch.

holder steadily. The upper part fulfills the actual purpose of placing a geometric structure around the sample on a cylinder.

The cylinder is 5.6 cm tall and 3.5 cm in diameter inside. The sample holder was designed with a 3D modeling software and printed using a 3D printer with accuracy of 0.08 to 0.2 mm on all three axes. The printing accuracy is important as the local coordinates of the spheres in the 3D model are used as reference points in the calibration algorithm.

The geometric structure embedded in the sample holder is a helix which is made up of 50 embedded aluminium spheres of 0.678 mm diameter. These spheres were fixed by hand on notches that were included in the design process of the holder. The spheres appear as circles on the detector images that will be segmented during calibration.

The helix can be parametrized by the following 3D parametric curve:

$$h(\tau) = \begin{pmatrix} u(\tau) \\ v(\tau) \\ w(\tau) \end{pmatrix} = \begin{pmatrix} r * cos(\rho * \tau + \phi) \\ r * sin(\rho * \tau + \phi) \\ \tau \end{pmatrix} \quad (1)$$
$$\tau, r, \rho, \phi \in \mathbb{R}$$

τ runs between the local w coordinates of the first sphere and the last sphere of the helix: $w_{min} < \tau < w_{max}$ where $w_{min}, w_{max} \in \mathbb{R}$.

The parameters r (radius), ρ (frequency) and ϕ (phase shift) parametrize the helix. They can be determined by fitting the sphere coordinates from the 3D model of the sample holder to eq. (1) with a least-squares term. The source code of this process can be found in the file *helix_fitter.py* in our repository [4].

The helix can be discretized by choosing a fixed number $H \in \mathbb{N}$ of points $\{\tau_i\}_{i=1,\dots,H} \in [w_{min}, w_{max}]$ for the free parameter τ:

$$h_i = \begin{pmatrix} u(\tau_i) & v(\tau_i) & w(\tau_i) \end{pmatrix}^\top . \quad (2)$$

C. Calibration

The calibration procedure tackles the issue that the robotic arm does not sufficiently accurately place the sample at the desired position due to inaccurate path planning and inaccurate electrical motors at its joints. With the calibration procedure we are able to identify the actual positions and orientations of the sample for the reconstruction step. For the calibration a sample holder with an embedded geometric structure that can be detected on the detector images is necessary. A suitable sample holder was introduced in section II-B.

The calibration is implemented in multiple steps (see fig. 3). The first step is the post-processing of the detector image. Its contrast is enhanced and a median filter with kernel size 5 is applied to reduce noise and improve the segmentation results. The calibration circles on the image are detected in the next step with the circle Hough transform algorithm [5]. The result is a set of 2D circle center coordinates $\hat{m}_j = \begin{pmatrix} d_{x,j} & d_{y,j} \end{pmatrix}^\top$ on the detector.

Eq. (1) and the current position of the robotic arm are now used to project a set of helix points h_i (eq. (2)) onto the detector image for comparison with the segmented points $\hat{m}_j$ and determining the geometry of the sample.

For this projection the intrinsic camera matrix K and the external parameters R and t are needed. K is fixed for the current X-ray CT setup and R, t are determined by the robotic arm's current position.

There are three critical coordinate systems in our setup. The first is fixed to the X-ray source with x, y and z-axis. The second is fixed to the center of the sample holder with u,v and w-axis and moves with the robotic arm as it is attached to the arm's fingers. The third is fixed to the detector with d_x and d_y axis.

The rotation R of the sample holder relative to the source can be parametrized w.l.o.g. by consecutive rotations about the z, y and x-axis:

$$R(\alpha, \beta, \gamma) = R_z(\alpha) R_y(\beta) R_x(\gamma) \quad (3)$$

t is the offset of the source center to the sample holder's center:

$$t = \begin{pmatrix} x & y & z \end{pmatrix} \tag{4}$$

K is fixed for the current setup and can be calculated with the parameters sdd (source to detector distance), $d_{x,p}, d_{y,p}$ (principal points on d_x and d_y-axis) and d_w, d_h (detector pixel width and height):

$$K = \begin{pmatrix} \frac{sdd}{d_w} & 0 & d_{x,p} \\ 0 & \frac{sdd}{d_h} & d_{y,p} \\ 0 & 0 & 1 \end{pmatrix} \tag{5}$$

We introduce the short notation $\zeta = (\alpha, \beta, \gamma, x, y, z)$ for the free parameters. The camera projection matrix P can now be calculated:

$$P(\zeta) = K \begin{pmatrix} R(\alpha, \beta, \gamma) & | & t \end{pmatrix} \in \mathbb{R}^{3\times 4}. \tag{6}$$

The projection matrix is now used to project a set of $H \in \mathbb{N}$ fixed points $h_i \in \mathbb{R}^4$ on the discretized helix from eq. (2) onto the detector:

$$\begin{pmatrix} d'_{x,i}(\zeta) \\ d'_{y,i}(\zeta) \\ d'_{z,i}(\zeta) \end{pmatrix} = P(\zeta) h_i \tag{7}$$

$$m_i(\zeta) = \begin{pmatrix} d_{x,i}(\zeta) \\ d_{y,i}(\zeta) \end{pmatrix} = \begin{pmatrix} d'_{x,i}(\zeta)/d'_{z,i}(\zeta) \\ d'_{y,i}(\zeta)/d'_{z,i}(\zeta) \end{pmatrix} \tag{8}$$

d'_i are the homogeneous detector pixel coordinates and m_i are the projected analytical helix points on the detector. These points resemble the expected position of the helix structure and they will be used for constructing an error term in the 2D detector image domain.

An appropriate cost function for comparing the error between the current and expected position of a measured circle center $\hat{m}_j$ and a projected point on the helix m_i is the *reprojection error*:

$$E(\zeta, \hat{m}_j, m_i) = \hat{m}_j - m_i(\zeta) \in \mathbb{R}^2 \tag{9}$$

Eq. (9) will only measure the error for a specific pair of points. In our case there are c (detected) circles on the current image, s spheres glued onto the holder and H projected points on the helix from eq. (1). It is important to note that $c \leq s$ because the segmentation algorithm might fail to detect all circles.

We now compare each of the c detected circle centers $\hat{m}_j$ to all H sampled and projected points m_i and choose the pair with the smallest distance.

We can formulate this algorithm as a least-squares problem:

$$\arg\min_{\zeta=(\alpha,\beta,\gamma,x,y,z)} \sum_{j=1}^{c} \min_{1\leq i\leq H} E(\zeta, \hat{m}_j, m_i) \tag{10}$$

The optimization problem is nonlinear due to the sine and cosine terms in the rotation parametrization. In our implementation we use the *Levenberg Marquardt* algorithm.

The resulting parameters $\alpha, \beta, \gamma, x, y, z$ can be used for the reconstruction as the geometry of the given acquisition.

D. Reconstruction

For tomographic reconstruction, the sinogram contained 1000 equidistant X-ray projections along a circular trajectory sized 720×720 pixels with a spacing of $600\mu m$. The reconstruction volume was sized $720 \times 720 \times 720$ with isotropic voxel spacing of $100\mu m$. Using our C++ reconstruction framework *elsa* [6], reconstruction was performed using an iterative conjugate gradient solver run for 50 iterations on a Tikhonov regularized weighted least squares problem, with the Josephs method for X-ray transform discretization and parallel beam geometry. Further iterations showed no improvement on the cost function.

E. Software Stack

The central part of our software stack is the *Robot Operating System* (*ROS*) [7]. Robot manipulation is accomplished with the *MoveIt!* framework [8], [9] and the *franka_ros* configuration package [10]. For image processing tasks and the circle segmentation we use *OpenCV* [11], for multithreading on the CPU *OpenMP* [12] and for the tomographic reconstruction *elsa* [6]. The scientific calculations in section II-C are implemented with *scipy* [13].

F. CT measurements

We conducted two experiments: a walnut was measured with the robotic arm and with a conventional rotational stage.

For each CT measurement, 1000 images were acquired with a source voltage of 30 kV, source power of 1445μA, and exposure time of 1s.

In fig. 4 a) the reconstruction of the walnut with the rotational stage is compared to the robotic arm as the sample holder. The two volumes were registered manually as we found automatic registration of the two discretized volumes to be unreliable. The slices were chosen manually for illustration purposes. The center slices of the volume from the top and the front view were extracted and cropped to the region of interest.

Our observation is that there is no qualitative difference between the results of the two different sample holders.

G. Future work

In future work the system can be improved in several ways.

The sample holder could be more flexible. Its size currently limits the size of the sample but this can be tackled in another design iteration by embedding the geometrical calibration structure into the base of the holder when it is positioned upwards of the base and compressed in its height. The cylindrical envelope surrounding the sample could be removed and as a consequence, the sample also doesn't strictly need to be inserted from the top.

Moreover, experiments with non-standard trajectories are subject of future work.

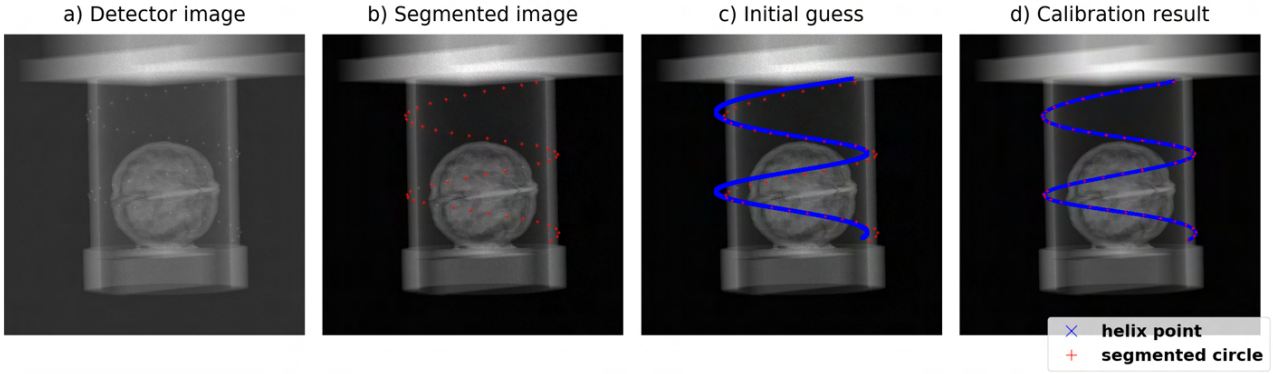


Fig. 3. Calibration procedure. In a) the flat-field corrected detector image is displayed. This image is contrast-enhanced and subsequently a circle detection algorithm is executed. The resulting image where the detected circle centers are marked with red crosses is displayed in b). Given the geometry of the sample holder and the robot's sensor readings when acquiring the image, an initial guess of the helix location (blue crosses) is projected onto the image plane (c). The parameters that define the rotation and translation of the helix are optimized in a least-squares problem in the 2D image domain. The resulting parameters are used to project the helix again to the image domain to display the final outcome of the calibration (d).

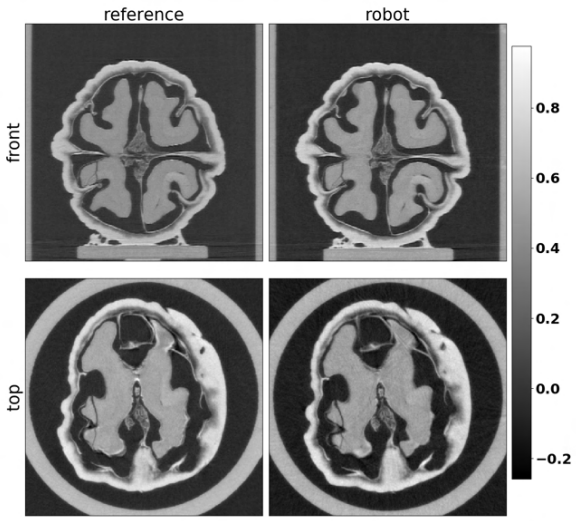


Fig. 4. Experimental results. A walnut was measured and reconstructed in order to compare the conventional rotational stage (*reference*) with the robotic sample holder (*robot*). The reconstruction volumes were registered and aligned but small differences are still visible. The detector images were binned with $4 * 4$ and the reconstruction volume has dimensions 720^3. The front slice is from the perspective of the x-ray source. The top slice is from the bird's eye view. Our observation is that the reconstruction quality is identical despite the fact that the volumes are not aligned perfectly and hence the contrast does not match.

Finally, the accuracy of the calibration algorithm could be improved by improving the circle detection algorithm that is run on the acquired images. Currently, we are using the circle Hough transform algorithm which could be replaced by a more precise algorithm with sub-pixel segmentation accuracy.

III. Conclusion

In this work we have demonstrated a geometric calibration method for the use of a seven degrees of freedom robot as a sample holder for X-ray computed tomography. Our findings have confirmed that this kind of robot can be used for computed tomography with consistent results when compared to more conventional sample holders. A suitably sized sample holder with a geometric structure that can be used for calibration must be provided.

Acknowledgments

We acknowledge financial support through the Center for Advanced Laser Applications (CALA).

References

[1] M. Wieczorek, F. Schaff *et al.*, "Anisotropic x-ray dark-field tomography: A continuous model and its discretization," *Physical review letters*, vol. 117, no. 15, p. 158101, 2016.

[2] M. Wieczorek, F. Schaff *et al.*, "Brain connectivity exposed by anisotropic x-ray dark-field tomography," *Scientific reports*, vol. 8, no. 1, pp. 1–6, 2018.

[3] F. EMIKA. (2020) Panda datasheet. [Online]. Available: https://s3-eu-central-1.amazonaws.com/franka-de-uploads/uploads/2019/04/Datasheet.pdf

[4] E. Pekel, "Robotic sample holder," https://gitlab.lrz.de/IP/robotic-sample-holder/robotic-sample-holder, 2021.

[5] D. H. Ballard, "Generalizing the Hough transform to detect arbitrary shapes," *Pattern Recognition*, vol. 13, no. 2, pp. 111–122, 1981. [Online]. Available: https://www.sciencedirect.com/science/article/pii/0031320381900091

[6] T. Lasser, M. Hornung, and D. Frank, "elsa - an elegant framework for tomographic reconstruction," in *15th International Meeting on Fully Three-Dimensional Image Reconstruction in Radiology and Nuclear Medicine*, S. Matej and S. D. Metzler, Eds., vol. 11072, International Society for Optics and Photonics. SPIE, 2019, pp. 570 – 573. [Online]. Available: https://doi.org/10.1117/12.2534833

[7] M. Quigley, K. Conley *et al.*, "Ros: an open-source robot operating system," in *ICRA workshop on open source software*, vol. 3, no. 3.2. Kobe, Japan, 2009, p. 5.

[8] D. Coleman, I. Sucan *et al.*, "Reducing the barrier to entry of complex robotic software: a moveit! case study," 2014.

[9] I. A. Sucan and S. Chitta, "MoveIt: [online]," http://moveit.ros.org/.

[10] E. FRANKA, "franka ros," https://github.com/frankaemika/franka_ros, 2021.

[11] G. Bradski, "The OpenCV Library," *Dr. Dobb's Journal of Software Tools*, 2000.

[12] L. Dagum and R. Menon, "Openmp: an industry standard api for shared-memory programming," 1998.

[13] P. Virtanen, R. Gommers *et al.*, "SciPy 1.0: Fundamental Algorithms for Scientific Computing in Python," *Nature Methods*, vol. 17, pp. 261–272, 2020.

A generalized total-variation-based image reconstruction method for limited-angle computed tomography

Xin Lu, Yunsong Zhao, and Peng Zhang

***Abstract*—Image reconstruction for limited-angle CT is challenging, because the acquired data are not complete. In order to improve the quality of the reconstructed images, this paper proposes a generalized total-variation (GTV) -based regularization model that multiplies the gradient operator ∇ by an adaptive weighting matrix T in the ℓ^1-norm regularization term. The weighting matrix T in GTV is related to the corresponding image gradient. So different weights are applied to the regularization term at each pixel. An alternating minimization type algorithm is derived to solve the proposed model. Numerical experiments on both simulated data and real data verify the effectiveness of the proposed imaging model and the solving algorithm. The experimental results show that the method can effectively eliminate image artifacts and blur caused by incomplete data. Image edges are accurately restored.**

***Index Terms*—limited-angle CT, image reconstruction algorithm, generalized TV, optimization model**

I. Introduction

In some practical applications, such as breast CT imaging, plate-shaped object detection, etc., due to the special shape of the scanned object and the influence of the scanning environment, it may be difficult to acquire full-angle projection data, instead, only limited-angle projection data can be acquired. Reconstructing images from such limited-angle projection data is called limited-angle CT problem.

The limited-angle CT problem is highly ill-posed because the projection data do not satisfy the data completeness condition [1]. Generally, images reconstructed from such data with traditional reconstruction algorithms, such as FBP [2], ART [3], etc., will have structural artifacts [4], i.e. there are strip artifacts along some specific directions closely related to the scanning angle; while at the same time, the images are blurred along some other directions.

Despite of the difficulty, limited-angle CT imaging has long been a hot research topic in view of its wide application requirements [5], [6], [7]. The ill-posedness as well as the artifact characteristics of limited-angle CT is deeply analyzed. Different kinds of reconstruction algorithms are proposed. Early methods try to restore the complete projection data by some kind of "extrapolation" on the limited-angle data, with smoothness prior or global properties of the projection data [8]. Recent methods try to incorporate various image priors, including geometrical shape, distribution of edges or densities (gray values) about the image, into iterative reconstruction algorithms [9]. With the development of compressed sensing (CS) theory, a series of limited angle reconstruction algorithms based on sparse transform prior and sparse optimization solution are proposed, among which, total variation (TV) regularization method that take the sparsity of image gradients as prior information have received extensive attention [10], and has obtained good reconstruction results in some specific scenarios. But TV regularization considers mainly on the image prior, while it is well known that the artifacts and blurring of the limited-angle CT image are scan angle dependent, so only sub-optimal results can be obtained with classic TV regularization. Later, being aware the dependence between image artifacts and the scan angle, anisotropic total variation (ATV) model is proposed and improves the reconstruction results [11]. In [6], a two-term regularization model, named AEDS, which plays edge-preserving diffusion in the x-direction and edge-preserving smoothing in the y-direction respectively, are proposed. In addition, a method based on similar ideas is also proposed for 3D computed laminography by Zhao et al [12]. Recently, a new state-of-the-art model, named DTV, is proposed by Zhang et al [13]. The model is convex, and therefore has better mathematical properties. In the paper, the authors also show satisfactory reconstruction results from projection data of a very small angular range.

In order to further improve the image reconstruction quality for limited-angle CT, in this paper, we propose a generalized total-variation (GTV) -based regularization model that multiplies the gradient operator ∇ by an adaptive weighting matrix T. The weighting matrix T in GTV is related to the corresponding image gradient. The main difference between the proposed model and that of [11] is adaptivity. The regularizer in [11] is a weighted summation of several directional derivative's ℓ^1-norms, where the weighting factors are constants selected empirically. While the regularizer in our proposed model is a weighted norm of the image gradients, in which the weighting factors are related to the corresponding image gradients and update with iterations. An alternating minimization algorithm is proposed to solve the model. Numerical experiments on both simulated data and real data verify the effectiveness of the proposed imaging model and

Xin Lu, Yunsong Zhao, and Peng Zhang are with School of Mathematical Sciences, Capital Normal University, Beijing, 100048, China, and with Beijing Advanced Innovation Center for Imaging Technology, Capital Normal University, Beijing, 100048, China.

E-mail: zhao_yunsong@cnu.edu.cn.

This work is supported in part by the National Natural Science Foundation of China (NSFC) (61871275, 61771324, and 61827809).

The authors are grateful to Beijing Higher Institution Engineering Research Center of Testing and Imaging as well as Beijing Advanced Innovation Center for Imaging Technology for funding this research work.

its reconstruction algorithm. The experimental results show that the proposed model can accurately reconstruct the image edges and obtain high-quality reconstruction results.

II. Materials and methods

A. The discrete imaging model of CT

The CT reconstruction problem can be described as solving a linear system:

$$Af = p \tag{1}$$

where $f = (f_1, \cdots, f_n, \cdots, f_N)^\top$ represents the image to be reconstructed, $p = (p_1, \cdots, p_m, \cdots p_M)^\top$ denotes the projection data, $A = (a_{m,n})$ is the system matrix of size $M \times N$, where M and N denote the total number of projections and image pixels, respectively. Element $a_{m,n}$ of matrix A represents the intersection length of the mth ray and the nth image pixel. Generally, for limited-angle CT imaging, the directions of available rays are constrained in a range which is generally smaller than π. Without loss of generality, we assume that the scanning angular range is symmetric relative to y-axis.

To improve the quality of the reconstructed images, the imaging problem is usually converted to an optimization problem, which can be expressed as:

$$f^* = \underset{f}{\operatorname{argmin}} \left\{ ||Af - p||_2^2 + \lambda R(f) \right\} \tag{2}$$

where $||Af - p||_2^2$ is the data fidelity term, measuring the difference between measured data p and the model data Af, $R(f)$ is the regularization term modeling some prior information, λ is the regularization parameter, f^* denotes the desired image to be reconstructed.

B. The proposed optimization model

The proposed model can be regarded as an extension of our AEDS model, which contains a two-term regularizer performing edge-preserving diffusion in the x-direction and edge-preserving smoothing in the y-direction respectively. The AEDS mode can achieve satisfactory results when the edges of the image are parallel to the horizontal or vertical directions. But it may fail to restore tilt edges as it focus mainly on the horizontal and vertical regularizations. In order to further improve the reconstruction quality for limited-angle CT, we propose our GTV model as follows:

$$f^* = \underset{f}{\operatorname{argmin}} \left\{ ||Af - p||_2^2 + \frac{\lambda}{2} ||T\nabla f||_{2,1} \right\} \tag{3}$$

where ∇ is the gradient operator and T is a weighting matrix can be expressed as:

$$T = \begin{bmatrix} \dfrac{1}{1 + \kappa |G_\sigma * \nabla_x f|} & 0 \\ 0 & \dfrac{1}{1 + \kappa |G_\sigma * \nabla_y f|} \end{bmatrix},$$

where κ is the tuning parameter, G_σ denotes the Gaussian convolution with the standard deviation σ in order to reduce the effect of the noise, $||\cdot||_2$ and $||\cdot||_1$ respectively denote the ℓ^2-norm and the ℓ^1-norm of a vector.

In (3), T controls the penalization of image gradient. As T is pixel-related, the strength of penalization on different pixels are generally different. Edge pixels with large gradients are penalized less, while non-edge pixels with small gradients are penalized more strongly.

C. Algorithm for solving the model

The model (3) is non-convex, so that we can only expect to get local minimizers. In our practice, we split the problem into two sub-problems which are solved in alternating iterations. Let $f^{(k)}$ denote the solution after k iterations, then the two sub-problems for obtaining $f^{(k+1)}$ can be expressed as:
Sub-problem 1:

$$f^{(k+1/2)} = \underset{f}{\operatorname{argmin}} \left\{ ||f^{(k)} - f||_2^2 + \frac{\lambda}{2} ||T\nabla f||_{2,1} \right\}.$$

Sub-problem 2:

$$f^{(k+1)} = \underset{f}{\operatorname{argmin}} \left\{ ||f^{(k+1/2)} - f||_2^2 + ||Af - p||_2^2 \right\}.$$

The alternating direction method of multipliers (ADMM) is used to solve sub-problem 1, detailed implementation can be found in [14]. For simplicity, the SART method is use to solve sub-problem 2, where $f^{(k+1/2)}$ is used as the initial value, and only one iteration is performed in our implementation.

III. Experiments and results

In this section, we perform numerical experiments with both simulated data and real data to validate the proposed imaging model and the corresponding image reconstruction algorithm for limited-angle CT. For simplicity, the experiments are restricted to the 2D fan-beam case.

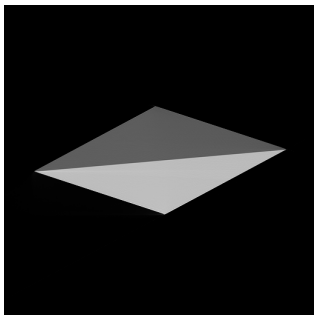


Fig. 1. The rhombus phantom used for simulation.

A. Simulated data experiments

1) Experiment setup: The phantom used in the simulation experiments is a rhombus, as is shown in Fig. 1. The rhombus is tilted to mimic an unideal placement. In addition, the upper and lower triangles of the rhombus are given different gray values. Note that it is a very challenge phantom, as the information of the common edge of the two triangles is lost almost completely from the limited-angle CT scan. To avoid the "inverse crime" problem, in our experiments, noise-free projection data of the phantom are simulated analytically. The parameters for the simulation are listed in table I.

TABLE I
GEOMETRIC PARAMETERS FOR SIMULATED DATA EXPERIMENTS.

Parameter	Value
Scanning angle range	$[\pi/6, 5\pi/6]$
Scanning angular interval	0.5 degree
Number of detector units	1024
Width of detector unit	0.127 mm
Distance of x-ray source to rotation center	311.49 mm
Distance of rotation center to detector	697.9082 mm

The AEDS is used as the comparison method, as it is a relatively new method, which is also proposed by us, so it is easier for us to give an objective and unbiased comparison.

2) Reconstruction results: Fig. 2 shows the reconstruction results with SART (the first row), AEDS (the second row) and the proposed method (the third row) from 120-degree projection data, where the left column is the reconstructed images and the second column is the residual images. The iterations for the three methods are 10, 2000, and 2000 respectively. The display window is set to [0, 3]. It can be seen from the image reconstructed with the SART method that the four edges of the rhombus are blurred to varying degrees as they are of different tilt angles, moreover the common edge between the two triangles are totally lost. The AEDS method improves the reconstruction result in some degree, but the edge are not restored accurately, especially for the common edge between the two triangles. In the contrary, the image reconstructed with the proposed method (Fig. 2(e)) restores all the edges of the rhombus correctly.

B. Real data experiments

1) Experiment setup: Real data is scanned with an industrial CT system developed by our laboratory, which is equipped with an YXLON-FXE-225.48 x-ray source and a Varian PS2520V detector. The scanning parameters are listed in table II. The projection data corresponding the central slice is extracted to mimic a fan-beam scan for the experiment.

TABLE II
SCANNING PARAMETERS FOR ACQUIRING THE REAL DATA .

Parameter	Value
Voltage	140 kV
Current	100 mA
Scanning angle range	$[0, 2\pi]$
Angle range for limited-angle reconstruciton	$[\pi/6, 5\pi/6]$
Scanning Angular Interval	0.5 degree
Number of detector units	1920
Width of detector unit	0.127 mm
Distance of X-ray source to rotation center	311.49 mm
Distance of rotation center to detector	697.9082 mm
Size of reconstruction image	512×512
Pixel size of the object	0.127×0.127 mm^2

Fig. 3(a) shows a photograph of the scanned phantom, which consists of a cylinder and two triangular prisms with different shapes. So the cross section contains a disk and two triangles. Full-angle projection data are acquired, while projection data corresponding to the angle range $[\pi/6, 5\pi/6]$ is used for limited-angle CT reconstruction.

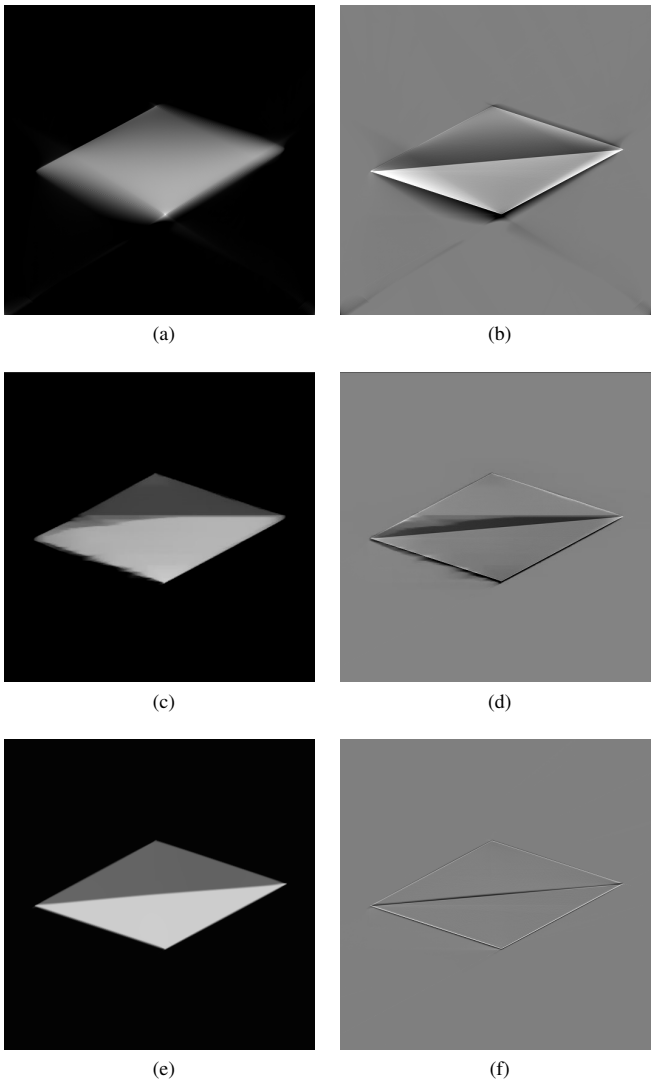
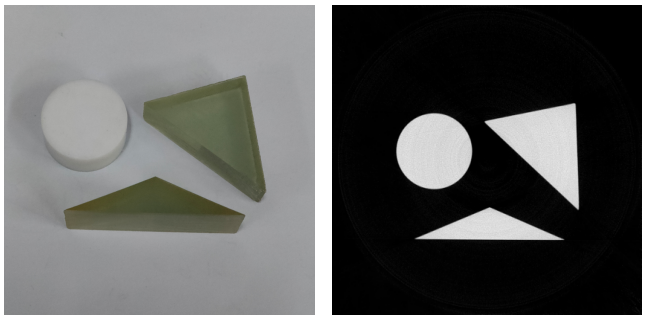


Fig. 2. The reconstruction results of the SART method (the first row), the AEDS methd (the second row), and the proposed mehtod (the last row). the left column is the reconstructed images and the right column is the residual images.

2) Reconstruction results: Fig. 3(b) shows the image reconstructed with SART (10 iterations) from full-angle projection data, which is also used as the reference image.

Fig. 4 shows the reconstruction results with SART (the first row), AEDS (the second row) and the proposed method (the third row) from 120-degree real projection data, where the left column is the reconstructed images and the right column is the residual images. The iterations for the three methods are 10, 2000, and 2000 respectively. The display window for the left and right columns are [0.02, 0.03] and [0, 0.01] respectively.

It can be seen that the image edges are severely blurred, and there are obvious strip artifacts shown in the image reconstructed with the SART method. The AEDS method improves the image quality to some extent. But the upper and lower edges of the disk and the sloping edges of the triangles are not reconstructed correctly, and new artifacts are



(a) (b)

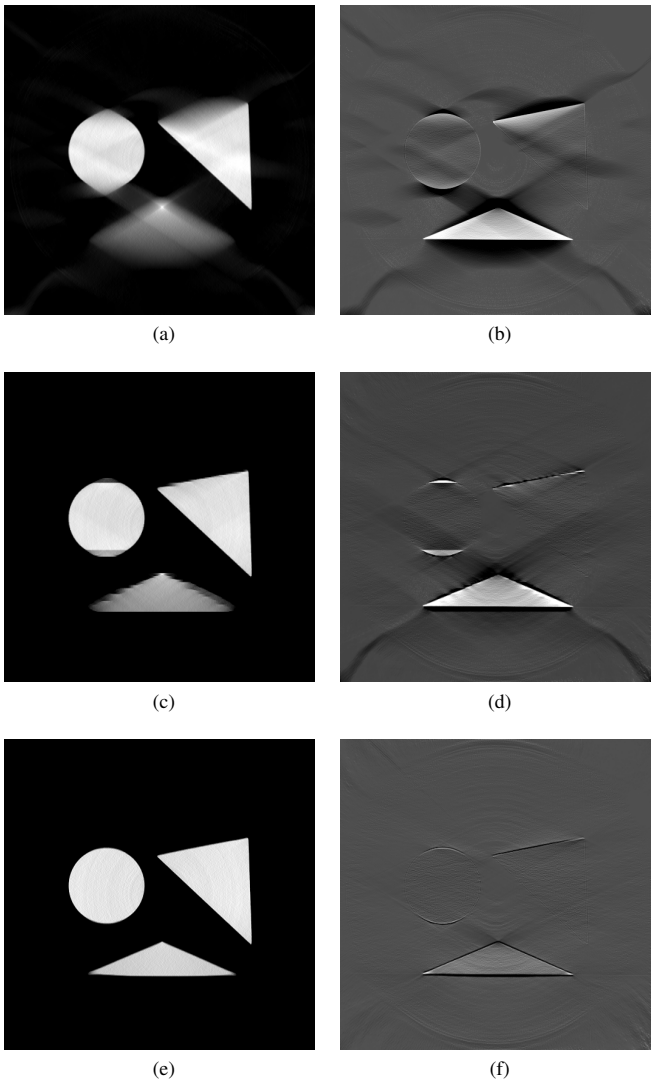
Fig. 3. A photograph of the phantom used for the experiment (a) and the corresponding CT image reconstructed with SART (10 iterations) from full-angle projection data (b). The display window is set to [0.02, 0.03].

introduced. On the contrary, our method correctly reconstructs all the edges of the image, furthermore, the artifacts are elevated effectively.

IV. Summary

A generalized TV-based regularization model is propose for limited-angle CT, in which the penalization of the image gradient for each image pixel is adaptively adjusted by a weighting matrix. Experiments on both simulation data and real data verify its effectiveness. It is shown that the proposed method has better performance on recovering the tile image edges than the AEDS method.

It is meaningful to give a thorough unbiased comparison of our method with DTV model [13]. Preliminary comparisons have been made and detailed results will be given after we confirm with the authors some details of the implementation of DTV.



(a) (b)

(c) (d)

(e) (f)

Fig. 4. The reconstruction results of the SART method (the first row), the AEDS methd (the second row), and the proposed mehtod (the last row) from 120-degree real projection data. the left column is the reconstructed images and the right column is the residual images.

References

[1] H. K. Tuy, "An inversion formula for cone-beam reconstruction," *SIAM Journal on Applied Mathematics*, vol. 43, no. 3, pp. 546–552, 1983.

[2] G. T. Gullberg, "The reconstruction of fan-beam data by filtering the back-projection," *Computer Graphics and Image Processing*, vol. 10, no. 1, pp. 30–47, 1979.

[3] R. Gordon, R. Bender, and G. T. Herman, "Algebraic reconstruction techniques (ART) for three-dimensional electron microscopy and x-ray photography," *Journal of theoretical Biology*, vol. 29, no. 3, pp. 471–481, 1970.

[4] F. Noo, M. Defrise, R. Clackdoyle, and H. Kudo, "Image reconstruction from fan-beam projections on less than a short scan," *Physics in medicine & biology*, vol. 47, no. 14, p. 2525, 2002.

[5] A. K. Louis, "Incomplete data problems in x-ray computerized tomography," *Numerische Mathematik*, vol. 48, no. 3, pp. 251–262, 1986.

[6] J. Xu, Y. Zhao, H. Li, and P. Zhang, "An image reconstruction model regularized by edge-preserving diffusion and smoothing for limited-angle computed tomography," *Inverse Problems*, vol. 35, no. 8, p. 085004, 2019.

[7] X. Xue, S. Zhao, Y. Zhao, and P. Zhang, "Image reconstruction for limited-angle computed tomography with curvature constraint," *Measurement Science and Technology*, vol. 30, no. 12, p. 125401, 2019.

[8] F. Natterer, *The mathematics of computerized tomography*. SIAM, 2001.

[9] N. S. O'Brien, R. P. Boardman, I. Sinclair, and T. Blumensath, "Recent advances in x-ray cone-beam computed laminography.," *Journal of X-ray science and technology*, vol. 24, no. 5, p. 691, 2016.

[10] E. Y. Sidky, C.-M. Kao, and X. Pan, "Accurate image reconstruction from few-views and limited-angle data in divergent-beam CT," *Journal of X-ray Science and Technology*, vol. 14, no. 2, pp. 119–139, 2006.

[11] Z. Chen, X. Jin, L. Li, and G. Wang, "A limited-angle CT reconstruction method based on anisotropic TV minimization," *Physics in Medicine & Biology*, vol. 58, no. 7, p. 2119, 2013.

[12] Y. Zhao, J. Xu, H. Li, and P. Zhang, "Edge information diffusion based reconstruction (EIDR) for cone beam computed laminography," *IEEE transactions on image processing : a publication of the IEEE Signal Processing Society*, 2018.

[13] Z. Zhang, B. Chen, D. Xia, E. Y. Sidky, and X. Pan, "Directional-TV algorithm for image reconstruction from limited-angular-range data," *Medical Image Analysis*, vol. 70, p. 102030, 2021.

[14] Z. Pang, Y. Zhou, T. Wu, and D. Li, "Image denoising via a new anisotropic total-variation-based model," *Signal Processing: Image Communication*, vol. 74, pp. 140–152, 2019.

Comparison of Energy Bin Compression Strategies for Photon Counting Detectors

Yirong Yang, Sen Wang, Debashish Pal, Norbert J. Pelc, Adam S. Wang

***Abstract*—Photon counting detectors (PCDs) with energy discrimination capabilities allow us to perform quantitative material decomposition with high spatial resolution. Although PCDs provide more spectral information than conventional energy integrating detectors (EIDs), it is more challenging for the system to transmit projection data from the detectors across the slip ring to the processing computer and store the data, due to the increased amount of data with increasing number of energy bins. To address this problem, many approaches have been proposed to compress the bin data while maintaining the image quality. In this work, we compare the performance of strategies to reduce projection data and determine the optimal choice of bin compression strategies and the number of measurements for multiple tasks.**

We first obtain the optimal thresholds for conventional energy bins, as determined by minimizing the Cramér–Rao lower bound (CRLB) for material decomposition tasks with a realistic silicon detector energy response. We then consider the case of reducing data from 8 native energy bins by forming weighted sums, either with binary weights or continuous weights, by minimizing the relative CRLB between the compressed measurements and the original 8 bins. We then evaluate their respective performance using Monte Carlo simulation for a head phantom. The results show that the continuous weights strategy is superior to others, with low bias and less than 10% variance penalty for 2 weighted sums, with a data reduction of 75% within a large material thickness space. The other strategies have up to 50% variance penalty compared with the original 8 bins and are less robust when there is photon starvation. With additional weighted measurements, the continuous weights method can achieve less than 1% variance penalty when reducing the 8 native energy bins to half the number of measurements. Overall, combining energy bins by forming weighted sums with continuous weights is an effective strategy for reducing data while preserving spectral information.

***Index Term*—photon counting detector, bin compression, Cramér–Rao lower bound**

I. Introduction

TRANSMITTING projection data from detector arrays on the rotating CT gantry through the slip ring to the data processing computer and storing them have always been challenging for CT systems. Photon counting detectors (PCDs) are advanced detectors that provide more spectral and spatial information than current dual-energy CT systems using energy integrating detectors (EIDs). For example, deep silicon PCDs with 8 native energy bins enable more precise material decomposition quantification by taking advantage of all the spectral information. However, the projection data transmission and storage become more challenging for PCDs due to the increased amount of data. There are several approaches to reducing the projection data by reducing the number of measurements. One direct approach is to reduce the number of native energy bins by setting fewer energy thresholds. Other strategies include splitting the native energy bins into N groups and summing them up based on a preset figure of merit [1]. Instead of transmitting and storing the native energy bins, the summed bins are then used for further processing so that the data is downsampled. A more generalized method is to combine the native energy bins into N measurements with binary weights. Wang *et al* had also proposed a weighting method to compress spectral information from infinite bins and an ideal detector energy response without information loss [2]–[4]. More recently, we proposed a generalized version of the continuous weights strategy for finite native energy bins under realistic non-ideal detector energy response and showed its potential in reducing projection data while maintaining material decomposition and virtual monoenergetic image quality [5].

In this work, we investigate the optimal strategy for bin compression that best preserves spectral information for specific tasks, such as material decomposition and virtual monoenergetic images, with high compression ratio. We compare the performance by assessing bias and noise in projection and image domains between four strategies – conventional bins, summed bins, binary weights, and continuous weights – with different compression ratios.

This work was supported by GE Healthcare.

Yirong Yang (yryangd@stanford.edu) and Adam S. Wang are with the Department of Electrical Engineering, Stanford University.

Yirong Yang, Sen Wang, Norbert J. Pelc, and Adam S. Wang are with the Department of Radiology, Stanford University.

Debashish Pal is with GE Healthcare, Menlo Park.

II. Methods

A. Experiment Settings

We used a published deep silicon detector model with 0.5×0.5 mm^2 pixels and 30 mm thickness to simulate the detected binned counts [6]. We used an incident spectrum of 120 kVp with 0.4 mAs per projection. The energy thresholds of the 8 native energy bins were set to be 4, 14, 30, 37, 47, 58, 67, and 79 keV to maximize the spectral information contained in the native binned counts. Performance of material decomposition and virtual monoenergetic images (VMI) was evaluated for a basis material space that spans 0 to 40 cm water and 0 to 5 cm of calcium in cortical bone. These settings were used throughout our study.

B. Energy Bin Compression Strategies

1) Conventional Bins

We use conventional bins to refer to native energy bins obtained through selection of N energy thresholds. We fix the lowest energy threshold to be 4 keV, which is consistent with our 8 native energy bins. To obtain the optimal remaining thresholds, we used an exhaustive search to find the thresholds that give the lowest average relative Cramér–Rao lower bound (rCRLB) over the material space. The CRLB represents the minimum variance of a task, such as basis material decomposition or VMIs, for an unbiased estimator given the measurements. The rCRLB is defined as the ratio of CRLB between the proposed methods and the 8 native energy bins,

$$\mathrm{rCRLB}=\frac{\mathrm{CRLB_W}}{\mathrm{CRLB_8}}, \quad (1)$$

where we use subscript W to represent the proposed bin reduction method and the subscript 8 to represent the native 8 bins. Because each method reduces the amount of information, $\mathrm{rCRLB} \geq 1$, but should be minimized to reduce the increased variance. An average rCRLB for the tasks of material decomposition and VMI (at 60 keV) across a range of M material thicknesses can be expressed as:

$$\overline{\mathrm{rCRLB}}=\frac{1}{3M}\sum_A \frac{\mathrm{CRLB_W(Ca)}}{\mathrm{CRLB_8(Ca)}}\Big|_A + \frac{\mathrm{CRLB_W(H_2O)}}{\mathrm{CRLB_8(H_2O)}}\Big|_A + \frac{\mathrm{CRLB_W(VMI)}}{\mathrm{CRLB_8(VMI)}}\Big|_A, \quad (2)$$

where A indexes the material thickness pairs.

2) Summed Bins

In this approach, we assume 8 native energy bins are acquired, and they each contribute once and only once to a summed bin. We used the same iterative method described in [1] to compress the 8 bins to different numbers of summed bins. In each iteration, the best combination of 2 bins is summed, using the average rCRLB as expressed in Eq. (2) as the figure of merit. The process and the results are shown in Figure 1.

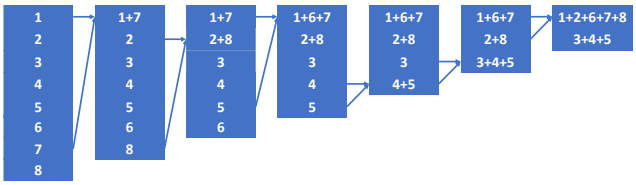


Figure 1: Optimization process of summed bins method.

Mathematically, we can use $W_{i,j}$ to represent N sets of weights, such that the jth summed bin can be written as $b_{W_j} = \sum_i W_{i,j} b_i$, where b_i represents the original binned counts. For summed bins, we have $W_{i,j} \in \{0, 1\}$, and $\sum_j W_{i,j} = 1$ so that each native bin contributes once and only once to a summed bin.

3) Binary Weights

A more generalized bin compression strategy is to use binary weights, where the elements of energy weights matrix $W_{i,j} \in \{0, 1\}$, but no longer has the constraint $\sum_j W_{i,j} = 1$.

To obtain $W_{i,j}$, we start with the energy weight matrix of summed bins $W^{\mathrm{S}} = [\boldsymbol{w_1}, \dots, \boldsymbol{w_8}]^T$, which compressed the 8 native energy bins to N measurements, and update it row by row (bin by bin) to select the best binary combinations. The pseudocode is listed below.

$i \leftarrow 0$
$W = [\boldsymbol{w_1}, \dots, \boldsymbol{w_8}]^T = W^{\mathrm{S}}$
while $i < 8$ **do**
  $i \leftarrow i + 1$
  $y_{i,j} \in \{0,1\}, \sum_j y_{i,j} > 1$
  for all possible $\boldsymbol{y_i}$ **do**
    if $Y = [\boldsymbol{w_1}, \dots, \boldsymbol{y_i}, \dots, \boldsymbol{w_8}]^T$ is full rank **then**
      Compute $\overline{\mathrm{rCRLB}}(\mathbf{b}_Y = Y^T\mathbf{b}, \mathbf{b}_W = W^T\mathbf{b})$
      if $\overline{\mathrm{rCRLB}} < 1$ **then**
        $W \leftarrow Y$

4) Continuous Weights

For continuous weights, we fully generalize the elements of the bin weights matrix to $W_{i,j} \in \mathbb{R}$. In this case, we use GlobalSearch in MATLAB (R2021a) to minimize the objective function described in Eq. (2).

III. Results

The optimal binary weights and continuous weights are presented in Figure for 2 measurements. The binary weights are identical to summed bins, except bin 5 contributes to both measurements, which was found to slightly improve performance. The continuous weights are normalized to 1 and show that the optimal solution leverages the flexibility of real-valued weights, including negative weights.

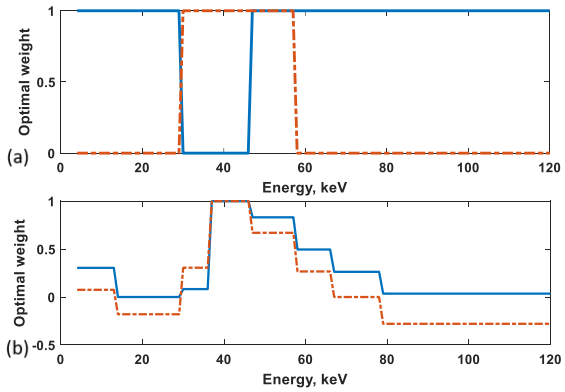


Figure 2: (a) Optimal binary weights and (b) optimal continuous weights for 2 measurements.

A. rCRLB Vs Compressed Measurements

We first show our analytical results of the average rCRLB, defined in Eq. (1), over the basis material space of 0 to 32 cm water and 0 to 4 cm of calcium in cortical bone obtained from the four compression methods for different numbers of measurements (Figure 3).

The results show that with increasing number of measurements, the noise performance improves for all four compression strategies. Of all strategies, the continuous weights method is superior to the other three methods, while

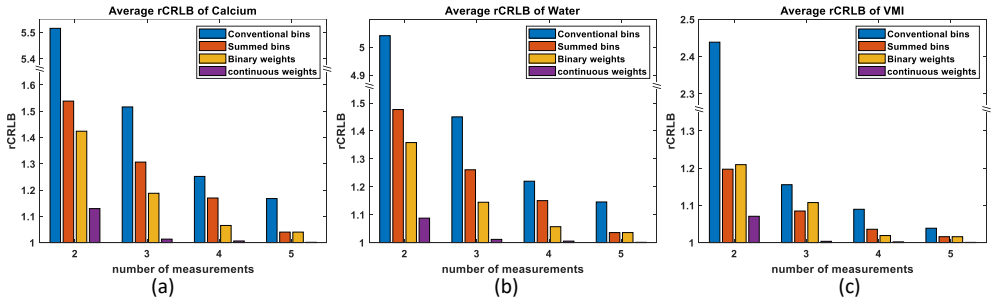


Figure 3: Average rCRLB of conventional bins, summed bins, binary weights, and continuous weights of (a) calcium and (b) water material decomposition and (c)VMI at 60 keV, with different number of measurements.

conventional bins perform the worst. The average rCRLB of 2 measurements generated from continuous weights (1.1294 and 1.0876 for calcium and water decomposition in projection domain) is less than that of 5 measurements from conventional bins (1.1675 and 1.1455 for calcium and water decomposition in projection domain), which indicates that with proper bin weighting combinations, spectral information from more native bins can be preserved better than simply reducing the number of native bins. Data reduction using summed bins and binary weights has less variance than conventional bins, but is substantially higher than continuous weights.

When we reach 4 measurements, the rCRLB obtained from continuous weights is 1.0061 for calcium decomposition, 1.0053 for water decomposition, and 1.0019 for VMI at 60 keV in the projection domain. The noise penalty of this strategy compared with the original data is less than 1% for all the tasks, while the amount of data is only half of the original. With continuous weights, the data from 4 compressed measurements performs essentially as well as the 8 native bins, which indicates that 4 measurements from continuous weights are sufficient to recover the information of 8 native energy bins across the full range of object sizes.

B. Projection Domain Monte Carlo Simulation

Monte Carlo simulation was conducted to validate the analytical results, using maximum likelihood estimation to perform material decomposition from the compressed measurements. 10^4 realizations were performed at each sample point of water and calcium thickness. The relative variance and bias of calcium thickness estimates are presented in Figures 4 and 5, respectively, for compressing to 2 measurements. The results of water thickness and VMI estimates are similar.

The variance penalty is defined as $\frac{\mathrm{Var_W}}{\mathrm{Var_8}} - 1$, where $\frac{\mathrm{Var_W}}{\mathrm{Var_8}}$ is the ratio between variances of basis material thicknesses estimated from the compressed data and the 8 native bins. We found that the variance penalties of 2 measurements with continuous weights for calcium decomposition in projection domain is less than 10% for most material thicknesses, with an average variance penalty of 5.55% in the central region, while that of the conventional bins is 499.67%, 55.14% for summed bins, and 54.29% for binary weights. The binary weights have lower variance penalty than summed bins for intermediate thicknesses, but higher variance penalty at greater thicknesses.

From Figure 5, we also observe that the bias of calcium material decomposition is small for compressing to 2 measurements with continuous weights, even for thicker material combinations. However, the other three strategies suffer increased bias when there are relatively few photons for thicker objects. The root mean squared deviation (RMSD) of

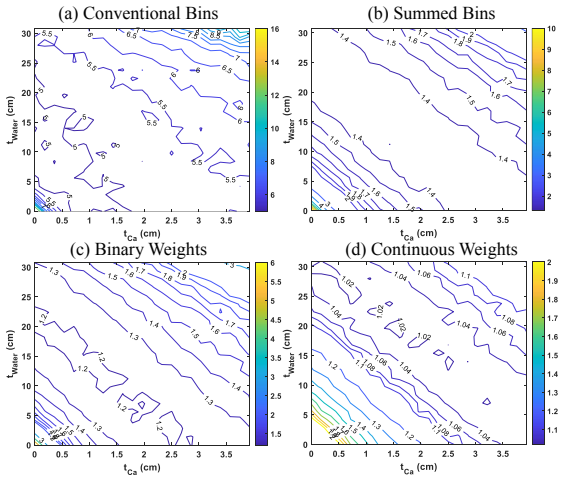


Figure 4: Relative variance between the bin compression strategies and 8 native bins for calcium decomposition.

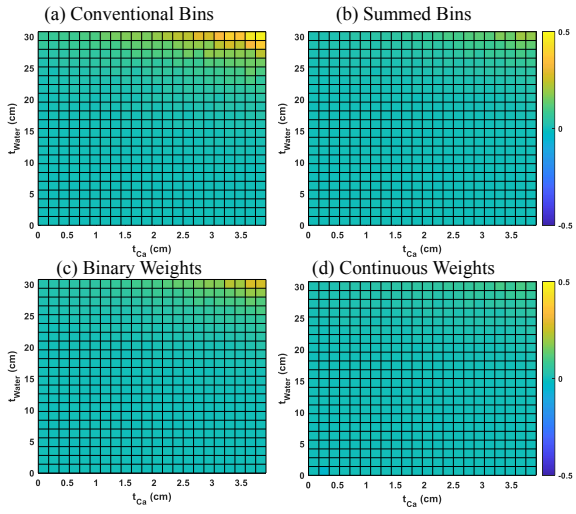


Figure 5: Bias of bin compression strategies for calcium decomposition.

the thickness bias is 0.0246 cm for continuous weights, which is a small increase from that of the 8 native energy bins (0.0153 cm). However, for summed bins, binary weights, and

conventional bins, the number increases to 0.0457, 0.0640, and 0.1095 cm, respectively.

C. Brain Phantom

We also applied these four strategies on a simulated brain phantom, and the resulting of water decomposition images for 2 measurements are shown in Figure 6. The average variance penalty of a region with uniform water is 352.35% for conventional bins, 33.54% for summed bins, 12.98% for binary weights, and 5.05% for continuous weights.

The increased noise due to bin compression is visibly obvious when we use conventional bins and summed bins. The image quality is the best when we use continuous weights and is comparable to that of 8 native bins (not shown). In addition, there is no structural bias observed between the images from compressed data and 8 native bins.

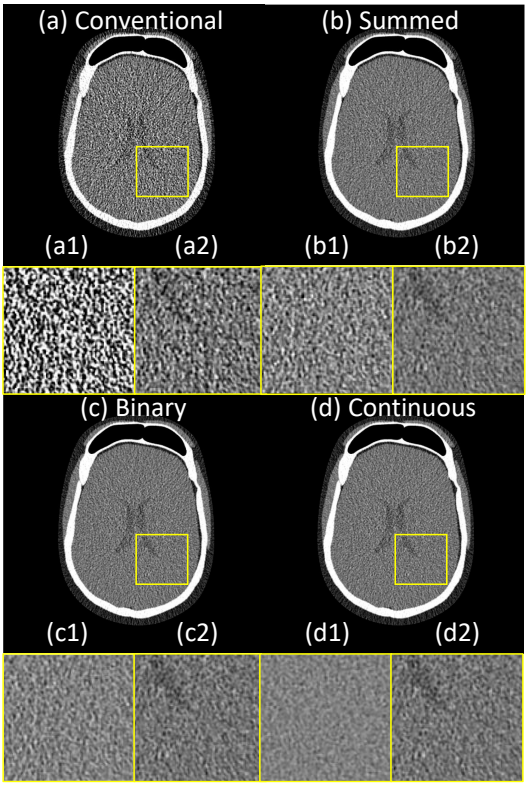


Figure 6: Phantom simulation results of water decomposition using data from two (a) conventional bins, (b) summed bins, (c) binary weights, and (d) continuous weights. (a2) - (d2) are the corresponding enlarged ROI. The display window is [0.9, 1.2] g/cc. The difference of the water decomposition between the proposed strategies and the 8 native bins is shown in (a1) - (d1), with display window [-0.05, 0.05] g/cc, showing the increased noise but no bias.

IV. Discussion

We have compared the bias and noise performance of 4 different energy bin compression methods in both projection and image domains. The results show that using continuous weights, the compressed 2 measurements can preserve almost as much spectral information as the original 8 native energy bins. It enables a 75% data reduction while preserving the native spectral information, with only a small increase of less than 10% in image variance over a large range of material thicknesses. Other strategies, such as conventional bins, summed bins, or binary weights, suffer from at least 50% increase in image variance when using 2 measurements. With 4 measurements using continuous weights, there is almost no image quality degradation.

The Monte Carlo simulation showed consistent results with the analytical CRLB predictions of noise performance. The bias of calcium thickness estimates is visibly higher at thicker material combinations when we use two conventional bins, summed bins, or binary weights, while the continuous weights and 8 native bins do not have this problem. From this observation, we infer that due to the lack of sufficient spectral information, if we use data from conventional bins, summed bins, and binary weights, the maximum likelihood estimator is no longer asymptotically unbiased when the detector collects fewer photon counts. We also conducted the same Monte Carlo simulation with 1 mAs per projection (not shown), which resulted in a smaller bias. This observation shows that the continuous weights method is more robust than the other three strategies in not only noise but also bias performance. It also indicates that extra calibration and higher exposures may be needed for large patients when only using 2 conventional bins. The silicon detector energy response is complex, with some high energy photons recorded in low energy bins due to Compton scattering. The 8 native bins and continuous weights with as few as 2 measurements contain the spectral information needed to account for the detector response, while other compression methods may struggle.

V. Conclusion

We found that compressing the binned data with continuous weights can best preserve the spectral information for material decomposition and virtual monoenergetic imaging from the original data of 8 native energy bins. Other bin compression strategies substantially increase image noise. With continuous weights, a compression ratio of 4 can be achieved with 2 measurements that mostly preserves the image quality of both basis materials and VMIs.

References

[1] Taly Gilat Schmidt *et al.*, "Energy-bin downsampling method for grayscale image reconstruction for a deep silicon photon-counting CT clinical prototype," presented at the RSNA Annual Meeting, 2021.

[2] A. S. Wang and N. J. Pelc, "Optimal energy thresholds and weights for separating materials using photon counting x-ray detectors with energy discriminating capabilities," in *SPIE Medical Imaging 2009: Physics of Medical Imaging*, Mar. 2009, vol. 7258, p. 725821.

[3] A. S. Wang and N. J. Pelc, "Sufficient Statistics as a Generalization of Binning in Spectral X-ray Imaging," *IEEE Transactions on Medical Imaging*, vol. 30, no. 1, pp. 84–93, Jan. 2011.

[4] A. S. Wang and N. J. Pelc, "Impact of photon counting detector spectral response on dual energy techniques," in *Medical Imaging 2010: Physics of Medical Imaging*, Mar. 2010, vol. 7622, p. 76223L.

[5] Yirong Yang, Sen Wang, Debashish Pal, Norbert J. Pelc, and Adam S. Wang, "Empirical Optimization of Energy Bin Weights for Compressing Measurements with Photon Counting X-ray Detectors," presented at SPIE Medical Imaging, 2022.

[6] M. Persson, A. Wang, and N. J. Pelc, "Detective quantum efficiency of photon-counting CdTe and Si detectors for computed tomography: a simulation study," *JMI*, vol. 7, no. 4, p. 043501, Jul. 2020.

A visible edge aware directional total variation model for limited-angle reconstruction

Yinghui Zhang, Ke Chen, Xing Zhao, Hongwei Li*.

***Abstract*—The directional total variation algorithm (DTV) reported in the literature achieves some promising results for limited angle reconstructions, especially when the scanning angular range is very small. However, the visible edge prior for limited-angle CT is not explicitly considered by DTV. In this paper, a variant of the DTV model is proposed which explicitly builds into the visible edge prior developed by Quinto et al. Numerical experiments show that the proposed model and algorithm produce very competitive results compared to DTV.**

***Index Terms*—limited-angle CT, directional total variation, iteration reconstruction.**

I. Introduction

In certain computed tomography (CT) applications, due to the restrictions on the scanning condition or the geometrical shapes of scanning objects, the projection data could be only acquired in a limited angular range which leads to the challenging limited-angle reconstruction problem. This happens in both medical diagnosis like breast imaging [7] and industrial inspection like the C-arm neuro imaging [4].

Conventional reconstruction methods like filtered back-projection (FBP) and (simultaneous) algebraic reconstruction technique ((S)ART) perform poorly with limited-angle data, introducing heavy image blurring along the directions perpendicular to the missing projection rays. The limited-angle reconstruction problem has been extensively studied for decades, including theoretical characterization and practical reconstruction algorithms.

An early method views it as a projection domain inpainting problem [5] and incorporate the smoothness prior of projection data into the reconstruction process. However, since a local extrapolation error in the projection domain may cause global artifacts in the image domain, this kind of methods suffers from severe stability issue. Another method is based on optimization models encoded with various hand-crafted priors. In certain applications, CT images can be approximated well by piecewise constant functions which should possess the gradient sparsity property. This property can be encoded by the total variation (TV) regularizer, extensively used in image processing. The first method adopting TV regularization is introduced in [10] for divergent CT reconstruction. Since then, various modifications and improvements have been proposed, including the adaptive steepest descent-projection onto a convex set method [11], prior image constrained compressed sensing method [2], adaptive-weighted TV model [6], TV-l_0 gradient minimization [14], etc. These methods could effectively improve the reconstruction quality and achieve promising results.

For limited-angle reconstruction, there is a vital prior described by the theory of visible and invisible boundaries [8] developed by Quinto et al. This prior is first considered by the anisotropic total variation (ATV) method [3]. Later on, the reweighted ATV [12] method takes projection directions as prior information and combines them into the TV formulation. Especially, the alternating edge-reserving diffusion and smoothing (AEDS) algorithm [13] takes the visible edges prior to its full advantage. By designing separated x- and y-direction regularizers, the AEDS model encodes explicitly the visible edges prior, and by adopting the alternating minimization technique, the AEDS algorithm decouples the x-direction regularization from the y-direction regularization such that the visible edges are protected and utilized to their full advantages.

A very recent algorithm named DTV (directional total variation) is proposed in [15], which shows very promising reconstructions, especially for very small scanning angular ranges. The energy functional associated with DTV can be seen as a reformulation or constrained version of the AEDS model when the regularizers are specified by x-direction TV (TV_x) and y-direction TV (TV_y). The workhorse of DTV is the primal-dual based Chambolle-Pock (CP) algorithm.

Motivated by the success of the DTV algorithm, we propose to reformulate the DTV model such that the new model treats the visible edges prior (corresponding to TV_x) differently from TV_y. This is achieved by exchanging the roles of TV_x and the data fidelity terms. Since the TV_x term goes into the energy functional while the TV_y term is specified as a constraint, the new model shall treat them differently. In this way, we think that the visible edges prior could be better utilized.

The remainder of this paper is organized as follows. We present our approach to the limited-angle CT reconstruction problem in Section II. In Section III, experiments are carried out to validate the proposed method. Finally, we conclude the paper in Section IV.

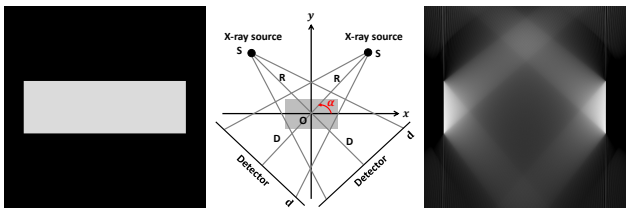
II. Method

A. The limited-angle CT reconstruction problem

Assume that the size of reconstruction image u is $M \times N$. The vector $\overrightarrow{u} \in \mathbb{R}^J, J = M \times N$ is a concatenate form along the columns of u, and u_i describes the ith entry of

Ke Chen is with the Department of Mathematical Sciences, University of Liverpool, L693BX, United Kingdom. Yinghui Zhang, Xing Zhao and Hongwei Li are with the School of Mathematical Sciences, Capital Normal University, Beijing, 100048, China.

E-mail: hongwei.li91@cnu.edu.cn.



(a) Phantom (b) Configuration (c) FBP

Fig. 1: Illustration of the theory of visible and invisible boundaries. (a) The rectangle phantom; (b) the scanning configuration; (c) the limited-angle reconstruction for the scanning angular range ($[\frac{\pi}{4}, \frac{3\pi}{4}]$).

$\overrightarrow{u}, i = 1, 2, \ldots, J$. The CT reconstruction problem could be formulated as solving a linear system

$$A\overrightarrow{u} = \overrightarrow{b} + \overrightarrow{e}, \tag{1}$$

where $A \in \mathbb{R}^{I\times J}$ is the system matrix and $\overrightarrow{b}$ is a vector of length $I = V \times D$ which represents the acquired projection data. V and D denote the number of projection views and the number of detector cells, respectively. $\overrightarrow{e}$ accounts for any measurement bias and additive noise.

For limited-angle data, the linear system (1) with $I \ll J$ is severely ill-posed, therefore, images reconstructed by conventional reconstruction algorithms will introduce streak and blurring artifacts. This is demonstrated in Fig.(1). Without loss of generality, here, we consider the fan-beam scanning with limited-angle range ($[\frac{\pi}{4}, \frac{3\pi}{4}]$). The rectangle phantom and the scanning configuration are shown in Fig.(1a) and Fig.(1b), respectively. Under this scanning configuration, according to the theory of visible and invisible boundaries, the edges close to vertical are visible and can be easily reconstructed, while for the edges close to be horizontal will be invisible and cannot be recovered well by conventional reconstruction algorithms like FBP or SART, as shown in Fig.(1c).

B. The DTV model

The DTV algorithm [15] is to solve the following minimization problem

$$\begin{aligned} &\overrightarrow{u}^\star = \min_u \frac{1}{2}\|A\overrightarrow{u} - \overrightarrow{b}\|_2^2 \\ &s.t.\, \|\nabla_x \overrightarrow{u}\|_1 \le t_x,\ \|\nabla_y \overrightarrow{u}\|_1 \le t_y,\ \overrightarrow{u} \ge 0, \end{aligned} \tag{2}$$

where ∇_x and ∇_y represent matrices size of $J\times J$, corresponding to the discrete x-direction and y-direction gradient operators, respectively, and t_x and t_y are two scalars, specifying the allowed total variations along the x-direction and y-direction, respectively. Since the model (2) is convex, the CP algorithm could be employed to compute a global minimizer. It should be noted that for limited-angle problems, the matrix A has a very large kernel space, so that the model (2) could possess multiple global minimizers. In this case, different parameterizations or different initializations could lead to different solutions.

C. The visilbe edge aware DTV model (VEA-DTV)

As mentioned earlier, to better utilize the visible edges prior, we reformulate the DTV model (2) as the following one

$$\begin{aligned} &\overrightarrow{u}^\star = \min_u \|\nabla_x \overrightarrow{u}\|_1 \\ &s.t.\, \|A\overrightarrow{u} - \overrightarrow{b}\|_2 \le \epsilon,\ \|\nabla_y \overrightarrow{u}\|_1 \le t_y,\ \overrightarrow{u} \ge 0. \end{aligned} \tag{3}$$

The parameter ϵ controls the noise-level of reconstructed image, which has a clear physical meaning [9]. It's easy to check that

$$\ker(A) \cap (\ker(\nabla_x) \cup \ker(\nabla_y)) = \{0\}, \tag{4}$$

so the proposed VEA-DTV model (3) is theoretically equivalent to the DTV model (2). However, since the two formulations are not the same, when applying the CP algorithm, the resulting solving algorithms would be different. As mentioned earlier, for limited-angle problems, the models (3) and (2) are not strictly convex and since the system matrix A has a large kernel space, each of the two models admits multiple solutions, in which case different algorithms might reach different global minimizers. So, starting from the formulation (3), the CP algorithm might compute a solution different from that of the DTV algorithm. This is also the case when comparing AEDS and DTV. The model AEDS(l_1, l_1) coincides exactly with that of DTV, since condition (4) is met. However, since AEDS and DTV employ different minimization algorithms, their performance could be different. In fact, the alternating minimization algorithm adopted by AEDS takes constant step-sizes, according to the framework of incremental methods [1], it only converges to a neighbourhood of the optimum. On the other hand, the CP algorithm can be proved to converges to a saddle point corresponding to a optimum.

D. Numerical algorithm

The CP algorithm is adapted to develop an iterative algorithm for solving (3) by

$$\begin{aligned} \overrightarrow{u}^\star = \min_u\ & \nu_1\|\nabla_x \overrightarrow{u}\|_1 + \delta_{\mathrm{Ball}(\epsilon)}(A\overrightarrow{u} - \overrightarrow{b}) \\ & + \delta_{\mathrm{Diamond}(\nu_2 t_y)}(\nu_2|\nabla_y \overrightarrow{u}|_1) + \delta_{\mathrm{P}}(\mu\overrightarrow{u}), \end{aligned} \tag{5}$$

where indicator functions $\delta_{\mathrm{Ball}}(\overrightarrow{x}), \delta_P(\overrightarrow{x}), \delta_{\mathrm{Diamond}(a)}(\overrightarrow{x})$ are defined as:

$$\delta_{\mathrm{Ball}(\epsilon)}(\overrightarrow{x}) = \begin{cases} 0, & \|\overrightarrow{x}\|_2 \le \epsilon \\ \infty, & \|\overrightarrow{x}\|_2 > \epsilon \end{cases}, \delta_{\mathrm{P}}(\overrightarrow{x}) = \begin{cases} 0, & \overrightarrow{x} \ge 0 \\ \infty, & \overrightarrow{x} < 0 \end{cases},$$

$$\delta_{\mathrm{Diamond}(a)}(\overrightarrow{x}) = \begin{cases} 0, & \|\overrightarrow{x}\|_1 \le a \\ \infty, & \|\overrightarrow{x}\|_1 > a \end{cases}.$$

Then, the min-max formulation of (5) is given by

$$\begin{aligned} &\left(\overrightarrow{u}^\star, \overrightarrow{w}^\star, \overrightarrow{p}^\star, \overrightarrow{q}^\star, \overrightarrow{t}^\star\right) = \min_{\overrightarrow{u}} \max_{\overrightarrow{w},\overrightarrow{p},\overrightarrow{q},\overrightarrow{t}} \langle \nu_1 \nabla_x \overrightarrow{u}, \overrightarrow{p}\rangle \\ &- \delta_{\mathrm{Box}(1)}(|\overrightarrow{p}|) + \langle A\overrightarrow{u} - \overrightarrow{b}, \overrightarrow{w}\rangle - \epsilon\|\overrightarrow{w}\|_2 + \langle \nu_2 \nabla_y \overrightarrow{u}, \overrightarrow{q}\rangle \\ &- \nu_2 t_y \|(|\overrightarrow{q}|)\|_\infty + \langle \mu\overrightarrow{u}, \overrightarrow{t}\rangle - \delta_{\mathrm{P}}(-\mu\overrightarrow{t}), \end{aligned} \tag{6}$$

where

$$\delta_{\mathrm{Box}(\epsilon)}(\overrightarrow{x}) = \begin{cases} 0, & \|\overrightarrow{x}\|_1 \le \epsilon \\ \infty, & \|\overrightarrow{x}\|_1 > \epsilon \end{cases}.$$

Applying the proximal point algorithm to solve (6) and taking an additional extrapolation step, we thus obtain the VEA-DTV algorithm described in **Algorithm 1**, where $\|\cdot\|_2$ is computed by the power method suggested in [9], $1_J \in \mathbb{R}^J$ denotes the constant vector with all elements set to 1, operator $\mathrm{sgn}(\cdot)$ returns the sign of a real number, and $l_1\mathrm{Ball}_a(\cdot)$ projects a vector onto the l_1 ball with radius of a. The symbol $\mathrm{neg}(\cdot)$ represents the negative thresholding function, i.e. projects any positive elements of its argument to zero.

Algorithm 1 (VEA-DTV)

Input $A, \overrightarrow{b}, \epsilon, t_y, a$

$L \leftarrow \|K\|_2, K = \left(A^\top, \nu_1\nabla_x^\top, \nu_2\nabla_y^\top, \mu I\right)^\top$, $\sigma \leftarrow \frac{1}{aL}$, $\nu_1 \leftarrow \frac{\|A\|_2}{\|\nabla_x\|_2}$, $\nu_2 \leftarrow \frac{\|A\|_2}{\|\nabla_y\|_2}$, $\mu \leftarrow \frac{\|A\|_2}{\|I\|_2}$

$k \leftarrow 0$

Initialize: $\overrightarrow{u}^{(0)}$, $\overrightarrow{w}^{(0)}$, $\overrightarrow{p}^{(0)}$, $\overrightarrow{q}^{(0)}$ and $\overrightarrow{t}^{(0)}$ to zero

$\overrightarrow{u}^{(0)} \leftarrow \overrightarrow{u}^{(0)}$

repeat

$$\overrightarrow{w}'^{(k)} = \overrightarrow{w}^{(k)} + \sigma\left(A\overrightarrow{u}^{(k)} - \overrightarrow{b}\right)$$
$$\overrightarrow{w}^{(k+1)} = \frac{\overrightarrow{w}'^{(k)}}{\|\overrightarrow{w}'^{(k)}\|_2}\max(\|\overrightarrow{w}'^{(k)}\|_2 - \sigma\epsilon, 0)$$

$$\overrightarrow{p}'^{(k)} = \overrightarrow{p}^{(k)} + \sigma\nu_1\nabla_x\overrightarrow{u}^k$$
$$\overrightarrow{p}^{(k+1)} = \frac{\overrightarrow{p}'^{(k)}}{\max(1_I, \overrightarrow{p}'^{(k)})}$$

$$\overrightarrow{q}'^{(k)} = \overrightarrow{q}^{(k)} + \sigma\nu_2\nabla_y\overrightarrow{u}^k$$
$$\overrightarrow{q}^{(k+1)} = \overrightarrow{q}'^{(k)} - \sigma\mathrm{sgn}\left(\overrightarrow{q}'^{(k)}\right) l_1\mathrm{Ball}_{\nu_2 t_y}\left(\frac{|\overrightarrow{q}'^{(k)}|}{\sigma}\right)$$

$$\overrightarrow{t}'^{(k)} = \overrightarrow{t}^{(k)} + \sigma\mu\overrightarrow{u}^k$$
$$\overrightarrow{t}^{(k+1)} = \mathrm{neg}(\overrightarrow{t}'^{(k)})$$

$$\overrightarrow{u}^{(k+1)} = \overrightarrow{u}^{(k)} - \tau(A^\top\overrightarrow{w}^{(k+1)} + \nu_1\nabla_x^\top\overrightarrow{p}^{(k+1)} + \nu_2\nabla_y^\top\overrightarrow{q}^{(k+1)} + \mu\overrightarrow{t}^{(k+1)})$$

$$\overrightarrow{u}^{(k+1)} = 2\overrightarrow{u}^{(k+1)} - \overrightarrow{u}^{(k)}$$

$k \leftarrow k + 1$

until stopping criterion is met

Output : $\overrightarrow{u}^{(k)}$

III. Experiments

Numerical experiments with simulated data against SART and the DTV algorithm are carried out to validate the effectiveness of the proposed reconstruction algorithm VEA-DTV. The simulated analytic projection data are acquired by the open source software CTSim (http://www.ctsim.org), while the astra toolbox (https://www.astra-toolbox.com/) is utilized to perform the forward and backward projections when they are required.

In terms of parameter selections, since the general CP framwork is adopted, there are totally three parameters subject to tuning, i.e. a, t_y and ϵ for applying VEA-DTV. Correspondingly, there are parameters: a, t_x, t_y, are involved in the DTV algorithm. Ideally, t_x, t_y should be computed in terms of the ideal image. In our experiments, we apply the SART method with 10 iterations on the full-angular data to provide an approximation which then acts like the ideal image. The parameter ϵ relies on noise estimation of the projection data, which might be not easy to acquire. In this work, we tune the parameters to arrive at best performance in terms of artifacts removal and structure-preserving by sampling the parameter space.

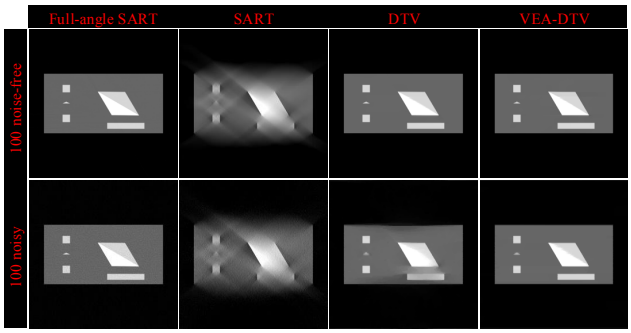


Fig. 2: The analytic rectangle phantom. From left to right, the images are reconstructed by full-angle SART, SART, DTV, VEA-DTV, respectively. The first row shows the reconstructions without noise, while the second row shows the results with added Poisson noise, with incidence intensity $I_0 = 1.5 \times 10^5$. The display window is set to [0, 0.5].

A. Inverse crime test

The inverse crime occurs when employing the same forward reconstruction model to generate, as well as to invert, synthetic data. To avoid the inverse crime, analytic projection data are acquired in CTSim. Both noise-free and noisy projection data are tested. The scanning angular range is set to $[\frac{2\pi}{9}, \frac{7\pi}{9}]$. Poisson noise with incidence intensity $I_0 = 1.5\times10^5$ is added to the analytic projection data.

The results are shown in Fig.2. From left to right, the columns 1 and 2 show the SART (10 iterations) reconstructions, with full data and limited data, respectively, and the columns 3 and 4 show the reconstructions of DTV and VEA-DTV, respectively. As Fig.1 has demonstrated, in the limited-angle reconstructions, the invisible edges are too blurred to be recognized. The first row and second row show the noise-free and noisy reconstructions, respectively. We can easily observe that for the noise-free case, DTV and VEA-DTV achieves similar high quality reconstructions, while for the noisy case, VEA-DTV demonstrates superior results. Distortions and blurring could be easily recognized in the DTV reconstructions, especially at the right bottom part. For the proposed VEA-DTV, blurring has been completely removed, and just small local distortions could be identified along the diagonal of the big parallel gram. Same conclusion could be drawn from the quantitative measures listed in Table I.

TABLE I: PSNR, SSIM and NRMSE for the analytic rectangle phantom.

	Index	SART	DTV	VEA-DTV
noise-free	PSNR	22.9650	38.7488	**39.0063**
	SSIM	0.81761	0.99175	**0.99195**
	NRMSE	0.12726	0.00336	**0.00317**
noisy	PSNR	21.8880	31.8155	**34.2304**
	SSIM	0.58396	0.93394	**0.99195**
	NRMSE	0.16308	0.02899	**0.00951**

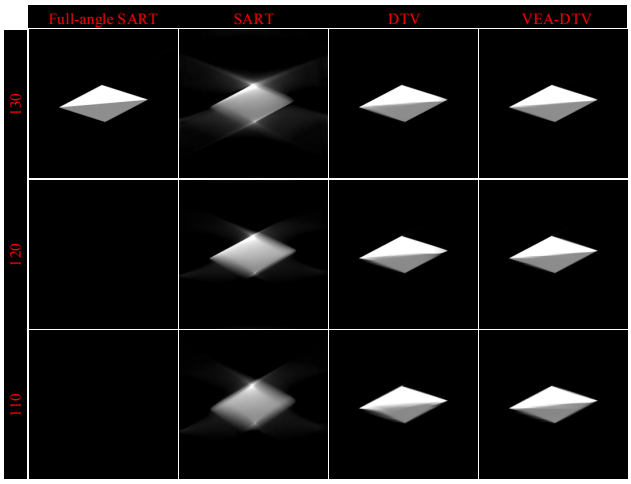


Fig. 3: The rhombus phantom. From left to right, the images are reconstruction results from full-angle SART, SART, DTV, VEA-DTV, respectively. From up to bottom, each row shows the reconstructions with different angular ranges. The display window is set to [0, 0.18].

B. Invisible edges recovery capability test

One rhombus phantom with tilt angle of 5 degrees is constructed in CTSim. Its projection data without noise are also acquired in CTSim. Both of them are analytic, which are used to test VEA-DTV's capacity to recover invisible edges. Since the boundary between the two triangles are completely invisible and not distributed along the axes, it is quite challenging to recover it.

The results are shown in Fig.3. From top to bottom, the images are reconstructed with angular ranges 130, 120 and 110 degrees, respectively. From left to right, the columns 2, 3, and 4 show the reconstructions by SART, DTV and VEA-DTV, respectively. As shown in the second column, the SART method fails to recover the invisible edge. For both DTV and VEA-DTV, the quality of the reconstructions decreases with reducing angular ranges, as demonstrated in the last two columns of Fig.3. When the scanning angular range is 130 degrees, both DTV and VEA-DTV recover the invisible edge nearly perfectly. However, when reducing the angular ranges, the performance of DTV deteriorates quickly, while the proposed VEA-DTV could demonstrate certain resistances to such changes. Same conclusion could be drawn from the quantitative measures listed in Table II.

TABLE II: PSNR, SSIM and NRMSE for the rhombus phantom.

Angular range	Index	SART	DTV	VEA-DTV
$[\frac{5\pi}{36}, \frac{31\pi}{36}]$	PSNR	24.2663	33.8890	**35.3223**
	SSIM	0.53365	0.99449	**0.99557**
	NRMSE	0.55199	0.01048	**0.00754**
$[\frac{\pi}{6}, \frac{5\pi}{6}]$	PSNR	21.8979	31.8155	**33.7639**
	SSIM	0.89156	0.99177	**0.99376**
	NRMSE	0.16742	0.01690	**0.01079**
$[\frac{7\pi}{36}, \frac{29\pi}{36}]$	PSNR	19.78663	29.9508	**30.5150**
	SSIM	0.85852	0.98734	**0.98829**
	NRMSE	0.27222	0.02597	**0.02279**

IV. Conclusion

We have proposed a visible edge aware convex model for limited-angle reconstruction which is derived by reformulating a DTV model. By treating the visible edges and the invisible ones differently, the proposed algorithm could make better use of the visible edges prior and achieve better reconstructions. Numerical experiments suggest that, compared to DTV, the proposed VEA-DTV demonstrates improved stability against noise and angular range reducing.

Acknowledgment

Thanks for the support of the National Natural Science Foundation of China (NSFC) (61971292, 61827809 and 61871275).

References

[1] Dimitri P Bertsekas et al. Incremental gradient, subgradient, and proximal methods for convex optimization: A survey. *Optimization for Machine Learning*, 2010(1-38):3, 2011.

[2] Guang-Hong Chen, Jie Tang, and Shuai Leng. Prior image constrained compressed sensing (PICCS): a method to accurately reconstruct dynamic CT images from highly undersampled projection data sets. *Medical Physics*, 35(2):660–663, 2008.

[3] Zhiqiang Chen, Xin Jin, Liang Li, and Ge Wang. A limited-angle CT reconstruction method based on anisotropic TV minimization. *Physics in Medicine & Biology*, 58(7):2119, 2013.

[4] Rebecca Fahrig, Robert Dixon, Thomas Payne, Richard L Morin, Arundhuti Ganguly, and Norbert Strobel. Dose and image quality for a cone-beam C-arm CT system. *Medical Physics*, 33(12):4541–4550, 2006.

[5] Hiroyuki Kudo and Tsuneo Saito. Sinogram recovery with the method of convex projections for limited-data reconstruction in computed tomography. *J. Opt. Soc. Am. A*, 8(7):1148–1160, Jul 1991.

[6] Yan Liu, Jianhua Ma, Yi Fan, and Zhengrong Liang. Adaptive-weighted total variation minimization for sparse data toward low-dose x-ray computed tomography image reconstruction. *Physics in Medicine & Biology*, 57(23):7923, 2012.

[7] Loren Niklason, Bradley Christian, Loren Niklason, Daniel Kopans, D Castleberry, B Opsahl-Ong, Cynthia Landberg, Priscilla Slanetz, Angela Giardino, R Moore, Douglas Albagli, M DeJule, Paul Fitzgerald, David Fobare, B Giambattista, Robert Kwasnick, J Liu, S Lubowski, G Possin, and R Wirth. Digital tomosynthesis in breast imaging. *Radiology*, 205(2):399–406, 1997.

[8] Eric Todd Quinto. Artifacts and visible singularities in limited data X-ray tomography. *Sensing and Imaging*, 18(1):9, 2017.

[9] Emil Y Sidky, Jakob H Jørgensen, and Xiaochuan Pan. Convex optimization problem prototyping for image reconstruction in computed tomography with the chambolle–pock algorithm. *Physics in Medicine and Biology*, 57(10):3065–3091, apr 2012.

[10] Emil Y Sidky, Chien-Min Kao, and Xiaochuan Pan. Accurate image reconstruction from few-views and limited-angle data in divergent-beam CT. *Journal of X-ray Science and Technology*, 14(2):119–139, 2006.

[11] Emil Y Sidky and Xiaochuan Pan. Image reconstruction in circular cone-beam computed tomography by constrained, total-variation minimization. *Physics in Medicine & Biology*, 53(17):4777, 2008.

[12] Ting Wang, Katsuhiro Nakamoto, Heye Zhang, and Huafeng Liu. Reweighted anisotropic total variation minimization for limited-angle CT reconstruction. *IEEE Transactions on Nuclear Science*, 64(10):2742–2760, 2017.

[13] Jinqiu Xu, Yunsong Zhao, Hongwei Li, and Peng Zhang. An image reconstruction model regularized by edge-preserving diffusion and smoothing for limited-angle computed tomography. *Inverse Problems*, 35(8):085004, 2019.

[14] Lingli Zhang, Li Zeng, and Yumeng Guo. l0 regularization based on a prior image incorporated non-local means for limited-angle x-ray ct reconstruction. *Journal of X-ray science and technology*, 26(3):481–498, 2018.

[15] Zheng Zhang, Buxin Chen, Dan Xia, Emil Y. Sidky, and Xiaochuan Pan. Directional-tv algorithm for image reconstruction from limited-angular-range data. *Medical Image Analysis*, 70:102030, 2021.

Dual-task Learning For Low-Dose CT Simulation and Denoising

Mingqiang Meng, Yongbo Wang, Manman Zhu, Xi Tao, Zhaoying Bian, Dong Zeng, and Jianhua Ma

***Abstract*—Deep learning (DL) are being extensively investigated for low-dose computed tomography (CT). The success of DL lies in the availability of big data, learning the non-linear mapping of low-dose CT to target images based on convolutional neural networks. However, due to the commercial confidentiality of CT vendors, there are very few publicly raw projection data available to simulate paired training data, which greatly reduces the generalization and performance of the network. In the paper, we propose a dual-task learning network (DTNet) for low-dose CT simulation and denoising at arbitrary dose levels simultaneously. The DTNet can integrate low-lose CT simulation and denoising into a unified optimization framework by learning the joint distribution of low-dose CT and normal-dose CT data. Specifically, in the simulation task, we propose to train the simulation network by learning a mapping from normal-dose to low-dose at different levels, where the dose level can be continuously controlled by a noise factor. In the denoising task, we propose a multi-level low-dose CT learning strategy to train the denoising network, learning many-to-one mapping. The experimental results demonstrate the effectiveness of our proposed method in low-dose CT simulation and denoising at arbitrary dose levels.**

***Index Terms*—Computed tomography, dual-task learning, denoising network, simulation network.**

I. Introduction

X-RAYcomputed tomography (CT) is widely used in clinical diagnosis and treatment because of its ability to image the body's three-dimensional anatomy in a non-invasive manner. However, ionizing radiation generated during CT scanning will accumulate in the human body, and high radiation doses will induce the risk of cancer in human tissues and organs [1]. Given these risks, efforts have been made on reducing the radiation dose and the principles of As Low As Reasonably Achievable (ALARA) is profoundly practiced in clinical CT imaging [2]. However, reducing the radiation dose inevitably increases the noise and artifacts of reconstructed CT images, which compromises the diagnostic performance. Consequently, improving the image quality of low-dose CT (LDCT) has become a hot topic in medical imaging over the past decade.

Recently, with the rapid development of deep learning (DL) technology, the LDCT imaging algorithm is dominated by convolutional neural network and has achieved unprecedented success. DL-based algorithm learns the mapping from LDCT projection/image to normal-dose CT (NDCT) ones by designing an elaborate convolutional neural network (CNN), such as RED-CNN [3], Wavelet networks [4], and Tensor-Net [5]. A key factor in the success of these supervised algorithms is the availability of big data, that is, a large amount of paired LDCT and NDCT images [6], [7]. Despite its superior denoising results, some issues still must be resolved before the DL models can be widely deployed in clinic. First, given the increase in total radiation dose, matched LDCT and NDCT data cannot be obtained in clinical practice. As a result, true NDCT and LDCT paired data are not available. Second, traditional LDCT simulation methods [8] usually insert noise into the raw projection data, however, very few raw data are publicly available to simulate paired training data, which degrades the generalization performance of the network. Third, most of DL-based models are designed for specific dose levels but perform poorly at lower doses.

To solve above problems, in this work, we propose a dual-task learning network (DTNet) for low-dose CT simulation and denoising at arbitrary dose levels. The presented DTNet can integrate LDCT simulation and denoising into a unified framework by learning the joint distribution of LDCT and NDCT data. Specifically, we first propose to train the simulation network by learning a mapping from NDCT to LDCT at different levels, where the dose level can be continuously controlled by a noise factor. In the denoising task, we present a multi-level LDCT learning strategy, which uses LDCT data of different levels to train the denoising network. Once trained, DTNet can be used for multi-level LDCT images simulation and denoising simultaneously, which greatly improves the applicability and generalization performance of the model.

II. Methodology

A. Low-dose CT Simulation Network

Let $x \in \mathbb{I}^{H\times W}$ denotes a NDCT image and $y \in \mathbb{I}^{H\times W}$ denotes the corresponding LDCT image. Low-dose CT simulation is an inverse process of LDCT denoising, which learns the opposite mapping from x to y. To precisely control the noise level of the generated LDCT image, we propose a noise control factor guided LDCT simulation scheme, as illustrated in Fig. 1 (a). Specifically, given a NDCT image and a mask image with value of 1, the simulator network encodes them into high-dimensional features for coupling and progressively decoding to reconstruct LDCT image of different levels, where the dose level can be continuously controlled by a noise factor. In addition, instead of directly applying the pixel-wise loss to

This work was supported in part by the NSFC under Grant U21A6005 and Grant U1708261, the National Key R&D Program of China under Grant No. 2020YFA0712200, and the China Postdoctoral Science Foundation under Grant 2021M701641. (Corresponding author: Jianhua Ma.).

M. Meng, Y. Wang, M. Zhu, X. Tao, Z. Bian, D. Zeng and J. Ma are with the School of Biomedical Engineering, Southern Medical University, Guangdong, China; and the Guangdong Artificial Intelligence and Digital Economy Laboratory (Guangzhou), Guangdong, China (e-mail: jhma@smu.edu.cn).

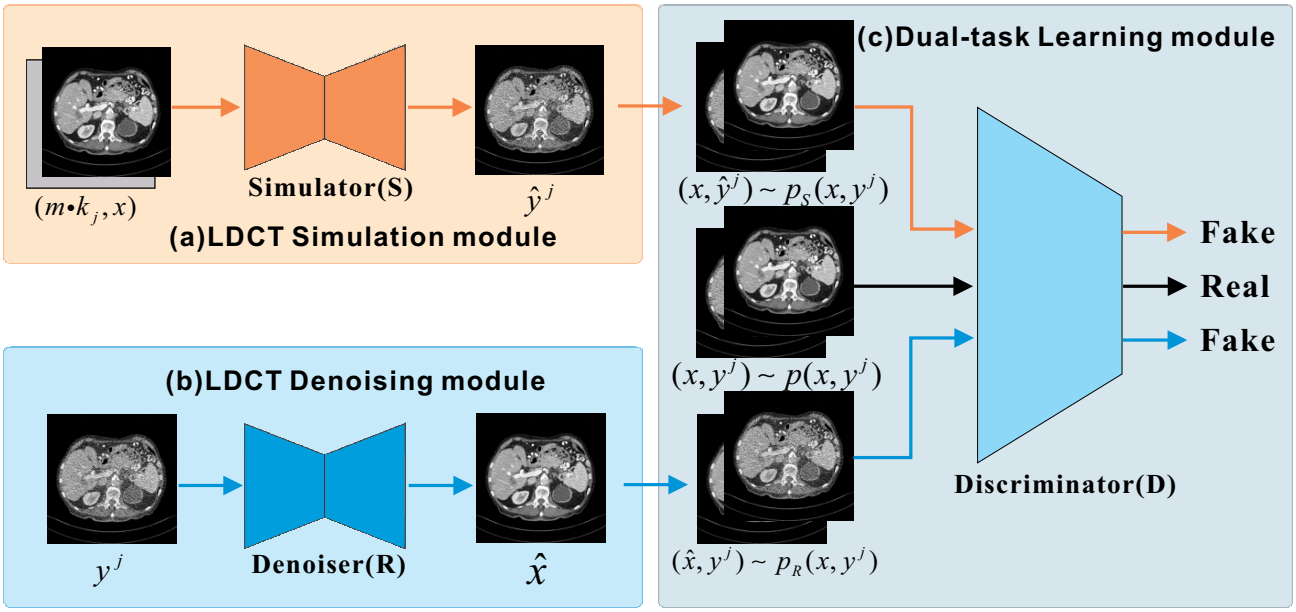


Fig. 1. The overall structure of the proposed DTNet framework. (a) Low-Dose CT simulation module, (b) Low-Dose CT denoising module, (c) Dual-task learning module. m is the mask image with value of 1. k_j is the noise scale factor.

the target image, we use a gaussian filter to extract the first-order statistics information of simulation and reference LDCT noise, and then constraint it with the MAE loss function, which can be formulated as:

$$L_S = arg\min_{\theta_S} \sum_{i=1}^{N} ||GF(S(x_i, m \cdot k_j) - x_i) - GF(y_i{}^j - x_i)||_1. \tag{1}$$

Here, m is the mask image with value of 1, $k_j = 1, ..., j$ is the noise factor that controls the simulated dose level j. θ_S represents the parameters of simulation network S. GF is a 2d Gaussian filter convolution kernel with a size of 5×5.

B. Low-dose CT Denoising Network

Typically, the DL-based denoising problem is to build a prediction network $R(\cdot)$ that learns the non-linear mapping from y to x, i.e., $R : y \rightarrow x$. In clinical, CT images of various dose levels may be obtained to meet the clinical diagnosis demands. In Fig. 1 (b), we propose a multi-level low-dose CT learning strategy to train the denoising network with different levels low-dose CT data by minimizing the mean absolute error (MAE) loss function, which is expressed as:

$$L_R = arg\min_{\theta_R} \sum_{i=1}^{N} ||R(y_i{}^j) - x_i||_1. \tag{2}$$

Here, $y_i{}^j, i = 1, ..., N, j = 1, ..., M$ represents the LDCT images, where N is the total number of LDCT images and M is number of dose levels. θ_R represent the parameters of denoising network R. Once trained, the denoising model can be applied to LDCT reconstruction at different dose levels in the clinic.

C. Dual-task Learning Network

To jointly optimize low-dose CT simulation and denoising tasks, we adopt the dual-task learning network (DTNet), which uses a joint discriminator to alternately optimize the simulation and denoising network by learning the joint distribution $p(x, y^j)$ of LDCT and NDCT data, as shown in Fig. 1 (c). Let $p_S(x, y^j)$ and $p_R(x, y^j)$ represent the pseudo joint distribution of simulation and denoising task, respectively. The dual adversarial loss can be defined as follows:

$$\begin{aligned} \min_{S,R} \max_{D}(S, R, , D) = & E_{(x,y^j)\backsim p(x,y^j)}[D(x, y^j)] \\ & - \lambda_S \cdot E_{(x,\hat{y}^j)\backsim p_S(x,y^j)}[D(x, \hat{y}^j)] \\ & - \lambda_R \cdot E_{(\hat{x},y^j)\backsim p_R(x,y^j)}[D(\hat{x}, y^j)]. \end{aligned} \tag{3}$$

Here, $E[\cdot]$ denotes the expectation operator, D represents the discriminator, which is used to receive the image pair (x, y^j), $(\hat{x}, y^j)$, $(x, \hat{y}^j)$ and distinguish them as real or fake samples. The hyper-parameters λ_R and λ_S controls the weight of GAN loss.

The final objective function for optimizing DTNet can be formulated as:

$$\min_{S,R} \max_{D}(S, R, D) + \alpha \cdot L_S + \beta \cdot L_R, \tag{4}$$

where α and β are the weight parameters that control the trade-off between adversarial loss and fidelity loss of simulation and denoising tasks.

III. Experiments

A. Dataset

The 2016 NIH-AAPM-Mayo Clinic Low Dose CT Grand Challenge dataset [9] published by Mayo Clinic is used to evaluate the effectiveness of the proposed DTNet model.

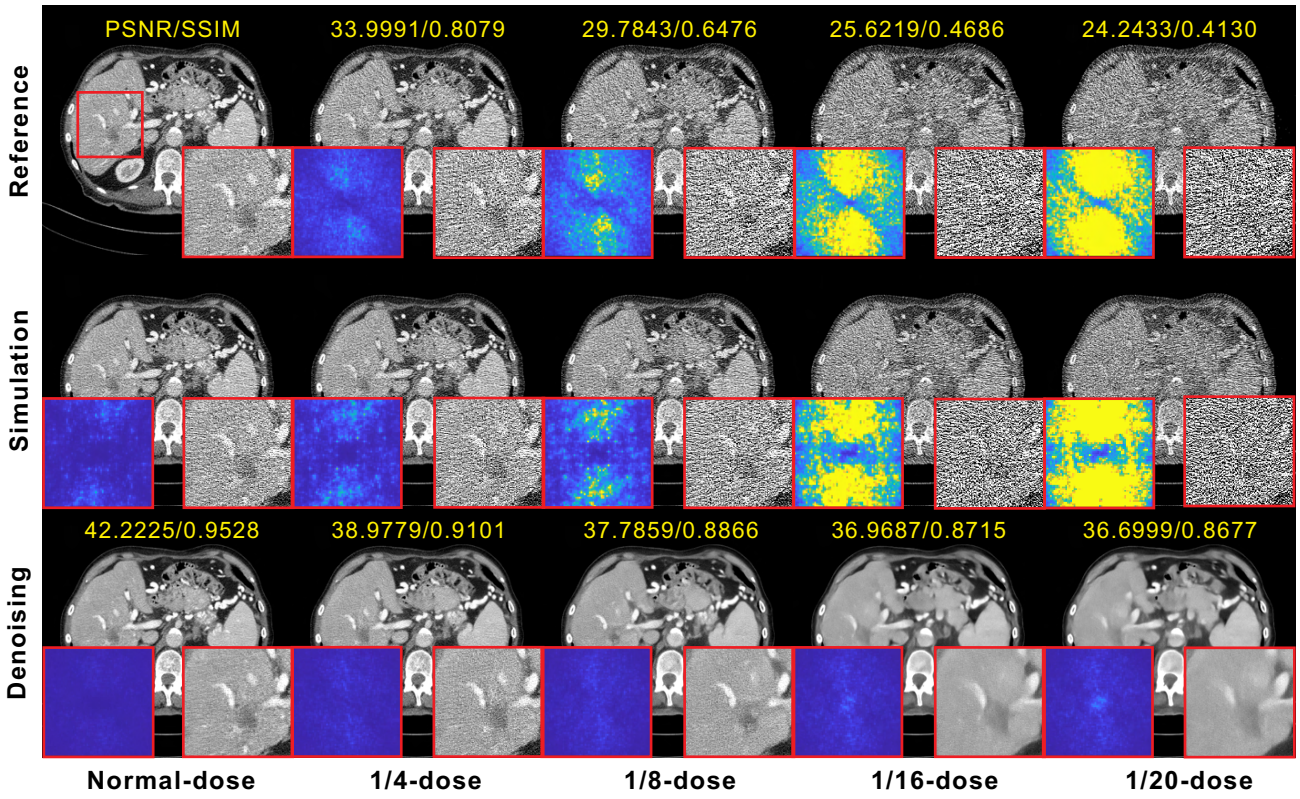


Fig. 2. The simulation and denoising results at different dose levels of proposed DTNet. The display window of images and zoomed-in ROIs is $[-140, 260]$ HU. The display window of NPS maps is [0 3000] HU^2mm^2.

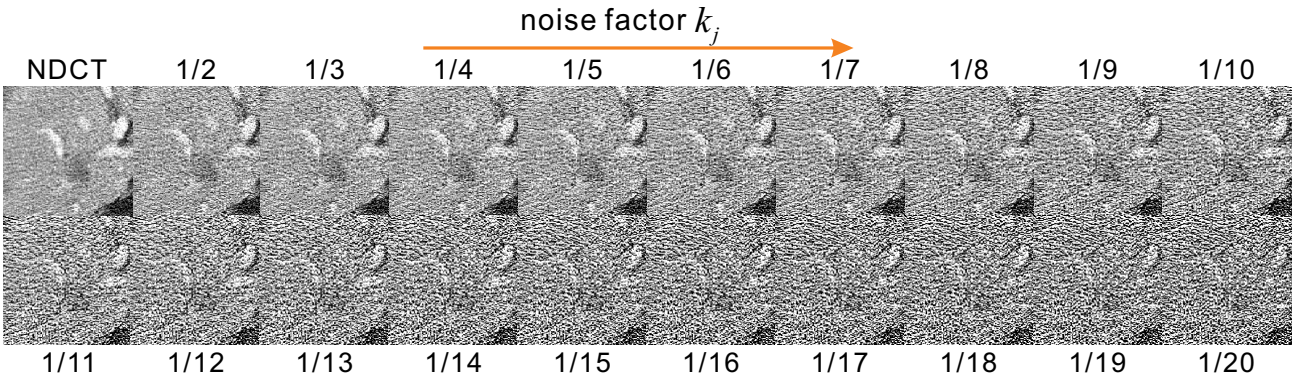


Fig. 3. Arbitrary dose levels simulation results of proposed DTNet. From the left to the right column, the noise level of CT image gradually increases as the noise factor k_j increases. The display window is $[-140, 260]$ HU

This dataset contains 10 anonymized patient normal-dose raw projection data, which were acquired using the Somatom Definition AS+ CT system under 100kV or 120kV and automatic exposure controlling mode, and simulated quarter-dose projection data. In order to obtain LDCT data at different dose levels, we re-simulated 1/4, 1/8, 1/16, 1/20-dose projection data using the corresponding simulation algorithm [8] to insert quantum and electronic noise into the normal-dose projection data. In this study, we selected seven patients with a total of 17056 image data pairs for training. Specifically, 4568 image pairs collected from two patients are used to validate the performance of DTNet, and 2100 image pairs from the remaining one patient are selected as the testing set.

B. Implementation details

In our experiments, the proposed DTNet model consists of three sub-networks: simulator S, denoiser R and discriminator D. For S and R, we use the same generator network structure UNet [10], which contains an encoder and decoder. The discriminator D has a similar structure to the PatchGAN [11]. The DTNet model is optimized in an alternating manner using Adam algorithm. The learning rates of S, R and D are set

to $1e^{-4}$, $1e^{-4}$, and $2e^{-4}$, respectively. The hyper-parameters of loss function are selected to be $\alpha = 100$, $\beta = 10$, $\lambda_R = \lambda_S = 0.5$. During training, we randomly extracted 4 patches of size 128×128 as input in each iteration, D is updated three times while S and R are updated once. All networks are implemented using Pytorch and trained with an GeForce RTX 3090 GPU.

IV. Results

Fig. 2 shows the simulation and denoising results of abdominal CT image at five different dose levels: normal-dose, 1/4, 1/8, 1/16, 1/20-dose. In the simulation task, the corresponding simulation noise factors are set to 0, 1/4, 1/8, 1/16, and 1/20, respectively. Note that the noise factor equal to 0 means that it does not insert any noise into the NDCT image. It can be observed that the proposed method can simulate LDCT images of different dose levels and the learned noise intensity and characterization are similar to the reference images. In addition, we also calculate the noise power spectrum (NPS) maps of magnified ROIs to evaluate the statistical property of noise. We can observe that the NPS of generated LDCT images is very close to the reference images. When the noise factor is set to 0, only little noise is embedded in the output image, which can be seen from the ROI and the corresponding NPS. To verify the robustness of DTNet in simulating other dose levels which are not including in training data, as shown in Fig. 3. It can be seen that the noise intensity of simulated LDCT images continuously increases with the increase of noise factor k_j. This demonstrates that the proposed DTNet has the ability to simulate the realistic LDCT images and can control the noise level well.

In the denoising task, the last row in Fig. 2 shows that DTNet can efficiently suppress noise and artifacts at difference dose levels. In particular, for normal-dose images with a small amount of noise, we can also remove the noise without smoothing the image content. And for ultra-low doses, such as 1/20-dose, some small structures are completely drowned out by noise and are difficult to recover well. Therefore, ultra-low dose scanning can be used for special imaging tasks where anatomical details are not important, such as localization imaging. To quantitatively analyze the denoising performance of DTNet, we calculate the peak-to-noise ratio (PSNR) and structural similarity (SSIM), as summarized in Table I. We can see that DTNet obtains the best quantitative values at different dose levels compared to RED-CNN [3] and WGAN-VGG [12], which is consistent with the visual evaluation.

TABLE I
PSNR and SSIM Quantitative comparison of DTNet denoising results at different dose levels.

Dose	RED-CNN	WGAN-VGG	DTNet
1/20	36.5572 ± 2.3307	33.3125 ± 2.5499	$\mathbf{36.6999 \pm 2.3983}$
	0.8665 ± 0.0573	0.7543 ± 0.0994	$\mathbf{0.8677 \pm 0.0583}$
1/16	36.8386 ± 2.3549	34.0276 ± 2.5127	$\mathbf{36.9687 \pm 2.4371}$
	0.8705 ± 0.0561	0.7795 ± 0.0908	$\mathbf{0.8715 \pm 0.0574}$
1/8	37.6845 ± 2.5182	35.8733 ± 2.5104	$\mathbf{37.7859 \pm 2.5292}$
	0.8839 ± 0.0533	0.8371 ± 0.0712	$\mathbf{0.8866 \pm 0.0515}$
1/4	38.6890 ± 2.6873	37.4532 ± 2.5651	$\mathbf{38.9779 \pm 2.7158}$
	0.9024 ± 0.0472	0.8789 ± 0.0557	$\mathbf{0.9101 \pm 0.0426}$

V. Conclusion

In this paper, we have presented a dual-task learning network (DTNet) for LDCT simulation and denoising tasks. In the simulation task, the simulation network encodes the NDCT image and mask image into high-dimensional features for coupling and decoding to generate LDCT images at different dose levels, where the dose levels can be controlled by a noise factor. The presented DTNet integrates the LDCT simulation and denoising tasks into a unified optimization model. In the denoising task, the multi-level LDCT learning strategy is used to train the denoising network, which can learn many-to-one end-to-end mapping. Both the quantitative and qualitative evaluation results have demonstrated the promising performance of DTNet in terms of LDCT simulation and denoising. In the feature, we will further improve the performance of DTNet by incorporating advanced network and prior information.

References

[1] R. Smith-Bindman, J. Lipson, R. Marcus, K.-P. Kim, M. Mahesh, R. Gould, A. Berrington de Gonzlez, and D. L. Miglioretti, "Radiation Dose Associated With Common Computed Tomography Examinations an0d the Associated Lifetime Attributable Risk of Cancer," *Archives of Internal Medicine*, vol. 169, no. 22, pp. 2078-2086, 2009.

[2] J. Valentin, "International Commission on Radiological Protection. The 2007 recommendations of the international commission on radiological protection," *Annals of the ICRP, ICRP Publication*, vol. 103, pp. 2-4, 2007.

[3] H. Chen, Y. Zhang, M. K. Kalra, F. Lin, Y. Chen, P. Liao, J. Zhou, and G. Wang, "Low-dose CT with a residual encoder-decoder convolutional neural network," *IEEE transactions on medical imaging*, vol. 36, no. 12, pp. 2524-2535, 2017.

[4] E. Kang, J. Min, and J. C. Ye, "A deep convolutional neural network using directional wavelets for lowdose Xray CT reconstruction," *Medical physics*, vol. 44, no. 10, pp. e360-e375, 2017.

[5] X. Tao, Y. Wang, L. Lin, Z. Hong, and J. Ma, "Learning to Reconstruct CT Images from the VVBP-Tensor," *IEEE Transactions on Medical Imaging*, 2021.

[6] G. Wang, J. C. Ye, and B. De Man, "Deep learning for tomographic image reconstruction," *Nature Machine Intelligence*, vol. 2, no. 12, pp. 737-748, 2020.

[7] G. Wang, M. Jacob, X. Mou, Y. Shi, and Y. C. Eldar, "Deep Tomographic Image Reconstruction: Yesterday, Today, and TomorrowEditorial for the 2nd Special Issue Machine Learning for Image Reconstruction," *IEEE transactions on medical imaging*, vol. 40, no. 11, pp. 2956-2964, 2021.

[8] L. Yu, M. Shiung, D. Jondal, and C. H. McCollough, "Development and validation of a practical lower-dose-simulation tool for optimizing computed tomography scan protocols," *Journal of computer assisted tomography*, vol. 36, no. 4, pp. 477-487, 2012.

[9] https://www.aapm.org/GrandChallenge/LowDoseCT/.

[10] O. Ronneberger, P. Fischer, and T. Brox, "U-net: Convolutional networks for biomedical image segmentation," *International Conference on Medical image computing and computer-assisted intervention*, pp. 234-241, 2015.

[11] S. Bera, and P. K. Biswas, "Noise Conscious Training of Non Local Neural Network Powered by Self Attentive Spectral Normalized Markovian Patch GAN for Low Dose CT Denoising," *IEEE Transactions on Medical Imaging*, 2021.

[12] Q. Yang, P. Yan, Y. Zhang, H. Yu, Y. Shi, X. Mou, M. K. Kalra, Y. Zhang, L. Sun, and G. Wang, "Low-dose CT image denoising using a generative adversarial network with Wasserstein distance and perceptual loss," *IEEE transactions on medical imaging*, vol. 37, no. 6, pp. 1348-1357, 2018.

Statistical Iteration Reconstruction based on Gaussian Mixture Model for Photon-counting CT

Danyang Li, Zheng Duan, Dong Zeng, Zhaoying Bian, and Jianhua Ma

***Abstract*—Photon-counting computed tomography (PCCT) can simultaneously obtain multi-energy data with abundant energy-dependent material-specific information of the scanned object. However, the photon counts in each energy bin are decreased and the collected data suffers from photon starvation effects, which degrades the quality of the reconstructed PCCT images. To solve it, many statistical iteration reconstruction (SIR) methods have been proposed by constructing data-fidelity and prior terms to suppress noise and remove artifacts. However, most of the current SIR methods assume the noise in PCCT images follows a Gaussian distribution, which deviates the real distribution of the noise in PCCT images. Therefore, we propose a new statistical iteration reconstruction method by considering more complex noise distribution in reality. Specifically, Gaussian mixture model (GMM), which is an universal approximator for any continuous density function, is utilized to model the noise in PCCT images. Moreover, the multi-energy PCCT images are treated as a 3-order tensor which is regularized by three dimensional total variation (3DTV) prior term. Finally, a statistical iteration reconstruction model based on GMM and 3DTV is established for PCCT imaging. For shorten, we call the presented reconstruction model as "GMM-3DTV". We then develop an expectation-maximization (EM) algorithm to solve the presented GMM-3DTV method. Numerical studies demonstrated the improvements of the presented GMM-3DTV method over the competing methods.**

***Index Terms*—Photon-counting CT, statistical iteration reconstruction, Gaussian mixture model, 3DTV.**

I. Introduction

RECENTLY, photon-counting computed tomography (PCCT) has been developed in clinics. It utilizes the photon counting detectors (PCDs) to simultaneously count photons in multi-energy bins. By obtaining abundant energy-dependent material-specific information and high contrast-to-noise ratio (CNR) for soft materials, PCCT has the advantages on material decomposition and lesion diagnostic [1], [2].

However, the narrow energy bin receives decreasing photons and the spectral data suffers from serious quantum noise [3]. This challenge significantly degrades the quality of the reconstructed PCCT images in multi-energy bins and effects the imaging performance of PCCT imaging. To obtain high quality PCCT images, several iteration reconstruction methods have been exploited. It incorporates the statistical property of X-ray photons and prior information of the desired PCCT images to build reconstruction model with data fidelity and regularization terms, respectively. For examples, Xu *et al.* introduced a statistical interior tomography method with TV regularization to reconstruction PCCT images [4], Kim et al. stacked similar image patches at the same position and utilized a low-rank regularization to suppress image noise [5]. Zhang *et al.* utilized a nonlocal mean regularization of the full-spectrum image to maintain the image details [6]. Yao *et al.* proposed to improve the reconstruction performance by utilizing a nonlocal spectral similarity of a weighted image [7]. Liu *et al*. developed a nonlocal total variation (TV) regularization term to construct the weights from the full-spectrum images [8]. Tao *et al.* utilized the structural redundancy between the base materials and the spectral images to establish a prior-knowledge-aware material decomposition method [9]. Zeng *et al.* analyzed the intrinsic tensor properties of the PCCT images and constructed a full-spectrum-knowledge-aware tensor model for PCCT imaging [10]. These methods have been shown great potential in preserving image structures and suppressing noise. Moreover, deep learning (DL) based methods have been utilized in spectral CT. For examples, Li *et al.* constructed a cascade DNN to estimate the high-energy image from low-energy image [11], and Cong *et al.* estimated the PCCT images from the data of energy-integrating detectors by using DNN based method [12].

However, most of the current SIR methods assume the noise in PCCT images follows a Gaussian distribution, which deviates the real distribution of the noise in PCCT images. Because the artifacts, which maybe induced by beam hardening effect in lower energy bins or photon starvation effect of high density materials, would complicate the noise distribution in the image and may invalidate the performances of the aforementioned iteration methods. In addition, the DL-based methods need quantity of paired data to obtain a desired network, and the collection of training data is time-consuming and the clinical PCCT data is hard to be obtained. To solve it, we propose to utilize Gaussian Mixture Model (GMM) [13], which is a universal approximator for any continuous density function, to model the noise in the PCCT images. Moreover, we treat the PCCT images as a 3-order tensor and serve the three dimensional total variation (3DTV) [14] as image prior. Finally, we construct a statistical iteration reconstruction method based on GMM and 3DTV. For shorten, we call the presented reconstruction method as "GMM-3DTV".

In summary, the main contributions of this work are:

- We present a statistical iteration reconstruction method, called GMM-3DTV, by modelling the complex distribution of the noise with GMM and serving the 3DTV as

This work was supported in part by the NSFC under Grant U21A6005 and Grant U1708261. *(Corresponding author: Zhaoying Bian and Jianhua Ma.)*

D. Li, Z. Duan, and D. Zeng are with School of Biomedical Engineering, Southern Medical University, Guangzhou 510515, China.

Z. Bian is with School of Biomedical Engineering, Southern Medical University, Guangzhou 510515, China (e-mail: zybian@smu.edu.cn).

J. Ma is with School of Biomedical Engineering, Southern Medical University, Guangzhou 510515, China (e-mail: jhma@smu.edu.cn).

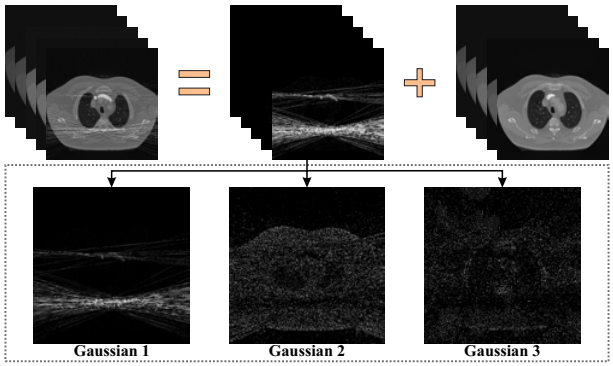


Fig. 1. Illustration of the approximated results by the GMM for complex noise in multi-energy bin PCCT images. First row (from left to right): Noisy PCCT images, noise and ground truth. Second row: three noise components of GMM in Bin 1.

the PCCT image prior.

- Considering the GMM parameters, we also employ an expectation-maximization (EM) algorithm to numerically optimize the presented GMM-3DTV method.
- We evaluated the presented GMM-3DTV on simulated and synthesized clinical data and demonstrated its effectiveness in terms of qualitative and quantitative metrics.

II. Methods

PCCT receives spectral data among multi-energy bins. Considering the spatial and energy dimensions of the measurements, the 3D PCCT imaging model can be expressed as follows:

$$\mathcal{Y} = \mathcal{A}\mathcal{X}, \tag{1}$$

where $\mathcal{Y} = \{y_n, n \leq N\}$ and $\mathcal{X} = \{x_n, n \leq N\}$ denote the measurements and desired PCCT images among multi-energy bins, respectively. N is the total number of the multi-energy bins. $\mathcal{A}$ is a linear projection operator. It should be note that additional constraints are incorporated to stable the solution of the above model.

Due to the variance of the material attenuations for X-ray along the whole spectrum, the feature of the noise is different among multi-energy-bins. Specifically, the reconstructed images in lower bins suffer from strip artifacts due to the beam hardening, and the ones in higher bins are corrupted by the photon starvation induced noise. Therefore, a single Gaussian is no sufficient to approximate the complex noise distribution in the PCCT images.

Fig. 1 shows the noisy PCCT images, noise and ground truth in Bin 1. It can be seen that the PCCT images has multiple modalities of noises, as illustrated by the three Gaussian components, and a simple Gaussian noise model may deviate the real cases. In order to solve the complex noise distribution, we model the noise in the PCCT images as a parametric probability distribution by GMM for more flexibly adapting different cases. Specifically, the noise term ε_n in the nth energy bin is expressed as follows:

$$\varepsilon_n \sim \sum_{k=1}^{K} \pi_{nk} \mathcal{N}\left(\varepsilon_n | \mu_{nk}, \sigma_{nk}^2\right), \tag{2}$$

where π_{nk}, μ_{nk} and σ_{nk}^2 denote the mixture rate, mean and variance values of the kth Gaussian compound in nth energy bin, respectively. In this work, the mean values are set to be zero. K is the total number of Gaussian compounds, and $\sum_{k=1}^{K} \pi_{nk} = 1$.

Considering the sparsity structures and low rank property of the PCCT images, a general 3D total variation (3DTV) term is utilized as the prior, as follows:

$$R_{3DTV}\left(\mathcal{X}\right) = \sum_{m=1}^{3} \|\nabla_m \mathcal{X}\|_1, \tag{3}$$

where ∇_m is different operations along spatial height, width and spectrum modes of $\mathcal{X}$, $\|\cdot\|_1$ is the L_1 norm. Therefore, coupling the GMM approximation for noise and sparse regularization for PCCT images, we can formulate a robust penalized weighted least squares method for PCCT imaging as follows:

$$\begin{aligned} &\min_{\mathcal{X}} \frac{1}{2}\|\mathcal{Y} - \mathcal{A}\mathcal{X}\|_2^2 + \alpha R_{3DTV}\left(\mathcal{X}\right), \\ &\varepsilon_n \sim \sum_{k=1}^{K} \pi_{nk} \mathcal{N}\left(\varepsilon_n | \mu_{nk}, \sigma_{nk}^2\right), n = 1, ..., N, \end{aligned} \tag{4}$$

where α is a hyper-parameter of the image prior term. Simply, we call the above method as "GMM-3DTV" method. Finally, by imposing the negative form of likelihood function of the GMM, Eq. (4) is rewritten as follows:

$$\begin{aligned} \min_{\mathcal{X},\Pi,\Sigma} &\frac{1}{2}\|\mathcal{Y} - \mathcal{A}\mathcal{X}\|_2^2 + \alpha R_{3DTV}\left(\mathcal{X}\right) \\ &- \beta \sum_{n=1}^{N} \log \sum_{k=1}^{K} \pi_{nk} \mathcal{N}\left(\varepsilon_n | 0, \sigma_{nk}^2\right), \end{aligned} \tag{5}$$

where β is the hyper-parameters of GMM likelihood terms. $\Pi = \{\pi_{nk}, n \leq N, k \leq K\}$ and $\Sigma = \{\sigma_{nk}^2, n \leq N, k \leq K\}$ are the sets of mixture coefficients and variance values, respectively. Moreover, we adopt an Expectation Maximization (EM) algorithm [15] to iteratively optimize the presented GMM-3DTV method.

The whole EM algorithm for optimizing Eq. (5) can be summarized in Algorithm 1.

Algorithm 1: Algorithm for Solving Eq. (5)

Input : The PCCT multi-energy measurements $\mathcal{Y}$, regularized parameters α and β, and stopping criteria ϑ.

Output: Reconstructed PCCT multi-energy images $\mathcal{X}$.

Initialization: $\mathbf{\Pi}, \mathbf{\Sigma}$, GMM number K;

while *not satisfy stopping criteria* ϑ **do**

  E step: calculate the expectation of posterior probability of GMM parameters;

  M step: maximum the augmented Lagrangian function.

end

Fig. 2. Reconstructed images of the presented and compared methods on XCAT phantom. The display windows from Bin 1 to 5 are [0.00, 0.02], [0.00, 0.010], [0.00, 0.017], [0.00, 0.015] and [0.00, 0.010] mm^{-1}, respectively. Zoomed ROIs indicated by the red box are displayed for better visualization.

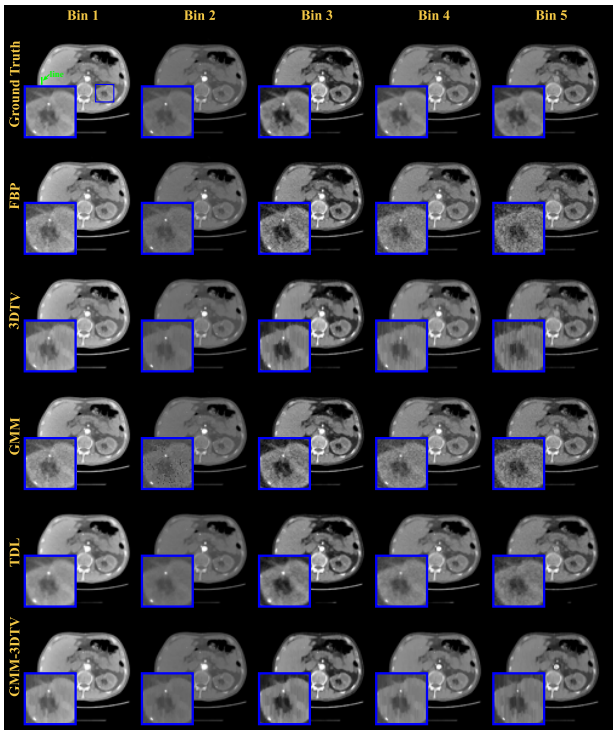


Fig. 3. Reconstructed images of the presented and compared methods on synthesized clinical phantom. The display windows from Bin 1 to 5 are [0.007, 0.0105], [0.0032, 0.0072], [0.003, 0.004], [0.0022, 0.0035] and [0.0018, 0.0027] mm^{-1}, respectively. Zoomed ROIs indicated by the blue box are displayed for better visualization.

III. Results

A. Implementation details

In this work, we compared the presented GMM-3DTV method with the filtered back projection (FBP) method using a ramp filter and a tensor-based dictionary learning regularization (TDL) method. In addition, two more methods were implemented to investigate different parts of the presented GMM-3DTV method. The one is GMM-based reconstruction model, and the other is 3DTV-based reconstruction model. The images at normal dose serve as the ground truth. Two numerical phantoms are utilized to validate the performance of the presented method. Specifically, XCAT phantom [16] contains alcohol, water, bone and gadolinium, and the synthesized clinical phantom with water, bone and iodine. The phantoms were scanned under the simulated 120 kVp X-ray spectrum with 1.6 mm Al filtration by SPEKTR toolbox [17]. Five energy bins with equal photon counts are determined by the thresholds: 25, 50, 60, 70 and 85 keV. The PCCT imaging parameters are set as follows: (1) source-to-detector and source-to-center distances are 1040.0 and 570.0 mm, respectively. (2) 1160 projection views are evenly scanned the objections over 360^o. (3) 816 detector channels are placed along the parallel X-ray beam. To generate the noisy projections, Poisson noise is applied into the simulated noise-free projections.

B. Qualitative analysis

Fig. 2 shows visual comparisons of the presented and compared methods on XCAT phantom. It can be observed that: 1) the FBP method suffers from noise-induced artifacts, and the other methods outperform the FBP method in terms of improving the image quality; 2) the 3DTV-based method suffers from blocky artifacts and the GMM-based method remains produces noisy images; 3) the TDL method smooths the images and enhances the texture of the noise-induced artifacts; 4) the presented GMM-3DTV method better handles the noise-induced artifacts and preserves the image details. Moreover, zoomed in regions-of-interest (ROIs) indicated by the red boxes in each images are illustrated. It can be observed that the presented GMM-3DTV maintains the details of the anatomic structures.

Fig. 3 illustrates the results of different methods on synthesized clinical phantom. Similar with the results on XCAT phantom, the TDL method fails to denoise among the multi-energy bins, the 3DTV induces additional block artifacts for the denoising results, and the GMM-based method hardly removes the noise. On the contrary, the presented GMM-3DTV method can effectively remove the noise-induced artifacts and recover the images details. For better visual inspection, ROIs indicated by the blue boxes are shown, which also demonstrates the same conclusion.

Fig. 4 shows the profiles of different methods on synthesized clinical phantom indicated by the green line in Fig 3. From the results, we can seen that the TDL and 3DTV-based methods produce blurry results, and the GMM-based method fails to suppress the noise. In contrast, the results of the presented GMM-3DTV method are the closest to the ground truth.

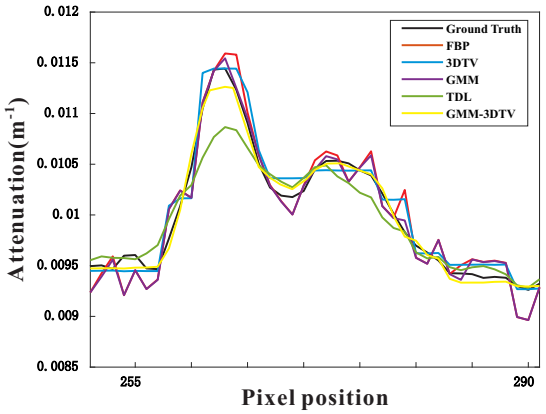


Fig. 4. Profiles of the results from different methods on synthesized clinical phantom. The location of the profiles is indicated by the green line in Fig. 3 for each reconstructed images in Bin 1.

C. Quantitative analysis

In this work, root-mean-square-error (RMSE) is calculated to quantify the errors between the ground truth and the results reconstructed by different methods. As shown in Table I and Table II, the presented GMM-3DTV method achieves consistently better metrics with smallest RMSE on both phantoms. In addition, it can be observed that the reconstruction performance is significantly heightened by fusing the GMM and 3DTV terms.

TABLE I
QUANTITATIVE MEASUREMENTS ON THE RECONSTRUCTION RESULTS ON XCAT PHANTOM FROM THE DIFFERENT METHODS

	RMSE ($\times 10^{-4}$)
FBP	14.066
3DTV	14.534
GMM	8.957
TDL	5.244
GMM-3DTV	**5.233**

TABLE II
QUANTITATIVE MEASUREMENTS ON THE RECONSTRUCTION RESULTS ON SYNTHESIZED CLINICAL PHANTOM FROM THE DIFFERENT METHODS

	RMSE ($\times 10^{-5}$)
FBP	11.528
3DTV	4.719
GMM	10.530
TDL	4.076
GMM-3DTV	**3.928**

IV. Discussion and Conclusion

Due to the beam-hardening and photon starvation effects, the reconstructed PCCT images suffer from noise-induced artifacts with complex noise distribution in image domain. Most of the SIR methods hardly handle the artifacts and produce suboptimal results. To address this issue, in this work, we presented a novel SIR PCCT reconstruction method. Specifically, we utilized GMM to approximate the complex noise distribution in the PCCT image domain. Moreover, a 3DTV term, which serves as the image prior, is also incorporated into the reconstruction model to encourage structural similarity of the PCCT images along the multi-energy bins. Experiments were conducted to demonstrate the effectiveness of the presented GMM-3DTV method. In the future, clinical and more scene studies would be included to further demonstrate the reconstruction performance of the presented GMM-3DTV method.

References

[1] M. J. Willemink, M. Persson, A. Pourmorteza, N. J. Pelc, and D. Fleischmann, "Photon-counting CT: Technical principles and clinical prospects," *Radiology*, vol. 289, no. 2, pp. 293-312, 2018.

[2] K. Taguchi, J. S. Iwanczyk, "Vision 20/20: Single photon counting x-ray detectors in medical imaging," *Medical Physics*, vol. 40, no. 10, p. 100901, 2013.

[3] S. Niu, G. Yu, J. Ma, and J. Wang, "Nonlocal low-rank and sparse matrix decomposition for spectral CT reconstruction,"*Inverse Problems*, vol. 34, no. 2, p. 024003, 2018.

[4] Q. Xu, H. Yu, J. Bennett, P. He, R. Zainon, R. Doesburg, and *et al.*, "Image reconstruction for hybrid true-color micro-CT," *IEEE Transactions on Bio-medical Engineering*, vol. 59, pp. 1711-1719, 2012.

[5] K. Kim *et al.*, "Sparse-view spectral CT reconstruction using spectral patch-based low-rank penalty," *IEEE Transactions on Medical Imaging*, vol. 34, no. 3, pp. 748-760, 2015.

[6] H. Zhang *et al.*, "Iterative reconstruction for dual energy CT with an average image-induced nonlocal means regularization," *Physics in Medicine & Biology*, vol. 62, no. 13, pp. 5556-5574, 2017.

[7] L. Yao *et al.*, "Multi-energy computed tomography reconstruction using a nonlocal spectral similarity model," *Physics in Medicine & Biology*, vol. 64, no. 3, Jan. 2019.

[8] J. Liu, H. Ding, S. Molloi, X. Zhang, and H. Gao, "TICMR: Total image constrained material reconstruction via nonlocal total variation regularization for spectral CT," *IEEE Transactions on Medical Imaging*, vol. 35, no. 12, pp. 2578-2586, 2016.

[9] S. Tao, K. Rajendran, C. H. McCollough, and S. Leng, "Material decomposition with prior knowledge aware iterative denoising (MDP-KAID)," *Physics in Medicine & Biology*, vol. 63, no. 19, 2018.

[10] D. Zeng, L. Yao, Y. Ge, S. Li, Q. Xie, H. Zhang, and *et al.*, "Full-spectrum-knowledge-aware tensor model for energy-resolved CT iterative reconstruction," *IEEE Transactions on Medical Imaging*, vol. 39, pp. 2831-2843, 2020.

[11] S. Li *et al.*, "Pseudo dual energy CT imaging using deep learning based framework: Initial study," 2017, *arXiv:1711.07118*. [Online]. Available:http://arxiv.org/abs/1711.07118

[12] W. Cong and G. Wang, "Monochromatic CT image reconstruction from current-integrating data via deep learning," 2017, *arXiv:1710.03784*. [Online]. Available: https://arxiv.org/abs/1710.03784

[13] G. J. McLachlan and K. E. Basford, "Mixture Models: Inference and Applications to Clustering," Marcel Dekker, 1988.

[14] S. Osher, M. Burger, D. Goldfarb, J. Xu, and W. Yin, "An iterative regularization method for total variation-based image restoration," *Multiscale Modeling & Simulation*, vol. 4, no. 2, pp. 460-489, 2005.

[15] A. P. Dempster, N. M. Laird, and D. B. Rubin, "Maximum likelihood from incomplete data via the em algorithm," *Journal of the Royal Statistical Society. Series B (methodological)*, pp. 1-38, 1977

[16] W. P. Segars, G. Sturgeon, S. Mendonca, J. Grimes, and B. M. W. Tsui, "4D XCAT phantom for multimodality imaging research," *Medical Physics*., vol. 37, no. 9, pp. 4902-4915, Aug. 2010.

[17] J. Punnoose, J. Xu, A. Sisniega, W. Zbijewski, and J. H. Siewerdsen, "Technical note: Spektr 3.0-A computational tool for X-ray spectrum modeling and analysis," *Medical Physics*, vol. 43, no. 8Part1, pp. 4711-4717, 2016.

Deep Learning Ring Artifact Correction in Photon-Counting Spectral CT with Perceptual Loss

Dennis Hein*‡, Konstantinos Liappis*‡, Alma Eguizabal†‡, and Mats Persson*‡
*Department of Physics, KTH Royal Institute of Technology, Stockholm, Sweden
†Department of Mathematics, KTH Royal Institute of Technology, Stockholm, Sweden
‡MedTechLabs, BioClinicum, Karolinska University Hospital, Solna, Sweden

***Abstract*—Photon-counting spectral CT is a novel technology with a lot of promise. However, one common issue is detector inhomogeneity which results in streak artifacts in the sinogram domain and ring artifacts in the image domain. These rings are very conspicuous and limit the clinical usefulness of the images. We propose a deep learning based image processing technique for ring artifact correction in the sinogram domain. In particular, we train a UNet using a perceptual loss function with VGG16 as feature extractor to remove streak artifacts in the basis sinograms. Our results show that this method can successfully produce ring-corrected virtual monoenergetic images at a range of energy levels.**

I. Introduction

Photon-counting spectral computed tomography (CT) is a promising novel technology for next-generation CT scanners [1], [2], [3]. Advantages of photon-counting detectors, compared to standard energy-integrating detectors, include higher contrast-to-noise ratio and spatial resolution, and improved low-dose imaging. One common issue in photon-counting CT is detector inhomogeneity, which results in energy threshold variation across detector elements, and, if not corrected for, leads to streak artifacts in the sinogram domain and ring artifacts in the image domain. This type of inhomogeneity can emerge due to an insufficiently calibrated forward model, temperature differences, and defective pixels [4]. Many methods have been suggested for artifact and noise reduction in CT imaging and lately there has been a shift towards deep learning as a way to tackle these problems [4], [5], [6], [7], [8], [9]. In this work, we add to this literature by training a deep neural network for ring artifact correction in the sinogram domain and demonstrating its effectiveness in reducing ring artifacts for virtual monoenergetic images at a range of energy levels in photon-counting spectral CT.

II. Method

A. Photon-counting spectral CT

1) Material decomposition: Consider a multi-bin system with $B > 2$ energy bins and, for simplicity, a 2-dimensional image space. The material decomposition starts with the ansatz that the X-ray linear attenuation coefficient $\mu(x, y; E)$ can be approximated by a linear combination of M basis materials

$$\mu(x, y; E) \approx \sum_{m=1}^{M} a_m(x, y)\tau_m(E), \tag{1}$$

where a_m and $\tau_m(E)$ are the basis coefficients and basis functions, respectively. It is usually resolved in the sinogram domain and thus the target variables are the material line integrals

$$A_m(\ell) = \int_\ell a_m(x, y)ds = \mathcal{R}(a_m), \tag{2}$$

where $\mathcal{R}$ denotes the Radon transform operator. The expected number of photons in energy bin j follows the polychromatic Beer-Lambert law

$$\lambda_j(\boldsymbol{A}) = \int_0^\infty \omega_j(E) \exp\left(-\sum_{m=1}^{M} A_m\tau_m(E)\right) dE. \tag{3}$$

This is our forward model. Finally, the measured data is the vector $\boldsymbol{y} := [y_1, ..., y_B]$ where for each j we assume that

$$y_j \sim \text{Poisson}(\lambda_j(\boldsymbol{A})). \tag{4}$$

Hence, the (non-linear) inverse problem is to map the photon counts $\boldsymbol{y}$ to the material line integrals $\boldsymbol{A} := [A_1, ..., A_M]$. The most common approach to this problem is maximum likelihood [10], [11], [12]. Setting up the objective as the log likelihood and simplifying yields

$$\begin{aligned} \min_{\mathbf{A}} \quad & \sum_{j=1}^{B} (\lambda_j(\boldsymbol{A}) - y_j \log(\lambda_j(\boldsymbol{A}))) \\ \text{s.t.} \quad & A_i \geq 0 \quad \forall i = 1, ..., M. \end{aligned} \tag{5}$$

This is subsequently solved using some iterative algorithm, e.g., the logarithmic barrier method [13].

2) Data generation: After generating numerical basis material phantoms (soft tissue, bone and iodine) by segmenting CT images from the KiTS19 dataset [9], photon-counting imaging was simulated using the `fanbeam` function in Matlab and a spectral response model of a photon-counting silicon detector [14] with 0.5×0.5 mm^2 pixels. The simulation was performed for 120 kVp and 200 mAs with 1579 detector pixels and 1600 view angles. After simulating Poisson noise, the maximum likelihood method was used for material decomposition of the simulated sinograms into bone and soft tissue basis sinograms, which were then reconstructed on a 583×583 pixel grid. To avoid streak artifacts due to photon starvation, a logarithmic barrier function was used to penalize large negative basis projection values. To simulate the effect of threshold variations, the simulation was performed with a

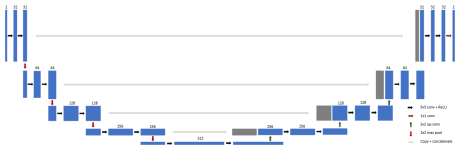


Fig. 1. Illustration of UNet [15].

random threshold shift ($\sigma = 0.5$ keV) applied independently to each of the eight thresholds of each detector pixel, and two material decompositions were performed: one with the actual bin thresholds that were used in the simulations, including the random shift, and one with the nominal bin thresholds, where the latter configuration yields images with ring artifacts.

B. Deep learning

1) Problem statement: We propose an image processing technique for ring artifact correction based on deep neural networks. More formally, let $\boldsymbol{x} \in \mathbb{R}^{M\times H\times W}$ denote the streak corrupted basis sinograms and $\boldsymbol{y} \in \mathbb{R}^{M\times H\times W}$ their streak artifact free counterpart, where M, H and W are the number of basis materials, view angles and detector pixels, respectively. Then our objective is to learn

$$f : \boldsymbol{x} \rightarrow \boldsymbol{y}. \tag{6}$$

We let $f_{\boldsymbol{\theta}}$ be a neural network and learn the map (6) by learning parameters $\boldsymbol{\theta}$.

2) Network architecture: UNet is an widely utilized architecture for a range of different tasks in biomedical imaging. The defining feature is the encoder-decoder structure. We use a version of the original UNet [15] shown in Fig. 1.

3) Loss functions: Mean square error (MSE) is perhaps the most commonly used loss function for applications of deep learning in biomedical imaging

$$\ell_{mse} = \frac{1}{MHW}||f_{\boldsymbol{\theta}}(\boldsymbol{x}) - \boldsymbol{y}||_2^2. \tag{7}$$

Using MSE loss will encourage the output to match the target pixel-by-pixel. This low-level per-pixel comparison is well known to produce output that is overly-smooth and lacking fine details that affect the perceptual quality [16], [17]. For several image transformation tasks, it has proved useful to instead employ a perceptual loss function which, instead of comparing pixel-by-pixel, compares high-level feature representations between the output and target. These feature representations are extracted from a pretrained convolutional neural network. We follow [16] and use VGG16 [18] pretrained on ImageNet [19] as feature extractor, or loss network. Let ϕ_j denote the j-th layer of VGG16, then our perceptual loss is defined as

$$\ell_{vgg} = \frac{1}{C_j H_j W_j}||\phi_j(f_{\boldsymbol{\theta}}(\boldsymbol{x})) - \phi_j(\boldsymbol{y})||_2^2, \tag{8}$$

where C_j is the number of channels in layer j. We will set $j = 9$ which corresponds to "relu2_2" in [16].

III. Training details

From each of the 1600×1579 basis sinograms we extract 20 256×256 patches. A total of 630 samples, yielding 12600 patches, are split 70/30 into a training and validation set. The network is trained using Adam [20] with $\beta_1 = 0.5, \beta_2 = 0.9$, and learning rate $\gamma = 1\times10^{-4}$ for 100 epochs with a batch size of 16 on one NVIDIA GeForce RTX 3070 Laptop GPU. We standardize the input by diving by the channel-wise standard deviation. We can obtain ring corrupted data with a range of different artifact magnitudes by taking a linear combination of streak corrupted and streak free basis sinograms. In this work, we are mainly concerned with the case when the rings are barely perceptible. Let w denote the weight given to the ring corrupted basis sinogram and $(1 - w)$ the weight given to its ring free counterpart. We found that $w = 0.4$ produces a realistic level for the artifacts and $w = 1$ a suitable level to train on.

IV. Results

A. Qualitative results

Qualitative results are available in Fig. 2 and 3. First, in Fig. 2, we have the results from the sinogram domain. Here, a pair of streak corrupted basis sinograms are passed through the network to produce the corresponding predicted pair. Note that despite training on 256×256 patches the network generalizes sufficiently to be able to deal with the entire 1600×1579 basis sinograms. The network does a fairly good job at removing the streaks. We subsequently reconstruct basis images from these sinograms and form virtual monoenergetic images at 40, 70, and 100 keV displayed in Fig. 3. Streak correction in the sinogram domain translates well into ring correction in the image domain. However, some residual rings are still visible. Note that, somewhat surprisingly, there is no significant difference in the performance of the network trained using MSE loss and the network trained using the perceptual loss.

B. Quantitative results

Quantitative results are available in table I. We employ the standard metrics used in this type of literature. Namely, structural similarity index measure (SSIM) [21] and peak signal-to-noise ratio (PSNR). However, we appreciate the fact that these are not necessarily great metrics of perceptual quality[1] and instead stress our qualitative results. Note that, surprisingly, the network trained with a perceptual loss achieves higher PSNR than the network trained with MSE loss. However, this difference is sufficiently small to reasonably be attributed to stochastic variation in the optimization procedure. We also investigate the resolution by adding a central circular insert in the KiTS19 phantoms and retrieving the edge spread function as an average over radial profiles in the region of interest (ROI). We then fit a Gaussian error function to estimate the resolution as its standard deviation. Both networks produce a slight decrease in resolution.

[1]See e.g., [16] for a brief discussion.

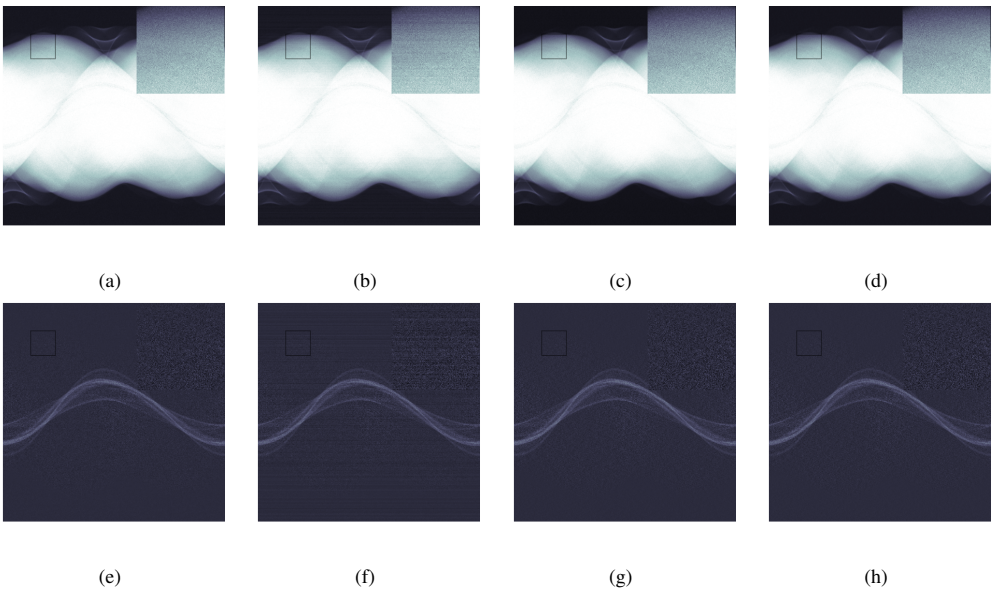


Fig. 2. Basis sinograms. The square in top right corner shows a magnification of the indicated ROI. (a-d) soft tissue, (a) truth, (b) observed, (c) observed + UNet_mse, (d) observed + UNet_vgg, (e-h) bone, (e) truth, (f) observed, (g) observed + UNet_mse, (h) observed + UNet_vgg.

TABLE I
QUANTITATIVE RESULTS

Network	SSIM	PSNR	Resolution (mm)
Truth	NA	NA	0.46
Observed	0.69	45.96	NA
UNet_mse	0.88	49.30	0.62
UNet_vgg	0.82	50.17	0.61

V. CONCLUSION

Detector inhomogeneity, a common issue in photon-counting spectral CT, results in streak artifacts in the sinogram domain and ring artifacts in the image domain. In this work, we propose a deep learning image processing technique for ring artifact correction in the sinogram domain. Artifact corrupted data is generated by solving the material decomposition problem with a correctly and an incorrectly calibrated forward model. We trained a deep neural network to remove the streaks in the basis sinograms, which are subsequently reconstructed to produce ring corrected basis images and virtual monoenergetic images. Instead of training a network to produce output that is similar to target pixel-by-pixel, we use a perceptual loss function that encourages the feature representation of the output to be similar to that of the target. Unexpectedly, we found that the network trained using the standard MSE loss essentially performs on par with the network trained using the perceptual loss. Future research will address the slight degradation in resolution caused by the networks, investigate why the networks perform so similarly, and further develop this method on a larger and more diverse dataset.

ACKNOWLEDGMENT

This study was financially supported by MedTechLabs and the Göran Gustafsson foundation. Mats Persson and Dennis Hein disclose research collaboration with GE Healthcare. Alma Eguizabal discloses consultancy with GE Healthcare.

REFERENCES

[1] E. Roessl and R. Proksa, "K-edge imaging in x-ray computed tomography using multi-bin photon counting detectors," *Physics in Medicine and Biology*, vol. 52, no. 15, pp. 4679–4696, jul 2007.

[2] M. J. Willemink, M. Persson, A. Pourmorteza, N. J. Pelc, and D. Fleischmann, "Photon-counting ct: Technical principles and clinical prospects," *Radiology*, vol. 289, no. 2, pp. 293–312, 2018, pMID: 30179101.

[3] M. Danielsson, M. Persson, and M. Sjölin, "Photon-counting x-ray detectors for CT," *Physics in Medicine & Biology*, vol. 66, no. 3, p. 03TR01, jan 2021.

[4] W. Fang, L. Li, and Z. Chen, "Removing ring artefacts for photon-counting detectors using neural networks in different domains," *IEEE Access*, vol. 8, pp. 42 447–42 457, 2020.

[5] Q. Yang, P. Yan, Y. Zhang, H. Yu, Y. Shi, X. Mou, M. K. Kalra, Y. Zhang, L. Sun, and G. Wang, "Low-dose ct image denoising using a generative adversarial network with wasserstein distance and perceptual loss," *IEEE transactions on medical imaging*, vol. 37, no. 6, pp. 1348–1357, 2018.

[6] B. Kim, M. Han, H. Shim, and J. Baek, "A performance comparison of convolutional neural network-based image denoising methods: The effect of loss functions on low-dose ct images," *Medical physics (Lancaster)*, vol. 46, no. 9, pp. 3906–3923, 2019.

[7] Z. Wang, J. Li, and M. Enoh, "Removing ring artifacts in cbct images via generative adversarial networks with unidirectional relative total variation loss," *Neural computing and applications*, vol. 31, no. 9, pp. 5147–5158, 2019.

[8] M. Nauwynck, S. Bazrafkan, A. V. Heteren, J. D. Beenhouwer, and J. Sijbers, "Ring artifact reduction in sinogram space using deep learning," in *6th International Conference on Image Formation in X-Ray Computed Tomography*, 2020.

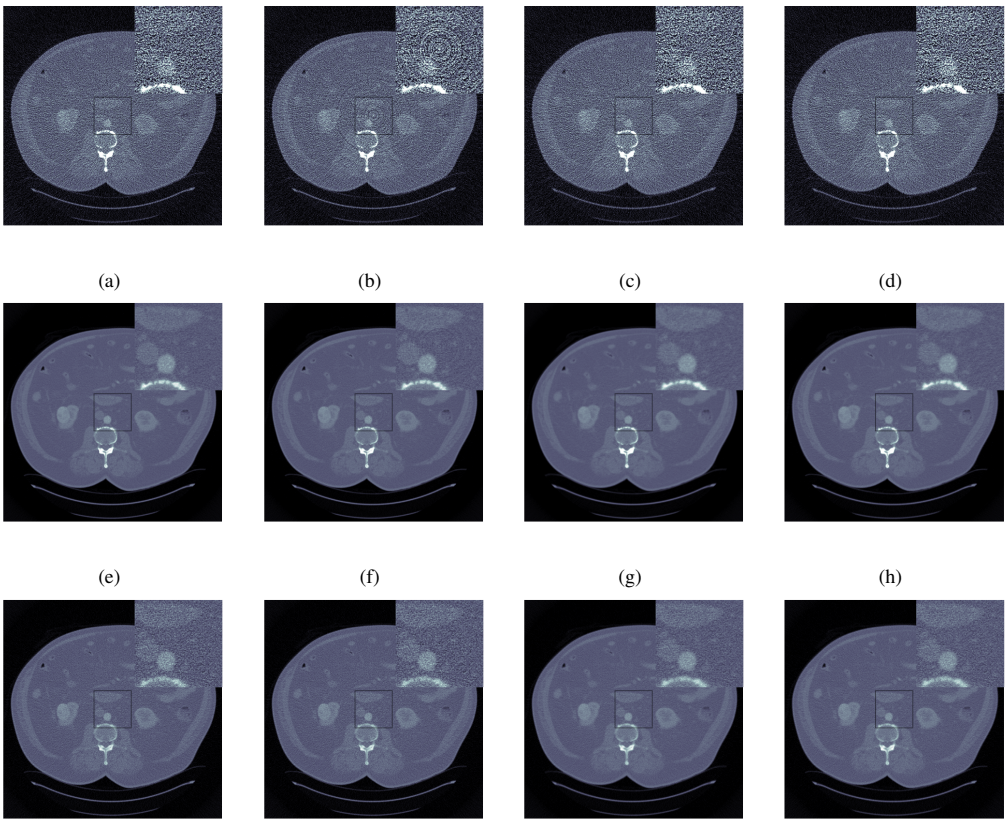


Fig. 3. Virtual monoenergetic images at 40, 70 and 100 keV. The square in top right corner shows a magnification of the indicated ROI. (a-d) 40 keV, (a) truth, (b) observed, (c) observed + UNet_mse, (d) observed + UNet_vgg, (e-h) 70 keV, (e) truth, (f) observed, (g) observed + UNet_mse, (h) observed + UNet_vgg, (i-l) 100 keV, (i) truth, (j) observed, (k) observed + UNet_mse, (l) observed + UNet_vgg.

[9] A. Eguizabal, M. U. Persson, and F. Grönberg, "A deep learning post-processing to enhance the maximum likelihood estimate of three material decomposition in photon counting spectral CT," in *Medical Imaging 2021: Physics of Medical Imaging*, H. Bosmans, W. Zhao, and L. Yu, Eds., vol. 11595, International Society for Optics and Photonics. SPIE, 2021, pp. 1080 – 1089.

[10] F. Grönberg, J. Lundberg, M. Sjölin, M. Persson, R. Bujila, H. Bornefalk, H. Almqvist, S. Holmin, and M. Danielsson, "Feasibility of unconstrained three-material decomposition: imaging an excised human heart using a prototype silicon photon-counting ct detector," *European radiology*, vol. 30, no. 11, pp. 5904–5912, 2020.

[11] R. E. Alvarez, "Estimator for photon counting energy selective x-ray imaging with multibin pulse height analysis," *Medical Physics*, vol. 38, no. 5, pp. 2324–2334, 2011.

[12] N. Ducros, J. F. P.-J. Abascal, B. Sixou, S. Rit, and F. Peyrin, "Regularization of nonlinear decomposition of spectral x-ray projection images," *Medical Physics*, vol. 44, no. 9, pp. e174–e187, 2017.

[13] S. Boyd and L. Vandenberghe, *Convex optimization*. Cambridge university press, 2004.

[14] M. Persson, A. Wang, and N. J. Pelc, "Detective quantum efficiency of photon-counting CdTe and Si detectors for computed tomography: a simulation study," *Journal of Medical Imaging*, vol. 7, no. 4, pp. 1 – 28, 2020.

[15] O. Ronneberger, P. Fischer, and T. Brox, "U-net: Convolutional networks for biomedical image segmentation," in *Medical Image Computing and Computer-Assisted Intervention – MICCAI 2015*, ser. Lecture Notes in Computer Science. Cham: Springer International Publishing, 2015, pp. 234–241.

[16] J. Johnson, A. Alahi, and L. Fei-Fei, "Perceptual losses for real-time style transfer and super-resolution," in *European Conference on Computer Vision*, 2016.

[17] C. Ledig, L. Theis, F. Huszár, J. Caballero, A. Cunningham, A. Acosta, A. Aitken, A. Tejani, J. Totz, Z. Wang, and W. Shi, "Photo-realistic single image super-resolution using a generative adversarial network," in *2017 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, 2017, pp. 105–114.

[18] K. Simonyan and A. Zisserman, "Very deep convolutional networks for large-scale image recognition," in *3rd International Conference on Learning Representations, ICLR 2015, San Diego, CA, USA, May 7-9, 2015, Conference Track Proceedings*, Y. Bengio and Y. LeCun, Eds., 2015.

[19] O. Russakovsky, J. Deng, H. Su, J. Krause, S. Satheesh, S. Ma, Z. Huang, A. Karpathy, A. Khosla, M. Bernstein, A. C. Berg, and L. Fei-Fei, "Imagenet large scale visual recognition challenge," *International journal of computer vision*, vol. 115, no. 3, pp. 211–252, 2015.

[20] D. P. Kingma and J. Ba, "Adam: A method for stochastic optimization," in *3rd International Conference on Learning Representations, ICLR 2015, San Diego, CA, USA, May 7-9, 2015, Conference Track Proceedings*, Y. Bengio and Y. LeCun, Eds., 2015.

[21] Z. Wang, A. Bovik, H. Sheikh, and E. Simoncelli, "Image quality assessment: from error visibility to structural similarity," *IEEE Transactions on Image Processing*, vol. 13, no. 4, pp. 600–612, 2004.

Photon Counting Detector-based Multi-energy Cone Beam CT Platform for Preclinical Small Animal Radiation Research

Xiaoyu Hu, Yuncheng Zhong, Kai Yang, and Xun Jia

***Abstract*—We reported our developments of a photon counting detector (PCD) based multi-energy cone beam CT (ME-CBCT) for preclinical small animal radiation research. The development was based on an existing preclinical small animal irradiator that includes a gantry mounted x-ray tube for imaging and radiation experiments in a self-shielded cabinet. We installed a PCD on the gantry and developed the data acquisition, processing, and image reconstruction pipeline to reconstruct CBCT images at three energy channels. We determined the optimal energy thresholds as 26, 56, and 90 keV to achieve uniform signal-to-noise ratio among energy channels. Pixel-based detector response calibration was performed to remove ring artifacts in the reconstructed CBCT images. The average difference between measured x-ray attenuation coefficients of targeted materials from ME-CBCT images and analytically calculated values was 10%. We decomposed the ME-CBCT images into images of water and bone material via an optimization model. The PCD-based ME-CBCT is expected to facilitate critical tasks in preclinical small animal irradiation researches, such as improved accuracy of radiation dose calculations in experiment planning.**

***Index Terms*—Photon counting detectors, Cone Beam CT, Small animal irradiation platform.**

I. Introduction

Dedicated image-guided small animal irradiation platform plays an essential role in advancing cancer radiation therapy and research [1]. In the current preclinical irradiation platforms, cone beam CT (CBCT) is widely used as an image guidance device to precisely guide the delivery of a radiation beam to the targeted area with sub-millimeter geometric accuracy [2]. Over decades, along with the advancements of preclinical radiobiology researches, there has been a strong growth of the desire for advanced and quantitative CBCT imaging to support tasks such as Monte Carlo-based radiation dose calculation in experimental planning [1], or identification of certain x-ray imaging contrast agents [3]. To meet these needs, we have previously developed multi-energy CBCT (ME-CBCT) imaging function on a SmART preclinical radiation platform (Precision X-ray Inc., North Branford, CT, USA) via a multiple-scan approach due to the restriction of using the existing flat-panel based CBCT imaging system [4]. However, the inevitably prolonged scan time leads to issues such as animal motion between scans that affects some imaging tasks such as reduced material decomposition accuracy, increased anesthesia use and hence the risks to the animal subject, as well as the increased x-ray radiation dose. One possible solution to overcome this limitation is photon-counting detector (PCD) based multi-energy CBCT. PCD has been asserted having a great potential in delivering more advanced CBCT functions than the widely-employed energy-integrating detectors (EIDs). Compared with EIDs, PCDs do not need to convert the photon into visible light but directly convert the x-ray photon energy to electrical charges in the form of electron-hole pairs [5]. The detected signal is then compared with a calibrated photon energy (threshold), and a count is added when the charge is greater than the specified threshold. This feature enabled by an application-specific integrated circuit (ASIC) allows the PCD-based CT system to simultaneously count photons above several selected energy thresholds, hence ME-CBCT in a single scan.

Over the years, significant efforts have been made in evaluating the performance of PCD-based imaging platforms [6] including preclincial imaging systems [7]. However, to our knowledge, PCD-based CBCT dedicated for preclinical radiation research has not been available, although several table-top systems have been previously built [8]. In this paper, we will report our recent developments of a PCD-based ME-CBCT on our preclinical small animal irradiator including the installation of the PCD on the gantry, the optimization of energy thresholds, and the development of a pipeline for data acquisition, processing, image reconstruction, and material decomposition.

II. Methods

A. ME-CBCT setup

The experiment in this work was based on the SmART preclinical radiation platform shown in Fig. 1, which included a rotating C-arm gantry and an small animal couch assembled in a self-shielded cabinet. The gantry can perform CBCT scanning using an x-ray tube (Comet iVario 225 kV, JME Ltd, Suffolk, UK) mounted on the gantry and an amorphous silicon flat-panel detector (FPD) (XRD 0820 AN3-ES, Perkin-Elmer, Wiesbaden, Germany). An XC-Thor Cadmium telluride (CdTe) PCD (Direct Conversion AB, Sweden) was mounted on an aluminum frame attached to the gantry. The PCD has an active sensor area of 51.2×100.0 mm^2 with a 0.1×0.1 mm^2 pixel size. The source to isocenter distance was 30.5 cm and

X. Hu, Y. Zhong, and X. Jia are with innovative Technology Of Radiotherapy Computations and Hardware (iTORCH) Laboratory, Department of Radiation Oncology, University of Texas Southwestern Medical Center, Dallas, TX 75390, USA. E-mail: Xun.Jia@UTsouthwestern.edu.

K. Yang is with the Division of Diagnostic Imaging Physics, Department of Radiology, Massachusetts General Hospital, Boston, MA 02114, USA. E-mail: KYANG11@mgh.harvard.edu

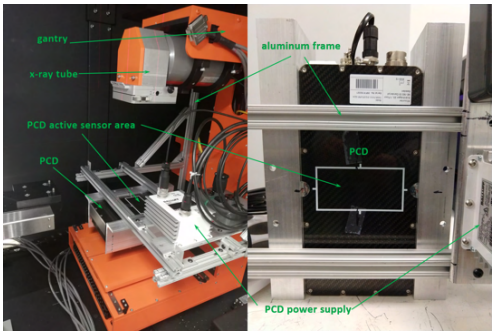


Fig. 1. Installation of the PCD on the SmART preclinical radiation platform (left) and the front view of of the PCD (right).

source to detector distance was 44.2 cm. This PCD was able to output the number of photon counts for each pixel above a user-specified energy threshold T for a time interval Δt.

We used a 100 kVp x-ray beam for CBCT data acquisition, with a rotation time of 1 min. The data acquisition of the PCD was controlled via the Application Programming Interface provided by the vendor. Specifically, we triggered the PCD to acquire projection data at a frequency of 30 frame/sec and each data frame receive photons for $\Delta t = 20$ msec. The remaining $\sim$ 13 msec time interval was allocated for data processing and readout. The energy threshold T was set to sequentially cycle through three energy thresholds T^l, $l = 1, \cdots, 3$. PCD operated under anti-coincidence high sensitivity mode for all acquisitions. For each projection, the projection angle was read out from the encoder of the gantry motor. With this setting, we acquired 600 projection images for each energy threshold.

The workflow covered by this study is summarized in Fig. 2. We will present details of key steps in subsequent sections.

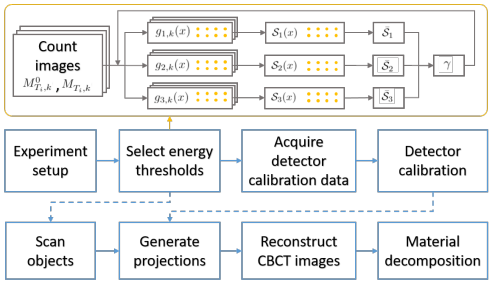


Fig. 2. Flowchart of PCD-based ME-CBCT imaging procedures.

B. Selection of energy thresholds

The lowest energy threshold T_1 was set to be 26 keV to avoid electronic noises [9]. The other two energy thresholds were determined to balance signal-to-noise ratio (SNR) in log-transformed projection images of all three energy channels, as the SNRs are directly related to those of the subsequently reconstructed images from each channel. As such, we acquired air projection images at 0 degree gantry angle under 3 mA for $N_t = 42$ thresholds T_i, $i = 1, \ldots, N_t$ ranged from 28 to 110 keV, with increment of 2 keV. Note that the thresholds were nominal input values to the PCD, therefore our search range went beyond the 100 kVp used. At every threshold T_i, we acquired $N_f = 200$ projections and denote the count image $M^0_{T_i,k}(x)$, where x is the index of each pixel, $k = 1, 2, ..., N_f$.

For a given threshold combination T^l, $l = 1, 2, 3$ selected among possible T_i, $i = 1, \ldots, N_t$ values, the count images at the three energy channels $E_1 = [T^1, T^2]$ keV, $E_2 = (T^2, T^3]$ keV and $E_3 = (T^3, 100]$ keV, were $C^0_{l,k}(x) = M^0_{T^l,k}(x) - M^0_{T^{l+1},k}(x)$ for $l = 1, 2$ and $C^0_{3,k}(x) = M^0_{T^3,k}(x)$. Note that we wrote the upper bound 100 keV in the third energy channel because of the 100 kVp x-ray beam, yet in practice this channel represents photons counts with energy above the threshold T^3. Following the same procedure, we acquired projection images of a CT calibration phantom, denoted as $C_{l,k}(x)$ for the energy channel l and frame k.

We then computed the log-transformed projection image $g_{l,k}(x) = -\ln[C_{l,k}(x)/\langle C^0_l(x)\rangle]$, where $\langle C^0_l(x)\rangle$ is the averaged counts of the air projection images over all frames. Using the repeatedly acquired 200 projections, we calculated the mean and standard deviation of $g_{l,k}(x)$, denoted as $\bar{g}_l(x)$ and $\sigma_{g_l}(x)$, respectively. Finally, the SNR at coordinate x was calculated as $\mathcal{S}_l(x) = \bar{g}_l(x)/\sigma_{g_l}(x)$. We averaged $\mathcal{S}_l(x)$ over 100 pixels inside the phantom region on the projection image, and denoted the averaged SNR as $\bar{\mathcal{S}}_l$. It was our objective to find proper thresholds such that $\bar{\mathcal{S}}_l$ approximately equal among all energy channels $l = 1, 2, 3$. As such, we considered the metric $\gamma = \max_l \bar{\mathcal{S}}_l / \min_l \bar{\mathcal{S}}_l$. We enumerated all possible combinations of thresholds T^2 and T^3, and found out the combination that minimized γ as the optimal thresholds for subsequent studies.

C. Detector calibration

To calibrate the non-uniform pixel responses of PCD, we acquired air scan under the threshold-sweeping mode as in the actual CBCT scan using the energy thresholds determined from Sec. II-B. We repeated this with x-ray tube current I in the range of [0, 3.4] mA with 0.2 mA increment. Let us denote the count image at energy threshold T^l, $l = 1, \ldots, 3$ with the jth tube current I^0_j, $j = 1, 2, ...$, as $M^0_{l,j}(x)$. Note that we repeatedly acquired 200 frames and took the average of these frames to reduce noise. Count images $C^0_{l,j}(x)$ at the three energy channels were computed by taking the difference between data with subsequent energy thresholds.

With the count images acquired for each energy channel $l = 1, 2, 3$ and tube current level $j = 0, 1, 2, ...$, for each pixel x and energy channel, we fit the tube current as a function of counts using I^0_j and $C^0_{l,j}(x)$ in a polynomial form. The coefficient of determination R^2 was used to evaluate the goodness of the fitting model. In the fitting procedure, we gradually increased the number of polynomial order, until the resulting R^2 reached $R^2_c = 0.9999$. If this was not achievable at 4^{th} order polynomial, the pixel was considered as a 'dead pixel'.

D. Image processing, reconstruction and material decomposition

After data acquisition of a CBCT scan, we converted raw count images $C_l(x, \theta)$ of energy channel l to the corresponding

images of current $I_l(x,\theta)$ using the established calibration model of each pixel. Here we explicitly write the projection angle θ. In this step, nearest-neighbor interpolation was applied to handle all 'dead pixels'. For each flat-fielded projection image at angle θ, we identified a region that was not blocked by the phantom and computed the averaged pixel value in this region as the air norm $I_l^0(\theta)$. Compared to using the nominal tube current in the scan as the air norm, this approach deriving $I_l^0(\theta)$ considered the tube current fluctuation among projections during the CBCT data acquisition.

After these stpes, we computed the projection image $g_l(x,\theta) = -\ln[I_l(x,\theta)/I_l^0(\theta)]$. Finally, CBCT image $\mathbf{f} = (f_1, f_2, f_3)^\top$ were reconstructed using the projection data and the GPU-based Feldkamp-Davis-Kress (FDK) reconstruction code [10], [11], where f_l, $l = 1, 2, 3$ represents reconstructed image of energy channel l.

In this study, we considered a two-material decomposition model that expressed each image voxel as a linear combination of two basis materials: water and bone. However, with three energy channels, in principle, it is possible to perform three-material decomposition. With our model, $\mathbf{f} = \mathbf{A}\mathbf{m}$, where $\mathbf{m} = (m_w, m_b)^\top$ is a two-component image representing mass density image of the two basis materials, and $\mathbf{A} \in \mathcal{R}^{(3\times 2)}$ is the system matrix. The two columns of $\mathbf{A}$ are x-ray mass attenuation coefficients of corresponding materials in the three energy channels, which was obtained in the reconstructed CT calibration phantom images. To solve for the material images without amplifying noise, we performed material decomposition by solving an optimization problem

$$\mathbf{m} = \underset{\mathbf{m}}{\operatorname{argmin}} \frac{1}{2}\|\mathbf{A}\mathbf{m} - \mathbf{f}\|_2^2 + \lambda\|\nabla\mathbf{m}\|_1, \qquad (1)$$

where the first term of the objective function ensured fidelity of the solution to the CBCT images, and the second one was a Total variation regularization term to reduce noise while preserving edges in the solution. λ is a weighting factor controlling the relative importance of the two terms. Note that the operator ∇ was applied to the spatial direction only. This model was solved with the Alternating Direction Method of Multipliers.

E. Evaluation

Two phantoms (SmART Scientific Solutions B.V., Maastricht, Netherlands) were utilized to calibrate the ME-CBCT system and evaluate the imaging procedure and image quality. The first one was a preclinical CT calibration insert phantom that contained 10 inserts of different materials plugged into a 30 mm diameter background base representing water. After the CBCT images were reconstructed, we picked two regions of interest with water and bone materials and used the average voxel values normalized by the known material density to construct the material decomposition system matrix $\mathbf{A}$. We then decomposed the CBCT images into the two material images using the model in Eq. (1). The second phantom was a plastinated mouse specimen. Due to the small size of the PCDs (5.0 cm along the rotation axis), only the head was scanned.

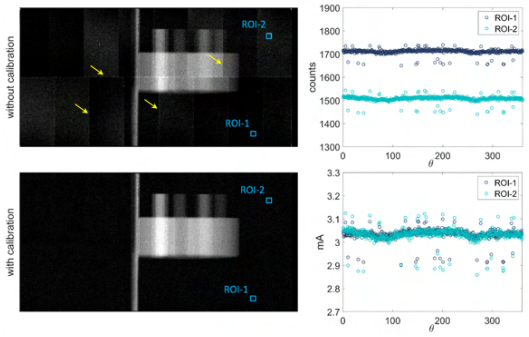


Fig. 3. Left: projection images at $\theta = 90$ degree in energy channel E_1 with and without detector calibrations; blue boxes show the location of two ROIs. Right-top: average counts of the two ROIs. Right-bottom: averaged mA of the two ROIs after detector calibration.

III. Results

A. Energy thresholds and detector calibration

With the method described in Sec. II-B, we computed the SNR ratios γ for all threshold combinations of (T^2, T^3). The SNR ratio that was closest to 1 was $\gamma = 1.05$, when the two energy thresholds were $T^2 = 56$ keV and $T^3 = 90$ keV. The corresponding SNRs of projection images were 7.07, 7.32 and 6.98 for the three energy channels, respectively.

Fig. 3 presents the impacts of detector calibration. Without calibration, we can clearly see different detector sub-panels to assemble the entire detector. The boundaries among panels would cause ring artifacts in the reconstructed images. We selected two regions of interest (ROIs) that were not blocked by the phantom. Due to flat x-ray beam fluence, the projection data in these two ROIs are expected to be the same. However, without calibration, we observed a large difference between the raw counts in the two ROIs across all projection angles. The calibration step was able to reduce this discrepancy. The fluctuation among projections was ascribed to x-ray tube output fluctuation, which was taken care of before image reconstruction by the choice of projection-specific air norm.

B. CBCT images and material decomposition

The reconstructed CBCT images of the CT calibration insert phantom and the mouse phantom with three energy channels are shown in Fig. 4. The object contrast reduced with increasing photon energies. For the calibration phantom, we compared measured x-ray attenuation coefficients and those computed by averaging the known x-ray attenuation data weighted by x-ray beam spectrum in the three energy channels. The difference was 10% on average. The difference can be ascribed to multiple factors including the reconstruction model, the discrepancy between nominal and actual energy thresholds, the ignored detector response variation over energy, etc.

To perform image-based material decomposition, we selected solid water and bone as two basis materials, and obtained the attenuation coefficients by averaging the voxel values over the corresponding ROIs indicated by the blue and red circles in Fig. 4. The mass attenuation coefficients

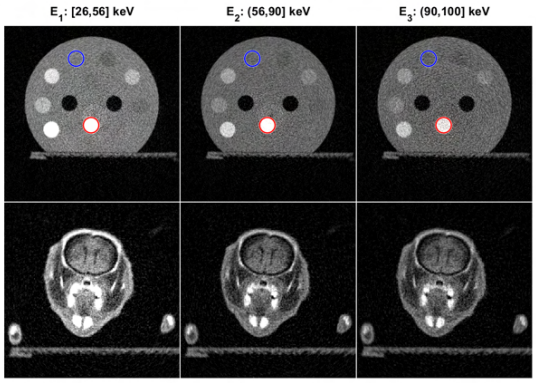


Fig. 4. CBCT images of the CT calibration insert phantom (top) and the plastinated mouse specimen (bottom) in the three energy channel. Display window: [0,0.5] cm^{-1}. The blue and red circles indicate the solid water and the bone basis materials.

in the three energy channels were $[0.240, 0.198, 0.186]$ cm^2/g and $[0.489, 0.296, 0.261]$ cm^2/g for water and bone materials, respectively.

Fig. 5 presents the material images of water and bone of CT calibration insert phantom and the plastinated mouse phantom in the axial plane. The two basis materials were differentiated clearly in the material images. Note that the animal skin also appeared in the bone image. The manufacture process of the plastinated mouse caused accumulation of polymer materials at the skin, which appeared as bone-like materials under x-ray imaging.

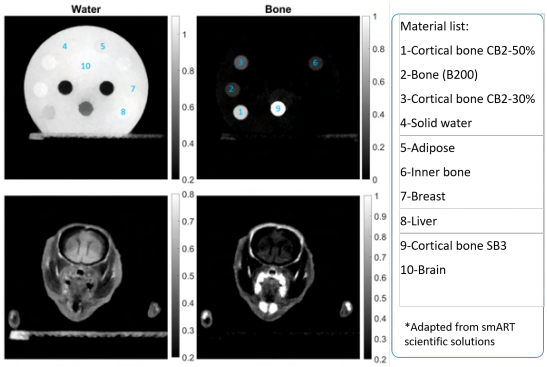


Fig. 5. Water (left) and bone (right) maps of the CT calibration insert phantom (top) and the plastinated mouse specimen in the axial plane (bottom).

IV. Conclusion and Discussions

In this work, we reported our initial study on a PCD-based CBCT for a preclinical small animal irradiation platform. We developed the data acquisition, processing, and image reconstruction pipeline to reconstruct CBCT images at three energy channels. We determined the optimal energy thresholds of 26 keV, 56 keV, and 90 keV to equalize SNR among energy channels. The average difference between measured x-ray attenuation coefficient from ME-CBCT images and calculated values was 10%. We decomposed ME-CBCT images into images of water and bone via an optimization model.

This study only serve as the preliminary step to enable PCD-based ME-CBCT function on a preclinical small animal irradiation platform. Despite the initial results achieved so far, there are a number of task down the road to further characterize and improve this system. For instance, the current CBCT reconstruction assumed an ideal circular trajectory, but the actual trajectory is known to deviate from this ideal case due to gravity. Geometry calibration is hence needed to improve geometry accuracy and resolution of reconstructed CBCT images. Additionally, the detector calibration can be improved via deep learning neural networks [12], [13]. The method of selecting energy thresholds proposed in this work was based on the objective of balancing the SNRs over all energy channels. In general, the choice of energy thresholds depends on the specific applications. One of the intended use of this ME-CBCT system is to visualize low-concentration gold nanoparticles, which requires optimization of energy thresholds for that task. Lastly, the utility of the CBCT system in terms of supporting preclinical radiobiology studies will be demonstrated in specific tasks, such as improvement in radiation dose calculations in experiment planning.

References

[1] F. Verhaegen, P. Granton, and E. Tryggestad, "Small animal radiotherapy research platforms," *Physics in Medicine & Biology*, vol. 56, no. 12, p. R55, 2011.

[2] R. Clarkson, P. Lindsay, S. Ansell, G. Wilson, S. Jelveh, R. Hill, and D. Jaffray, "Characterization of image quality and image-guidance performance of a preclinical microirradiator," *Medical physics*, vol. 38, no. 2, pp. 845–856, 2011.

[3] J. F. Hainfeld, F. A. Dilmanian, D. N. Slatkin, and H. M. Smilowitz, "Radiotherapy enhancement with gold nanoparticles," *Journal of pharmacy and pharmacology*, vol. 60, no. 8, pp. 977–985, 2008.

[4] Y. Huang, X. Hu, Y. Zhong, Y. Lai, C. Shen, and X. Jia, "Improving dose calculation accuracy in preclinical radiation experiments using multi-energy element resolved cone-beam ct," *Physics in Medicine & Biology*, vol. 66, no. 24, p. 245003, 2021.

[5] S. Fritz, "Characterization of photon counting czt detectors for medical x-ray imaging and spectroscopy," Ph.D. dissertation, Louisiana State University, 2011.

[6] S. Leng, W. Zhou, Z. Yu, A. Halaweish, B. Krauss, B. Schmidt, L. Yu, S. Kappler, and C. McCollough, "Spectral performance of a whole-body research photon counting detector CT: quantitative accuracy in derived image sets," *Physics in Medicine & Biology*, vol. 62, no. 17, 2017.

[7] J. P. Schlomka, E. Roessl, R. Dorscheid, S. Dill, G. Martens, T. Istel, C. Bäumer, C. Herrmann, R. Steadman, G. Zeitler, A. Livne, and R. Proksa, "Experimental feasibility of multi-energy photon-counting k-edge imaging in pre-clinical computed tomography," *Physics in Medicine and Biology*, vol. 53, no. 15, pp. 4031–4047, jul 2008.

[8] D. Richtsmeier, C. Dunning, K. Iniewski, and M. Bazalova-Carter, "Multi-contrast k-edge imaging on a bench-top photon-counting ct system: acquisition parameter study," *Journal of Instrumentation*, vol. 15, no. 10, p. P10029–P10029, 2020.

[9] T. Flohr, S. Ulzheimer, M. Petersilka, and B. Schmidt, "Optimization of k-edge imaging with spectral ct," *Chinese Journal of Academic Radiology*, vol. 2, pp. 19–34, 2020.

[10] L. A. Feldkamp, L. C. Davis, and J. W. Kress, "Practical cone-beam algorithm," *Josa a*, vol. 1, no. 6, pp. 612–619, 1984.

[11] X. Jia, Y. Lou, R. Li, W. Y. Song, and S. B. Jiang, "Gpu-based fast cone beam ct reconstruction from undersampled and noisy projection data via total variation," *Medical physics*, vol. 37, no. 4, pp. 1757–1760, 2010.

[12] M. Li, D. S. Rundle, and G. Wang, "X-ray photon-counting data correction through deep learning," 2020.

[13] C. Shen, D. Nguyen, Z. Zhou, S. B. Jiang, B. Dong, and X. Jia, "An introduction to deep learning in medical physics: advantages, potential, and challenges," *Physics in Medicine & Biology*, vol. 65, no. 5, p. 05TR01, 2020.

Design of Novel Loss Functions for Deep Learning in X-ray CT

Obaidullah Rahman, Ken D. Sauer, Department of Electrical Engineering, University of Notre Dame
Madhuri Nagare, Charles A. Bouman, School of Electrical and Computer Engineering, Purdue University
Roman Melnyk, Jie Tang, Brian Nett, General Electric Healthcare

***Abstract*—Deep learning (DL) shows promise of advantages over conventional signal processing techniques in a variety of imaging applications. The networks' being trained from examples of data rather than explicitly designed allows them to learn signal and noise characteristics to most effectively construct a mapping from corrupted data to higher quality representations. In inverse problems, one has options of applying DL in the domain of the originally captured data, in the transformed domain of the desired final representation, or both.**

X-ray computed tomography (CT), one of the most valuable tools in medical diagnostics, is already being improved by DL methods. Whether for removal of common quantum noise resulting from the Poisson-distributed photon counts, or for reduction of the ill effects of metal implants on image quality, researchers have begun employing DL widely in CT. The selection of training data is driven quite directly by the corruption on which the focus lies. However, the way in which differences between the target signal and measured data is penalized in training generally follows conventional, pointwise loss functions.

This work introduces a creative technique for favoring reconstruction characteristics that are not well described by norms such as mean-squared or mean-absolute error. Particularly in a field such as X-ray CT, where radiologists' subjective preferences in image characteristics are key to acceptance, it may be desirable to penalize differences in DL more creatively. This penalty may be applied in the data domain, here the CT sinogram, or in the reconstructed image. We design loss functions for both shaping and selectively preserving frequency content of the signal.

***Index Terms*—Deep learning, neural network, X-ray CT, novel loss functions, spectral shaping.**

I. Introduction

ARTIFICIAL neural networks (ANN) have been increasingly finding success in X-ray computed tomography (CT) [1]–[10]. ANN in imaging are designed by adjusting strengths of interconnections among artificial neurons with the goal of making the network's output, on the average, as close as possible to the ideal form of the image. This ideal form may be well known in training phase of the ANN, in which one may start with a perfect signal as the "target" and then corrupt it according to the character of noises and artifacts typically encountered in application. Alternatively, the target image may be imperfect, but far less afflicted with error than those encountered as measurements. In training, simple multiplicative coefficients or other representations of neural interconnections are iteratively adjusted to minimize some average measured error, or loss, between an ensemble of network-processed input data and their respective target images, as represented in Figure 1. The measured loss is backpropagated through the ANN to provide gradients to correct the connections and reduce loss, thus "learning" the inverse operator. Following training, the network may be applied to new data sets in order to reduce their content of error as described by the system's loss function. The process is, with increasing frequency, titled "deep learning" because more powerful computational resources have allowed more layers in the ANN, hence a "deeper" network.

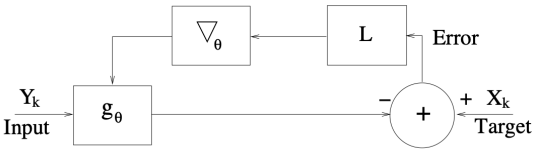


Fig. 1: Training of neural network. Parameters governing system behavior are denoted by θ. The gradient of the loss function's penalization of error (L), as a function of θ, is used to improve the averaged match between target and output of network during training.

Probably the most common loss function applied has been mean-squared error. Let us define Y as the input data, which we model as a function of some ideal, target image X, or $Y = h(X)$. The task of the ANN is to extract from Y a rendering close to the unknown, ideal image. If we define $g = h^{-1}$, our training would seek to learn g to produce $X = g(Y)$. Equality is seldom achievable due to noise or other corruption, and we optimize in the sense of average, possibly weighted, error. If we use the variable k to index among training pairs, n to index entries in vectors X_k and Y_k, and θ to represent the variable parameters of the ANN, our DL-trained mapping g_θ for the mean-squared error case may be expressed in terms of

$$\theta = \underset{\theta}{\mathit{argmin}} \sum_{k,n} w_{k,n}[X_{k,n} - (g_\theta(Y_k))_n]^2 \qquad (1)$$

in which the weightings $w_{k,n}$ may be fixed in either or both variables, or may be adapted according to relative local characteristics of data. This weighted, mean-squared penalty on the standard error, $S_k \triangleq X_k - g_\theta(Y_k)$, has a number of potential advantages, including being statistically well-matched to Gaussian noise. In cases where less severe penalization of large errors is desired, squared error may be replaced by absolute error, similarly to penalty adjustment in edge-preserving regularization.

While simple norms such as expressed above provide highly useful loss metrics, it has long been recognized in the image

processing community that they may be less than ideal for applications in which the final receiver for the system's output is a human observer. Various metrics for perceptual loss have been designed in hopes of optimizing the elusive human-interpreted quality of audio [1] and visual data [2]. For diagnostic CT imaging, in which much analysis is performed by radiologists, more subjective quality metrics are applied by the end users of the technology, and spectral content of residual noise, plateauing of image levels in low-contrast areas and other context-dependent evaluations must be addressed.

This work consists of a novel class of loss metrics which may expand the usefulness of DL in X-ray CT. We generalize the sense of optimality to

$$\hat{\theta} = \underset{\theta}{\mathit{argmin}} \sum_k L[X_k, Y_k, g_\theta(Y_k)], \tag{2}$$

where L is now a function that may capture any number of spectral and spatial characteristics in the error. In the X-ray CT arena, we may choose to improve the signal in either the sinogram domain, where measurements are made directly, or in the image domain after reconstruction by any existing algorithm. The signal and error statistics in these two differ, leading to designs tailored for each case. In the following, we describe one embodiment of the design.

II. Method

Conventional, point-wise mean-squared error as loss may be thought of as a flat spectral penalty. However, in cases where we wish to focus on removing artifacts with low or medium spatial frequency content, penalizing all frequencies equally may be counter-productive. Given that many well-developed, edge-preserving techniques are available for removing high-frequency noise, particularly in the image domain, low-signal correction in CT may in some cases be better served by training the network to remove errors only in lower frequencies. In this case, we propose a loss function L in eq. (3) that may take the form

$$L[S_k] \triangleq \phi[f_1(S_k)], \tag{3}$$

where ϕ is a suitable error metric applied only within the passband of the lowpass filter f_1. The higher frequency error becomes a "don't care" element for the network. Alternatively, band-pass or high-pass filtering may focus loss on those portions of the error spectrum. Particularly in three-dimensional image vectors, frequencies may be treated differently along the three axes. This forms the first part of our novel loss function.

The discussion above is most commonly addressed to conventional CT imagery in two or three dimensions, in which spatial frequency has roughly equivalent meaning in all dimensions. However, the present methods are intended at least as importantly for use in the native domain of the data, the sinogram. Application of the type of loss function in eq. (3) in the sinogram requires modeling behavior in such coordinates as row, channel and view, where the first two index in the detector panel of the CT gantry, and the last indexes the distinct rotating, two-dimensional views of patient or object. In this case, the error filtering operation will need to be spatially adapted, as statistics of both the underlying signal and the corrupting noise vary spatially in the sinogram domain.

It has been widely observed in the DL community that networks appear to have a strong tendency toward elimination of high frequencies in the output and this may occur even when the penalized loss is restricted to low frequency error as in eq. (3). An example application is using DL for low signal correction, where some of the most problematic artifacts are of low to medium spatial frequency. Here, it may be advantageous to retain parts of the error spectrum in the output when the correction network is applied in the sinogram domain. Powerful, adaptive denoisers in the image domain can capitalize on the relatively stationary underlying image statistics to remove higher frequency noise with little damage to edge resolution. Thus, we may wish to actively discourage suppression of this part of the error signal in the first stage of processing in order to preserve both resolution and desirable texture. We propose a second part of the loss function that will penalize removal of components of the signal Y_k according to their spectral content. This component of the loss may be expressed similarly to eq. (3), but with the argument redefined as

$$T_k \triangleq Y_k - g_\theta(Y_k) \tag{4}$$

$$L[S_k,\ T_k] \triangleq \phi[f_1(S_k) + \alpha f_2(T_k)], \tag{5}$$

An realization of the system is shown below in Figure 2. It includes the two loss functions discussed previously. The first loss, realized by the right branch, penalizes the error from eq. (3) filtered by f_1. The left branch features the error from eq. (4), where a different portion of the spatial frequency spectrum of error within the passband of f_2 is penalized. The two types of error signals are combined before the application of the norm ϕ and the gradient for backpropagation. The weighting factor α could be any positive value, with increase resulting in more of the desired frequency components preserved in the output. The responses of filters f_1 and f_2 plus the parameter

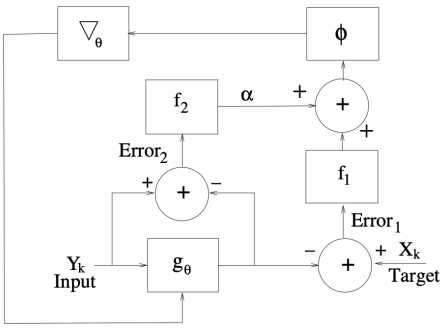


Fig. 2: Training of system to encourage the output to mimic the target content as selected by filter f_1, but refrain from removal of input signal content as selected by filter f_2

.

α appear to provide a great deal of control over the inference behavior of the network. In an extremely conservative case,

with $f_1 = f_2 = 1.0 \ \forall \ \omega$ and $\alpha = 1$, the composite error becomes

$$X_k + Y_k - 2g_\theta(Y_k), \tag{6}$$

which will simply place the optimum output midway between the target and the input.

III. Results

Parts of this method have been preliminarily tested with phantom and clinical data. Below are a few results with the latter. In this configuration, the training loss was the weighted sum of low pass (LP) filtered error between output and target, and high pass (HP) filtered error between input and output. The filters are shown in the Figure 3. The DL network was trained to operate in the original domain i.e. counts domains. Training data consisted of high-dosage Kyoto phantom scans as targets, with synthetic photon counting and electronic noise added to form input sinograms. We can see in Figure 4 the increase in fine-grain texture i.e. high frequency components in the output with increase in α.

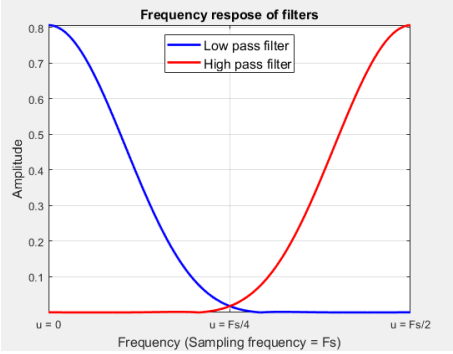
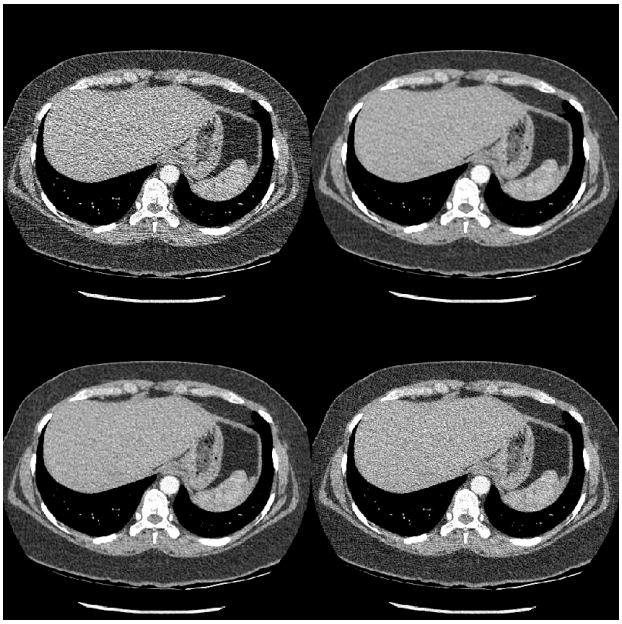


Fig. 3: Filters used. The LP filter is f_1 and the HP filter is f_2

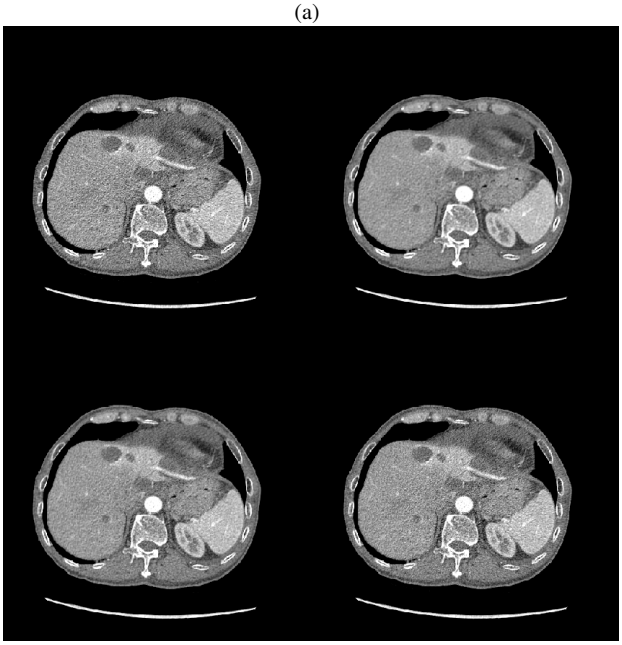
The noise power spectra (NPS), shown in Figure 5, were measured in the liver region of reconstructed clinical images. The NPS resulting from the use of only low pass in the loss function ($\alpha = 0$) can be seen to lack much high frequency content. Use of the high pass filter on the error between the input and the output preserves some of the high frequency components, retaining resolution along with high-frequency noise. The value of α can be adjusted based on the balance between NPS qualities and noise tolerance in the image. To assess the flatness of the NPS curve, entropy measurement was performed as

$$Entropy = \sum_{\omega_i=0}^{\omega_s/2} NPS(\omega_i) log_2 \frac{1}{NPS(\omega_i)}, \tag{7}$$

where ω_i is the discrete spatial frequency and ω_s is the spatial sampling frequency. It can be seen in Table I that the flatness of the NPS increases with α as far as 0.8, but it suffers from excessive high frequency emphasis for α of 1.0. This case exhibits undesirable streaks in the image as well.



(a)



(b)

Fig. 4: Reconstructed image (Upper left) Uncorrected; (Upper right) corrected with low pass filter loss ($\alpha = 0$); (Lower left) $\alpha = 0.6$; (Lower right) $\alpha = 0.8$. (a) Chest exam. (b) Liver exam

IV. Conclusion

This paper presents a combination of two frequency-weighted loss function components for a deep network, furnishing potentially better control of the behavior of the network in removing signal corruption. The first part of the DL loss function employed here restricts training loss to lower frequency error between a target data set and the input set

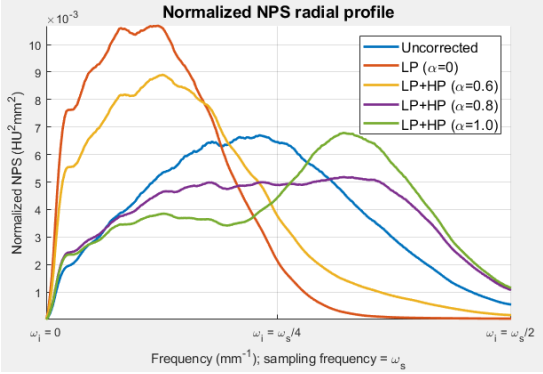


Fig. 5: Normalized NPS curves

Flatness metric (entropy in bits of information)					
Ideal case	Uncorrected	$\alpha = 0$	$\alpha = 0.6$	$\alpha = 0.8$	$\alpha = 1.0$
8.00	7.76	7.07	7.46	7.90	7.87

TABLE I: Entropy as a measurement of flatness of NPS curves. Higher value indicates flatter, more desirable NPS

processed by the network. The second component of the loss ensures the preservation of select error content from the uncorrected data, with the intent of delegating any removal of that error to a later stage of processing. This results in network's ability to retain desired traits in the data according to chosen models for training loss. In our example application, improvement in the texture of the reconstructed image was observed and confirmed with the NPS metric. Further work will test the value of this design in improving the noise/resolution trade-off in the presence of image-domain postprocessing. We have developed this novel DL loss function design for X-ray CT imaging, but it can easily find application in other areas.

REFERENCES

[1] Ishwarya Ananthabhotia, Sebastian Ewert, and Joseph A. Paradiso. "Towards a perceptual loss: Using a neural network codec approximation as a loss for generative audio models." Proceedings of the 27th ACM International Conference on Multimedia. 2019.

[2] Qingsong Yang, et al. "Low-dose CT image denoising using a generative adversarial network with Wasserstein distance and perceptual loss." IEEE transactions on medical imaging 37.6 (2018): 1348-1357.

[3] Kenji Suzuki, "Transforming projection data in tomography by means of machine learning." U.S. Patent Application No. 15/646,119.

[4] Qingsong Yang, et al. "CT image denoising with perceptive deep neural networks." arXiv preprint arXiv:1702.07019 (2017).

[5] Muhammad Usman Ghani and W. Clem Karl. "CNN based sinogram denoising for low-dose CT." Mathematics in Imaging. Optical Society of America, 2018.

[6] Tzu-Cheng Lee, Jian Zhou, and Zhou Yu. "Deep learning based adaptive filtering for projection data noise reduction in x-ray computed tomography." 15th International Meeting on Fully Three-Dimensional Image Reconstruction in Radiology and Nuclear Medicine. Vol. 11072. International Society for Optics and Photonics, 2019.

[7] Huizhuo Yuan, Jinzhu Jia, and Zhanxing Zhu. "Sipid: A deep learning framework for sinogram interpolation and image denoising in low-dose ct reconstruction." 2018 IEEE 15th International Symposium on Biomedical Imaging (ISBI 2018). IEEE, 2018.

[8] Ge Wang, Wenxiang Cong, and Qingsong Yang. "Tomographic image reconstruction via machine learning." U.S. Patent No. 10,970,887. 6 Apr. 2021.

[9] J.-B. Thibault, S. Srivastava, J. Hsieh, C.A. Bouman, D. Ye and K. Sauer, "Image Generation Using Machine Learning," US Patent 11126914, issued 21 Sept. 2021

[10] Obaidullah Rahman, Madhuri Nagare, Ken D. Sauer, Charles A. Bouman, Roman Melnyk, Brian Nett, and Jie Tang. "MBIR Training for a 2.5 D DL network in X-ray CT." in 16th International Meeting on Fully Three-Dimensional Image Reconstruction in Radiology and Nuclear Medicine, 2021, pp. 411–413

Effect of Attenuation Model on Iodine Quantification in Contrast-Enhanced Breast CT

Mikhail Mikerov, Koen Michielsen, James G. Nagy, and Ioannis Sechopoulos

***Abstract*—Accurate models of the x-ray attenuation process are required for quantitative estimation of iodine concentration with model-based reconstruction methods. The choice of model is influenced not only by the accuracy sought but also by the increasing complexity when more free parameters need to be reconstructed. The applicability of three attenuation models was investigated in a single pixel problem using either two or three monochromatic beams near the K-edge energy of iodine.**

We found that an empirical model with 5 components, proposed by Midgley, leads to the lowest error when modeling iodine free materials and small error in estimating iodine concentration (0.1% and 3.39%), whereas the decomposition into contributions due to photoelectric effect and incoherent scatter results in more accurate estimation of the iodine concentration (0.72%) but has larger error (8.9%) when reconstructing iodine free materials. Decomposition into base materials shows the worst results on both objectives (8.9% and 62%).

I. Introduction

Tumor characterization through quantitative functional imaging may allow for better treatment decisions in patients with breast cancer [1]. Dynamic contrast-enhanced breast CT is a new imaging modality being developed with aim to provide such functional information at good spatial and temporal resolutions. However, to maximize its clinical potential, accurate estimation of iodine concentration in the breast CT images is crucial. Current knowledge based on body CT imaging and computer simulations of contrast-enhanced breast CT indicate that the iodine concentration in the areas of interest, especially the tumor, can be expected to be in the range of 0.5 to 3.5 mg I per mL blood [2]. Coupled with sparse projections of typical breast CT systems and low photon energies that are required to increase the contrast (typically around 30 keV), the estimation of iodine concentration is a challenging task. In our implementation of contrast-enhanced breast CT, to save acquisition time and dose to the patient, individual projections are acquired only once with one of the x-ray spectra being used. This makes the use of decomposition methods in the projection domain not applicable.

Model-based methods are well suited to solve this reconstruction problem since they can use all available information about the acquisition, such as system geometry, utilized spectra, and physics models of attenuation processes. The latter determines, among other things, the number of free parameters that need to be estimated. Accurate modeling of the attenuation process is more difficult at low photon energies, where, in addition to photoelectric effect and incoherent scatter, coherent scatter also plays a role. Various parameterization schemes with different amounts of free parameters have been proposed to model attenuation.

In this work, we present the results of two experiments in which we examine the performance of three different parameterization schemes for energy dependent linear attenuation coefficients in a single pixel reconstruction problem. We focus on biological materials in the breast at x-ray energies below 49 keV and on accuracy of constrast quantification after adding iodine in the attenuation models, so we can determine which parameterization is most suitable to extend for our application, and include in our reconstruction method for quantitative breast CT imaging.

M. Mikerov, K. Michielsen, and I. Sechopoulos are with the Department of Medical Imaging, Radboudumc, Nijmegen, The Netherlands, e-mail: mikhail.mikerov@radboudumc.nl.

J. G. Nagy is with the Department of Mathematics, Emory University, Atlanta, GA, USA

I. Sechopoulos is with the Dutch Expert Centre for Screening (LRCB), Nijmegen, The Netherlands

I. Sechopoulos is with the Technical Medicine Centre, University of Twente, Enschede, The Netherlands

II. Methods

To avoid confounding influences, we focus on estimation of the energy dependent linear attenuation coefficient between 10 keV and 49 keV in a single pixel with monochromatic beams in dual and triple energy systems. Therefore, we are not solving the geometric aspect of the CT reconstruction problem, but are rather showing the adequacy of possible models of the energy dependency of linear attenuation.

A. Forward model

All values in the projection domain are obtained using the Beer-Lambert law:

$$p_E = I_0 \exp\left(-L \cdot \mu_E\right), \tag{1}$$

where μ_E is the linear attenuation coefficient at energy E, I_0 is the signal before attenuation, and L is the intersection length of the ray with the pixel of interest. The values of I_0 and L were set to 1 for all experiments.

B. Solution of linear systems

Limiting the estimation of the linear attenuation coefficient to a single pixel allows us to solve the linear system $A\mathbf{x} = \mathbf{b}$ using two analytical methods. The first is non-negative least squares, which is applied when the linear system has either full rank or is overdetermined [3]. We solve underdetermined systems of equations using the conjugate gradient method [3],

an iterative method that requires a good initial guess. It can be shown that for underdetemined systems of equations, conjugate gradient methods applied to $A^TA\mathbf{x} = A^T\mathbf{b}$, such as CGLS and LSQR, will converge to the minimum norm solution $A^T(AA^T)^{-1}\mathbf{b}$, making their application feasible [3].

C. Models of attenuation

We consider three different models for the energy dependent attenuation. The first model consists of a decomposition into base materials. In breast imaging, decomposition into adipose, fibro-glandular, and iodine components is the most evident choice. The system of equations for the single pixel problem in dual energy systems then takes the following form:

$$-\begin{pmatrix} \mu_a^L & \mu_g^L & \mu_i^L \\ \mu_a^H & \mu_g^H & \mu_i^H \\ -1 & -1 & -1 \end{pmatrix} \begin{pmatrix} f_a \\ f_g \\ f_i \end{pmatrix} = \begin{pmatrix} \ln p_L \\ \ln p_H \\ 1 \end{pmatrix} \tag{2}$$

The subscripts L and H refer to low and high energies. The last row puts a constraint on the otherwise underdetermined system of equations by enforcing conservation of volume, possibly causing undesirable behavior when the selected base materials are suboptimal to represent all expected tissues. When more than two spectra are used to acquire the images, additional rows can be added, making this system overdetermined.

The second model makes use of the physical processes underlying the attenuation of x rays. It decomposes the attenuation profile into contributions due to photoelectric effect and incoherent scatter. The energy dependency of the photoelectric effect is usually modeled using the power law

$$\Phi(E) = 1/E^3, \tag{3}$$

where E is the x-ray energy, whereas the Klein-Nishina equation is employed to describe incoherent (Compton) scatter

$$\begin{aligned} \Theta(E) =& \frac{(1+\alpha)}{\alpha^2}\left(\frac{2(1+\alpha)}{1+2\alpha} - \frac{\ln(1+2\alpha)}{\alpha}\right) \\ &+ \frac{\ln(1+2\alpha)}{2\alpha} - \frac{1+3\alpha}{(1+2\alpha)^2}, \end{aligned} \tag{4}$$

where $\alpha = E/511$ keV. However, this parameterization cannot model K-edges. Thus, in order to use this model to estimate the concentration of contrast agents, the attenuation characteristics of iodine must be included in the model. Consequently, the linear system of equations becomes:

$$-\begin{pmatrix} \phi^L & \Theta^L & \mu_{iodine}^L \\ \phi^M & \Theta^M & \mu_{iodine}^M \\ \phi^H & \Theta^H & \mu_{iodine}^H \end{pmatrix} \begin{pmatrix} f_\phi \\ f_\Theta \\ f_i \end{pmatrix} = \begin{pmatrix} \ln p_L \\ \ln p_M \\ \ln p_H \end{pmatrix}, \tag{5}$$

meaning that measurements at three different energies are required for the linear system to have full rank.

This model can be used in dual-energy setups following the method that was proposed by Depypere et al [4], [5]. If one can write ϕ and Θ as piece-wise linear functions of some parameter, e.g., the attenuation coefficient at a given energy, only two different measurements are needed. Hence, two parameters — the attenuation coefficient at a fixed energy and the iodine volume fraction — can be reconstructed using dual-energy setups. The difficulty of applying this method for breast imaging is that it is not possible to write the attenuation coefficient as a piece-wise linear function, as indicated in figure 1. Blood, skin, adipose and glandular tissues have very similar properties. Since any possible combination of materials in the triangle adipose-glandular-blood is theoretically possible, a systematic error is introduced as soon as the background tissue does not lie on the chosen line.

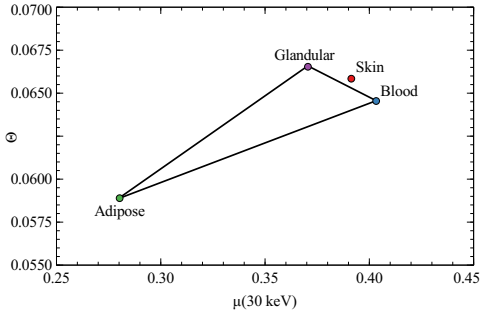


Fig. 1. Fitted contribution due to incoherent scatter as function of linear attenuation coefficient at 30 keV for materials present in the breast.

The third parameterization we included in our experiment was proposed by Midgley [6]. It decomposes the linear attenuation coefficient into energy-dependent S-parameters that are weighted by composition-dependent a-parameters:

$$\mu_E = \sum_{k=1}^{5} a_k S_k. \tag{6}$$

The number of S-parameters is not fixed and more parameters will lead to more accurate models. However, five S-parameters are sufficient to keep the maximum error below 2% in the energy range up to 50 keV [6]. Moreover, following that $a_{k+1} \geq a_k$, μ_E monotonically decreases with energy. The S-parameters for biological tissues in breast imaging are shown in figure 2. They were calculated by taking into account that a-parameters are functions of atomic number and electron density. Thus, the S-parameters are obtained by solving a least-squares problem [7], in our instance for elements that are found in breast tissues, specifically H, B, C, N, O, Na, Al, Si, P, S, Cl, Ar, K, Ca and Fe. As with the previous model, this parameterization scheme cannot account for K-edges. Thus, we extended this model to include the attenuation coefficient of iodine

$$\mu_E = \sum_{k=1}^{5} a_k S_k (1-f) + f\mu_{iodine}. \tag{7}$$

Assuming that the iodine volume fraction f is very small, the resulting system of equations in a dual-energy setup can

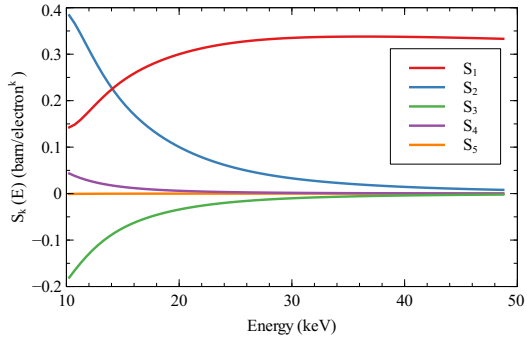


Fig. 2. The S-parameters were calculated using compositions of materials found in the breast.

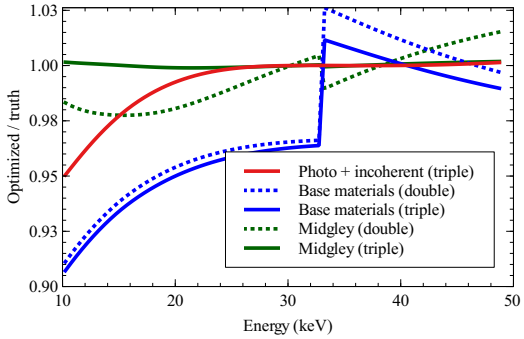


Fig. 3. Proportional error between reconstructed and true linear attenuation coefficient of skin when using different models of attenuation and number of spectra.

be written as

$$-\begin{pmatrix} S_1^L & S_2^L & S_3^L & S_4^L & S_5^L & \mu_{iodine}^L \\ S_1^H & S_2^H & S_3^H & S_4^H & S_5^H & \mu_{iodine}^H \end{pmatrix} \begin{pmatrix} a_1 \\ a_2 \\ a_3 \\ a_4 \\ a_5 \\ f \end{pmatrix} = \begin{pmatrix} \ln p_L \\ \ln p_H \end{pmatrix}. \tag{8}$$

D. Experiments

We examine the effect of the attenuation model on the accuracy of skin and iodine reconstruction in dual- and triple-energy acquisitions. We use skin as a good example of a background material that needs to be modeled correctly in order to avoid systematic errors that will otherwise propagate into the estimated iodine concentration in model-based reconstruction. Since model-based reconstruction tries to minimize the mismatch in the projection domain, incorrect linear attenuation coefficients in background material must be compensated by adjusting iodine concentration, which leads to lower accuracy.

The tissue background for the experiment including iodine contrast consists of a mixture of 50% adipose tissue, 50% fibro-glandular tissue; the volume fraction of blood is 30%. Varying amounts of iodine (0–20 mg/mL) were then added to the blood fraction. Our contrast agent is modeled as pure iodine. We assume that the attenuation of the contrast agent suspension medium is equal to that of blood. All materials were modeled with corresponding tissue substitutes [8]–[10]. and the attenuation profiles of elements were obtained from XrayDB [11].

The energies of the low and high energy spectra in the dual-energy setup were set to 30 keV and 34 keV, and the third energy for the triple-energy setup was set to 38 keV. Water was used as initialization for Midgley's parameterization. No initialization was needed for the other models since the resulting systems of equations were not underdetermined. The decomposition into photoelectric effect and incoherent scatter was considered only in the triple-energy setting.

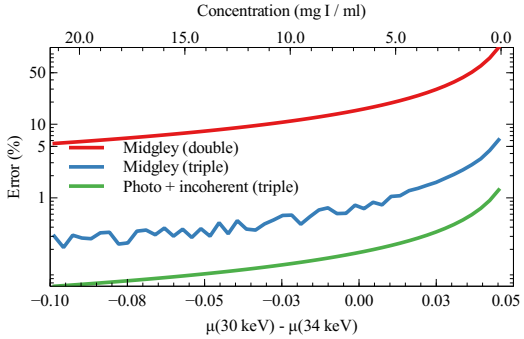


Fig. 4. Error in estimated iodine concentration as function of the difference between the linear attenuation of the mixture below and above the K-edge of iodine and concentration in blood. The negative region indicates rise in the attenuation due to K-edge. The background tissue was 50% adipose and 50% fibro-glandular tissues; volume fraction of blood was 30%.

III. Results

A. Reconstruction of skin

Figure 3 shows the relative error of the reconstructed attenuation coefficient of skin with three methods. The largest deviation (8.9% at 10 keV and $1.44{\cdot}10^{-4}$ volume fraction of iodine) results from the model that decomposes the background into adipose and glandular tissues. The most accurate result is achieved with Midgley's model in a triple-energy setup (0.1% at 10 keV and $-2.27{\cdot}10^{-6}$ volume fraction of iodine). Complete results are shown in table I.

TABLE I
MEAN AND MAXIMUM ABSOLUTE RELATIVE ERRORS ON ATTENUATION BETWEEN 10 KEV AND 49 KEV, AND RECONSTRUCTED IODINE VOLUME FRACTION FOR SKIN.

attenuation model	mean error	maximum error	iodine volume fraction
PE + Compton (double)	0.71%	4.95%	0.00
Base materials (double)	3.50%	8.88%	$1.44{\cdot}10^{-4}$
Base materials (triple)	3.50%	9.28%	$1.14{\cdot}10^{-4}$
Midgley (double)	1.13%	2.25%	$-3.80{\cdot}10^{-5}$
Midgley (triple)	0.07%	0.18%	$-2.27{\cdot}10^{-6}$

B. Estimation of iodine concentration

Figure 4 shows the error in iodine concentration estimation as a function of the difference between linear attenuation at 30 keV and 34 keV, resulting from the iodine K-edge when reconstructed using Midgley's model in dual- and triple-energy setups. The corresponding concentration of iodine is shown on top. The error is 62% and 3.39% at 0.97 mg I per mL of blood with double-energy and triple-energy setups, respectively. An even lower error (0.72%) is achieved with the decomposition into photoelectric effect and incoherent scatter with monochromatic radiation.

IV. Discussion

The aim of this study was to find the most accurate parameterization scheme for model-based reconstructions methods for breast CT with a focus on quantitative accuracy of the estimation of iodine concentration. Unsurprisingly, the base material decomposition into iodine, adipose and glandular tissues is the least accurate method to reconstruct skin. The error is substantial as it clearly shows that the iodine concentration in skin will not be zero. The accuracy improves only slightly when using a triple-energy system.

Decomposition into contributions due to photoelectric effect and incoherent scatter on the other hand, is very accurate above 25 keV. However, its accuracy drops noticeably in the lower energy range. This can be explained by the contribution due to coherent scatter in this low-energy range, which is absent from this model.

Finally, Midgley's parameterization in the triple-energy setup leads to the most accurate results for skin. The parameterization with three energies achieves better results than the one with two. In particular, triple energy optimization has negligible iodine signal in skin even though the linear system remains underdetermined. This is most likely explained by our initialization which assigns a good approximation of the relative contributions of the scatter processes. Parameters not well constrained in the problem would then not end up far from a reasonable value. Further examination of how the initialization with water holds up for calcifications will give an indication on the limitations of using a simple homogeneous initialization in patient images.

Our experiment shows that the decomposition into contributions due to photoloelectric effect and incoherent scatter is equally good as Midgley's parameterization at energies above 25 keV. However, the result shown in figure 3 represents the ideal case. Introduction of real polychromatic spectra will complicate the situation. Besides, one of the spectra must have mean average energy below the K-edge of iodine to capture it, which leads to nonzero fluence in the low energy range. Taking into account that the attenuation coefficients are highest in this range, it is favorable to avoid systematic errors that are introduced due to less accurate models when using model-based reconstruction methods. Nevertheless, the mono-energetic simplification is the main limitation of this study that needs to be addressed.

The triple-energy system leads to better results in the relevant concentration range in our application, i.e., below 3.5 mg I per mL blood. Notably, it is not enough to just have beam energies on either side of the K-edge. Accurate quantification requires the K-edge that causes an increase in attenuation at relevant energies, otherwise, the accuracy quickly decreases when the K-edge is not causing such increase as can be observed in figure 4. Since any allowed combinations of S-parameters are monotonically decreasing functions, the K-edge does not need to be modeled in order to connect the points on the linear attenuation profile at relevant energies. The continuous decrease in accuracy can be explained by different slopes of iodine before and after the K-edge, which means that there is still some information about iodine content present. The situation changes when the measurement at the third energy is added. Now, the only way to account for rise in attenuation and different slopes below and above the K-edge is to add the correct fraction of iodine.

V. Conclusion and outlook

In this simplified numerical study we have shown that the attenuation model influences accuracy of iodine reconstruction in contrast-enhanced breast CT at energies below 49 keV. Such models could lead to underdetermined linear systems. Nevertheless, an accurate solution can be found if triple-energy systems and appropriate initialization are used. In our future research, we will examine if these conclusions remain valid for polychromatic spectra before incorporating the preferred attenuation model in our model-based reconstruction for breast CT.

Acknowledgment

This study was supported by ERC grant 864929.

References

[1] J. Wu, G. Cao, X. Sun, J. Lee, D. L. Rubin, S. Napel, A. W. Kurian, B. L. Daniel, and R. Li, "Intratumoral spatial heterogeneity at perfusion mr imaging predicts recurrence-free survival in locally advanced breast cancer treated with neoadjuvant chemotherapy," *Radiology*, vol. 288, no. 1, pp. 26–35, 2018, pMID: 29714680. [Online]. Available: https://doi.org/10.1148/radiol.2018172462

[2] M. Caballo, R. Mann, and I. Sechopoulos, "Patient-based 4D digital breast phantom for perfusion contrast-enhanced breast CT imaging," *Medical Physics*, vol. 45, no. 10, pp. 4448–4460, Oct. 2018. [Online]. Available: https://onlinelibrary.wiley.com/doi/10.1002/mp.13156

[3] Å. Björck, *Numerical Methods for Least Squares Problems*. SIAM, 1996.

[4] M. Depypere, J. Nuyts, N. van Gastel, G. Carmeliet, F. Maes, and P. Suetens, "An iterative dual energy CT reconstruction method for a K-edge contrast material," N. J. Pelc, E. Samei, and R. M. Nishikawa, Eds., Lake Buena Vista, Florida, Mar. 2011, p. 79610M. [Online]. Available: https://doi.org/10.1117/12.878162

[5] B. De Man, J. Nuyts, P. Dupont, G. Marchal, and P. Suetens, "An iterative maximum-likelihood polychromatic algorithm for CT," *IEEE Transactions on Medical Imaging*, vol. 20, no. 10, pp. 999–1008, Oct. 2001. [Online]. Available: http://ieeexplore.ieee.org/document/959297/

[6] S. M. Midgley, "A parameterization scheme for the x-ray linear attenuation coefficient and energy absorption coefficient," *Physics in Medicine and Biology*, vol. 49, no. 2, pp. 307–325, Jan. 2004. [Online]. Available: https://iopscience.iop.org/article/10.1088/0031-9155/49/2/009

[7] ——, "A method for estimating radiation interaction coefficients for tissues from single energy CT," *Physics in Medicine and Biology*, vol. 59, no. 23, pp. 7479–7499, Dec. 2014. [Online]. Available: https://iopscience.iop.org/article/10.1088/0031-9155/59/23/7479

[8] "ICRP, 2009. Adult Reference Computational Phantoms. ICRP Publication 110. Ann. ICRP 39 (2)."

[9] D. R. White, J. Booz, R. V. Griffith, J. J. Spokas, and I. J. Wilson, "Report 44," *Journal of the International Commission on Radiation Units and Measurements*, vol. os23, no. 1, pp. NP–NP, 04 2016. [Online]. Available: https://doi.org/10.1093/jicru/os23.1.Report44

[10] J. W. Byng, J. G. Mainprize, and M. J. Yaffe, "X-ray characterization of breast phantom materials," *Physics in Medicine and Biology*, vol. 43, no. 5, pp. 1367–1377, may 1998. [Online]. Available: https://doi.org/10.1088/0031-9155/43/5/026

[11] "Xraydb," https://github.com/xraypy/XrayDB.

Motion Compensated Weighted Filtered Backprojection Considering Rebinning Process

Nora Steinich, Johan Sunnegårdh, and Harald Schöndube

***Abstract*—Motion during CT scans causes artifacts that can severely degrade the image quality. We propose a motion compensation algorithm that can be combined with reconstruction algorithms that contain a rebinning step like the weighted filtered backprojection algorithm. Therefore, we assume that the motion present during acquisition is known, and we extend the backprojection step in the reconstruction to consider this motion, reducing the motion artifacts in the resulting images. Furthermore, we propose a combination of our motion compensation algorithm with two versions of an iterative weighted filtered backprojection algorithm.**

***Index Terms*—Computed Tomography, Motion Compensation, Rebinning, Weighted Filtered Backprojection, Iterative Weighted Filtered Backprojection**

I. INTRODUCTION

MOTION during image acquisition in a CT scan causes artifacts that can severely degrade the image quality and lower the diagnostic value of the reconstructed images. The motion leads to an inconsistency in the acquired data and hence to artifacts like blurring, streaking or ghost images. The motion is often caused by patient movement, however it can also originate in motion of the scanner itself, for example in mobile scanner systems like a moving or sliding gantry. Since a complete prevention of motion during the scan can be difficult, a solution is required that compensates the present motion and thus improves the quality of the resulting images.

There are many algorithms proposed in the literature that attempt motion compensation in CT reconstruction. One method worth mentioning is the one by Hahn et al. [1], based on partial angle reconstruction. Another interesting motion compensation method is proposed by Bhagalia et al. [2].

In this paper we focus on cases where the motion present during acquisition, or an approximation thereof, is already known. The motion can either be measured with sensors or estimated using a motion estimation method like the one proposed by Bruder et al. [3].

A commonly used method for motion compensation in CT reconstruction with known motion is the method described by Schäfer et al. [4]. Their motion correction algorithm is applied in the reconstruction process during the backprojection step. Every voxel of the voxel volume that should be reconstructed is virtually shifted according to the motion present at the moment of acquisition. They included their method into the reconstruction algorithm of Feldkamp, Davis, and Kress [5] and achieved very good results.

However, there are several popular reconstruction algorithms, like the weighted filtered backprojection (WFBP) algorithm by Stierstorfer et al. [6] which rely on a rebinning step that, when combined with the method of Schäfer et al., will cause artifacts remaining in the final result.

In the following we propose an extension of the algorithm by Schäfer et al. that can also handle reconstruction algorithms containing a rebinning method. Furthermore, we propose the combination of the described motion compensation algorithm with an iterative reconstruction method.

N. Steinich, J. Sunnegårdh, and H. Schöndube are with the Siemens Healthineers, Forchheim, Germany.

II. METHODS

A. Motion Compensated Weighted Filtered Backprojection

In this paper we combine the motion compensation method proposed by Schäfer et al. [4] with the WFBP algorithm of Stierstorfer et al. [6], a 3D filtered backprojection algorithm for multislice spiral CT. However, our proposed motion compensation method works with any CT reconstruction algorithm using rebinning and backprojection steps.

In the rebinning step of the WFBP algorithm the projection images acquired in the cone beam geometry are virtually rearranged to form semi-parallel projection images better suited for the backprojection step. The cone beam geometry is described using α as the rotation angle, β as the horizontal opening angle of a ray and q as the cone parameter describing the vertical cone angle of a ray. For the semi-parallel projection images the rotation angle is described as θ and (p, q) denote the detector columns and rows, respectively. The geometry is only semi-parallel, as the rays are only parallel along the detector rows, but not along the columns. In the rebinning step the rays of the initially obtained cone beam projection images $P(\alpha, \beta, q)$ are rearranged to form semi-parallel projection images $P(\theta, p, q)$ using the rebinning formulas

$$\theta = \alpha + \beta \qquad p = R \sin(\beta) \tag{1}$$

with R describing the distance from source to isocenter.

The backprojection step in the WFBP algorithm iterates through all rebinned and filtered projection images $P(\theta, p, q)$ and for each projection image it iterates through all voxels $\boldsymbol{x}$ of

the voxel volume to be reconstructed. For each voxel $\boldsymbol{x}$, the corresponding position (p, q) in the projection image $P(\theta, p, q)$ the voxel has been projected to during acquisition is calculated. The value at position (p, q) is then added to voxel $\boldsymbol{x}$ in the voxel volume. This is performed for all voxels and all projection images until the entire voxel volume has been reconstructed.

For our motion compensation method we assume that the motion present during acquisition is known. For every projection image $P(\alpha, \beta, q)$ a motion state $\boldsymbol{M}_\alpha$ is required. Each motion state $\boldsymbol{M}_\alpha$ holds the information about the position of the scanned object relative to the scanner at the moment of the acquisition of $P(\alpha, \beta, q)$.

Our proposed algorithm is based on the method by Schäfer et al. [4], extending it to work with reconstruction algorithms containing a rebinning step like the WFBP algorithm. When combining the WFBP algorithm with Schäfer's method the backprojection step of the reconstruction algorithm is extended by moving every voxel $\boldsymbol{x}$ in the voxel volume to be reconstructed according to the known motion pattern. As described above, the backprojection is performed by iterating for each projection image $P(\theta, p, q)$ through all voxels $\boldsymbol{x}$. Knowing the motion state $\boldsymbol{M}_\alpha$ present during the acquisition of the projection image $P(\alpha, \beta, q)$ obtained at rotation angle α, we can use this information to move each voxel $\boldsymbol{x}$ accordingly, assuming $\alpha = \theta$. For each projection image $P(\theta, p, q)$ each voxel position $\boldsymbol{x}$ is recalculated to $\boldsymbol{x}_{\boldsymbol{M},\alpha}$ considering the motion state $\boldsymbol{M}_{\alpha=\theta}$. Using the new position $\boldsymbol{x}_{\boldsymbol{M},\alpha}$ the corresponding position (p, q) in the projection image $P(\theta, p, q)$ is calculated and the value of this position is added to the value at voxel position $\boldsymbol{x}$. This is then repeated for all voxels in the voxel volume and for all obtained projection images. This leads to a more accurate reconstruction, as the motion present during acquisition is considered. However, when combining the WFBP algorithm with the motion compensation method proposed by Schäfer et al. in this manner, some motion artifacts remain in the final result, as the rebinning process is not considered.

Due to the rebinning the assumption that for every projection image $P(\theta, p, q)$ one motion state $\boldsymbol{M}_\alpha$ needs to be considered during backprojection is not accurate. This assumption only holds for the projection images $P(\alpha, \beta, q)$. The rebinning process however rearranges the rays such that the rebinned projection images each consist of information obtained at different time points and hence depicting different motion states. The available motion state $\boldsymbol{M}_\alpha$ is no longer correct for the entire projection image $P(\theta, p, q)$, but only for the column where $p = 0$, as only there $\theta = \alpha$ holds. Using our geometry, the columns right of the column $p = 0$ contain previous motion states and columns towards the left contain subsequent motion states. Hence using a motion state $\boldsymbol{M}_\alpha$ for an entire projection image $P(\theta, p, q)$ in the backprojection will lead to motion artifacts remaining in the final result.

We propose a motion compensation method as described in algorithm 1 extending Schäfer's method to consider the rebinning process, enabling a combination with the WFBP reconstruction algorithm.

Algorithm 1 Motion Compensated WFBP

```
for all projection images P(θ,p,q) do
    for all voxels x of the voxel volume do
        1) Calculate new voxel position x_{M,θ} using M_{α=θ}
        2) Calculate (p,q) using x_{M,θ}
        repeat
            3) Calculate M' using p/p'
            4) Calculate x'_{M,θ} using M'
            5) Calculate (p',q') using x'_{M,θ}
        until convergence
        6) Add value at position (p',q') to voxel x
    end for
end for
```

Steps 1) and 2) are the steps as described in the method of Schäfer et al., while steps 3)-6) are our proposed extension. Our method uses the motion state $\boldsymbol{M}_{\alpha=\theta}$ to calculate a new position $\boldsymbol{x}_{\boldsymbol{M},\theta}$ of the voxel $\boldsymbol{x}$, which is then used to calculate an initial position (p, q) in the projection image $P(\theta, p, q)$ just as described by Schäfer et al. This position (p, q) is then used to choose a more accurate motion state $\boldsymbol{M}'$. Using the column position p and the rebinning formulas described in equation (1) the cone beam projection image $P(\alpha, \beta, q)$ from which the information in this column p was taken in the rebinning and hence which motion state $\boldsymbol{M}'$ was present can be calculated. With this new more accurate motion state $\boldsymbol{M}'$ the voxel position $\boldsymbol{x}$ can be recalculated to $\boldsymbol{x}'_{\boldsymbol{M},\theta}$, followed by the calculation of the new position (p', q') in the projection image $P(\theta, p, q)$. This more accurate position (p', q') can then be used to write its values to the voxel $\boldsymbol{x}$. The steps 3) to 5) can be repeated to achieve even more accuracy, however experiments have shown that even one iteration is able to yield good results.

For a more efficient implementation, we propose a modification of the algorithm stated above. Instead of using the exact motion state $\boldsymbol{M}'$ that was calculated for every position p in the projection image $P(\theta, p, q)$, we suggest choosing 3 motion states per projection image, whilst interpolating the motion states in between. Specifically, we chose to use the motion state at position $p = 0$ which corresponds to the initial motion state $\boldsymbol{M}_{\alpha=\theta}$ as well as one motion state for one column left and one column right of $p = 0$. However, we are not choosing the left and rightmost columns in the projection image $P(\theta, p, q)$, but the columns furthest to the left and right that are actually reached in a projection. These columns need to be calculated individually, as they depend on several parameters. For the three chosen columns the corresponding motion states are calculated using the rebinning equations (1). For a position (p, q) in the projection image $P(\theta, p, q)$ the corresponding motion state is then obtained from the motion states of the two closest chosen columns using linear interpolation of the motion values.

B. Motion Compensated Iterative Weighted Filtered Backprojection

The method described above can also be combined with an iterative weighted filtered backprojection algorithm, as the one

described by Sunnegårdh et al. [7]. The iterative reconstruction algorithm can be described with the vectors $\boldsymbol{p}$ and $\boldsymbol{f}$ denoting the projection data and the resulting image data, and the matrices $\boldsymbol{Q}$ and $\boldsymbol{P}$ denoting backprojection and forward projection operators, respectively. In our case, $\boldsymbol{Q}$ is the weighted filtered backprojection by Stierstorfer et al. [6] and $\boldsymbol{P}$ is the forward projection by Joseph [8]. The iterative reconstruction process can then be described as

$$\boldsymbol{f}_{k+1} = \boldsymbol{f}_k - \alpha \boldsymbol{Q}(\boldsymbol{P}\boldsymbol{f}_k - \boldsymbol{p}). \tag{2}$$

An initial result $\boldsymbol{f}_0 = \boldsymbol{Q}\boldsymbol{p}$ is calculated by backprojecting the rebinned and filtered projection data. The initial result $\boldsymbol{f}_0$ is then forward projected and the difference to the initial projection data $\boldsymbol{p}$ is calculated. This difference image is backprojected and subtracted from the previous result $\boldsymbol{f}_0$ using a factor α. This is repeated multiple times, improving the result with every iteration.

Including the proposed motion compensation method into this iterative reconstruction process consists of two parts. First, the motion compensation is included into the backprojection operator $\boldsymbol{Q}$ as described above. Second, the motion also needs to be considered in the forward projection operator $\boldsymbol{P}$. However, in the forward projection the motion is included again rather than removed, as the forward projected result is compared to the original not motion compensated projection data $\boldsymbol{p}$. Including the motion in the forward projection operator $\boldsymbol{P}$ is performed similar to the motion compensation in the backprojection operator $\boldsymbol{Q}$. In the forward projection, we are assuming an implementation that iterates through all positions (p, q) in the projection image $P'(\theta, p, q)$ that should be calculated. For each position (p, q), the corresponding ray through the voxel volume is calculated and the values along its path are accumulated. This is repeated for all projection images $P'(\theta, p, q)$. Knowing the position (p, q) the corresponding motion state at this position can be interpolated as described above. The motion state can then be included in the calculations of each ray changing its path through the voxel volume accordingly. The value accumulated along the rays path is written to the initial position (p, q) of the projection image $P'(\theta, p, q)$. However, the motion compensation calculations need to be included in every backprojection and every forward projection step of every iteration.

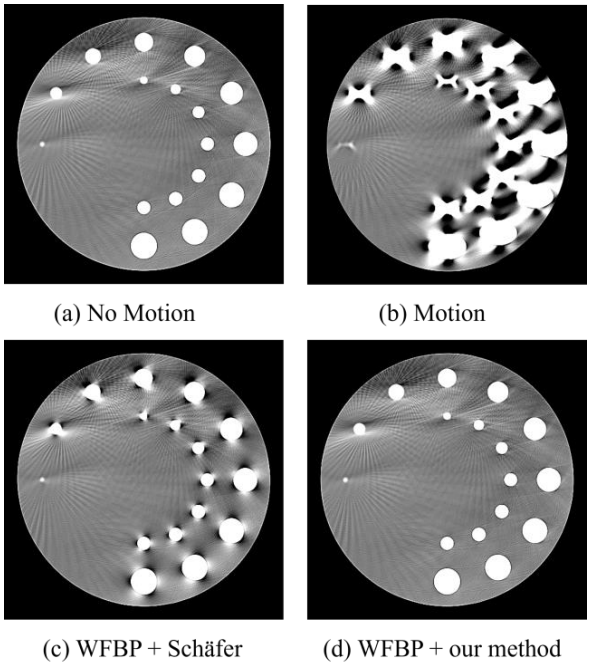
Hence, we propose to use a modification of the iterative reconstruction algorithm used above, as the one described by Sunnegårdh et al. [9], that allows for a more efficient implementation of the proposed motion compensation algorithm. The algorithm stated in equation (2) is adapted by considering the combination of forward projection followed by backprojection to be the sum of a low-pass filter $\boldsymbol{L}$ and a linear operator $\boldsymbol{A}$ that causes artifacts:

$$\boldsymbol{Q}\boldsymbol{P} = \boldsymbol{L} + \boldsymbol{A}. \tag{3}$$

Hence the iterative reconstruction equation (2) can be rewritten to

$$\boldsymbol{f}_{k+1} = \boldsymbol{f}_0 - (\boldsymbol{Q}\boldsymbol{P} - \boldsymbol{L})\,\boldsymbol{f}_k \tag{4}$$

as described in more detail by Sunnegårdh et al. [9]. This modification allows to include the motion compensation algorithm only in the first backprojection step, without the need to repeat the motion compensation calculations in every iteration. Furthermore, there is no need to adapt the forward projection operator to consider motion.



(a) No Motion (b) Motion

(c) WFBP + Schäfer (d) WFBP + our method

Fig. 1. Example reconstructions of the Turbell clock phantom (a) without introducing motion, (b) introducing a translation in x-direction performing a 5mm sinus without motion compensation, (c) introducing the same motion and compensating it with the combination of Schäfer's method and the WFBP algorithm, and (d) introducing the same motion and compensating it with our proposed method. Greyscale window C: 0 HU, W: 100 HU.

III. Results

To evaluate the proposed motion compensation algorithm the Turbell clock phantom [10] is used, simulating a helical cone beam CT scan, followed by reconstruction with and without motion compensation.

To evaluate the motion compensation the phantom was simulated to be moving along a predefined path. In this case a rigid motion was introduced, moving the entire phantom via translation, rotation, and moving of the rotation center. For the acquisition of each projection image $P(\alpha, \beta, q)$ a motion state $\boldsymbol{M}_\alpha$ of the phantom was defined, consisting of values for translation, rotation, and the position of the rotation center. The values of all motion states $\boldsymbol{M}_\alpha$ used for the simulation are saved for later use in the motion compensated reconstruction as it is assumed throughout this paper that the motion present during acquisition is known.

Fig. 1 (a) shows the phantom simulated without any motion and reconstructed using the WFBP algorithm [6] as reference.

Fig. 1 (b) shows the WFBP reconstruction of the phantom for a simulated scan including motion. The motion included in this example is a translation in x-direction describing a sinus curve with an amplitude of 5mm and one oscillation per scanner rotation. Fig. 1 (c) shows the reconstruction of a scan of the phantom with the same motion, now including the motion compensation method proposed by Schäfer et al. [4] but combined with the WFBP algorithm. Fig. 1 (d) shows the reconstruction result of the scan simulated with the same motion, using our proposed motion compensation method. In this case only one iteration of steps 3) to 5) of our proposed algorithm was performed. It can be seen that for the combination with the WFBP algorithm our proposed method shows significant improvement regarding the motion artifact reduction, when compared to the combination of the WFBP algorithm with Schäfer's method, even if only one iteration of our proposed algorithm is performed.

Fig. 2 shows the results for the two motion compensated iterative reconstruction methods. Again Fig. 2 (a) and (b) show the reference images without and with the introduced motion, the same translation introduced in the experiments shown in Fig. 1. Fig. 2 (c) shows the result for the iterative reconstruction where the motion compensation is included in every backprojection and every forward projection step. Fig. 2 (d) shows the iterative reconstruction result, for the version, where the motion compensation is only included in the first backprojection step.

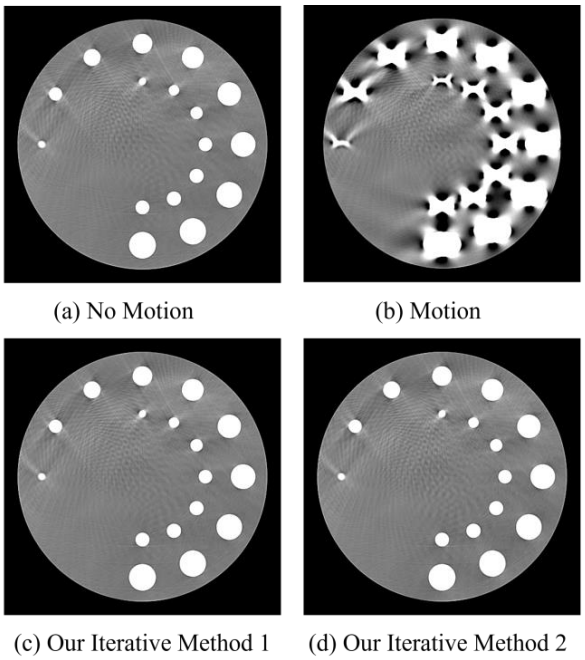


Fig. 2. Iterative reconstructions of the Turbell clock phantom (a) without introducing motion, (b) introducing a translation in x-direction performing a 5mm sinus without motion compensation, (c) introducing the same motion and compensating it with the motion being considered in every backprojection and every forward projection, and (d) introducing the same motion and compensating it only in the first backprojection. Greyscale window C: 0 HU, W: 100 HU.

IV. Conclusion

With their motion compensation algorithm Schäfer et al. [4] proposed a solution for motion artifacts in CT images that yields very good results, if the motion present during acquisition is known. However, they proposed the combination of their motion compensation with the reconstruction algorithm of Feldkamp, Davis, and Kress [5]. Combining their method with reconstruction algorithms containing a rebinning step like the WFBP algorithm [6] does not yield similarly good results as different motion states are mixed.

In this paper we propose an extension of the method by Schäfer et al. that considers the rebinning process in the backprojection step, making a combination with the WFBP reconstruction algorithm possible. Furthermore, we propose the combination of our motion compensation method with iterative weighted filtered backprojection.

Our experiments show promising motion compensation results both with the WFBP algorithm and the iterative reconstruction algorithm that are comparable to the results achieved with the method by Schäfer et al. combined as suggested with the reconstruction algorithm of Feldkamp, Davis, and Kress.

In our experiments we focused on the introduction and compensation of rigid motion, however the proposed algorithm should also work with non-rigid motion, considering the motion states of individual voxels, rather than the entire phantom.

References

[1] J. Hahn, H. Bruder, C. Rohkohl, T. Allmendinger, K. Stierstorfer, T. Flohr, and M. Kachelrieß, "Motion compensation in the region of the coronary arteries based on partial angle reconstructions from short-scan CT data," *Medical physics, 44(11),* 2017.

[2] R. Bhagalia, J. D. Pack, J. V. Miller, and M. Iatrou, "Nonrigid registration-based coronary artery motion correction for cardiac computed tomography," *Medical physics, 39(7Part1),* pp. 4245-4254, 2012.

[3] H. Bruder, C. Rohkohl, K. Stierstorfer, and T. Flohr, "Compensation of skull motion and breathing motion in CT using data-based and image-based metrics, respectively," *Proc. SPIE 9783, Medical Imaging 2016: Physics of Medical Imaging, 97831E,* 22 March 2016.

[4] D. Schafer, J. Borgert, V. Rasche, and M. Grass, "Motion-compensated and gated cone beam filtered back-projection for 3-D rotational X-ray angiography," *IEEE Transactions on Medical Imaging, vol. 25, no. 7,* pp. 898-906, July 2006.

[5] L. A. Feldkamp, L. C. Davis, and J. W. Kress, "Practical cone-beam algorithm," *J. Opt. Soc. Am. A 1,* pp. 612-619, 1984.

[6] K. Stierstorfer, A. Rauscher, J. Boese, H. Bruder, S. Schaller, and T. Flohr, "Weighted FBP—a simple approximate 3D FBP algorithm for multislice spiral CT with good dose usage for arbitrary pitch," *Physics in Medicine & Biology, 49(11),* p. 2209, 2004.

[7] J. Sunnegardh, and P.E. Danielsson, "Regularized iterative weighted filtered backprojection for helical cone-beam CT," *Medical physics, 35(9),* pp. 4173-4185, 2008.

[8] P.M. Joseph, "An improved algorithm for reprojecting rays through pixel images," *IEEE transactions on medical imaging, 1(3),* pp. 192-196, 1982.

[9] J. Sunnegardh and K. Stierstorfer, "A New Method for Windmill Artifact Reduction," *The 12th International Meeting on Fully Three-Dimensional Image Reconstruction in Radiology and Nuclear Medicine,* pp. 420-423, 2013.

[10] H. Turbell, "Cone-beam reconstruction using filtered backprojection," *Diss. Linköping University Electronic Press,* 2001.

On the use of voxel-driven backprojection and iterative reconstruction for small ROI CT imaging

Leonardo Di Schiavi Trotta, Dmitri Matenine, Margherita Martini, Yannick Lemaréchal, Pierre Francus, and Philippe Després

***Abstract*—In this work related to the use of a commercial medical CT scanner for the non-destructive analysis of highly attenuating materials (mineral samples), the effect of backprojection techniques and truncation artifacts corrections were explored. For small ROIs, the CT couch interferes significantly in images of small samples (few centimeters). An iterative reconstruction algorithm (OSC-TV) was used to perform reconstructions from uncorrected raw projection data made available through a collaboration with the CT vendor, who provided binaries and methods to remove low-level, proprietary data corrections (for beam hardening). The OSC-TV algorithm is customizable, allowing for the use of different forward-projection and backprojection techniques. Reconstruction parameters were tuned by performing simulations in a virtual phantom involving highly attenuating materials. Strategies to reconstruct small ROIs were also explored, with the objective of reducing truncation artifacts. Three samples were scanned to compare a ray-driven backprojection and a voxel-driven backprojection technique based on bilinear interpolation. The voxel-driven approach led to better results in terms of noise and reconstruction artifacts. An iterative ROI reconstruction technique was used to reconstruct small ROIs. This technique allows obtaining a sinogram with the projections of the ROI only. With that, truncation artifacts were reduced, which led to images with less blurring.**

***Index Terms*—tomographic reconstruction algorithm, Iterative reconstruction, Proprietary data format, backprojection by bilinear interpolation, GPU acceleration**

I. Introduction

X-ray Computed Tomography (CT) is nowadays ubiquitous in medicine for diagnosis, treatment planning and treatment responses purposes. This technology is also increasingly used for non-medical purposes in many fields, providing several advantages such as: (i) non-destructive testing and (ii) high spatial and density resolution [1].

In clinical CT scanners, raw acquisition data are stored as sinograms and are typically processed by proprietary methods, notably to reduce beam hardening artifacts in reconstructed images. Raw sinograms are relatively large from a storage perspective and are typically not kept on the long term, as opposed to reconstructed images which are normally sent to a PACS. Even though it is possible to archive sinograms, their proprietary format typically prevents users from reading them. The sinograms could also be preprocessed (*e.g.* calibration, beam-hardening correction), and in this respect might not truly represent *raw* CT data, *i.e.* measures of the attenuation along a ray. In some cases, including for research purposes, it might be desirable to access this raw attenuation data.

Previous works have shown the importance of the forward projection model on image quality [2]. The aim of this work was to explore how the reconstruction of small samples with small voxel size (*e.g.* 512 pixels $\times$ 512 pixels, and a pixel size in x and y of 0.00977 cm), are affected by the backprojection technique implemented on a iterative reconstruction algorithm, OSC-TV (Ordered subset convex algorithm with total variation minimization) [3].

Besides, a technique used to reconstruct regions-of-interest with this class of algorithms was also analyzed [4], as they allow reducing truncation artifacts caused by objects outside the reconstruction matrix.

II. Materials and methods

We have developed a framework where an in-house iterative algorithm can be used to reconstruct images based on genuinely raw attenuation data. First, the sinogram data in proprietary format are converted through the use of binaries provided by the manufacturer. The conversion generates usable sinogram data, and also provides associated geometry data. These data provided by the vendor is then used to perform the reconstruction with the iterative algorithm OSC-TV. Two backprojection techniques were evaluated in this work with regards to their impact on image features: a Siddon-based (ray-driven) [5] and a bilinear interpolation (voxel-driven) approach.

A. Projection and backprojection techniques

In the OSC-TV algorithm, the estimated image is forward-projected and backprojected several times, depending on the number of iterations and subsets [3].

In this work, the estimated image is forward projected using Siddon's algorithm [5], which is a ray-based technique that can be efficiently implemented on the GPU [3]. Backprojection was performed in this work with two techniques: (i) ray-based or (ii) voxel-driven by bilinear interpolation.

Sentinel North program of Université Laval, made possible by the Canada First Research Excellence Fund, and *Fonds de recherche du Québec - Nature et technologies* [2018-PR-206076].

L. Di Schiavi Trotta, Y. Lemaréchal and P. Després are with the Department of Physics, Physical Engineering, and Optics, Université Laval, 1045, av. de la Médecine G1V 0A6, Québec, Québec, Canada. e-mail: leonardo.di-schiavi-trotta.1@ulaval.ca

D. Matenine was with the Département de génie des systèmes, École de technologie supérieure,1100 Rue Notre-Dame Ouest H3C 1K3, Montréal, Québec, Canada.

M. Martini and P. Francus are with the Centre Eau Terre Environnement, Institut National de la Recherche Scientifique, 2605 Boulevard du Parc Technologique G1P 4S5, Québec, Québec, Canada.

In the voxel-driven backprojection by bilinear interpolation (BLI), the center of each voxel is projected on the detector and the reading corresponds to the weighted sum of the four neighboring pixels (the position is calculated relative to the detector pixel centers).

On the other hand, for a ray-driven backprojection, the detector readings are smeared back across the image. The voxels that are incremented in this process depend on the ray path, from the detector reading to the source. Typically, a finite number of rays are defined (e.g. one ray per detector). In this approach, depending on the angle between rays and the voxel size, some voxels might not be traversed by a given ray, or they are under-utilized, for the intersection length is negligible. This problem is not unique to the backprojection, but it is also present in the forward-projection in different techniques (*e.g.* Siddon, Joseph's method and bilinear interpolation) [2].

B. Working with proprietary format

For this work, a Siemens SOMATOM Definition AS+ 128 scanner was used. This device is installed at the *Institut national de la recherche scientifique*, in Québec City, Canada. This platform is used for several non-medical applications, including material characterization and custom beam hardening corrections with dual-energy techniques [6], [7].

The process of using raw data from this medical CT scanner is illustrated in Fig. 1. In summary, a vendor-provided calibration table is used for acquisitions to cancel any vendor-specific beam hardening corrections (detector calibration is still applied). The raw data is stored in the host system and copied to a different machine for archiving purposes, along with calibration data. These raw data, free of vendor-specific beam hardening corrections, can be read with vendor-provided binaries. These genuinely raw data (except for detector calibration), can thereafter be used in custom reconstruction algorithms designed to handle corrections from first principles, notably through dual-energy approaches [7].

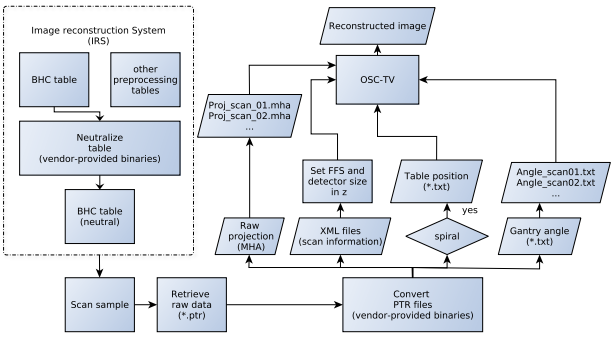


Fig. 1. Flowchart depicting how raw data is processed to perform tomographic reconstructions with OSC-TV.

C. Convergence of OSC-TV in numerical simulations

In order to verify the convergence of the reconstructed image, a virtual phantom was defined (see Fig. 2). It is composed of a water cylinder with 25.2 cm of diameter and 9 cylinder rods of distinct materials, each one with a diameter of 2.4 cm. The X-ray absorption properties of the phantom, defined by the linear attenuation coefficient, was retrieved from the NIST XCOM database [8].

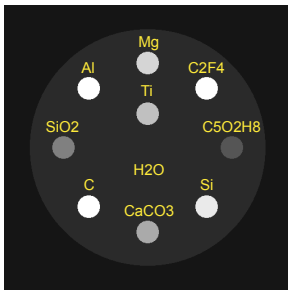


Fig. 2. Virtual phantom.

A noise-free monoenergetic step-and-shoot acquisition of the virtual phantom at 83 keV was simulated by using the geometry of the medical CT scanner Siemens SOMATOM Definition AS+ 128.

The convergence of the reconstruction to an optimal result is tuned by the reconstruction parameters: number of iterations, number of subsets, final number of subsets, and initial image; regularization parameters also play an important role in decreasing the overall noise and controlling the spatial resolution: gradient steps and strength of regularization (rms) [3].

A high number of subsets, equivalent to half the number of projections and a final number of subsets representing 1/10 of that value, can be used to achieve optimal convergence, as already suggested with few-view acquisitions [9].

Different combinations of reconstruction parameters are used to reconstruct the virtual phantom in order to assess the convergence: 5, 9 and 12 iterations; 84, 576 (1/4 projections) and 1152 (1/2 projections) subsets. A total of 9 reconstructions are performed (3×3), with the regularization constant and gradient steps fixed at 0.02 and 20, respectively.

D. Scanning protocols

Four samples were imaged in the experimental protocol (see Fig. 3): (a) water phantom, (b,c) small mineral sample (approximately 5 cm) of granite and chalcopyrite, respectively, and (d) a sandstone, with a diameter of 10.0 cm, henceforth called BEC A196-6.

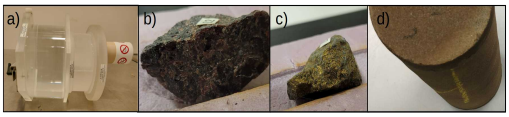


Fig. 3. Samples: (a) water phantom, (b) granite, (c) chalcopyrite, (d) BEC A196-6.

Different scan protocols were used for the samples, and for some cases proprietary beam-hardening correction (BHC) was neutralized (neutral): water phantom and chalcopyrite (neutral) at 140 kVp, BEC A196-6 (neutral) and granite (neutral) at 100 kVp. Tomographic reconstructions performed with the

Siemens algorithm, filtered backprojection, used the B30s (smooth) kernel.

E. Conversion calibration

Tomographic reconstruction performed by the OSC-TV algorithm are inherently in units of linear attenuation (cm^{-1}), while the ones obtained with the Siemens algorithm are in Hounsfield units (HU). The relation between these quantities is given by: $\mu = \left(\frac{HU}{1000} + 1\right) \times \mu_{water}$. Reconstruction of the water phantom, at 100 kVp and 140 kVp, are performed with OSC-TV, and the mean value in a ROI of each case is calculated. The average of the results provides μ_{water}. This value is independent of the tube voltage, for the projections are always normalized for water, so HU is always close to 0 (proprietary preprocessing of raw data).

F. Reconstruction of raw data using OSC-TV

Following the workflow depicted in Fig. 1, and the numerically-determined optimal reconstruction parameters defined in Section II-E, images were reconstructed with well-defined parameters of the OSC-TV code.

For the cases where small voxels are used with a regular grid (512 pixels $\times$ 512 pixels), a modified iterative reconstruction of a region-of-interest (IR ROI) technique based on the work of Ziegler *et al.* 2008 [4] was applied. The CT table lies outside the image reconstruction matrix for some cases. This technique allows us to perform reconstruction with a normal grid and small pixel size (e.g. 0.00977 cm), avoiding truncation artifacts caused by objects outside the reconstruction matrix.

Contrariwise, high-resolution reconstructions (e.g. 2048 pixels $\times$ 2048 pixels) would have to be made to cover the entire field-of-view (FOV) (*e.g.* 50 cm), for the table presents an important attenuation, even though a small object (few centimeters) is being reconstructed.

Variations of the OSC-TV algorithm are identified by acronyms, where BLI stands for backprojection by bilinear interpolation (voxel-driven), Siddon for ray-driven backprojection, and ROI for the IR ROI technique

III. Results and Discussion

A. Convergence of OSC-TV in numerical simulations

The visual convergence of the OSC-TV algorithm in terms of number of iterations and subsets is depicted in Fig. 4. As one can notice, 84 subsets is insufficient even when 12 iterations are performed, giving rise to beam-hardening-like artifacts. As those images were reconstructed from monochromatic projections (83 keV), such artifacts were not expected, and so are due to a non-optimal convergence. By increasing the number of subsets (576, or 1/4 of the number of projections), and the number of iterations, such artifacts are decreased. Finally, it is only when a high number of subsets (1152, which is equivalent to half the number of projections) is used that those artifacts are removed, as a result of high rate convergence.

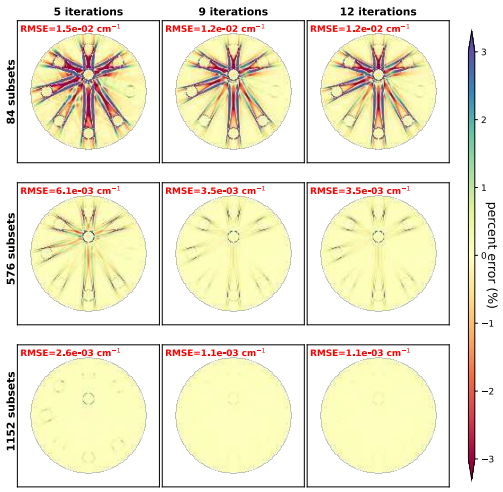


Fig. 4. Convergence of virtual phantom. Percent error maps.

B. Samples

From the measurements in the central slice of the water phantom, at 100 kVp and 140 kVp, it was obtained an average value $\mu_{water} = 0.1918\ cm^{-1}$, which allows conversion from HU to linear attenuation coefficient to be performed (see Fig. 5).

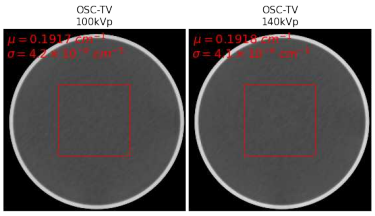


Fig. 5. Water phantom: (left) 100 kVp, (right) 140 kVp. Window: [0.18:0.22]

C. Reconstruction of ROI and small samples: OSC-TV ROI

The importance of the backprojection technique for artifact mitigation and the IR ROI technique for removing truncation artifacts is shown in Fig 6. In (a), the CT table was not included in the reconstruction matrix. It is worth noting that the setup uses a custom table, with much more important absorption properties than a medical one. Secondly, in (b), the image is reconstructed using Siddon backprojection and IR ROI technique, so the table is removed from the projection data. As not all voxels are incremented by the correspondent detector read during backprojection, more artifacts arise, producing a noisy image. Finally, in (c), the IR ROI and the voxel-driven backprojection by bilinear interpolation are combined. These two techniques are capable of mitigating both artifacts: truncation and lack of data in voxel increment during backprojection.

In Figs. 7 and 8, the reconstruction of a small sample of granite and chalcopyrite (approximately 5 cm) is shown with different techniques: (a) Siemens with the B30s kernel

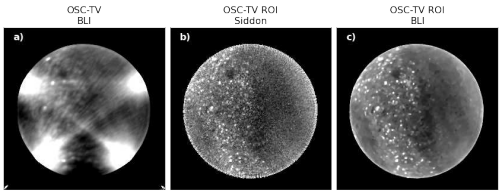


Fig. 6. BEC A196-6 100 kVp, reconstructed with OSC-TV applying different techniques: (a) backprojection by bilinear interpolation, (b) ray-driven backprojection and region-of-interest strategy, (c) backprojection by bilinear interpolation and region-of-interest strategy. Window [0.5:0.6] cm^{-1}.

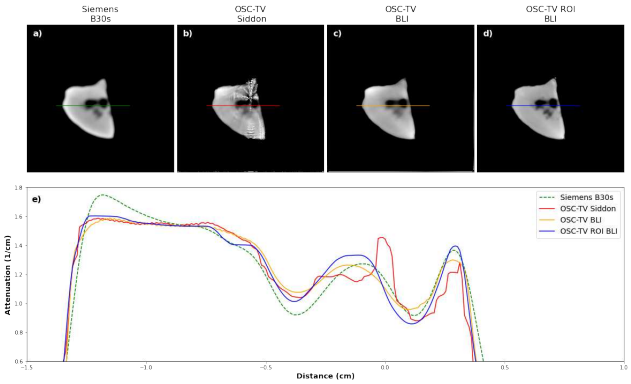


Fig. 7. Granite 100 kVp: (a) Siemens algorithm (B30s), (b) OSC-TV and the Siddon backprojection, (c) OSC-TV and backprojection by bilinear interpolation, (d) OSC-TV, backprojection by bilinear interpolation and region-of-interest strategy, (e) plot of the line profiles. Window [1.0:1.8]

(smooth); (b) OSC-TV with Siddon's backprojection; (c) OSC-TV with backprojection by bilinear interpolation; (d) OSC-TV with IR ROI technique and backprojection by bilinear interpolation. As a smooth kernel was selected for the Siemens reconstruction, its natural its OSC-TV counterpart (d) is less blurry. Image (b) suffers from the inherent problem of the Siddon's backprojection, where some voxels are not properly incremented by the backprojection and the artifacts are scattered from the center of the image, and the presence of noise. When (c) and (d) are compared, it is clear that the first is smoother. The increased blurring is also caused by truncation artifacts due to the CT scanner table, which is outside the reconstruction matrix. Contrariwise, the edges are smoother in general with the OSC-TV technique. The iterative technique also provides images with less beam-hardening artifacts: at the left edge of the granite in Fig 7, pixels values are steady, while a smooth decrease is observed in the Siemens reconstruction.

IV. Conclusion

In this work, a framework for working with proprietary data from a commercial CT scanner was presented. Binaries provided by the manufacturer allows research to be performed in both the projection and the image space, where beam-hardening preprocessing can also be neutralized, so custom BHC can also be applied. Ray-driven backprojection, with a limited number of rays, is insufficient for small geometries and samples. With a backprojection by bilinear interpolation

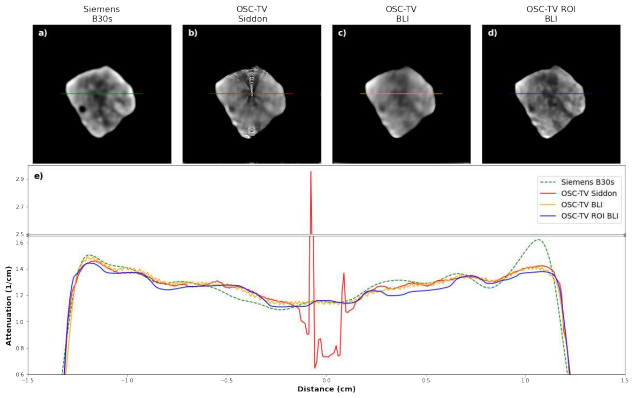


Fig. 8. Chalcopyrite 140 kVp: (a) Siemens algorithm (B30s), (b) OSC-TV and the Siddon backprojection, (c) OSC-TV and backprojection by bilinear interpolation, (d) OSC-TV, backprojection by bilinear interpolation and region-of-interest strategy, (e) plot of the line profiles. Window [1.0:1.6]

(voxel-driven), results with less noise and free of reconstruction artifacts were obtained for studied cases. Strategies to reconstruct regions-of-interest were also applied, where truncation artifacts were reduced, and less blurring was observed.

Acknowledgment

Karl Stierstorfer (software support), Philippe Letellier and Mathieu des Roches (support in scanning the samples), Stéphanie Larmagnat (providing samples), Jérôme Landry and Pascal Bourgault (algorithm development).

References

[1] G. v. Kaick and S. Delorme, "Computed tomography in various fields outside medicine," *Eur Radiol Suppl*, vol. 15, no. 4, pp. d74–d81, Nov. 2005. [Online]. Available: https://link.springer.com/article/10.1007/s10406-005-0138-1

[2] K. Hahn, H. Schöndube, K. Stierstorfer, J. Hornegger, and F. Noo, "A comparison of linear interpolation models for iterative CT reconstruction," *Medical Physics*, vol. 43, no. 12, pp. 6455–6473, Dec. 2016. [Online]. Available: https://www.ncbi.nlm.nih.gov/pmc/articles/PMC5106434/

[3] D. Matenine, Y. Goussard, and P. Després, "GPU-accelerated regularized iterative reconstruction for few-view cone beam CT," *Med Phys*, vol. 42, no. 4, pp. 1505–1517, Apr. 2015.

[4] A. Ziegler, T. Nielsen, and M. Grass, "Iterative reconstruction of a region of interest for transmission tomography," *Medical Physics*, vol. 35, no. 4, pp. 1317–1327, Apr. 2008.

[5] R. L. Siddon, "Fast calculation of the exact radiological path for a three-dimensional CT array," *Medical Physics*, vol. 12, no. 2, pp. 252–255, 1985. [Online]. Available: http://link.aip.org/link/?MPH/12/252/1

[6] M. Martini, P. Francus, L. Di Schiavi Trotta, and P. Després, "Identification of Common Minerals Using Stoichiometric Calibration Method for Dual-Energy CT," *Geochemistry, Geophysics, Geosystems*, vol. 22, no. 11, p. e2021GC009885, 2021. [Online]. Available: https://onlinelibrary.wiley.com/doi/abs/10.1029/2021GC009885

[7] L. Di Schiavi Trotta, D. Matenine, M. Martini, K. Stierstorfer, Y. Lemaréchal, P. Francus, and P. Després, "Beam-hardening corrections through a polychromatic projection model integrated to an iterative reconstruction algorithm," *NDT & E International*, p. 102594, Dec. 2021. [Online]. Available: https://www.sciencedirect.com/science/article/pii/S0963869521001936

[8] C. Suplee, "XCOM: Photon Cross Sections Database," Sep. 2009. [Online]. Available: https://www.nist.gov/pml/xcom-photon-cross-sections-database

[9] F. J. Beekman and C. Kamphuis, "Ordered subset reconstruction for x-ray CT," *Physics in medicine and biology*, vol. 46, no. 7, pp. 1835–1844, 2001. [Online]. Available: http://stacks.iop.org/0031-9155/46/1835

A Decomposition Method for Directional Total Variation With Application to Needle Reconstruction in Interventional Imaging

Marion Savanier, Cyril Riddell, Yves Trousset, Emilie Chouzenoux and Jean-Christophe Pesquet

***Abstract*—In interventional radiology, 3D reconstruction of devices such as needles would increase the precision of procedures. Doing so with CBCT is time-consuming and increases the X-ray dose. Needles being sparse, a compressed-sensing reconstruction approach seems viable. We thus investigate the interest of directional total variation as an adequate prior for anisotropic devices. We introduce a decomposition method that allows several a priori directions to be considered at once as well as excludes the anatomical background that is not sparse. The capacity of the method is illustrated on simulations of limited-angle acquisitions. It is shown to allow good reconstruction of the needles from a small angular coverage, even if the anatomical background cannot be recovered.**

***Index Terms*—CT, directional total variation, limited-angle reconstruction, optimization-based reconstruction**

I. Introduction

FLAT-panel based C-arm systems provide real-time 2D imaging to guide the navigation of therapeutic devices during minimally invasive vascular or percutaneous procedures. In the following, we shall focus on interventions that make use of metallic needles, such as vertebroplasty, radiofrequency ablations, or biopsies. Cone-beam computed tomography (CBCT) is available through the rotation of the C-arm around the patient. It allows the 3D reconstruction of highly attenuating metallic devices together with the patient background anatomy. In the guidance phase of the procedure, the patient is positioned to optimize the real-time visualization of the device and its trajectory. At any time, CBCT scans could be performed to precisely assess the position of the device according to the planned trajectory. But such repeated acquisitions increase the X-ray dose received by the patient. Furthermore, the patient may need to be moved to another position to avoid collisions during the $200°$ rotation required by tomography. Rotating over a smaller angular coverage would reduce the number of times changing the position of the patient is needed, and would reduce the number of projections, and the amount of X-ray dose. Model-based iterative reconstruction methods (MBIR) have proven useful to reconstruct soft tissues from a reduced number of projections over a full tomographic angular coverage. In [1], a least-squares criterion regularized with total variation (TV) was used to remove the undersampling streaks of dense objects over the less dense soft tissues. However, TV regularization is isotropic: in 2D, it penalizes the ℓ_1 norm of the image partial derivatives along the vertical and horizontal directions equally. TV is not able to recover the edges along directions not sampled by the limited angular coverage. Only edges and details tangent to the projection directions are recovered [2]. For piecewise constant geometrical objects, successful results have been obtained with the anisotropic total variation (ATV). ATV assigns different weights to the vertical and horizontal partial derivatives of the image. This strategy allows for considering the angular range as an additional prior information [3]. Since the non convex ℓ_0 pseudo-norm is the most direct measure of sparsity of an object, ℓ_1-reweighting strategies and heuristics have been investigated to incorporate the idea of an independence on the magnitude of the ℓ_0-norm into the ATV approach. But, due to non-convexity, it is not clear that the resulting optimization methods converge to a global minimum [3]–[5]. Recently, ATV constrained formulations (instead of regularization-based ones) allowed for the reconstruction of complex patterns from limited angle acquisitions [6]. The anisotropic regularizer proposed in [7] is particularly suited to thin objects like needles because it emphasizes one specific direction. In the following, we call it directional total variation (DTV). With DTV, the gradient norm is computed along one selected direction that is not necessarily aligned with the pixel grid. Applications on denoising and reconstructing images of fiber materials have been successful [8]. Additionally, DTV with a spatially varying direction and strength [9]–[11], that includes higher order derivatives [8], [12] has been proposed to extend the applicability of DTV, for instance to vessels and fingerprints. In this work, we consider the simple geometric shape of needles that is very sparse and can be reconstructed from a limited-angle acquisition. However, we here allow more directions than one and superimpose them over an anatomical background that is not assumed sparse. We propose to adopt an image decomposition approach that applies DTV over multiple directions for the needles and TV to approximate the background. Decomposition was first proposed for texture-geometry decomposition [13] and has also been applied to CT imaging to decompose the reconstruction into three components of object, subsampling artifacts and noise [14].

In Section II, we review the DTV regularization for incorporating directional information and the decomposition method to selectively apply separate directional constraints on separate components as well as to exclude the anatomical background.

M. Savanier, C. Riddell and Y. Trousset are with GE Healthcare, Buc, France. E-mail: first.last@ge.com

E. Chouzenoux and J.-C. Pesquet are with Université Paris-Saclay, CentraleSupélec, Centre pour la Vision Numérique, Inria, Gif-sur-Yvette, France. E-mail: first.last@centralesupelec.fr

Numerical experiments are then provided and discussed in Section III to illustrate the potential of the proposed method.

II. Method

A. Scanning model

Data collected in X-ray tomographic imaging can be modeled by the following discrete linear system

$$\mathbf{y} = \mathbf{H}\mathbf{x} + \mathbf{b} \tag{1}$$

where $\mathbf{H} \in \mathbb{R}^{M\times N}$ is the discretized model of forward projection [15], $\mathbf{x} \in \mathbb{R}^N$ is the unknown attenuation image, $\mathbf{y} \in \mathbb{R}^M$ represents the log-transform of the data measured by the detector and $\mathbf{b} \in \mathbb{R}^M$ is an additive noise term.
In a limited angle setting, (1) is a severely underdetermined system of linear equations. The lack of data must be compensated by a priori knowledge that constrains the problem by limiting the space of feasible solutions. To estimate $\mathbf{x}$, we consider the sum of a least-squares data fidelity term and a convex regularizer g embedding this prior information, in particular sparsity and direction:

$$\underset{\mathbf{x}\in\mathbb{R}^N}{\text{minimize}} \ \frac{1}{2}\|\mathbf{y} - \mathbf{H}\mathbf{x}\|^2_{\mathbf{D}} + g(\mathbf{x}). \tag{2}$$

Matrix $\mathbf{D}$ stands for the ramp filter which provides faster convergence through an approximate inversion [1], [16], [17]. We now discuss the choices for g.

B. Directional total variation (DTV)

DTV enforces the prior that the object is piecewise constant and follows one main direction. For an image $\mathbf{x} \in \mathbb{R}^N$, its DTV can be defined as $\mathrm{DTV}_\Omega(\mathbf{x}) = \sum_{n=1}^N \|(\nabla_\Omega \mathbf{x})_n\|_1 = \|\Lambda \mathbf{R}_\theta(\nabla \mathbf{x})\|_{1,1}$ where $\nabla_\Omega \in \mathbb{R}^{2\times N}$ contains two directional derivatives at pixel n, $\Delta_n^\theta \mathbf{x}$ and $\Delta_n^{\theta+\pi/2}\mathbf{x}$, parameterized by direction angle $\theta \in [0°, 180°[$, and a so-called stretching factor $s \in]0, 1]$ for anisotropy, $\Omega = \{\theta, s\}$, i.e.

$$\begin{aligned}(\nabla_\Omega \mathbf{x})_n &= \begin{pmatrix} \Delta_n^\theta \mathbf{x} \\ s\Delta_n^{\theta+\pi/2}\mathbf{x} \end{pmatrix} = \Lambda \mathbf{R}_\theta \begin{pmatrix} \Delta_n^{\mathrm{h}} \mathbf{x} \\ \Delta_n^{\mathrm{v}} \mathbf{x} \end{pmatrix} \\ &= \begin{pmatrix} 1 & 0 \\ 0 & s \end{pmatrix} \begin{pmatrix} \cos\theta & \sin\theta \\ -\sin\theta & \cos\theta \end{pmatrix} \begin{pmatrix} \Delta_n^{\mathrm{h}} \mathbf{x} \\ \Delta_n^{\mathrm{v}} \mathbf{x} \end{pmatrix},\end{aligned} \tag{3}$$

with $\Delta_n^{\mathrm{h}} \in \mathbb{R}^N$, $\Delta_n^{\mathrm{v}} \in \mathbb{R}^N$, respectively, the horizontal and vertical discrete gradient operators at location n. These quantities can be obtained by applying a forward finite difference scheme with zero boundary condition.
Given that a set of needles makes a very sparse image, we add an ℓ_1 penalty so that $g(\mathbf{x}) = g_\Omega(\mathbf{x}) = \rho \mathrm{DTV}_\Omega(\mathbf{x}) + \alpha\|\mathbf{x}\|_1$, $(\alpha, \rho) \in]0, +\infty[^2$ in (2).

C. Image decomposition

With a non-sparse background, the lack of data cannot be compensated and the problem does not have a sparse solution. Decomposing $\mathbf{x}$ into a linear combination of several components aims to restore sparsity in all components that can then be recovered from the limited data. A component is thus defined by its specific sparsity prior. To direct the interfering background into a single component, a different sparse approximation is used. Then instead of estimating the sum directly, we solve the minimization problem for each of these components simultaneously.
Here we decompose $\mathbf{x}$ into the anatomical background component $\mathbf{x}_{\mathrm{B}}$ penalized with TV and $I \in \mathbb{N}$ directional components $\mathbf{x}_{\Omega_i}$ penalized with DTV of direction $\theta_i \in [0°, 180°[$ and stretching parameter $s_i \in]0, +\infty[$ such that

$$\mathbf{x} = \mathbf{x}_{\mathrm{B}} + \sum_{i=1}^I \mathbf{x}_{\Omega_i} \tag{4}$$

where $\Omega_i = \{\theta_i, s_i\}$, $i \in \{1, \dots, I\}$.
Altogether, we must solve the following convex problem:

$$\begin{aligned}\underset{\mathbf{x}_{\mathrm{B}}, (\mathbf{x}_{\Omega_i})_{i=1}^I \in \mathbb{R}^N}{\text{minimize}} \ & \frac{1}{2}\|\mathbf{y} - \mathbf{H}(\mathbf{x}_{\mathrm{B}} + \sum_{i=1}^I \mathbf{x}_{\Omega_i})\|^2_{\mathbf{D}} \\ & + \sum_{i=1}^I g_{\Omega_i}(\mathbf{x}_{\Omega_i}) + g_{\mathrm{TV}}(\mathbf{x}_{\mathrm{B}}).\end{aligned} \tag{5}$$

Note that each directional component $\mathbf{x}_{\Omega_i}$ can actually capture a needle or a group of needles of about the same direction.

D. Optimization algorithm

To minimize Problem (5), we reformulate it. Let $\mathbf{z} = \begin{bmatrix}\mathbf{x}_{\mathrm{B}}^\top & \mathbf{x}_{\Omega_1}^\top & \dots & \mathbf{x}_{\Omega_I}^\top\end{bmatrix}^\top \in \mathbb{R}^{(I+1)N}$. Let $\tilde{\mathbf{H}} = \Pi\mathbf{H}$ and $\tilde{\mathbf{D}} = \Pi\mathbf{D}$ where $\Pi\colon \mathbf{M} \mapsto \begin{bmatrix}\mathbf{M}^\top & \dots & \mathbf{M}^\top\end{bmatrix}^\top \in \mathbb{R}^{((I+1)L)\times L}$. We then write:

$$\underset{\mathbf{z}\in\mathbb{R}^{(I+1)N}}{\text{minimize}} \ \frac{1}{2}\|\mathbf{y} - \tilde{\mathbf{H}}\mathbf{z}\|^2_{\tilde{\mathbf{D}}} + h(\mathbf{z}), \tag{6}$$

with $h\colon \mathbf{z} \mapsto \sum_i^I g_{\Omega_i}(\mathbf{x}_{\Omega_i}) + g_{\mathrm{TV}}(\mathbf{x}_{\mathrm{B}})$.
The cost function in (6) is convex but non-smooth. In this context, FISTA algorithm [18] is attractive due to its simplicity and low-memory requirements. It relies on the use of proximal operators. Let us recall that the proximity operator of h at $\mathbf{x}$ is defined as $\mathrm{prox}_h(\mathbf{x}) = \underset{\mathbf{z}\in\mathbb{R}^N}{\mathrm{argmin}} \ \big(h(\mathbf{z}) + \frac{1}{2}\|\mathbf{x} - \mathbf{z}\|^2\big)$.
Let a be a positive real number such that $a > 2$. The k-th iteration of FISTA applied to (6) reads:

$$\begin{cases} \beta_k = k/(k+1+a) \\ \tilde{\mathbf{z}}^k = \mathbf{z}^k + \beta_k(\mathbf{z}^k - \mathbf{z}^{k-1}) \\ \mathbf{z}^{k+1} = \mathrm{prox}_{\tau h}(\tilde{\mathbf{z}}^k - \tau\tilde{\mathbf{H}}^\top\tilde{\mathbf{D}}(\mathbf{H}\tilde{\mathbf{z}}^k - \mathbf{y})) \end{cases} \tag{7}$$

which includes a momentum step to accelerate the convergence. Thanks to the separability in each component of $\mathbf{z}$, we derive an update rule for each map:

$$\begin{cases} \beta_k = k/(k+1+a) \\ \tilde{\mathbf{x}}_{\mathrm{B}}^k = \mathbf{x}_{\mathrm{B}}^k + \beta_k(\mathbf{x}_{\mathrm{B}}^k - \mathbf{x}_{\mathrm{B}}^{k-1}) \\ \text{For } i \in \{1, \dots, I\}: \\ \quad \tilde{\mathbf{x}}_{\Omega_i}^k = \mathbf{x}_{\Omega_i}^k + \beta_k(\mathbf{x}_{\Omega_i}^k - \mathbf{x}_{\Omega_i}^{k-1}) \\ \mathbf{x}^k = \tilde{\mathbf{x}}_{\mathrm{B}}^k + \sum_i^I \tilde{\mathbf{x}}_i^k \\ \mathbf{x}_{\mathrm{B}}^{k+1} = \mathrm{prox}_{\tau g_{\mathrm{TV}}}(\tilde{\mathbf{x}}_{\mathrm{B}}^k - \tau\mathbf{H}^\top\mathbf{D}(\mathbf{H}\mathbf{x}^k - \mathbf{y})) \\ \text{For } i \in \{1, \dots, I\}: \\ \quad \mathbf{x}_{\Omega_i}^{k+1} = \mathrm{prox}_{\tau g_{\Omega_i}}(\tilde{\mathbf{x}}_{\Omega_i}^k - \tau\mathbf{H}^\top\mathbf{D}(\mathbf{H}\mathbf{x}^k - \mathbf{y})) \end{cases} \tag{8}$$

The convergence of Algorithm (8) is guaranteed for $0 < \tau \leqslant 1/|||\tilde{\mathbf{H}}^\top\tilde{\mathbf{D}}\tilde{\mathbf{H}}||| = 1/|||(I+1)\mathbf{H}^\top\mathbf{D}\mathbf{H}|||$ where $|||\cdot|||$ denotes

the spectral norm of the input matrix.
The proximity operators of $\mathbf{x} \mapsto \tau g_{\mathrm{TV}}(\mathbf{x})$ and $\mathbf{x} \mapsto \tau g_{\Omega_i}(\mathbf{x})$ do not have a closed form, hence they are both approximated by using inner iterations of the dual forward-backward (DFB) algorithm [19] with warm-restart. In particular, for DTV, $\hat{\mathbf{x}} = \mathrm{prox}_{\tau g_{\Omega_i}}(\tilde{\mathbf{x}})$ is estimated using the following sub-iteration:

$$\begin{cases} \mathbf{x}^n = \tilde{\mathbf{x}} - \nabla_\Omega^\top \mathbf{u}^n \\ \mathbf{u}^{n+1} = \mathrm{proj}_{\|\cdot\|_{\infty,\infty} \leqslant \beta}(u_n + \gamma \nabla_\Omega \mathbf{x}^n) \end{cases} \tag{9}$$

where $\gamma < 2/|||\nabla_\Omega|||^2$.

III. Experiments

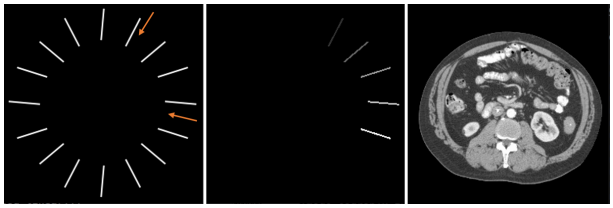


Fig. 1: Reference images. From left to right: Phantom (A) with needles of intensity 3500 HU, Image with needles of growing intensity from 3000 HU up to 5000 HU, Anatomical background [1800-2200 HU].

A. Simulations

We carry out simulations in parallel geometry using two numerical phantoms on a 256×256 grid. Hounsfield units (HU) are shifted such that air has value 0 HU and water is 1000 HU. Phantom (A) is purely geometric and represents a set of needles of intensity 3500 HU covering 8 directions (5°, 27.5°, 50°, 72.5°, 95°, 107.5°, 130°, 152.5°) as shown in Figure 1 (angles start at twelve o'clock and grow clockwise). Phantom (B) is the sum of an axial CT slice of an abdomen (see Figure 1) with a subset of needles of varying intensity (3000-5000 HU).
A needle is within the scanning arc if the projection data contain its so-called bull's eye view i.e., the view orthogonal to its axis. We computed simulated data of these phantoms over a circular arc of amplitude $\nu = 66°$ from angle $\theta_{\min} = 29°$ to angle $\theta_{\max} = 95°$ (indicated by the arrows in Figure 1) so that the projection data contains the bull's eye view of three needles. Noise term $\mathbf{b}$ was i.i.d. Gaussian of mean 0 and standard deviation 50. The angular sampling was uniform with a step of 2°. Reconstruction with TV regularization was taken as a baseline. FBP reconstruction followed by a thresholding of the intensity was added to the comparison. We performed 100 iterations of DFB and 5000 iterations of FISTA.
First, we analyze the performance of our decomposition method (5) for the reconstruction of a subset of the needles of Phantom (A) thanks to four DTV of direction $\{5°, 27.5°, 72.5°, 107.5°\}$ (i.e., $I = 4$). Then we show the applicability of the method to the more complex case of a background and needles of different intensities by reconstructing Phantom (B). This time, a set of $I = 3$ directions is used: $\{27.5°, 72.5°, 107.5°\}$. In all these simulations, the needles have the same size, so we use the same stretching parameter $s = 0.001$. Needles with the same intensity have the same regularization parameters ρ and α. TV and DTV parameters are thresholds that are homogeneous to HU intensity values of the image.

B. Background-free needles

Figure 2 shows the reconstructions of Phantom (A) with FBP, TV ($\beta = 50$) and DTV ($\rho = 50$, $\alpha = 1$). First, we see that with FBP, only partial reconstruction of the three needles within the scanning arc is achieved. Figure 3 displays the four reconstructed directional components. Both DTV and TV regularization lead to similar reconstructions for the three needles in the scanning arc. For the two needles of direction close to $\theta_{\min}$, TV yielded a partial recovery only, whereas DTV fully recovered 12 out of 16 needles, because their directions were sufficiently close to the imposed a priori directions. The four remaining missing needles show that there is no recovery without a priori directional information.

C. Needles with background

Figure 4 shows the reconstruction of the needles of Phantom (B) with FBP, TV and DTV (sum of all needle maps). First, as expected, the anatomical background cannot be recovered with a sparse prior in this limited angle settings. The needles reconstructed with FBP are distorted and the intensity values are not recovered. With TV, only the three needles within the scanning arc remains after thresholding whereas five needles are recovered with DTV. Figure 5 shows that the decomposition method coupled with directional information separate the three sets of needles from the background map.

IV. Discussion and conclusion

The potential of DTV for reconstructing geometrical objects from limited data acquisition is confirmed in our results where needles are recovered even when their bull's eye view is not sampled. A decomposition method was developed to benefit from this capacity with multiple a priori directions and in presence of an anatomical background that cannot be recovered by total variation priors. The proposed approach provides the equivalent of a background subtraction which thus allows DTV to enhance each needle along its a priori known direction and recover them from small scanning arcs. This approach is promising for increasing the precision of interventional radiology procedures through limited-angle acquisition. Our study is however limited to the interference due to the background, as it is based on nearly perfect data. Data corruption due to physical effects and stronger noise, in particular metallic artifacts, is another source of interference that may limit the capacity of sparse priors and must be dealt with before clinical application.

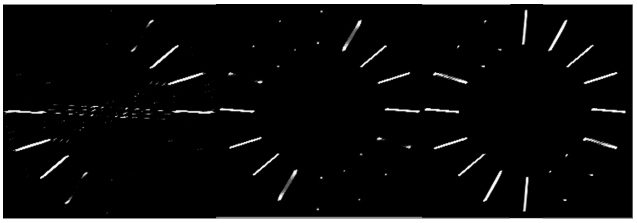


Fig. 2: Reconstructed images for $\nu = 66°$. From left to right: FBP, TV, DTV.

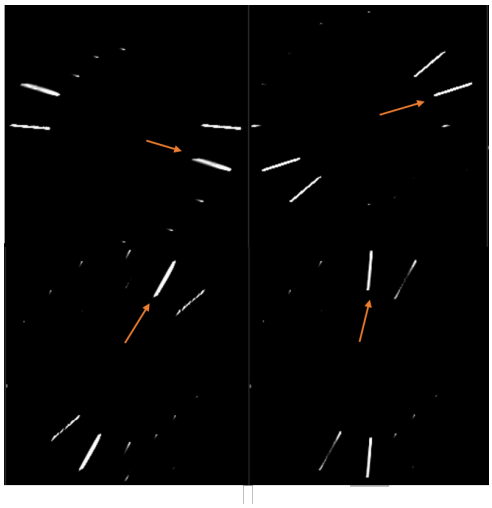


Fig. 3: Directional components obtained with DTV model on Phantom (A). Top: from left to right, $\theta_1 = 107.5°$, $\theta_2 = 72.5°$. Bottom: from left to right, $\theta_3 = 27.5°$, $\theta_4 = 5°$.

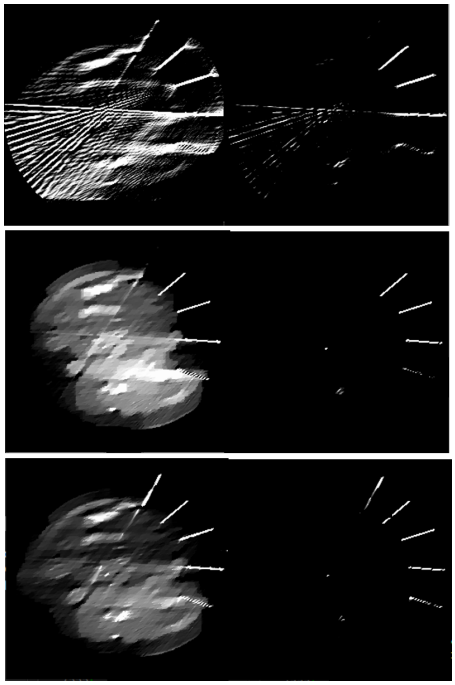


Fig. 4: Reconstruction of needles of Phantom (B) in the presence of a background for $\nu = 66°$. Left: needles with background. Right: needles map. From top to bottom: FBP, TV, DTV.

Fig. 5: Reconstructed needles maps obtained with DTV model on Phantom (B). From left to right: $\mathbf{x}_{\Omega_1}$ ($\theta_1 = 107.5°$), $\mathbf{x}_{\Omega_2}$ ($\theta_2 = 27.5°$), $\mathbf{x}_{\Omega_3}$ ($\theta_3 = 72.5°$).

REFERENCES

[1] H. Langet, C. Riddell, A. Reshef, Y. Trousset, A. Tenenhaus, E. Lahalle, G. Fleury, and N. Paragios. Compressed-sensing-based content-driven hierarchical reconstruction: Theory and application to C-arm cone-beam tomography. *Medical Physics*, 42(9):5222–5237, 2015.

[2] E. T. Quinto. Tomographic reconstructions from incomplete data numerical inversion of the exterior radon transform. *Inverse Problems*, 4(3):867, 1988.

[3] T. Wang, K.H. Nakamoto, Heye Zhang, and H. Liu. Reweighted anisotropic total variation minimization for limited-angle ct reconstruction. *IEEE Transactions on Nuclear Science*, 64(10):2742–2760, 2017.

[4] W. Yu, C. Wang, and M. Huang. Edge-preserving reconstruction from sparse projections of limited-angle computed tomography using ℓ_0-regularized gradient prior. *Review of Scientific Instruments*, 88(4):043703, 2017.

[5] J. Xu, Y. Zhao, H. Li, and P. Zhang. An image reconstruction model regularized by edge-preserving diffusion and smoothing for limited-angle computed tomography. *Inverse Problems*, 35(8):085004, 2019.

[6] Z. Zhang, B. Chen, D. Xia, E. Y. Sidky, and X. Pan. Directional-tv algorithm for image reconstruction from limited-angular-range data. *Medical Image Analysis*, 70:102030, 2021.

[7] I. Bayram and M. Kamasak. Directional total variation. *IEEE Signal Processing Letters*, 19:781–784, 2012.

[8] R. D. Kongskov, Y. Dong, and K. Knudsen. Directional total generalized variation regularization. *BIT Numerical Mathematics*, May 2019.

[9] H. Zhang and Y. Wang. Edge adaptive directional total variation. *The Journal of Engineering*, 2013(11):61–62, 2013.

[10] O. Merveille, B. Naegel, H. Talbot, and N. Passat. nD variational restoration of curvilinear structures with prior-based directional regularization. *IEEE Transactions on Image Processing*, 28(8):3848–3859, 2019.

[11] S. Parisotto and C.-B. Schönlieb. *Total Directional Variation for Video Denoising*. Springer International Publishing, 2019.

[12] S. Parisotto, J. Lellmann, S. Masnou, and C.-B. Schönlieb. Higher-order total directional variation: Imaging applications. *SIAM Journal on Imaging Sciences*, 13(4):2063–2104, 2020.

[13] J.-F. Aujol, G. Gilboa, T. F. Chan, and S. Osher. Structure-texture image decomposition—modeling, algorithms, and parameter selection. *International Journal of Computer Vision*, 67:111–136, 2006.

[14] J. Li, C. Miao, Z. Shen, G. Wang, and H.Yu. Robust frame based x-ray ct reconstruction. *Journal of Computational Mathematics*, 34:683–704, 2016.

[15] M. Savanier, C. Riddell, Y. Trousset, E. Chouzenoux, and J.-C. Pesquet. Magnification-driven B-spline interpolation for cone-beam projection and backprojection. *Medical Physics*, 48(10):6339–6361.

[16] M. Savanier, E. Chouzenoux, J.-C. Pesquet, and C. Riddell. Unmatched preconditioning of the proximal gradient algorithm. *IEEE Signal Processing Letters (submitted)*, 2022.

[17] E. Chouzenoux, J.-C. Pesquet, C. Riddell, M. Savanier, and Y. Trousset. Convergence of proximal gradient algorithm in the presence of adjoint mismatch. *Inverse Problems*, 37(6), 2021.

[18] A. Chambolle and C. Dossal. On the convergence of the iterates of "fista". *Journal of Optimization Theory and Applications*, 166:25, 2015.

[19] P. L. Combettes and J.-C. Pesquet. Proximal Splitting Methods in Signal Processing. In *Fixed-Point Algorithms for Inverse Problems in Science and Engineering*, pages 185–212. Springer, New York, NY, 2011.

New Reconstruction Methodology for Chest Tomosynthesis based on Deep Learning

F Del Cerro. C, Galán. A, García-Blas. J, Desco. M, Abella M.

***Abstract*—Tomosynthesis offers an alternative to planar radiography providing pseudo-tomographic information at a much lower radiation dose than CT. The fact that it cannot convey information about the density poses a major limitation towards the use of tomosynthesis in chest imaging, due to the wide range of pathologies that present an increase in the density of the pulmonary parenchyma. Previous works have attempted to improve image quality through enhanced analytical, iterative algorithms, or including a deep learning-based step in the reconstruction, but the results shown are still far from the quantitative information of a CT. In this work, we propose a reconstruction methodology consisting of a filtered back-projection step followed by post-processing based on Deep Learning to obtain a tomographic image closer to CT. Preliminary results show the potential of the proposed methodology to obtain true tomographic information from tomosynthesis data, which could replace CT scans in applications where the radiation dose is critical.**

***Index Terms*—Chest tomosynthesis, Computed Tomography Deep Learning, FDK-based reconstruction, Transfer Learning.**

I. Introduction

CHEST radiography is widely used for the diagnosis of lung diseases. It has a high spatial resolution, but being a projection image, tissues are shown overlapped without depth information and with a masking of low-density structures behind the bones. This limitation has been partially solved by tomosynthesis, a technique mainly used in mammography, but it still cannot convey density information. This is a major limitation for the use of tomosynthesis in chest imaging, due to the wide range of pathologies with increased lung parenchymal density. The only available technique to retrieve a density map is computed tomography (CT), but its use is limited by issues of availability and/or radiation dose control.

Several reconstruction strategies have been proposed to obtain an improved reconstruction of tomosynthesis data. In [1], the authors proposed a Digital Breast Tomosynthesis (DBT) reconstruction based on the filtered back-projection (FBP) algorithm by a modification of the ramp filter, while [2, 3] propose iterative reconstruction (IR) algorithms. Nevertheless, although these approaches increase image quality, the results shown are still far from the quantitative information of a CT.

During the last few years, image reconstruction methods have incorporated Deep Learning strategies as: (1) a post-processing step, after reconstruction [4], (2) a pre-processing of the projection data [5], or (3) within an IR algorithm, in the so-called unrolled methods [6, 7]. Nevertheless, despite further improving image quality over traditional IR, none of the proposed methods show true tomographic results.

In this work, we propose a reconstruction methodology to obtain tomographic images (close to a CT) from chest tomosynthesis. The methodology consists of an FDK-based reconstruction step followed by post-processing based on Deep Learning.

II. Database generation

The database consists of tomosynthesis reconstructions as inputs and CT scans as a reference. We used 55 chest CT studies, 12 provided by the Hospital General Universitario Gregorio Marañón and 43 extracted from the Medical Imaging and Data Resource Center (MIDRC) portal from three different datasets. All studies were interpolated to a common isotropic voxel size, $1.25\times1.25\times1.25$ mm, which is larger than that of any of the original CT volumes.

To simulate the tomosynthesis projections and perform the reconstructions, we used FuxSim [8], a software that allows simulating different acquisition geometries based on a density map and system parameters. The simulations were based on a standard chest tomosynthesis protocol: 61 projections with a total source-to-detector distance of 1800 mm and a scan angle of 30 degrees. The resulting projections had a matrix size of 1072×1072 pixels with a pixel size of 0.4×0.4 mm. Since the patient is positioned next to the detector in the acquisition of chest tomosynthesis data, we cropped the volumes to remove the excess of air from the CT field of view and eliminated the bed and other patient supports, as shown in Fig. 1.

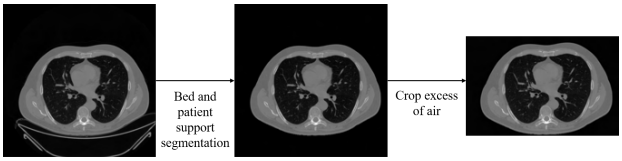


Fig. 1. Pre-processing of CT data to simulate patient position in tomosynthesis system: removal of the bed, patient support and excess of air.

To avoid truncation artifacts in the tomosynthesis projections, caused by the spatial limitation of the chest CT scans in the vertical direction, we extended the CT volumes as shown in Fig. 2.

C. F Del Cerro, A. Galán, M. Abella and M. Desco are with the Dpto. Bioingeniería e Ingeniería Aeroespacial, Universidad Carlos III de Madrid and the Instituto de Investigación Sanitaria Gregorio Marañón, Madrid, España (e-mail: mabella@ing.uc3m.es, cfernandez@hggm.es, agalan@hggm.es).

J. García Blas is with the Dpto. Arquitectura de Computadores, Comunicaciones y Sistemas, Universidad Carlos III de Madrid, Madrid, España (e-mail: fjblas@inf.uc3m.es)

M. Abella and M. Desco are with the Centro Nacional de Investigaciones Cardiovasculares Carlos III (CNIC), Madrid, España (e-mail: manuel.desco@cnic.es).

M. Desco is with the Centro de investigación en red salud mental (CIBERSAM), Madrid, España. (e-mail: desco@hggm.es).

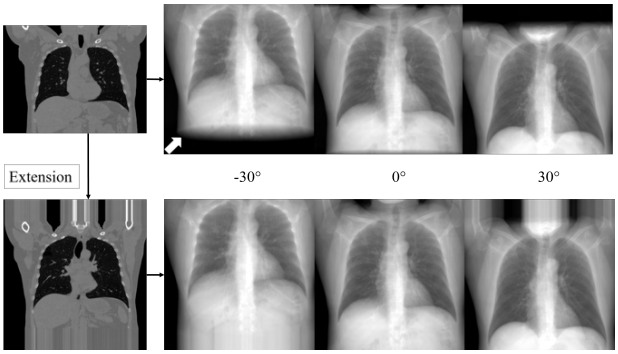


Fig. 2. Tomosynthesis projections from CT for three different positions before (top) and after (bottom) extension. White arrow shows truncation artifact.

The simulated tomosynthesis data were reconstructed with an FDK-based algorithm, modifying the minimum value of the ramp filter to avoid the loss of mean value. For each study, we obtained a volume of $275\times192\times99$ corresponding to the central part of the thorax, which is the ROI evaluated for diagnosis, with $1.25\times1.25\times1.25$ mm voxel size. Fig. 3 shows the central coronal slice of the tomosynthesis reconstruction together with that of the reference CT for three of the volumes.

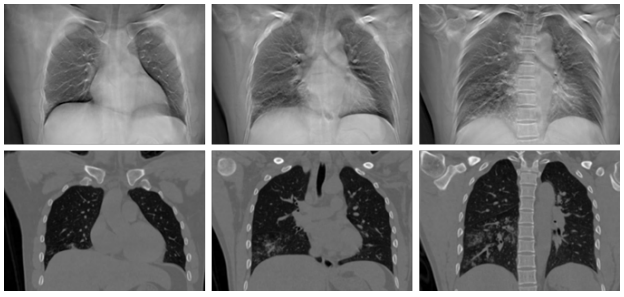


Fig. 3. Coronal slices of the result of FDK-based reconstruction from the tomosynthesis data (top) and their corresponding CT (bottom).

The database was divided into the following sets:

- Training set: Composed of 38 volumes, giving a total of 3762 coronal slices and 7296 axial slices.
- Validation set: Composed of 10 volumes, giving a total of 990 coronal slices and 1920 axial slices.
- Test set: Composed of 6 volumes, giving a total of 594 coronal slices and 1152 axial slices.

III. Post-processing step

The post-processing step has as input the FDK-based reconstruction and as output a volume with enhanced tomographic information. The proposed architecture is a modification of the original version of the U-Net architecture [9] to include new strategies recently proposed in the literature. The encoder was replaced by ResNet-34 [10], due to the improved performance offered by the residual blocks over the original version of the U-Net encoder. The decoder consists mainly of oversampling blocks composed of a convolutional layer, followed by an average pooling layer (APL), a batch normalization (BN) layer, a ReLU, and two convolution layers (Conv2D) followed by a ReLU each. At the output of the first convolutional layer, the pixel shuffle operation with ICNR initialization [11] is applied. The complete network is shown in Fig. 4.

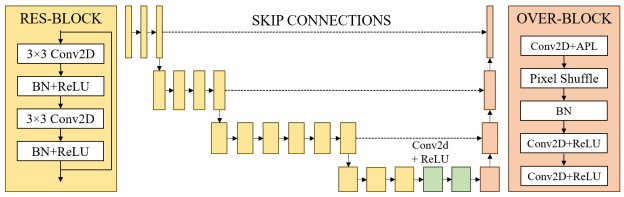


Fig. 4. Proposed U-net network architecture.

For the initialization of the encoder, we used the pre-trained version of Resnet-34 for *Imagenet* [12]. The decoder was initialized randomly adapting it to the output of our problem. The training methodology was based on Transfer Learning [13], first freezing the encoder to train only the decoder and then training the end-to-end network for fine-tuning.

Perceptual Loss (PL) [14] was chosen as the cost function to assure the recovery of fine details. We chose Adam [15] as the optimizer, due to its higher convergence speed, and a weight decay equal to 10^{-3} as regularization strategy. To determine the learning rate, we used the test presented by Leslie N. Smith [16], which resulted in an optimal learning rate of 10^{-3} for decoder training and 10^{-5} for end-to-end network training.

Although the model processes 2D images, we incorporate 3D information by alternating coronal and axial slices in a multi-stage training strategy. As shown in Fig. 5, in the first stage we trained with the coronal view of the FDK-based reconstruction as inputs and the coronal CT view as target. In the next round, the coronal prediction of the first stage was resliced to obtain the axial view, this being the input at this stage and the target being the CT axial view. The axial view of the new prediction was then resliced to obtain the coronal view again and a new training was performed.

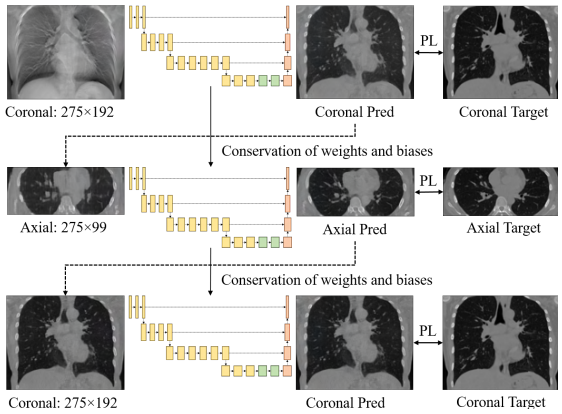


Fig. 5. Scheme of the training strategy.

IV. Results

Fig. 6 shows a good recovery of general structure and realistic texture with the proposed reconstruction methodology, especially in areas with well-defined in the FDK-based reconstruction, such as the spine.

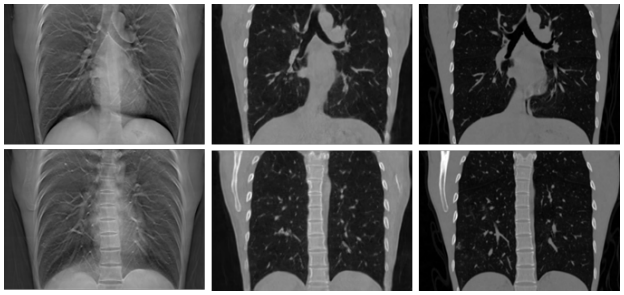


Fig. 6. Coronal slices of: FDK-based reconstruction (left), prediction (center) and CT (right).

Fig. 7 shows the mean of the root mean square error (RMSE) with respect to the target CT for the six test volumes. We can see that our proposed methodology reduces the RMSE compared to FDK-based reconstruction in whole volume, but greater in the slices with tissues with higher contrast, such as spine and ribs (see coronal slices shown).

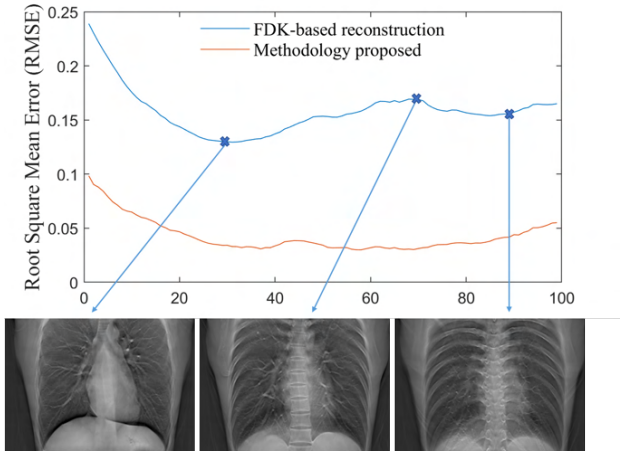


Fig. 7. Mean value of RMSE for the six test volumes along the different coronal slices with three FDK-based reconstruction examples.

Fig. 8 show a zoom of coronal slices in Figure 6, where we can see a blurring, and more importantly, structures that do not exactly match those in the original CT (see white arrows).

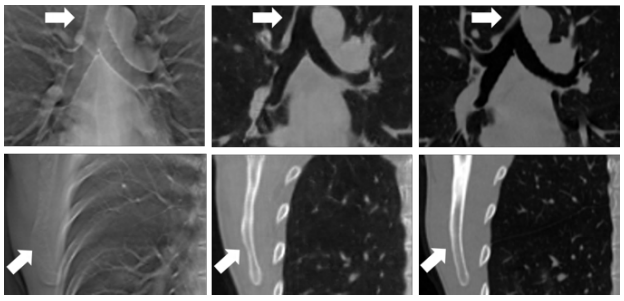


Fig. 8. Zoomed ROIs from coronal slices in Fig. 6: FDK-based reconstruction (left), prediction (center) and CT (right). White arrow highlights structures wrongly.

Fig. 9 shows the predicted image after each stage. After the first stage, we can still see artifacts in the axial view from the FDK-based reconstruction, because the prediction is done on the coronal view and the FDK. The second stage, working on the axial view, these artifacts are removed, the contours are recovered, and texture is improved. The last stage further improves texture improves without altering shapes already found in previous stages.

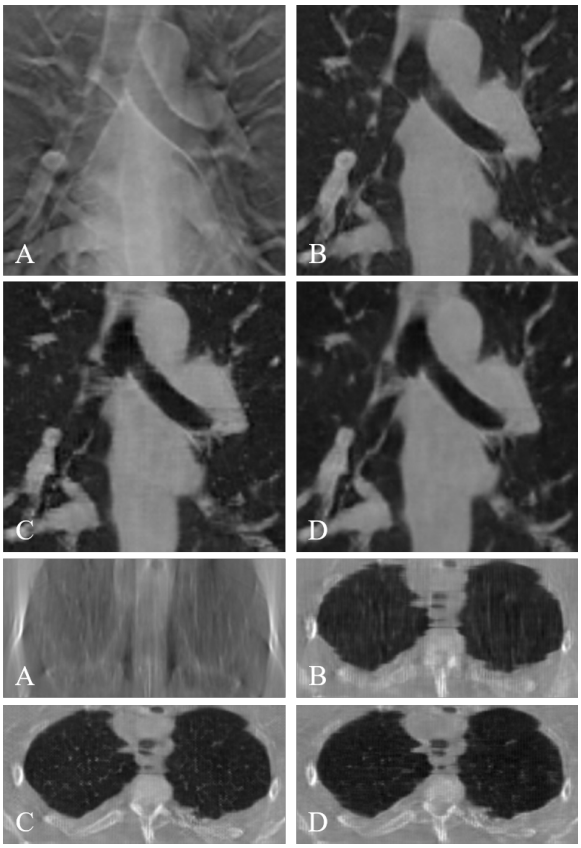


Fig. 9. Coronal (top) and axial (bottom) slices of: FDK-based reconstruction (A), the first (B), second (C) and third (D) stages.

V. Discussion

We have presented a reconstruction methodology for chest tomosynthesis based on FDK followed by post-processing based on deep learning. We use the transfer learning workflow, with U-Net architecture having a ResNet34 as encoder. Although our network processes 2D images, we incorporate 3D information by alternating coronal and axial slices in a multi-stage training strategy. The first stage of training uses the coronal view because it is the one with most information in FDK-based reconstruction. The evaluation shows a good reconstruction, especially in areas with high contrast and well-defined structures in the FDK-based reconstruction, such as the spine, and a very realistic tomographic texture. However, in the predicted images, we can find structures that do not match exactly with the original CT (biases). Looking at the result of the different stages, we can see that the biases

appear in the first stage and are maintained through subsequent stages. Future work will evaluate the combination of both prediction of previous stage together with the original FDK-based reconstruction at the input of the second and third stages to minimize the propagation of bias along the stages.

In this work we use a voxel size bigger than that of the original CTs due to computational limitations. This might be together with the subsampling layers of our U-Net might be responsible of the loss of spatial resolution. Future work will try to overcome the memory limitation by distributed training in using multiple GPUs.

Acknowledgment

This work has been funded by the Programme of support for the implementation of interdisciplinary R&D projects for young researchers of the Universidad Carlos III de Madrid 2019-2020 in the framework of the Multiannual Agreement Community of Madrid- Universidad Carlos III de Madrid (DEEPCT-CM-UC3M project) and by CRUE Universities, CSIC and Banco Santander (Supera Fund, RADCOV19 project). The CNIC is funded by the Ministry of Science, Innovation and Universities and the PRO-CNIC foundation and is a Severo Ochoa centre of excellence (SEV-2015-0505).

The imaging and associated clinical data downloaded from MIDRC (The Medical Imaging Data Resource Center) and used for research in this publication was made possible by the National Institute of Biomedical Imaging and Bioengineering (NIBIB) of the National Institutes of Health under contracts 75N92020C00008 and 75N92020C00021.

References

[1] Sean D Rose, Emil Y Sidky, Ingrid S Reiser, and Xiaochuan Pan. Filtered back-projection for digital breast tomosynthesis with 2d filtering. In *Medical Imaging 2019: Physics of Medical Imaging*, volume 10948, page 1094851. International Society for Optics and Photonics, 2019.

[2] Sean D Rose, Adrian A Sanchez, Emil Y Sidky, and Xiaochuan Pan. Investigating simulation-based metrics for characterizing linear iterative reconstruction in digital breast tomosynthesis. *Medical physics*, 44(9):e279–e296, 2017.

[3] Alejandro Rodriguez-Ruiz, Jonas Teuwen, Suzan Vreemann, Ramona W Bouwman, Ruben E van Engen, Nico Karssemeijer, Ritse M Mann, Albert Gubern-Merida, and Ioannis Sechopoulos. New reconstruction algorithm for digital breast tomosynthesis: better image quality for humans and computers. *Acta Radiologica*, 59(9):1051–1059, 2018.

[4] Donghoon Lee and Hee-Joung Kim. Restoration of full data from sparse data in low-dose chest digital tomosynthesis using deep convolutional neural networks. *Journal of digital imaging*, 32(3):489–498, 2019.

[5] Tsutomu Gomi, Hidetake Hara, Yusuke Watanabe, and Shinya Mizukami. Improved digital chest tomosynthesis image quality by use of a projection-based dual-energy virtual monochromatic convolutional neural network with super resolution. *PLoS One*, 15(12):e0244745, 2020.

[6] Nikita Moriakov, Koen Michielsen, Jonas Adler, Ritse Mann, Ioannis Sechopoulos, and Jonas Teuwen. Deep learning framework for digital breast tomosynthesis reconstruction. In *Medical Imaging 2019: Physics of Medical Imaging*, volume 10948, page 1094804. International Society for Optics and Photonics, 2019.

[7] Ting Su, Xiaolei Deng, Jiecheng Yang, Zhenwei Wang, Shibo Fang, Hairong Zheng, Dong Liang, and Yongshuai Ge. Dir-dbtnet: Deep iterative reconstruction network for three-dimensional digital breast tomosynthesis imaging. *Medical Physics*, 2021.

[8] Monica Abella, Estefania Serrano, Javier Garcia-Blas, Ines García, Claudia De Molina, Jesus Carretero, and Manuel Desco. Fux-sim: Implementation of a fast universal simulation/reconstruction framework for x-ray systems. *PloS one*, 12(7):e0180363, 2017.

[9] Olaf Ronneberger, Philipp Fischer, and Thomas Brox. U-net: Convolutional networks for biomedical image segmentation. In *International Conference on Medical image computing and computer-assisted intervention*, pages 234–241. Springer, 2015.

[10] Kaiming He, Xiangyu Zhang, Shaoqing Ren, and Jian Sun. Deep residual learning for image recognition. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pages 770–778, 2016.

[11] Wenzhe Shi, Jose Caballero, Ferenc Huszár, Johannes Totz, Andrew P Aitken, Rob Bishop, Daniel Rueckert, and Zehan Wang. Real-time single image and video super-resolution using an efficient sub-pixel convolutional neural network. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pages 1874–1883, 2016.

[12] Olga Russakovsky, Jia Deng, Hao Su, Jonathan Krause, Sanjeev Satheesh, Sean Ma, Zhiheng Huang, Andrej Karpathy, Aditya Khosla, Michael Bernstein, et al. Imagenet large scale visual recognition challenge. *International journal of computer vision*, 115(3):211–252, 2015.

[13] Simon Kornblith, Jonathon Shlens, and Quoc V Le. Do better imagenet models transfer better? In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 2661–2671, 2019.

[14] Justin Johnson, Alexandre Alahi, and Li Fei-Fei. Perceptual losses for real-time style transfer and super-resolution. In *European conference on computer vision*, pages 694–711. Springer, 2016.

[15] Diederik P Kingma and Jimmy Ba. Adam: A method for stochastic optimization. *arXiv preprint arXiv:1412.6980*, 2014.

[16] Leslie N Smith. Cyclical learning rates for training neural networks. In *2017 IEEE winter conference on applications of computer vision (WACV)*, pages 464–472. IEEE, 2017.

Iterative Intraoperative Digital Tomosynthesis Image Reconstruction using a Prior as Initial Image

Fatima Saad, Robert Frysch, Tim Pfeiffer, Sylvia Saalfeld, Jessica Schulz, Jens-Christoph Georgi, Andreas Nürnberger, Guenter Lauritsch, and Georg Rose

***Abstract*—The efficacy of interventional treatments highly relies on an accurate identification of the target lesions and the interventional tools in the guidance images. Whereas X-ray radiography poses low doses to the patient, its weakness is in the superposition of the different image structures in a 2D image. Cone-beam computed tomography (CBCT) might look ideal providing exact 3D information, however this is at the cost of a higher radiation dose, longer imaging time, and more space requirements in the operating room. Introducing some depth information with relatively low dose, and requiring less space, digital tomosynthesis (DTS) is a potential candidate for guiding interventions. However, due to the few number of projections and to the limited angle acquisition, DTS has poor depth resolution. Since high quality patient-specific prior CT scans are usually performed prior to the intervention for diagnosis or to plan the intervention, and given that such images share a fair amount of information with the intraoperative DTS images, we propose in this work a prior-based iterative reconstruction framework to improve the intraoperative DTS image quality. The framework is based on registering the prior CT image to an intermediate low-quality intraoperative DTS image, then iteratively re-reconstructing the intraoperative DTS image using the coregistered prior CT as the starting image. We acquired prior CT and intraoperative CBCT data of a liver phantom and simulated some intraoperative DTS projection images using a spherical ellipse scan geometry. Our results show a great improvement in the DTS image quality with the proposed method and prove the importance of choosing a good starting point for the iterative DTS reconstruction.**

***Index Terms*—Digital tomosynthesis, iterative reconstruction, prior information**

I. Introduction

IMPOSING less space restrictions and less radiation dose than cone-beam computed tomography (CBCT) on the one hand, and introducing depth information to X-ray radiography on the other hand, digital tomosynthesis (DTS) could potentially make a leap in guiding interventions. During this procedure, the C-arm acquires a sequence of few projections over a limited angular range and provides a quasi 3D image. On top of the streaking and ringing artifacts due to the low number of projections, DTS images suffer from geometric distortions due to the limited angle acquisition. The resolution in the reconstructed volume is anisotropic, more precisely, DTS images have poor resolution in the depth direction. For this reason, DTS has not reached yet its application in real clinical interventions where having high quality guidance images is crucial in order to navigate the interventional tool (e.g. needle) to the target location or to confirm the tool-in-lesion. In this context, we recently proposed a new C-arm-based multi-directional spherical ellipse DTS scan trajectory to improve the depth resolution of the standard linear DTS [1]. In this work we investigate the further improvement of intraoperative DTS image quality by exploiting patient-specific prior information. Usually, prior to the surgical intervention, a high quality patient-specific CT scan is performed for diagnosis or for treatment planning. A key observation is that the prior CT image and the intraoperative one share a fair amount of anatomical information. The only deformations between the two images are caused by patient motion, and the positioning of the surgical instrument. Many iterative prior-based reconstruction approaches have been proposed for CT problems with insufficient data [2], [3]. However, the priors have been always included in the body of the iteration whereas the initialization has been overlooked. As we are dealing with a highly under-determined problem with DTS reconstruction where the nullspace is huge and only a tiny part of the image information is measured, the choice of the starting image is potentially crucial. Therefore, we propose in this work a framework based on the registration of the prior CT image to an intermediate intraoperative DTS image followed by an iterative reconstruction of the intraoperative DTS image using the co-registered prior as a first guess.

F. Saad, R. Frysch, T. Pfeiffer, and G. Rose are with the Institute for Medical Engineering, Otto-von-Guericke University, Magdeburg, Germany and also with the Forschungscampus STIMULATE, Otto-von-Guericke University, Magdeburg, Germany.

S. Saalfeld is with the Faculty of Computer Science, Otto-von-Guericke University, Magdeburg, Germany and also with the Forschungscampus STIMULATE, Otto-von-Guericke University, Magdeburg, Germany.

J. Schulz, J.C. Georgi, and G. Lauritsch are with Siemens Healthcare GmbH, Forchheim, Germany.

A. Nürnberger is with the Data and Knowledge Engineering Group, Faculty of Computer Science, Otto-von-Guericke University, Magdeburg, Germany.

II. Materials and Methods

A. Data

The Quant CT-Training Phantom from Cascination[1] was used in this work. It includes a liver with some lesions, a spine, the ribs, and a portal vein. The prior CT volume was acquired at 70 KVp with a Siemens SOMATOM X.cite scanner (Siemens Healthcare GmbH, Erlangen, Germany). The acquired image dimensions were $768 \times 768 \times 266$ voxels with a voxel size of $0.479mm \times 0.479mm \times 1mm$. The intraoperative CBCT volume was acquired using a Siemens ARTIS Icono C-arm scanner (Siemens Healthcare GmbH, Erlangen, Germany) at 87.5 KVp. The CBCT image dimensions were $512 \times 512 \times 368$ voxels with a voxel size of

[1]https://www.cascination.com/en/quant-training-phantom

$0.49mm \times 0.49mm \times 0.49mm$. To mimic an interventional scenario, an ablation needle was inserted in the phantom before acquiring the intraoperative CBCT volume.

To simulate the intraoperative DTS projection images, the CBCT volume was forward-projected according to a newly proposed spherical ellipse scan geometry [1]. This protocol assumes that the source and the detector are mounted on a robotic C-arm and each moves along a spherical ellipse orbit. Fig. 1 illustrates the spherical ellipse acquisition setup. Mounted in opposite to the X-ray source, the detector performs an in-plane rotation in a way its rows are kept tangent to the scan trajectory. The detector is composed of 616×480 pixels with a 0.616 mm pixel pitch. The source to iso-center distance is set to 785 mm and the source to detector distance is set to 1200 mm. This trajectory is defined by two tomographic angles. In this work, we fixed the angles to $\pm 15°$ and $\pm 23°$ and simulated 72 projection views. The system acquisition has been implemented in the open-source Computed Tomography Library CTL[2] [4].

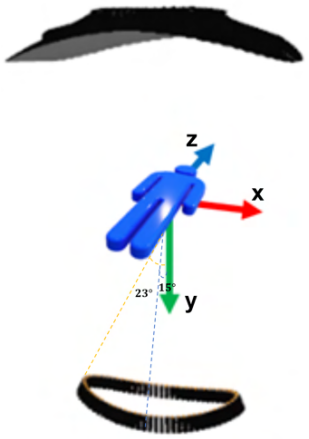


Fig. 1. Illustration of the spherical ellipse DTS scan trajectory.

B. Registration and Reconstruction

The proposed framework is illustrated in Fig. 2. To register the prior CT volume to the current intraoperative anatomy, an intermediate intraoperative DTS image is reconstructed from the intraoperative DTS projections. An iterative algebraic reconstruction technique (ART) with an ordered-subsets scheme [5] is used. The relaxation parameter is estimated as in [6]. At this stage the initial image voxels are set to zero. It was shown in [7] that the complexity of DTS resolution characteristics limits the accuracy of DTS-CT registration and a DTS-DTS registration is more precise. Therefore, from the prior CT volume we generated a set of reference digitally reconstructed radiographs (DRRs) using the same intraoperative DTS scan geometry described above and we reconstructed a reference DTS image. At this step the prior CT image

[2]code available at: https://gitlab.com/tpfeiffe/ctl/-/tree/master/examples/TomosynthesisEllipticalTrajectory

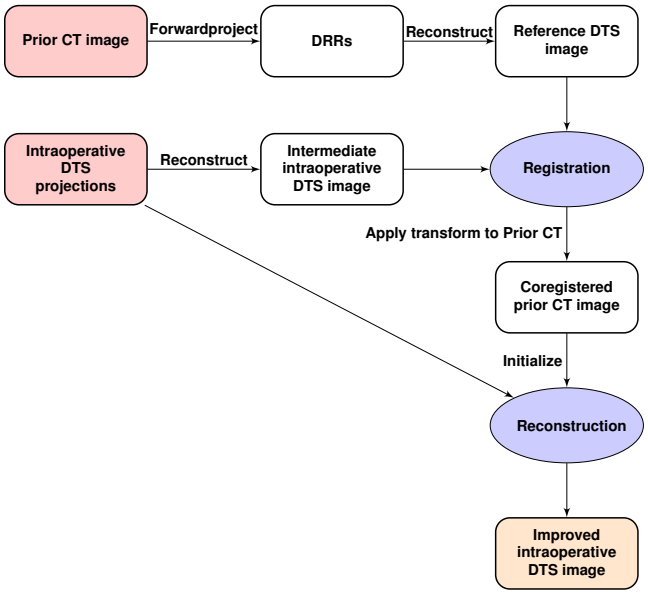


Fig. 2. Flowchart of the proposed reconstruction method.

was downsampled to $512 \times 512 \times 177$ voxels with a voxel size of $0.7187mm \times 0.7187mm \times 1.5mm$. Here it worth be noted that the reconstruction of the reference DTS image can be performed before the intervention as soon as the prior CT image is available. We performed the registration of the reference DTS image to the intraoperative DTS image using a rigid registration employing the Mattes mutual information as a similarity measure. We used the registration toolbox ANTsPy [8]. The registration transform was then applied to the prior CT image. The intraoperative DTS projections are iteratively re-reconstructed using the co-registered prior CT as a first guess and including only a positivity constraint.

III. Results

The registration results are shown in Fig. 3. One axial slice and one coronal slice where the needle tip and a liver lesion exist are shown in the upper row and the lower row respectively. Fig. 3(a) and 3(b) illustrate the prior CT volume and the intermediate intraoperative DTS volume (zero-initialized ART) respectively. The co-registered prior CT volume is shown in Fig. 3(c). For reference the same slice of the intraoperative CBCT volume is shown in Fig. 3(d). Despite that we cannot examine the difference image due to the difference in intensity values between the CT and the CBCT images, it is evident from the visual inspection that the prior CT volume is well registered to the intraoperative volume. The liver, the spine, and the lesions seem accurately aligned in (c) and (d).

Fig. 4 illustrates the reconstruction results. For comparison the same axial slice is shown in the upper row and the same coronal slice is shown in the lower row. Fig. 4(a) shows the DTS reconstruction with zero-initialized ART, without including the prior. Fig. 4(b) illustrates the reconstructed image with our proposed method using the co-registered prior CT as a starting image. For reference the intraoperative CBCT

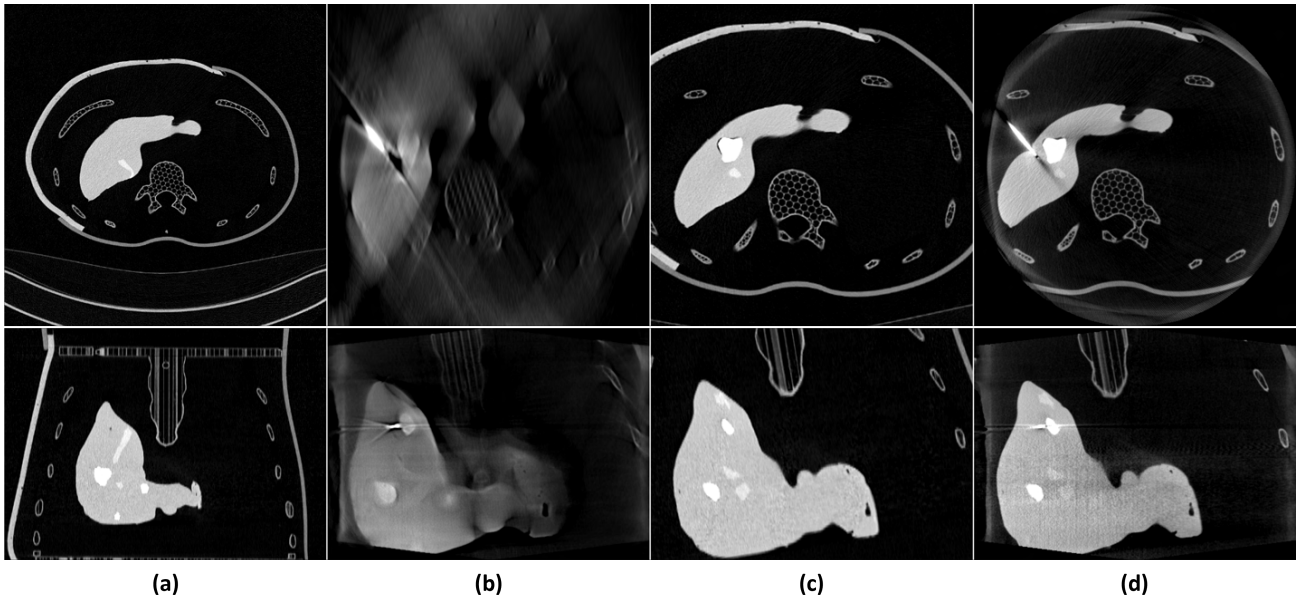


Fig. 3. Registration results. (a) Prior CT, (b) intermediate intraoperative tomosynthesis, (c) coregistered prior CT, and (d) intraoperative CBCT. Upper row: one axial slice. Lower row: one coronal slice. Same slices are shown in (b), (c), and (d). The display window range is [-1000HU, 700HU].

image is shown in Fig. 4(c). As expected, without including the prior, DTS exhibits poor resolution in the depth direction, the lesion is blurred and barely visible and the spine as well. A considerable improvement in image quality is observed with the prior-initialized reconstruction. It looks much similar to the intraoperative CBCT. These observations are further validated in the line profiles plotted in Fig. 5 corresponding to the vertical (upper graph) and the horizontal (lower graph) lines highlighted in yellow in Fig. 4(c).

IV. Discussion and Conclusion

In this work we proposed a prior-based iterative reconstruction framework to improve intraoperative digital tomosynthesis (DTS) reconstruction. We considered a high-quality prior CT volume is available prior to the intervention. We simulated intraoperative DTS projection images using a spherical ellipse DTS geometry [1]. The proposed framework is composed of two steps. First, the registration of the prior CT to the current anatomy, at this stage a reference prior DTS image simulated from the prior CT and an intermediate low-quality reconstructed intraoperative DTS image are co-registered, the registration transform is then applied to the prior CT. Second, an iterative ART reconstruction using the co-registered prior CT as a first guess is performed to further improve the intraoperative DTS image quality. The reference prior DTS can be reconstructed beforehand prior to the intervention as soon as the prior CT is available. While prior information have been incorporated in the iterative process itself in previous works, the initialization was mostly overlooked. In this work, we showed how important is the starting point of the iterative reconstruction in such a highly under-determined problem. A considerable improvement in the different structures visibility has been achieved. A still open question is how sensible is the reconstruction to insufficiently accurate registration. By the design of the method, information measured from the tomosynthesis trajectory are included in the prior-initialized ART and all missing information are inherited from the co-registered prior. In case of inaccuracies in the registration, this may lead to wrong conclusions while interpreting the images. Therefore, clinical experts will be included in the evaluation process in the future. In the current work, we only considered a rigid motion between the prior and intraoperative states, this is probably valid for head interventions, however in many other interventions, such as in the lungs, the abdomen, and the liver, theses organs are prone to elastic motions. Considering such complex motions is the subject of current works. Moreover, how low can we go with the angular range and the number of projections will be further investigated.

Acknowledgment

This work was in part conducted within the context of the International Graduate School MEMoRIAL at the Otto von Guericke University (OVGU) Magdeburg, Germany, kindly supported by the ESF (project no. ZS/2016/08/80646) and was partly funded by the Federal Ministry of Education and Research within the Forschungscampus STIMULATE under grant number 13GW0473A. The concepts and information presented in this paper are based on research and are not commercially available.

References

[1] F. Saad, R. Frysch, T. Pfeiffer, J.-C. Georgi, T. Knetsch, R. F. Casal, A. Nürnberger, G. Lauritsch, and G. Rose, “Spherical ellipse scan trajectory for tomosynthesis-assisted interventional bronchoscopy,” in

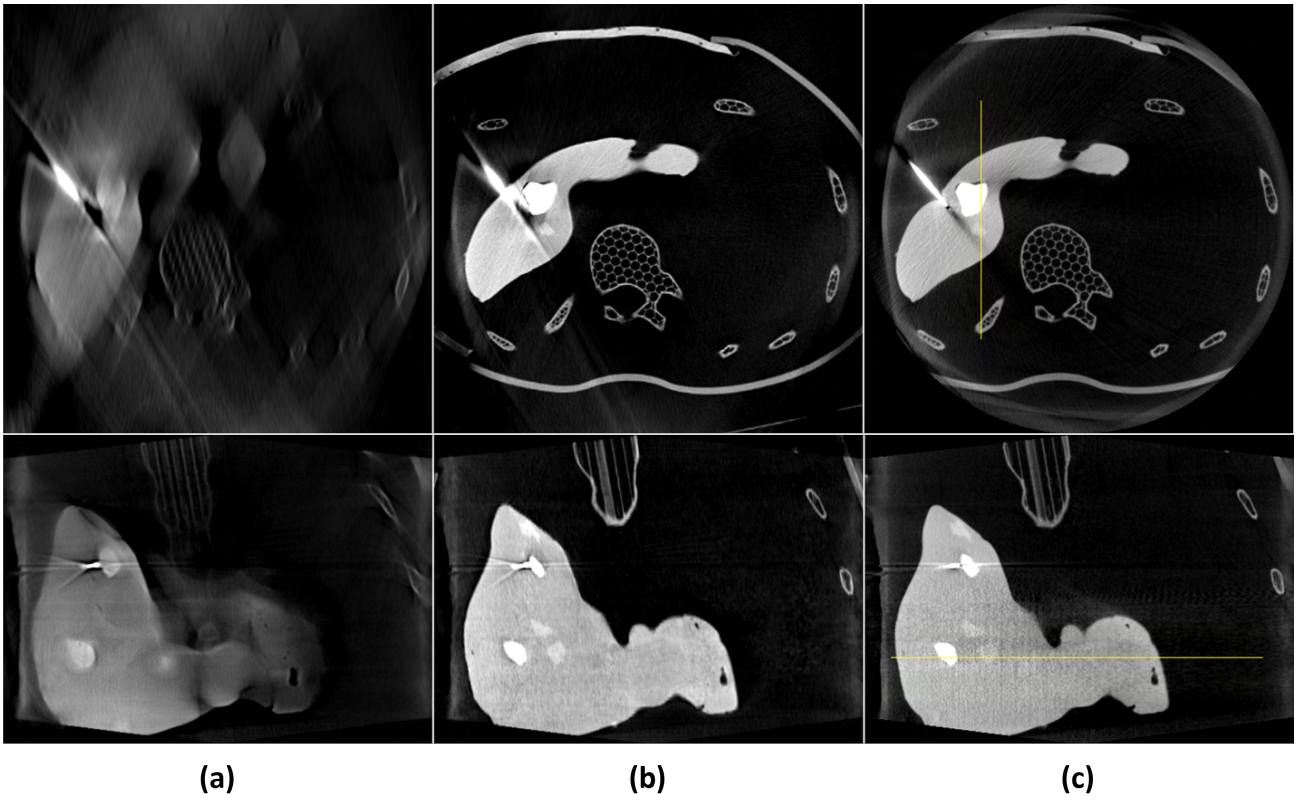


Fig. 4. Reconstruction results. (a) zero-initialized ART, (b) proposed method (prior-initialized ART), and (c) intraoperative CBCT. Upper row: one axial slice. Lower row: one coronal slice. Same slices are shown in (a), (b), and (c). The display window range is [-1000HU, 700HU].

Proceedings of the 16th Virtual International Meeting on Fully 3D Image Reconstruction in Radiology and Nuclear Medicine, G. Schramm, A. Rezaei, K. Thielemans, and J. Nuyts, Eds., 2021, pp. 352–356.

[2] G.-H. Chen, J. Tang, and S. Leng, "Prior image constrained compressed sensing (piccs): a method to accurately reconstruct dynamic ct images from highly undersampled projection data sets," *Medical physics*, vol. 35, no. 2, pp. 660–663, 2008.

[3] E. Y. Sidky, C.-M. Kao, and X. Pan, "Accurate image reconstruction from few-views and limited-angle data in divergent-beam ct," *Journal of X-ray Science and Technology*, vol. 14, no. 2, pp. 119–139, 2006.

[4] T. Pfeiffer, R. Frysch, R. N. Bismark, and G. Rose, "Ctl: modular open-source c++-library for ct-simulations," in *15th International Meeting on Fully Three-Dimensional Image Reconstruction in Radiology and Nuclear Medicine*, vol. 11072. International Society for Optics and Photonics, 2019, p. 110721L.

[5] D. Kim, D. Pal, J.-B. Thibault, and J. A. Fessler, "Improved ordered subsets algorithm for 3d x-ray ct image reconstruction," *Proc. 2nd Intl. Mtg. on image formation in X-ray CT*, pp. 378–81, 2012.

[6] S. Bannasch, R. Frysch, R. Bismark, G. Warnecke, and G. Rose, "An Optimal relaxation of the Algebraic Reconstruction Technique for CT Imaging," in *Fully Three-Dimensional Image Reconstruction in Radiology and Nuclear Medicine : proceedings*, Newpoer, Rhode Island, USA, Jun. 2015, pp. 622–625.

[7] D. J. Godfrey, L. Ren, H. Yan, Q. Wu, S. Yoo, M. Oldham, and F. F. Yin, "Evaluation of three types of reference image data for external beam radiotherapy target localization using digital tomosynthesis (dts)," *Medical physics*, vol. 34, no. 8, pp. 3374–3384, 2007.

[8] B. B. Avants, N. Tustison, G. Song *et al.*, "Advanced normalization tools (ants)," *Insight j*, vol. 2, no. 365, pp. 1–35, 2009.

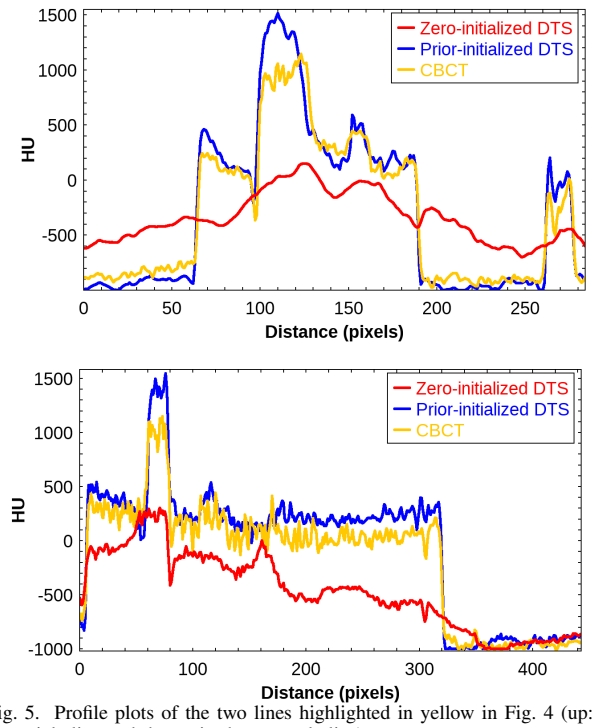


Fig. 5. Profile plots of the two lines highlighted in yellow in Fig. 4 (up: in the axial slice and down: in the coronal slice).

Learning CT Scatter Estimation Without Labeled Data – A Feasibility Study

Joscha Maier, Luca Jordan, Elias Eulig, Fabian Jäger, Stefan Sawall, Michael Knaup, and Marc Kachelrieß

***Abstract*—Since x-ray scattering is a major cause of artifacts, its correction is a crucial step in almost any CT application. Most existing approaches, however, are based on complex theoretical models that need to be tailored to that particular application. To perform scatter estimation in absence of such models, we propose the unsupervised deep scatter estimation (uDSE). Here, uDSE combines a scatter estimation network that operates in projection domain with a scatter correction layer and CT reconstruction layer. In that way scatter estimation can be trained using an unsupervised Wassersten GAN (WGAN) setup in which the parameters of the scatter estimation network are optimized such that the resulting scatter corrected reconstructions cannot be distinguished from samples of a true artifact-free reference set. To demonstrate the feasibility of the proposed approach, uDSE is evaluated for simulated CBCT scans. Applied to the corresponding test data, uDSE is able to remove most of the present scatter artifacts and yields similar CT value accuracy (mean error of 27.9 HU vs. 24.7 HU) as a state-of-the-art supervised scatter estimation approach. Thus, uDSE may be used in the future to learn scatter estimation in cases where labels are not available or cannot be generated with sufficient accuracy.**

***Index Terms*—CT, scatter estimation, deep learning, unsupervised learning.**

I. Introduction

THE contribution of scattered x-rays to the acquired projection data leads to a violation of CT reconstruction criteria, and thus, to the introduction of CT artifacts. In particular this holds true for cone-beam CT (CBCT), where scatter-to-primary ratios may easily be in the order of 1 and above. Therefore, scatter correction is a crucial preprocessing step to achieve diagnostic image quality. Typically, existing approaches implement such a correction by deriving an estimate of the present scatter distribution and by subtracting it subsequently from the acquired raw data. Here, the scatter distribution can either be estimated using dedicated hardware such as beam blockers or primary modulation grids [1]–[5],or using software-based approaches. The latter rely on physical, empirical, or consistency-based models to predict x-ray scattering [6]–[13],or more recently, on neural networks that make use of such models during training, e.g. by being trained to reproduce the output of Monte Carlo simulations [14]–[18]. While these approaches have proven great potential in terms of accuracy and computation time, their performance highly depends on the quality and the availability of labeled data. Since several applications may lack such data, this study proposes the unsupervised deep scatter estimation (uDSE) which is able to overcome this limitation. Instead of using a supervised setup, uDSE relies on a Wasserstein GAN (WGAN) setup that can be trained without labeled data or prior knowledge about the CT scanner's x-ray and scatter properties [19]. Here we demonstrate the feasibility of the proposed approach using simulated CBCT data and compare the results against DSE, our supervised scatter estimation approach [14], [15].

Dr. J. Maier, L. Jordan, and Dr. M. Knaup are with the German Cancer Research Center (DKFZ), Heidelberg, Germany.
E. Eulig, F. Jäger, Dr. S. Sawall and Prof. Dr. M. Kachelrieß are with the German Cancer Research Center (DKFZ), Heidelberg, Germany and with the Heidelberg University, Heidelberg, Germany
Corresponding author: Dr. Joscha Maier (joscha.maier@dkfz.de)

II. Material and Methods

A. Deep Scatter Estimation (DSE)

The basic idea of the DSE approach is illustrated in figure 1. Here, DSE uses a U-net-like architecture to predict scatter as a function of the acquired projection data. To learn the corresponding mapping, DSE is trained to reproduce Monte Carlo simulations, i.e. the U-net's weights are determined by minimizing the following loss function:

$$L_{\mathrm{DSE}}(\theta) = \sum_{n}^{B} \left| \frac{\mathrm{DSE}_\theta(I_n) - S_n}{S_n} \right|, \tag{1}$$

where θ denotes the parameter vector, n the sample number within a batch of size B, I_n the flat field-corrected intensities, and S_n the Monte Carlo scatter estimate. It has to be noted that the first layer of the DSE network performs a "pep"-transform

$$T_{\mathrm{pep}} : I \to -I \cdot \ln(I), \tag{2}$$

to be consistent with our DSE publication [14].

B. Unsupervised Deep Scatter Estimation (uDSE)

The proposed uDSE approach, shown in figure 1, extends the concept of DSE to cases where labeled data are not available. To do so, it is composed of a generator network and a critic network. Here, the generator combines the DSE network with a scatter correction layer and a Feldkamp reconstruction layer, such that it is able to map acquired intensities I of a CT scan to scatter-corrected CT reconstructions. The critic network, in turn, is designed to distinguish between the generator's output and true scatter-free CT reconstructions. Thus, letting the critic network act as loss function for the generator allows to

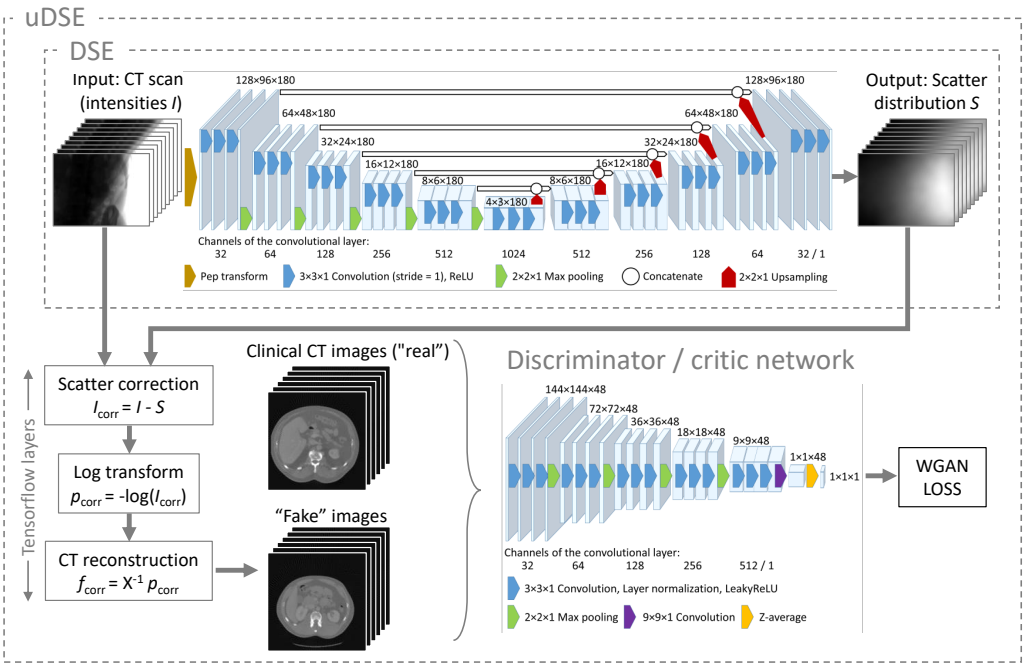


Fig. 1. Schematic of the DSE and the proposed uDSE approach. DSE uses a U-net-like architecture to predict scatter as a function of the acquired projection data. In a supervised setup DSE is trained to reproduce a ground truth Monte Carlos scatter distribution. uDSE, in contrast, uses a Wasserstein GAN (WGAN) setup. Here, a scatter correction followed by a CT reconstruction is performed. Subsequently, the correction is evaluated by a critic network that is trained simultaneously to recognize scatter artifact-free images. By optimizing the weights of the scatter estimation network to fool the critic network, scatter estimation can be learned without labels.

learn CT scatter estimation without labeled or paired data, respectively.

Here, the corresponding optimization is performed using a WGAN setup in which the generator netork $G_{\theta_g}(I)$ and the critic network $C_{\theta_c}(f)$ are optimized in an alternating manner according to the following loss functions:

$$L_{\text{critic}}(\theta_c) = \sum_{n}^{B} C_{\theta_c}(G_{\theta_g}(I_n)) - C_{\theta_c}(f_{\text{real, n}}), \quad (3)$$

$$L_{\text{gen}}(\theta_g) = -\sum_{n}^{B} C_{\theta_c}(G_{\theta_g}(I_n)), \quad (4)$$

where θ_c and θ_g denote the parameter vectors of the generator and the critic network, n is the sample number within a batch of size B, and f_{real} corresponds to a sample from a set of almost scatter-free clinical CT reconstructions.

C. Datasets

In the present study, DSE as well as uDSE were trained and tested on simulated CBCT data. Therefore, clinical CT reconstructions of 65 patients were used as prior. Based on the corresponding voxel volumes f_{prior}, CBCT scans with 360 views and an angular coverage of 360° were simulated at five different z-positions within the abdomen region using a tube voltage of 120 kV, a source-to-isocenter distance of 700 mm, a source-to-detector distance of 1100 mm, and a 1024×768 flat detector with an isotropic pixel spacing of 0.39 mm. Furthermore, a shifted detector was used to increase the field of measurement to about 380 mm. Using this setup, three datasets were generated using different patients: one scatter-corrupted dataset (30 patients) from which input data for the generator network were sampled during training, one scatter-free dataset (30 patients) that was used to provide ideal reference CT reconstructions for the critic network, and a scatter-corrupted dataset (5 remaining patients) for testing. In any case, the scatter-corrupted data were simulated as $I = I_{\text{p}} + S_{\text{MC}}$, where I_p corresponds to a polychromatic forward projection of the prior volume and S_{MC} is a scatter distribution that was generated using our in-house Monte Carlo simulation. The ideal reference data, on the other hand, correspond to a CBCT reconstruction of only the primary intensities, i.e. $f_{\text{real}} = X^{-1}(-\ln(I_{\text{p}}))$, with X^{-1} being the CBCT reconstruction operator. It has to be noted that f_{real} is not sampled directly from the set of clinical CT reconstructions, but is generated via forward and backprojection, to have the same spatial resolution and the same field of measurement as the reconstructions provided by the generator network.

D. Training and Evaluation

The uDSE approach was trained using the datasets described in section II-C. However, to avoid memory issues as well as to increase the computational performance, a 2-fold angular downsampling to 180 views and an 8-fold spatial downsampling to a detector size of 128 × 96 was performed in advance. This downsampling can be

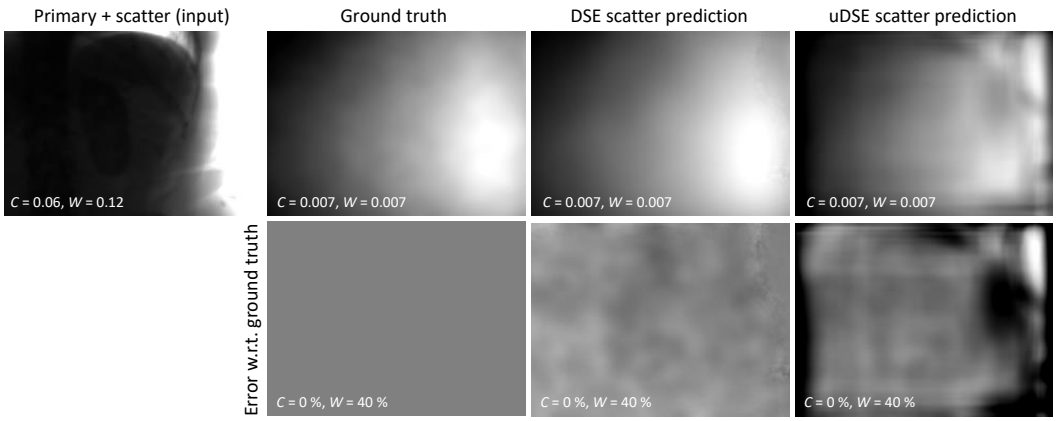


Fig. 2. DSE and uDSE scatter estimates for an exemplary shifted detector projection of the test dataset (top row), as well as the corresponding error with respect to the ground truth Monte Carlo scatter distribution (bottom row).

justified by the fact that scatter distributions are known to be of low frequency. Therefore, we only expect a minor degradation of accuracy compared to a training that uses the full size projection data. Given the small detector size, the internal reconstruction operations were also performed on a low resolution grid with $144 \times 144 \times 48$ voxels and an isotropic spacing of 2.7 mm.

The corresponding optimization was implemented according to section II-B using the Tensorflow framework. Here, all hyperparameters except for the batch size (this study: $B = 16$) and the number of iterations (this study: $N_{\text{itr}} = 5000$) were chosen following to the original WGAN publication, i.e. five updates of the critic network are followed by one update of the generator network using an RMSProp optimizer with a learning rate of 0.00005 and a weight clipping to $[-0.01, 0.01]$ in the critic network.

Finally, uDSE was applied to the test data and compared against DSE which was trained in a supervised setting using the Monte Carlo scatter distributions a labels. In contrast to the training data, the scatter correction was performed on the full size projection data. Therefore, all scatter predictions were upsampled to the original detector size of 1024×768 pixels prior to scatter correction.

III. Results

To estimate scatter, DSE and uDSE were trained as described in section II-D and evaluated for the five patients of the test dataset. Exemplary scatter predictions are shown in figure 2. Here, the DSE approach yields scatter estimates that are almost equal to the Monte Carlo ground truth while uDSE shows slightly higher deviations. This can be attributed to the fact the uDSE loss function is evaluated in image domain. Therefore, the scatter estimate is less reliable in regions where it has a low impact on image quality, i.e. in regions with low scatter-to-primary ratio. In particular this explains the poor accuracy of uDSE scatter estimates in air regions without patient intersection.

A quantitative evaluation of the scatter estimates in terms of the mean absolute percentage error with respect to the ground truth yields similar trends. Here, the average error of DSE for all 9000 projections of the test dataset is 3.5 % while the average error of uDSE is 9.8 %.

CT reconstructions with and without scatter correction are shown in figure 3. Here, DSE as well as uDSE are able to remove most of the artifacts that are present in the uncorrected reconstruction and provide CT images that are almost equal to the scatter-free ground truth. Quantitatively, the application of the scatter correction improves the mean absolute error of the CT values from 160.7 HU (no correction) to 24.7 HU (DSE) or 27.9 HU (uDSE), respectively.

IV. Discussion and Conclusion

This study introduces a novel approach to learn scatter estimation without labeled data. To do so, the proposed uDSE makes use of a WGAN setup in which the generator network is optimized such that its output, i.e. scatter corrected reconstructions, cannot be distinguished from samples of an artifact-free reference set. Here, we demonstrate the feasibility of uDSE using CBCT simulations as input and clinical CT reconstructions as reference. However, it has to be noted that uDSE is not restricted to this particular choice but can be trained with any tomographic input and any scatter-free reference as long as both distributions are sufficiently equal after scatter correction. The fact that this condition is perfectly met here, i.e. input and reference distributions only differ by scatter, can be considered as a limitation of this study. Practically, both distribution may additionally differ by the amount of beam hardening, the dynamic range, or the contrast media distribution for instance. Investigating uDSE's performance in such cases is subject to further research and may require the incorporation of additional constraints and correction layers. In the current setup, however, uDSE is able to remove most of the present scatter artifacts and yields similar CT value accuracy (mean error of 27.9 HU vs. 24.7 HU) as a state-of-the-art supervised scatter estimation approach. Thus, uDSE

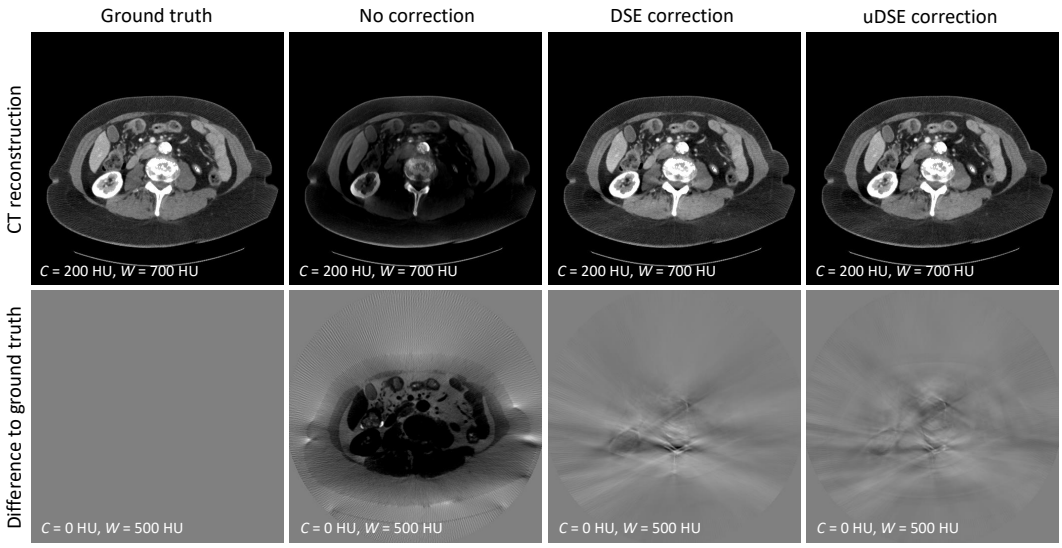


Fig. 3. CT reconstruction without and with DSE and uDSE scatter correction (top row), as well as difference images to the scatter free ground truth (bottom row).

has the potential to extend the concept of neural network-based scatter estimation and correction to scenarios where labels are not available or cannot be generated with sufficient accuracy.

Acknowledgment

Parts of the reconstruction software RayConStruct IR and of the Monte Carlo software RayConStruct MC were provided by RayConStruct® GmbH, Nürnberg, Germany.

References

[1] R. Ning, X. Tang, and D. Conover, "X-ray scatter correction algorithm for cone beam CT imaging," *Medical Physics*, vol. 31, no. 5, pp. 1195–1202, apr 2004.

[2] L. Zhu, N. R. Bennett, and R. Fahrig, "Scatter Correction Method for X-Ray CT Using Primary Modulation: Theory and Preliminary Results," *IEEE Transactions on Medical Imaging*, vol. 25, no. 12, pp. 1573–1587, dec 2006.

[3] R. Grimmer, R. Fahrig, W. Hinshaw, H. Gao, and M. Kachelrieß, "Empirical Cupping Correction for CT Scanners with Primary Modulation (ECCP)," *Medical Physics*, vol. 39, no. 2, pp. 825–831, jan 2012.

[4] L. Ritschl, R. Fahrig, M. Knaup, J. Maier, and M. Kachelrieß, "Robust primary modulation-based scatter estimation for cone-beam CT," *Medical Physics*, vol. 42, no. 1, pp. 469–478, jan 2015.

[5] S. Hsieh, "Estimating scatter in cone beam CT with striped ratio grids: A preliminary investigation," *Medical Physics*, vol. 43, no. 9, pp. 5084–5092, aug 2016.

[6] W. Swindell and P. M. Evans, "Scattered radiation in portal images: A Monte Carlo simulation and a simple physical model," *Medical Physics*, vol. 23, no. 1, pp. 63–73, jan 1996.

[7] B. Ohnesorge, T. Flohr, and K. Klingenbeck-Regn, "Efficient object scatter correction algorithm for third and fourth generation CT scanners," *European Radiology*, vol. 9, no. 3, pp. 563–569, mar 1999.

[8] G. Poludniowski, P. M. Evans, V. N. Hansen, and S. Webb, "An efficient Monte Carlo-based algorithm for scatter correction in keV cone-beam CT," *Physics in Medicine and Biology*, vol. 54, no. 12, pp. 3847–3864, jun 2009.

[9] J. Star-Lack, M. Sun, A. Kaestner, R. Hassanein, G. Virshup, T. Berkus, and M. Oelhafen, "Efficient scatter correction using asymmetric kernels," in *SPIE Medical Imaging Conference*, E. Samei and J. Hsieh, Eds., vol. 7258, feb 2009, p. 72581Z.

[10] M. Sun and J. M. Star-Lack, "Improved scatter correction using adaptive scatter kernel superposition," *Physics in Medicine and Biology*, vol. 55, no. 22, pp. 6695–6720, nov 2010.

[11] E. Meyer, C. Maaß, M. Baer, R. Raupach, B. Schmidt, and M. Kachelrieß, "Empirical scatter correction (ESC): A new CT scatter correction method and its application to metal artifact reduction," in *IEEE Medical Imaging Conference*. IEEE, oct 2010, pp. 2036–2041.

[12] M. Baer and M. Kachelrieß, "Hybrid scatter correction for CT imaging," *Physics in Medicine and Biology*, vol. 57, no. 21, pp. 6849–6867, nov 2012.

[13] A. Maslowski, A. Wang, M. Sun, T. Wareing, I. Davis, and J. Star-Lack, "Acuros CTS: A fast, linear Boltzmann transport equation solver for computed tomography scatter - Part I: Core algorithms and validation," *Medical Physics*, vol. 45, no. 5, pp. 1899–1913, may 2018.

[14] J. Maier, E. Eulig, T. Vöth, M. Knaup, J. Kuntz, S. Sawall, and M. Kachelrieß, "Real-time scatter estimation for medical CT using the deep scatter estimation: Method and robustness analysis with respect to different anatomies, dose levels, tube voltages, and data truncation." *Medical Physics*, vol. 46, no. 1, pp. 238–249, jan 2019.

[15] J. Maier, S. Sawall, M. Knaup, and M. Kachelrieß, "Deep Scatter Estimation (DSE): Accurate Real-Time Scatter Estimation for X-Ray CT Using a Deep Convolutional Neural Network," *Journal of Nondestructive Evaluation*, vol. 37, no. 3, pp. 57:1–9, sep 2018.

[16] Y. Nomura, Q. Xu, H. Shirato, S. Shimizu, and L. Xing, "Projection-domain scatter correction for cone beam computed tomography using a residual convolutional neural network," *Medical Physics*, p. mp.13583, jun 2019.

[17] Y. Jiang, C. Yang, P. Yang, X. Hu, C. Luo, Y. Xue, L. Xu, X. Hu, L. Zhang, J. Wang, K. Sheng, and T. Niu, "Scatter correction of cone-beam CT using a deep residual convolution neural network (DRCNN)," *Physics in Medicine and Biology*, vol. 64, no. 14, p. 145003, jul 2019.

[18] J. Erath, T. Vöth, J. Maier, E. Fournié, M. Petersilka, K. Stierstorfer, and M. Kachelrieß, "Deep learning-based forward and cross-scatter correction in dual-source CT," *Medical Physics*, vol. 48, no. 9, pp. 4824–4842, sep 2021.

[19] M. Arjovsky, S. Chintala, and L. Bottou, "Wasserstein Generative Adversarial Networks," in *Proceedings of the 34th International Conference on Machine Learning*, ser. Proceedings of Machine Learning Research, D. Precup and Y. W. Teh, Eds., vol. 70. PMLR, 2017, pp. 214–223.

Implementations of Statistical Reconstruction Algorithm for CT Scanners with Flying Focal Spot

Robert Cierniak and Jarosław Bilski and Piotr Pluta

***Abstract*—This paper presents some practical realizations of image reconstruction methods for spiral cone-beam tomography scanners in which an X-ray tube with a flying focal spot is used. These methods are related to the original formulated 3D statistical model-based iterative reconstruction approach for tomography with flying focal spot. The conception proposed here is based on principles of a statistical model-based iterative reconstruction (MBIR) methodology, where the reconstruction problem is formulated as a shift-invariant system (a continuous-to-continuous data model). We adopted nutating reconstruction-based approaches, i.e. the advanced single slice rebinning methodology (usually applied in CT scanners with X-ray tubes with a flying focal spot), and a procedure compliant with the FDK scheme. We showed that our methods significantly improve the quality of obtained images compared to the traditional FBP algorithms. Consequently, it can allow for a reduction in the x-ray dose absorbed by a patient. Additionally, we show that our approach can be competitive in terms of the time of calculations, especially if we consider commercially used statistical reconstruction systems.**

***Index Terms*—spiral computed tomography flying focal spot, statistical iterative reconstruction algorithm.**

I. Introduction

DESPITE the long history of medical computed tomography, the search for new designs of CT scanners still continues, and at the beginning of the XXIth century, one such new design was the spiral scanner with an X-ray tube with a flying focal spot [1]. The intention of this new technique was to increase the sampling density of the integral lines in the reconstruction planes, and the density of simultaneously acquired slices in the longitudinal direction. Of course, this technique is realized in multidetector row CT (MDCT) scanners but there it allows for view-by-view deflections of the focal spot in the rotational α-direction (αFFS) and/or in the longitudinal z-direction (zFFS). Thanks to this, it is possible to improve the quality of the reconstructed images, mainly throughout decreasing the influence of the aliasing effect in the reconstruction plane and in the z-direction, and additionally, causing a reduction of the so-called windmill artifacts. Obviously, implementation of the FFS entailed the necessity of formulating new reconstruction methods that allow for the use of measurements obtained from scanners equipped with this technique. In practice, manufacturers decided primarily to modify the adaptive multiple plane reconstruction (AMPR) method for this purpose. The AMPR conception belongs to the class of so-called nutating reconstruction methods, and is a development of the advanced single slice rebinning (ASSR) algorithm (for details see e.g. [**?**]). Generally, nutating methods have several serious drawbacks, among others, there is a problem with obtaining equi-spaced resolution of the slices in z-direction, due to the constant change in the position of successive reconstruction planes. Another big problem is its limited ability to suppress noise, caused by the linear form of the filters that are used during signal processing in those methods. This means that it cannot be considered for CT systems that aim to reduce the dose of X-ray radiation absorbed by patients during examinations. Recent research in the area of X-ray computed tomography, including our own investigations, is mostly focused on this challenge, because of the extremely harmful effect that CT examinations have on human health. On the other hand, nowadays, the most interesting research directions in the area of CT are statistical reconstruction approaches, especially those belonging to the model-based iterative reconstruction (MBIR) class of methods [3], and some commercial solutions have been developed, which conduct the reconstruction process in an iterative way, which aims to suppress noise in the obtained images. Unfortunately, the MBIR methods used commercially (the iterative coordinate descent (ICD) algorithm described comprehensively in [4]) have some serious drawbacks, namely, the calculation complexity of the problem is approximately proportional to I^4, where I is the image resolution, and the iterative reconstruction procedure based on this conception necessitates simultaneous calculations for all the voxels in the range of the reconstructed 3D image. Moreover, the size of the forward model matrix $\mathbf{A}$ is extremely large and it has to be calculated online. The reconstruction problem used there is also extremely ill-conditioned [5]. The methodology mentioned above is classified as a method based on the discrete-to discrete (D-D) data model. All those drawbacks can be reduced by using an approach that is formulated based on a continuous-to-continuous (C-C) data model. It should be underlined, that our conception has some significant advantages over that based on the D-D model. First of all, in our method, the forward model is formulated as a shift invariant system, which allows for the use of FFT algorithms in the most computationally demanding elements of the reconstruction algorithm (realization of the 2D convolutions in the frequency domain). Furthermore, we can pre-calculate the model matrix (coefficients), i.e. we establish it before the algorithm is started. Additionally, the reconstruction process can be carried out in only one plane in 2D space, which greatly simplifies the reconstruction problem, and it is possible to obtain every slice of the body separately. Our approach also outperforms the D-D method already at the

The project financed under the program of the Polish Minister of Science and Higher Education under the name "Regional Initiative of Excellence" in the years 2019 - 2022 project number 020/RID/2018/19 the amount of financing 12,000,000 PLN

Czestochowa University of Technology, Department of Intelligent Computer Systems, Armii Krajowej 36, Czestochowa, Poland, 42-200.

level of problem formulation, regarding the better condition number [5]. Finally, the presented here concept using the statistical reconstruction approach developed by us, can be in easy way used for spiral CT scanners with a flying focal spot technique.

II. Reconstruction Algorithms

Our reconstruction approach is based on the well-known maximum-likelihood (ML) estimation, where an optimization formula is consistent with the C-C data model, as follows:

$$\mu_{\min} = \arg\min_{\mu} \left(\int_x \int_y \left(\int_{\bar{x}} \int_{\bar{y}} \mu\left(\bar{x},\bar{y}\right) \cdot h_{\Delta x,\Delta y} d\bar{x} d\bar{y} - \tilde{\mu}\left(x,y\right) \right) \right) \tag{1}$$

where $\tilde{\mu}\left(x,y\right)$ is an image obtained by way of a back-projection operation (without filtration), and the coefficients $h_{\Delta i,\Delta j}$ can be precalculated according to the following relation:

$$h_{\Delta x,\Delta y} = \int_0^{2\pi} int\left(\Delta x \cos\alpha + \Delta y \sin\alpha\right) d\alpha, \tag{2}$$

and $int\left(\Delta s\right)$ is a linear interpolation function.

The presence of a shift-invariant system in the optimization problem (1) implies that this system is much better conditioned than the least squares problems present in the referential approach [5].

According to the formulated by us iterative approach to the reconstruction problem, decribed by Eqs (1)-(2), it is possible to show a practical model-based statistical method of image reconstruction, as follows:

$$\mu_{\min} = \arg\min_{\mu} \left(\sum_{i=1}^{\mathrm{I}} \sum_{j=1}^{\mathrm{I}} \left(\sum_{\bar{i}=1}^{\mathrm{I}} \sum_{\bar{j}=1}^{\mathrm{I}} \mu^{*}\left(x_{\bar{i}},y_{\bar{j}}\right) \cdot h_{\Delta i,\Delta j} - \tilde{\mu}\left(x_i,y_j\right) \right)^2 \right), \tag{3}$$

where I is a resolution of the reconstructed image, $\tilde{\mu}\left(i,j\right)$ is an image obtained by way of a back-projection operation, and coefficients $h_{\Delta i,\Delta j}$ are determined according to the following formula:

$$h_{\Delta i,\Delta j} = \frac{1}{\Delta s^2} \Delta_{\alpha} \sum_{\psi=0}^{\Psi-1} int\left(\Delta i \cos\psi\Delta_{\alpha} + \Delta j \sin\psi\Delta_{\alpha}\right), \tag{4}$$

wherein $int\left(\Delta s\right)$ is an interpolation function used in the back-projection operation, and $\Delta_s = R_f * \tan\Delta_{\beta}$.

Figure 1 depicts this general algorithm after discretization and implementation of FFT which significantly accelerates the calculations.

Between all approaches to image reconstruction problem using projections obtained in the spiral scanner with flying focal spot there is only one difference: determination of the image obtained after back-projection operation. Up to now, we have formulated the following approaches to determine the image $\tilde{\mu}\left(i,j\right)$: method which uses spiral cone-beam projections directly, similar to FDK-type algorithms [6], methods belonging to the class of nutating reconstruction algorithms (based on the advanced single slice rebinning) methodology, using projections from different focal spots separately [7], and together (unpublished).

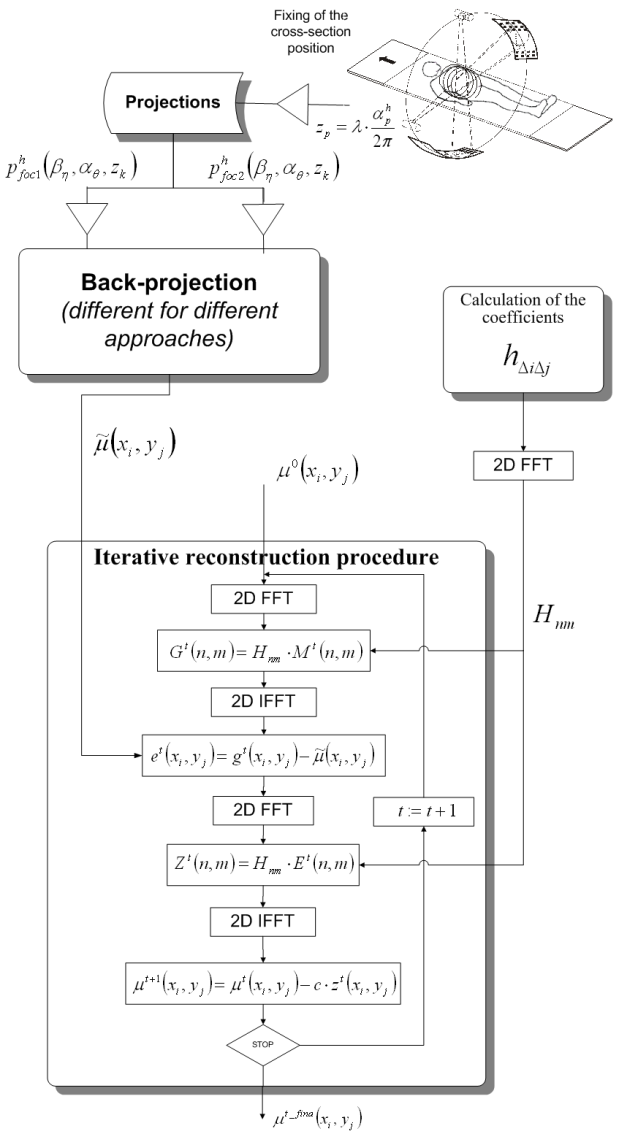


Fig. 1. General statistical reconstruction algorithm for spiral cone-beam scanner with flying focal spot.

III. Experimental results

All our experiments were carried out using projections obtained from a Somatom Definition AS+ (helical mode) scanner with the following parameters: reference tube potential 120kVp and quality reference effective 200mAs. The geometrical parameters of this scanner are as follows: $R_{fd} =$

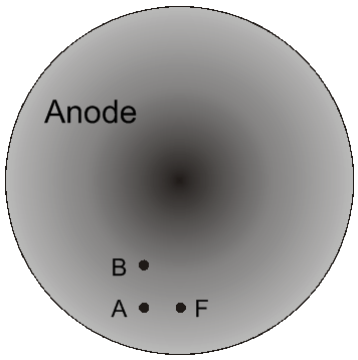


Fig. 2. The movement of focal spot with flying spot technique used in our experiments.

TABLE I
COMBINATION OF α-DIRECTION AND z-DIRECTION DEVIATIONS

Parameter/Focal spot	R^{fs}	α	$\dot{z}$
A	0	$\delta\alpha^A$	0
B	δR^B	$\delta\alpha^B$	$\delta\dot{z}^B$

$1085.6mm$, $\mathrm{R}_f = 595\mathrm{mm}$, number of views per rotation $\Psi = 1152$, number of pixels in detector panel 736, detector dimensions $1.09\mathrm{mm} \times 1.28\mathrm{mm}$. During experiments, we used the spiral CT scanner with only two flying spot positions, as shown in Fig. 2.

The focus position A is involved with movement in α-direction, and the focus position B corresponds to the focal movement both in α-direction and in z direction. The deviations of the geometrical parameters are summarized in Table I.

The size of the reconstructed image was fixed at 512×512 pixels. A discrete representation of the matrix $h_{\Delta x,\Delta y}$ was established in a computational way before the reconstruction process was started. These coefficients were fixed (transformed into the frequency domain) and used for the whole iterative reconstruction procedure. A prepared result of an FBP reconstruction algorithm was chosen as the starting point of the iterative reconstruction procedure (using projections obtained from the focal spot position A).

A crucial parameter for the practical implementation of a reconstruction method is the actual computation time of the reconstruction procedure. We have implemented our iterative reconstruction procedure using some hardware configurations, namely: a computer with 10 cores, i.e. with an Intel i9-7900X BOX/3800MHz processor (our iterative procedure was implemented at assembler level), using different GPUs (see III). In table II, we show time result for application which is working only on CPU which is develop in Assembler (special vector registers AVX 512 used). In turn, in table III, we present time result for application which is working only on GPU accelerators. There are compared those accelerators. It is worth noting that it is very stable time, because deviation is extremely small and that application is very susceptible to parallelisation, because time for one iteration it is getting smaller

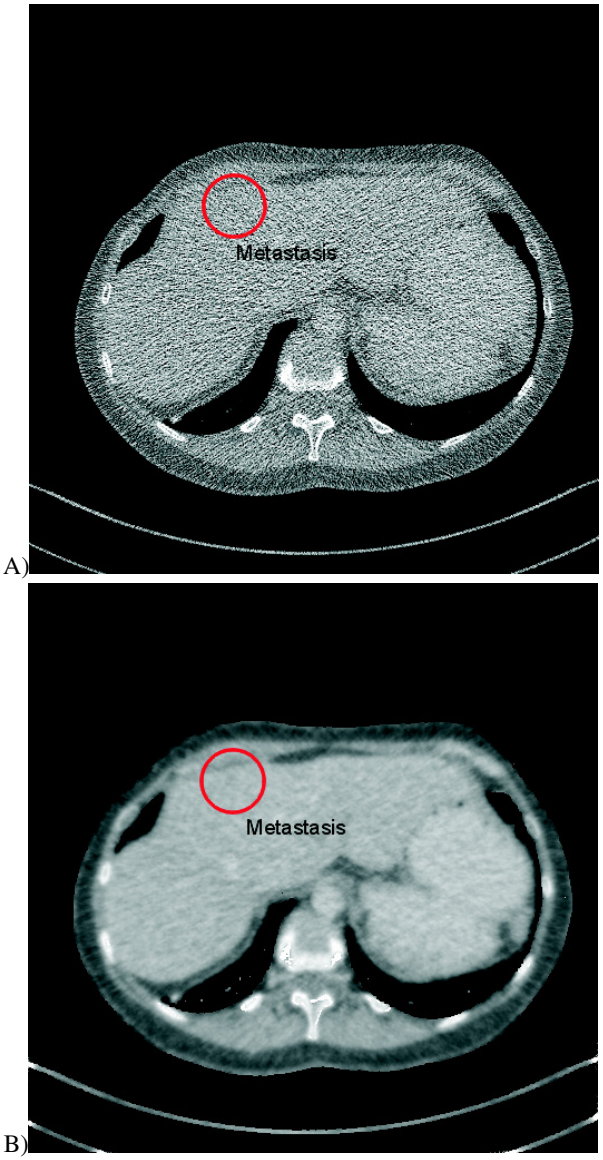


Fig. 3. Obtained images (a case with relative small pathological change in the liver) using quarter-dose projections with application of: the standard FDK algorithm (focal spot position A) (A); the statistical method presented in this paper (FDK scheme with both focal spot positions) (B).

with on more CUDA Cores assembled in GPU Accelerator.

According to an assessment of the quality of the obtained images by a radiologist, 7000 iterations are enough to provide an acceptable image for medical purposes. One can compare the results obtained by assessing the views of the reconstructed images in Figure 3, where where the quarter-dose projections were used.

IV. CONCLUSION

In this paper, it has been shown that our statistical approach, which was originally formulated for CT scanner with parallel beam geometry, can be adapted for helical scanner with flying focal spot technique. We have presented a fully feasible

TABLE II
TIMES RESULTS OF RECONSTRUCTION IMAGE ON MULTI THREADING CPU: INTEL I9-7900X (10-CORES, 20-TREADS). AN APPLICATION CREATED IN THE ASSEMBLER PROGRAMMING LANGUAGE WITH MULTITHREADING

Threads:	4	8	10	16	20
Avg. time 30000[ms]	63 724,36	33 571,42	29 836,34	30 532,14	27 905,62
Avg. time 20000[ms]	42 482,91	22 380,95	19 890,89	20 354,76	18 603,75
Avg. time 10000[ms]:	21 241,45	11 190,47	9 945,45	10 177,38	9 301,87
Time 1 iteration [ms]:	2,124145	1,119047	0,994545	1,017738	0,930187
HT effectiveness:	-	-	-	0,909468	0,935290
Median for 30000:	63 694	33 542,5	29 800	30 566	27 854
Deviation std.:	135,69	117,32	217,58	193,88	391,76

TABLE III
TIMES RESULTS OF RECONSTRUCTION IMAGE ON DIFFERENT MODELS GPU ACCELERATOR. AN APPLICATION CREATED IN THE CUDA PROGRAMMING LANGUAGE

GPU:	MSI GTX 1050	ASUS GTX 1080 Ti	nVidia Titan V
Avg. time 30000[ms]	2 562 175,10	49 699,71	28 858,40
Avg. time 20000[ms]	170 845,28	33 132,52	19 224,48
Avg. time 10000[ms]:	85 467,24	16 593,00	9 616,75
Time 1 iteration [ms]:	8,540583	1,656657	0,961947
Median for 30000:	256 229,55	49 703,68	28 861,24
Deviation std.:	0,160806	0,310476	0,010239

statistical reconstruction for cone-beam CT. Comprehensive experiments have been performed, which prove that our reconstruction method is relatively fast (thanks to the use of FFT algorithms) and gives satisfactory results with suppressed noise. It should be noted that approximately the same results were achieved for both hardware implementations: the iterative reconstruction procedure takes less than 7s, mainly thanks to the use of an FFT algorithm in the iterative reconstruction procedure and to the use of the efficient programming techniques. These are rewarding results regarding possibilities of the commercial Veo system (referential MBIR technique), where reconstruction times range between 10 to 90 minutes depending on the number of reconstructed slices [8]. It means an unacceptable delay between data acquisition and availability for interpretation for emergent indications. Additionally, all formulated by us reconstruction algorithms are very easy to implement and open to use multisource techniques regardless the kind of the focal spot movement (both z and angle flying).

ACKNOWLEDGMENT

The authors thank Dr. Cynthia McCoullough and the American Association of Physicists in Medicine for providing the Low-Dose CT Grand Challenge dataset.

This work was partly supported by The National Centre for Research and Development in Poland (Research Project POIR.01.01.01-00-0463/17).

REFERENCES

[1] M. Kachelrieß, M. Knaup, C. Penßel, W. Kalender, *Flying focal spot (FFS) in cone-beam CT*, IEEE Transactions on Nuclear Science 53 (3), 1238–1247, 2006.

[2] M. Kachelrieß, S. Schaller, W. A. Kalender, *Advanced single-slice rebinning in cone-beam spiral CT*, Med. Phys. 27, 754–773, 2000.

[3] Y. Zhou, J.-B. Thibault, C.A. Bouman, J. Hsieh, K.D. Sauer, *Fast model-based x-ray CT reconstruction using spatially non-homogeneous ICD optimization*, IEEE Tran. Im. Proc. 20, 161–175, 2011.

[4] J. -B Thibault, K. D. Sauer, C. A. Bouman, J. Hsieh, *A three-dimensional statistical approach to improved image quality for multislice helical CT*, Med. Phys. 34, 4526–4544, 2007.

[5] R. Cierniak, A. Lorent, *Comparison of algebraic and analytical approaches to the formulation of the statistical model-based reconstruction problem for x-ray computed tomography*, Computerized Medical Imaging and Graphics 52, 19–27, 2016.

[6] R. Cierniak, et al, *A new statistical reconstruction method for the computed tomography using an x-ray tube with flying focal spot*, Journal of Artificial Intelligence and Soft Computing Research 11, 271–286, 2021.

[7] R. Cierniak, P. Pluta, A. Kaźmierczak, *A practical statistical approach to the reconstruction problem using a single slice rebinning method*, Journal of Artificial Intelligence and Soft Computing Research 10, 137–149, 2020.

[8] L. L. Geyer, et al, *State of the art: iterative CT reconstruction techniques*, Radiology 276 (2), 339–357, 2015.

Multiple Linear Detector Off-Line Calibration

Sasha Gasquet[1,2], Laurent Desbat[1], and Pierre-Yves Solane[2]

***Abstract*—Imaging systems require to be calibrated. The geometric calibration consists in estimating several parameters describing the projection geometry. Just like in computer vision for cameras, intrinsic parameters characterize the internal parameters of x-ray projection system. The extrinsic parameters define the orientation and position of the acquisition system. In x-ray computed tomography (CT), the acquisition systems are generally composed of a detector and a x-ray source. The object to be reconstructed lies in-between. A perfect knowledge of the calibration parameters is needed for the reconstruction algorithm to reduce artefacts. In this paper, we focus on off-line calibration methods for 1D linear x-ray detector systems. We first introduce a calibration method for systems composed of a single linear detector. This method solves the problem of calibration in two steps using calibration objects based on four co-planar lines. Moreover, we generalize the single linear detector geometric calibration method to a multi-linear detector system. We compare four different numerical models and methods. Three are based on non-linear equation systems. Finally, we propose an adaptative calibration object.**

***Index Terms*—Calibration, linear detector, multi-linear detectors, computed tomography**

I. Introduction

A high accuracy geometric calibration of the acquisition system is required to perform the 3D reconstruction of an object from its projections. The algorithms rely on the perfect knowledge of the intrinsic and extrinsic parameters of the system. In x-ray cone-beam CT (CBCT), these parameters describe the relation between a 3D point and its projection point on the detector image plane. Therefore, inaccurate estimations will lead to a poor reconstruction.

Off-line calibration methods of systems using 2D detector are well-known in the computer vision literature [1]. Many methods are based on the data acquired using a perfectly known geometrical object called the calibration object. In computer vision, a well-known calibration object is the Tsai grid [2]. In x-ray CBCT, calibration methods are adapted from computer vision. Calibration objects, well suited to the circular CB geometry, composed of several balls of high density material, have been designed. The projections of the balls form an ellipse from which the calibration parameters can be estimated [3] [4] [5].

Many computer vision methods are based on a pinhole camera model. The model relates a 3D point (X, Y, Z) lying in the scene to a 2D point (u, v) on the detector. The intrinsic parameters $\alpha_u = k_u f$, $\alpha_v = k_v f$, u_0 and v_0 are contained in the calibration matrix $\mathbf{K}_{2D} \in \mathbb{R}^{3\times4}$ where f is the focal distance, k_u and k_v are the densities of pixels along the image axes **u** and **v**, respectively, and (u_0, v_0) is the principal point on the detector in the image coordinates.

$$\mathbf{K}_{2D} = \begin{bmatrix} \alpha_u & 0 & u_0 & 0 \\ 0 & \alpha_v & v_0 & 0 \\ 0 & 0 & 1 & 0 \end{bmatrix} \tag{1}$$

We define the rotation matrix $\mathbf{R} \in \mathbb{R}^{3\times3}$ and the translation vector $\mathbf{t} \in \mathbb{R}^{3\times1}$ representing the orientation and position of the camera. Therefore, the pinhole camera model is defined in the Eq. (2). The parameter s is a scale factor.

$$\begin{bmatrix} su \\ sv \\ s \end{bmatrix} = \mathbf{K}_{2D} \begin{bmatrix} \mathbf{R} & \mathbf{t} \\ \mathbf{0}^T & 1 \end{bmatrix} \begin{bmatrix} X \\ Y \\ Z \\ 1 \end{bmatrix} \tag{2}$$

Geometric calibration is the identification of $\mathbf{K}_{2D}$, **R** and **t**. In computer vision, $\mathbf{K}_{2D}$, **R** and **t** are often estimated from sufficient projections (u, v) of 3D world points (X, Y, Z) using (2). However, these model and methods must be adapted to calibrate a linear detector. A calibration object composed of lines is more suitable to linear cameras or detectors [6]. Horaud *et al.* proposed a two-step calibration method based on four co-planar lines calibration objects. The adaptation of this method to a x-ray system with a linear detector is straight forward as both systems can be described with the same geometric pinhole model.

In this paper, we adapt and generalize the Horaud *et al.* computer vision calibration method for linear camera to a multiple linear detector x-ray system. We propose a calibration object with a minimal number of opaque lines. A total of four different methods are proposed. In addition, we present a calibration object which can be adapted to different configurations of detectors. Finally, we show the performances of the proposed methods and calibration object in numerical simulations.

II. Theory

A. Geometry

The system is composed of n_D linear detectors, denoted D_l, $l = 1, \ldots, n_\mathrm{D}$, and an unique x-ray source denoted S. We denote $(\mathbf{O}, \mathbf{x}, \mathbf{y}, \mathbf{z})$ the world coordinate system centred at the origin **O**. An illustration of such a system is given in the Fig. 1. Furthermore, we introduce the detector coordinate system associated to the l^{th} detector $(\mathbf{O}_l, \mathbf{u}_l, \mathbf{v}_l, \mathbf{w}_l)$ where $\mathbf{v}_l$ is the image axis, $\mathbf{w}_l$ is the axis perpendicular to $\mathbf{v}_l$ pointing towards the source and $\mathbf{u}_l = \mathbf{v}_l \times \mathbf{w}_l$. The origin of the system is the point $\mathbf{O}_l$ which is the orthogonal projection of the source S on the detector D_l. The real v_l is the coordinate of $\mathbf{O}_l$ along the linear detector, along $\mathbf{v}_l$, relative to the pixel 0 of D_l.

[1]Univ. Grenoble Alpes, CNRS, UMR 5525, VetAgro Sup, Grenoble INP, TIMC, 38000 Grenoble, France

[2]TIAMA, 215 chemin du Grand Revoyet, F-69230 Saint-Genis-Laval, France

Contact : s.gasquet@tiama.com

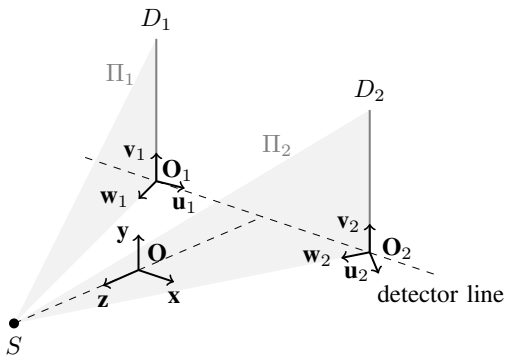


Fig. 1. A two linear detector system.

B. Single detector calibration

We start by introducing the single linear detector system calibration method derived from Horaud *et al* [6].

1) Pinhole linear camera model: Let's consider a point (X, Y, Z) expressed in the world coordinates system and its projection v on the linear detector D. To be seen, the point has to belong to the viewing plane Π which is defined in the Eq. (3) using three real parameters p, q and r.

$$X = pY + qZ + r \tag{3}$$

Moreover, we adjust the pinhole camera model presented in the Eq. (2) to the system using a 1D detector. The rotation matrix $\mathbf{R} \in \mathbb{R}^{3\times3}$ and the translation vector $\mathbf{t} \in \mathbb{R}^{3\times1}$ are representing the rigid transformation from the world to the source coordinates. The calibration matrix $\mathbf{K}_{2D} \in \mathbb{R}^{3\times4}$ defined in the Eq. (1) becomes $\mathbf{K}_{1D} \in \mathbb{R}^{2\times4}$.

$$\mathbf{K}_{1D} = \begin{bmatrix} 0 & \alpha_v & v_0 & 0 \\ 0 & 0 & 1 & 0 \end{bmatrix} \tag{4}$$

The pinhole linear camera model is given by:

$$\begin{bmatrix} sv \\ s \end{bmatrix} = \mathbf{K}_{1D} \begin{bmatrix} \mathbf{R} & \mathbf{t} \\ \mathbf{0}^T & 1 \end{bmatrix} \begin{bmatrix} X \\ Y \\ Z \\ 1 \end{bmatrix} \tag{5}$$

Within the plane Π, *i.e.* using Eq. (3), we can rewrite the Eq. (5) such that the calibration problem is reduced to the estimation of five real parameters n_1, n_2, n_3, n_4, n_5 from :

$$(n_1 - n_4 v)Y + (n_2 - n_5 v)Z + n_3 - v = 0 \tag{6}$$

and the three parameters p, q, r from Eq. (3).

2) Calibration object: The calibration object is made of four co-planar lines. Three of them are parallel and the last one is oblique. For example, we can define these lines within the plane $Z = 0$ by the following equations.

$$\begin{aligned} (L_1) \quad & Y = 0 \\ (L_2) \quad & Y = \xi_1 \\ (L_3) \quad & Y = \xi_2 \\ (L_4) \quad & Y = \alpha X + \beta \end{aligned} \tag{7}$$

The intersections of these lines with the plane Π and their projections on D are used to solve the calibration problem. A key idea introduced by Horaud *et al* is to use the intersections of (L_1), (L_2) and (L_3) with the plane Π and their projections on D to estimate the intersection point of (L_4) and Π using a projective invariant : the cross-ratio [1]. Therefore, by translating this object several times along the **y** and/or **z** axes, we can acquire enough data to solve the calibration problem.

3) Calibration problem: We denote $\{Y_i^p, Z_i^p, v_i^p\}$ and $\left\{X_j^o, Y_j^o, Z_j^o\right\}$ the sets of known data related to the parallel and oblique lines, respectively, with $i = 1, \ldots, n_\mathrm{P}$ and $j = 1, \ldots, n_\mathrm{O}$, where $n_\mathrm{P} \geq 5$ and $n_\mathrm{O} \geq 3$ are the numbers of parallel and oblique lines, respectively. In Fig. 2 we show a calibration object containing the minimal number of eight calibration lines. Using (6) and the parallels lines set of data, we can estimate the parameters n_1, n_2, n_3, n_4, n_5 by solving a system of n_P equations in the least square sense. Likewise, we estimate the viewing plane parameters p, q, r by solving a system of n_O equations based on the Eq. (3).

4) Calibration parameters estimation: The intrinsic and extrinsic parameters can easily be extracted from n_1, n_2, n_3, n_4 and n_5 using p, q and r (see [6]). Then, the source position (X_S, Y_S, Z_S) can be estimated by solving the linear system (8) where v^* and $v^\diamond$ are two different detector pixels. Indeed, the source belongs to all the backprojection lines and the viewing plane Π.

$$\begin{cases} (n_1 - n_4 v^*)Y_S + (n_2 - n_5 v^*)Z_S + n_3 - v^* = 0 \\ (n_1 - n_4 v^\diamond)Y_S + (n_2 - n_5 v^\diamond)Z_S + n_3 - v^\diamond = 0 \\ -X_S + pY_S + qZ_S + r = 0 \end{cases} \tag{8}$$

C. Multi-detector calibration

The first obvious idea to calibrate a multiple linear detector system would be to use the previous method on each detector. However, all subsystems share the same source. In this section, we present four different methods exploiting this property.
In the following, as in section II-A, the index l refers to the detector D_l, $l = 1, \ldots, n_\mathrm{D}$.

1) Method 1 (M1): We use the previous method on each subsystem to calibrate the system. We then agregate the equations (8) associated to each detector $D_l, l = 1, \ldots, n_\mathrm{D}$ for estimating an unique source position. The agregated system of equations (9) is composed of $3n_\mathrm{D}$ equations and can be easily solved by least squares.

$$\begin{cases} (n_{l,1} - n_{l,4} v_l^*)Y_S + (n_{l,2} - n_{l,5} v_l^*)Z_S = v_l^* - n_{l,3} \\ (n_{l,1} - n_{l,4} v_l^\diamond)Y_S + (n_{l,2} - n_{l,5} v_l^\diamond)Z_S = v_l^\diamond - n_{l,3} \\ -X_S + p_l Y_S + q_l Z_S = -r_l \end{cases} \tag{9}$$

2) Method 2 (M2): From the combination of the equations (3) and (9), we get a system of $(n_\mathrm{O} + 3)n_\mathrm{D}$ non-linear equations (10) in the parameters p_l, q_l, r_l, X_S, Y_S and Z_S, $l = 1, \ldots, n_\mathrm{D}$.

$$\begin{cases} (n_{l,1} - n_{l,4} v_l^*)Y_S + (n_{l,2} - n_{l,5} v_l^*)Z_S = v_l^* - n_{l,3} \\ (n_{l,1} - n_{l,4} v_l^\diamond)Y_S + (n_{l,2} - n_{l,5} v_l^\diamond)Z_S = v_l^\diamond - n_{l,3} \\ -X_S + p_l Y_S + q_l Z_S + r_l = 0 \\ p_l Y_{l,j}^o + q_l Z_{l,j}^o + r_l = X_{l,j}^o, \;\; j = 1, \ldots, n_\mathrm{O} \end{cases} \tag{10}$$

This system can be solved using the Gauss-Newton algorithm. We suggest to initialize the algorithm with the results of M1.

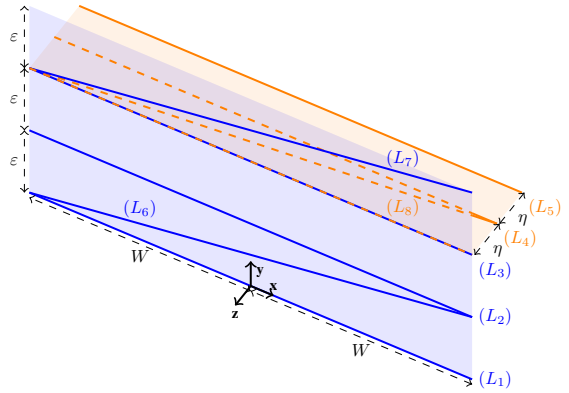


Fig. 2. The minimal calibration object.

3) Method 3 (M3): Similarly to M2, by combining the equations (6) and (9), we build a system of $(n_{\mathrm{P}}+3)n_{\mathrm{D}}$ non-linear equations (11) in the parameters $n_{l,1}$, $n_{l,2}$, $n_{l,3}$, $n_{l,4}$, $n_{l,5}$, X_S, Y_S and Z_S, $l=1,\ldots,n_{\mathrm{D}}$.

$$\begin{cases} (n_{l,1}-n_{l,4}v_l^*)Y_S+(n_{l,2}-n_{l,5}v_l^*)Z_S+n_{l,3}=v_l^* \\ (n_{l,1}-n_{l,4}v_l^\diamond)Y_S+(n_{l,2}-n_{l,5}v_l^\diamond)Z_S+n_{l,3}=v_l^\diamond \\ -X_S+p_lY_S+q_lZ_S=-r_l \\ (n_{l,1}-n_{l,4}v_{l,i}^p)Y_{l,i}^p+(n_{l,2}-n_{l,5}v_{l,i}^p)Z_{l,i}^p+n_{l,3}=v_{l,i}^p, \\ \qquad\qquad i=1,\ldots,n_{\mathrm{P}} \end{cases} \tag{11}$$

We use the same numerical method as for M2 for solving Eq. (11).

4) Method 4 (M4): The last method combines the equations of M2 and M3. Consequently, the system of $(n_{\mathrm{P}}+n_{\mathrm{O}}+3)n_{\mathrm{D}}$ equations (12) to solve is non-linear in the parameters p_l, q_l, r_l, $n_{l,1}$, $n_{l,2}$, $n_{l,3}$, $n_{l,4}$, $n_{l,5}$, X_S, Y_S and Z_S, $l=1,\ldots,n_{\mathrm{D}}$.

$$\begin{cases} (n_{l,1}-n_{l,4}v_l^*)Y_S+(n_{l,2}-n_{l,5}v_l^*)Z_S+n_{l,3}=v_l^* \\ (n_{l,1}-n_{l,4}v_l^\diamond)Y_S+(n_{l,2}-n_{l,5}v_l^\diamond)Z_S+n_{l,3}=v_l^\diamond \\ -X_S+p_lY_S+q_lZ_S+r_l=0 \\ p_lY_{l,j}^o+q_lZ_{l,j}^o+r_l=X_{l,j}^o,\ \ j=1,\ldots,n_{\mathrm{O}} \\ (n_{l,1}-n_{l,4}v_{l,i}^p)Y_{l,i}^p+(n_{l,2}-n_{l,5}v_{l,i}^p)Z_{l,i}^p+n_{l,3}=v_{l,i}^p, \\ \qquad\qquad i=1,\ldots,n_{\mathrm{P}} \end{cases} \tag{12}$$

The resolution can be done using the Gauss-Newton algorithm as for M2 and M3.

D. Calibration object

The use of sufficient four co-planar lines calibration objects presented in the section II-B2 could be enough to generate the data required to solve the calibration problem. Nevertheless, we present in this section several improvements.

1) Minimal calibration object: The first improvement is to construct a minimal calibration object composed of 5 parallel lines and 3 oblique lines. The parallel lines are shared among the oblique lines such that 3 groups of 3 parallel plus one

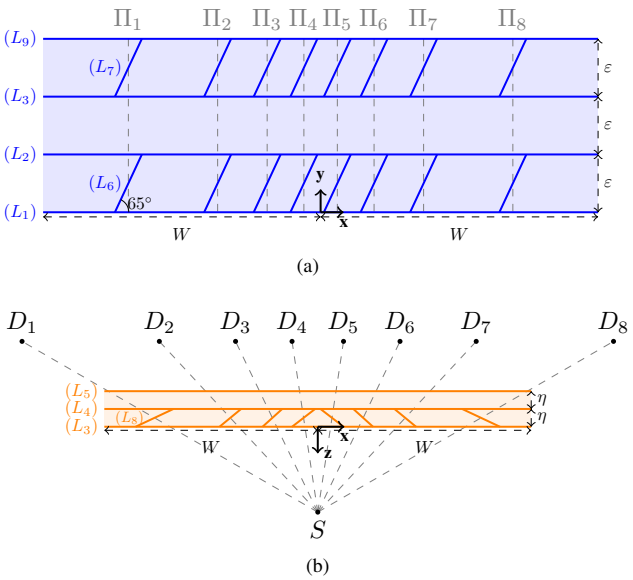


Fig. 3. (a) Vertical plane and (b) horizontal plane of the object used for the 8-detectors system calibration.

oblique line, as in section II-B2, can be provided. The 5 parallel lines are defined by the equations (13).

$$\begin{array}{lll} (L_1) & Z=0, & Y=0 \\ (L_2) & Z=0, & Y=\varepsilon \\ (L_3) & Z=0, & Y=2\varepsilon \\ (L_4) & Z=-\eta, & Y=2\varepsilon \\ (L_5) & Z=-2\eta, & Y=2\varepsilon \end{array} \tag{13}$$

The three oblique lines are defined by the equations (14).

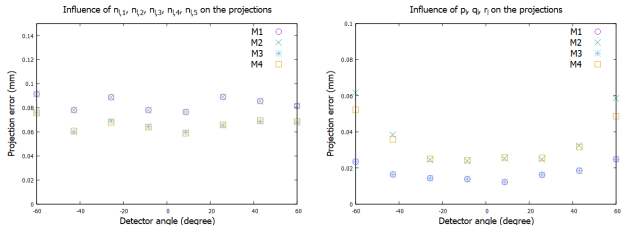
$$\begin{array}{llll} (L_6) & Z=0, & Y=\dfrac{\varepsilon}{2W}X+\dfrac{\varepsilon}{2} \\ (L_7) & Z=0, & Y=\dfrac{\varepsilon}{2W}X+\dfrac{5\varepsilon}{2} \\ (L_8) & Y=2\varepsilon, & Z=-\dfrac{\eta}{2W}X-\dfrac{\eta}{2} \end{array} \tag{14}$$

The minimal calibration object is illustrated in the Fig. 2. One can notice that the group of lines in the plane $Y=2\varepsilon$ doesn't result from a translation as suggested in the section II-B2. The cross-ratio can be adapted in this plane. Besides, we remark that the lines are positioned such that their projections can be spanned all over the detectors using an adequate value of ε.

2) Adapted object: First, we observed that adding to the minimal calibration object the parallel line (L_9) defined in the Eq. (15) improved significantly the accuracy of the calibration.

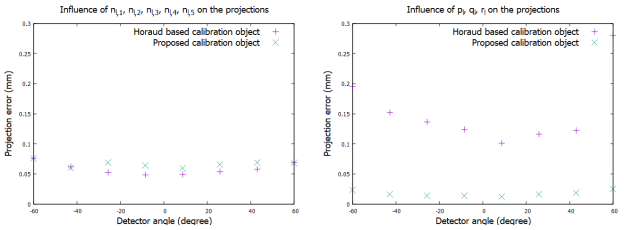
$$(L_9)\quad Z=0,\quad Y=3\varepsilon \tag{15}$$

Then, another improvement is the positioning of the oblique lines relatively to each detector D_l, $l=1,\ldots,n_{\mathrm{D}}$. The lines are placed such that the intersections of the oblique lines and the viewing planes are at equal distance of the two closest co-planar parallel lines. We denote these lines $(L_{6,l})$, $(L_{7,l})$ and $(L_{8,l})$, $l=1,\ldots,n_{\mathrm{D}}$. The lines $(L_{6,l})$ and $(L_{7,l})$ are positioned in the plane $Z=0$ between the lines (L_1) and (L_2), and the lines (L_3) and (L_9), respectively. They are



(a) (b)

Fig. 4. Methods comparison. The graphics show the average projection errors of 2 oblique lines on the 8 detectors using the estimated values of (a) the parameters $n_{l,1}$, $n_{l,2}$, $n_{l,3}$, $n_{l,4}$, $n_{l,5}$ (b) the parameters p_l, q_l, r_l, with $\sigma = 0.24px$, $l = 1, \ldots, n_{\mathrm{D}}$.



(a) (b)

Fig. 5. Calibration objects comparison. The graphics show the average projection errors of 2 oblique lines on the 8 detectors using the estimated values of (a) the parameters $n_{l,1}$, $n_{l,2}$, $n_{l,3}$, $n_{l,4}$, $n_{l,5}$ (b) the parameters p_l, q_l, r_l, with $\sigma = 0.24px$, $l = 1, \ldots, n_{\mathrm{D}}$.

inclined at a fixed angle λ. The lines $(L_{8,l})$ are positioned in the plane $Y = 2\varepsilon$, between the lines (L_3) and (L_4). The inclinations of these lines are specific to each detector. Nevertheless, the position of the lines $(L_{8,l})$ doesn't impact the results as much as the two others obliques lines. Thus, every way of positioning the lines can be used providing that the intersection of the lines and the viewing planes are at $Z = -\frac{\eta}{2}$. An example of such a calibration object is given in the Fig. 3.

III. Simulation

We consider a 8-linear-detector system. The distance from the source to the plane containing the detectors is $480mm$. The detectors are spaced at a constant angle from the source. The calibration object is positioned halfway between the source and the detectors line. The detectors have 1920 pixels of height $0.4mm$. The parameters of the calibration object are set to $W = 600mm$, $\varepsilon = 125mm$ and $\eta = 50mm$. The two planes containing line segments of the calibration object are illustrated in the Fig. 3a and 3b, respectively.

We begin with the comparison of the methods presented in the section II-C. The methods are compared on the average projection errors observed on 100 simulations where we add Gaussian noise $\mathcal{N}(0, \sigma^2)$ to the data. The value of σ is set as a fraction of the pixels size. The average projection errors are calculated by computing the absolute value of the difference between the theoretical projections of 2 oblique lines and the projections obtained using successively the Eq. (3) to compute the intersections of the oblique lines and Π_l, and the Eq. (6) to compute the projections of the intersections points. Two projections are calculated using the estimated values of either the parameters $n_{l,1}$, $n_{l,2}$, $n_{l,3}$, $n_{l,4}$, $n_{l,5}$ or the parameters p_l, q_l, r_l. The results are presented in the Fig. 4. It can be observed in the Fig. 4a that M3 and M4 improve slightly the estimation of the parameters $n_{l,1}$, $n_{l,2}$, $n_{l,3}$, $n_{l,4}$, $n_{l,5}$. We can see in the Fig. 4b that M2 and M4 fail to improve the estimation of the viewing planes parameters whereas M1 and M3 have the lowest errors.

Finally, we compare the results of the proposed calibration object with those obtained with the object presented in the section II-B2. We consider here only one wide object for all the detectors. We set $\xi_1 = \xi_2 = 100mm$, $\alpha = 0.25$ and $\beta = 75$. The object is initially positioned in-between the source and the detectors line. Then, it is shifted twice on the **y** and **z** axes. Exactly, we shift the object successively by $Y_{shift} = 200mm$ along the **y** axis and by $Z_{shift} = -150mm$ along the **z** axis. We use M3 to solve the problem. The results are presented in the Fig. 5. We can see in the Fig. 5a that the objects achieves comparable results on the estimation of the parameters $n_{l,1}$, $n_{l,2}$, $n_{l,3}$, $n_{l,4}$ and $n_{l,5}$. However, we can observe in the Fig. 5b that the estimation of the viewing planes parameters are much better with the calibration object proposed in II-D.

IV. Conclusion

We have extended the linear camera geometric calibration method introduced by Horaud et al [6] (see section II-B) to a multiple linear detector system. We have proposed a minimal calibration phantom of 8 opaque lines. But we have observed that adding one opaque line improves highly the accuracy and the stability of the calibration parameter estimation. This calibration object has been adapted to a multiple linear detector system in order to preserve a sufficient oblicity for the oblique lines. We have proposed and evaluated 4 numerical methods exploiting the fact that all subsystems share the same x-ray source.

Our numerical simulations have shown that estimating the geometric parameters of multi linear detector systems taking into account that they share the same x-ray source, improves the geometric calibration.

References

[1] R. Hartley, and A. Zisserman, "Multiple View Geometry in Computer Vision", *Cambridge University Press*, second edition, 2004.

[2] R. Tsai, "A versatile Camera Calibration Technique for High-Accuracy 3D Machine Vision Metrology Using Off-the-Shelf TV Cameras and Lenses", *IEEE Journal of Robotics and Automation*, vol. RA-3, no. 4, pp. 323-344, 1987.

[3] F. Noo, R. Clackdoyle, C. Mennessier, T. White and T. Roney, "Analytic method based on identification of ellipse parameters for scanner calibration in cone-beam tomography", *Physics in Medicine and Biology*, vol. 45, no. 11, pp. 3489-3508, 2000.

[4] Y. Cho, D. Moseley, J. Siewerdsen and D. Jaffray, "Accurate technique for complete geometric calibration of cone-beam computed tomography systems", *Medical Physics*, vol. 32, no. 4, pp. 968-983, 2005.

[5] M. Daly, J. Siewerdsen, Y. Cho, D. Jaffray and J. Irish, "Geometric calibration of a mobile c-arm for intraoperative cone-beam ct", *Medical Physics*, vol. 35, no. 5, pp. 2124-2136, 2008.

[6] R. Horaud, R. Mohr and B. Lorecki, "On single-scanline camera calibration", *IEEE Transactions on Robotics and Automation*, vol. 1, no. 9, pp. 71-75, 1993.

Iodine-enhanced Liver Vessel Segmentation in Photon Counting Detector-based Computed Tomography using Deep Learning

Sumin Baek, Okkyun Lee*, and Dong Hye Ye*

***Abstract*—Liver vessel segmentation is important in diagnosing and treating liver diseases. Iodine-based contrast agents are typically used to improve liver vessel segmentation by enhancing vascular structure contrast. However, conventional computed tomography (CT) is still limited with low contrast due to energy-integrating detectors. Photon counting detector-based computed tomography (PCD-CT) shows the high vascular structure contrast in CT images using multi-energy information, thereby allowing accurate liver vessel segmentation. In this paper, we propose a deep learning-based liver vessel segmentation method which takes advantages of the multi-energy information from PCD-CT. We develop a 3D UNet to segment vascular structures within the liver from 4 multi-energy bin images which separates iodine contrast agents. The experimental results on simulated abdominal phantom dataset demonstrated that our proposed method for the PCD-CT outperformed the standard deep learning segmentation method with conventional CT in terms of dice overlap score and 3D vascular structure visualization.**

***Index Terms*—Photon Counting Detector, Computed Tomography, Liver Vessel Segmentation, Deep Learning**

I. Introduction

VASCULAR structures segmentation within the liver plays an important role in diagnosing and treating the patient's liver condition. For example, if there is a tumor in the patient's liver, determining the location of the tumor and the distribution of vascular structures in the liver before making liver resection will help planning for liver surgery [1], [2]. In particular, visualizing blood vessels in three dimensions before radio frequency ablation treatments or minimally invasive surgery can provide a road map for analyzing the location of blood vessels related to lesions [2]. In addition, in the case of liver transplantation, the vascular structure of the liver donor can be identified in advance, helping to accurately verify whether it is suitable for liver donors [3], [4].

As such, the segmentation of the vascular structure in the liver is critical for the diagnosis and treatment of liver diseases. Liver vessel segmentation is challenging due to the anatomical variability of the structure of the liver and blood vessels. To facilitate the liver vessel segmentation, iodine-based contrast agents are usually injected into the patient to highlight the vascular structures in the liver. However, the contrast of the iodine-enhanced vessels to the surrounding tissue is limited in the conventional computed tomography (CT) with energy-integrating detectors [5]. Even though there are promising CT organ segmentation results using deep learning methods such as UNET [6]–[8], it still remains challenging for deep learning methods to segment the liver vascular structures in the conventional CT images due to low contrast.

Meanwhile, photon counting detector-based computed tomography (PCD-CT) is an emerging medical imaging technology, which has potential advantages over conventional CT. Especially, PCD-CT has the ability to enhance the contrast to noise ratio with iodine contrast agents, generate multi-energy images simultaneously, reduce the noise and increase the spatial resolution [9]. The degree of attenuation for each material by X-ray depends on the energy and properties of the material [10]. In particular, when the iodine is injected, the attenuation in the vessels increase, resulting in an energy-dependent difference in attenuation between the iodine-enhanced vessels and surrounding tissue in the liver. Thus, we hypothesize that the energy discrimination capability of the PCD-CT has benefit for the blood vessel segmentation compared to the conventional CT.

In this paper, we propose a 3D UNET for liver vessel segmentation in PCD-CT, which trains bin-wise images at once in a deep convolutional neural network with a single labeled image. We compare our proposed method with the same UNET structure trained by a conventional CT image (except the multi-channel input for the proposed one). Datasets are generated with abdominal simulation phantom with iodine-enhanced for the liver blood vessels. We apply a low concentration level of iodine, reducing the risk of contrast-induced nephropathy into a liver and vessel, with a relatively low level of iodine diffused in the liver [11]. The experimental results using the PCD-CT dataset are compared and analyzed with those using the conventional CT dataset through dice score and 3D vascular structure visualizations.

The paper is organized by the following. In Section II, we describe the PCD-CT measurement models, and UNET architecture for liver vessel segmentation. Section III presents the experimental setup and the quantitative/qualitative assessment for liver vessel segmentation results with conventional CT and PCD-CT. The conclusion and discussion are provided in Section IV.

S. Baek and O. Lee are with the Department of Robotics Engineering, Daegu Gyeongbuk Institute of Science and Technology, Daegu, 42988 Korea.

D. Ye is with the Department of Electrical and Computer Engineering, Marquette University, Milwaukee, WI, 53233.

*Corresponding author (contact: oklee@dgist.ac.kr, donghye.ye@marquette.edu)

This work was supported by Institute for Information & Communications Technology Promotion(IITP) grant funded by the Korea government (MSIP) (No.2020-0-01524, Development of deep learning-based metal artifact reduction and organ segmentation techniques using spectral CT)

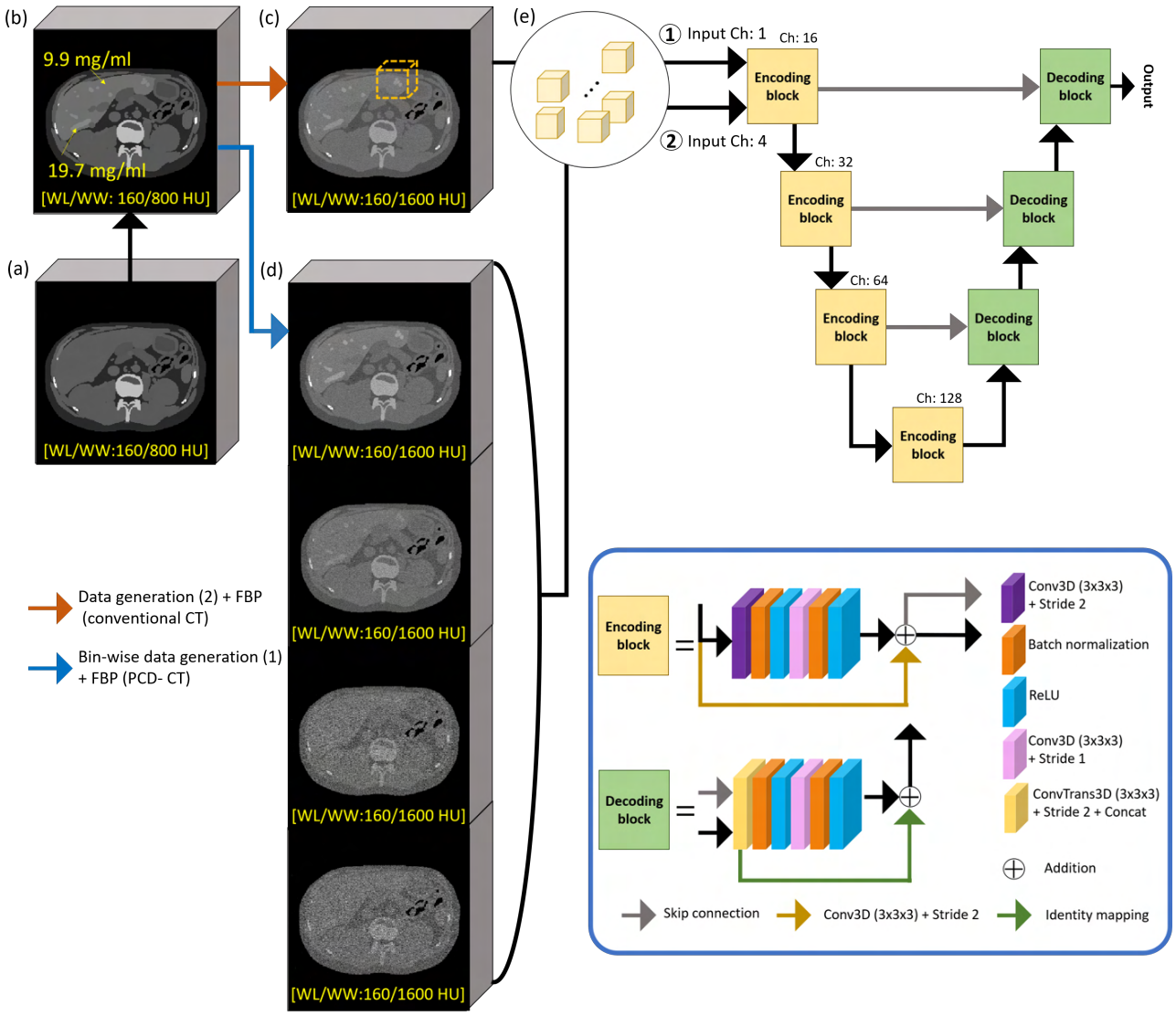


Fig. 1: The overview of liver vessel segmentation pipeline using the simulated PCD-CT and conventional CT dataset. (a) 3D volume of the ground truth before the addition of iodine, (b) after the addition of iodine, (c) 3D volume conventional CT data, (d) 3D volume of PCD-CT dataset, and (e) 3D patches randomly extracted from each data.

II. METHODS

Fig. 1 describes the overview of our liver vessel segmentation method using the PCD-CT. We first add the iodine on ground-truth 3D CT and generate 4-bin images based on the PCD-CT measurement model. We train the UNET from 3D random patches from 4-bin images and compare the liver vessel segmentation performance with the UNET trained on the corresponding conventional CT dataset.

A. PCD-CT measurement model

We assume that the bin-wise photon counts are independent Poisson random variables, and the expected counts $\bar{y}_b$ for the b-th energy bin of PCD-CT is given by

$$\bar{y}_b = \int_0^\infty I(E) S_b(E) e^{-\int \mu(r,E) dr} dE, \quad for\ b = 1, 2, \cdots, B, \tag{1}$$

where $I(E)$ represents the X-ray incident spectrum, and $S_b(E)$, which includes the spectral distortion in PCD, is a bin sensitivity function. $\int \mu(r, E) dr$ is the line integral of the linear attenuation coefficient.

An experiment was conducted using a conventional CT dataset as a comparison technique to the experiment using the PCD-CT dataset. We also assume that the measurement from the conventional CT, a single outcome from a detector, follows Poisson distribution, but with an expected counts $\bar{\lambda}$ as

follows:

$$\bar{\lambda} = \int_0^\infty I(E) \left\{ \sum_{b=1}^{B} S_b(E) \right\} e^{-\int \mu(r,E) dr} dE, \qquad (2)$$

For a given object, we use the measurements models of (1) and (2) to generate data sets for the PCD-CT and the conventional one, respectively. We applied filtered backprojection (FBP) with Shepp-Logan filter to the sinograms to obtain the data sets. Sample images of datasets for conventional CT and PCD-CT measurement models are shown in Fig. 1(c) and (d), respectively.

For the PCD-CT model, threshold values of energy bins set to the PCD are 20, 67, 101, 113 keV. An aluminium pre-filtered incident spectrum with 140kVp was used, and the total number of photons is 2.5x10^5. The spectral distortion is modeled in $S_b(E)$ [10], and water beam hardening correction is applied before we perform the FBP. The number of PCD channels and views are 601 and 1000, respectively, and the ASTRA toolbox is used for the projection and FBP with fan-beam geometry [12]. The distance from the source to the iso-center and the distance from the iso-center to the detector are 500 mm, respectively.

B. UNET architecture

We use UNET for both PCD-CT and conventional CT. It is worth noting that UNET for PCD-CT takes 4-channel input unlike the conventional CT one to take advantage of multi-energy information in PCD-CT. The UNET consists of an encoding process and a decoding process. The context of the image is obtained in the encoding process, which is combined with the feature map of the decoding process to enable more accurate localization [13]. As shown in Fig. 1(e), 3D patches are extracted from the 3D volume images of each subject and used for training.

The UNET structure is mainly composed of 3x3x3 convolution filter, and two convolutions are performed for each block. The first layer of the encoding block is convolution with stride 2, reducing the size of the image. The first layer of the decoding block is the transposed convolution with stride 2, and the feature map of the same image size obtained during the encoding process is concatenated by the skip connection process. Starting from the second layer, batch normalization, ReLU, convolution with stride 1, batch normalization, and ReLU are performed in the order. The residual unit is applied to both blocks, 3D convolution with stride 2 is performed in encoding block to reduce the feature size, and then added to the output feature of the encoding block, and the identity mapping is used in the decoding block [14]. The parameters are updated using the softmax function combined with dice loss. We obtain two channel outputs for both PCD-CT and conventional CT, the foreground and background. They are compared pixel-wise to make a binary image.

III. EXPERIMENTS

The experiment is conducted using a total of 56 abdominal simulation phantoms consisting of 33 adult men and 23 adult women [15]. A three-dimensional CT dataset consisting of about 100 slices including the liver region is used, and each slice is generated at 1.5 mm intervals.

As shown in Fig. 1(b), the concentration of iodine in the liver is set to 9.9 mg/ml and the concentration of iodine in the liver vessel is set to 19.7 mg/ml. The iodine concentration is chosen for low dose agents that can clinically reduce the risk of contrast-induced nephropathy, which is a challenging case to segment liver vascular structures [11].

For UNET training, we generate 150 patches having a size of 64x64x64 pixels for each subject. Thus, the PCD-CT dataset has four input channels in Fig. 1(d), and the conventional CT has one input channel in Fig. 1(c). Our experiment runs on GeForce RTX 3090. We implemented UNET using a MONAI library, a Pytorch-based open-source framework [16], which consists of 16, 32, 64, and 128 channels, and uses 2 strides and 3x3x3 kernel sizes, and the learning rate was 10^{-3}. The batch size was 1 and the epoch size was 1500. Training dataset, test dataset, and validation dataset were divided into 6:2:2 ratios, and performance was evaluated through 5 fold cross-validation.

A. Quantitative assessment

To evaluate the quantitative performance of liver vessel segmentation in PCD-CT and conventional CT, we measure the dice similarity coefficient (DSC). DSC is defined by the following equation:

$$DSC = \frac{2|A \cap B|}{|A| + |B|}, \qquad 0 \leq DSC \leq 1, \qquad (3)$$

where A is the segmented (binary) image for each method and B is also a binary image of the ground truth of the target. $|\cdot|$ indicates the number of elements in the set. The more similar the segmented results to ground truth, the higher the DSC value, and the more accurate segmentation is obtained.

The average and standard deviation of DSC scores from 5-fold cross validation on conventional CT and PCD-CT datasets (56 subjects) are reported in Table I. Overall, DSC scores are under 0.9 because the low number of photons and low concentration iodine are used toward the realistic simulation. On average, UNET with PCD-CT significantly improves the DSC by 0.011 (2.6 times of 0.043 standard deviation) compared with UNET of conventional CT. For all folds, PCD-CT shows the consistent improvement of liver vessel segmentation performance over conventional CT. This indicates that PCD-CT enhances the contrast of iodine-injected vascular structures

	Conventional CT	PCD-CT
Fold 1	0.861 ± 0.024	0.869 ± 0.025
Fold 2	0.848 ± 0.053	0.860 ± 0.063
Fold 3	0.866 ± 0.024	0.874 ± 0.022
Fold 4	0.839 ± 0.044	0.853 ± 0.041
Fold 5	0.844 ± 0.058	0.858 ± 0.052
Average	0.852 ± 0.042	0.863 ± 0.043

TABLE I: The table of dice similarity coefficients for each fold and average of 56 subjects (mean ± standard deviation).

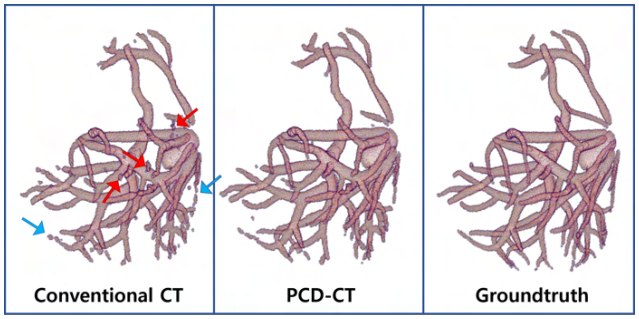


Fig. 2: Liver vessel segmentation of the female subject. (red arrow: mis-segmented area, blue arrow: disconnected area)

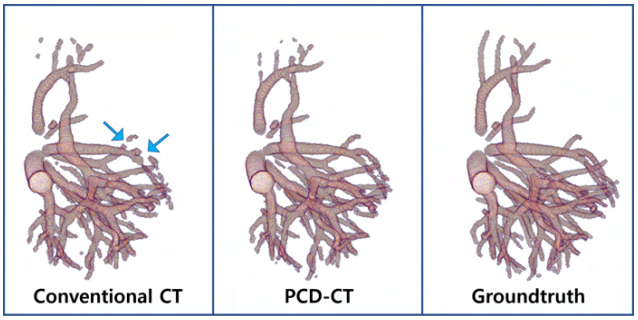


Fig. 3: Liver vessel segmentation of the male subject. (blue arrow: disconnected area)

in the liver, benefiting liver vessel segmentation using deep learning.

B. Qualitative assessment

For qualitative assessment, we visualize the three-dimensional vascular structure from the UNET-based liver vessel segmentation from both conventional CT and PCD-CT in Fig. 2 and 3. We also show the ground truth vessel structure from the corresponding phantom as reference. We note that each person has a very different appearance and size of vessels, posing the challenge in deep learning segmentation. Compared to the conventional CT, UNET on PCD-CT produces the more accurate segmentation similar to ground truth. In the segmentation results from conventional CT, it is observed that there are disconnected vessels (marked with blue arrow) and false vessel structures (marked with red arrow). This highlights the benefit of using PCD-CT for diagnosing and treating liver disease over conventional CT.

IV. CONCLUSION

In this paper, we generated PCD-CT datasets by utilizing the characteristics of energy-discrimination capabilities of PCD-CT, and segmented liver vascular structures by learning them using the deep neural network. The experimental performance was compared and analyzed with those of the conventional CT dataset through dice score and 3D vascular structure visualization. The dice score of the PCD-CT dataset were significantly higher (2.6 times of standard deviation) than that of the conventional CT dataset. For most testing subjects, the incorrect segmentation of the peripheral part of the vessel was reduced using UNET on PCD-CT. This showed that UNET segmentation using PCD-CT outperformed the one using conventional CT, improving diagnosis and management of liver diseases.

REFERENCES

[1] F. Conversano, R. Franchini, C. Demitri, L. Massoptier, F. Montagna, A. Maffezzoli, A. Malvasi, and S. Casciaro, "Hepatic vessel segmentation for 3d planning of liver surgery: Experimental evaluation of a new fully automatic algorithm," *Academic Radiology*, vol. 18, no. 4, pp. 461–470, 2011.

[2] E. Fishman, B. Kuszyk, D. Heath, L. Gao, and B. Cabral, "Surgical planning for liver resection," *Computer*, vol. 29, no. 1, pp. 64–72, 1996.

[3] G. Low, E. Wiebe, A. Walji, and D. Bigam, "Imaging evaluation of potential donors in living-donor liver transplantation," *Clinical radiology*, vol. 63, no. 2, pp. 136–145, 2008.

[4] M. Bassignani, A. Fulcher, W. C. R.A. Szucs, U. Prasad, and A. Marcos, "Use of imaging for living donor liver transplantation," *Radiographics*, vol. 21, no. 1, pp. 39–52, 2001.

[5] J. Kaftan, H. Tek, and T. Aach, "A two-stage approach for fully automatic segmentation of venous vascular structures in liver CT images," *Medical imaging 2009: image processing*, vol. 7259, p. 725911, 2009.

[6] W. Yu, B. Fang, Y. Liu, M. Gao, S. Zheng, and Y. Wang, "Liver vessels segmentation based on 3d residual U-NET," *2019 IEEE International Conference on Image Processing*, pp. 250–254, 2019.

[7] Q. Huang, J. Sun, H. Ding, X. Wang, and G. Wang, "Robust liver vessel extraction using 3D U-Net with variant dice loss function," *Computers in biology and medicine*, vol. 101, pp. 153–162, 2018.

[8] T. Kitrungrotsakul, X. Han, Y. Iwamoto, L. Lin, A. H. Foruzan, W. Xiong, and Y. Chen, "Vesselnet: A deep convolutional neural network with multi pathways for robust hepatic vessel segmentation," *Computerized Medical Imaging and Graphics*, vol. 75, pp. 74–83, 2019.

[9] S. Leng, M. Bruesewitz, S. Tao, K. Rajendran, A. F. Halaweish, N. G. Campeau, J. G. Fletcher, and C. H. McCollough, "Photon-counting detector CT: system design and clinical applications of an emerging technology," *Radiographics*, vol. 39, no. 3, pp. 729–743, 2019.

[10] J. P. Schlomka, E. Roessl, R. Dorscheid, S. Dill, G. Martens, T. Istel, C. Bäumer, C. Herrmann, R. Steadman, G. Zeitler, A. Livne, and R. Proksa, "Experimental feasibility of multi-energy photon-counting K-edge imaging in pre-clinical computed tomography," *Phys. Med. Biol.*, vol. 53, pp. 4031–4047, 2008.

[11] M. Radon, M. Kaduthodil, J. Jagdish, S. Matthews, C. Hill, M. Bull, and S. Morcos, "Potentials and limitations of low-concentration contrast medium (150 mg iodine/ml) in CT pulmonary angiography," *Clinical radiology*, vol. 66, no. 1, pp. 43–49, 2011.

[12] W. van Aarle, W. J. Palenstijn, J. D. Beenhouwer, T. Altantzis, S. Bals, K. J. Batenburg, and J. Sijbers, "The ASTRA Toolbox: A platform for advanced algorithm development in electron tomography," vol. 157, pp. 35–47, 2015.

[13] O. Ronneberger, P. Fischer, and T. Brox, "U-net: Convolutional networks for biomedical image segmentation," pp. 234–241, 2015.

[14] E. Kerfoot, J. Clough, I. Oksuz, J. Lee, A. King, and J. Schnabel, "Left-ventricle quantification using residual u-net," pp. 371–380, 2018.

[15] W. P. Segars, M. Mahesh, T. J. Beck, E. C. Frey, and B. M. W. Tsui, "Realistic CT simulation using the 4D XCAT phantom," *Med. Phys.*, vol. 35, no. 8, pp. 3800–3808, Aug. 2008.

[16] T. M. Consortium, "Project monai," 2020.

Deep Learning-based Prior toward Normalized Metal Artifact Reduction in Computed Tomography

Jeonghyeon Nam, Dong Hye Ye*, and Okkyun Lee*

***Abstract*—X-ray computed tomography (CT) often suffers from scatter and beam-hardening artifacts in the presence of metal. These metal artifacts are problematic as severe distortions in the CT images deteriorate the diagnostic quality in clinical applications such as orthopedic arthroplasty. The normalized metal artifact reduction (NMAR) method effectively reduces the artifacts by normalizing the sinogram with the metal traces through the forward projection of the prior image. Because the prior image is the thresholded CT image with the values of the air and soft tissues replaced, the image is noticeably different from the ideal CT thereby making normalized sinogram not completely flat. In this paper, we propose the novel NMAR method with the deep learning-enhanced prior image which is denoised by learning the relationship between NMAR and clean image without metal artifact. The denoised prior image is then forward projected to correct the sinogram with the metal trace. The experimental results on simulated rat phantom dataset demonstrate that our proposed deep prior NMAR achieves higher structural similarity index (SSIM) and peak signal-to-noise ratio (PSNR) than the original NMAR.**

***Index Terms*—Metal Artifact Reduction, Computed Tomography, Deep Learning Prior**

I. Introduction

COMPUTED tomography (CT) suffers from the artifact when imaging objects with metal objects are present. Due to the metal's high x-ray absorption rate, the detectors receive low signals when the x-ray goes through the metal objects. This results in scattering and beam-hardening effects in the CT image. Metal artifacts make it difficult to observe small details around the metal objects and affect throughout the image by severe streaks and shading. Therefore, it is necessary to reduce the metal artifacts in CT images for reliable diagnosis in clinical applications such as orthopedic arthroplasty.

Various metal artifact reduction (MAR) methods have been developed to deal with the metal artifacts. For example, filtering [1], [2] or interpolation [3], [4] methods directly corrects the sinogram in the metal trace. These sinogram-based methods show the limited performance as the image information is not taken account and back projection of the corrected sinogram often causes the secondary artifacts in non-metal areas. Iterative methods [5], [6] reconstruct the MAR image through the Bayesian optimization with both forward (sinogram) and prior (image) model, but they are computationally expensive due to the iterative optimization. Recently, deep learning-based approaches [7], [8] have been also applied to reduce metal artifacts and demonstrated its effectiveness. However, they are processing only in the image domain without taking account into the sinogram information, thereby showing blurring in the MAR image.

The normalized metal artifact reduction (NMAR) is one of the efficient and effective MAR methods which utilizes both sinogram and image information. The NMAR normalizes the sinogram with the metal traces before the interpolation to remove metal artifacts effectively while minimizing other artifacts caused by the interpolation. The sinogram is normalized using the prior image, which is the thresholded image of the CT image with metal artifacts. Since the normalization is the most important process, the NMAR method is highly dependent on the quality of the prior image. The NMAR method may have residual artifacts in the results due to the inaccurate prior image, depending on the metal size and locations.

In this paper, we combine the NMAR with a deep learning to further reduce the metal artifacts while preserving the details in the tissue. We use the NMAR results as the input to the deep denoising network and use the output as the prior image for the second trial of the NMAR. The deep learning-based denoising effectively reduces the residual streak and shading artifacts after the NMAR. However, as other deep learning-based MAR methods, it may cause a blurring in the outcome and lose the details in the tissue and not be sufficient for the clinical purposes. Instead, we propose to use it as the improved prior image and perform the NMAR again to further reduce the residual artifacts. We validate the proposed method with the simulation dataset of rat phantom. We demonstrate that the proposed method shows less artifacts in terms of improved peak signal-to-noise ratio (PSNR) and structural similarity index (SSIM) compared to the original NMAR.

II. Methods

Fig. 1 illustrates the overall procedures of our proposed deep prior NMAR. We first apply the conventional NMAR by generating the thresholded prior image. We then denoise the NMAR image with the trained deep neural network and feed the denoised image as the prior image for further metal artifact reduction. In following, we describe the existing NMAR and proposed deep prior NMAR.

J. Nam and O. Lee are with the Department of Robotics Engineering, Daegu Gyeongbuk Institute of Science and Technology, Daegu, 42988 Korea.

D. Ye is with the Department of Electrical and Computer Engineering, Marquette University, Milwaukee, WI, 53233.

*Corresponding authors (contact: donghye.ye@marquette.edu, oklee@dgist.ac.kr)

This work was supported by Institute for Information & Communications Technology Promotion(IITP) grant funded by the Korea government (MSIP) (No.2020-0-01524, Development of deep learning-based metal artifact reduction and organ segmentation techniques using spectral CT)

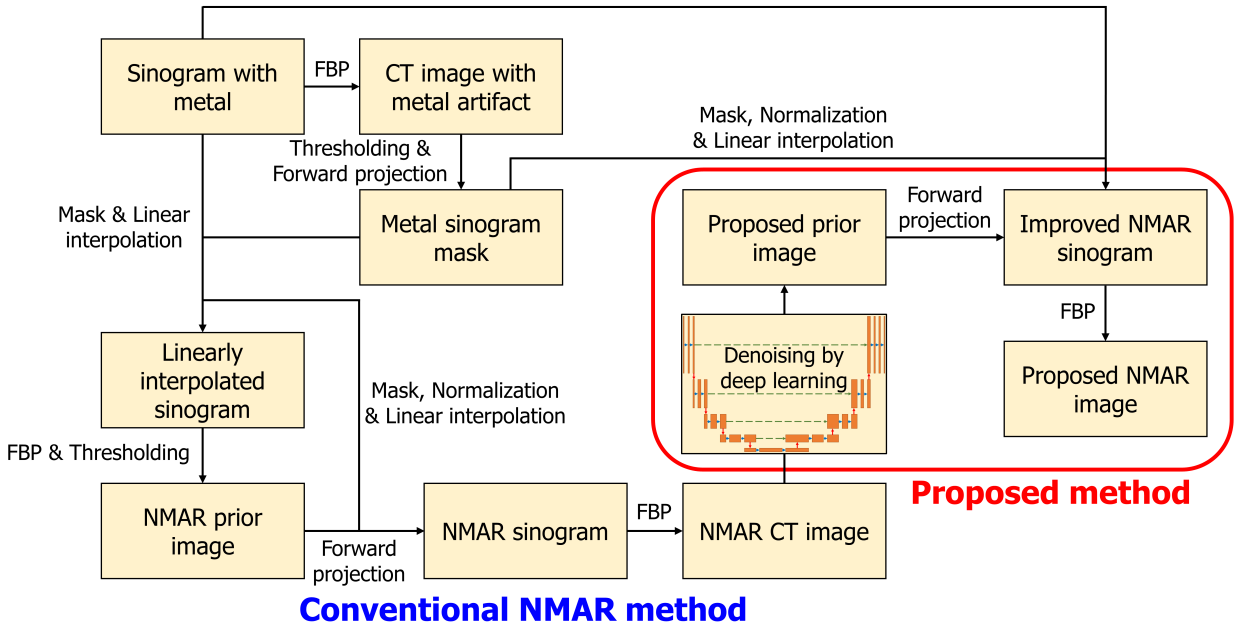


Fig. 1. The overview of the proposed method. The sinogram with metal trace is processed with the original NMAR method to obtain the artifacts reduced image. Then, the NMAR result is processed with the deep learning network to obtain the prior image for the proposed method. With the denoised prior image, the NMAR method is performed again.

A. Normalized Metal Artifact Reduction (NMAR) [4]

As described in Fig. 1 (blue arrow), the NMAR creates the metal trace mask for sinogram by thresholding the CT image to find the location of the metal objects, then forward-projecting the metal location. For the forward projection, we use the counting model. The expected number of x-ray photons is given by,

$$\bar{\lambda} = \int_0^\infty S(E) \exp\left(-\int \mu_a(\vec{r}, E) d\vec{r}\right) dE, \qquad (1)$$

where $S(E)$ is the system model, including the x-ray incident spectrum and the detector response [9]. We omit the specific x-ray path in the model for the sake of simplicity. X-ray with a 90 kVp incident spectrum is used with 2mm Aluminum pre-filtration. Total of 130,000 incident photons are used, and photons with less than 20 keV were ignored. We use 511 detectors (0.25mm /channels) with 768 views per rotation. We apply the fanbeam geometry (equi-distance) and used the ASTRA toolbox [10], [11].

The obtained metal trace is then linearly interpolated to get the prior sinogram. Filtered back projection (FBP) of the prior sinogram gives the CT image with less metal artifacts. The CT image is thresholded to segment it into three materials, air, soft tissue, and bone, and then replace the pixel values of air to -1000 [HU] and soft tissue to 0 [HU] to make the prior image. The original sinogram with the metal traces are divided pixel-by-pixel by the forward projection of the prior image. The normalized sinogram has similar values except the metal traces which gives an advantage to the linear interpolation. The normalized sinogram is masked and interpolated in the same way as the prior sinogram and then, de-normalized to get the final outcome followed by FBP.

B. Proposed Deep Prior NMAR

The proposed method is an extension to the original NMAR method. The NMAR method effectively removes the metal artifacts, but there are severe streak and shade artifacts remained depending on the size and location of the metals. The remaining artifacts are caused by the fluctuation in the normalized sinogram due to the difference between the original sinogram and the sinogram of the prior image. To utilize the benefit of the normalization fully, better prior image is required. Thus, we propose the deep prior NMAR method which we replace the prior images with the denoised NMAR results by the deep neural network. As shown in Fig. 1, we train the deep learning network to obtain the denoised image, focusing on removing shading and streak artifacts after the NMAR. By using the denoised image as the prior image, we can acquire the image with the metal artifacts reduced better than the original NMAR method with details remained. Note that we have not applied any thresholding to the prior image (denoised one) before the projection.

III. EXPERIMENTS

The methods are validated on the simulated rat phantom. For each slice of the phantom, 25 images were generated with two stainless steel objects with the diameter of 3mm inserted in the random places; there were 100 slices, the width of each slice is 1 mm, in the chest and abdomen area, making 2,500 images for the dataset. The images were 512×512 pixels (0.125 mm/pixel).

We use U-Net structure to denoise the NMAR images [12]. The output activation function is removed for denoising purpose. The inputs are the NMAR images and the labels are the CT images without the metal artifacts. The network is trained for 300 epochs with L1 loss. From the 2,500 images dataset, 2,000 images were used for the training and the other 500 images were used as the test set. In order to make the test set include all parts of the body, images from every 5th slices from the 100 slices were set as the test set. The part of the training set, 400 images, are randomly chosen for validation dataset to find the best performing epoch. The network is trained with the GeForce RTX 3090 GPU.

A. *Qualitative Evaluation*

The normalized sinograms are displayed in Figure 2. The normalized sinogram of the proposed deep prior NMAR method is flatter than that of the original NMAR method. Since the quality of the metal artifact reduction is improved with the flatness of the normalized sinogram, the proposed method can reduce the metal artifacts better than the original method.

Figure 3 illustrates the experimental results of the original NMAR and the proposed deep prior NMAR method, respectively. The original CT images without and with metal objects are displayed as reference in Fig.3 (a, b). The difference images are obtained by subtracting the generated MAR image with the ground-truth CT image without metal objects. In Fig. 3(e, f), the original NMAR result still shows significant amount of artifacts, reflecting less flattened normalized sinogram. The prior image of the NMAR method (Fig. 3(c)) is computed by the thresholding method, therefore, it affects the normalization process in the NMAR method.

The deep learning denoised image (Fig. 3(g)) presents reduced metal artifacts, but the quality of the image is blurred from the deep learning process. The proposed deep prior NMAR method has noticeably reduced metal artifacts with image quality similar to the ground truth CT image without metal insertion.

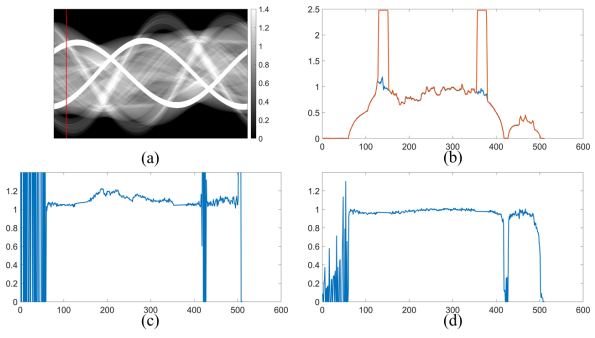


Fig. 2. (a) The original sinogram with metal traces, (b) the profiles with and without metal traces, (c) the profile of the normalized and interpolated sinogram of the original NMAR method, (d) the profile of the normalized and interpolated sinogram of the proposed method. (All the profiles correspond to the specific angle view indicated by the red line in (a).)

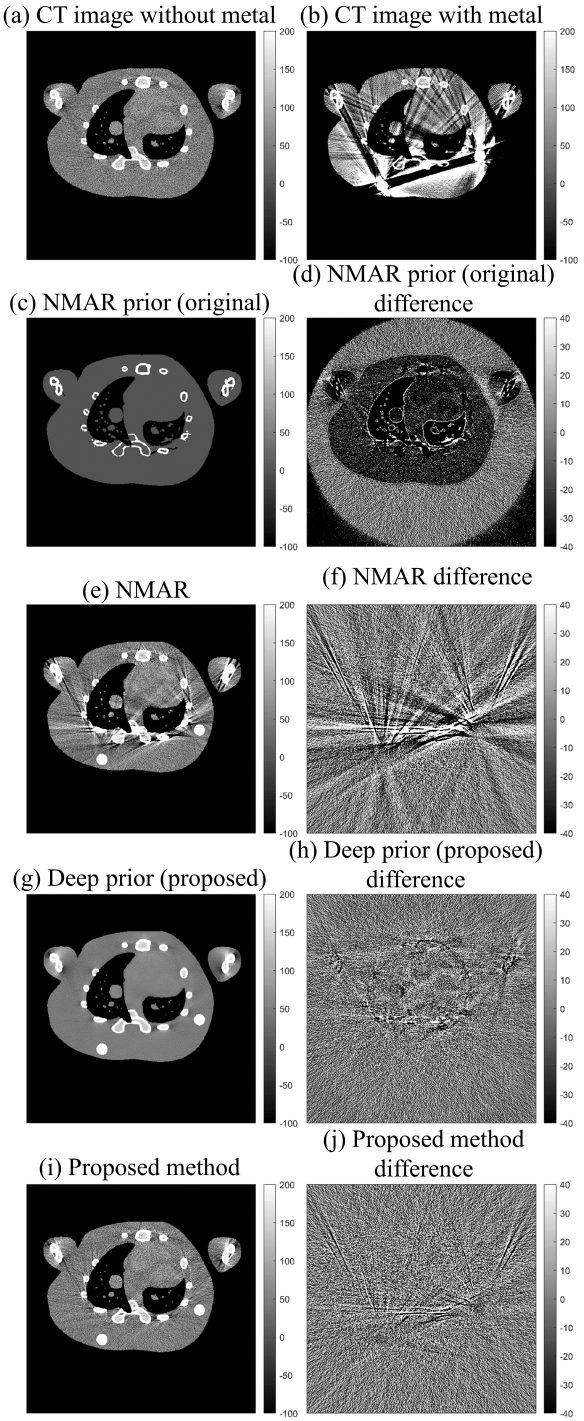


Fig. 3. The results and the difference images using the original NMAR and the propose method with the associated prior images. The difference images are compared with the CT image without the metal objects. (a) The CT image without the metal objects, (b) the CT images with the metal objects, (c, d) the prior image for the NMAR method and the difference image, (e, f) the CT image from the NMAR method and the difference image, (g, h) the prior image for the proposed method and the difference image, (i, j) the CT image from the proposed method and the difference image.

TABLE I
MEAN AND STANDARD DEVIATION OF SSIM AND PSNR OF THE NMAR AND THE PROPOSED METHODS.

	NMAR	Proposed method
SSIM	0.8832 ± 0.0266	0.9038 ± 0.0137
PSNR	36.28 ± 2.901	39.12 ± 1.351

B. Quantitative Evaluation

The tables I report the SSIM and PSNR values of the conventional NMAR and the proposed deep prior NMAR. The proposed deep prior NMAR increases SSIM by 0.0206 and PSNR by 2.84dB compared with the conventional NMAR, while decreasing the standard deviation. This indicates that our proposed methods effectively reduces the metal artifact while preserving the fine details in the CT image by taking advantage of the denoised prior image through deep learning, showing potential for clinical applications such as orthopedic arthroplasty.

IV. CONCLUSION

In this paper, we proposed the deep prior NMAR method to further reduce the metal artifacts from the conventional NMAR result. The proposed method denoises the NMAR image with the deep learning network and uses it as a prior image for flattened normalized sinogram. The experimental results on simulated rat phantom data showed that our deep prior NMAR method improved the MAR performance compared with the conventional NMAR method while maintaining the fine details of the original CT image. In addition, we validate the stability and robustness of our proposed method which will be useful for clinical applications like orthopedic arthroplasty.

REFERENCES

[1] J. Hsieh, "Adaptive streak artifact reduction in computed tomography resulting from excessive x-ray photon noise," *Medical Physics*, vol. 25, no. 11, pp. 2139–2147, Nov. 1998. [Online]. Available: http://doi.wiley.com/10.1118/1.598410

[2] R. M. Rangayyan and R. Gordon, "Streak Preventive Image Reconstruction with ART and Adaptive Filtering," *IEEE Transactions on Medical Imaging*, vol. 1, no. 3, pp. 173–178, Nov. 1982. [Online]. Available: http://ieeexplore.ieee.org/document/4307569/

[3] W. A. Kalender, R. Hebel, and J. Ebersberger, "Reduction of CT artifacts caused by metallic implants." *Radiology*, vol. 164, no. 2, pp. 576–577, Aug. 1987. [Online]. Available: http://pubs.rsna.org/doi/10.1148/radiology.164.2.3602406

[4] E. Meyer, R. Raupach, M. Lell, B. Schmidt, and M. Kachelrieß, "Normalized metal artifact reduction (NMAR) in computed tomography," *Medical Physics*, vol. 37, no. 10, pp. 5482–5493, Sep. 2010. [Online]. Available: http://doi.wiley.com/10.1118/1.3484090

[5] Ge Wang, D. Snyder, J. O'Sullivan, and M. Vannier, "Iterative deblurring for CT metal artifact reduction," *IEEE Transactions on Medical Imaging*, vol. 15, no. 5, pp. 657–664, Oct. 1996. [Online]. Available: http://ieeexplore.ieee.org/document/538943/

[6] M. Oehler and T. M. Buzug, "Statistical Image Reconstruction for Inconsistent CT Projection Data," *Methods of Information in Medicine*, vol. 46, no. 03, pp. 261–269, 2007. [Online]. Available: http://www.thieme-connect.de/DOI/DOI?10.1160/ME9041

[7] W.-A. Lin, H. Liao, C. Peng, X. Sun, J. Zhang, J. Luo, R. Chellappa, and S. K. Zhou, "Dudonet: Dual domain network for ct metal artifact reduction," 2019.

[8] H. Wang, Y. Li, H. Zhang, J. Chen, K. Ma, D. Meng, and Y. Zheng, "Indudonet: An interpretable dual domain network for ct metal artifact reduction," 2021.

[9] J. P. Schlomka, E. Roessl, R. Dorscheid, S. Dill, G. Martens, T. Istel, C. Bäumer, C. Herrmann, R. Steadman, G. Zeitler, A. Livne, and R. Proksa, "Experimental feasibility of multi-energy photon-counting K-edge imaging in pre-clinical computed tomography," *Physics in Medicine and Biology*, vol. 53, no. 15, pp. 4031–4047, Aug. 2008. [Online]. Available: https://iopscience.iop.org/article/10.1088/0031-9155/53/15/002

[10] W. van Aarle, W. J. Palenstijn, J. Cant, E. Janssens, F. Bleichrodt, A. Dabravolski, J. D. Beenhouwer, K. J. Batenburg, and J. Sijbers, "Fast and flexible x-ray tomography using the ASTRA toolbox," *Opt Express*, vol. 24, no. 22, pp. 25 129–25 147, October 2016.

[11] W. van Aarle, W. J. Palenstijn, J. D. Beenhouwer, T. Altantzis, S. Bals, K. J. Batenburg, and J. Sijbers, "The ASTRA Toolbox: A platform for advanced algorithm development in electron tomography," *Ultramicroscopy*, vol. 157, pp. 35–47, October 2015.

[12] O. Ronneberger, P. Fischer, and T. Brox, "U-net: Convolutional networks for biomedical image segmentation," 2015.

Optimization of Empirical Beamhardening Correction Algorithm

Andriy Andreyev, Faguo Yang, Lars Omlor, and Matthew Andrew

***Abstract*—Metal artifacts are one of the most common reasons for reduced image quality and usability in polychromatic cone-beam CT. In this work, we revisit empirical beam hardening correction algorithm and propose a few practical optimizations to simplify its application. First, fuzzy C-means segmentation method is used to perform an automatic segmentation of the metal component. Second, a minimum variance optimization technique provides a suitable combination of correction basis images. Finally, a sub-volume (spatially varying) optimization method is used to account for a varying contribution of metal artifacts through the image. We apply the modified algorithm to datasets from cone-beam CT and evaluate its performance.**

***Index Terms*— Computed Tomography, Image Reconstruction, Beam Hardening, Metal Artifacts, Fuzzy C-Means.**

I. Introduction

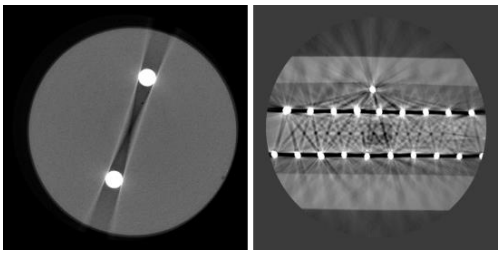
METAL artifacts are one of the primary causes for image quality degradation in Computed Tomography. The streaks and shadows caused by the artifacts obscure the useful information content, prevent robust segmentation and feature detection in medical as well as industrial CT imaging. Numerous algorithmic methods have been proposed over the years to resolve the problem of metal artifact reduction (MAR). These include but not limited to segmentation-based [1, 3, 5], inpainting or projection completion based [6], iterative model-based [3],[7], machine-learning based [8], and so on. More comprehensive overview of various metal artifact correction can be found in [9].

In this work we revisit an empirical beam hardening correction method (EBHC) [1] and propose a few practical modifications and enhancements, in order to reduce its reliance on manual user intervention. We apply the modified method to a number of cases from CBCT and discuss the results.

II. Methods

The polychromatic nature of most X-ray sources used in CT leads to artifacts in the reconstructed images. These are most evident whenever studied objects and samples demonstrate high variability in atomic number (Z). The artifacts take the form of dark streaks and halos as shown in a) b)

Figure 1. While it is possible to minimize the severity of metal artifacts by modifying the acquisition conditions, the most common artifact correction solutions are algorithmic. Here, we try to build upon and add a few modifications to an empirical beam hardening correction method which has proven its effectiveness despite its relative simplicity [1]. The primary goal of the proposed modifications is to make the EBHC application more autonomous, without requiring user intervention.



a) b)

Figure 1. Examples of samples in CBCT that lead to creation of metal artifacts: a) cylindrical phantom with plastic filling and two 4mm diameter brass rod inserts. b) HDMI connector.

EBHC consists of the following steps:

1. The acquired dataset g is being reconstructed to create 3D tomographic image as $f = F^{-1}g$, where F^{-1} is the tomographic reconstruction algorithm of choice; no metal artifact related correction is necessary at this stage, however, in many cases simple polynomial based beam hardening correction may reduce cupping artifacts and enhance the image structures which will help with the next step.

2. The reconstructed CT image f is segmented to extract only the metal component m.

3. Results of the segmentation is then forward projected in the matching geometry to the original acquisition. Forward projection can be done in monochromatic fashion, however it is important that it follows the geometry as close as possible.

$$h^j = Fm$$

4. The forward projected data h is then combined with the original data g to create a set of basis data p_{ij}:

$$p_{ij} = g^i h^j,$$

where i and j can be lower integer values ranging from 0 to 5 as an example. Not all possible combinations of i and j need to be considered due to increased computational burden and

The authors are with Carl Zeiss X-Ray Microscopy, Dublin, CA 94588 USA (e-mail: andriy.andreyev@zeiss.com).

somewhat overlapping nature of these correction terms.

5. Each data combination p_{ij} is then reconstructed using reconstruction operation F^{-1} to create corresponding set of images f_{ij}. We further denote uncorrected reconstructed image as f_{10}.

6. The higher order basis images that include forward projected metal component are effectively able to replicate the metal artifact streaks. As such, the generated set of basis images can be combined using a certain optimal set of weighting coefficients to create a final corrected image that can effectively mitigate the effect of metal artifacts. In this work, original uncorrected image always uses the weight of 1.0, and the rest of the images can have variable weighting coefficients (negative values are also possible). This workflow is shown in figure 2.

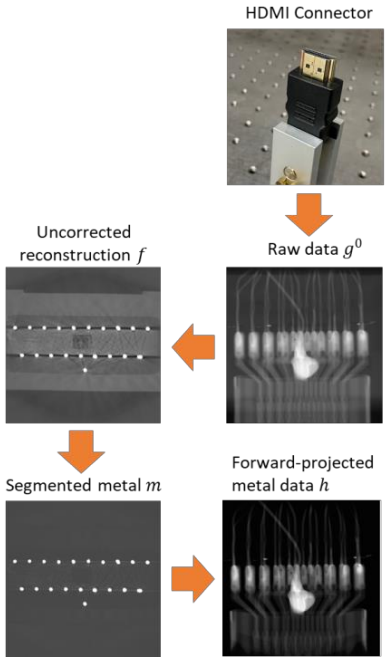


Figure 2. Diagram of the EBHC method, describing the first part of the method, leading to the forward projected data generation.

In original EBHC method soft thresholding was used to segment out the high-density components. In this work we found that fuzzy C-means segmentation can be very effective method as implemented in [10]. We used two material segmentation with starting values for the low-Z material segmentation set at zero, and high-Z material set at 80% of the maximum value of f. To improve robustness of the segmentation it is also advisable to prefilter the image before the segmentation with edge preserving filter, such as median filter.

The basis functions are combined according to the following equation:

$$\hat{f} = \begin{cases} f_{10} & m > 1 \\ f_{10} + \sum_{j,j} w_{ij} f_{ij} & m = 0 \end{cases}$$

In this way no correction has been applied inside the metal parts, as we seek to mitigate the streak artifacts in between metal/higher-Z components.

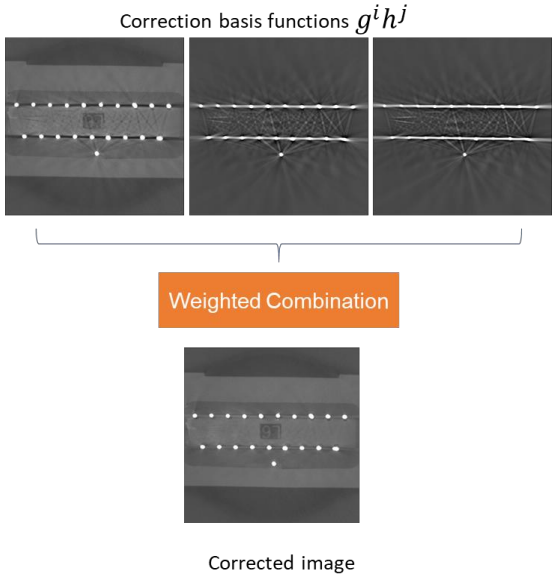


Figure 3. Second part of the EBHC method: optimal weighted sum of basis images leads to a reduction in metal artifact expression.

Optimal weight determination is then the significant outstanding challenge. One obvious way is to do this manually, ad hoc going through all possible combinations of w_{ij} that can provide artifact streak reduction while maintaining the image quality. The success of such method may vary and depends on the operator training. Additionally, any manual, operator-controlled optimization method is tedious to perform with more than two or three basis images, and any optimization is global, with a single set of weights determined for the entire image (in contrast to an automated technique which may be allowed to vary spatially).

In some cases, to simplify the optimization process, it may also be feasible to define the region in the reconstructed image that is known to be flat but has been imbued with spurious signal from the artifact streaks. Some examples of such flat region include uniform plastic enclosure surrounding the metal wires, or soft tissue surrounding metal implants or bone matter for biological samples. The assumption is that proper combination of basis images will minimize the streaks and will make such region more uniform. In this work, we propose to use minimum variance-based optimizer, estimated over the entire volume excluding metal/high-Z components. This also means that entire EBHC workflow can now be performed fully automatically (referred to in the text as automatic EBHC (AEBHC)). The further advantage of such automatic optimization method is that it can take arbitrary number of basis functions as an input without any complication for the user (other than prolonged reconstruction time).

As the objects can be highly non-uniform, with different amount of beamhardening, scatter, and metal artifact presence in each slice, a globally optimized single set of combination weights may result in under- or overcorrection of certain sub-volumes (slices). Likewise, it is not practically feasible to have those hand-tuned for each sub-volume, as it would be a very tedious task. We propose that optimization weights are recalculated for every sub-volume in AEBHC. Here we recalculate the weight every 64 slices, while averaging the individually optimized sub-volumes into the final volume using 50% overlap.

Overall, all the discussed modifications allow for a high degree of EBHC autonomy, allowing single-click metal artifact reductions.

III. Results and Discussions

To provide some examples of the reconstruction, use tomographic 3D X-ray microscopy data from Zeiss Xradia Versa (Carl Zeiss X-Ray Microscopy, Dublin, CA).

The first example is the cylindrical phantom from Figure 1. Correction with AEBHC was able to drastically reduce the severity of the metal artifacts and the visual conspicuity of a small void indicated was much improved as shown in Figure 4.

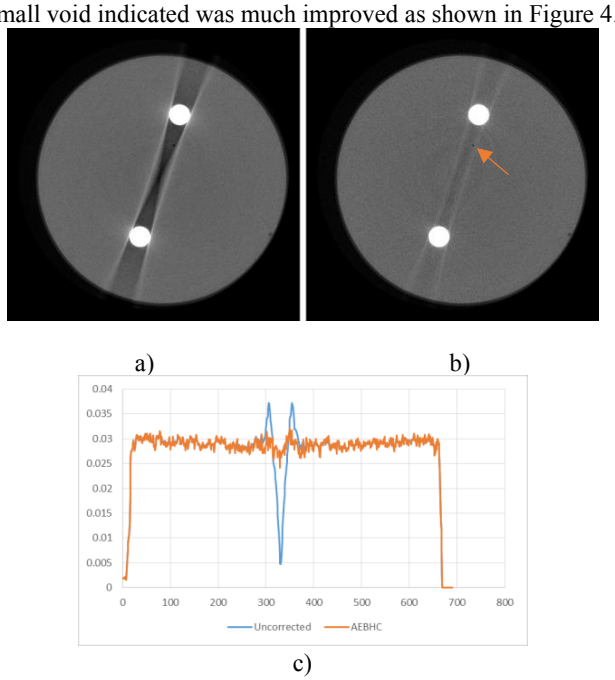


Figure 4. Cylindrical phantom with plastic filling and two 4mm diameter brass rod inserts: a) uncorrected FDK reconstruction; b) corrected with AEBHC; c) reconstructed intensity profiles, drawn across the phantom from approximately 10 to 4 o'clock.

Three scans of a standard HDMI connector have been performed to test the performance of auto-EBHC method at different acquisition conditions. Acquisition conditions included 160 and 100 kVp, as well as stronger (equivalent to approximated 2 mm of Cu) and medium filter. In Figure 5, we show both uncorrected and AEBHC corrected images. From the images it is clear that higher kVp and stronger filtered X-ray spectrum (removing lower energy part of the spectrum) helps to reduce the severity of metal artifacts even in the uncorrected data. The artifacts are more severe at lower kVp, showing that acquisition parameters play a strong role in AEBHC effectiveness. AEBHC performs better at higher X-ray energy setting, and stronger filter (Figure 5), however, corrected reconstructions outperform uncorrected reconstructions at all conditions.

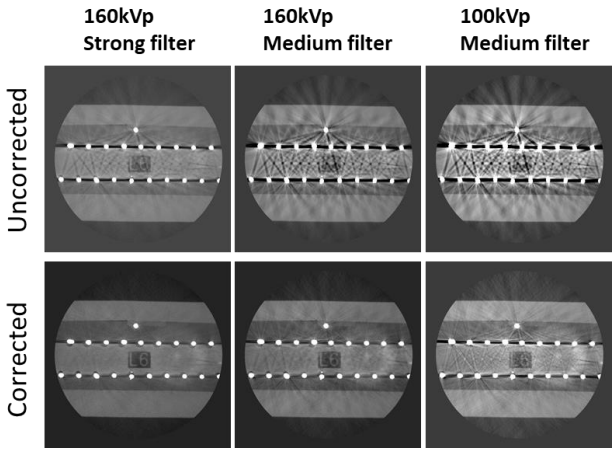


Figure 5. Reconstructions of HDMI cable connector, performed at different 160 and 100 kVp, as well as with stronger (~2mm Cu thickness) and medium (1mm Cu thickness) filtering material. Uncorrected and corrected reconstructions are in top and bottom row respectively.

The advantage of sub-volume optimized AEBHC method is demonstrated in Figure 6. Here, we use simple phantom consisting of steel rods, plastic tubes inserted into the piece of plastic foam. In Figure 6 we show two reconstructed slices through the phantom, with both AEBHC and manually optimized EBHC method. Manual optimization of weights for EBHC was done over the slice shown in the top, and then applied globally to the entire volume. That leads to overcorrection for metal artifacts in the slice 200 shown in the bottom row. AEBHC was able to perform more consistently avoiding overcorrection across the entire volume.

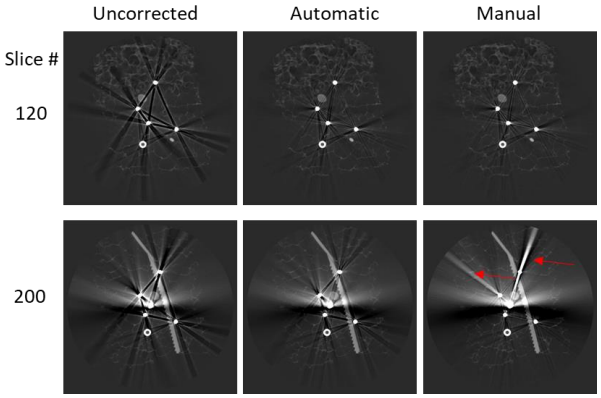


Figure 6. Reconstructions of simple steel rods phantom. Manual optimization of global weighting correction parameters was performed for the slice 120.

IV. Conclusion

We have demonstrated an improvement an empirical beam hardening correction method, combining three techniques to make the EBHC method fully automatic, with a good approximation for metal artifacts. First, fuzzy C-means was used to perform an automatic segmentation of the metal component. Second, a minimum variance optimization method was used to provide a suitable combination of correction basis functions. Finally, a sub-volume (spatially varying) optimization method was used to account for a varying contribution of metal artefacts through the image. The proposed method has been tested across the variety of samples and acquisition conditions and was able to noticeably diminish the severity of metal artifacts. It was also able to perform similarly to manually optimized method, as well as outperform manual method globally. We also foresee that some of the proposed modifications can be applied to other variations of beamhardening correction algorithms.

Acknowledgements

The authors are grateful to Dr. Martin Krenkel and Dr. Daniel Weiss from Carl Zeiss Industrial Solutions (Oberkochen, Germany) for useful ideas and discussions.

References

[1] Y. Kyriakou, E. Meyer, D. Prell, and M. Kachelrieß, “Empirical beam hardening correction (EBHC) for CT,” Medical Physics, 37, 5179, 2010.

[2] G. Van Gompel, K. Van Slambrouck, M. Defrise, K. J. Batenburg, J. de Mey, J. Sijbers, J. Nuyts, “Iterative correction of beam hardening artifacts in CT,” Med. Phys. 38, S36 (2011).

[3] E. Meyer, R. Raupach, M. Lell, B. Schmidt, M. Kachelrieß, “Frequency split metal artifact reduction (FSMAR) in computed tomography,” Med. Phys., 39(4), pp. 1904-16, 2012.

[4] G.L. Zeng, M. Zeng, “Reducing metal artifacts by restricting negative pixels,” Vis Comput Ind Biomed Art., 4(1):17, 2021.

[5] G.L. Zeng, “A projection-domain iterative algorithm for metal artifact reduction by minimizing the total-variation norm and the negative-pixel energy,” Vis Comput Ind Biomed Art, 2022 Jan 2;5(1):1.

[6] Y. Zhang, Y.F. Pu, J.R. Hu, Y. Liu, J.L. Zhou, “A new CT metal artifacts reduction algorithm based on fractional-order sinogram inpainting,” J X-Ray Sci Technol. 2011;19(3):373–384.

[7] J. W. Stayman, Y. Otake, J. L. Prince, A. J. Khanna, and J. H. Siewerdsen, ‘‘Model-based tomographic reconstruction of objects containing known components,’’ IEEE Trans. Med. Imag., vol. 31, no. 10, pp. 1837–1848, Oct. 2012.

[8] S. Xu, H. Dang, “Deep residual learning enabled metal artifact reduction in CT,” Proc. SPIE 10573, Medical Imaging 2018.

[9] L. Gjesteby, B. De Man; Y. Jin; H. Paganetti, J. Verburg, D. Giantsoudi; G. Wang, "Metal Artifact Reduction in CT: Where Are We After Four Decades?," in IEEE Access, vol. 4, pp. 5826-5849, 2016

[10] M. N. Ahmed, S. M. Yamany, N. Mohamed, A. A. Farag and T. Moriarty, "A modified fuzzy c-means algorithm for bias field estimation and segmentation of MRI data," in IEEE Transactions on Medical Imaging, vol. 21, no. 3, pp. 193-199, March 2002

On use of augmentation for the DNN-based CT image denoising

Prabhat Kc, Kyle J. Myers, M. Mehdi Farhangi, Rongping Zeng

***Abstract*—Augmentation strategies have been suggested to overcome issues related to limited training data for the Deep Learning (DL)-based CT image denoising problem. Although augmentation is, indeed, a good machine learning practice, the extent of improvements achieved by DL-based CT denoising solvers through augmentation pipelines remains to be quantified. Accordingly, in this work, we make use of two different deep neural networks (the REDCNN, the DnCNN) to quantify gains in CT image quality through augmentation. The augmentation strategies considered include computer vision inspired strategies (like scaling, rotating, flipping, image-blending) and the CT forward model-based noise (radiation dose) insertion. Likewise, image qualities considered in this work include visual perception- and data fidelity-based global metrics (like the PSNR, the SSIM, the RSE) which are common in the computer vision literature, and CT bench tests (like the NPS, the MTF) and a task-based detectability assessment (the LCD) from the CT imaging literature. Our preliminary results indicate that the DL solvers trained on augmented data show gains in the global metrics, in low-frequency noise texture components and in image resolution as compared to the ones trained on non-augmented data. However, when the augmented DL-solvers were compared against their low-dose counterparts their performance - with respect to the noise frequency components, resolution, and detection task - was not all improved, and in some cases even worse.**

***Index Terms*—Data Augmentation, Noise Insertion, CT Image Denoising, Deep Learning, Image Quality, Transfer Learning.**

I. Introduction

ONE of the good machine learning (ML) practices to achieve a generalizable performance by a ML model is to train the model with a large dataset. However, accessibility to sufficiently large data repositories in medical applications – that span across different regions, population demographics, practice of medicine and acquisition protocols – continues to be a challenging prospect. As such, a surrogate data generation methodology called the data augmentation technique is frequently employed to increase the available training data. The prime objective of the data augmentation technique is to ensure that a Deep Learning (DL) method does not overtly train to properties specific only to its training data and rather trains to general properties exhibited by a larger patient population. In the machine learning language, a data augmentation strategy seeks to alleviate issues related to overfitting (when a model performs well on the training dataset but fails to perform on an unseen dataset) and underfitting (when a model fails to perform well on the training dataset). Some of the commonly employed augmentation strategies – borrowed from computer vision-based applications to medical imaging problems – include scaling, rotation, flipping, cropping, and contrast blending. Likewise, there are a broad spectrum of augmentation strategies relevant for medical applications. These strategies can be roughly divided into two categories. These include strategies that replicate variations in (i) image acquisition/practice of medicine and (ii) object/subject/patient being imaged. The category (i) type data augmentations are deduced by making use of physics-based forward models that reflect different medical image acquisition methods. This procedure aids in increasing a given training dataset with respect to covariates such as radiation dose, sparse sampling, scanning parameters etc. For the category (ii) type data augmentation, distribution learning ML models like the Generative Adversarial Networks (GANs) are being explored to synthesize training/testing data that properly replicate variations in image textures found in the patient population (without explicitly performing patient-based acquisitions).

Having said this, the extent of improvements in CT image quality through augmentation strategies for medical applications still remain elusive. In this contribution, we seek to quantify gains of DL-based denoising solvers trained on augmented data. We make use of computer vision-inspired image augmentations and CT physics-based noise insertion as the two data augmentation techniques. The latter augmentation methodology is performed under the framework of transfer learning. Likewise, we make use of two differing DL-based convolutional neural networks (CNNs) for the denoising purpose. These include the feed-forward denoising CNN (DnCNN: that incorporates very deep CNNs to formulate a total of 17 layers) [1] and the Residual Encoder-Decoder CNN (REDCNN: that incorporates encoder-decoder layers to constitutes a total of 10 layers) [1]. Finally, we make use of global metrics (such as the Relative Squared Error (RSE), the Peak-Signal to Noise Ratio (PNSR), the Structural Similarity Index (SSIM)), as well as CT bench tests (the contrast-dependent Modulation Transfer Function (MTF), the Noise Power Spectrum (NPS)) and a task-based assessment (the Low-Contrast Detectability (LCD)) to estimate changes in the image qualities of the DL-based denoised solutions trained on the augmented vs non-augmented data.

II. Method

A. DL framework and architecture

Network weights, θ, of the two DNNs, i.e. the DnCNN and the REDCNN considered in this paper, are estimated by solving the following objective function:

$$\hat{\theta} \leftarrow \arg\min_{\theta} \ell(f(\mathbf{X};\theta),\mathbf{Y}), \tag{1}$$

where $\ell(\cdot)$ denotes a loss function and $\mathbf{X},\mathbf{Y} \in \mathbf{R}^{m\times n}$ represent, complimentary, Low-Dose CT (LDCT) and Normal-Dose CT (NDCT) images used to train the two networks. These images are sequestered from the Low-Dose Grand Challenge (LDGC) dataset. The training set comprises 1560 CT images

Corresponding author: Prabhat KC, prabhat.kc@fda.hhs.gov
Affiliations: P. KC, M. M. Farhangi and R. Zeng are from DIDSR/OSEL/FDA. K. J. Myers is FDA, retired.

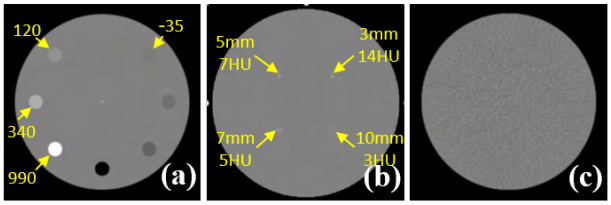


Fig. 1. Simulated phantoms that mimic (a) contrast levels in the CATPHAN600 for measuring the MTF, (b) the CCT189 low contrast body phantom for measuring the LCD, and (c) the cylindrical water phantom for the NPS estimation.

of size 512×512 obtained from 6 different patients. This training data is artificially increased through various procedures (detailed in section II-B). Likewise, our validation/tuning set consists of 36 CT images randomly pre-selected from the same 6 patients before formulating the training set. Finally, our test sets include 223 CT images obtained from a seventh patient, the CATPHAN600 phantom (fig. 1(a)), the CCT189 phantom (fig. 1(b)) and the cylindrical water phantom (fig. 1(c)). These test sets are not used in any of the network training procedures.

B. Dose augmentation framework

The LDCT images in the LDGC dataset correspond to quarter-dose CT (QDCT & $\mathbf{X} = \mathbf{X}_{\text{QD}}$) acquisitions. Accordingly, we make use of a CT-physics based forward model [2] to synthesize radiation dose acquisitions corresponding to 95% dose ($\mathbf{X}_{95}$) and 75% dose ($\mathbf{X}_{75}$). Here we note that the normal-dose and quarter-dose projection measurements provided in the LDGC were acquired in helical mode. However, in absence of the vendor specific forward model for the helical acquisition ($\text{A}_{\text{Helical}}$), we make use of the NDCT images and a CT fan-beam based forward model ($\mathbf{A}_{\text{fan}}$) to synthesize projections in which we insert dose-based noise (η). Mathematically, $\mathbf{X}_{95} = \mathbf{A}_{\text{fan}}^{-1}(\mathbf{A}_{\text{fan}}\mathbf{Y} + \eta_{95\%})$ and $\mathbf{X}_{75} = \mathbf{A}_{\text{fan}}^{-1}(\mathbf{A}_{\text{fan}}\mathbf{Y} + \eta_{75\%})$. The inversion is performed using the Filtered BackProjection (FBP) method. We validated the noise components of the outputs from our realistic dose-based augmentation procedure using the NPS. A representative scan determined from this realistic dose augmentation procedure is provided in fig. 2(b).

Next, we make use of a transfer learning framework to sequentially train the two DNNs on the augmented and non-augmented datasets. This work draws its motivation from the traditional multi-scale based tomographic reconstruction to improve the quality of the restored image. More concretely, rather than performing a straightforward network learning between the quarter- and normal-dose pairs $(\mathbf{X}_{\text{QD}}, \mathbf{Y})$, we sequentially train networks on a series of increasing radiation dose gaps between the input and target pairs. In other words, $\hat{\theta}_{95\text{-ND}}$ is estimated from training pairs $(\mathbf{X}_{95}, \mathbf{Y})$; then $\hat{\theta}_{95\text{-ND}}$ is used as the initializer to estimate $\hat{\theta}_{75\text{-ND}}$ from training pairs $(\mathbf{X}_{75}, \mathbf{Y})$; and finally, $\hat{\theta}_{75\text{-ND}}$ is used as the initializer to estimate $\hat{\theta}_{\text{QD-ND}}$ from the training pairs $(\mathbf{X}_{\text{QD}}, \mathbf{Y})$.

C. Computational optimization

In ref. [3], we showed that the performance of a DNN depends on several choices. These choices include patch size (e.g. P-55 for patch 55×55), learning rate (lr), mini-batch size (mi-b), loss function (e.g. MSE, MAE, MSE with L_1 image prior/ total variation (TV) image prior/ weight decay term,

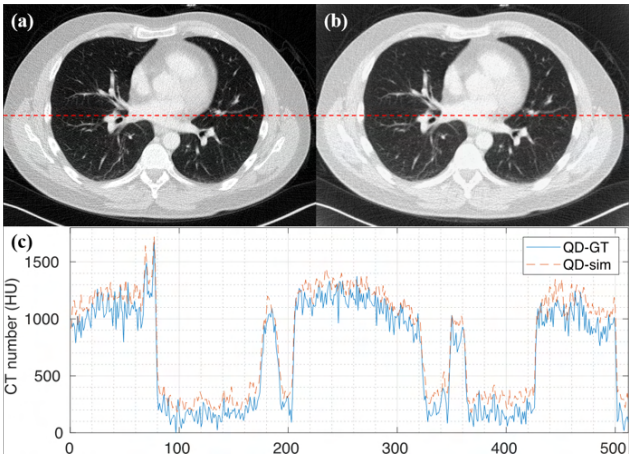


Fig. 2. (a) A QDCT lung image from the LDGC dataset. (b) Corresponding QDCT image from our realistic noise insertion model. (c) Line plot comparison of the two images along the dotted red-line. The display window of images in (a) & (b) is (W: 1300 L: -370).

MS-SSIM), pre-processing choices (in terms of normalization vs no normalization (*normF*)), and pseudo data augmentation (rotation, scaling, flipping, image-blending). This overall optimization procedure can be compartmentalized into the following three stages:

(i) Stage 1 (*without/ aF*): At this stage, we do not perform any augmentation (*aF*) on the training data. Next, the training data for both the DNNs is normalized to the range $[0, 1]$ (i.e., *unity* normalization). The two DNNs are trained to yield the best values w.r.t the global metrics - PSNR and SSIM - on the tuning set. At the end of this stage, the optimized choices for the two networks include REDCNN: (P-96 | lr:10^{-4} | mi-b:32 | MSE | *unity-aF*), DnCNN: (P-55 | lr:10^{-4} | mi-b:32 | MSE+λ·L_1 | λ:10^{-7}| *unity-aF*).

(ii) Stage 2 (*Pseudo/ PaT*): At this stage, we perform the scaling, rotation, flipping & image-blending augmentations (*PaT*) on the training data. Also, normalization type is tuned to yield the best performance for the two DNNs w.r.t the HU accuracy. Accordingly, the optimized choices for these two networks include: REDCNN: (P-96 | lr:10^{-5} | mi-b:64 | MSE+λ·TV| λ:10^{-3}| *normF-PaT*), DnCNN: (P-55 | lr:10^{-4} | mi-b:32 | MSE+λ·L_1 | λ:10^{-7}| *unity-PaT*).

(iii) Stage 3 (*Realistic/ RaT*): At this stage, we train the two DNNs on realistically dose augmented datasets $((\mathbf{X}_{95}, \mathbf{Y}), (\mathbf{X}_{75}, \mathbf{Y}))$ - estimated from the CT noise based forward model - under the framework of transfer learning as detailed in our previous section II-B. The corresponding choices for the two DNNs include: REDCNN: (P-96 | lr:10^{-5} | mi-b:64 | MSE+λ·TV| λ:10^{-4}| *normF-RaT*), DnCNN: (P-55 | lr:10^{-4} | mi-b:32 | MSE+λ·L_1 | λ:10^{-7}| *unity-RaT*).

In summary, stage 1 depicts the scenario where the DL is performed in the absence of any forms of augmentation. Stage 2 and stage 3 depict scenarios where the DL is supplemented by the computer vision-inspired and by the realistic CT noise-based augmentation strategies respectively.

III. Results & Discussions

A. Global Metrics

Table I lists the RSE, PSNR and SSIM values from the two DNNs applied on the QDCT test images, relative to their NDCT counterparts. Columns 3 & 4 in the table indicate that even when the two DNNs are trained *without-augmentation* (*aF*), they show significant improvement in their global metric values as compared to that from the QDCT. There is no further gain in the global metric values for the DnCNN at the stage 2 (*pseudo or PaT*) augmentation and the stage 3 (*realistic or RaT*) augmentation (besides the SSIM value that is best at the *PaT* stage). On the other hand, REDCNN outputs its best global metric-based performance at the *RaT* stage (besides the SSIM value that is best at the *PaT* stage). Overall, this evaluation suggests that the DnCNN (with its 17 convolutional layers that encapsulate batch normalization layers for a faster convergence) does not gain any new CT imaging related information through the augmentation pipeline. On the contrary, the REDCNN (with its encoder-decoder based 10 weight layers that incorporate a rich set of skip connections) shows a gain in its performance through the augmentation pipeline.

Table I: RSE, PSNR and SSIM values corresponding to the QDCT images and the DnCNN & REDCNN outputs at different augmentation stages.

Network	Metric	Quarter Dose	Stage 1 (aF)	Stage 2 (PaT)	Stage 3 (RaT)
DnCNN	RSE	0.0141 (±0.003)	**0.0087 (±0.005)**	0.0096 (±0.005)	0.0089 (±0.005)
	PSNR	32.615 (±1.075)	**35.129 (±2.106)**	34.839 (±2.236)	35.095 (±2.35)
	SSIM	0.7483 (±0.039)	0.9053 (±0.014)	**0.9062 (±0.019)**	0.9039 (±0.019)
REDCNN	RSE	0.0141 (±0.003)	0.0089 (±0.005)	0.0078 (±0.005)	**0.0075 (±0.004)**
	PSNR	32.615 (±1.075)	35. 097 (±2. 159)	35.755 (±2. 248)	**35.868 (±2.17)**
	SSIM	0.7483 (±0.039)	0. 9048 (±0. 019)	**0.9061 (±0.019)**	0.9050 (±0.019)

B. MTF

To evaluate the CT bench tests from the phantoms, first, we generated a noiseless sinogram from the CATPHAN600 (fig. 1 (a)). The sinogram was inverted using the FBP method - paired with a sharp kernel that was used to reconstruct the training data - and the two DNNs were applied on reconstructed CATPHAN600. Subsequently, MTF50% values at the four contrast disks $(990, 340, 120, -35$ HU) were estimated to represent image resolution (figs. 3 (a,b)). Both the DNNs show significant improvement in their MTF50% values at the *pseudo-augmentation* stage. For the DnCNN, there is a further minor increment in the MTF50% values at the *realistic-augmentation* stage. On the contrary, for the REDCNN, there is a minor decay in the MTF50% values at the *realistic-augmentation* stage as compared to that at the *pseudo-augmentation* stage. Broadly speaking, after the *pseudo-* or *realistic-augmentation* stages, the maximum MTF50% values for the two DNNs – for all four contrast disks – plateaus at or right below their corresponding MTF50% values from the FBP-sharp method.

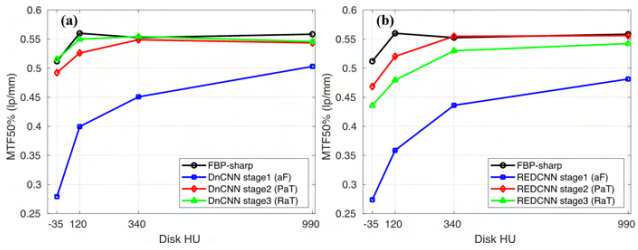


Fig. 3. MTF50% plots of the (a) DnCNN and (b) REDCNN applied on the reconstructed CATPHAN600. Each plot illustrates MTF50% values obtained from the FBP method and the different stages of augmentation.

C. NPS

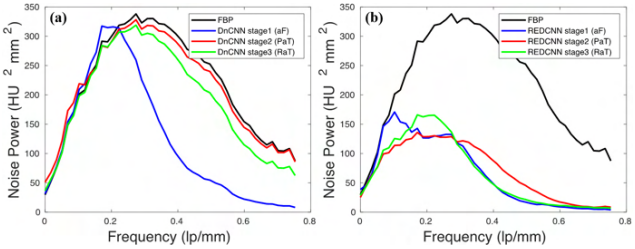


Fig. 4. Radial 1D NPS plots evaluated from (a) the DnCNN and (b) the REDCNN applied on the noisy realizations of the reconstructed cylindrical phantom. Each plot includes 1D NPS curves corresponding to the different augmentation strategies.

The NPS was estimated making use of 30 noisy scans/realizations from the cylindrical water phantom that were reconstructed using the FBP method with a sharp kernel. The resulting radial 1D NPS plots for the two DNNs are depicted in fig. 4. In the case of the DnCNN, we notice that for frequency bands above 0.3 lp/mm, there is a significant gain in noise power at the *pseudo-* or *realistic-augmentation* stages as compared to its performance at the *without-augmentation* stage. Also, the DnCNN's performance at the *pseudo-* or *realistic-* stages is similar to that from the FBP method. For the REDCNN, the overall nature of its radial 1D NPS curves - at different augmentation stages - remain mostly similar to one another; thereby suggesting that the REDCNN does not learn to improve noise content information through the augmentation pipeline.

D. LCD

For the LCD test, we simulated 200 noisy scans from the CCT189 phantom, and 100 noisy scans from the uniform cylindrical water phantom at each of the five dose levels, i.e. $(30\%, 50\%, 70\%, 85\%, 100\%)$ of incident flux (3×10^5). For each of the four inserts, one signal-present (SP) ROI was extracted from the CCT189 phantom image and five signal-absent (SA) ROIs were extracted at the vicinity of the insert location from each of the uniform phantom images. As a result, a total of 200 SP and 500 SA ROIs for each low contrast insert were created to evaluate its corresponding detectability. The LCD plots of the two DNNs for the 3 mm & 10 mm inserts are depicted in fig. 5. For the DnCNN (*aF, PaT, RaT*) applied on the 3 mm insert, we do not observe any gain in the detectability of the insert at any of the five dose levels. On the contrary, there is a decay in the detectability performance between 0% to 5% for the DnCNN (*aF, PaT*). We observe similar results for the DnCNN (*aF, PaT, RaT*) applied on 5 mm & 7 mm inserts. However, for the DnCNN (*RaT*) applied on the 10 mm insert, we observe a gain in the LCD values between 0% to 5%, at each dose level, albeit with a large

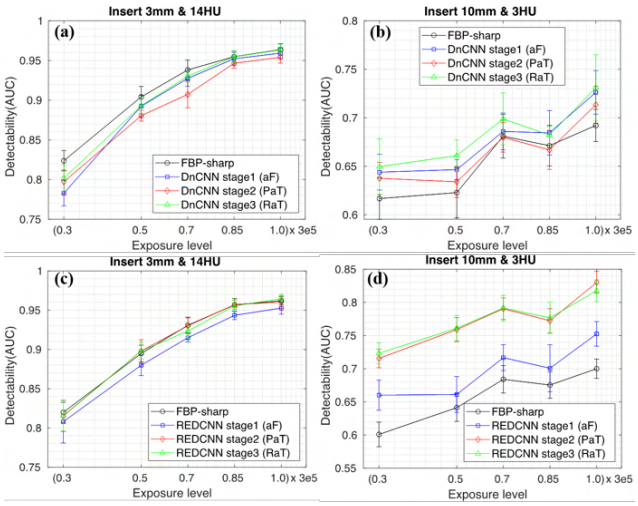


Fig. 5. LCD of the DnCNN for (a) insert 3mm, (b) insert 10mm and of the REDCNN for (c) insert 3mm, (d) insert 10mm. Each plot depicts LCD values obtained from the FBP method and the different stages of augmentation.

spread in their corresponding error bars. For the REDCNN (*aF, PaT, RaT*) applied on the 3 mm insert, their corresponding detectability values relative to the five dose levels, are around or right below to those obtained from the FBP method. Similar results are seen for the REDCNN (*aF, PaT, RaT*) applied on the 5 mm insert. For the 10 mm insert, we observe a gain in the detectability values ranging from 2% to 5% for the REDCNN (*aF*) and between 11% to 15% for the REDCNN (*PaT, RaT*).

IV. Conclusions

In this contribution, we set out to quantify the improvements in the DL-based CT denoising solvers through augmentation strategies. Correspondingly, for the DnCNN, its very deep network architecture was such that its global metrics-based performance peaks at the *without-augmentation* stage. There is a small drop in its global metrics-based performance at the *pseudo* and *realistic-augmentation* stages. From the LCD viewpoint (that is relevant to diagnostic accuracy), DnCNN yields sub-par performance at *without-augmentation* and *pseudo-augmentation* stages. The gain in its LCD performance was only seen at the *realistic-augmentation* stage, albeit for only large inserts (7 mm & 10 mm) and with a caveat that the detectability of the DnCNN (*RaT*) for low-dose was never as good as that of the FBP at high-dose.

For the REDCNN, its global metric values sequentially increase until the *realistic-augmentation* stage. However, from the LCD viewpoint, the REDCNN's performance saturates at the *pseudo-augmentation* stage.

Next, from the resolution viewpoint, for both the DNNs, the *pseudo-augmentation* stage yields significantly higher MTF50% values than that obtained at the *without-augmentation* stage (these values are at or near to the FBP-sharp limit). However, the presence of edge-like structures throughout absolute difference plots (in figs. 6(d-i)) from both the DNNs (for all augmentation stages) points at the limitation of using the digital CATPHAN600 (in particular, using its locally constant regions i.e., disks) to estimate resolution. A proper account of the true nature of the resolution capacity of the two DNNs, relative to different augmentation strategies, will require a further resolution analysis with the aid of a line/wire pattern phantom. Note that the CT scan considered to determine the difference plots is a challenging image for the DL solvers (as it exhibits bone, muscles fibers and fatty regions). Also, the contrast window width of the difference plots is set to properly depict edges in these plots.

Finally, from the NPS viewpoint, it appears that the DnCNN learns not to denoise scans that are devoid of anatomical structures with the sequential augmentation steps (fig. 4(a)). Yet, the visible distinction between the low-contrast outer regions (outside the dotted window (w) in fig. 6) and the high-contrast inner regions (inside the dotted window (w) in fig. 6), and the presence of similar edge-like features (inside (w)) in figs. 6(d-f)) suggest that augmentations (*pseudo* or *realistic*) do not lead to any improvement in low, as well as high, frequency noise components. For the REDCNN, its 1D NPS curves (fig. 4(b)) paired with difference plots (fig. 6(g-i)) indicate that augmentations (*pseudo* or *realistic*) do not lead to any change in high frequency noise components.

In conclusion, when the DL-based denoising solvers were compared against their low-dose counterparts, we did not find any improvement in the solvers' LCD performance, for the 3 mm, 5 mm, & 7 mm, that merits to a statistically significant level. Similarly, the solvers' improvements w.r.t the NPS (evaluated using the uniform phantom) and the MTF (estimated from the contrast disks) do not translate in the noise content information and the resolution of the denoised CT outputs.

In future, we seek to extend this preliminary work with a larger patient data training, and perform more generalizability tests (w.r.t dose levels between QD & ND) and subgroup analyses (by body regions for CT scans vs CT bench testing).

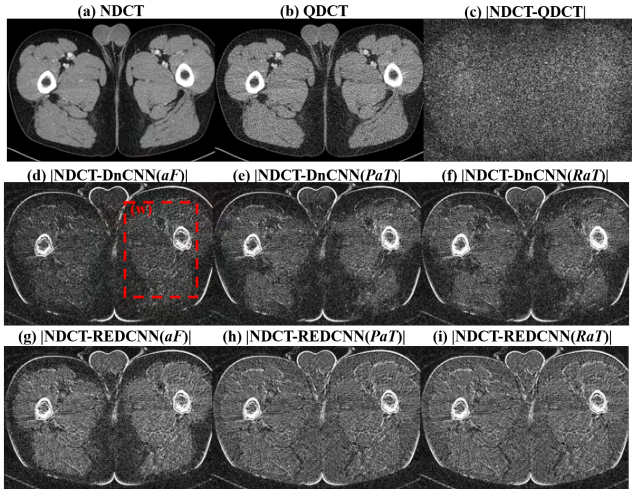


Fig. 6. CT images in plots (a) & (b) with the display window of (W:400 L:50) HU. Absolute difference images in plots (c-i) with the display window of [0, 88] HU. Dotted red window (w) is used indicate the central muscular mass with bone & edge contrasts in plots (d-i).

References

[1] S. Escalera, S. Ayache, J. Wan, M. Madadi, U. Güçlü, and X. Baró, *Inpainting and Denoising Challenges*. Springer, 2019.

[2] D. Zeng, J. Huang, Z. Bian, S. Niu, H. Zhang, Q. Feng, Z. Liang, and J. Ma, "A simple low-dose x-ray ct simulation from high-dose scan," *IEEE transactions on nuclear science*, vol. 62.

[3] P. KC, R. Zeng, M. M. Farhangi, and K. J. Myers, "Deep neural networks-based denoising models for CT imaging and their efficacy," in *Medical Imaging 2021: Physics of Medical Imaging*, vol. 11595.

Joint Multi-channel Total Generalized Variation Minimization and Tensor Decomposition for Spectral CT Reconstruction

Huihua Kong, Xiangyuan Lian, Jinxiao Pan, and Hengyong Yu, *Member, IEEE*

Abstract—**Photon-counting detector based spectral computed tomography (CT) has great potential in material decomposition, tissue characterization, lesion detection, and other applications. For a fixed total photon number or radiation dose, the increase of the number of channels will lead to the decrease of the number of photons in a single channel, resulting in degraded image quality of the reconstructed image. This is difficult to meet the practical applications for material decomposition. etc. To improve the quality of image reconstruction, we propose a spectral CT reconstruction algorithm based on joint multi-channel total generalized variation (TGV) minimization and tensor decomposition. On one hand, the algorithm takes joint multi-channel TGV function as regularization. The regularization extends total generalized variation to the vector, and the sparsity of singular values is used to promote the linear dependence of the image gradient. On the other hand, the multi-channel images share the same physical structure, and the algorithm employs the non-local feature similarity in the image domain. Similar image blocks are clustered into a four-order tensor group, and the noise was reduced by sparse representation of high-dimensional tensors. Experiment results show the proposed algorithm can further improve the quality of reconstructed image and preserve the edge and details of the image for spectral CT.**

Index Terms—**spectral CT, image reconstruction, tensor decomposition, total generalized variation.**

I. Introduction

Photon-counting detector (PCD) based spectral CT has attracted an increasing attention [1]. However, a single energy bin contains only a fraction of the total photon, and most PCDs can only accommodate a limited count rate, the multiple projection datasets obtained by PCD usually contain very strong Poisson noise [2]. This makes it difficult to meet the challenges of practical applications.

Inspired by the image domain non-local feature similarity, Zhang *et al.* extended the traditional vectorized dictionary learning to tensor dictionary learning (TDL) for spectral CT reconstruction. The TDL based reconstruction algorithms for spectral CT have some limitations [3]. When the noise is too large, it is impossible to distinguish the noise from the organizational structures. And block artifacts appear in the reconstructed images. Recently, tensor decomposition methods are widely used in image denoising. A high dimensional tensor can be represented approximately by the weighted sum of a series of low-rank tensor data, leading to effective noise suppression and artifact reduction.

The correlation of reconstructed image in each spectral channel has attracted more and more attention. Rigie and Rivier developed a spectral CT reconstruction method based on vector total variation (VTV) [4], which combines the nuclear norm to promote the sparsity of the multi-channel gradient vector field. Knoll *et al*. used the second-order total generalized variation as a special multi-channel regularization function (MTGV), and the structural information was shared in the process of reconstruction while the unique differences were retained [5].

In order to make better use of sparsity in the image domain and the correlation of the reconstructed image for multi spectral channels, we propose a spectral CT reconstruction algorithm based on joint multi-channel total generalized variation and tensor decomposition, and we call it MTGV-TD.

II. Models and Method

A. Joint multi-channel TGV regularization based on nuclear norm

In 2014, Rigie and Patrick extended the single-channel gradient vector to the vector field and defined the discrete Jacobian matrix in the following form:

$$\mathrm{J}_{\boldsymbol{u}}(i,j)=\begin{bmatrix}(\boldsymbol{u}_1)_x(i,j) & (\boldsymbol{u}_1)_y(i,j)\\ (\boldsymbol{u}_2)_x(i,j) & (\boldsymbol{u}_2)_y(i,j)\\ \vdots & \vdots\\ (\boldsymbol{u}_L)_x(i,j) & (\boldsymbol{u}_L)_y(i,j)\end{bmatrix} \tag{1}$$

This work was supported in part by the National Science Foundation of China under Grant 61971381, 61871351.

Huihua Kong is with School of Science, North University of China, Taiyuan , China, 030051, （huihuak@163.com）,

Xiangyuan Lian is with School of Science, North University of China, Taiyuan , China, 030051, （1393550566@qq.com）,

Jinxiao Pan is with School of Science, North University of China, Taiyuan , China, 030051 （panjx@nuc.edu.cn）,

Hengyong Yu is with Department of Electrical and Computer Engineering, University of Massachusetts Lowell, Lowell, MA 01854, USA, (hengyong-yu@ieee.org).

where $\boldsymbol{u}=(\boldsymbol{u}_1,\boldsymbol{u}_2,\cdots,\boldsymbol{u}_L)^{\mathrm{T}}\in U^L$.

When two images have the same curve, the two images have the same direction gradient and the converse is also true [6]. If all the gradient vectors of each channel image are parallel or antiparallel, then the rank of the Jacobian matrix will be 1. Hence, there will be only one non-zero singular value. Based on those facts, if the image gradients of each channel are parallel, the nuclear norm will be minimized. Our algorithm will extend the total generalized variation to the vector:

$$\mathrm{R_N}(\boldsymbol{u})=\alpha_0\left\|\nabla\boldsymbol{u}-\boldsymbol{w}\right\|_{1,\mathrm{nuc}}+\alpha_1\left\|\varepsilon(\boldsymbol{w})\right\|_{1,\mathrm{Frob}} \tag{2}$$

where

$$\left\|\nabla\boldsymbol{u}-\boldsymbol{w}\right\|_{1,\mathrm{nuc}}=\left\|\left\|(\nabla\boldsymbol{u}-\boldsymbol{w})(i,j)\right\|_{\mathrm{nuc}}\right\|_1,$$

$$\left\|\varepsilon(\boldsymbol{w})\right\|_{1,\mathrm{Frob}}=\left\|\left\|\varepsilon(\boldsymbol{w})(i,j)\right\|_{\mathrm{Frob}}\right\|_1,$$

$$\left\|(\nabla\boldsymbol{u}-\boldsymbol{w})(i,j)\right\|_{\mathrm{nuc}}=\left\|\boldsymbol{\sigma}\right\|_1.$$

$\boldsymbol{\sigma}$ is the singular value vector of the $(\nabla\boldsymbol{u}-\boldsymbol{w})(i,j)$ matrix, and $\|\cdot\|_{\mathrm{Frob}}$ is the sum of squares of all elements in the matrix (the Frobenius norm of the matrix). The discrete gradient operator of the vector field is $\nabla: U^L\to U^{L\times 2}$, The symmetric gradient operator is $\varepsilon: U^{L\times 2}\to U^{L\times 4}$, and $\nabla\mathbf{u}(i,j)=\mathrm{J}_{\mathbf{u}}(i,j)$, $\boldsymbol{w}\in U^{L\times 2}$, $\varepsilon(\boldsymbol{w})\in U^{L\times 4}$.

if

$$\boldsymbol{w}=\begin{bmatrix} w_{11} & w_{21}\cdots w_{L1} \\ w_{12} & w_{22}\cdots w_{L2}\end{bmatrix}^{\mathrm{T}}, \tag{3}$$

then

$$\varepsilon(\mathbf{w})=\begin{bmatrix} \partial_x^- w_{11} & \dfrac{\partial_x^- w_{11}+\partial_y^- w_{12}}{2} & \dfrac{\partial_y^- w_{11}+\partial_x^- w_{12}}{2} & \partial_y^- w_{12} \\ \partial_x^- w_{21} & \dfrac{\partial_x^- w_{21}+\partial_y^- w_{22}}{2} & \dfrac{\partial_y^- w_{21}+\partial_x^- w_{22}}{2} & \partial_y^- w_{22} \\ \vdots & \vdots & \vdots & \vdots \\ \partial_x^- w_{L1} & \dfrac{\partial_x^- w_{L1}+\partial_y^- w_{L2}}{2} & \dfrac{\partial_y^- w_{L1}+\partial_x^- w_{L2}}{2} & \partial_y^- w_{L2}\end{bmatrix} \tag{4}$$

where ∂_x^-, ∂_y^- Represents the first order backward difference between horizontal and vertical directions, $\mathbf{w}_{ij}\in U$.

B. Tensor Decomposition

Tensor decomposition has been widely used in image reconstruction and image processing, *etc*. Usually, there are two main tensor decomposition methods: the Tucker decomposition and the Candecomp/Parafac(CP) decomposition. Here, we only address the CP decomposition.

CP decomposition can decompose a N$^{\mathrm{th}}$ tensor $\boldsymbol{\chi}$ into the weighted sum of rank-one tensors, which can be expressed as

$$\begin{aligned}\boldsymbol{\chi}&=\sum_{h=1}^{H}\lambda_h a_{1h}\circ a_{2h}\cdots\circ a_{nh}\circ\cdots\circ a_{Nh}\\&=\left[\left[\lambda;A_1,A_2,\cdots,A_n,\cdots,A_N\right]\right]\end{aligned} \tag{5}$$

where λ_h is the weight function, $a_{nh}\in R^{1n},(\mathrm{n}=1,2,\cdots,\mathrm{N})$ is a normalized vector and the symbol " $\circ$ " represents the outer product. The sparsity level of representation can be controlled by adjusting the parameter H. We can apply tensor decomposition for image restoration. The mathematical model can then be expressed as

$$\arg\min\left\|\boldsymbol{\chi}-\bar{\boldsymbol{\chi}}\right\|_F^2 \quad s.t.\ \bar{\boldsymbol{\chi}}=[[\lambda;A_1,A_2,\cdots,A_n,\cdots,A_N]] \tag{6}$$

where $\boldsymbol{\chi}$ and $\bar{\boldsymbol{\chi}}$ are the corrupted and restored images,respectively. $\bar{\boldsymbol{\chi}}$ can be obtained using the ALS method by solving each A_n alternatively [7].

C. Mathematical Model for MTGV-TD

In this algorithm, tensor decomposition is used to improve the image blocks quality, and joint multi-channel TGV function is sharing information between channels. Combining multi-channel TGV function and tensor decomposition a spectral CT reconstruction algorithm is proposed. Its objective function can be expressed as the following convex minimization problem:

$$\arg\min_{x,\{Z_c(\vec{x})\}}\frac{\lambda}{2}\sum_{\mathrm{s}=1}^{\mathrm{S}}\|\boldsymbol{A}\boldsymbol{x}-\boldsymbol{y}\|_2^2+R_N(\boldsymbol{x})+\frac{\beta}{2}(\|\sum_{c=1}^{C}Z_c(\boldsymbol{x})-Z_c(\vec{\boldsymbol{x}})\|_F^2) \tag{7}$$

$$s.t.\ R_N=\alpha_0\|\nabla\boldsymbol{x}-\boldsymbol{\omega}\|_{1,\mathrm{nuc}}+\alpha_1\|\frac{(\nabla\boldsymbol{\omega}+(\nabla\boldsymbol{\omega})^T)}{2}\|_{1,Frob}$$

$$\overrightarrow{Z_c(\boldsymbol{x})}=[[\lambda_c;A_1,A_2,A_3,A_4]]$$

where $\boldsymbol{x}=(\boldsymbol{x}_1,\boldsymbol{x}_2,\cdots,\boldsymbol{x}_S)$, $\boldsymbol{y}=(\boldsymbol{y}_1,\boldsymbol{y}_2,\cdots,\boldsymbol{y}_S)$, S indicates the number of channels, $\boldsymbol{x}_i$ $(i=1,2,\cdots S)$ represents the reconstructed image in i$^{\mathrm{th}}$ energy channel, $\boldsymbol{y}_i$ $(i=1,2,\cdots S)$ represents the projection data in i$^{\mathrm{th}}$ energy channel, C and Z_c denote the group number and group extraction operator.

D. Solution

Refer to (7), there are two variables that need to optimization. We divides it into the following two sub-problems and then adopts the method of alternating optimization to solve them :

$$\arg\min_{\boldsymbol{x}}\frac{\lambda}{2}\|\boldsymbol{A}\boldsymbol{x}-\boldsymbol{y}\|_2^2+R_N(\boldsymbol{x})+\frac{\beta}{2}(\|\sum_{c=1}^{C}Z_c(\boldsymbol{x})-Z_c(\vec{\boldsymbol{x}})\|_F^2), \tag{7a}$$

$$\arg\min_{\{Z_c(\boldsymbol{x})\}}\frac{\beta}{2}(\|\sum_{c=1}^{C}Z_c(\boldsymbol{x})-Z_c(\vec{\boldsymbol{x}})\|_F^2). \tag{7b}$$

By introducing variables, equation (7a) is transformed into a constrained problem, which can be summarized as:

$$\begin{aligned}&\arg\min_{\boldsymbol{x}}\frac{\lambda}{2}\|\boldsymbol{A}\boldsymbol{x}-\boldsymbol{y}\|_2^2+R_N(\boldsymbol{x})+\frac{\beta}{2}(\|\sum_{c=1}^{C}Z_c(\boldsymbol{x})-Z_c(\vec{\boldsymbol{x}})\|_F^2)\\&\text{s.t.}\quad\boldsymbol{u}=\boldsymbol{x}\end{aligned} \tag{8}$$

. Equation (8) can be converted into another unconstrained optimization function

$$\begin{aligned}&\arg\min_{\boldsymbol{x},\boldsymbol{u}}\frac{\lambda}{2}\|\boldsymbol{A}\boldsymbol{x}-\boldsymbol{y}\|_2^2+R_N(\boldsymbol{u})+\frac{\mu}{2}\|\boldsymbol{x}-\boldsymbol{u}\|_2^2\\&+\frac{\beta}{2}(\|\sum_{c=1}^{C}Z_c(\boldsymbol{x})-Z_c(\vec{\boldsymbol{x}})\|_F^2)\end{aligned} \tag{9}$$

:

Equation (9) contains two variables, which are optimized by using the method of alternating iteration. Equation (9) is divided into two sub-problems:

$$\arg\min_{\boldsymbol{x}} \frac{\lambda}{2}\|\boldsymbol{A}\boldsymbol{x}-\boldsymbol{y}\|_2^2+\frac{\mu}{2}\left\|\boldsymbol{x}-\boldsymbol{u}\right\|_2^2 + \frac{\beta}{2}(\|\sum_{c=1}^{C} Z_c(\boldsymbol{x})-Z_c(\vec{\boldsymbol{x}})\|_F^2) \quad , \tag{10}$$

$$\arg\min_{\boldsymbol{u}} R_N(\boldsymbol{u})+\frac{\mu}{2}\left\|\boldsymbol{x}-\boldsymbol{u}\right\|_2^2 . \tag{11}$$

Gradient descent method is adopted to solve equation (10), then:

$$\boldsymbol{x}^{(\mathrm{k}+1)}=\boldsymbol{x}^{(\mathrm{k})}-\lambda\boldsymbol{A}^T(\boldsymbol{A}\boldsymbol{x}-\boldsymbol{y})+\mu(\boldsymbol{x}^{(\mathrm{k})}-\boldsymbol{u}^{(\mathrm{k})}) - \beta\sum_{c=1}^{C} Z_C^T(\mathrm{Z}_C(\boldsymbol{x}^{(\mathrm{k})})-(Z_c(\vec{\boldsymbol{x}}))^k) \quad . \tag{12}$$

A first-order Primal-Dual algorithm is used to approximate the global optimal solution of equation (11) [8].

Because each group is independent regarding the optimization process, equation (7b) can be rewritten into

$$\arg\min_{\overrightarrow{Z_C(\boldsymbol{x})}} \left\|Z_c(\boldsymbol{x}^{(\mathrm{k}+1)})-\overrightarrow{Z_c(\boldsymbol{x})}\right\|_F^2 \quad s.t.\ \overrightarrow{Z_c(\boldsymbol{x})}=[[\lambda_c;A_1,A_2,A_3,A_4]] \tag{13}$$

Equation (13) is solved by using ALS method [7].

III. Experimental Results

The major goal of this paper is to evaluate the performance of the MTGV-TD for spectral CT. The algorithms of SART, TV, TDL and NTGV are implemented for comparison [9,10] . All the algorithms are implemented in a hybrid mode of Matlab and C++. While the interface is implemented in Matlab, all the extensive computational parts are implemented in C++ and complied via MEX function.

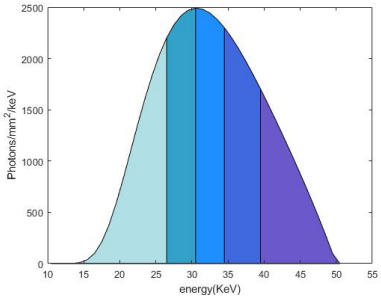


Fig.1 Spectrum used for numerical simulation

A numerical mouse thorax phantom generated by the MOBY software is used for simulation experiments and 1.2% (by volume) iodine contrast agent is introduced into the blood circulation. The model size is 20 × 20 mm^2 , and the resolution is 512 × 512. The scanning radius is 100 mm, and the length of virtual detector located at the center of the object is 20 mm. There are 320 detector units in total, each of which is 0.0625mm. The voltage is 50kVp and energy spectrum is divided into five energy channels: WE1 = {11keV-26keV}, WE2 = {27keV-30keV},WE3 = {31keV-34keV}, WE4= {35keV-39keV} and WE5 = {40keV-50keV}, as shown in figure 1. In the simulation experiments, an isometric fan-beam scanning is adopted. For each X-ray path, 50000 photons are assumed emitted from the X-ray source. The projection data with Poisson noises are generated with expectations being the number of photons received in the corresponding noise-free case. The parameters used in this algorithm are λ=50 , β=0.8 , μ=0.2 , C=120 .

In this experiment, Table 1 shows the comparison results of PSNR, NRMSE and SSIM values between reconstructed images and comparison images in representative channels with full scan 360 projections. It can be seen from table 1 that the reconstruction accuracy of MTGV-TD is superior to the comparison algorithms for all the channels. Fig. 2 gives the corresponding reconstructed images of channels 1, 3 and 5.

From figure 2, one can see that SART algorithm has strong noise in image reconstruction. Image blurring and detail missing appear in TV algorithm, and it is difficult to distinguish noise from details. The overall denoising ability of TDL is weaker than that of MTGV. The reconstructed image of the MTGV algorithm shows that it can guarantee clearer edges while denoising. The proposed MTGV-TD algorithm in this paper reconstructs the image with clear edges and obvious details, and it achieves better noise reduction effect and detail preservation ability. Fig.3 shows the comparison results of PSNR, NRMSE and SSIM values between reconstructed images and comparison images with 360, 180, 120, and 90 uniformly sampled full scan projections. It can be seen that the reconstruction effect will become better with the increase of projection number.

IV. Conclusion

In this paper we propose a spectral CT reconstruction algorithm based on tensor decomposition and joint multi-channel total generalized variation. The algorithm improves the reconstructed image quality by using the image domain sparse condition and the information correlation among channels. The experiment results show this method can not only effectively suppress the noise, but also protect the edge and detail features of the images.

References

[1]K Taguchi, J S Iwanczyk. “Vision 20/20: Single photon counting X-ray d etectors in medical imaging,” Med Phys, vol. 40, No. 10, pp.100901, 2 013, DOI. 10.1118/1.4820371.

[2]K Taguchi, J S Iwanczyk, “Vision 20/20: Single photon counting x-ray de tectors in medical imaging,”Medical physics, vol. 40, No.10, pp.10090 1, 2013, DOI. 10.1118/1.4820371

[3]Y B Zhang, X Q Mou, G Wang, H Y Yu, “Tensor-Based Dictionary Lear ning for Spectral CT Reconstruction,” IEEE Transactions on Medical I maging, vol. 36, No. 1, pp. 142–154, 2017, DOI. 10.1109/TMI.2016.26 00249.

[4] D S Rigie, J L R Patrick, “Joint reconstruction of multi-channel, spectr al CT data via constrained total nuclear variation minimization,” Phys Med Biol, vol. 60, No. 5, pp. 1741–17621, 2015, DOI. 10.1088/0031-9 155/60/5/1741.

Table.1 Quantitative evaluation results of different algorithms

	WE1			WE3			WE5		
	PSNR	MSE	SSIM	PSNR	MSE	SSIM	PSNR	MSE	SSIM
FBP	19.4523	0.0314	0.8148	17.9566	0.0151	0.8570	16.2239	0.0108	0.8732
SART	37.8556	0.0038	0.9877	32.2237	0.0029	0.9910	28.7681	0.0026	0.9924
TV	40.5474	0.0028	0.9954	36.1795	0.0019	0.9968	33.5872	0.0015	0.9974
TDL	41.7601	0.0024	0.9951	40.6116	0.0011	0.9986	38.3339	0.0008	0.9990
MTGV	41.6307	0.0024	0.9961	38.4977	0.0014	0.9980	36.3173	0.0011	0.9985
MTGV-TD	42.4184	0.0022	0.9965	40.9714	0.0011	0.9988	39.4804	0.0007	0.9992

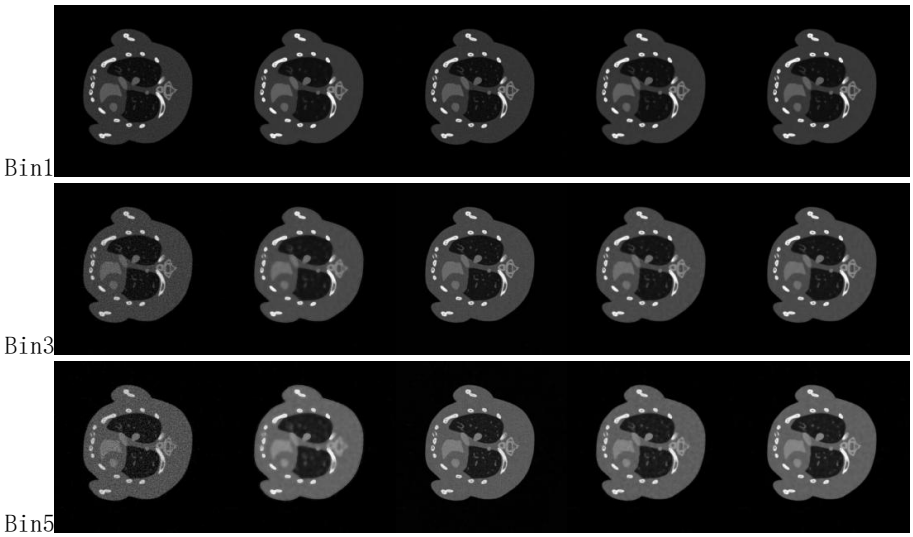
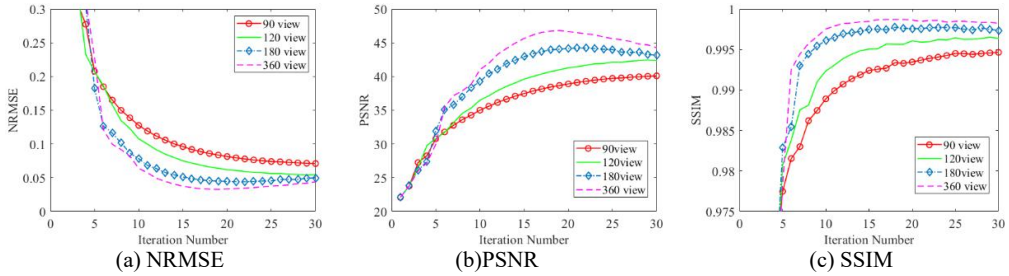


Fig.2 The reconstruction results of the mouse model form 360 projections. From left to right, the rows are SART, TV, TDL, MTGV and MTGV_TD. From top to bottom, the columns are 1st, 3rd and 5th energy channels.



(a) NRMSE (b)PSNR (c) SSIM

Fig.3 Convergence curves of (a)NRMSE, (b)PSNR and (c) SSIM as a function of iterations under different angle

[5] F Knoll, M Holler, T Koesters, etc, "Joint MR-PET Reconstruction Using a Multi-Channel Image Regularizer," IEEE Transactions on Medical Imaging, vol. 36, No. 1, pp. 1–16, 2017, DOI. 10.1109/TMI.2016.2564989.

[6] J E Matthias, R A Simon, "Vector-valued image processing by parallel level sets,"IEEE Transactions on Image Processing, Vol. 23, No.1, pp. 9-18, 2014,DOI. 10.1109/TIP.2013.2277775.

[7] D L Hu, W W Wu, M R Xu, et al, "SISTER: spectral-image similarity-based tensor with enhanced-sparsity reconstruction for sparse-view multi-energy CT," IEEE Trans Comput Imaging, vol. 6, pp. 477–490, 2020, DOI. 10.1109/TCI.2019.2956886.

[8] J H Jorgensen, E Y Sidky, X Pan, " Convex optimization problem prototyping for image reconstruction in computed tomography with the Chambolle–Pock algorithm,"Physics in Medicine and Biology, Vol. 57, No.10, pp. 3065-3091, 2012, DOI. 10.1088/0031-9155/57/10/3065.

[9] W K Zhang, H M Zhang,L Y Wang, et al, "Limited angle CT reconstruction by simultaneous spatial and Radon domain regularization based on TV and data-driven tight frame," Nucl Instr Meth Phys Res A, Vol. 880, pp. 107-117. 2018, DOI. 10.1016/j.nima.2017.10.056.

[10] X Y Lian, H H Kong, J X Pan, et al, "Joint multi-channel total generalized variational algorithm for spectral CT reconstruction," Opto-Electron Eng, Vol. 48, N0. 9, pp. 210211, 2021, DOI: 10.12086/oee.2021.210211.

Trade-offs between redundancy and increased rank for tomographic system matrices

Feriel Khellaf and Rolf Clackdoyle

***Abstract*—In this paper we examine whether increasing the rank of a tomographic system matrix by slightly changing the measurement lines might improve the reconstructed image quality in region-of-interest (ROI) tomography. The image quality is assessed by the size of the reconstructible region and by the reconstructed noise of the pixels in that region. A theorem is presented to specify cases when changing elements in the system matrix leads to solving new pixels. Numerical simulations were performed on several modest sized examples to measure the bias and variance of the reconstructed images using system matrices with the same size but different rank. The results show that while increasing the rank might help to solve new pixels, it can also lead to increases in the pixel variances of the initial ROI.**

I. Introduction

Region-of-interest (ROI) tomography consists of reconstructing only a part of the whole image from truncated data [1]. Truncated data arises when the patient is not completely inside the field of view (FOV). ROI tomography can be useful to reduce the radiation dose given to the patient or for partial reconstruction when the patient is larger than the FOV. The reconstructible ROI depends on the acquired projection lines, described in the system matrix mapping the image to the measurements.

We investigate how small modifications of the acquired projection lines might improve ROI reconstruction. One improvement could be the size of the ROI, i.e. the number of reconstructible pixels. Another improvement might be in the reconstruction noise, as described by the pixel variances in the ROI. Our previous work [2] has shown that a tiny change in the system matrix elements might lead to a large improvement of the variance in the ROI when this change causes the rank of the system matrix to drop. Those results were in line with a previous theorem about the the impact of new data on the reconstruction quality [3]. In this paper, we derive a theorem specifying how changes in the system matrix can increase the size of the ROI, and show that a trade-off must be made between the size of the ROI and the reconstruction noise.

II. Materials and Methods

A. Theory

Let us consider the $m \times n$ system matrix S defined such that $rank(S) < n$. The linear system

$$y = Sx \tag{1}$$

This work was partially supported by grants ANR-17-CE19-0006 (ROIdoré project) and ANR-21-CE45-0026 (SPECT-Motion-eDCC) from the Agence Nationale de la Recherche (France).

F. Khellaf, and R. Clackdoyle are with Univ. Grenoble Alpes, CNRS, Grenoble INP, TIMC-IMAG, 38000 Grenoble, France.

with $x \in \mathbb{R}^n$ the image vector and $y \in \mathbb{R}^m$ the measurements vector, does not have a unique solution, although some components of x (some pixels) might be uniquely determined. Those elements correspond to the ROI pixels. In this model, S_{ij} represents the length of the intersection of line i with pixel j, assuming unit pixels. Note that there might be linear dependencies between the rows of S, i.e. the rank of S could be smaller than m. Now we consider the $m \times n$ matrix C defining the linear system

$$z = Cx \tag{2}$$

(for the same x as in equation (1)), such that S and C are equal except for one element. Consequently,

$$\text{rank}(S) - 1 \leq \text{rank}(C) \leq \text{rank}(S) + 1 \tag{3}$$

i.e. changing one element in S might increase the rank by one, decrease it by one, or not change the rank. We establish the following theorem

Theorem. *Let S be a $m \times n$ matrix. If there exists a $m \times n$ matrix C such that $C_{ij} = S_{ij}$ for all $(i,j) \neq (a,b)$ and $\text{rank}(C) = \text{rank}(S) + 1$, then (i) C can uniquely determine at least the b^{th} component of x, and (ii) S cannot uniquely determine the b^{th} component of x.*

Note that C might also have a non-trivial nullspace, (i.e. $\text{rank}(C) < n$), but x_b would nevertheless be solvable. Note also, that only one element of S is allowed to change. The example in the appendix shows that the theorem fails if the rank increases by one, but that S and C differ by more than one element.

Proof. (i) We first note that the change in rank means that S and C cannot be the same matrix, and therefore $S_{ab} \neq C_{ab}$. We will show that there exist scalars $\{\beta_i\}_{i=1,2...m}$ such that

$$x_b = \sum_{i=1}^{n} \beta_i z_i \tag{4}$$

and thus x_b, can be uniquely determined from (2).

The only difference between y and z is the a^{th} measurement, given by

$$y_a = \sum_{j=1}^{n} S_{aj} x_j \tag{5}$$

for y and

$$\begin{aligned} z_a &= \sum_{j=1}^{n} C_{aj} x_j \\ &= \sum_{\substack{j=1 \\ j\neq b}}^{n} S_{aj} x_j + C_{ab} x_b \end{aligned} \tag{6}$$

for z. Subtracting (6) from (5) gives

$$y_a - z_a = (S_{ab} - C_{ab}) x_b \tag{7}$$

On the other hand, since S and C are the same size, if $\mathrm{rank}(C) = \mathrm{rank}(S) + 1$, then there is a linear dependency between the rows of S. More specifically, if a change only in row a increases the rank, then that row is a linear combination of the other rows. (Recall that the rank of a matrix equals the dimension of the row-space.) Writing $S^{(i)}$ for the i^{th} row of S, we have, for some $m-1$ scalars $\{\alpha_i\}_{i=1,2\ldots m; i\neq a}$,

$$S^{(a)} = \sum_{\substack{i=1 \\ i\neq a}}^{m} \alpha_i S^{(i)} \tag{8}$$

and since $y_i = x \cdot S^{(i)}$ we see that

$$\begin{aligned} y_a = x \cdot S^{(a)} &= \sum_{\substack{i=1 \\ i\neq a}}^{m} \alpha_i x \cdot S^{(i)} = \sum_{\substack{i=1 \\ i\neq a}}^{m} \alpha_i x \cdot C^{(i)} \\ &= \sum_{\substack{i=1 \\ i\neq a}}^{m} \alpha_i z_i \end{aligned} \tag{9}$$

So, the difference between the two measurements is

$$y_a - z_a = \sum_{\substack{i=1 \\ i\neq a}}^{m} \alpha_i z_i - z_a \tag{10}$$

Combining (7) and (10) gives

$$x_b = \frac{1}{S_{ab} - C_{ab}} \Big(\sum_{\substack{i=1 \\ i\neq a}}^{m} \alpha_i z_i - z_a \Big) = \sum_{i=1}^{m} \beta_i z_i \tag{11}$$

where the β_i are defined by the right hand equality of (11) since $S_{ab} - C_{ab} \neq 0$. Therefore, the component x_b can be determined from the z_i measurements.

(ii) We will prove the contrapositive: if x_b can be found using (1) then changing S to C will not increase the rank.

We define $e_b \in \mathbb{R}^n$ to be all zeros, except the b^{th} component, where it is one. Thus $x \cdot e_b = x_b$. Now since x_b can be solved from (1), there must exist scalars $\{\gamma_i\}$ such that

$$\begin{aligned} x_b &= \sum_{i=1}^{m} \gamma_i y_i = \sum_{i=1}^{m} \gamma_i (x \cdot S^{(i)}) \\ &= x \cdot (\sum_{i=1}^{m} \gamma_i S^{(i)}) \end{aligned} \tag{12}$$

and since (12) holds for all $x \in \mathbb{R}^n$, we must have

$$e_b = \sum_{i=1}^{m} \gamma_i S^{(i)} \tag{13}$$

which equivalently says that row-reducing the matrix S will result in one of the rows becoming e_b.

We now observe that $C^{(a)}$ is a linear combination of $S^{(a)}$ and e_b, and note that e_b is a linear combination of the rows of S (see (13)). Every row of C can therefore be obtained from a linear combination of the rows of S, so the rank of C cannot exceed the rank of S (the dimension of the row-space of C cannot exceed the dimension of the row-space of S).

We have shown that, if x_b is solvable from S (and S and C are the same matrix except for element C_{ab}) then C cannot have higher rank than S. Equivalently, if $\mathrm{rank}(C) = \mathrm{rank}(S) + 1$, then x_b cannot be uniquely determined from S. □

Corollary. *If C is such that $C_{ij} = S_{ij}$ for $(i,j) \neq (a_k, b_k)$ for $k \in \{1 \ldots K\}$ and $\mathrm{rank}(C) = \mathrm{rank}(S) + K$, then C can uniquely determine at least the K b_k components of x, and these K components were not uniquely solvable using S. Furthermore the K matrix elements that changed all lie on different rows and different columns ($a_k \neq a_l$ and $b_k \neq b_l$ for $k, l \in \{1 \ldots K\}$).*

Proof. For the rank to increase by K, at least K rows of the matrix must have changed. But with only K changes of the matrix, each changed element must have been on a different row. Similarly, since the rank is also equal to the dimension of the column space, the same argument applies to the columns, and each changed element is in a different column. The corollary is now obtained by successively applying the theorem. We recall that changing one element of a matrix can cause the rank to change by one at most, i.e. if $K = 1$, $\mathrm{rank}(C) \leq \mathrm{rank}(S) + 1$. Therefore, as K changes in the matrix cause $\mathrm{rank}(C) = \mathrm{rank}(S) + K$, each changed element (a_k, b_k) must have increased the rank by 1, and, according to the theorem, lets us determine the b_k component of x, and b_k was not solvable before the change. □

Remark. *It is easily seen that if S can uniquely determine some components of x, then C can also determine those same components, as well as the additional component(s). In short, under the hypotheses of the theorem, matrix C can always determine more components than S can.*

In the following sections, we illustrate the theorem and the corollary with several examples.

B. Small-scale examples

We define the following 8×9 system matrix

$$S = \begin{pmatrix} 1 & 1 & 1 & 0 & 0 & 0 & 0 & 0 & 0 \\ 1 & 0 & 0 & 1 & 0 & 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 & 0 & 1 & 0 & 0 & 1 \\ 0 & A & 0 & A & 0 & 0 & A & 0 & 0 \\ A & 0 & 0 & 0 & A & 0 & 0 & A & 0 \\ 0 & 0 & A & 0 & A & 0 & 0 & A & 0 \\ 0 & 0 & 0 & A & 0 & 0 & A & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & A & 0 & 0 & A \end{pmatrix} \tag{14}$$

with $A = \sqrt{5}/2$, which allows exact reconstruction of the first three elements of x (almost the same matrix was used in [2]).

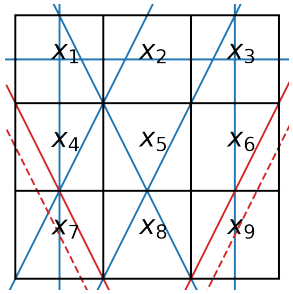


Fig. 1: Projection lines corresponding to the system matrix S. The red lines correspond to the 7^{th} and 8^{th} rows, and the dotted lines to the new measurements after changing one element in each row.

The 8 corresponding projection lines are shown in Fig. 1. The matrix S is rank deficient with $\text{rank}(S) = 6$. Now we define a matrix C such that

$$\begin{cases} C_{ij} = A/2 & \text{if } (i,j) = (8,6) \\ C_{ij} = S_{ij} & \text{otherwise.} \end{cases} \tag{15}$$

and $\text{rank}(C) = 7$. This change corresponds to a downwards half pixel translation of the eighth projection line (see Fig. 1).

We took a $n \times 1$ image vector x with values between 0 and 10, and simulated noiseless measurements according to equations (1) and (2). The values in the 3×3 image were

$$x = \begin{pmatrix} 5.75 & 7.43 & 0.50 \\ 9.88 & 3.57 & 8.79 \\ 0.37 & 1.24 & 6.32 \end{pmatrix}. \tag{16}$$

The estimated image $\hat{x}$ was reconstructed by applying the pseudo-inverse of the system matrices. The absolute bias

$$|\hat{x}_j - x_j| \tag{17}$$

for each reconstruction was computed in order to establish which pixels x_j were correctly determined, a bias of zero corresponding to a correct reconstruction. The results, shown in Table I confirmed that, as expected, pixels $\{x_1, x_2, x_3\}$ can be reconstructed from both S and C. Also, the matrix C uniquely determines pixel x_6 as predicted by the theorem, given that the only different element between S and C was $(8,6)$. In addition, pixel x_9 was fortuitously reconstructed exactly, since the eighth projection line only crossed pixels x_6 and x_9.

In order to evaluate the variance of the estimated pixels, the measurements were repeated 10^5 times with additional Gaussian noise of variance $\sigma^2 = 1$. In Table I, it is observed that the variance of pixels x_1 to x_3, which are the only ones correctly reconstructed by both S and C, increased when C was used for reconstruction. The increase was by as much as 50% for pixel x_3. The variances of the new reconstructible pixels, x_6 and x_9, were larger with C, but their previous variances were not relevant since they were not accurately reconstructible anyway (the pseudo-inversion produced a large bias).

In order to illustrate the corollary, we performed the same experiment with C defined as

$$\begin{cases} C_{ij} = A/2 & \text{if } (i,j) = \{(7,4),(8,6)\} \\ C_{ij} = S_{ij} & \text{otherwise.} \end{cases} \tag{18}$$

TABLE I: Bias and variance of the estimated solution using either S or C.

	S		C (Eq. 15)		C (Eq. 18)	
	Bias	Var.	Bias	Var.	Bias	Var.
Pixel 1	$< 10^{-14}$	0.47	$< 10^{-15}$	0.47	$< 10^{-14}$	0.49
Pixel 2	$< 10^{-14}$	0.80	$< 10^{-14}$	0.89	$< 10^{-14}$	1.08
Pixel 3	$< 10^{-15}$	0.62	$< 10^{-14}$	0.93	$< 10^{-14}$	1.02
Pixel 4	4.75	N/R	4.75	N/R	$< 10^{-14}$	6.45
Pixel 5	1.16	N/R	1.16	N/R	1.16	N/R
Pixel 6	1.23	N/R	$< 10^{-16}$	10.9	$< 10^{-13}$	11.35
Pixel 7	4.75	N/R	4.75	N/R	$< 10^{-13}$	3.99
Pixel 8	1.16	N/R	1.16	N/R	1.16	N/R
Pixel 9	1.23	N/R	$< 10^{-16}$	5.14	$< 10^{-16}$	5.25

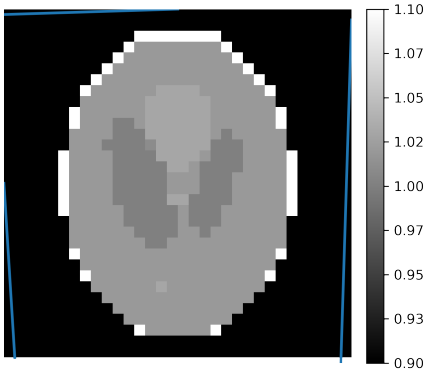


Fig. 2: Shepp-Logan phantom. The three blue lines correspond to the projection lines in S that were changed.

resulting in $\text{rank}(C) = 8$ (see Fig. 1).

According to the corollary, the matrix C should be able to determine pixels x_4 and x_6, which was confirmed by the zero bias for these two pixels in Table I. Two additional pixels, x_7 and x_9, were also accurately reconstructed. On the other hand, the variances for pixels $\{x_1, x_2, x_3\}$ have increased even more (by as much as 65% for pixel x_3).

C. Medium scale example

We performed the same experiment but for a larger 32×32 image of the Shepp Logan phantom (Fig. 2). We built a 396×1024 system matrix S of rank 369 which was able to accurately reconstruct a 16×16 ROI in the upper left corner of the image. Similar to the small-scale example, we built the matrix C by changing K elements (one per row). We chose $K = 3$, such that $\text{rank}(C) = 369 + 3 = 372$. Specifically, the changes occurred in columns 257, 272, and 513. The three projection line that were changed in S are shown in Fig. 2.

The bias maps between the reconstructions (without simulated noise) using either S or C and the reference phantom are shown in Figure 3. The pixels in the upper left corner of the reconstructed image using S appear in blue, corresponding to a very small bias, hence they are accurately reconstructed. As expected, the 3 pixels corresponding to the changed columns 257, 272, and 513, were also accurately reconstructed with matrix C. Additionally, figure 3 shows that pixel 514 (to the right of pixel 513) was also accurately reconstruced using the matrix C.

The differences between the variances of the pixels in the upper left 16×16 ROI reconstructed using either S or C are

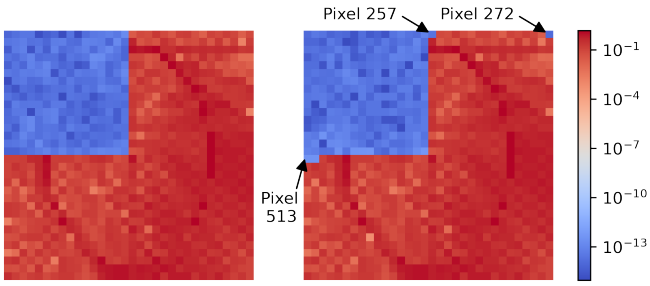


Fig. 3: Absolute bias between the reconstructed image using S (left) or C (right) and the reference Shepp Logan phantom. Note the logarithmic scale.

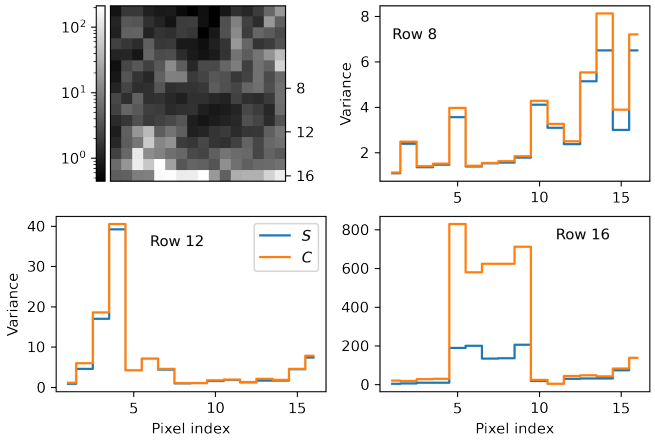


Fig. 4: Top left: variance in the 16×16 ROI reconstructed with S. Top right, bottom left, bottom right: profiles through the 8^{th}, 12^{th}, and 16^{th} row of the variance map using either S or C.

shown in Fig. 4. For all pixels in the ROI, the variance was larger when C was used to reconstruct the image.

III. Discussion

Our results have shown that, while increasing the rank of a matrix by changing its elements might increase the number of reconstructible pixels, it can also lead to a larger variance in the initial ROI. Therefore, a trade-off must be made between the size of the ROI and the reconstruction noise. For an $m \times n$ matrix, increasing the rank means increasing the number of independent equations and potentially solving a new variable. On the other hand, a rank deficient matrix contains redundant information, i.e. a possible smaller variance for the variables that can be solved.

The proposed theorem specifies when changes in the matrix will solve new variables, but is limited to one change per row. In practice, only projection lines close to the edge of the image can be modified since only one value in the system matrix is allowed to change. In our examples, the K new reconstructible pixels predicted by the theorem were all at the edges of the image, although other pixels can also be reconstructed. Nevertheless, increasing the rank of a matrix (while keeping its size constant) will increase the chance of solving new variables, while reducing the number of redundant measurements that might have been used to reconstruct ROI pixels.

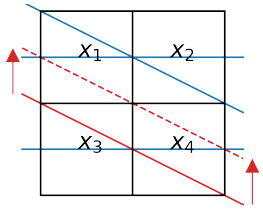


Fig. 5: Projection lines corresponding to the 4×4 system matrices. The full red line represents the last row of S and the dotted line the last row of C.

IV. Conclusion

A theorem specifying when a change in a tomographic system matrix can help solve new pixels has been proven. The study of the bias and the variance of the reconstructed images has shown that the additional reconstructible pixels might come at the cost of a reduced variance in the original ROI.

V. Appendix

Here we show that the hypothesis of the theorem, that only one element must differ between S and C, cannot be relaxed for the conclusions of the theorem to hold. We give an example of matrices S and C of size 4×4 such that $\text{rank}(S) = 2$ and $\text{rank}(C) = 3$ but no component of x can be uniquely reconstructed using either S or C. Two of the matrix elements of C differ from those of S. This example illustrates that, to apply the theorem, only one element of the matrix is allowed to change, when the rank increases by 1.

The example corresponds to a very simple tomographic system with x representing a 2×2 image. Using the same value of $A = \sqrt{5}/2$, we define S as

$$S = \begin{pmatrix} 1 & 1 & 0 & 0 \\ 0 & 0 & 1 & 1 \\ A & A & 0 & 0 \\ 0 & 0 & A & A \end{pmatrix} \tag{19}$$

and for the matrix C the last row is replaced by $(A\ 0\ 0\ A)$. The system is shown in Figure 5.

It is easy to see that $\text{rank}(S) = 2$ and $\text{rank}(C) = 3$ and that no component of x can be determined using matrix S. For the matrix C, the one-dimensional nullspace is spanned by the vector $(1, -1, 1, -1)$ and therefore no component of x can be uniquely resolved when using the matrix C.

References

[1] R. Clackdoyle and M. Defrise, Tomographic reconstruction in the 21st century, *IEEE Signal Processing Magazine*, vol. 27, no. 4, pp. 60-80, Jul. 2010.

[2] F. Khellaf, R. Clackdoyle, S. Rit and L. Desbat, "Tiny changes in tomographic system matrices can cause large changes in reconstruction quality", *Phys. Med. Biol.*, 2022.

[3] M. Defrise, R. Clackdoyle, L. Desbat, F. Noo, and J. Nuyts, "Do additional data improve region-of-interest reconstruction?" *The 16th International Meeting on Fully Three-Dimensional Image Reconstruction in Radiology and Nuclear Medicine*, Leuven, Belgium. July 19-23, 2021. Pages: 38-43.

Stationary X-ray Tomography for Hemorrhagic Stroke Imaging - Sampling and Resolution Properties

A. Lopez-Montes, T. McSkimming, W. Zbijewski, J. H. Siewerdsen, A. Skeats, B. Gonzales, A. Sisniega

Abstract— Hemorrhagic stroke accounts for up to 20% of all stroke cases, and requires a treatment pathway drastically different to ischemic stroke. Prompt triage is therefore crucial and often only attainable with neuroimaging for intracranial hemorrhage (ICH) evaluation, for which MDCT is the frontline modality. Availability of ICH dedicated imaging in the pre-hospital setting, with portable CT systems, would facilitate early ICH diagnosis. However, current CT or cone-beam CT (CBCT) approaches often use conventional x-ray sources mounted on a rotating gantry limiting their minimum weight and footprint. Recent advances on cold-cathode, compact x-ray sources, based on carbon nanotube (CNT) technology, enable the development of ultra-compact designs based on source-array arrangements on stationary configurations. However, such geometrical arrangements show limited angular sampling, and sparse, non-stationary, volume sampling.

In this work we present first investigation of geometric configuration and effects of 3D sampling pertinent to the task of ICH detection on an ultra-portable stationary CBCT for ICH imaging. The baseline configuration included 31 CNT sources on a curved array illuminating a curved panel detector (871 mm length), on a compact geometrical configuration (SDD = 690 mm). Metrics of sampling completeness, sampling density, and MTF shape and band-width integral were explored for configurations varying in source angular span (30°-170°), source array and detector curvature radius (250 mm to flat), use of 2D matrix source arrangements, and multi-acquisition protocols. The results show that sufficient sampling and resolution can be achieved with a combination of moderate curvature (~450 mm radius) of the source array and detectors, with better sampling properties for approximately matched curvature radii (up to 30% BWI improvement). Improved image quality was demonstrated with configurations featuring matrix source arrangements in combination with multi-acquisition protocols (around a 6% of improvement in sampling completeness).

***Index Terms*—Brain imaging, CBCT, Tomosynthesis.**

I. Introduction

Intracranial hemorrhage (ICH) presents a medical emergency often caused by head trauma or hemorrhagic stroke [1]. ICH accounts for up to 20% of all strokes [2], affecting 2 million people/year [3]. Promising therapeutic options have opened for treatment of acute ICH, involving aggressive supportive care and reversal of coagulopathy [4], with greater benefit for immediate treatment start, following the concept of "time is brain" [4]. Effective triage between hemorrhagic and ischemic stroke, based on neuroimaging, should be done as early as possible, preferably in the ambulatory or emergent, pre-hospital setup. To meet this need for pre-hospital imaging, several approaches to portable systems have been proposed in previous work, often based on cone beam CT (CBCT) designs some suitable to installation on mobile units [5], or the emergency room [6]. Such designs were often based on conventional CT or CBCT designs with enhanced portability, but required a rotating gantry system and used conventional x-ray sources that limited their portability and footprint. Recent advances in cold-cathode x-ray sources, such as those based on carbon-nanotube technologies (CNT), yielded improved stability and durability, and allowed the design of compact, light-weight sources that can be combined into linear or matrix arrangements [7]. Such arrangements can be integrated into stationary systems with extremely compact acquisition geometry and simple mechanical design, with recent examples in baggage inspection [7] and dental imaging [8]. Further reduction in system footprint could be attained in combination with novel curved detector designs. However, the capability of such systems to provide sufficient image quality for ICH visualization remains an open question, largely attributable to limited angular sampling; sparse, non-stationary, volume sampling; and effects of scatter in such compact configurations.

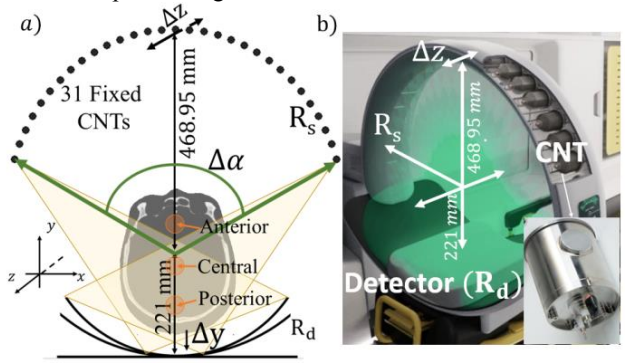


Fig. 1. a) Schematic representation of the stationary scanner configuration, featuring a curved 1D array (with extension to 2D matrix configurations) of CNT sources placed in opposition to a static flat-panel detector with variable curvature. The scanner configuration was optimized as a function of the variable parameters listed in the schematics ($\Delta\alpha$, RD, Rs, Δz, Δy). b) Artistic depiction of an example configuration of the stationary head scanner, featured as installed on an ambulance or mobile stroke unit. The CNT sources (see detail) are arranged in a curved array opposed to a curved flat-panel detector, following the general configuration in a).

In this work we present first investigation of geometric configuration and effects of 3D sampling pertinent to the task of ICH detection on an ultra-portable stationary CBCT for ICH imaging, designed for installation in mobile setups (Fig. 1B). Sampling, artifacts, and spatially-varying, directional, image resolution were characterized for a range of system configurations. Alternative linear and matrix source arrays and multi-exposure acquisition protocols were evaluated in

This work was supported by academic-industry partnership between the Australian Stroke Alliance, Micro-X Ltd., and Johns Hopkins University.

A. Lopez-Montes, T. McSkimming, W. Zbijewski, J. H. Siewerdsen, and A. Sisniega are with the Department of Biomedical Engineering, Johns Hopkins University. T. McSkimming, A. Skeats, and B. Gonzales are with Micro-X Ltd.

simulation studies.

II. Methods

A. Simulation setup for system optimization.

The configuration of the stationary head CBCT is illustrated in Fig.1A. The design included 31 CNT x-ray sources arranged along a curved array, at equiangular increments. The source arrangement was placed in direct opposition to a curved flat-panel detector of 837 mm size in the lateral direction (x) and 421 mm in the superior-inferior (SI) direction (z), with 1395 x 702 pixels (0.6 mm x 0.6 mm pixel size). The center of the world coordinate system was placed at the center of the patient head. To allow sufficient space for positioning of the head on a head rest, the center was placed 221 mm away from the detector surface (analogous to the axis-to-detector distance – ADD – often invoked in conventional CBCT). The distance between the center of the coordinate system and the source array (viz., source-to-axis distance – SAD – in CBCT) was set to 469 mm, yielding a source-to-detector distance (SDD) of 690 mm, compatible with the stringent footprint requirements for on-board systems on mobile medical units.

Variable parameters in the optimization studies included: i) the curvature of the source array, denoted R_s, defined as the radius of the circumference arc followed by the array (Fig. 1A), and ranging from a linear array with $R_s = \infty$ to a compact curved array with $R_s = 250$ mm; ii) the detector curvature, denoted R_d (Fig. 1A), defined in an analogously to R_s, and ranging from a conventional flat-detector configuration ($R_d = \infty$) to a curved detector with $R_d = 250$ mm, while maintaining a constant length of 837 mm for the curved segment; and, iii) the angular span of the source array ($\Delta\alpha$, see Fig. 1A) varied between 30° and 165° for a nominal source curvature of $R_s = 469$ mm (resulting in an arrangement centered at the origin of coordinates).

Configurations featuring matrix source arrangements in the form of two parallel curved arrays were simulated by translating each source in consecutive pairs in opposite directions along the SI axis (z) by a distance Δz ranging from 0 mm to 40 mm. Even (odd) sources were shifted towards positive (negative) z values. Final configurations studied potential improvement from multi-acquisition protocols in which two datasets were obtained with a linear translation of the patient head in the antero-posterior (AP) direction (y) between 0 mm and $\Delta y = 20$ mm.

Optimization studies used a realistic simulator for primary signal [9], including polychromatic source spectra and energy-dependent material properties and detector response. Current studies were focused on pure sampling effects and did not include effects of x-ray scatter and system noise. The simulations used a head phantom imaged with high-resolution MDCT and featuring natural bone skull and sinuses structure and a homogeneous brain parenchyma with 0 HU attenuation. ICH was simulated with synthetic spherical bleed inserts with 60 HU contrast and 8 mm diameter, placed on a regular grid lattice with 16 mm spacing. The head was positioned with a 45° tilt along the z axis, aligning the central scan plane with a plane joining the posterior skull base with a frontal brain region. The x-ray source spectrum was simulated with the Spektr [10] TASMICS model for 100 kV (+2 mm Al, + 0.2 mm Cu added filtration). The detector model featured a flat-panel detector with a 250 mg/cm^2 CsI:Tl scintillator.

B. Model Based Iterative Reconstruction.

Image reconstruction used a Penalized Weight Least Squares (PWLS) algorithm that can accommodate non-conventional sampling patterns [11], such as those arising in the proposed stationary configurations. Reconstructions were obtained minimizing the following PWLS cost-function. [12]

$$\mu = \arg_{\mu \geq 0} \min \left|\left|\mathbf{A}\mu - l\right|\right|_w^2 + \beta R(\mu) \qquad (1)$$

where $\mathbf{A}$ denotes the projection operator, l are the raw line integrals, and w are stochastic weighting terms that minimize the contribution of noisy measurements. In this work we model w with the raw measurements, commonly used as a surrogate of measurement variance. To avoid sharp transitions in presence of truncation, w values adjacent to the detector lateral edge were tapered following a ERF function with $\sigma = 2$ mm[13]. $R(\mu)$ is an image roughness quadratic regularization term weighted by the scalar β, set to a nominal value of $\beta = 10^{-3}$. To induce consistent regularization across configurations, the value of β was scaled by a factor of 2 for multi-acquisition protocols to compensate for the double number of views included in such protocols. Reconstructions were obtained on a 390 x 696 x 696 voxels grid (0.5 mm isotropic size) for 1000 iterations.

C. Metrics of sampling completeness and image quality.

Imaging performance was assessed in terms of voxel sampling density, sampling completeness, and local spatial resolution properties. Local sampling density was estimated as the number of rays (i.e. source views) contributing to an individual voxel j, normalized by a nominal value of 360 for a full conventional CBCT scan, yielding $\rho_{samp}(j) = N_{view}(j)/360$.

Effects of limited angular sampling were evaluated by the local sampling completeness (ΔA), estimated as the angular span covered by the rays contributing to a voxel, normalized by the 180° span required for artifact-free tomographic reconstruction. Defining θ_0 and θ_f as the angle between the x axis and the first and last views contributing to voxel j, completeness was computed as $\Delta A(j) = |\theta_0(j) - \theta_f(j)|/180$.

While spatial resolution properties of the volume obtained with model-based iterative reconstruction methods are non-stationary and generally difficult to predict, approximations invoking local stationarity have been shown to provide accurate estimations of the local MTF in PWLS with locally smooth (quadratic) penalty designs. Local MTF was estimated using the model in [14], which for a voxel j inside a homogeneous region is given by:

$$MTF_j = \frac{\mathcal{F}\{ROI_j\{\mathbf{A^T W A} e_j\}\}}{\mathcal{F}\{ROI_j\{\mathbf{A^T W A} e_j + \mathbf{R} e_j\}\}} \qquad (2)$$

where $\mathbf{A^T}$ is the backward projection operator, $\mathbf{W}$ is the diagonal matrix containing the PWLS weighs, and $\mathbf{R}$ is the Hessian of the PWLS quadratic penalty. $\mathbf{B}$ is a system blurring factor, obtained as in [14], and e_j represents a Kronecker delta at voxel j. To evaluate the resolution properties across configurations we qualitatively assessed the shape of the resulting MTF, while quantitative assessment was obtained with a version of a modified version of the Band Width Integral (BWI) in [15]:

$$\text{BWI} = 100 \times \frac{\int_{\omega_0}^{\omega_f} MTF^2 d\omega}{\int_{\omega_0}^{\omega_f} d\omega^2} \qquad (3)$$

The integration limits were set to ω_0 = 0.2 mm^{-1} and ω_f = 0.6 mm^{-1} pertinent to visualization of mid-size hemorrhage modelled with Gaussian kernel with σ = 4 mm. Sampling and resolution metrics, were evaluated for three regions of interest (ROIs), at the anterior, central, and posterior regions of the brain (see Fig. 1A).

III. Results

A. Stationary Tomography Geometrical Configuration.

Fig. 2A illustrates voxel sampling completeness (ΔA) for the three ROIs as a function of the angular span ($\Delta\alpha$) of the source array, for a fixed R_s = 469, and three detector curvatures: i) a flat-panel configuration ($R_d = \infty$); ii) a moderate curvature (R_d = 650 mm); and iii) a curved, compact, detector (R_d = 450 mm). Large values of $\Delta\alpha$ resulted in significant the truncation of the anterior regions of the head for certain views, evidenced as a reduction of ΔA for $\Delta\alpha > 80°$ and $\Delta\alpha > 100°$ for $R_d = \infty$ in the anterior and central ROIs, respectively. Curved detectors offered improved sampling completeness for those regions, showing a plateau at $\Delta A \cong 0.7$ for $\Delta\alpha > 120°$, for the anterior ROI, compared to $\Delta A \cong 0.6$ with $R_d = \infty$, with slight better performance for R_d = 450 mm. A similar trend was observed for the central and posterior ROIs.

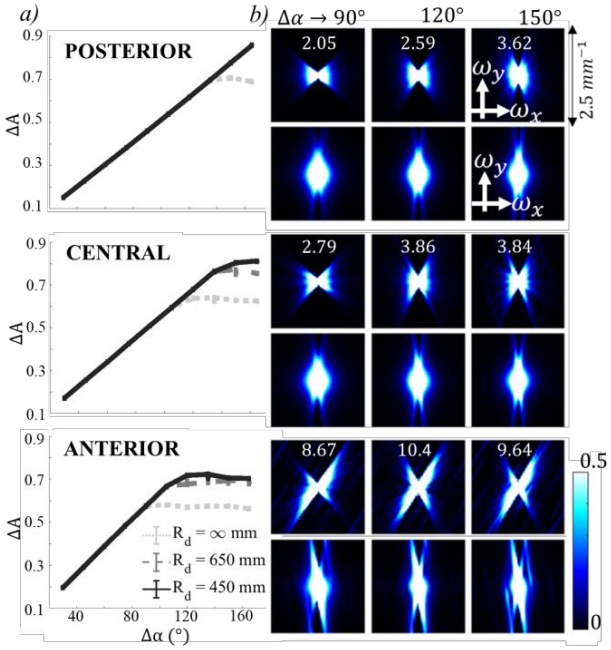


Fig. 2. ***a)*** Voxel sampling completeness as a function of sources angular span, for three detector curvatures, $R_d = \infty$, R_d = 450 mm, and R_d = 650 mm. Curves represent the mean value while error bars mark the 80% confidence interval within the ROI. ***b)*** Axial and coronal MTF for varying source angular span at a fixed detector curvature of R_d = 450 mm. Numerical values in the MTFs show the mid-band BWI.

MTF distributions in the axial and coronal planes are shown in Fig. 2B, for R_d = 450 mm, for the three ROIs, and for angular spans around the region where sampling completeness becomes flat in Fig. 2A: $\Delta\alpha$ of 90°, 120°, and 150°. As hinted by trends in ΔA, we observe a reduction of the MTF shade region between $\Delta\alpha$ = 90° and $\Delta\alpha$ = 120°, yielding an increase of 10%, 30%, and 25% in mid-band integrated MTF for the anterior, central, and posterior ROIs, respectively. The gains in directional resolution are less apparent for angular span $\Delta\alpha > 120°$, except for the posterior region, that sees increased completeness and a reduction of the MTF shade region. However, such gains come accompanied by the appearance of high-frequency structures in the frontal region, associated to streak artifacts from increased truncation of intricate frontal maxillofacial bone anatomy.

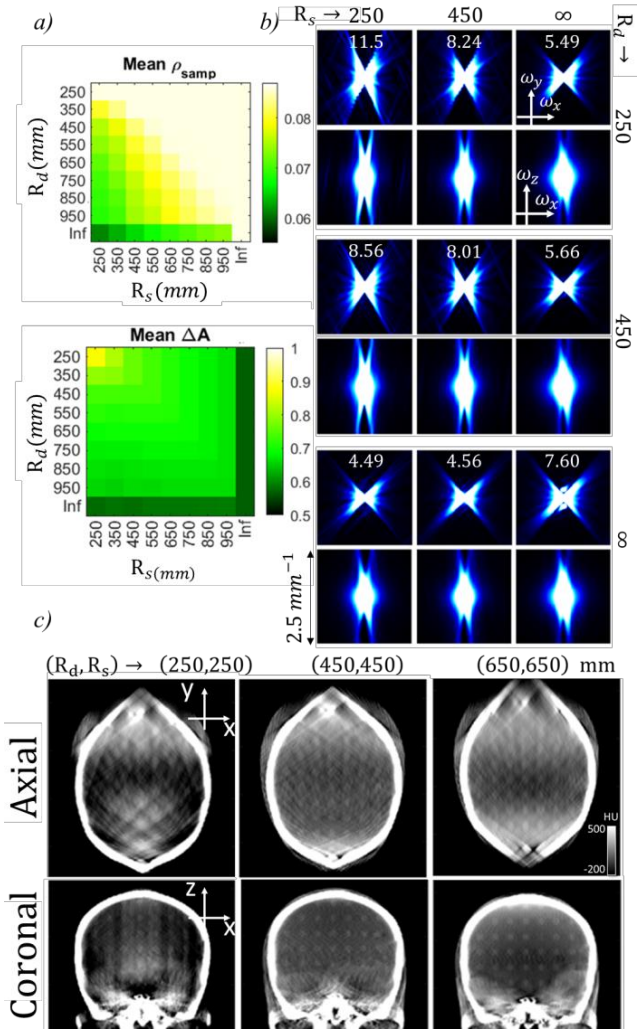


Fig. 3. *a)* Mean sampling density and angular span for the anterior ROI b) MTF as a function of sources and detector curvature for the anterior ROI. Numerical values in the MTFs correspond to the BWI along a mid-frequency band (0.2 mm^{-1} - 0.6 mm^{-1}). c) Reconstructed images for different combinations of sources and detector curvatures, for a very compact, curved geometry (left), a moderately curved geometry (middle), and a configuration with a linear array of sources and a flat-detector yielding a linear tomosynthesis configuration.

Fig. 3 illustrates the impact of varying the curvature of the detector and arc of sources for a configuration with an arc length of the compromise setup with $\Delta\alpha$ = 120° arising from results in Fig. 2. The range of curvature explored ranged for very compact arrangements ($R_s = R_d$ = 250 mm), to extended arrangements pertinent to linear tomosynthesis ($R_s = R_d = \infty$) The results in Fig. 3A suggest increased sampling density and completeness for very compact configurations, increased truncation of anterior bone regions result in high frequency streak patterns in the MTF that can be observed as increased high frequency artifacts and non-uniformity in Fig. 3C (top row). Consistent with monotonic reduction of sampling

completeness (Fig 3A) with increased R_s and R_d, Fig. 3B show a similar loss of frequency information for such combinations of large R_s and R_d. For example, using a linear tomosynthesis configuration leads to a reduction of ~10% mid-band MTF compared to a moderate configuration with $R_s = R_d = 450$ mm. This loss of information for very large radii resulted in limited angle artifacts usually observed in tomosynthesis imaging (see Fig. 3C). The results in Fig. 3 suggest that approximate match of the detector and source curvature is associated to better sampling properties, with mid-band MTF increases up to 25%, compared to configurations with large mismatch between detector and source curvatures.

B. Array vs matrix source configurations.

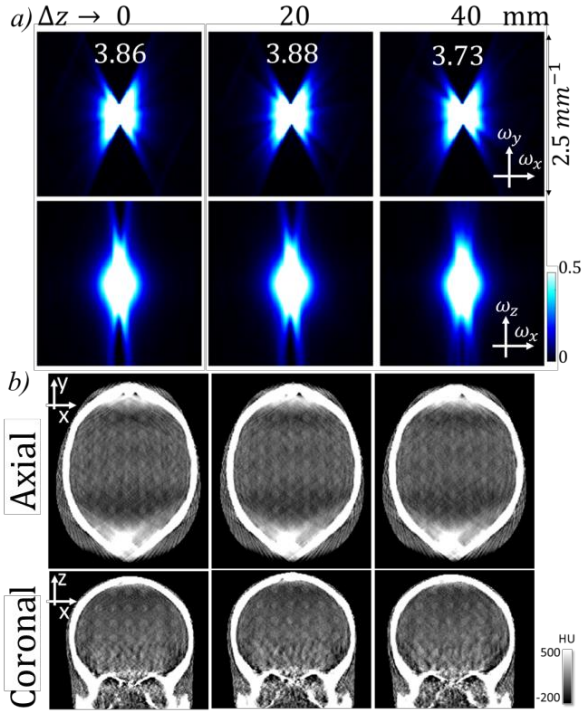


Fig. 4. Results of array ($\Delta z = 0$ mm) vs matrix source configurations with $\Delta z =$ 20 mm and $\Delta z = 40$ mm. ***a)*** MTF for the central region of the brain. Numerical values in the MTFs correspond to the BWI along a band pass filter between 0.2 mm^{-1} and 0.6 mm^{-1}. ***b)*** Reconstructed images for different distance between parallel arcs. Slight improvement in shading artifacts attributable to cone beam effects can be observed at frontal regions close and above the orbital bone for matrix arrangements and large separation. However, reduced shading was associated to reduced contrast for bleed inserts at central regions of the brain.

Figure 4 compares MTF and image results for 1-dimensional array and matrix configurations of the source arrangement. Matrix arrangements with an interleaved distribution of sources resulted in increased resolution in the SI direction, as evidenced by the reduction of the shade region in the coronal MTF and reduction of cone-beam artifacts above and around the orbital structures at the anterior region of the brain. However, such improvements were accompanied by a reduction in the overall contrast of the blood inserts, attributable to a reduction of local sampling density across the brain region.

C. Multi-acquisition protocol.

As illustrated in previous results, one limiting factor for image quality is the increased truncation of bone anatomy and reduced sampling density at anterior regions of the head for compact configurations, causing high-frequency streak-like artifacts in the reconstructed images. As shown in Fig. 5, the sampling completeness is increased by ~6% for the multi-acquisition protocol with $\Delta y > 12$ mm. This improvement leads to better visualization of blood inserts, particularly in the anterior region of the brain, as seen in Fig. 5B.

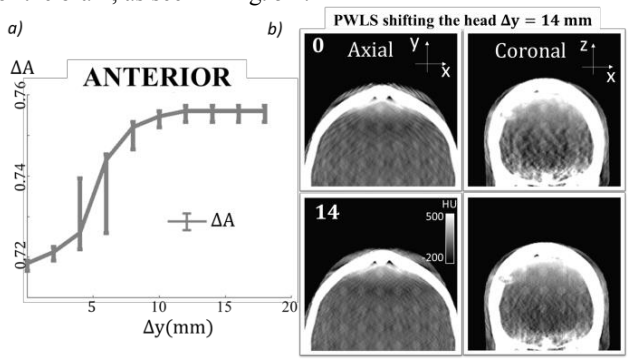


Fig. 5. Performance of the multi-acquisition protocol as a function of the head shift distance. ***a)*** Sampling completeness as a function of head translation. ***b)*** Anterior region of the reconstructed images for the single acquisition protocol (top) and for a multi-acquisition protocol with $\Delta y = 14\ mm$.

IV. Discussion and Conclusion

This work studied the tradeoffs in sampling and resolution for a comprehensive range of geometrical configurations in stationary tomographic systems for imaging of hemorrhagic stroke. The results show that sampling and resolution sufficient for the task of ICH detection can be achieved with a combination of curved array sources arrangements and curved panel detectors. Better sampling properties was observed for approximately matched curvature radii, as evidenced by increase in sampling completeness and density aggregated metrics as well as increases in radial homogeneity of the MTF distributions. Improved image quality can be achieved by combination of matrix source configurations and multi-acquisition protocols. The studies presented in this work did not consider the effects of x-ray scatter and quantum and instrumentation noise on image quality and focused solely on effects of sampling and frequency response. Scatter and noise are expected to pose a complex multi-variate scenario, in combination with sampling considerations, and are the subject of ongoing work.

References

[1] Ghajar, J. Lancet 356(9233), 923–929 (2000).
[2] Feigin, V. L et al. The Lancet Neurology, 8(4), 355-369 (2009).
[3] Krishnamurthi, et al. Lancet Glob. Heal. 1(5), e259–e281 (2013).
[4] Cordonnier, C. et al., Lancet 392(10154), 1257–1268 (2018).
[5] Ebinger, M. et al., JAMA311(16), 1622–1631 (2014).
[6] Wu, P. et al., Med. Phys. 0(0), 1–16 (2020).
[7] Gonzales, B. et al., IEEE Access 2, 971–981 (2014).
[8] Spronk, D. et al. Medical Imaging(Vol. 11312, p. 1131237) 115 (2020).
[9] Wu, P. et al. arXiv preprint arXiv:2108.09229. (2021).
[10] Punnoose, J. et al., *Medical Physics*, *43*(8), 4711–4717. (2016).
[11] Stayman, J. W. et al. J. Med. Imaging 6(02), 1 (2019).
[12] Sisniega, A. et al., PMB, 66(5). (2021).
[13] Zhang, Y. et al. *Journal of Computer Assisted Tomography*, *33*(3), 426–435. (2009).
[14] Wang W. et al. Med. Phys. 46 65–80. (2019).
[15] Wagner R.F. et al. Med. Phys. 6, 83 (1979);

Angular normalized glandular dose coefficient in breast CT: clinical data study

Hsin Wu Tseng, Andrew Karellas, and Srinivasan Vedantham
Department of Medical Imaging, The University of Arizona, Tucson, AZ 85724, USA

***Abstract*—The goal of this study is to understand how the normalized glandular dose coefficient (DgN^{CT}) varies with projection angle in dedicated cone-beam breast computed tomography (CBBCT). Seventy five CBBCT clinical datasets from a research database were used for this study. All samples were segmented into skin, adipose and fibroglandular tissues. The segmented volumes were used in a Monte Carlo simulation package (GATE 8.0) to estimate the radiation dose at 10 angles in a full scan. An analytical model is proposed and this model predicted that the angular DgN^{CT} follows a sine wave and the maximum is related to the center of geometry of the fibroglandular tissue (COG_f). The angular DgN^{CT} from Monte Carlo simulations was consistent with our model and follows a sine wave with amplitude of 0.0376. The maximum of the wave occurs when the x-ray source is approximately at head position, which is consistent with our model. Our results indicate that the higher angular DgN^{CT} occurs when the x-ray source is superior to the breast. This suggests using a x-ray source trajectory inferior to the breast for short-scan CBBCT design.**

***Index Terms*—breast CT, mean glandular dose, Monte Carlo, radiation dose**

I. Introduction

The mean glandular dose in cone-beam breast CT (CBBCT) can be estimated by the product of normalized glandular dose (DgN^{CT}) coefficient and the air kerma (AK) measured at the axis of rotation (AOR) without any object. Monte Carlo (MC) simulations are the most common method to compute the energy deposited in the fibroglandular tissues [1]–[8]. The DgN^{CT} depends on the x-ray spectrum and the breast model. The semi-ellipsodal breast model with a homogeneous distribution of fibroglandular tissue [7], [9], [10] and patient-specific breast model [6], [7] are the two common models used in CBBCT studies. The former model uses the effective chest-wall diameter (D_{eff}), the chest wall-to-nipple length (CNL), and fibroglandular fraction (f_g) of the breast to create a semi-ellipsoidal model, and every voxel, except the skin, has the same f_g. The second model segments each 3D reconstructed breast volume into skin, adipose, and fibroglandular tissues, in addition to air. The homogeneous semi-ellipsoidal model was found to overestimate DgN^{CT} because this model overestimates the amount of the fibroglandular tissue along the periphery of the breast. Since the DgN^{CT} homogeneous semi-ellipsoidal model can be described by a fitting function [11], here we focus on the patient-specific model in this study.

To our best knowledge, studies in literature only considered DgN^{CT}, but none have considered the variation of the DgN^{CT} with the projection angles, i.e., angular DgN^{CT}. We are particularly interested in DgN^{CT} because we have developed feasible image reconstruction algorithms for short-scan and sparse-view acquisitions [12], [13], and we would like to understand which angular range should be used for short-scan acquisition to reduce the radiation to the breast either for prone or upright patient-position CBBCT systems.

A cohort of 75 CBBCT datasets from a research database of subjects who had participated in a prior IRB-approved clinical trial was used in this study. A validated MC simulation code in our recent publication [10] and following the guideline of the Task Group No. 268 of the American Association of Physics in Medicine (AAPM)[14] was used here to compute the angular DgN^{CT}.

II. Materials and Methods

A. CBBCT system

The projections were acquired by a CBBCT system that is a Pre-FDA approval prototype (KBCT1000, Koning Corp., West Henrietta, NY). The operated x-ray tube was RAD71-SP (Varex Imaging, Salt Lake City, UT), and the x-ray was operated at 49 kV with a pulse-width of 8 milliseconds. 300 projections with uniform 1.2 deg/view angular sampling in a full scan (360 deg) were performed. A CsI:Tl scintillator coupled, amorphous silicon-based flat-panel detector (PaxScan 4030CB, Varex Imaging, Salt Lake City, UT) used used in the CBBCT system. The operating pixel size of this detector is 0.388 mm, and the dimension of the detector is 1024×768. The distance between the source and the AOR is 650 mm, and the source-to-detector distance is 898 mm.

B. Angular DgN^{CT} computation

The breast images were all first reconstructed by our developed deep learning-based algorithm, multi-slice residual dense network (MS-RDN) [13], that reduces image noise. Then all MS-RDN reconstructed images were segmented into air, skin, adipose, and fibroglandular tissues (Fig. 1) using a previously reported method [15]. The CNL, D_{eff}, and f_g can be estimated from segmented images for a semi-ellipsoidal

homogeneous breast model.

The DgN^{CT} (mGy/mGy) of the patient-specific breast model can be calculated as [5], [6]

$$DgN_{hete}^{CT} = \frac{E_{g,dep}}{n_g m_g \times AK(E)} \quad (1)$$

where the subscript, *hete* represents heterogeneous tissue distribution, $E_{g,dep}$ is the total energy deposited in all fibroglandular tissue voxels, n_g indicates the total number of fibroglandular tissue voxels, m_g is the mass of a fibroglandular tissue voxel, and $AK(E)$ is the air kerma at the breast center with the energy, E, of the incident photons. The same computational method was used for calculating the angular DgN^{CT}. In MC simulations, all photons were radiated to the breast from the assigned x-ray source position (angle). The MC simulations were performed using the MC code (GATE 8.0) validated in our previous study [10]. The number of photons was 10^6 as suggested by literature [7], [16], and resulted in a variation of less than 0.7%.

The angular DgN^{CT} of the homogeneous breast of a semi-elliptical shape is the same as its DgN^{CT} at any angle because of the rotational symmetry. The angular DgN^{CT} of this model can be expressed as a fitting function (standard deviation is 1.13%) [11]

$$DgN_{Fit}^{CT} = \left[1.0758247 - 0.2353669 \times \ln\left(D_{eff}\right) - 0.1253462 f_g\right] \times \left[0.1153 \times \ln\left(\frac{CNL}{D_{eff}}\right) + 1.0818\right] \quad (2)$$

C. Simplified Math Model

The angular DgN^{CT} in Eq. (1) is related to the energy deposited on the fibroglandular tissues, which is proportional to the pathlengths (PL) of photons through the fibroglandular tissues. To easily demonstrate this concept, a 2D circle of a finite size presents the fibroglandular tissues here. The center of the circle is the center of the geometry of the fibroglandular tissues ($\mathrm{COG_f}$). The PL can be analytically solved as

$$PL = 2\left\{\frac{(L + R\cos\beta)^2(1 + \tan\alpha\tan(\alpha+\theta))^2}{1 + tan^2(\alpha+\theta)} + r^2 - (L + R\cos\beta)^2(1 + tan^2\alpha)\right\}^{1/2} \quad (3)$$

where L, R, r, and β, α, and θ are the distance between the x-ray source and AOR, the radius of the orbit of the 2D circle, the radius of the circle, the polar angle of the circle in the orbit, the angle between the central line of the circle and the central line of the fan-beam, the angle deviated from the central line of the circular object, respectively. The figure is depicted in Fig. 2. Similar to the concept of the *Radon* transform, the curve of the PL is a sine wave varying with projection angle.

III. RESULTS

A. Total Simulation Time

All MC simulations were performed on a Dell workstation 7810 with Intel Xeon CPU (3.20 GHz) and 32 GB RAM. For each angle, the MC simulation took approximtely 40 minutes, resulting in 40×10×75=30000 minutes for all 75 samples.

B. Numerical results of the simplified math model

Numerical results of two particular examples of our simplified math model are when the $\mathrm{COG_f}$ is located at the center and above the AOR (i.e. β=0 in Eq. (3) and Fig. 2). Let the radius of the $\mathrm{COG_f}$ be 15 mm with (1) 0 mm (reference) and (2) 50 mm distance away from the AOR. Without losing any generality, $\theta = 1^\circ$ (or -1°) was considered here. The results (Fig. 3) show that the PL is a sine wave as predicted and in this particular example, the minimum of the curve occurs when the x-ray source is at ϕ=180 deg (feet position). The average PL of the sine wave is 0.1296 *mm* less than the reference. The sine wave has the same PL as that of the reference at ϕ=92.2042 deg and 267.7958 deg, which can be analytically solved by Eq. (3).

C. Center of the geometry of fibroglandular tissues

In MC simulations, the center of the geometry of the entire breast $\mathrm{COG_b}$ is at the AOR. The deviation of $\mathrm{COG_f}$ from $\mathrm{COG_b}$ for each sample was shown in Fig. 4. The breast laterality is factored with 90 deg representing medial and 270 deg representing lateral aspects of the breast. It was found that 62.67% of samples have $\mathrm{COG_f}$ between 36 deg and 324 deg. From the previous section, the numerical results show that the minimum of the PL curve would happen at 180 deg if the $\mathrm{COG_f}$ is exactly at 0 deg. Thus, the angular DgN^{CT} in this particular dataset should be a sine wave with a minimum at approximately within the range of 144 to 216 deg.

D. AngularDgN^{CT}

For each breast volume, the angular DgN^{CT} was normalized by the DgN^{CT} from the homogeneous semi-ellipsoidal model (reference) to characterize its deviation. The average of this normalized angular DgN^{CT} from the 75 breast CT volumes is shown in Fig. 5. Consistent with the prediction from section III. B and III. C above, the curve of the angular DgN^{CT} is approximately a sine wave with a minimum between 144 deg and 216 deg. The curve can be fitted to sine wave of the form $0.0376\sin(\emptyset + 83^\circ)$, with a root mean square error of 0.0106.

IV. DISCUSSION AND CONCLUSIONS

In this study, we have investigated the variation in DgN^{CT} with x-ray projection angle for real breasts using both numerical study of PL calculation and using MC simulations. Although the PL calculation is based on the 2D geometry, it provides us an easy way to understand how the DgN^{CT} changes when a real 3D breast is scanned in the CBBCT system. The MC results were consistent with our simplified math model and $\mathrm{COG_f}$ analysis. As predicted by our theory, there is higher energy deposition in most of the patients' breasts when the x-ray source is approximately superior to the breast, i.e., between the

shoulders and above the breast. Thus, to reduce the radiation dose to the patients in short-scan CBBCT acquisition, it is preferable to avoid acquiring projections superior to the breast. This implies an x-ray source trajectory that is inferior to the breast. Further, this design would allow the detector to traverse below the chin, which could avoid the neck strain reported in a prior study [17] due to the need to turn the patient's head to accommodate the x-ray source trajectory. Development of such an upright CBBCT system is in progress. The determination of the angular sampling interval for the short-scan acquisition should depend on the design and the implementation of the image reconstruction algorithms.

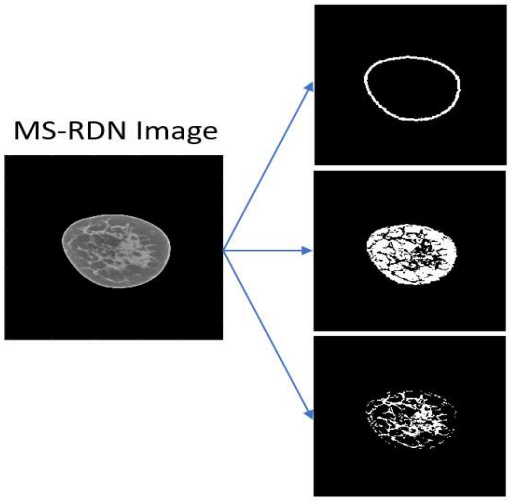


Fig. 1. Segmentation of the image. The images were reconstructed by MS-RDN deep learning algorithm. Each image was then segmented into air, skin, adipose, and fibroglandular tissues.

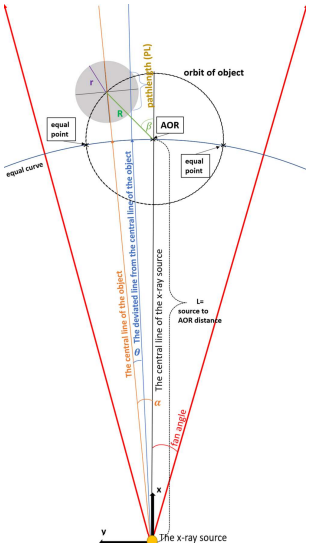


Fig. 2. Simplified mathematical model. A 2D circular object in a fan-beam geometry.

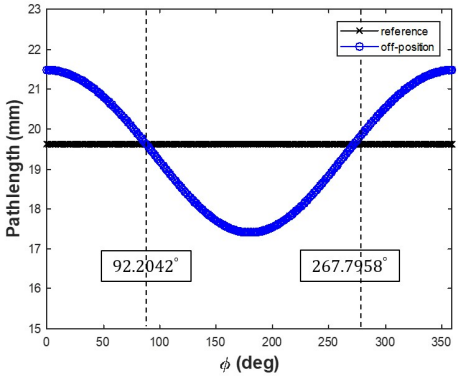


Fig. 3. The pathlength of the object off positioned above from the AOR. ϕ is the angle of the x-ray source.

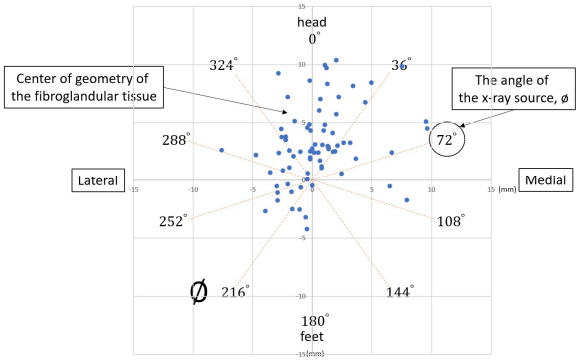


Fig. 4. Center of the geometry of the fibroglandular tissues (COG_f) deviated from the center of the geometry of the entire breast (COG_b). ϕ is the angle of the x-ray source.

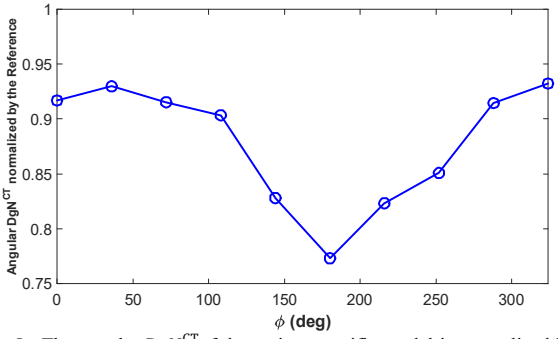


Fig. 5. The angular DgN^{CT} of the patient-specific model is normalized by the reference (homogeneous breast of semi-elliptical shape).

ACKNOWLEDGMENT

This work was supported in part by the National Cancer Institute (NCI) of the National Institute of Health (NIH) grant R01 CA 199044 and R01 CA 241709. The contents are solely the responsibility of the authors and do not represent the official views of the NCI or the NIH.

REFERENCES

[1] J. M. Boone, "Glandular breast dose for monoenergetic and high-energy X-ray beams: Monte Carlo assessment," *Radiology*, vol. 213, no. 1, pp. 23–37, Oct. 1999.

[2] J. M. Boone, N. Shah, and T. R. Nelson, "A comprehensive analysis of DgN(CT) coefficients for pendant-geometry cone-beam breast computed tomography," *Med Phys*, vol. 31, no. 2, pp. 226–235, Feb. 2004.

[3] J. M. Boone, "Normalized glandular dose (DgN) coefficients for arbitrary X-ray spectra in mammography: computer-fit values of Monte Carlo derived data," *Med Phys*, vol. 29, no. 5, pp. 869–875, May 2002.

[4] D. R. Dance, "Monte Carlo calculation of conversion factors for the estimation of mean glandular breast dose," *Phys Med Biol*, vol. 35, no. 9, pp. 1211–1219, Sep. 1990.

[5] A. M. Hernandez, A. E. Becker, and J. M. Boone, "Updated breast CT dose coefficients (DgNCT) using patient-derived breast shapes and heterogeneous fibroglandular distributions," *Med Phys*, vol. 46, no. 3, pp. 1455–1466, Mar. 2019.

[6] A. Sarno, D. R. Dance, R. E. van Engen, K. C. Young, P. Russo, F. Di Lillo, G. Mettivier, K. Bliznakova, B. Fei, and I. Sechopoulos, "A Monte Carlo model for mean glandular dose evaluation in spot compression mammography," *Med Phys*, vol. 44, no. 7, pp. 3848–3860, Jul. 2017.

[7] I. Sechopoulos, S. S. J. Feng, and C. J. D'Orsi, "Dosimetric characterization of a dedicated breast computed tomography clinical prototype," *Med Phys*, vol. 37, no. 8, pp. 4110–4120, Aug. 2010.

[8] I. Sechopoulos, J. M. Sabol, J. Berglund, W. E. Bolch, L. Brateman, E. Christodoulou, M. Flynn, W. Geiser, M. Goodsitt, A. K. Jones, J. Y. Lo, A. D. A. Maidment, K. Nishino, A. Nosratieh, B. Ren, W. P. Segars, and M. Von Tiedemann, "Radiation dosimetry in digital breast tomosynthesis: report of AAPM Tomosynthesis Subcommittee Task Group 223," *Med Phys*, vol. 41, no. 9, p. 091501, Sep. 2014.

[9] S. C. Thacker and S. J. Glick, "Normalized glandular dose (DgN) coefficients for flat-panel CT breast imaging," *Phys Med Biol*, vol. 49, no. 24, pp. 5433–5444, Dec. 2004.

[10] H. W. Tseng, A. Karellas, and S. Vedantham, "Radiation dosimetry of a clinical prototype dedicated cone-beam breast CT system with offset detector," *Med Phys*, vol. 48, no. 3, pp. 1079–1088, Mar. 2021.

[11] S. Vedantham, L. Shi, A. Karellas, A. M. O'Connell, and D. L. Conover, "Personalized estimates of radiation dose from dedicated breast CT in a diagnostic population and comparison with diagnostic mammography," *Phys. Med. Biol.*, vol. 58, no. 22, pp. 7921–7936, Nov. 2013.

[12] H. W. Tseng, A. Karellas, and S. Vedantham, "Sparse-view, short-scan, dedicated cone-beam breast computed tomography: image quality assessment," *Biomed Phys Eng Express*, Sep. 2020.

[13] Z. Fu, H. W. Tseng, S. Vedantham, A. Karellas, and A. Bilgin, "A residual dense network assisted sparse view reconstruction for breast computed tomography," *Sci Rep*, vol. 10, p. 21111, Dec. 2020.

[14] I. Sechopoulos, D. W. O. Rogers, M. Bazalova-Carter, W. E. Bolch, E. C. Heath, M. F. McNitt-Gray, J. Sempau, and J. F. Williamson, "RECORDS: improved Reporting of montE CarlO RaDiation transport Studies: Report of the AAPM Research Committee Task Group 268," *Med Phys*, vol. 45, no. 1, pp. e1–e5, Jan. 2018.

[15] S. Vedantham, L. Shi, A. Karellas, and A. M. O'Connell, "Dedicated breast CT: Fibroglandular volume measurements in a diagnostic population," *Med Phys*, vol. 39, no. 12, pp. 7317–7328, Dec. 2012.

[16] S. Vedantham, L. Shi, A. Karellas, and F. Noo, "Dedicated breast CT: radiation dose for circle-plus-line trajectory," *Med Phys*, vol. 39, no. 3, pp. 1530–1541, Mar. 2012.

[17] A. O'Connell, D. L. Conover, Y. Zhang, P. Seifert, W. Logan-Young, C.-F. L. Lin, L. Sahler, and R. Ning, "Cone-Beam CT for Breast Imaging: Radiation Dose, Breast Coverage, and Image Quality," *American Journal of Roentgenology*, vol. 195, no. 2, pp. 496–509, Aug. 2010.

Estimating the accuracy and precision of quantitative imaging biomarkers as endpoints for clinical trials using standard-of-care CT

Paul Kinahan[1], Darrin Byrd[1], Hao Yang[2], Hugo Aerts[3], Binzhang Zhao[2], Andrey Fedorov[3], Lawrence Schwartz[2], Tavis Allison[2], Chaya Moskowitz[4]

Abstract - Quantitative imaging biomarkers (QIBs) hold enormous potential to improve the efficiency of clinical trials that use standard-of-care CT imaging. Examples of QIBs include size, shape, intensity histogram characteristics, texture, radiomics, and more. There is, however, a well-recognized gap between discovery and the translation to practice of QIBs, which is driven in part by concerns about their repeatability and reproducibility in the diverse clinical environment. Our goal is to characterize QIB repeatability and reproducibility by using virtual imaging clinical trials (VICTs) to simulate the full data pathway. We start by estimating the probability distribution functions (PDFs) for patient-, disease-, treatment-, and imaging-related sources of variability. These are used to forward-model sinograms that are reconstructed and then analyzed by the QIB under evaluation in a virtual imaging pipeline. By repeatedly sampling from the variability PDFs, estimates of the bias, variance, repeatability and reproducibility of the QIB can be generated by comparison with the known ground truth. These estimates of QIB performance can be used as evidence of the utility of QIBs in clinical trials of new therapies.

Keywords: Imaging biomarkers, quantitative imaging, repeatability and reproducibility, clinical trials

I. Introduction

Clinical trials are a cornerstone of developing more effective cancer therapies. However, traditional clinical trials are often slow, expensive, and inefficient. Imaging of disease with standard-of-care CT plays a pivotal role in the management of patients with cancer and is used to measure endpoints in cancer drug trials to quantify efficacy in candidate compounds. There is a tremendous potential for quantitative imaging biomarkers (QIBs) to make clinical trials more efficient and informative. Examples of QIBs include size, shape, intensity histogram characteristics, and texture. Taking advantage of this potential is imperative since in the era of targeted therapies, studies will be smaller, more fractionated, with more expensive therapies. There is, however, a well-recognized gap between discovery and translation to practice for biomarkers in general and specifically for quantitative imaging biomarkers used in clinical trials. This gap arises for reasons that have been described including, among other items, a lack of data for testing and validation, a lack of rigor in the experimental design, inconsistent algorithm implementation, incomplete reporting, and a lack of appreciation for the requirements for adoption of quantitative imaging biomarkers. Addressing the lack of knowledge about the a priori distributions of random effects in imaging scenarios that should be evaluated, providing a rigorous methodology for evaluation, and ensuring pathways for adoption for all stakeholders can overcome these barriers.

To do so we propose to build a measurement error model by using virtual imaging clinical trials (VICTs) [1] to simulate the entire data pathway from patient models through image generation to QIBs. As a first step in this process, virtual imaging clinical trials (VICTs) are an emerging methodological adjunct to clinical trials using imaging. A VICT is essentially an extension to a clinical trial simulation in that the population of human subjects is replaced with a population of virtual digital subjects; imaging systems are replaced with physics-based virtual imaging simulators; and clinical interpretations are replaced with AI-derived image analyses. A VICT offers a feasible and efficient means to conduct experimentation in medical imaging by providing the practical ability to systematically assess and optimize a host of trial design factors and imaging parameters in the development and evaluation of imaging technologies, a task not possible through diagnostic clinical trials. While time, cost efficiency, and ethical feasibility are the main advantages of VICTs, VICTs offer one additional attribute; ground truth can be perfectly known and precisely controlled. As the condition of the patient is defined a priori, a VICT makes it possible to ascertain how an image analysis metric represents the ground truth. This is a unique capability that can never be assured in clinical trials. Of course, a VICT cannot predict the impact of a novel therapy on a type of disease in a specific patient. However, VICTs can predict the range of outcomes to be expected for a pre-determined (i.e. plausible) domain of known variables, e.g. baseline tumor size and the

[1] Department of Radiology, University of Washington, Seattle WA
[2] Department of Radiology, Columbia University; New York Presbyterian Hospital, New York NY
[3] Department of Radiology, Harvard University, Boston MA
[4] Department of Biostatistics, Memorial Sloan Kettering Cancer Center, New York NY

subsequent shrinkage due to a postulated therapy. Over the last several years there has been a steady improvement of the realism of human and imaging system models. The growing maturation of VICTs as useful tools is demonstrated by multiple publications in mammography, CT, and PET, and even FDA approvals based on VICT studies of some aspects of image technology.

Our goal is to use VICTs to characterize the accuracy of QIBs using standard-of-care CT in oncology trials. From this we can develop a guide for implementation in clinical trials and also a roadmap for adoption by regulatory bodies, industry, oncologists, cooperative oncology groups and professional societies.

II. Methods

The virtual imaging pipeline component is the computational core, which uses the XCAT patient model [2] as an input to the CT-simulator CatSim [3]. The sources of variability can be grouped into categories along the pathway of the virtual imaging pipeline: (1) patient variability, (2) tumor characteristics, (3) CT acquisition, (4) image reconstruction, and (5) the QIB algorithm.

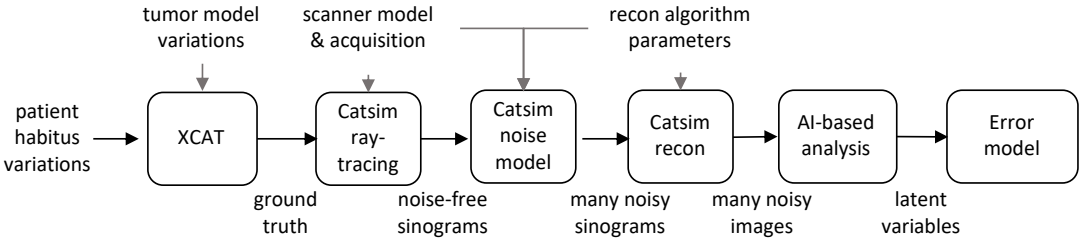


Fig. 1. Data flow in the virtual imaging pipeline.

Data available from the VELOUR clinical trial (NCT00561470) [6], one of the Vol-PACT cohorts [4], are used to define probability density functions. Some of the distributions of scanner- patient-, disease-, and imaging-related sources of variability are shown in Figs. 2 and 3.

Aquilion	Brilliance 6	Emotion 16	LightSpeed	Mx8000 Dual	Sensation 64
Aquilion ONE	Brilliance 64	Emotion 16 (2007)	LightSpeed Plus	Mx8000 IDT 10	Sensation Cardia
Asteion	Brilliance Big Bore	Emotion 6	LightSpeed Pro 16	Mx8000 IDT 16	Sensation Cardiac
AVE1	Brilliance iCT	Emotion Duo	LightSpeed Pro 32	MxTwin	Sensation Cardiac 64
Balance	Brilliance16	Esprit+	LightSpeed QX/i	null	Sensation Open
Biograph 40	CT/e	Gemini TF	LightSpeed RT	Philips CT Aura	Somaris/5 3D
Biograph 6	CT/e Dual	GEMINI TF TOF 64	LightSpeed RT16	PQ2000	SOMATOM Definition Flash
Biograph 64	Definition	GeminiGXL 16	LightSpeed Ultra	ProSpeed	SOMATOM PLUS 4
Biograph40	Definition AS	HiSpeed	LightSpeed VCT	SCT-7000TS	Spirit
BrightSpeed	Definition AS+	HiSpeed CT/i	LightSpeed Xtra	SCT-7800TX	Volume Access
BrightSpeed S	Discovery CT750 HD	HiSpeed Dual	LightSpeed16	Sensation 10	Volume Zoom
Brilliance 16	Discovery ST	HiSpeed NX/i	Manufacturer Model	Sensation 16	Xpress/GX
Brilliance 16P	Discovery STE	HiSpeed QX/i	MX 6000 Dual	Sensation 4	
Brilliance 40	Emotion	iCT 256	Mx8000	Sensation 40	

Fig. 2. List of the CT scanner models used in the VELOUR trial as recorded in the DICOM image headers

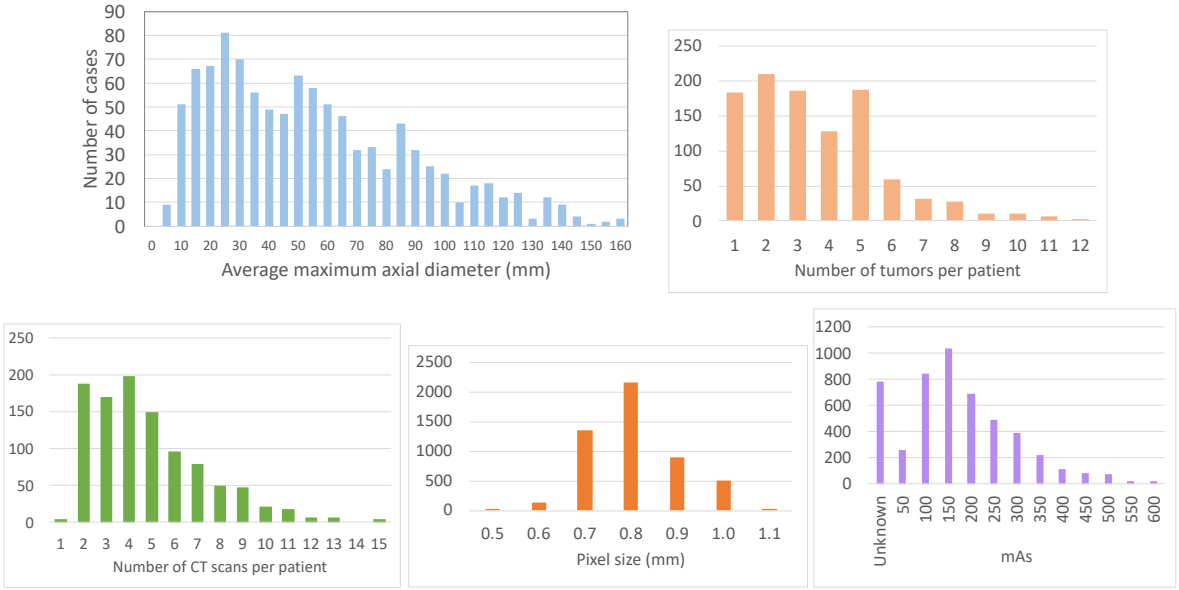


Fig. 3: Some of the sources of variability for the multicenter VELOUR trial data [6]. Shown are the average baseline tumor diameter, the number of tumors per patient at baseline, the number of standard-of-care CT scans per patient, the reconstruction pixel size and mAs per scan.

We used a VICT based on a two-arm trial (control and treatment) as shown in Fig. 4. that uses a baseline and follow-up scan to determine reduction in average tumor volume.

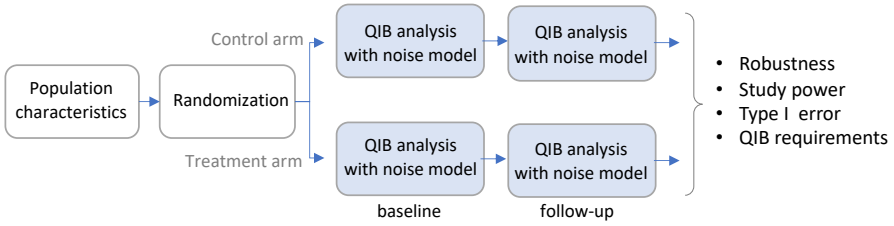


Fig. 4 Virtual imaging clinical trial (VICT) with multi-center baseline and follow-up CT scans. The impact of variability of AI-derived quantitative imaging biomarkers (AI-QIBs) on study power as a function of patient numbers, effect size, and measurement type is assessed.

For a range of effect sizes and trial sizes, we computed study power as a function of QIB variability. The error model used a generalized linear approach for bias and variance of a QIB. In this case we used prior tumor volume estimates (12.5% CoV, but over 25% has been reported). There were 1,000 simulations for each parameter combination to evaluate the QIB in terms of standard error, Type I error, and Type II error (i.e. 1 - study power).

III. Results

Simulated data to be used as plausible ground truth was generated using correlated log-normal distributions modeled on the measured data (Fig. 5). Goodness of fit was checked with Q-Q plots and other statistical tests.

Initial results of study power (Fig. 6) demonstrate the impact of QIB variance in clinical trials using multicenter standard-of-care CT imaging, which features heterogeneity in imaging systems across sites.

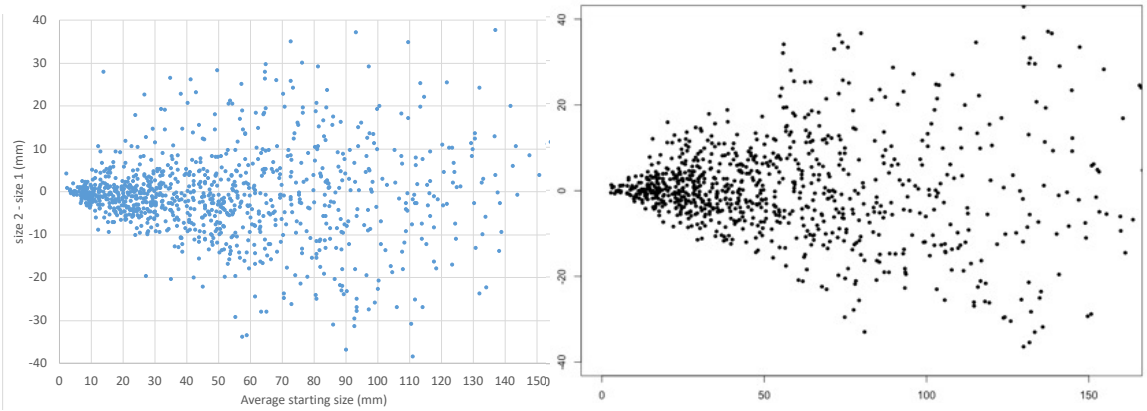


Fig. 5. Tumor size difference (i.e. follow-up - baseline) as a function of baseline tumor size. Left: Measured data from the VELOUR trial with 1,043 patients. Right: Simulated results from multivariate log-normal distributions displayed using the same scales.

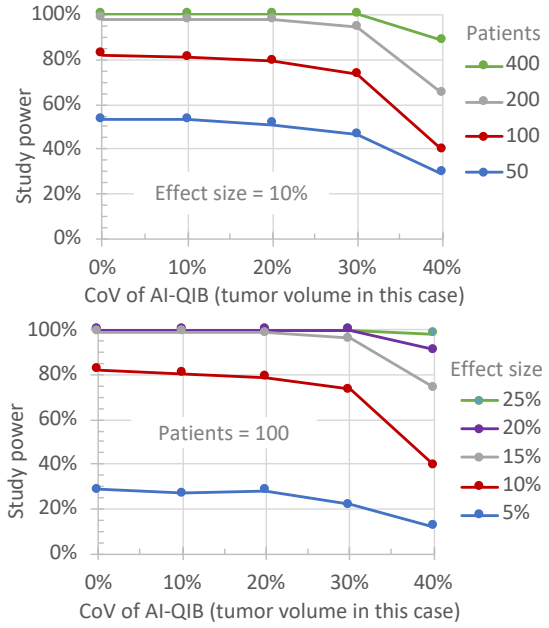


Fig. 6. Study power for the clinical trial illustrated in Fig. 4 showing the importance of understanding the variability of the QIB as a function of sample size and the true difference between arms. Top: impact of QIB coefficient of variation (CoV) and study size. Bottom: Importance of QIB CoV for small studies, i.e. targeted and/or expensive therapies. Data for 100 patients and effect size = 10% is common to both plots, showing the importance of controlling the CoV for a typical study power of 80%.

IV. Discussion

Reliable smaller-*n* studies are imperative for clinical trials that are smaller, more fractionated, and use more expensive therapies. Understanding the application of QIBs to reduce the number of patients while retaining study power (and knowledge of the expected study power) is important for these trials to be successful in the advancement of more effective therapies. These methods are based on data from prior clinical trials, and in turn will provide feedback on the robustness of more effective QIBs and guidance for their use in clinical trials.

References

[1] E. Abadi, W. P. Segars, B. M. W. Tsui, P. E. Kinahan, N. Bottenus, A. F. Frangi, A. Maidment, J. Lo, and E. Samei, "Virtual clinical trials in medical imaging: a review.," J. Med. Imag, vol.7 2020.

[2] W. P. Segars, G. Sturgeon, S. Mendonca, J. Grimes, and B. M. W. Tsui, "4D XCAT phantom for multimodality imaging research.," Medical Physics, vol. 37, no. 9, pp. 4902–4915, Sep. 2010.

[3] B. De Man, S. Basu, N. Chandra, B. Dunham, P. Edic, M. Iatrou, et al., "CatSim: a new computer assisted tomography simulation environment," Proceedings of SPIE, vol. 6510, p. 65102G, 2007.

[4] L. Dercle, D. E. Connors, Y. Tang, et al., "Vol-PACT: A Foundation for the NIH Public-Private Partnership That Supports Sharing of Clinical Trial Data for the Development of Improved Imaging Biomarkers in Oncology.," *JCO Clinical Cancer Informatics*, vol. 2, pp. 1–12, Dec. 2018.

[5] H. Yang, L. H. Schwartz, and B. Zhao, "A Response Assessment Platform for Development and Validation of Imaging Biomarkers in Oncology.," Tomography, vol. 2, pp. 406–410, Dec. 2016.

[6] J. Tabernero, E. Van Cutsem, R. Lakomý, J. Prausová, P. Ruff, G. A. van Hazel, V. M. Moiseyenko, D. R. Ferry, J. J. McKendrick, K. Soussan-Lazard, S. Chevalier, and C. J. Allegra, "Aflibercept versus placebo in combination with fluorouracil, leucovorin and irinotecan in the treatment of previously treated metastatic colorectal cancer: prespecified subgroup analyses from the VELOUR trial.," Eur. J. Cancer, vol. 50, no. 2, pp. 320–331, Jan. 2014.

Reconstructing Invariances of CT Image Denoising Networks using Invertible Neural Networks

Elias Eulig, Björn Ommer, and Marc Kachelrieß

***Abstract*—Long lasting efforts have been made to reduce radiation dose and thus the potential radiation risk to the patient for CT acquisitions without severe deterioration of image quality. To this end, different reconstruction and noise reduction algorithms have been developed, many of which are based on iterative reconstruction techniques, incorporating prior knowledge in the image domain. Recently, deep learning-based methods have shown impressive performance, outperforming many of the previously proposed CT denoising approaches both visually and quantitatively. However, with most neural networks being black boxes they remain notoriously difficult to interpret and concerns about the robustness and safety of such denoising methods have been raised. In this work we want to lay the fundamentals for a post-hoc interpretation of existing CT denoising networks by reconstructing their invariances.**

***Index Terms*—low dose, denoising, explainable, invariances.**

I. Introduction

In recent years, deep learning methods have been employed for many problems in medical image formation, including image-based and projection-based noise reduction, image reconstruction, scatter estimation, and artifact reduction. While the results of deep neural-network (DNN) based methods often excel those of conventional algorithms both qualitatively and quantitatively, they lack interpretability due to most DNNs being black boxes. Particularly for low dose CT imaging, recent advancements in generative methods such as generative adversarial networks (GANs) [1] and variational autoencoders (VAEs) [2] demonstrated impressive performance, providing competitive image quality compared to commercial iterative reconstruction techniques [3].

In this work, instead of focusing on the actual denoising performance of DNN-based methods for CT imaging, we want to lay the fundamentals for a post-hoc analysis of such networks in terms of their interpretability and robustness. To this end, we investigate what they have learned to represent and to ignore (i.e. their invariances) at different layers and argue that robust and non-robust denoising networks are invariant to different input features. Note, that this type of analysis is not restricted to CT and similar methods can be applied to denoising networks for other imaging modalities (e.g. magnetic resonance imaging or positron emission tomography).

Elias Eulig is with the German Cancer Research Center (DKFZ), Heidelberg, Germany.

Prof. Dr. Marc Kachelrieß is with the German Cancer Research Center (DKFZ), Heidelberg, Germany and with the Heidelberg University, Heidelberg, Germany.

Prof. Dr. Björn Ommer is with the University of Munich and with the Interdisciplinary Center for Scientific Computing at Heidelberg University.

Corresponding author: Elias Eulig (elias.eulig@dkfz.de)

II. Background

A. CT Image Denoising with DNNs

In this work we assume to have high-dose images $y \in \mathbb{R}^{m\times n}$ as well as low dose images $x \in \mathbb{R}^{m\times n}$ during training time. The aim of any deep-learning based denoising method is then to find a function $f(\cdot\,;\theta)$ with parameters θ, such that

$$\arg\min_{\theta} \|f(x;\theta) - y\| \,, \tag{1}$$

where f is realised by a DNN. In recent years most improvements on finding an optimal f focused on alterations of the architecture and training scheme. While earlier work utilized pixelwise losses (in image or feature space) which lead to smooth predictions and lack high-frequency information [4, 6], many recent methods are being trained as GANs, leading to extremely realistic denoising results [3, 5].

B. Invariances of DNNs

Our work is based on reference [7], where the authors seek to reconstruct and interpret the invariances of image classification DNNs using invertible neural networks (INNs).

Given a network $f(x)$ we can analyze any internal latent representation z thereof by decomposing f into $f(x) = \Psi(z) = \Psi(\Phi(x))$. To then explain z we need to know what information of the input x is captured in z and to what information Φ is invariant to (and is thus missing in z). To this end, the authors of [7] employ a VAE comprised of an encoder E and a decoder D that is trained to learn a complete data representation $\bar{z} = E(x)$ by reconstructing the input from $\bar{z}$ s.t. $\|D(E(x)) - x\|$ is minimized.

Since the complete data representation $\bar{z}$ now not only contains the information captured in z but also its invariances v, we need to disentangle v and z by learning a mapping

$$t(\cdot|z) : \bar{z} \to v = t(\bar{z}|z)\,. \tag{2}$$

Here, it is assumed that invariances v can be sampled from a Gaussian distribution, i.e. $p(v) = \mathcal{N}(v|0,1)$, and the mapping t is realized through a normalizing flow [8–10], a sequence of INNs between the simple (normal) distribution $p(v)$ and the complex distribution $p(\bar{z})$.

Since t is invertible, we can generate new $\bar{z}$ that only differ in their realization of the invariances by first sampling $v \sim p(v)$ and then applying the inverse mapping of t

$$t^{-1}(\cdot|z) : v \to \bar{z} = t^{-1}(v|z)\,. \tag{3}$$

To visualize $\bar{z}$ in the low dose image space we can reconstruct them using the previously trained decoder D: $\bar{x} = D(\bar{z})$.

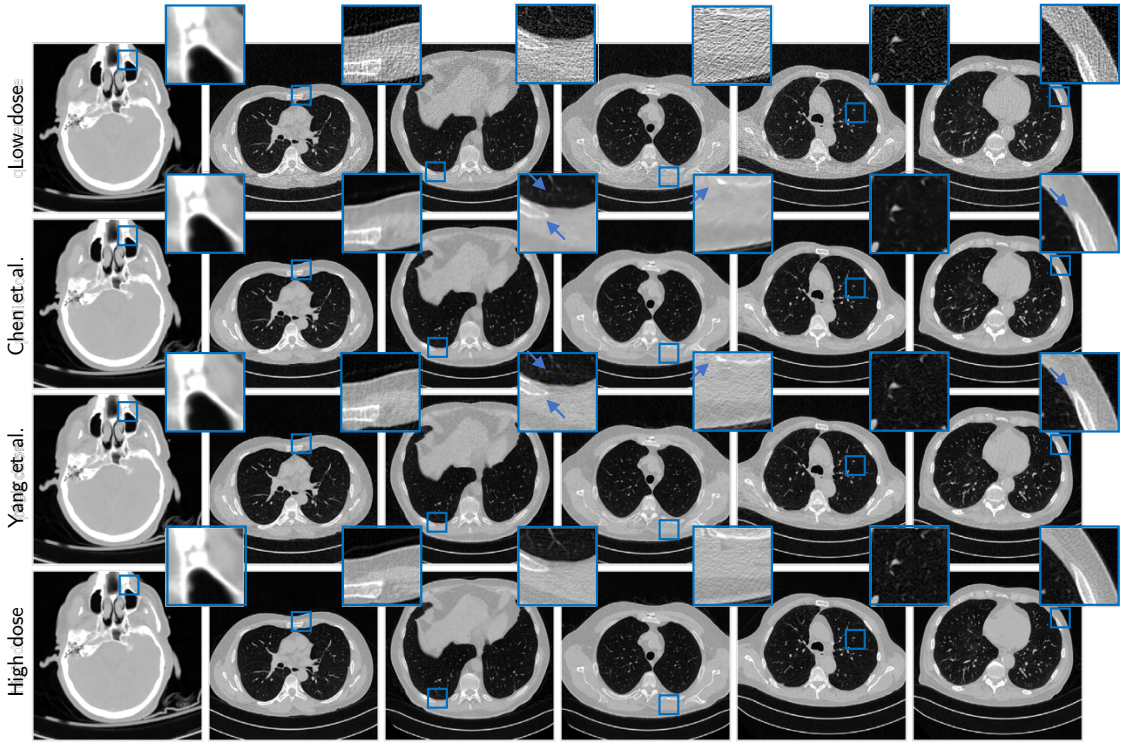


Fig. 1: Denoising performance of Chen et al. [4] and Yang et al. [5] for six different dataset samples (columns). Blue arrows indicate regions where the networks produced errors in the reconstruction of anatomical details.

III. Methods

A. Dataset

For all our studies the Low Dose CT Image and Projection dataset [11] is employed. The dataset comprises 50 head scans, 50 chest scans, and 50 abdomen scans acquired at routine dose levels with a SOMATOM Definition Flash (Siemens Healthineers, Forchheim, Germany) CT scanner. Additionally the dataset provides simulated low dose reconstructions (at $25\,\%$ dose for abdomen/head and at $10\,\%$ dose for chest scans) which were used as input to the denoising networks. We split the dataset into $70\,\%/20\,\%/10\,\%$ for training/validation/test across all patients and trained with a weighted sampling scheme such that slices from each patient were sampled with equal probability.

To make results between different methods comparable we trained and validated all denoising networks as well as the invariance reconstruction method on the same training/validation split of our data.

B. Denoising Methods

While our method can be used to provide post-hoc invariance analysis for any (trained) DNN-based denoising method, for simplicity, we here focus on interpreting the invariances of two well-known denoising methods:

Chen et al. [4] proposed a simple three-layer convolutional neural network which was trained to minimize (1) using an L_2 loss. The authors trained their network on patches of size 33×33 using an SGD optimizer and showed that their method can outperform conventional state-of the art methods.

Yang et al. [5] improved on previous works by training a Wasserstein GAN (WGAN) [12] in combination with a perceptual loss [13] in feature space. Furthermore, they utilize a deeper generator compared to [4] and train the network on larger patches of size 64×64.

We trained both [4] and [5] on the dataset described in Sec. III-A using the hyperparameters as described in the original papers. Whenever hyperparameters were not stated by the authors, we ran a grid-search and used the parameters that result in the lowest validation loss.

TABLE I: Overview of generator architectures used in Chen et al. [4] and Yang et al. [5]. Kernel sizes of the 2D convolutions are indicated by k and their number of filters by f. Final nonlinearities of the original architectures were omitted to accommodate for the normalization of our data.

Layer	Chen et al. [4]	Yang et al. [5]
1	Conv k9 f64	Conv k3 f32
	ReLU	ReLU
3	Conv k3 f32	Conv k3 f32
	ReLU	ReLU
5	Conv k3 f1	Conv k3 f32
⋮		⋮
15		Conv k3 f1

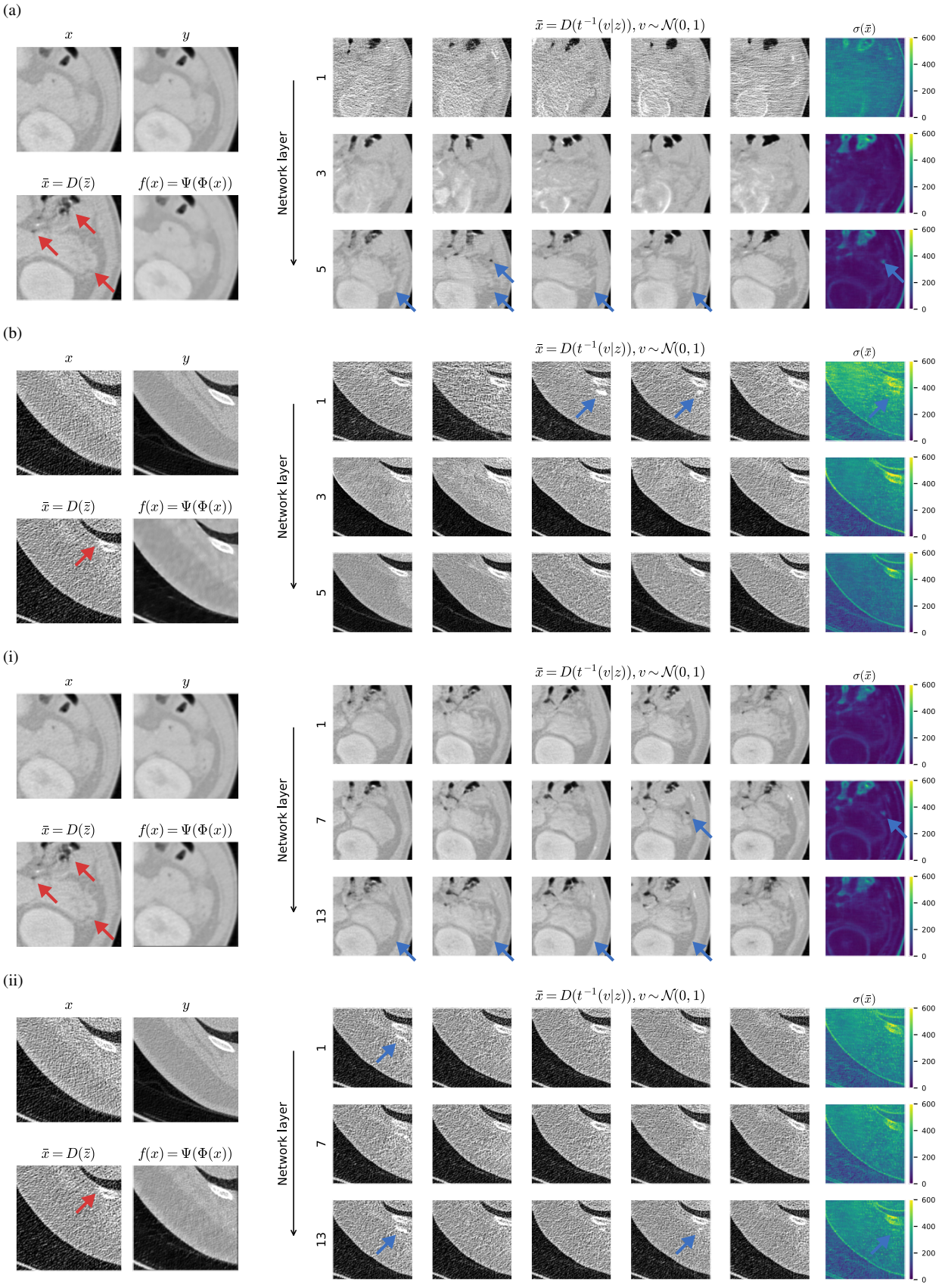


Fig. 2: Best viewed in color. Analysis of Chen et al. [4], (a) & (b), and Yang et al. [5], (i) & (ii). Provided are low dose input image x, high dose ground truth image y, VAE network reconstruction $\bar{x}$ (Sec. III-C), denoised image $f(x)$, five reconstructed samples from the space of invariances, and the standard deviation over 250 invariance samples. Red arrows highlight errors in the VAE reconstruction $D(\bar{z})$ and blue arrows highlight regions in the reconstructed invariances $D(t^{-1}(v|z))$.

C. Recovering Invariances

Similar to reference [7] we first learn a complete data representation $\bar{z} = E(x)$ for a given low dose image x by training a VAE $g(x) = D(E(x))$. Our encoder is based on a ResNet-101 [14] and our decoder on a BigGAN [15] where the conditioning on the class is replaced by a conditioning on the latent representation $\bar{z}$. To improve reconstruction quality the VAE is trained together with a critic C as a WGAN and instead of training it on entire 512×512 pixels images we train it on 128×128 pixels patches.

For both of the two denoising networks evaluated, we train three conditional INNs (cINNs) to learn to reconstruct invariances at three different layers in the network. For Chen et al. [4] we do so at layer 1, 3, and 5 and for Yang et al. [5] at layer 1, 7, and 13 (refer Tab. I). Each of the cINNs, t is composed of four invertible blocks, where each block is composed of coupling blocks [16], actnorm layers [17], and shuffling layers. For each invertible block, the conditioning on the denoising network representation z is realized by concatenating an embedding $h = H(z)$, where H is a shallow network, with the input to the respective block.

For each network and layer we then reconstruct different samples of the invariances $\bar{x} = D(t^{-1}(v, z)), v \sim \mathcal{N}(0, 1)$. Additionally, we can compute the standard deviation over a large set (here 250) of samples to highlight regions with high variation across the reconstructed invariances.

IV. Results

A. Denoising Methods

We find that the results from both denoising networks are similar to those reported in the respective original papers (Fig. 1). Due to the L_2 loss in image space the results from [4] appear smooth and lack structural fidelity. This is alleviated by training with an adversarial loss and consequently our results for [5] look much more realistic with higher details and noise structures very similar to those present in the high dose images.

However, we find that both methods are unable to correctly reconstruct anatomical details in several cases (refer Fig. 1, blue arrows). This is particularly problematic when the network is trained in an adversarial setting, where those false anatomies can look very convincing to the radiologist.

B. Reconstructed Invariances

The reconstructed invariances for both networks and two different samples (ref. Sec. III-C) are provided in Fig. 2. For each sample we also show the low dose input image x, the high dose ground truth image y, the reconstruction of the complete data representation $\bar{x} = D(\bar{z})$, and the denoised image $f(x)$.

From this we find that both denoising methods are invariant to several anatomical features to some extent (Fig. 2; blue arrows). We also find a higher overall variance of the invariances in homogeneous regions of the image for [4], indicating that it is more invariant to the specific realization of noise in the low dose input image. However, when inspecting the VAE reconstructions $\bar{x} = D(\bar{z})$ we also find major deviations from the original low dose image x (Fig. 2, red arrows), which may explain some of the differences between the reconstructed invariances and x.

V. Conclusion

In this work we analyzed deep neural networks for CT image denoising regarding their invariances to anatomical features in the low dose image domain. To reconstruct those invariances we adapted a method from prior work on interpretable AI and sampled reconstructions of invariances for two CT denoising networks. Upon analysis of the reconstructed invariances, we find that the representations of both networks at different layers are invariant to several anatomical features.

While this work demonstrated the potential of an invariance-based analysis of DNNs for CT image denoising, the ability to interpret those invariances is currently limited due to reconstruction errors from the embedding $\bar{z}$ and the complex, high-dimensional structure of the invariance images $\bar{x}$. Overcoming this drawback by improving the embedding $\bar{z}$ as well as mapping the sampled invariances to a semantically meaningful space remains part of future work.

Acknowledgment

This work was supported in part by the Helmholtz International Graduate School for Cancer Research, Heidelberg, Germany.

References

[1] I. Goodfellow, J. Pouget-Abadie, M. Mirza, B. Xu, D. Warde-Farley, S. Ozair, A. Courville, and Y. Bengio, "Generative adversarial nets," *NeurIPS*, vol. 2, pp. 2672–2680, 2014.

[2] D. P. Kingma and M. Welling, "Auto-encoding variational Bayes," *ICLR*, 2014.

[3] H. Shan, A. Padole, F. Homayounieh, U. Kruger, R. D. Khera, C. Nitiwarangkul, M. K. Kalra, and G. Wang, "Competitive performance of a modularized deep neural network compared to commercial algorithms for low-dose CT image reconstruction," *Nature Machine Intelligence*, vol. 1, no. 6, pp. 269–276, 2019.

[4] H. Chen, Y. Zhang, W. Zhang, P. Liao, K. Li, J. Zhou, and G. Wang, "Low-dose CT denoising with convolutional neural network," *International Symposium on Biomedical Imaging (ISBI)*, pp. 143–146, 2017.

[5] Q. Yang, P. Yan, Y. Zhang, H. Yu, Y. Shi, X. Mou, M. K. Kalra, Y. Zhang, L. Sun, and G. Wang, "Low-dose CT image denoising using a generative adversarial network with wasserstein distance and perceptual loss," *IEEE TMI*, vol. 37, no. 6, pp. 1348–135, 2018.

[6] H. Chen, Y. Zhang, M. K. Kalra, F. Lin, Y. Chen, P. Liao, J. Zhou, and G. Wang, "Low-dose ct with a residual encoder-decoder convolutional neural network," *IEEE TMI*, vol. 36, no. 12, pp. 2524–2535, 2017.

[7] R. Rombach, P. Esser, and B. Ommer, "Making sense of CNNs: Interpreting deep representations & their invariances with INNs," *ECCV*, 2020.

[8] D. J. Rezende and S. Mohamed, "Variational inference with normalizing flows," *ICML*, pp. 1530–1538, 2015.

[9] L. Dinh, D. Krueger, and Y. Bengio, "NICE: non-linear independent components estimation," *ICLR*, 2015.

[10] J. S.-D. Dinh, Laurent and S. Bengio, "Density estimation using real NVP," *ICLR*, 2017.

[11] C. McCollough, B. Chen, D. Holmes, X. Duan, Z. Yu, L. Yu, S. Leng, and J. Fletcher, "Data from low dose ct image and projection data [data set]," *The Cancer Imaging Archive*, 2020.

[12] M. Arjovsky, S. Chintala, and L. Bottou, "Wasserstein generative adversarial networks," *ICML*, vol. 15, pp. 214–223, 2017.

[13] J. Johnson, A. Alahi, and L. Fei-Fei, "Perceptual losses for real-time style transfer and super-resolution," *ECCV*, pp. 214–223, 2016.

[14] K. He, X. Zhang, S. Ren, and J. Sun, "Deep residual learning for image recognition," *CVPR*, pp. 770–778, 2016.

[15] A. Brock, J. Donahue, and K. Simonyan, "Large scale GAN training for high fidelity natural image synthesis," *ICLR*, 2019.

[16] L. Ardizzone, J. Kruse, S. Wirkert, D. Rahner, E. W. Pellegrini, R. S. Klessen, L. Maier-Hein, C. Rother, and U. Köthe, "Analyzing inverse problems with invertible neural networks," *ICLR*, 2019.

[17] D. P. Kingma and P. Dhariwal, "Glow: Generative flow with invertible 1x1 convolutions," *NeurIPS*, vol. 31, 2018.

Local Linearity Analysis of Deep Learning CT Denoising Algorithms

Junyuan Li*, Wenying Wang*, Matt Tivnan*, Jeremias Sulam*, Jerry L Prince†, Michael McNitt-Gray‡, J. Webster Stayman* and Grace J. Gang*

*Department of Biomedical Engineering, Johns Hopkins University, Baltimore, MD 21218

†Department of Electrical and Computer Engineering, Johns Hopkins University, Baltimore, MD 21218

‡Department of Radiological Science, University of California Los Angeles, Los Angeles, CA 90024

***Abstract*—The rapid development of deep-learning methods in medical imaging has called for an analysis method suitable for non-linear and data-dependent algorithms. In this work, we investigate a local linearity analysis where a complex neural network can be represented as piecewise linear systems. We recognize that a large number of neural networks consists of alternating linear layers and rectified linear unit (ReLU) activations, and are therefore strictly piecewise linear. We investigated the extent of these locally linear regions by gradually adding perturbations to an operating point. For this work, we explored perturbations based on image features of interest, including lesion contrast, background, and additive noise. We then developed strategies to extend these strictly locally linear regions to include neighboring linear regions with similar gradients. Using these approximately linear regions, we applied singular value decomposition (SVD) analysis to each local linear system to investigate and explain the overall nonlinear and data-dependent behaviors of neural networks. The analysis was applied to an example CT denoising algorithm trained on thorax CT scans. We observed that the strictly local linear regions are highly sensitive to small signal perturbations. Over a range of lesion contrast from 0.007 to 0.04 mm^{-1}, there is a total of 33992 linear regions. The Jacobians are also shift-variant. However, the Jacobians of neighboring linear regions are very similar. By combining linear regions with similar Jacobians, we narrowed down the number of approximately linear regions to four over lesion contrast from 0.001 to 0.08 mm^{-1}. The SVD analysis to different linear regions revealed denoising behavior that is highly dependent on the background intensity. Analysis further identified greater amount of noise reduction in uniform regions compared to lesion edges. In summary, the local linearity analysis framework we proposed has the potential for us to better characterize and interpret the non-linear and data-dependent behaviors of neural networks.**

I. Introduction

Recent years we have seen rapid development of deep learning algorithms in the field of medical imaging. For CT, a popular application of deep learning lies in "denoising" of CT reconstructions. Many network architectures have been proposed in literature and demonstrated potential for reducing image noise and improving signal to noise ratio. At the same time, the nonlinear and data-dependent nature of such algorithms have raised questions over how to systematically characterize their performance. While positive results have been reported in many cases, we have also observed undesirable behavior where critical diagnostic features (e.g., lesion contrast, size, etc.) can be misrepresented [1]. Therefore, an analysis framework that allows systematic examinations of network performance is essential in understanding the advantages and limitations of deep learning algorithms.

An increasing number of investigations have been devoted to characterizing the performance of deep learning algorithms. So far, most studies have relied on evaluating traditional medical image quality measures (e.g, resolution, noise, and detectability index) using specific phantoms or clinical images [2] [3]. While these studies elucidated many interesting and important dependencies in deep learning algorithms, it is difficult to generalize such retrospective analysis in a systematic manner.

Local linear approximation is a common analysis method for nonlinear systems and has been applied to deep learning as well to investigate network stability [4], derive adversarial examples [5], etc. In this work, we seek to identify locally linear representations of a deep learning CT denoising network. Using such representations, we then apply linear system analysis tools to different local linear systems to explain the overall nonlinear and data-dependent behavior of the network.

II. Methods

A. Piecewise Linear Neural Networks

In this work, we consider common deep learning networks consisting of alternating linear layers (e.g., fully connected layer, convolutional layer, residual blocks) and nonlinear activation functions. Furthermore, we focus on the popular Rectified Linear Unit (ReLU) activation function, which comprises two piece-wise linear functions. We designate each linear function of the ReLU by its *activation indicator*, o. Denoting the input to each ReLU as z,

$$o = 1 \ if \ z \geq 0; \quad o = 0 \ if \ z \leq 0 \tag{1}$$

For a trained network with such structure, each input and output pair is governed by a particular linear system determined by the weights and biases in the linear layers and the activation indicator of each ReLU. Following [6], we define *activation pattern*, $\mathbb{O}$, as the collective activation indicators of each ReLU in the network:

$$\mathbb{O} = \{o^1, o^2, ...o^N | o^n \in \{0, 1\} \quad \forall \quad n \in N\} \tag{2}$$

where N is the total number of ReLUs in the network. Inputs that trigger the same activation pattern are governed by the same linear system and belong to a locally *linear region* in the input space. Thus, we can express the network as the following piecewise linear system :

$$\boldsymbol{\mu}_{\text{out}} = H(\boldsymbol{\mu}_{\text{in}}) = h_L \circ \ldots \circ h_2 \circ h_1(\boldsymbol{\mu}_{\text{in}}) \tag{3}$$

where the function associated with layer l, h_l, is:

$$h_l(\boldsymbol{\mu}_{l-1}) = \mathbf{O}_l(\mathbf{W}_l \boldsymbol{\mu}_{l-1} + \boldsymbol{b}_l) \tag{4}$$

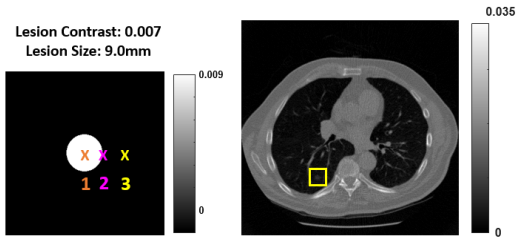


Fig. 1. The operating point $\boldsymbol{\mu}_o$ was chosen as a uniform spherical lesion in the lung region of a thorax CT scan. We present the local Jacobian for three locations at the center, edge, and background of the lesion.

Here, $\mathbf{W_l}$ and $\boldsymbol{b_l}$ denote the weights and biases associated with the linear layers, and $\mathbf{O}_l$ is a diagonal matrix where its diagonal is the vector of activation indicators for the lth layer.

The piecewise linear nature of these networks theoretically allows us to use linear analysis tools to completely characterize the system response for each locally linear system. One such measure that is convenient to compute is the Jacobian, i.e., for any given input or operating point, $\boldsymbol{\mu}_o$, we may write down the corresponding linear system as:

$$H(\boldsymbol{\mu}_k) = H(\boldsymbol{\mu}_o) + \mathbf{J}_o(\boldsymbol{\mu}_k - \boldsymbol{\mu}_o). \tag{5}$$

The Jabocian, $\mathbf{J_o}$, is defined as

$$\mathbf{J}_{o,ij} = \frac{dH(\boldsymbol{\mu}_o)_i}{d\boldsymbol{\mu}_{o,j}}. \tag{6}$$

where i and j are indices of output and input voxels, respetively. This equation holds for all $\boldsymbol{\mu}_k$ that belongs to the same linear region as $\boldsymbol{\mu}_o$.

B. Extent of strictly locally linear regions

While the piecewise linear interpretation of neural networks is convenient, the question remains whether it is practical to analyse each linear region separately. In particular, deep networks tend to partition the input space into a large number of linear regions [7]. It is also possible for networks to have unstable gradient - i.e., small perturbations in the inputs resulting in large changes in gradient or Jacobians. To investigate these behaviors for inputs relevant to CT denoising, we use example CT images as operating points and gradually insert perturbations of interest. For high dimensional input spaces in neural networks, there are many potential types of perturbations that may be explored. Here, we choose clinically relevant perturbations like lesion features of interest (e.g., contrast, shape, texture), noise, or background the lesion is embedded in. We record the activation pattern associated with each perturbation and report the number of changes in activation indicators from the operating point and the total number of activation patterns through the range of perturbations. Furthermore, we compare the Jacobians in neighboring linear regions. In this work, the operating point was chosen as a region of interest (ROI) containing a spherical, uniform lesion with diameter 9.0 mm and contrast 0.007 mm^{-1} in the lung region of thorax CT scan as shown in Fig 1. Results shown below pertain to lesion contrast from 0.007 to 0.04 mm^{-1} in small increments of 2×10^{-7} mm^{-1} .

C. Extent of approximately locally linear regions

Through initial experimentation, we observed that the Jacobians for neighbouring linear regions are similar. We therefore investigate whether we can extend the boundary of locally linear regions to include multiple strictly linear regions with approximately the same Jacobian. Using the same operating point $\boldsymbol{\mu}_o$ and perturbation scheme in the previous section, we compute the output of perturbed inputs $\boldsymbol{\mu}_k$ using the Jacobians for $\boldsymbol{\mu}_o$ (the right hand side of Eq.5) and compare it with the true CNN output. For initial investigation in this work, we compare the maximum (over voxels) absolute percent error between the two and set a threshold below which the two outputs are considered similar enough and that the inputs fall within the same linear region. We developed strategies to choose different operating points so that the percentage error throughout the range of perturbations falls below the threshold.

For results shown below, we present the linear regions and associated operating points for an input space encompassing two types of perturbations - lesion contrast ranging from 0.001 to 0.08 mm^{-1} and the intensity of a uniform background the lesion is embedded in from 0 to 0.04 mm^{-1}.

D. Neural network analysis based on locally linear regions

Using the approximately linear regions identified above, we may apply linear system analysis tools to each linear region to understand and explain some of the overall nonlinear and data-dependent behaviors of neural network algorithms. To ensure generality without assumptions of shift-invariance, we performed SVD of the Jacobians for each approximately linear region. Then, by projecting inputs of interest $\boldsymbol{\mu}_{\text{in}}$ onto the singular basis vectors, we can analyze which features of the inputs that are considered preserved, denoted as $\boldsymbol{\mu}_{\text{in}}^p$ (i.e., with singular values above a certain threshold) and which features are attenuated, denoted as $\boldsymbol{\mu}_{\text{in}}^a$ (i.e., with singular values below a certain threshold). Mathematically:

$$\boldsymbol{\mu}_{\text{in}}^p = \sum_{i \text{ for } s_i > \epsilon} s_i(\boldsymbol{v}_i^T \boldsymbol{\mu}_{\text{in}})\boldsymbol{v}_i; \quad \boldsymbol{\mu}_{\text{in}}^a = \sum_{i \text{ for } s_i \leq \epsilon} s_i(\boldsymbol{v}_i^T \boldsymbol{\mu}_{\text{in}})\boldsymbol{v}_i \tag{7}$$

where s_i and $\boldsymbol{v}_i$ are the ith singular value and basis vector, and ϵ is the threshold on singular values, chosen as 0.10 in this work.

We applied the SVD analysis to noisy input images to visualize how the network "denoises". In this case, the perturbation, $\boldsymbol{\mu}_k - \boldsymbol{\mu}_o$, is noise. We generated 100 different noise realizations at a noise level comparable to the training dataset, and decomposed each realization to the "preserved" and "attenuated" components according to Eq. 7. The mean and standard deviation over all noise realizations are presented to visualize how neural network reduces noise.

E. Experimental setup

In this work, we applied the above analysis to a network based on the REDCNN architecture [8]. We identified 2900 slices from thorax CT scans in the LIDC database [9] and use them as ground truth to generate the training data. The normal and low dose training pairs were generated from filtered-backprojection reconstruction using a barebeam fluence of $I_0 = 10^5$ and $I_0 = 1.25 \times 10^4$, respectively.

III. Results

We first present results showing the extent of strictly linear regions as a function of lesion contrast from the operating point in Fig.1. Combining the number of cumulative activation patterns in Fig.2a and the number of indicator change in Fig.2b, we may infer how many strictly linear regions there are within the range of perturbations investigated. Zooming in on two small range of contrast in Fig.2c, both the cumulative number of activation patterns and the number of indicator changes have the same trend. The perturbation increment was chosen small enough that the input either stays in the same linear region, or transition to a neighboring linear region with at most one ReLU change. Note that while the cumulative activation patterns either stays the same or increases (by definition) for each contrast increment, the number of indicator change may also decrease. Overall, the neural network has seen 33992 strictly locally linear regions for the range of lesion contrast investigated. We further present a profile through the Jacobian matrix that passes through each of the three locations identified in Fig.1. The Jacobians are different from lesion center, to edge, to background, indicating a linear but shift-variant system. Comparing amongst the different strictly linear regions in each portion of the plot, the Jacobians are very similar despite belonging to different linear systems.

Investigations into strictly locally linear regions reveal that the transition between neighboring linear region is sensitive to small changes in perturbations but the Jacobians are similar. We therefore investigate whether the input space can be partitioned into fewer approximately linear regions using methods in Sec.II-C. We perturbed both the lesion contrast and the lesion background intensity and plotted the maximum absolute percent error between local linear approximation and CNN output. Four example inputs at different contrast and background (labeled μ_k) are presented showing comparisons between the empirical CNN output, a linear approximation using Jacobians at an operating point (labeled μ_o), and the percentage error map between the two. Figure 3(a) shows the approximation errors for just one operating point shown in Fig.1. The error is 0 at the operating point and increases as we move further away. Figure 3(b) and (c) shows the approximations error improving as we use more operating points. Compared to 33992 strictly locally linear regions in just one dimension (Fig.2), the linear approximation method yields a much small number of linear systems that is practical to analyze.

Using the approximate local regions identified in Fig.3(c), we performed SVD analysis on two locally linear systems with operating points $\boldsymbol{\mu}_o^1$ and $\boldsymbol{\mu}_o^3$. We chose inputs belonging to each linear region and added noise as perturbations. The inputs contains lesions of the contrast, noise magnitude, and noise correlation; the only difference is the background intensity. Figure 4(a) shows the FBP input, empirical CNN output, and its linear approximation for both a sample noise realization and standard deviation maps over 100 noise realizations. Good agreement was observed between the CNN outputs and linear approximations for the $\boldsymbol{\mu}_o^3$ case, indicating that noise perturbations can be approximated by the same linear system at the operating point. The noise magnitude was well-approximated for the $\boldsymbol{\mu}_o^1$ case but the spatial distribution could be improved, which suggests that the criteria for linear approximation should be revisited for noise prediction. Comparing the two linear regions, the one based on μ_o^3 imparts greater noise reduction seen from the lower standard deviation magnitude. Figure 4b shows the "preserved" and "attenuated" input features (Eq.7) for four sample noise realizations as well as the mean and standard deviation maps over 100 noise realizations. The attenuated portion is high frequency and appears noise-like for both systems - consistent with the "denoising" purpose of the network. The preserved signal is smoother outside the lesion but contains more mid- to high- frequency variations inside the lesion. Both the preserved and attenuated signals are space-variant, with more noise removal in uniform regions (outside and inside the lesion) compared to the edges. This behavior is more obvious in the mean and standard deviation images.

IV. Conclusions

In this work, we presented a method for analyzing piecewise linear neural networks. We observed rapid transitions between strict locally linear regions and introduced an approximation method to make the analysis more tractable. Linear system analysis tools such as the SVD were applied to explain some of the nonlinear and data-dependent behavior of an example denoising network, specifically, what input features can be preserved and which are not.

The most significant challenge with this type of analysis is the high dimensional input space. We chose to use clinically relevant image features as "perturbations" or search directions to map out the locally linear regions. Future work will encompass a wider range of perturbations so that neural network performance can be analyzed in relation to whether image features important for diagnosis can be preserved. Furthermore, we will investigate strategies to identify maximally separated operating points in the input space such that the analysis remains tractable.

References

[1] G. J. Gang, X. Guo, and J. W. Stayman, "Performance analysis for nonlinear tomographic data processing." SPIE-Intl Soc Optical Eng, 5 2019, p. 124.

[2] J. Solomon, P. Lyu, D. Marin, and E. Samei, "Noise and spatial resolution properties of a commercially available deep learning-based ct reconstruction algorithm," *Medical Physics*, vol. 47, pp. 3961–3971, 9 2020.

[3] P. KC, R. Zeng, M. M. Farhangi, and K. J. Myers, "Deep neural networks-based denoising models for ct imaging and their efficacy." SPIE-Intl Soc Optical Eng, 2 2021, p. 16.

[4] G.-H. Lee, D. Alvarez-Melis, and T. S. Jaakkola, "Towards robust, locally linear deep networks," 7 2019. [Online]. Available: http://arxiv.org/abs/1907.03207

[5] I. J. Goodfellow, J. Shlens, and C. Szegedy, "Explaining and harnessing adversarial examples," 12 2014. [Online]. Available: http://arxiv.org/abs/1412.6572

[6] M. Raghu, B. Poole, J. Kleinberg, S. Ganguli, and J. S. Dickstein, "On the expressive power of deep neural networks," 2017.

[7] G. Montúfar, R. Pascanu, K. Cho, and Y. Bengio, "On the number of linear regions of deep neural networks," 2 2014. [Online]. Available: http://arxiv.org/abs/1402.1869

[8] H. Chen, Y. Zhang, M. K. Kalra, F. Lin, Y. Chen, P. Liao, J. Zhou, and G. Wang, "Low-dose ct with a residual encoder-decoder convolutional neural network," *IEEE Transactions on Medical Imaging*, vol. 36, pp. 2524–2535, 12 2017.

[9] M. C. Hancock and J. F. Magnan, "Lung nodule malignancy classification using only radiologist-quantified image features as inputs to statistical learning algorithms: probing the Lung Image Database Consortium dataset with two statistical learning methods," *Journal of Medical Imaging*, vol. 3, no. 4, p. 044504, 2016.

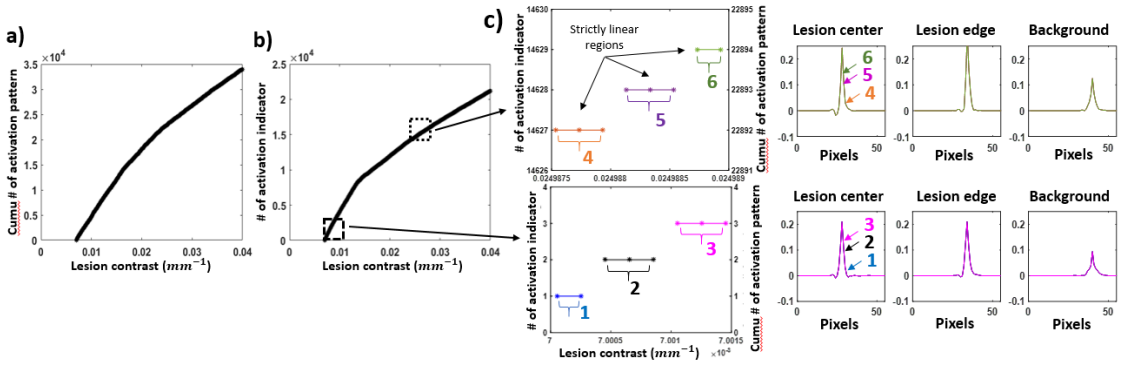


Fig. 2. a) Number of cumulative activation pattern over the range of lesion contrast $0.007 \sim 0.04mm^{-1}$ b) Number of activation indicator change over the same range of lesion contrast c) Top row: Zoom-in view of the curve in b), which shows three strict linear regions and overlays their local Jacobians at three lesion locations. Bottom row: Similar contents for the other zoom-in region (lower lesion contrast)

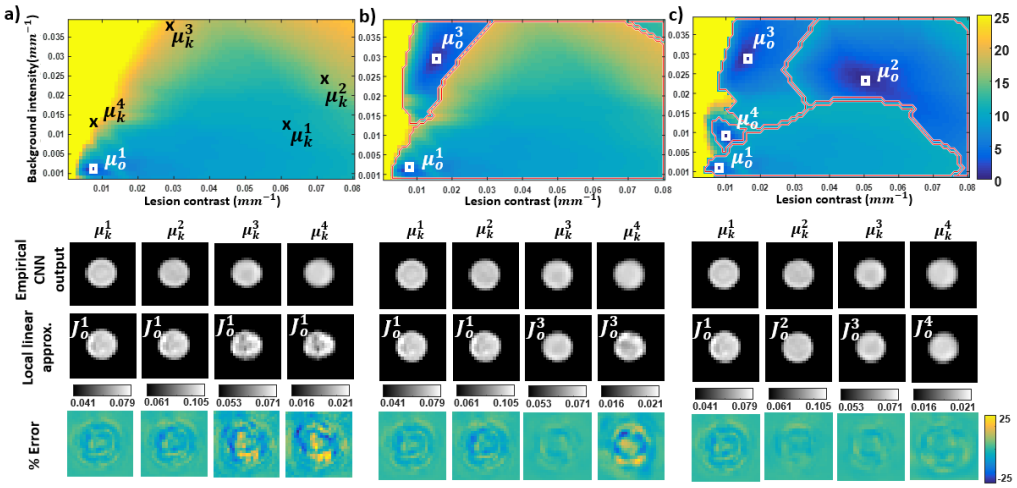


Fig. 3. a) Top row: Map of maximum local linear approximation error with increasing lesion contrast and background intensity for system with operating point $\boldsymbol{\mu}_o^1$; Bottom row: four lesion inputs of interest with increasing approximation error. b) Top row: Input space being partitioned by applying two operating points; Bottom row: lesion inputs of interest evaluated by two systems. c) Top row: Input space being partitioned by applying four operating points; Bottom row: lesion inputs of interest evaluated by each of the system

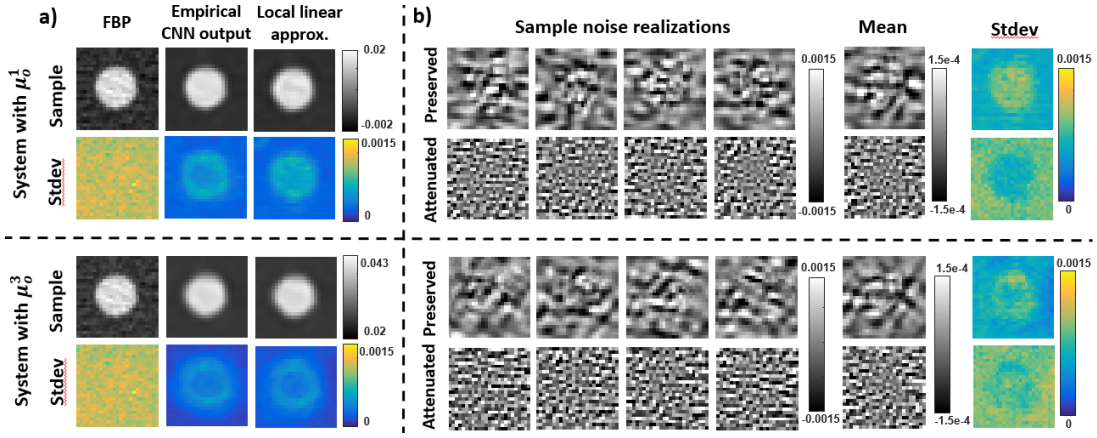


Fig. 4. a) Top row: Sample noisy realization and standard deviation map across 100 different realizations for noisy FBP input, CNN output and local linear approximation by using system with operating point $\boldsymbol{\mu}_o^1$; Bottom row: Similar contents by using system with $\boldsymbol{\mu}_o^3$. b) Top row: Preserved and attenuated features from SVD analysis for sample noise realizations, mean and standard deviation maps across 100 noise realizations by using system with operating point $\boldsymbol{\mu}_o^1$; Bottom row: Similar contents for system with $\boldsymbol{\mu}_o^3$

Evaluation of deep learning-based CT reconstruction with a signal-Laplacian model observer

Gregory Ongie, Emil Y. Sidky, Ingrid S. Reiser, & Xiaochuan Pan

***Abstract*—Recent studies have proposed to optimize deep learning-based CT reconstruction methods for signal detectability performance. However, obtaining objective measures of signal detectability performance of the trained reconstruction networks is challenging due to the non-linear nature of the reconstruction. We propose a simple evaluation metric based on the model observer framework. The metric is based on the performance of a specific linear observer on signal-known-exactly/background-known-exactly task. The linear observer uses the signal Laplacian as a template, which we hypothesize is a better proxy for a human model observer than the ideal/Hotelling observer. We illustrate that the proposed metric can be used to select training hyper-parameters for a CNN-model used to reconstruct synthetic sparse-view breast CT data.**

***Index Terms*—Deep learning, Model observers, CT Reconstruction**

I. Introduction

There has been a surge of interest in training convolutional neural networks (CNNs) to reconstruct low-dose/sparse-view CT data. Most current approaches train the CNN by minimizing a pixel-wise mean-squared error (MSE) or similar loss function over a training set of images. However, these losses are insensitive to small and/or low-contrast features that are critical for screening and diagnosis (e.g., tumor spiculations or microcalcifications in breast imaging), and these subtle features can be significantly degraded in the reconstructions.

To address this issue, recent work has proposed modified CNN training procedures inspired by the model observer framework to enhance the detectability of weak signals in the reconstructions [1], [2], [3]. The model observer framework, based on signal detection theory, offers an objective means to evaluate how well a reconstruction method preserves fine details in the reconstructions at a statistical level.

A major challenge with these approaches – and most other non-linear CT reconstruction techniques – is how to select various tuning parameters. For example, [3] relies on a regularization parameter that trades-off between mean-squared error of the reconstructions and signal detectability performance.

One potential approach, investigated in [3], is to measure signal detectability performance of the reconstructions in terms of the ideal observer, or a close proxy, such as the (channelized) Hotelling observer, on a signal-known-exactly/background-known-exactly (SKE/BKE) task. However, there are several issues with this approach. First, finding the ideal observer for a non-linear reconstruction method is challenging. Second, ideal observer performance is known to correlate poorly with human observer performance. Indeed, if the goal is to maximize performance according to the ideal observer, the optimal strategy is to not process the data at all.

Instead, in this study, we propose evaluating the signal detectability performance using a different type of model observer. The proposed observer model uses a linear test statistic using the discrete Laplacian of the signal as the template. We hypothesize that this observer model is a better proxy for human observer performance.

We illustrate the proposed observer model on simulated data in two settings: a simple denoising setting, and reconstruction of sparse-view breast CT data. In both cases, we demonstrate empirically that there is an identifiable peak in detectability performance of the signal-Laplacian observer when varying tuning parameters, unlike the ideal observer. We find this peak correlates well with our own subjective assessment of preservation of fine details in the reconstructions.

II. Methods

The focus of this work is evaluation of learning-based reconstruction models for sparse-view CT reconstruction. First, we briefly describe the CNN training approach proposed in [3] that is also used in this study. Then we describe the proposed evaluation metric based on an observer model.

A. CNN Training with Observer Regularization

Let $f_{\boldsymbol{\theta}} : \mathbb{R}^d \to \mathbb{R}^d$ denote a CNN depending on parameters $\boldsymbol{\theta} \in \mathbb{R}^p$ mapping noisy sparse-view FBP images $\mathbf{y} \in \mathbb{R}^d$ to reconstructed images $\mathbf{x} \in \mathbb{R}^d$, which we call the *reconstruction network*. Let $\{(\mathbf{x}_i, \mathbf{y}_i)\}_{i=1}^N$ be a collection of training pairs, where each $\mathbf{y}_i$ is a noisy, sparse-view FBP image and $\mathbf{x}_i$ is the corresponding ground truth image, and let $\mathcal{D}$ denote the corresponding empirical distribution of these training pairs. We train the parameters $\boldsymbol{\theta}$ of the reconstruction network $f_{\boldsymbol{\theta}}$ by attempting to minimize the loss given by

$$\min_{\boldsymbol{\theta}} \mathbb{E}_{(\mathbf{x},\mathbf{y})\sim\mathcal{D}} \left[\| f_{\boldsymbol{\theta}}(\mathbf{y}) - \mathbf{x} \|^2 \right] + \lambda \cdot \text{ObsReg}(\boldsymbol{\theta}) \qquad (1)$$

The "observer regularizer" $\text{ObsReg}(\boldsymbol{\theta})$ term is defined with respect to a user-specified distribution of random signals to be planted within the training images. In particular, we assume that pairs $(\mathbf{s}, \hat{\mathbf{s}})$ can be randomly generated, where $\mathbf{s}$ is the signal in input space (e.g., its sparse-view FBP) and $\hat{\mathbf{s}}$ is the same signal represented in output space (e.g., its gridded reconstruction). Then we define $\text{ObsReg}(\boldsymbol{\theta})$ as

$$\text{ObsReg}(\boldsymbol{\theta}) = -\mathbb{E}[(f_{\boldsymbol{\theta}}(\mathbf{y} + \mathbf{s}) - f_{\boldsymbol{\theta}}(\mathbf{y}))^\top \hat{\mathbf{s}}] \qquad (2)$$

G. Ongie is with the Department of Mathematical and Statistical Sciences, Marquette University, Milwaukee, WI, USA e-mail: gregory.ongie@marquette.edu

E. Sidky, I. Reiser, & X. Pan are with the Department of Radiology, University of Chicago, IL, USA

where the expectation above is taken with respect to both the noisy sparse-view FBP training images $\mathbf{y}$ and the random signal pairs $(\mathbf{s}, \hat{\mathbf{s}})$. The observer regularizer measures the correlation between the difference of the reconstructions with signal present/signal absent and the true signal. Minimizing this quantity maximizes their positive correlation. Intuitively, this should enhance signal detectability in the reconstructed images.

B. Evaluation of CNN Reconstruction Methods Using Observer Models

A challenge in deploying the above CNN-based reconstruction scheme is choosing the "best" regularization parameter λ in equation (1). This parameter trades-off between denoising capabilities of the reconstruction network and signal detectability: large values of λ enhance signal detectibilty performance at the expense of more noise in the reconstructions.

One approach, adopted in [3], is to measure the ability of the reconstruction network to preserve small signals on a SKE/BKE task. In [3] a channelized Hotelling observer (CHO) is used as a proxy for the ideal observer. For a range of regularization parameter settings, the CHO is estimated and its AUC is estimated empirically. However, in [3] it was shown that the AUC as determined by the CHO increased monotonically with regularization parameter λ, plateauing for sufficient large λ where the CNN output reconstructions nearly identical to the input noisy FBP image. Therefore, according to this metric, the "optimal" reconstruction is the noisy FBP image. While this may be optimal from an information-theoretic point of view, we conjecture this is not optimal for human observers. The main contribution of this abstract is to investigate an alternative evaluation metric that we conjecture correlates better with human model observer performance.

C. Proposed Observer Model

To measure performance on the SKE/BKE task we propose using a linear observer, i.e., a linear test statistic of the form $t(\mathbf{y}) = \langle \mathbf{w}, \mathbf{y} \rangle$, where $\mathbf{w}$ is a fixed template image. We propose to use the discrete Laplacian of the signal as the template: $\mathbf{w} = \Delta \mathbf{s}$. where $\mathbf{s}$ is the signal as used in the SKE/BKE task, and Δ is the discrete Laplacian computed using centered finite differences.

III. Results

In order to motivate the use of the signal-Laplacian as a template for the human model observer in a SKE/BKE detection task, we consider a simple imaging system of a signal in white noise. We then apply this observer model to parameter tuning for the CNN-based image reconstruction algorithm.

a) Smoothing of image containing a signal in a white noise background: We consider a 256x256 pixel noisy image where the noise follows an uncorrelated Gaussian distribution with uniform pixel standard deviation of 2.0. Furthermore, a detection task is considered with a smoothed-disk signal centered in the middle of the image of radius 5.725 pixels and amplitude 1.0 and the image background is zero. The detection task is tantamount to classifying a shown image into either signal-present or signal-absent image hypotheses, and the confounding factor is the image noise. The observer also has the ability to apply Gaussian smoothing to the image with a full-width half-max (FWHM) parameter w as measured in units of the signal FWHM, 11.45 pixels.

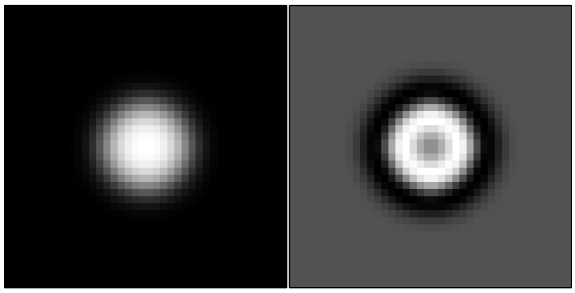


Fig. 1: Blow-up of candidate templates for human model observer signal detection on a 20×20 pixel grid: (Left) the signal itself and (Right) the signal-Laplacian. The considered imaging system is a signal in white noise on a 256×256 pixel array, and the detection task is SKE/BKE.

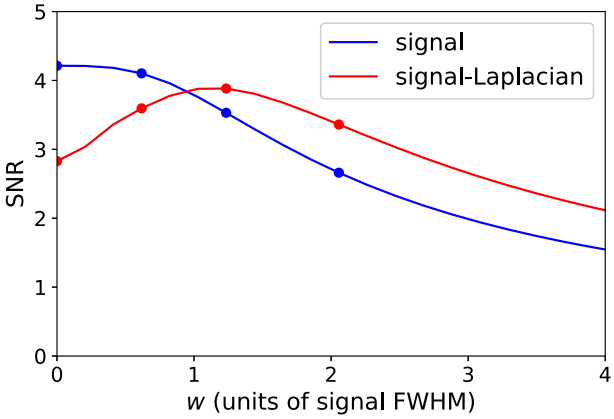


Fig. 2: SNR for SKE/BKE signal detection as a function of smoothing strength for the signal in white noise imaging system. The SNRs are computed with two possible human model observer templates shown in Fig. 1. For no smoothing $w = 0$, the signal template is the ideal observer template, and accordingly the corresponding SNR value of 4.213 is the maximum possible SNR. The solid circles indicate values that correspond to the images in Fig. 3.

The ideal observer for this simple imaging system uses a test statistic that involves the dot product between the image and a template that is the signal itself, because the noise distribution is uncorrelated and uniform. Furthermore, the ideal observer would not perform smoothing at all as its SNR for detection is maximal already with $w = 0$. It is, however, not clear that this signal template is the one that would model a human observer. We hypothesize that a template that focuses on the edges of the signal may be more representative of a human's strategy and we formulate this edge-focused model as the dot product of the image with the signal-Laplacian. To illustrate these two strategies, images of the signal and signal-Laplacian templates

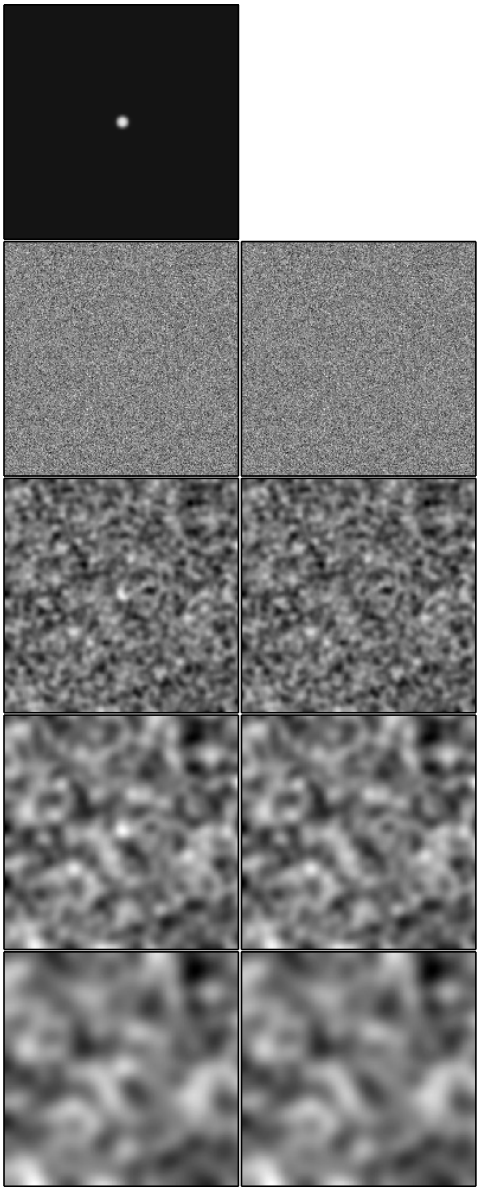


Fig. 3: Noisy signal-present (Left column) and signal-absent (Right column) images for the signal in white noise imaging system. The top row shows the noiseless signal image on the left. The subsequent rows show noisy images for different levels of Gaussian smoothing, $w = 0$, 0.62, 1.23, and 2.06 signal-widths going from top to bottom. For the noisy background the same noise realization is used for all eight panels so that it is easy to observe the difference between the signal-present and signal-absent images. The gray scale for all of the noisy images are determined by the minimum and maximum pixel values in the signal-present images.

are shown in Fig. 1. We note that the signal-Laplacian has a "center-surround" structure that has been associated with human observer 2D templates for detection [4], where a middle region of positive weights is surrounded by a ring of negative weights.

In Fig. 2, the SNR is computed for both signal and signal-Laplacian templates and different levels of Gaussian smoothing. The two curves have quite different behavior with signal and signal-Laplacian SNRs peaking at $w = 0$ and $w = 1.23$ signal-widths, respectively. In order to establish correspondence of these results with visualization, noisy image realizations for both signal-present and signal-absent hypotheses are shown in Fig. 3 for different levels of smoothing. As this figure is only illustrative, it shows a relatively large signal that is easy to detect and the same noise realization is used on all of the noisy images; it is not intended to be representative of a true two-alternative forced-choice experiment. Starting at the top with $w = 0$, it is difficult to distinguish between the signal-present and signal-absent images because the noise amplitude is large compared with the signal. As w increases the noise amplitude is decreased relative to the signal because it is wider than the speckle structure due to the noise, and it becomes easier to see the signal. In the bottom row of the figure, for $w = 2.06$ signal-widths, the smoothing significantly degrades the signal amplitude and the signal once again becomes lost in the noise. Thus, the visual trend of Fig. 3 supports the SNR trend of the signal-Laplacian template from Fig. 2. We note that human-observer experiments would be needed to establish this correspondence quantitatively. For this work, we go ahead and apply the observer model, specified by the dot product with the signal-Laplacian template, to determine parameter settings for the CNN-based image reconstruction algorithm.

A. Evaluation of CNN's for Sparse-View CT Reconstruction

We focus on a sparse-view setting using synthetic breast CT phantoms. For training data, we generate random phantom images using a structured fibro-glandular tissue model. An initial image is generated on a 2048×2048 pixel grid, from which we numerically simulate noisy 128-view sinogram data under a 2D circular, fan-beam scanning geometry, which is representative of the mid-plane slice of a 3D circular cone-beam scan. Noise-free ground truth images are formed by downsampling the initial image to a 512×512 pixel grid. We also compute an initial FBP reconstruction from the simulated sparse-view sinogram data, which is passed as input to the CNN. We generate 1000 FBP and ground truth image pairs in this way to use for training. We use a U-net architecture [5] for the reconstruction network in all our experiments. We modify the standard U-net slightly by adding a residual "skip" connection with trainable weights.

We set up a SKE/BKE task to measure signal detectability performance on a hold-out test set of images. The test set consists of 1000 signal present realizations and 1000 signal absent realizations, all sharing the same fixed background image. To facilitate computation of the signal detectability metrics, we fix the location of the test signal to the center of the image. The signal strength is set so that the data domain AUC is 0.86.

In addition to the signal observer and the proposed signal-Laplacian observer, we also compare against a channelized Hotelling observer that uses a hybrid of pixel and Laguerre-Gauss channels (hybrid-CHO), which have been found effective in estimating signal detectability performance of other nonlinear reconstruction methods [6]. As our figure-of-merit,

we compute the area under the ROC curve (AUC) of each observer. This is estimated empirically using the two-alternative forced choice (2-AFC) calculation over the reconstructed test images.

Figure 4 shows the AUCs obtained by the model observers. We observe that AUCs of the signal observer and the hybrid-CHO observer roughly monotonically increase with increasing λ, plateauing at an AUC close to 0.80. The proposed signal-Laplacian observer, though giving much lower AUCs overall, reaches a peak AUC for a small value of the regularization parameter, similar to the denoising experiment above.

In Figure 5 we illustrate the correspondence between visual image quality and signal detectability metrics by reconstructing a test image containing an additional contrast-detail (CD) insert. The CD insert consists of an 8×8 grid of point-like signals of varying widths and contrasts. Visually comparing the reconstructions obtained from different networks trained with different λ, the signal-Laplacian AUC maximizer ($\lambda = 0.005$) gives a more faithful reconstruction of the CD insert than lower values of λ, while still suppressing noise.

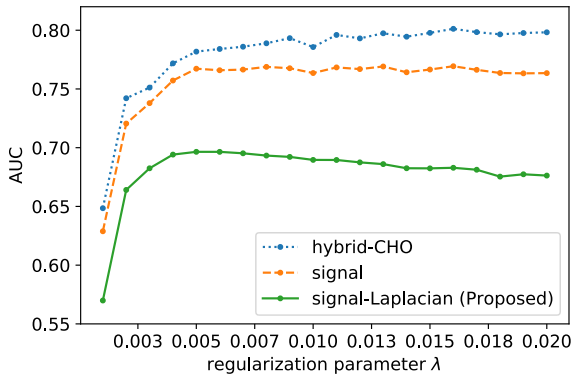


Fig. 4: AUC of observer models vs. observer regularization parameter λ used in training a CNN reconstruction network. Observe that the AUC of proposed signal-Laplacian observer reaches a maximum at $\lambda \approx 0.005$, while the AUC for the Hybrid-CHO and signal observer increases roughly monotonically with λ increasing.

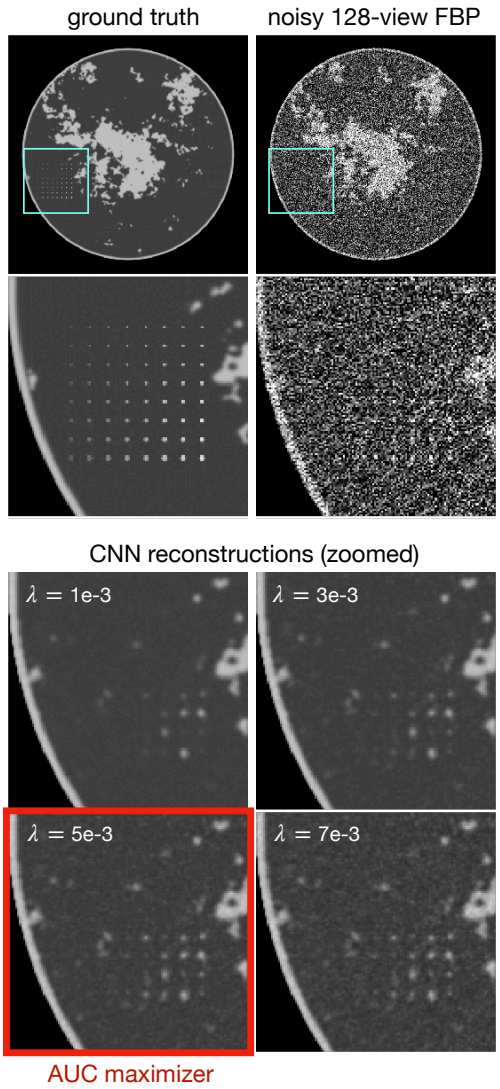


Fig. 5: CNN reconstructions of noisy 128-view breast CT phantom with contrast detail insert. A family of CNNs is trained over a range regularization parameters λ in (1). Highlighted in red is the output from the CNN trained with the choice of λ that maximizes the AUC of the signal-Laplacian observer.

IV. Conclusion

We propose a model observer approach to assess signal detectability performance of non-linear CT reconstruction using CNNs. The proposed model observer is based on the signal-Laplacian, which we hypothesize is a reasonable proxy for a human model observer. We demonstrate its potential to aid in selecting hyper-parameters when training a CNN to reconstruct synthetic sparse-view breast CT data.

References

[1] W. Wang, G. J. Gang, and J. W. Stayman IV, "A CT denoising neural network with image properties parameterization and control," in *Medical Imaging 2021: Physics of Medical Imaging*, vol. 11595. International Society for Optics and Photonics, 2021, p. 115950K.

[2] M. Han, H. Shim, and J. Baek, "Low-dose CT denoising via convolutional neural network with an observer loss function," *Medical physics*, vol. 48, no. 10, pp. 5727–5742, 2021.

[3] G. Ongie, E. Y. Sidky, I. S. Reiser, and X. Pan, "Optimizing model observer performance in learning-based CT reconstruction," in *Medical Imaging 2022: Physics of Medical Imaging*. International Society for Optics and Photonics, 2022.

[4] C. K. Abbey, M. A. Lago, and M. P. Eckstein, "Comparative observer effects in 2D and 3D localization tasks," *J. Med. Imag.*, vol. 8, p. 041206, 2021.

[5] O. Ronneberger, P. Fischer, and T. Brox, "U-net: Convolutional networks for biomedical image segmentation," in *International Conference on Medical image computing and computer-assisted intervention*. Springer, 2015, pp. 234–241.

[6] J. P. Phillips, E. Y. Sidky, G. Ongie, W. Zhou, J. Cruz-Bastida, I. S. Reiser, M. A. Anastasio, and X. Pan, "A hybrid channelized hotelling observer for estimating the ideal linear observer for total-variation-based image reconstruction," in *Medical Imaging 2021: Image Perception, Observer Performance, and Technology Assessment*, vol. 11599. International Society for Optics and Photonics, 2021, p. 115990D.

Tunable Neural Networks for Multi-Material Image Formation from Spectral CT Measurements

Matthew Tivnan, Grace Gang, Peter Noël, Jeremias Sulam, and J. Webster Stayman

***Abstract*—Quantitative estimation of multi-material density images is an important goal for Spectral CT imaging. However, material decomposition is a poorly-conditioned nonlinear inverse problem. Maximum-likelihood model-based material decomposition results in very noisy material density image estimates. One increasingly popular strategy for noise reduction is to apply deep neural networks for multi-material image formation. The most common loss function is mean squared error with respect to supervised target images such as ground truth or higher-dose cases. However, we believe that the mean-squared error loss function has several issues for multi-material image formation. In this work, we present a new loss function which includes multiple noise realizations with separate weights on covariance and bias for joint denoising of all material bases. By modulating these weights, it is possible to tune the image quality of neural network output images. To demonstrate our proposed approach, we conducted a simulation of a water/calcium/gadolinium spectral CT imaging scenario using a deep neural network for multi-material image denoising. Our results show that by changing the weights of our proposed loss function, it is possible to control the tradeoff between variance and bias for individual materials as well as the control over the bias coupling between materials.**

***Index Terms*—Tunable Neural Networks, Spectral CT, Multi-Energy CT, Material Decomposition, Multi-Material Imaging**

I. Introduction

Spectral CT imaging systems incorporate multiple photon energy sensitivity channels into one data acquisition which makes it possible to estimate material composition based on the differences in mass attenuation spectra. Quantitative estimation of material density images has many clinical applications ranging from iodine concentration estimation, soft-tissue/bone decomposition, and k-edge imaging of one or more contrast agents [1] [2]. One physics-driven approach to multi-material image formation is direct model-based material decomposition (MBMD). However, maximum-likelihood MBMD is a poorly conditioned nonlinear inverse problem and results in noisy material density image estimates.

An increasingly popular strategy is to train a deep neural network (DNN) for multi-material noise reduction. There are many possible ways to incorporate a DNN into the spectral CT image reconstruction and material decomposition data processing chain. Some examples include DNNs for projection-domain correction of spectral distortions [3], image-domain material decomposition and noise suppression [4] [5], virtual-non-contrast image estimation [6], and direct reconstruction and material decomposition [7], [8]. Most multi-material image formation DNNs (including all these cited examples) are trained using a mean-squared error (MSE) loss function. We believe there are several problems with MSE for this application. First, it places equal weight on density errors for different materials despite the fact they may have very different attenuation coefficients. Also, MSE is a fixed objective function which offers no way to trade-off between variance and bias in output images. If the network learns to reduce noise in a way that systematically distorts low-contrast anatomical features, that can be a major problem for many clinical applications.

In this work, we propose a modification to the MSE loss function which uses multiple noise realizations for training with separate weights on covariance and bias of the DNN output image distribution. We present a demonstration of our proposed loss function involving a simulated water/calcium/gadolinium spectral CT imaging scenario using a population of digital anthropomorphic phantoms. Our results show that by modifying these weights, it is possible to control the relative penalty on bias and variance for individual materials as well as combinations of materials such as virtual monoenergetic attenuation images. This loss function provides a new capability for tunable image quality in DNN output images. The approach is an extension of the single-energy denoising approach presented in [9].

Matthew Tivnan, Grace Gang, Jeremias Sulam and Matthew Tivnan were with the Department of Biomedical Engineering, Johns Hopkins University, Baltimore, MD. Peter B. Noël was with the Department of Radiology, Hospital of the University of Pennsylvania, Philadelphia, PA.

This work is supported, in part, by Canon Medical Systems and NIH grant R01EB030494

II. Methods

A. Probabilisitic Model of Multi-Material Imaging

Consider an imaging scenario where a patient, $\mathbf{X}$, undergoes a spectral CT scan resulting in the measurements, $\mathbf{Y}$. A multi-material image formation algorithm is then applied to produce the reconstruction, $\hat{\mathbf{X}}$, which is an estimation of $\mathbf{X}$. In this work, we use a probabilistic model to describe the variation within a population of patients as well as the variation due to noise in both measurements and reconstructions, for a given patient. Therefore, we define $\mathbf{X}$ and $\hat{\mathbf{X}}$ as random vectors representing voxelized multi-material density images and $\mathbf{Y}$ as a random vector representing channelized projection-domain spectral CT measurements. The joint probability density function for this imaging scenario is therefore given by

$$p(\mathbf{x}, \mathbf{y}, \hat{\mathbf{x}}) = p(\mathbf{x})p(\mathbf{y}|\mathbf{x})p(\hat{\mathbf{x}}|\mathbf{y}). \tag{1}$$

We also assume that the reconstructed images are a deterministic function of the measurements and therefore $\hat{\mathbf{X}}$ is conditionally indpendent of $\mathbf{X}$ given $\mathbf{Y}$. That is, $\hat{\mathbf{X}} = \mathcal{H}_{\boldsymbol{\theta}}(\mathbf{Y})$ where the function, $\mathcal{H}_{\boldsymbol{\theta}}$, represents the deterministic multi-material reconstruction algorithm parameterized by $\boldsymbol{\theta}$. Therefore, we have $p(\hat{\mathbf{x}}|\mathbf{x}, \mathbf{y}) = p(\hat{\mathbf{x}}|\mathbf{y}) = \delta(\hat{\mathbf{x}} - \mathcal{H}_{\boldsymbol{\theta}}(\mathbf{y}))$.

Depending on the choice of image reconstruction and material decomposition algorithm, the parameters, $\boldsymbol{\theta}$ could be the hyper-parameters of a model-based iterative reconstruction algorithm (MBIR) or the weights and biases of a deep neural network (DNN). The purpose of this work is to provide method for selecting/training $\boldsymbol{\theta}$ to control image quality of multi-material density estimates.

B. Mean, Bias, and Covariance in Multi-Material Images

The probabilistic end-to-end model of a spectral CT imaging system, including data acquisition, image reconstruction, and material decomposition is given by the conditional distribution $p(\hat{\mathbf{x}}|\mathbf{x})$. Samples from this distribution are the reconstructed images, $\hat{\mathbf{X}}$, provided to clinicians with the intention to represent some ground-truth patient, $\mathbf{X}$. An ideal imaging system would produce image estimates which are both *accurate* and *precise* representations of the ground truth. For a given patient, $\mathbf{X} = \mathbf{x}$, accuracy is quantified by the expectation of error, or bias, and precision is quantified by the noise covariance of $\hat{\mathbf{X}}|\mathbf{x}$. Since $\hat{\mathbf{X}} = \mathcal{H}_{\boldsymbol{\theta}}(\mathbf{Y})$, we can define the mean, bias, and covariance in terms of expectations over $\mathbf{Y}|\mathbf{x}$ as follows:

$$\boldsymbol{\mu}_{\hat{\mathbf{x}}|\mathbf{x}} = \int p(\mathbf{y}|\mathbf{x}) \;\; \mathcal{H}_\theta(\mathbf{y}) \;\; \mathbf{dy} \quad , \tag{2}$$

$$\mathbf{b}_{\hat{\mathbf{x}}|\mathbf{x}} = \int p(\mathbf{y}|\mathbf{x}) \Big[\mathcal{H}_\theta(\mathbf{y}) - \mathbf{x} \Big] \mathbf{dy} = \boldsymbol{\mu}_{\hat{\mathbf{x}}|\mathbf{x}} - \mathbf{x} \quad , \tag{3}$$

$$\boldsymbol{\Sigma}_{\hat{\mathbf{x}}|\mathbf{x}} = \int p(\mathbf{y}|\mathbf{x}) \;\; \Big[(\mathcal{H}_\theta(\mathbf{y}) - \boldsymbol{\mu}_{\hat{\mathbf{x}}|\mathbf{x}})(\mathcal{H}_\theta(\mathbf{y}) - \boldsymbol{\mu}_{\hat{\mathbf{x}}|\mathbf{x}})^T \Big] \;\; \mathbf{dy}. \tag{4}$$

In practice, these quantities can be approximated by computing sample expectations for multiple measurement noise realizations. Examples of bias, $\mathbf{b}_{\hat{\mathbf{x}}|\mathbf{x}}$, in multi-material images include spatial blur, beam-hardening artifacts, systematic errors due to miscalibration of spectral sensitivity, cross-material edge artifacts, and more. The covariance matrix, $\boldsymbol{\Sigma}_{\hat{\mathbf{x}}|\mathbf{x}}$ provides a description of the noise power or variance for each voxel and material estimate as well as all cross-voxel and cross-material noise correlations.

C. Spectral CT Physical Model

We define the measurement likelihood $p(\mathbf{y}|\mathbf{x})$ based on a physical model of spectral CT. We assume that the measurements represent channelized spectral measurements (e.g. from a photon-counting detector, multi-layer detector, etc.). We also assume $\mathbf{Y}|\mathbf{x}$ follows a multivariate Gaussian distribution $\mathcal{N}(\bar{\mathbf{y}}(\mathbf{x}), \boldsymbol{\Sigma}_{\boldsymbol{y}}(\mathbf{x}))$ where the mean model, $\bar{\mathbf{y}}(\mathbf{x})$ and measurement noise covariance, $\boldsymbol{\Sigma}_{\boldsymbol{y}}$ are given by

$$\bar{\mathbf{y}}(\mathbf{x}) = \mathbf{S} \exp\Big(-\mathbf{Q}\mathbf{A}\mathbf{x} \Big) \quad , \quad \boldsymbol{\Sigma}_{\boldsymbol{y}}(\mathbf{x}) = D\Big\{ \bar{\mathbf{y}}(\mathbf{x}) \Big\} \quad . \tag{5}$$

The matrix $\mathbf{A}$ is a CT forward-projector to model the line integral sampling geometry, the matrix $\mathbf{Q}$ contains the mass attenuation coefficient for each material and each photon-energy bin, and the matrix $\mathbf{S}$ contains the system sensitivity spectra for each projection and each spectral channel. The term, $D\{\bar{\mathbf{y}}(\mathbf{x})\}$ represents the diagonal matrix with $\bar{\mathbf{y}}(\mathbf{x})$ as the diagonal elements. In this work, the mean and variance of $\mathbf{Y}|\mathbf{x}$ are Gaussian approximations of Poisson noise.

D. Model-Based Material Decomposition

A physics-driven approach to image reconstruction and material decomposition is to use the maximum-likelihood estimator, $\hat{\mathbf{x}}_{\text{ML}}(\mathbf{y}) = \text{argmin}_{\mathbf{x}} \; \{-\log p(\mathbf{y}|\mathbf{x})\}$. Plugging in our spectral CT physics model for $p(\mathbf{y}|\mathbf{x})$ leads to the nonlinear weighted least-squares objective function,

$$\hat{\mathbf{x}}_{\text{ML}}(\mathbf{y}) = \underset{\mathbf{x}}{\text{argmin}} \;\; \frac{1}{2} (\mathbf{y} - \bar{\mathbf{y}}(\mathbf{x}))^T D^{-1}\Big\{ \bar{\mathbf{y}}(\mathbf{x}) \Big\} (\mathbf{y} - \bar{\mathbf{y}}(\mathbf{x})). \tag{6}$$

This estimator can be implemented via iterative numerical optimization such as the SPS algorithm described in [10].

E. Multi-Material Image Formation with Neural Networks

Machine learning models using deep neural networks (DNNs) are increasingly popular for CT image reconstruction and material decomposition. A common training training strategy for these regression models is mean-squared error (MSE) which is given by

$$\text{MSE}_{\hat{\mathbf{x}},\mathbf{X}} = E_{\hat{\mathbf{X}},\mathbf{X}}[(\hat{\mathbf{X}} - \mathbf{X})^T(\hat{\mathbf{X}} - \mathbf{X})] = E_{\mathbf{X}}[\text{MSE}_{\hat{\mathbf{X}}|\mathbf{x}}]. \tag{7}$$

In practice, the expectation, $E_{\mathbf{X}}$ is typically approximated by computing the sample mean for a batch of random patients from the training dataset. The conditional MSE represents the expectation over noise realizations of the DNN output, $\hat{\mathbf{X}}$ for a given patient $\mathbf{X} = \mathbf{x}$ is defined as

$$\text{MSE}_{\hat{\mathbf{X}}|\mathbf{x}} = E_{\hat{\mathbf{X}}|\mathbf{x}}[(\hat{\mathbf{X}} - \mathbf{x})^T(\hat{\mathbf{X}} - \mathbf{x})] \quad . \tag{8}$$

MSE can be decomposed into a variance component and a bias component as follows:

$$\text{MSE}_{\hat{\mathbf{X}}|\mathbf{x}} = \text{tr}\Big\{ \boldsymbol{\Sigma}_{\hat{\mathbf{x}}|\mathbf{x}} \Big\} + \mathbf{b}_{\hat{\mathbf{x}}|\mathbf{x}}^T \mathbf{b}_{\hat{\mathbf{x}}|\mathbf{x}}. \tag{9}$$

For a situation where the training dataset consists of only one noise realization of the input for each ground-truth output, $\text{MSE}_{\hat{\mathbf{X}}|\mathbf{x}}$ must be approximated from one sample as the squared error for one sample, $(\hat{\mathbf{x}} - \mathbf{x})^T(\hat{\mathbf{x}} - \mathbf{x})$. However, if multiple noise realizations are available, the expectation, $E_{\hat{\mathbf{X}}|\mathbf{X}}$, can be approximated by a sample mean over noise realizations. It is also possible to compute sample mean, sample bias, and sample covariance to decompose the error, as shown in (9), into variance and bias components.

F. Weighted Covariance and Bias (WCB) Loss Function

To address some of the issues with MSE as a loss function for DNN-based multi-material image formation models we propose a new loss function called Weighted Covariance and Bias (WCB) which we define as

$$\text{WCB}_{\hat{\mathbf{x}}|\mathbf{x}} = \text{tr}\Big\{ \mathbf{W_C} \boldsymbol{\Sigma}_{\hat{\mathbf{x}}|\mathbf{x}} \mathbf{W_C}^T \Big\} + \mathbf{b}_{\hat{\mathbf{x}}|\mathbf{x}}^T \mathbf{W_B}^T \mathbf{W_B} \mathbf{b}_{\hat{\mathbf{x}}|\mathbf{x}}. \tag{10}$$

The WCB loss function in (10) is a modification of (9) with the inclusion of bias weighting matrix, $\mathbf{W_B}$, and covariance weighting matrix, $\mathbf{W_C}$. Note that the $\text{WCB}_{\hat{\mathbf{x}}|\mathbf{x}}$ is a patient-conditional loss but the weights $\mathbf{W_B}$ and $\mathbf{W_C}$ are constant for all patients. Similar to (7), batches of patient samples are used during training to minimize the sample expectation $E_{\mathbf{X}}[\text{WCB}_{\hat{\mathbf{x}}|\mathbf{x}}]$.

For the special case, $\mathbf{W_B} = \mathbf{W_C} = \mathbf{I}$, the WCB loss function is equivalent to MSE. One simple modification would $\mathbf{W_B} = \sqrt{\alpha} \;\; \mathbf{I}$, $\mathbf{W_C} = \sqrt{1-\alpha} \;\; \mathbf{I}$ so that the parameter α can be used for bias/variance trade-off as introduced in [9] for single-energy CT. For multi-material imaging, one possibility is to use diagonal weighting matrices which scale the material densities by their mass attenuation coefficient at a certain photon energy. That way, all errors can be evaluated on the same scale rather than summing density errors for different materials. Another possibility is to penalize noise and bias in virtual monoenergetic attenuation images which are the sum of all materials weighted by their mass attenuation coefficients. Finally, we hypothesize it will be possible to control the relative importance of variance or bias in any individual material or linear combination of materials by modifying the relative magnitude of $\mathbf{W_C}$ and $\mathbf{W_B}$.

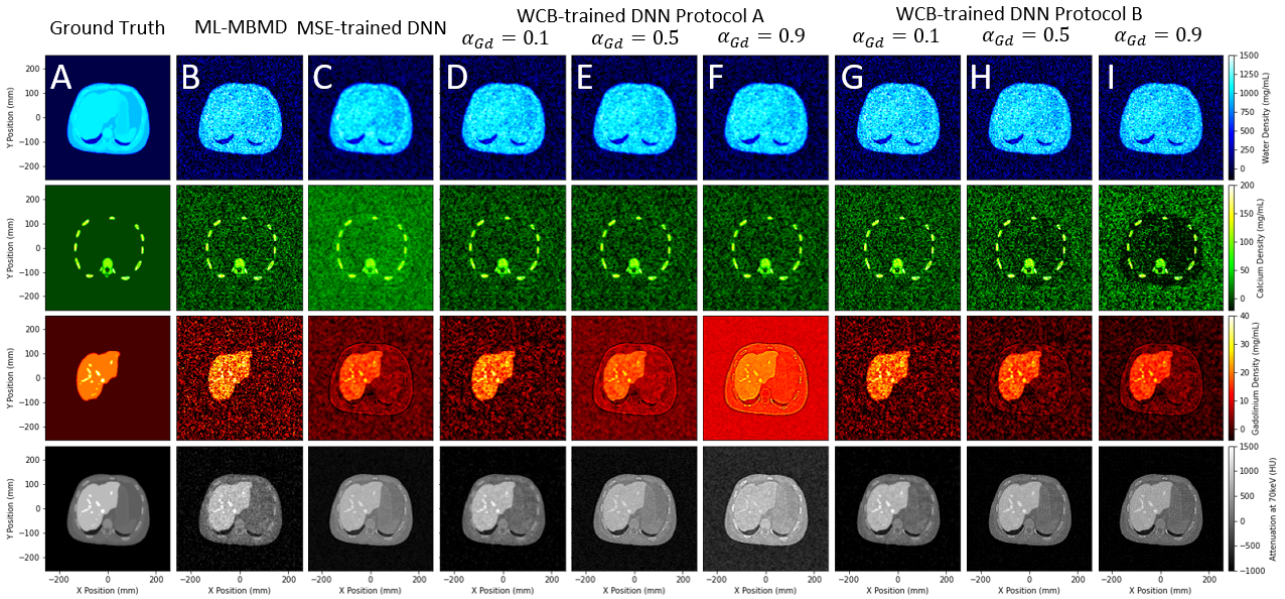


Fig. 1. Water, calcium, and gadolinium density images and total attenuation images at 70keV for A) the ground truth, B) maximum likelihood MBMD which is the CNN input, C) MSE-trained CNN, D-F) WCB-trained CNN for protocol A with D) $\alpha_{Gd} = 0.1$, E) $\alpha_{Gd} = 0.5$, F) $\alpha_{Gd} = 0.9$, G-I) WCB-trained CNN for protocol B with G) $\alpha_{Gd} = 0.1$, H) $\alpha_{Gd} = 0.5$, and I) $\alpha_{Gd} = 0.9$.

G. Simulation Study: DNN for Multi-Material Denoising

To evaluate our proposed WCB loss function, we conducted a simulation of spectral CT for gadolinium contrast-enhanced liver imaging with water/calcium/gadolinium decomposition. The multi-material image formation model has two parts. First, we apply maximum-likelihood MBMD, which produces low-bias, high-variance material density estimates. Second, we apply a neural network for post-processing. Therefore, the network is trained with supervised pairs of noisy MBMD reconstructions and ground-truth material density images.

We generated a population of 50 female and 50 male adult patients as 3D voxelized images with 1.0 mm cubic voxels using the XCAT anthropomorphic digital phantom generator [11] with randomized anatomical features using the XCAT male/female defaults $\pm$ 5% half-width at half-maximum of an independent Gaussian distribution. Two attenuation images were generated at 100 keV and 150 keV to which we applied image-domain decomposition to produce ground truth water and calcium density images. The XCAT activity feature was also used to generate the biodistribution of exogenous gadolinium contrast enhancement with 20.0 $\pm$ 1.0 mg/mL in the liver and 40.0 $\pm$ 2.0 mg/mL in the veins to roughly represent portal venous phase contrast enhancement. From these 100 3D patient models, we sampled 1,000 3-material density image slices, downsampled to size 128×128, randomly scaled by 100 $\pm$ 10%, and randomly rotated by 0 $\pm$ 2 degrees.

We simulated spectral CT measurements for a fan-beam geometry with 1200 mm source-to-detector distance, 800 mm source-to-axis distance, 360 views per rotation, and 1.0 mm detector pixel spacing. For the system spectral sensitivity, we simulated a 120kVp source spectrum with 10^7 photons per pixel per view and ideal photon-counting energy-binning detectors with bin edges at 40, 60, 80, and 100 keV. We used this 5-channel data to reconstruct 3-material water/calcium/gadolinium images using 1000 iterations of the separable parabolic surrogates algorithm from [10] with the cross-material preconditioner from [12]. We repeated this process for 32 noise realizations so the final training data consisted of noisy reconstructed multi-material images with shape $1000 \times 32 \times 3 \times 128 \times 128$ and ground-truth multi-material images with shape $1000 \times 1 \times 3 \times 128 \times 128$. We used 80% of the data for training and reserved 20% for validation.

Our DNN architecture consists of five 3×3 convolutional layers between the input and each layer. Each layer used a ReLU activation function except the last layer which is linear and the number of output channels for each layer, in order, are 128, 64, 32, 16, and 3. We also used skip connections to concatenate the input images to the output of each convolutional layer.

We trained the this DNN architecture repeatedly using the WCB loss function. We used only block-diagonal cross-material weighting matrices. We denote the blocks of $\mathbf{W_B}$ as $\tilde{\mathbf{W}}_B$ and the blocks of $\mathbf{W_C}$ as $\tilde{\mathbf{W}}_C$. Note there is one identical cross-material block matrix per voxel but these blocks do not need to be square; they must have three columns for water, calcium, and gadolinium, but they can have any number of rows. We use cross-material blocks which follow the parameterized form:

$$\tilde{\mathbf{W}}_B = D\{\sqrt{\boldsymbol{\beta}}\}D\{\sqrt{1-\boldsymbol{\alpha}}\}\mathbf{P}D\{\boldsymbol{\mu}\} \tag{11}$$

$$\tilde{\mathbf{W}}_C = D\{\sqrt{\boldsymbol{\beta}}\}D\{\sqrt{\boldsymbol{\alpha}}\}\mathbf{P}D\{\boldsymbol{\mu}\} \tag{12}$$

where $\boldsymbol{\mu}$ is a vector of mass attenuation coefficients at 70 keV, $\mathbf{P}$ is a 4×3 matrix which is an identity matrix for the first 3 rows and all ones for the fourth row to represent a virtual monoenergetic attenuation image, $\boldsymbol{\alpha}$ is a 4×1 vector of parameters controlling the level of noise suppression for each material (or combination), and $\boldsymbol{\beta}$ controls the relative weights. Tuning the output image quality is done by modulating the 8 parameters in $\boldsymbol{\alpha}$ and $\boldsymbol{\beta}$ before training the neural network.

We started by training the network with the Adam optimizer with learning rate 0.0001 for 1000 epochs using

$\boldsymbol{\alpha} = [0.1, 0.1, 0.1, 0.1]$ and $\boldsymbol{\beta} = [1.0, 1.0, 1.0, 0.0]$. From this reference point, we tuned the network with 100 training epochs for different weighting matrices choices. We conducted two one-dimensional parameters sweeps keeping all other weights constant. Protocol A is defined as $\boldsymbol{\alpha} = [0.1, 0.1, \alpha_{Gd}, 0.1]$, $\boldsymbol{\beta} = [1.0, 1.0, 1.0, 0.0]$ and protocol B is defined as $\boldsymbol{\alpha} = [0.1, 0.1, \alpha_{Gd}, 0.1]$, $\boldsymbol{\beta} = [1.0, 1.0, 10.0, 10.0]$, where the single parameter α_{Gd} controls the level of noise reduction in the gadolinium channel.

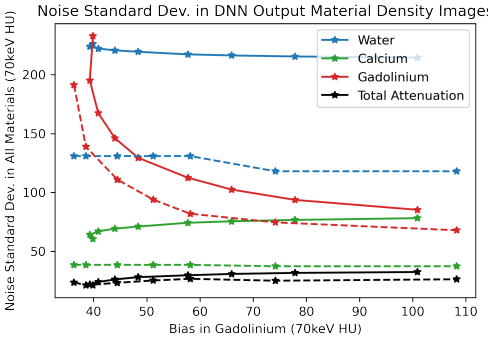


Fig. 2. Noise in DNN output material density images as a function of gadolinium bias which is controlled by α_{Gd}. The dashed line shows protocol A and the solid line shows protocol B.

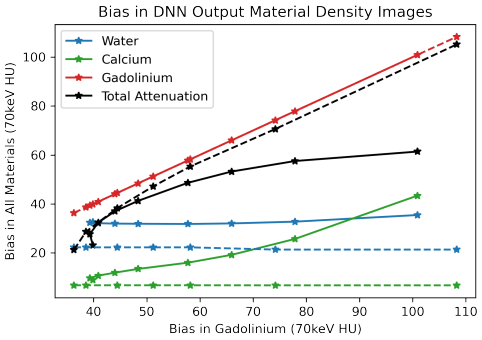


Fig. 3. Bias in DNN output material density images as a function of gadolinium bias which is controlled by α_{Gd}. The dashed line shows protocol A and the solid line shows protocol B.

III. Results

Multi-material density images as well as virtual monoenergetic images at 70 keV are shown in Figure 1. The ground truth image in column A was sampled from the validation dataset. Column B shows a maximum-likelihood MBMD reconstruction which has low bias but high noise. Column C shows mean-squared error training which reduces noise but introduces bias in all four images. Columns D-F show the results for protocol A for three different values of α_{Gd}. For protocol A, changing α_{Gd} controls the bias/variance trade-off for the gadolinium image but has almost no effect on the other other two material images. Large α_{Gd} leads to less noise but more bias in the gadolinium channel as well as more bias in the monoenergetic attenuation images. Columns G-I show the results for protocol B which strongly penalizes bias in the monoenergetic attenuation image. As a result, a higher value of α_{Gd} leads to increased bias in both the gadolinium and calcium channels but in such a way that the weighted sum of materials produces low bias in the monoenergetic attenuation images. These visual observations are also validated by the noise and bias plots in Figures 2 and 3.

Figure 2 shows that changing α_{Gd} can be used to trade-off between variance and bias in the gadolinium density image estimates with very little effect on the noise level for the other materials. The dashed lines in Figure 3 show that, for protocol A, increasing α_{Gd} leads to more bias in gadolinium and the total attenuation images with very little impact on the other materials. The solid lines in Figure 3 show that bias in monoenergetic attenuation images has been reduced for protocol B relative to protocol A. However, this comes at the cost of additional bias in the calcium channel which is needed to cancel out the gadolinium bias. For applications requiring low bias in the calcium channel, we recommend protocol A and for applications requiring low bias in the monoenergetic images we recommend protocol B.

IV. Conclusion

In this work, we have presented a novel loss function for DNNs using multiple noise realizations and explicit weights on covariance and bias of multi-material image estimates. The results of our simulation study confirm that these weights can be used to tune neural network output image quality. For joint image denoising in multi-material imaging, we have shown these weights can be used to emphasize variance and bias in individual materials as well as linear combinations thereof.

One limitation of our proposed method is that it requires multiple noise realizations of the network input. For our simulation study, we used XCAT phantoms and a spectral CT physics simulator to generate multiple noise realizations per patient. To implement WCB in cases where only a single noise realization is available, it may be possible to generate many lower-dose images from one higher-dose image as described by [13]. Another limitation is the relatively simple DNN architecture we used since better performance is likely possible using the latest network architectures. We note that the WCB loss function is very general and can be applied to all the same regression problems which would use MSE. In future studies, we are interested in applying the WCB loss to projection-domain denoising and deep reconstruction models which directly estimate material densities from the data.

The proposed training method offers an new capability for tunable neural networks for multi-material image formation. The specific image quality goals will vary for different tasks which will inform the weights on covariance and bias. The WCB loss function will provide imaging scientists with a flexible new tool to tune neural networks to optimize multi-material image quality for many clinical applications of spectral CT.

References

[1] Schirra, C. O *et al.*, *Contrast media & molecular imaging*, 2014.
[2] Taguchi, K and Iwanczyk, J. S, *Medical physics*, 2013.
[3] Touch, M *et al.*, *Physics in Medicine & Biology*, 2016.
[4] Li, Z *et al.*, in *2020 IEEE 17th International Symposium on Biomedical Imaging (ISBI)*. IEEE, 2020, pp. 651–655.
[5] Xu, Y *et al.*, *Comp. and mathematical methods in medicine*, 2018.
[6] Poirot, M. G *et al.*, *Scientific reports*, vol. 9, no. 1, pp. 1–9, 2019.
[7] Zhang, Y *et al.*, in *Medical Imaging*, 2020.
[8] Wu, D *et al.*, in *Fully3D*, 2019.
[9] Tivnan, M *et al.*, in *SPIE. In Print.*, 2022.
[10] Tilley II, S *et al.*, *Physics in Medicine & Biology*, 2019.
[11] Segars, W. P *et al.*, *Medical physics*, 2010.
[12] Tivnan, M *et al.*, *arXiv preprint arXiv:2010.01371*, 2020.
[13] Zeng, D *et al.*, *IEEE transactions on nuclear science*, 2015.

Self-supervised nonlocal spectral similarity-induced material decomposition network for dual-energy CT

Lei Wang, Yongbo Wang, Zhaoying Bian, Dong Zeng, and Jianhua Ma

***Abstract*—Dual-energy computed tomography (DECT) imaging plays an important role in clinical diagnosis applications due to its material decomposition capability. However, in the cases of low-dose DECT imaging and ill-conditioned issue, the direct decomposed material images from DECT images would suffer from severe noise-induced artifacts, leading to low quality and accuracy. In this paper, we propose a self-supervised Nonlocal Spectral Similarity-induced Decomposition Network (NSSD-Net) to produce decomposed material images with high quality and accuracy in the low-dose DECT imaging. Specifically, we first build the model-driven iterative decomposition model and optimize the objective function by the iterative shrinkage-thresholding algorithm (ISTA) with the convolutional neural network. Considering the intrinsic characteristics information (i.e., structural similarity and spectral correlation) underlying DECT images, which can be used as the prior information to improve the accuracy of the decomposed material images, we construct the nonlocal spectral similarity-based cost function by using the prior information and incorporating it into the iterative decomposition network to guarantee stability. The proposed NSSD-Net method was validated and evaluated with real clinical data. Experimental results showed that the presented NSSD-Net method outperforms the other competing methods in terms of noise-induced artifacts reduction and decomposition accuracy.**

***Index Terms*—Dual-energy computed tomography, deep learning, material decomposition, low-dose, self-supervised learning.**

I. Introduction

DUAL-energy computed tomography (DECT) has been widely used in clinical diagnosis, including kidney stone characterization, iodine quantitative examination, and many other applications. Compared to traditional CT, DECT provides two sets of attenuation measurements by exploiting two different energy spectra to achieve the material decomposition and energy-selective imaging, which opens up new diagnosis possibilities. However, the direct decomposed material images from DECT images would suffer from severe noise and noise-induced artifacts, especially under low-dose scan protocols. The main reason is that the material decomposition process suffers an ill-conditioned problem that leads to severe noise in the decomposed material images [1]. To tackle this problem, many advanced algorithms have been proposed to improve the signal-to-noise ratio of decomposed material images.

These material decomposition algorithms can be generally divided into two categories: model-driven material decomposition methods and data-driven material decomposition methods. The model-driven methods [2] generally build a model on the physical properties and the attenuation characteristics, then optimize the objective function to get the decomposed material images. These methods also can be characterized into one-step inversion, projection-domain, and image-domain decomposition methods. One-step inversion is mathematically the most elegant with one-step matrix inversion, but they are very computationally expensive. Projection-domain methods directly decompose the projection data into the basis material sinograms and then reconstruct them with the filter back-projection algorithm. Although the decomposition performance is significant, they require accurate system calibrations that use nonlinear models. Image-domain methods reconstruct each energy bin image and then directly perform material decomposition on image data, which may lead to beam hardening artifacts. On the other hand, the data-driven methods [3] train the convolutional neural network (CNN) in an end-to-end manner to obtain accurate decomposed material images. However, these methods have two main problems. First, these methods do not take into account the inherent physics mechanism, their unexplained property obstacles deep learning techniques to be widely applied in clinical. Second, most of these existing algorithms are supervised learning methods. The decomposition performance depends on the quality and quantity of training data pairs. However, high-quality training data pairs are difficult to acquire in clinic, which limits the accuracy of decomposed material images and the generalization performance of models.

To address these problems, we propose a nonlocal spectral similarity-induced material decomposition network (NSSD-Net) for DECT. The NSSD-Net couples the model-based material decomposition model with the deep network and optimizes by the iterative shrinkage-thresholding algorithm (ISTA). Considering the structural similarity and spectral correlation underlying DECT images, the nonlocal spectral similarity-based cost function is designed to guarantee network convergence in a self-supervised manner. The main contributions of the proposed NSSD-Net can be summarized as follows: the first one is that the proposed method fully considers the material decomposition physician mechanism and combines the benefits of model-driven methods and data-driven methods. The second one is that the proposed methods use the nonlocal spectral similarity features as prior information and optimize hyper-parameters by self-supervised learning strategy with no ground-truth data, which enhances the generalization of the proposed model. Experiments on clinical data have confirmed the significant decomposition performance of the NSSD-Net.

This work was supported in part by the NSFC under Grant U21A6005 and Grant U1708261, the National Key R&D Program of China under Grant No.2020YFA0712200. *Corresponding author: J. Ma.*

L. Wang, Y. Wang, Z. Bian, D. Zeng, and J. Ma are with School of Biomedical Engineering, Southern Medical University, Guangzhou 510515, China.(e-mail: jhma@smu.edu.cn)

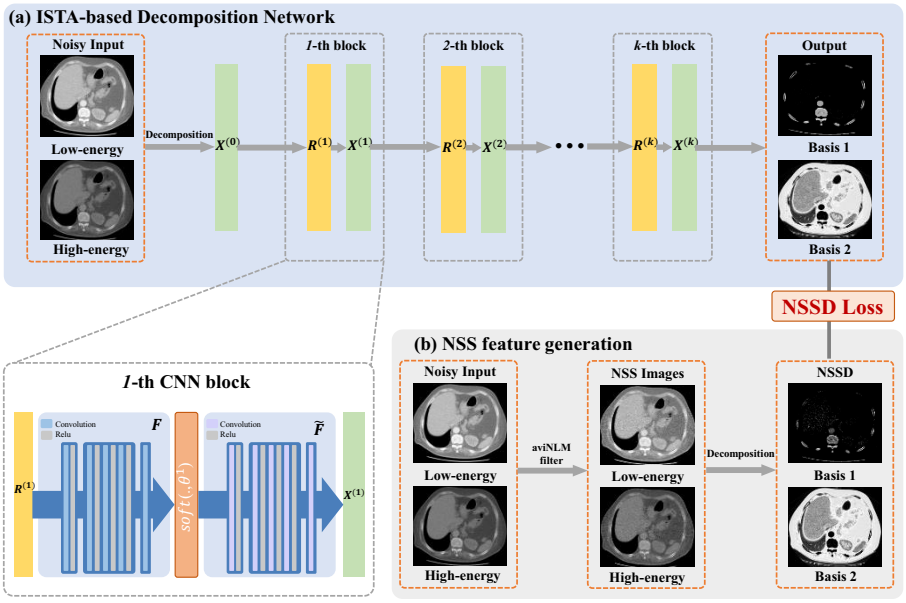


Fig. 1. The overall framework of the proposed NSSD-Net.

II. Methods

A. Model-driven material decomposition model

According to image-based decomposition theory, the linear combination of pixel values in the basis material images can represent the linear attenuation coefficient of each pixel in the input images. Based on the above assumption, the formulation of the material composition for DECT can be written as:

$$U = AX, \tag{1}$$

where $U = [\mu_H, \mu_L]^T$ denotes the linear attenuation coefficient of high and low energy spectrum, $X = [x_1, x_2]^T$ denotes the normalized densities of the basis materials, A is the decomposition matrix and can be represented as follows:

$$A = \begin{pmatrix} \mu_{M1H} I & \mu_{M2H} I \\ \mu_{M1L} I & \mu_{M2L} I \end{pmatrix}, \tag{2}$$

where μ_{Mij} denotes the mass attenuation coefficient of the basis materials, I is an identity matrix with the dimension of $2N$-by-$2N$, and N is the total number of pixels in one CT image.

Direct decomposition generating decomposed material images with severely degraded SNR, especially at low-dose levels. This is because there is a large condition number on the matrix A, and this can be treated as an ill-posed problem, which makes the decomposition process is sensitive to the noise on the raw CT images. To alleviate this situation, least-square estimation model with regularization $\mathcal{R}(X)$ is introduced to suppress decomposed material images noise, which can be expressed as follows:

$$X^* = \underset{X}{\operatorname{argmin}} \|AX - U\|_2^2 + \lambda \mathcal{R}(X), \tag{3}$$

where $\mathcal{R}(X) = \|X\|_1$ and λ is the hyper-parameters of regularization term.

ISTA is a prevailing framework to optimize the decomposition solvation process with non-smooth regularizers. Each iteration of ISTA involves gradient descent update the decomposed material images followed by a shrinkage-threshold step:

$$X^{(k)} = \arg\min_X \left\| X - R^{(k)} \right\|_2^2 + \lambda \|X\|_1, \tag{4}$$

$$R^{(k)} = X^{(k-1)} - \rho A^T \left(AX^{(k-1)} - U \right), \tag{5}$$

where ρ is the step size, $X^{(k)}$ and $R^{(k)}$ denote the intermediate variables of the decomposed results, and λ is the hyper-parameter of the regularization term. However, the model-based material decomposition methods are computationally expensive, and it is difficult to select the optimal parameters (i.e., ρ and λ) and the global optimal solution in practical applications.

B. The proposed NSSD-Net

To address these problems, we propose a self-supervised nonlocal spectral similarity-induced material decomposition network. As shown in Fig. 1, the proposed NSSD-Net consists of two parts: (a) ISTA-based decomposition network and (b) NSS feature generation.

(a) ISTA-based decomposition network: Inspired by the powerful representational capability and universal approximation of CNN, we use the CNN model $F(\cdot)$ stands for the material decomposition nonlinear function to suppress the noise of images. Thus, the ISTA-based material decomposition model can be formulated as:

$$X^{(k)} = \arg\min_X \left\| F(X) - F\left(R^{(k)} \right) \right\|_2^2 + \theta \|F(X)\|_1, \tag{6}$$

$$R^{(k)} = X^{(k-1)} - \rho A^T \left(AX^{(k-1)} - U \right), \tag{7}$$

where θ is a learnable parameter. Therefore, the k-th iteration of the decomposed material images $X^{(k)}$ can be efficiently computed in closed-form as follows:

$$X^{(k)} = \tilde{F}\left(\text{soft}\left(F\left(R^{(k)}\right), \theta^{(k)}\right)\right), \tag{8}$$

where $\tilde{F}(\cdot)$ denotes the left inverse of $F(\cdot)$, and has a symmetrical structure with $F(\cdot)$.

(b) NSS feature generation: Considering the intrinsic characteristics underlying DECT images, i.e., structural similarity and spectrum correlation, which characterizes the priors of desired DECT images. Based on the success of the previous work [4], in this work, we utilize this prior information to generate the nonlocal spectral similarity (NSS) features to enhance the stability of the ISTA-based iterative decomposition network. Specifically, the proposed method is designed by averaging the acquired DECT images $U = [\mu_H, \mu_L]^T$ to obtain the average image U_{avi}, which takes the image similarity within the two energies into consideration. Then, we use the aviNLM method [4] to denoise the noisy DECT images U and estimate the decomposed material image X_p by the image-domain-based decomposition algorithm. The decomposed material image X_p can provide the prior information to guide the ISTA-based iterative decomposition network. Thus, the proposed nonlocal spectral similarity-based cost function can be represented as:

$$L_{NSSD} = \frac{1}{N}\sum_{i=1}^{N}\left\|{X_i}^{(N_t)} - X_p\right\|_2^2, \tag{9}$$

where N and N_t are the number of training data and basic CNN module, respectively. ${X_i}^{(N_t)}$ is the decomposed result of the NSSD-Net.

C. Loss Function

NSSD-Net takes the noisy DECT image $\{U_i\}_{i=1}^N$ and NSS feature-induced decomposed material image $\{X_p\}_{i=1}^N$ as input, and generates the decomposed material image, denoted by ${X_i}^{(N_t)}$ as output. We optimize the proposed network by the self-supervised learning strategy and the total loss function can be expressed as:

$$L_c = \frac{1}{N_t N}\sum_{i=1}^{N}\sum_{k=1}^{N_t}\left\|\tilde{F}^{(k)}\left(F^{(k)}\left(X_i\right)\right) - X_p\right\|_2^2, \tag{10}$$

$$L_{\text{total}} = L_{NSSD} + \tau L_c. \tag{11}$$

where τ is a constant and we set $\tau = 1$ in this work.

D. Experimental Dataset and Settings

The clinical patient study was conducted to assess the proposed NSSD-Net performance. After writing the informed consent from 29 volunteer patients, clinical data were acquired from the GE Discovery CT750 HD scanner with 140kVp and 80kVp. The virtual monochromatic spectral (VMS) images were generated using the GE commercial software. In this work, the VMS images at 90kev and 140kev were selected as the DECT images. In addition, the linear attenuation coefficient of basis materials according to the VMS images reported by the National Institute of Standards and Technology (NIST). Then, we simulated the noisy DECT data by the previous work [5] to further analyze the noise suppression performance among different material decomposition methods.

In the experiment, 1000 simulated noisy VMS image pairs (i.e., 90kev and 140kev) from 22 patients were used for training and 188 simulated noisy VMS images pairs from the remaining patients were used for testing. The network was implemented on an NVIDIA Tesla P40 GPU based on the PyTorch framework and using the Adam optimizer. The initial learning rate, batch size, epoch, and iterations were set to $1e^{-3}$, 1, 50, and 100, respectively. During training, the number of CNN basic module was set to 8. According to experience, the initial gradient descent step $\rho = 0.5$, the hyper-parameter $\tau = 1$, and the threshold $\theta = 0.1$.

III. Results

Fig. 2 shows the decomposed material images of different methods for two test cases. The first column is the ground-truth decomposed material images. It can be observed that both direct matrix inversion method (DIMD) and statistic iterative based decomposition method (Iterative MD) failed to properly distinguish basis materials, and the decomposed material images were heavily corrupted by the noise and noise-induced artifacts. Compared to the DIMD and Iterative MD, the data-based material decomposition methods (i.e., CD-Convnet and ISTA-Net) remove the noise and artifacts further and provide accurately estimates of different materials. However, compared to NSSD-Net, the data-based material decomposition methods tend to over-smoothen the images, especially in the structural edge and soft tissue area. The results demonstrate that the self-supervised nonlocal spectral similarity prior can provide the spectral data structure feature information and effectively guide the material decomposition process. Fig. 3 shows the profile comparison indicated by the orange lines in Fig. 2. It can be judged that the NSSD-Net achieves a good balance between bias control and noise suppression.

In order to quantitatively validate the decomposed performance of the proposed NSSD-Net, the peak signal-to-noise ratio (PSNR), structure similarity index measure (SSIM) and root mean square error (RMSE) metrics are calculated. As shown in Tab. I, both NSSD-Net results (i.e. water and bone equivalent fractions), achieve relative superior quantitative metrics than the other decomposition methods expect for the ISTA-Net. The possible reason is that the ISTA-Net utilize the fully supervised manner to train the network, while the proposed NSSD-Net trained without the labeled data. It is worth noting that the proposed method outperforms the ISTA-Net in terms of the SSIM and confirm that the NSSD-Net can provide the global and local structure information from DECT images.

IV. Conclusion

In this paper, we propose a novel material decomposition framework for DECT based on the self-supervised nonlocal

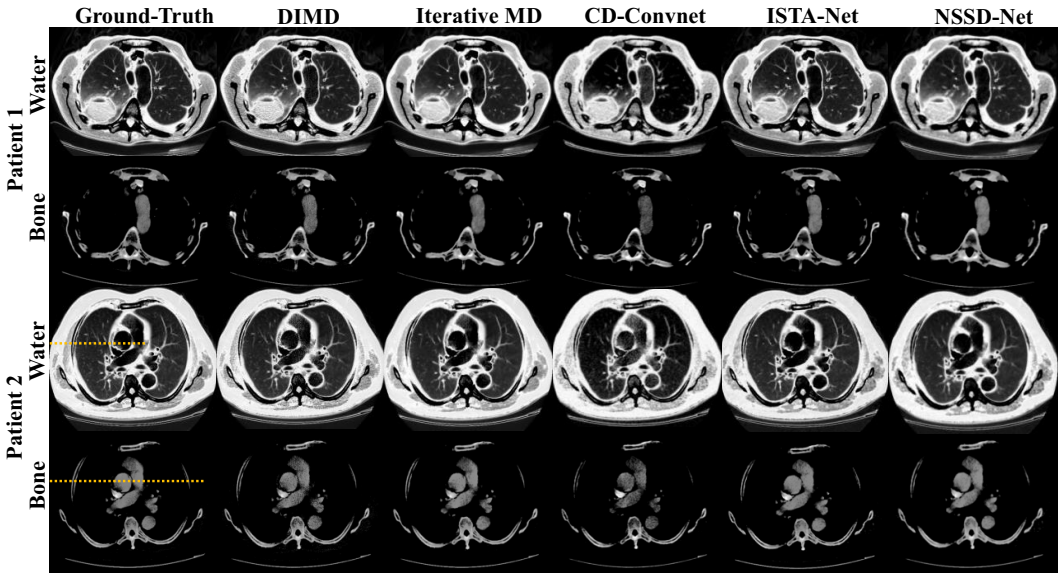


Fig. 2. The decomposed material images of different methods: the first and third row shows the water decomposed images with the display window [0, 1] $g.cm^{-3}$; the second and fourth row shows the bone decomposed images with the display window [0.5 1.5] $g.cm^{-3}$.

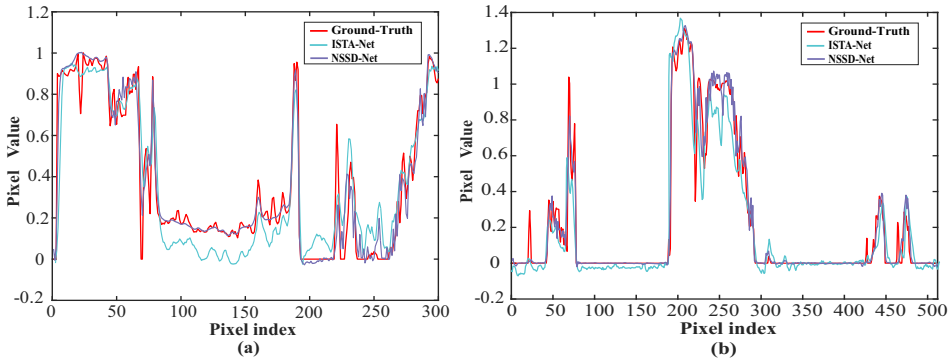


Fig. 3. Profile comparison of the ISTA-Net and NSSD-Net along the orange line indicated in the Fig. 2: (a) water decomposed image, (b) bone decomposed image.

TABLE I. Quantitative results of different material decomposition methods.

Methods		PSNR	SSIM	RMSE
DIMD	bone	22.1648 ± 1.8508	0.7946 ± 0.0995	8.8992 ± 3.0836
	water	21.6688 ± 2.0852	0.8072 ± 0.0866	9.1729 ± 2.9362
Iterative MD	bone	27.6527 ± 1.0160	0.9103 ± 0.0392	3.6492 ± 1.1476
	water	24.0065 ± 0.8603	0.8932 ± 0.0385	4.3312 ± 1.0572
CD-Convnet	bone	28.1554 ± 0.8808	0.9264 ± 0.0198	4.0299 ± 0.8422
	water	29.2132 ± 0.3138	0.8647 ± 0.0365	6.1902 ± 2.0852
ISTA-Net	bone	**32.5106 ± 1.0177**	0.9602 ± 0.0142	**1.4208 ± 0.6827**
	water	**32.6545 ± 0.7508**	0.9684 ± 0.0251	**2.6060 ± 0.8472**
NSSD-Net	bone	31.4440 ± 1.0940	**0.9631 ± 0.0215**	1.6329 ± 0.6989
	water	31.3831 ± 0.8861	**0.9701 ± 0.0527**	2.9008 ± 0.5268

spectral similarity prior. Specifically, the proposed NSSD-Net combines the traditional model-based decomposition model with the data-based method and optimizes the objective function by the nonlocal spectral similarity prior. The experimental results with clinical data show that the NSSD-Net can obtain accurate decomposed material images of dual-energy.

REFERENCES

[1] W. A. Kalender, E. Klotz, and L. Kostaridou, "An algorithm for noise suppression in dual energy CT material density images." *IEEE Transactions on Medical Imaging*, vol. 7, no. 3, pp. 218-224, 1988.

[2] N. Ducros, J. Abascal, and B. Sixou, "Regularization of nonlinear decomposition of spectral x-ray projection images." *Medical Physics*, vol. 44, no. 9, pp. e174-e187, 2017.

[3] Z. Shi, H. Li, and Q. Cao, "A material decomposition method for dual-energy CT via dual interactive Wasserstein generative adversarial networks." *Medical Physics*, vol. 48, no. 6, pp. 2891-2905, 2021.

[4] H. Zhang, D. Zeng, J. Lin, and J. Ma, "Iterative reconstruction for dual energy CT with an average image-induced nonlocal means regularization." *Physics in Medicine and Biology*, vol. 62, no. 13, pp. 5556-5574, 2017.

[5] D. Zeng, J. Huang, Z. Bian, and J. Ma, "A Simple Low-Dose X-Ray CT Simulation From High-Dose Scan." *IEEE Transactions on Nuclear Science*, vol. 62, no. 5, pp. 2226-2233, 2015.

Likelihood-based bilateral filtration in material decomposition for photon counting CT

Okkyun Lee*

***Abstract*—The maximum likelihood (ML) principle has been a gold standard for estimating basis line-integrals due to the optimal statistical property. However, the estimates are sensitive to noise from large attenuations or low dose levels. One may apply filtering in the estimated basis sinograms or model-based iterative reconstruction. Both methods effectively reduce noise, but the degraded spatial resolution is a concern. In this study, we propose a likelihood-based bilateral filter (LBF) for the estimated basis sinograms to reduce noise while preserving spatial resolution. It is a post-processing filtration applied to the ML-based basis line-integrals, the estimates with a high noise level but minimal degradation of spatial resolution. The proposed filter considers likelihood in neighbours instead of weighting by pixel values as in the original bilateral filtration. Two-material decomposition (water and bone) results demonstrate that the proposed method shows improved noise-to-spatial resolution tendency compared to conventional methods.**

***Index Terms*—Photon counting CT, material decomposition, maximum likelihood, bilateral filter**

I. Introduction

Energy discriminating photon counting detector (PCD) enables estimating line-integrals of basis materials. It provides various potential applications in PCD-based computed tomography (CT) compared to the conventional energy integrated detector-based CT [1]. The maximum likelihood (ML) principle has been a gold standard for estimating basis line-integrals in PCD-CT [2, 3]. It has the optimal statistical property–unbiased and achieves the Cramér-Rao lower bound (CRLB), the minimum noise for an unbiased estimator. However, the noise level in the estimated basis sinograms and the associated reconstructed basis images may not be satisfactory for clinical purposes [4, 5]. When the dose level is low, or the attenuation is large, the estimates greatly suffer from noise [4, 6]. One may apply low-pass filtering (LPF) in the estimated basis sinograms or statistical iterative reconstruction (SIR) by regularizing neighboring pixels in the image domain while exploiting noise statistics in the forward model [4, 5, 7]. Both approaches effectively reduce the noise in the basis images, but the degraded spatial resolution is a concern.

In this study, we propose a likelihood-based bilateral filter (LBF) to reduce noise while preserving spatial resolution. Bilateral filter (BF) is an effective edge-preserving method for image denoising[8, 9]. In a region of interest (ROI; the central pixel is the one we want to correct), it gives more weight for a neighborhood pixel whose value is more similar to that of the centered one, in addition to a pixel distance weight (e.g., Gaussian shape). The method is effective when the central pixel in ROI is a reliable reference for filtering purposes. However, the estimated basis sinograms may be too noisy to apply in practice (we validate it in the simulation study later). Instead of the pixel value-dependent weighting, we apply the neighboring likelihood-dependent one. The proposed method first estimates basis sinograms using the ML principle. It then calculates log-likelihoods using the estimates in the neighboring pixels for given PCD measurements at the centered one. It gives more weight to the neighbors whose likelihoods are greater than others. It is possible that pixel values that are very different from the center one (the ML estimates) may produce likelihoods close to the ML, so the method reflects the statistical distance rather than the distance between noisy estimates.

We validate the proposed method for a two-material decomposition (water and bone) and compare it to conventional methods. The ML-based estimation for basis sinograms is the starting point for all the methods we consider here. We used Gaussian and the original bilateral filters in the estimated basis sinograms as the conventional methods. We also applied model-based iterative reconstruction with edge-preserved penalty in the image domain. Both noise-reduced images (basis images and CT images) and noise-FWHM (full width at half maximum) behavior demonstrate that the proposed method effectively reduces noise without sacrificing spatial resolution compared to the conventional ones.

*O. Lee is with the Department of Robotics Engineering, Daegu Gyeongbuk Institute of Science and Technology (DGIST), Daegu, Republic of Korea (oklee@dgist.ac.kr).

II. Methods

A. The ML-based Estimation for Basis Sinograms in PCD-CT

The log-likelihood of unknown basis line-integrals for the j-th sinogram point ($\mathbf{x}_j = [x_{1,j}, x_{2,j}, \cdots, x_{M,j}]^T \in \mathbb{R}^{M\times 1}\,[\mathrm{cm}]$) for a given PCD measurements $\mathbf{y}_j = [y_{1,j}, y_{2,j}, \cdots, y_{B,j}]^T \in \mathbb{R}^{B\times 1}\,[\mathrm{counts}]$ is given by (without constant terms),

$$\ln \mathcal{L}(\mathbf{x}_j|\mathbf{y}_j) = \sum_{b=1}^{B} y_{b,j} \ln \lambda_b(\mathbf{x}_j) - \lambda_b(\mathbf{x}_j), \tag{1}$$

where $y_{b,j}$ is the number of photons measured at the b-th energy bin of the PCD pixel, B is the total number of energy bins, and M is the number of basis materials. We assume that the counts in energy bins are independent Poisson random variables with the expected value ($\lambda_b(\mathbf{x}_j)$) as follows:

$$\lambda_b(\mathbf{x}_j) = \int S_b(E) e^{-\sum_{m=1}^{M} x_{m,j}\phi_m(E)} dE, \tag{2}$$

for $b = 1, 2, \cdots, B$, where $S_b(E)\,[\mathrm{counts/keV}]$ is the bin-wise energy-dependent function, including incident spectrum, and detector response function [2] and $\phi(E)\,[\mathrm{cm}^{-1}]$ is the

energy-dependent basis function. The ML-based estimation can be described as follows:

$$\hat{\mathbf{x}}_j^{ML} = \arg\min_{\mathbf{x}} -\ln \mathcal{L}(\mathbf{x}|\mathbf{y}_j), \text{ s.t. } x_m^{\min} \leq x_m \leq x_m^{\max}, \quad (3)$$

for $m = 1, 2, \cdots, M$. It includes a box constraint to the line-integrals to restrain excessively large estimates. After the estimation for all detector channels and projection views, one can reconstruct the basis images by applying e.g., filtered backprojection (FBP).

B. Proposed Likelihood-based Bilateral Filtering (LBF)

The conventional bilateral filtration (BF) we applied to the basis sinograms is given by,

$$\hat{\mathbf{x}}_j^{BF} = \alpha(j) \sum_{i \in \mathcal{N}_j} \hat{\mathbf{x}}_i^{ML} \mathcal{G}(i, j, \sigma_1) \mathcal{H}(i, j, \sigma_2), \quad (4)$$

where $\mathcal{N}_j$ is the set of neighboring sinogram pixels centered by the j-th pixel and $\alpha(j)$ is a normalization factor. The first weighting function $\mathcal{G}(i, j, \sigma_1)$ is a spatial low-pass filter (LPF), and we used a Gaussian filter with a standard deviation σ_1. The second weighting function is given by $\mathcal{H}(i, j, \sigma_2) = \exp\left\{-\frac{\|\hat{\mathbf{x}}_i^{ML} - \hat{\mathbf{x}}_j^{ML}\|^2}{\sigma_2^2}\right\}$, which has a control parameter σ_2, where $\|\cdot\|$ denotes the standard ℓ_2 norm (Euclidean distance).

We propose the likelihood-based bilateral filtration (LBF) as follows:

$$\hat{\mathbf{x}}_j^{LBF} = \beta(j) \sum_{i \in \mathcal{N}_j} \hat{\mathbf{x}}_i^{ML} \mathcal{G}(i, j, \sigma_1) \mathcal{H}_\ell(i, j, \sigma_2), \quad (5)$$

where the second weighting function is changed to $\mathcal{H}_\ell(i, j, \sigma_2) = \exp\left\{-\frac{(\ln \mathcal{L}(\mathbf{x}_i^{ML}|\mathbf{y}_j) - \ln \mathcal{L}(\mathbf{x}_j^{ML}|\mathbf{y}_j))^2}{\sigma_2^2}\right\}$, and again $\beta(j)$ is a normalization factor.

C. Simulation Settings and Conventional Methods

We used an incident x-ray spectrum ($140\,\mathrm{kVp}$) with a $5\,\mathrm{mm}$ aluminium pre-filtration [10]. A numerical thorax phantom was used, as shown in Fig. 1 (512×512 with a pixel size $0.75 \times 0.75\,\mathrm{mm}$) [11]. We used water ($1\,\mathrm{g/cm^3}$) and bone ($0.925\,\mathrm{g/cm^3}$) as the two basis functions for a two-material decomposition. We applied a fan-beam geometry (equidistance) and used the ASTRA toolbox for the projection and backprojection [12]. The source to center and center to detector were both $50\,\mathrm{cm}$. The number of views per rotation was 1000. The number of PCD pixels was 998 with a detector pitch of $0.1\,\mathrm{cm}$. We used four energy bins $20, 67, 77, 95\,\mathrm{keV}$ with minimum CRLB when the basis line-integrals of water and bone were 30 and $6\,\mathrm{cm}$, respectively [13]. The total number of incident photons was 5×10^5 for each ray. The minimum photon numbers measured at PCD with the phantom were around 350 (summed for all energy bins).

As illustrated in Fig. 1 (the first row), we measured FWHM for the reconstructed water and CT images (indicated by arrows) and also standard deviation indicated by the rectangles (ROIs 1 and 2). The ROI2 was selected to examine the severe streak artifacts in the image. The standard deviation was calculated for a single noise trial for each method and the FWHM using an average image from 100 noise trials. We also showed the reconstructed two basis images and the synthesized CT image for a single noise trial and corresponding bias images. The ground truth was set as the FBP (Shepp-Logan) images following the ML estimates using a substantially high number of incident photons (10^9).

We compared the proposed LBF to the ML estimates, the original BF, and the LPF given by,

$$\hat{\mathbf{x}}_j^{LPF} = \sum_{i \in \mathcal{N}_j} \hat{\mathbf{x}}_i^{ML} \mathcal{G}(i, j, \sigma_1). \quad (6)$$

For the ML estimates, we used the derivative-free Nelda-Mead method with the box constraint for water and bone, $-2 \leq x_{water} \leq 40\,\mathrm{cm}$ and $-5 \leq x_{bone} \leq 15\,\mathrm{cm}$, respectively. We set $\sigma_1 = 0.9\,[\mathrm{pix}]$ with a 3×3 window for BF and LBF, and we varied σ_2. For LPF, we also used a 3×3 window but with different values of σ_1. Once the filtration was applied to the basis sinograms, we also combined them to synthesize a CT image (at $65\,\mathrm{keV}$).

We also applied statistical iterative reconstruction (SIR) to the ML estimates as described in [7]. We used edge preserved pseudo-Huber penalty [14] with the separable surrogate-based monotonic algorithm [15]. The detailed algorithm is given in [5] (Appendix C). We set the maximum iterations as 2000 and terminated when the relative change in the cost function is less than 2×10^{-4}. There are two control parameters in SIR: δ for the shape of the pseudo-Huber penalty and λ for the regularization parameter (weight on the penalty term). We tried various combinations of δ and λ and selected the best one by observing the reconstructed images and the FWHM and noise measures.

III. Results

We present the reconstructed basis images and the CT image overlapped by the bias image in Fig. 1 for various methods. For basis images (water and bone), we selected the results showing similar noise levels in ROI1 for all the methods (except the ML estimates; see the arrows in Fig. 2(a)). For the CT image, we also selected the ones with similar noise levels in ROI2 (see the arrows in Fig. 2(b)). Thus, the CT image in Fig. 1 may not represent the linear combination of the water and bone images as shown in the same figure. The images from the ML estimates are noisier than other methods, but there is no bias at edges inside the phantom. The LPF is effective in reducing the streak artifacts in the basis images, but it shows noticeable bias at edges for all basis and CT images. The images using the BF also show substantial bias at the edges, and the streak artifact is larger than the LPF. However, the bias at the edges from the proposed LBF is smaller than those of the LPF and BF, and the reduced streak artifact is qualitatively similar to the LPF. It also clearly demonstrates the effectiveness of the proposed LBF over the original BF (statistical difference versus pixel value difference). The SIR shows substantial bias at the edges in the CT image compared to the basis images. Note that all the CT images have similar noise levels at ROI2, and it clearly shows the trade-off between spatial resolution and noise, compared to the results in LBF.

We show the noise-FWHM plots measured as indicated in Fig. 1 for the water and CT images in Fig. 2(a) and (b),

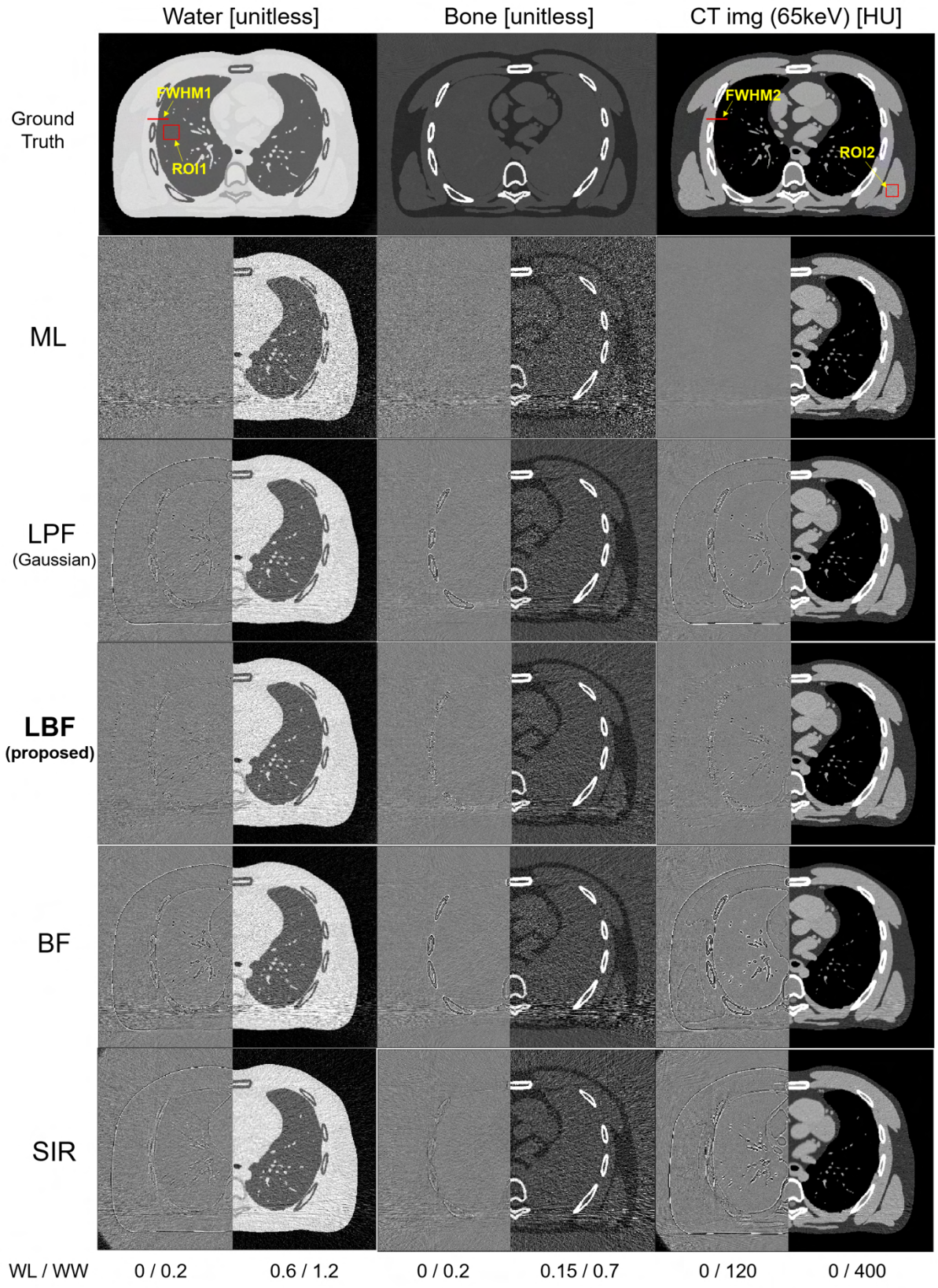


Fig. 1: Reconstructed basis and CT images using various methods. For each image, the left half is the bias calculated from 100 noise trials, and the right half is the result of a single noise trial.

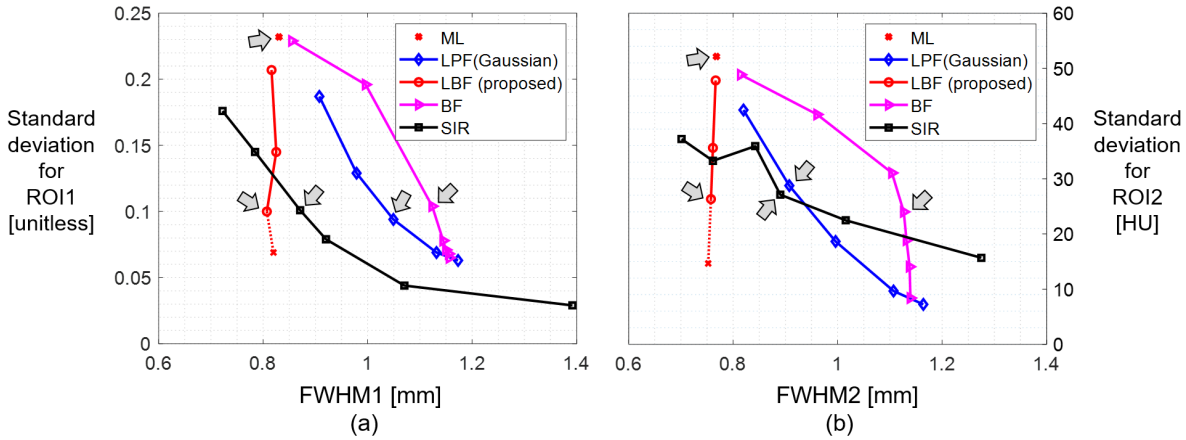


Fig. 2: Noise-FWHM plots calculated from the reconstructed water image (ROI1 and FWHM1) and CT image (ROI2 and FWHM2) for various methods. The arrows indicate the points we selected to show in Fig. 1.

respectively. For the water image (FWHM1 and ROI1), the FWHM from the ML is around $0.8\,\mathrm{mm}$, but it has the largest noise among the methods. The LPF can substantially reduce the noise by sacrificing spatial resolution. The tendency is similar for the BF, but it degrades the resolution more than the LPF for reducing the same amount of noise. Interestingly, the proposed LBF reduces noise without sacrificing the resolution (even slight improvement). However, for a large value of σ_2, we could see slight but visible artifacts between bones in the average image (from 100 noise trials). It corresponds to the last point in the curve, and we indicated it with a cross mark (dotted line). The SIR presents better property on the noise-FWHM than the LPF, but the proposed method shows a slightly smaller resolution than the SIR at some point (indicated by arrow). The overall tendency in the CT image (FWHM2 and ROI2–streak artifacts; Fig. 2(b)) is similar to the water image. However, the noise-FWHM relation of SIR is not as superior as shown in Fig. 1(a). We searched for all the possible combinations of regularization parameters used for water and bone basis images, but we might have missed the (unknown) optimal parameters.

IV. Conclusions

We proposed the likelihood-based bilateral filtration in basis sinograms for photon counting CT. The method can be applied to the estimated basis sinograms from the ML principle. It considers the neighborhood likelihoods in ROI for the weighting function instead of the pixel values (estimates). We demonstrated that the proposed method is more effective in reducing noise without sacrificing the spatial resolution, compared to the original bilateral filter and other conventional methods.

References

[1] K. Taguchi and J. S. Iwanczyk, "Vision 20/20: single photon counting x-ray detectors in medical imaging," *Med. Phys.*, vol. 40, no. 10, pp. 100 901(1–19), Oct. 2013.

[2] J. P. Schlomka, E. Roessl, R. Dorscheid, S. Dill, G. Martens, T. Istel, C. Bäumer, C. Herrmann, R. Steadman, G. Zeitler, A. Livne, and R. Proksa, "Experimental feasibility of multi-energy photon-counting K-edge imaging in pre-clinical computed tomography," *Phys. Med. Biol.*, vol. 53, pp. 4031–4047, 2008.

[3] O. Lee, S. Kappler, C. Polster, and K. Taguchi, "Estimation of basis line-integrals in a spectral distortion-modeled photon counting detector using low-rank approximation-based x-ray transmittance modeling: K-edge imaging application," *IEEE Trans. on Medical Imaging*, vol. 36, no. 11, pp. 2389–2403, Nov. 2017.

[4] S. Rit, C. Mory, and P. B. Noël, "Image formation in spectral computed tomography," in *Spectral, Photon Counting Computed Tomography*, K. Taguchi, I. Blevis, and K. Iniewski, Eds. Boca Raton, FL: CRC Press, 2020, pp. 355–371.

[5] S. Baek and O. Lee, "A data-driven maximum likelihood classification for nanoparticle agent identification in photon-counting CT," *Phys. Med. Biol.*, vol. 66, pp. 145 004:1–19, 2021.

[6] J. Nam and O. Lee, "Technical Note: The nearest neighborhood-based approach for estimating basis line-integrals using photon-counting detector," *Med. Phys.*, vol. 48, no. 10, pp. 6531–6535, Oct. 2021.

[7] C. O. Schirra, E. Roessl, T. Koehler, B. Brendel, A. Thran, D. Pan, M. A. Anastasio, and R. Proksa, "Statistical reconstruction of material decomposed data in spectral CT," *IEEE Trans. on Medical Imaging*, vol. 32, no. 7, pp. 1249–1257, July 2013.

[8] A. Buades, B. Coll, and J. M. Morel, "Neighborhood filters and PDE's," *Numerische Mathematik*, vol. 105, no. 1, pp. 1–34, 2006.

[9] H. Zhang, D. Zeng, H. Zhang, J. Wang, Z. Liang, and J. Ma, "Applications of nonlocal means algorithm in low-dose X-ray CT image processing and reconstruction: A review," *Med. Phys.*, vol. 44, no. 3, pp. 1168–1185, March 2017.

[10] J. Punnoose, J. Xu, A. Sisniega, W. Zbijewski, and J. H. Siewerdsen, "Technical Note: spektr 3.0-a computational tool for x-ray spectrum modeling and analysis," *Med. Phys.*, vol. 43, no. 8, pp. 4711–4717, Aug. 2016.

[11] W. P. Segars, M. Mahesh, T. J. Beck, E. C. Frey, and B. M. W. Tsui, "Realistic CT simulation using the 4D XCAT phantom," *Med. Phys.*, vol. 35, no. 8, pp. 3800–3808, Aug. 2008.

[12] W. van Aarle, W. J. Palenstijn, J. Cant, E. Janssens, F. Bleichrodt, A. Dabravolski, J. D. Beenhouwer, K. J. Batenburg, and J. Sijbers, "Fast and flexible x-ray tomography using the ASTRA toolbox," *Opt. Express*, vol. 24, no. 22, pp. 25 129–25 147, October 2016.

[13] E. Roessl and C. Herrmann, "Cramér-Rao lower bound of basis image noise in multiple-energy x-ray imaging," *Phys. Med. Biol.*, vol. 54, pp. 1307–1318, 2009.

[14] P. Charbonnier, L. Blanc-Féraud, G. Aubert, and M. Barlaud, "Deterministic edge-preserving regularization in computed imaging," *IEEE Trans. on Image Processing*, vol. 6, no. 2, pp. 298–311, Feb. 1997.

[15] H. Erdoğan and J. A. Fessler, "Monotonic algorithms for transmission tomography," *IEEE Trans. on Medical Imaging*, vol. 18, no. 9, pp. 801–814, Sept. 1999.

Experimental Evaluation of Polychromatic Reconstruction for Quantitative CBCT

Michał Walczak, Pascal Paysan, Mathieu Plamondon, Stefan Scheib

***Abstract*—Polychromatic reconstruction is a promising technique for quantitative cone-beam computed tomography in radiation therapy. In this study, we have implemented polychromatic forward projection into our reconstruction framework to directly reconstruct relative electron density volumes without the need for additional HU calibration. The underlying spectral model takes beam hardening into account by design. Thereby this extended reconstruction framework is a natural step in the direction of spectral imaging, albeit without any hardware modifications. Reconstructed relative electron density volumes from phantom scans show sufficiently good agreement with ground truth for photon dose calculation; relative errors for most inserts are below 3%. We also demonstrate beam hardening artifact reduction in virtual monoenergetic images obtained from polychromatic reconstruction as compared to an established iterative reconstruction using water-based correction. Similarly, polychromatic reconstruction shows potential for mitigating metal artifacts in a clinical scan acquired for a patient with bilateral hip implants.**

***Index Terms*—Beam hardening, polychromatic reconstruction, quantitative CBCT.**

I. Introduction

High quality cone-beam computed tomography (CBCT) images are essential in image guided radiation therapy for tasks such as soft tissue based patient positioning, structure delineation, and dose calculation. However, due to the polyenergetic nature of X-ray sources, image quality in available single energy reconstructions deteriorates as a result of beam hardening and metal artifacts. Another limitation is that, to determine relative electron density (RED), different HU-RED calibrations are necessary for different acquisition protocols, which is a common obstacle in clinical use. Spectral imaging, e.g., dual energy or photon counting technology, addresses these limitations. However, it generally requires specific hardware solutions. By contrast, leveraging prior knowledge about the polychromatic characteristic of the X-ray beam as well as attenuation properties of different materials can be seen as a middle ground software-based solution. This concept is at the core of polychromatic reconstruction algorithms in the literature [1]–[6].

To directly reconstruct RED volumes, we expanded our non-clinical iterative CBCT (iCBCT) [7] reconstruction pipeline (iTools Reconstruction, Varian Medical Systems, Palo Alto, CA, USA) by implementing a polychromatic forward projection that maps RED volume to intensities in projection space, as described by Mason *et al.* [4], [5]. The core component of the polychromatic forward projection is a piecewise linear approximation of the relation between polychromatic attenuation coefficients and RED values fitted for different materials. Together with an energy-resolved air norm, this polychromatic attenuation model accounts for beam hardening by construction. The model also eliminates the need for using different HU-RED calibrations for different scanning protocols and can be further used to convert reconstructed RED volumes to virtual monoenergetic images (VMI).

Unlike Mason *et al.* [5], we do not use a polyenergetic kernel-based scatter correction, but instead utilize Acuros® CTS, a polychromatic object scatter estimate based on the linear Boltzmann transport equation [8], [9], available in our iCBCT reconstruction pipeline.

In this paper, we demonstrate that our CBCT reconstruction pipeline augmented with the polychromatic model can be used to reliably determine RED directly. Further, we show the benefits of polychromatic reconstruction for reducing beam hardening artifacts and mitigating metal artifacts, as compared to an established iterative reconstruction method.

M. Walczak, P. Paysan, M. Plamondon, and S. Scheib are employees of Varian, a Siemens Healthineers company, Taefernstrasse 7, 5405 Daettwil, Switzerland (e-mail: Michal.mw.Walczak@varian.com)

II. Method

A. Iterative polychromatic reconstruction

At the core of the iterative polychromatic reconstruction implementation is a polychromatic forward projection, adapted from [4]:

$$\mathbf{p}_\alpha = \sum_{j=1}^{N_E} \mathbf{r}_0(E_j) \exp(-[\mathbf{A}\hat{\mu}(\boldsymbol{\rho}_e, E_j)]_\alpha) \tag{1}$$

where $\mathbf{p}_\alpha$ is simulated projection under angle α, $\mathbf{r}_0(E_j)$ is spectral air norm, $\mathbf{A}$ is forward-projection operator, and $\hat{\mu}(\boldsymbol{\rho}_e, E_j)$ is a polychromatic attenuation model transforming the relative electron density volume $\boldsymbol{\rho}_e$ into attenuation coefficients μ. The polychromatic forward projection combines thus prior spectral knowledge about the imaging system (spectral air norm) and material characteristics (polychromatic attenuation model).

The spectral air norm $\mathbf{r}_0(E_j)$ resolves the (energy-integrated) air norm after bowtie filtration $\mathbf{r}_{BT}$ according to spectral characteristics of the imaging system:

$$\mathbf{r}_0(E_j) = \frac{\mathbf{r}_{BT} \circ \mathbf{S}(E_j)}{\sum_{j=1}^{N_E} \mathbf{S}(E_j)} \tag{2}$$

where $\circ$ denotes element-wise product, $\mathbf{S}(E_j)$ is the system sensitivity. It combines source and detector properties:

$$\mathbf{S}(E_j) = I_s(E_j) F_s(E_j) \mathrm{e}^{-\mu_{\mathrm{Ti}}(E_j) l_{\mathrm{Ti}} - \mu_{\mathrm{Al}}(E_j) \mathbf{l}_{\mathrm{Al}}} \eta_d(E_j) E_j \tag{3}$$

where $I_s(E_j)$ is the spectrum flux at the source, $F_s(E_j)$ is inherent filtration at the source, μ_{Ti} and l_{Ti} are respectively attenuation coefficient and thickness of the titanium filter, μ_{Al} and $\mathbf{l}_{\text{Al}}$ are respectively attenuation coefficient and thickness profile of the aluminium bowtie filter, and $\eta_d(E_j)$ is the detector efficiency.

The polychromatic attenuation model takes the form of a piecewise linear fit [4]:

$$\hat{\mu}(\boldsymbol{\rho}_e, E_j) = \sum_{i=1}^{N_f} f_i(\boldsymbol{\rho}_e) \circ [\alpha_i(E_j)\boldsymbol{\rho}_e + \beta_i(E_j)] \qquad (4)$$

where $f_i(\boldsymbol{\rho}_e)$ is a material class identification function, $\circ$ denotes element-wise product, $\alpha_i(E_j)$ and $\beta_i(E_j)$ are energy-dependent fit coefficients. The material class identification function maps relative electron density values to corresponding fit segments [4]:

$$f_i(\boldsymbol{\rho}_e) = \begin{cases} 1 & \text{if } k_{i-1} \leq \boldsymbol{\rho}_e \leq k_i \\ 0 & \text{otherwise} \end{cases} \qquad (5)$$

where k_i marks a break point between the i and $i+1$ fit segments, with $k_0 = 0$ and $k_{N_f} = \infty$. As described in [4], model parameters $\alpha_i(E_j)$, $\beta_i(E_j)$ and k_i are obtained via a fit to human tissue materials from ICRP Publication [10], albeit in extended energy range from 15 keV to 140 keV.

Similarly to our established iCBCT reconstruction [7], for RED volume $\boldsymbol{\rho}_e$ optimization, we use a penalized likelihood cost function with total variation (TV) regularization [11] with a modified expression for the data-fidelity gradient:

$$\nabla_\alpha^{\text{DF-PL}} = \mathbf{A}^{\text{T}}(\mathbf{q}_\alpha - \mathbf{p}_\alpha) \qquad (6)$$

where $\mathbf{A}^{\text{T}}$ is the back-projection operator acting on the difference between measured projection $\mathbf{q}_\alpha$ and the polychromatic forward projection $\mathbf{p}_\alpha$ defined in (1). To accelerate the polychromatic reconstruction, we utilize separable quadratic surrogate cost function [12], ordered subsets [13], and Nesterov momentum method [14].

B. Reconstruction pipeline

The polychromatic reconstruction pipeline is a natural extension of our iCBCT reconstruction pipeline, which is a two-pass approach [7]. In the first pass of the iCBCT pipeline, the initial HU volume is reconstructed with algebraic reconstruction technique (ART) from fASKS [15] scatter-corrected projections. In the second pass, the initial volume is further refined with statistical reconstruction based on penalized maximum likelihood (PL) using projection images scatter-corrected with Acuros® CTS.

The iterative polychromatic reconstruction (IPR) requires pre-processed projection images. The pre-processing step comprises object and hardware scatter corrections. In our two-pass polychromatic reconstruction pipeline, depicted in Fig. 1, scatter correction is done at two stages. The first pass fASKS scatter correction is carried out identically as in the iCBCT pipeline, and provides input for initial ART reconstruction. In the second pass, scatter correction is done by Acuros® CTS, which consumes the initially reconstructed HU volume to solve the linear Boltzmann transport equation. The scatter-corrected projection images are then fed into IPR to reconstruct the final RED volume.

III. Results

A. Direct RED reconstruction

To demonstrate the use of the polychromatic model for direct RED reconstruction, we scanned the head insert of a CBCT electron density phantom (Model 062MA, Computerized Imaging Reference Systems, Inc., Norfolk, VA, USA) on a Varian Halcyon™ machine (Varian Medical Systems, Palo Alto, CA, USA). The scan was acquired in full-fan geometry at 125 kVp and with kV blades in the longitudinal direction collimated to the height of the phantom electron density plugs. RED volume was reconstructed with 5 initial iCBCT ART iterations, followed by 50 IPR iterations on projections scatter-corrected with Acuros® CTS. Number of iterations was determined empirically for better RED accuracy and image quality.

Central slice through reconstructed RED volume is depicted in Fig. 2a and the resulting difference to ground truth (GT) values relative to water is shown in Fig. 2b. As GT, we assigned nominal RED values from the phantom specification. Mean reconstructed RED values within each plug (areas indicated by dashed circles in Fig. 2a) are listed in Table I. Except for 'Lung inhale' and 'Dense bone 800' plugs, tissue materials were reconstructed with relative RED error below 3%, thus enabling accurate treatment dose calculation.

TABLE I: Mean reconstructed RED values for inserts at central slice and their differences to the ground truth (GT) relative to water

Insert	GT RED	reconstructed RED	relative difference %
(A) Trabecular bone 200	1.12	1.146	-2.6
(B) Liver	1.048	1.038	1.0
(C) Lung exhale	0.504	0.505	-0.1
(D) Adipose	0.95	0.937	1.3
(E) Solid dense bone 800	1.441	1.533	-9.2
(F) Muscle	1.051	1.038	1.3
(G) Lung inhale	0.185	0.148	3.7
(H) Breast	0.977	0.967	1.0
(I) Distilled water	1.0	0.999	0.1

B. Beam hardening artifact reduction

To illustrate the benefit of polychromatic reconstruction for beam hardening artifact reduction, we scanned the CIRS CBCT electron density phantom with bone inserts (from left to right: solid dense bone 800, solid dense bone 1750, solid trabecular bone 200, solid dense bone 800, solid dense bone 1250) aligned in one row. The phantom was scanned on a Varian Halcyon™ machine in full-fan beam geometry at 125 kVp with kV blades in the longitudinal direction collimated to the

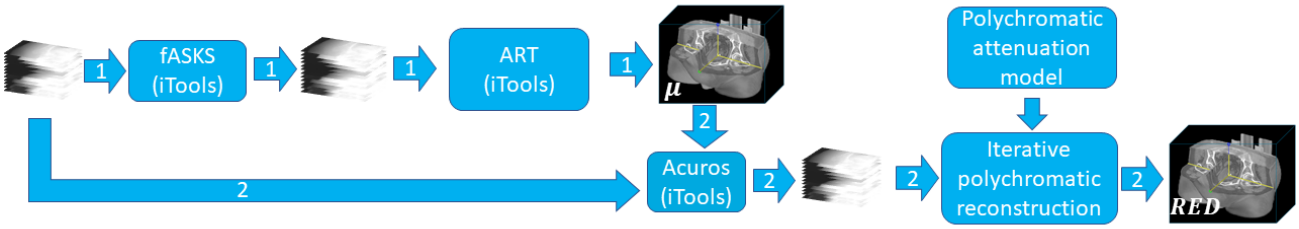
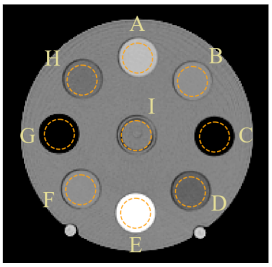
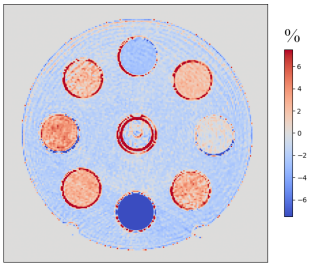


Fig. 1: Diagram showing a two-pass polychromatic reconstruction pipeline. In the first pass, algebraic reconstruction technique (ART) consumes projections corrected for scatter with a kernel-based approach (fASKS). HU volume from first-pass provides input for Acuros® CTS scatter correction for subsequent second-pass iterative polychromatic reconstruction (IPR).

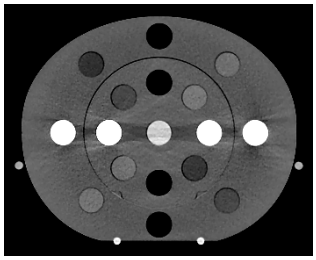


(a) Reconstructed RED

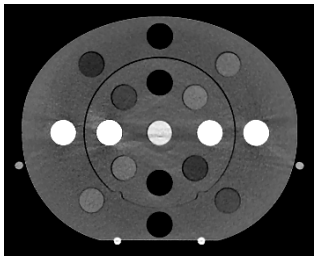


(b) Difference to GT relative to water

Fig. 2: Central slice of reconstructed RED volume (left; W/L = 0.6/1.0) compared to ground truth (GT) RED (right; W/L = 15/0 %). Positive difference (red) values in right panel denote lower reconstructed values w.r.t. GT. Dashed circles indicate regions for mean RED value calculations.



(a) iCBCT



(b) VMI (66.5 keV)

Fig. 3: Beam hardening artifact near bone plugs is reduced in the VMI from polychromatic reconstruction (right) as compared to an iCBCT reconstruction (left). W/L = 500/75 HU

insert height for scatter reduction. We compared two reconstructions shown in Fig. 3: a standard iCBCT reconstruction (5 ART iterations, followed by Acuros® CTS scatter correction and 10 PL iterations) and a polychromatic reconstruction (5 iCBCT ART iterations, followed by Acuros® CTS scatter correction and subsequent 50 IPR iterations; with number of iterations determined empirically for better RED accuracy and image quality. To compare HU values for both reconstructions, we transformed the RED volume from the polychromatic reconstruction to a VMI using (4) at $E_j = 66.5$ keV, corresponding to the mean beam energy at 125 kVp.

The shading near bone plugs present in the iCBCT reconstruction (Fig. 3a) caused by beam hardening is reduced in the polychromatic reconstruction (Fig. 3b).

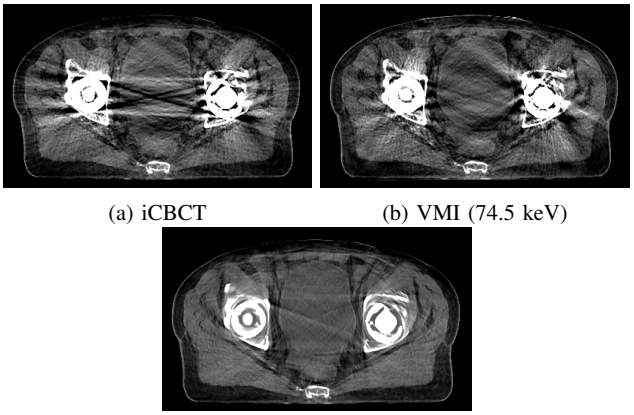
C. Metal artifact mitigation

We also tested the ability of polychromatic reconstruction to reduce metal artifacts. For that purpose, we selected a scan of a patient with bilateral metal hip implants (data courtesy of Queen's Hospital in Romford). The scan was acquired on a Varian Halcyon™ machine in half-fan beam configuration at 125 kVp. We compared three reconstructions shown in Fig. 4: a standard iCBCT reconstruction (5 ART iterations, followed by Acuros® CTS scatter correction and 5 PL iterations), an iCBCT reconstruction with metal artifact reduction (MAR) and a polychromatic reconstruction (5 iCBCT ART iterations, followed by Acuros® CTS object scatter correction and subsequent 25 IPR iterations). To compare HU values for both reconstructions, we transformed the RED volume from the polychromatic reconstruction to a VMI at $E_j = 74.5$ keV. The polychromatic model used for this reconstruction was extended by a fit to common hip replacement alloys [16] in addition to typical human tissues [10] and comprised of $N_f = 4$ segments.

The shading between hip joints caused by metal implants visible in the iCBCT reconstruction (Fig. 4a) is reduced in the VMI from the polychromatic reconstruction (Fig. 4b), improving the visibility of the bladder. Some level of metal artifacts still remains, however, we point out that, unlike in the iCBCT MAR reconstruction (Fig. 4c) the result of the polychromatic reconstruction is based solely on the underlying polychromatic model without metal inpainting in projections.

IV. Conclusion

We have implemented physics-based polychromatic forward projection within our CBCT reconstruction pipeline. Preliminary results on phantom data demonstrate the potential of iterative polychromatic reconstruction for direct RED volume reconstruction with reduced beam hardening artifacts. In the presence of metal implants, polychromatic reconstruction



(a) iCBCT (b) VMI (74.5 keV)

(c) iCBCT MAR

Fig. 4: Metal artifact within bladder region is reduced in the VMI from polychromatic reconstruction (top right) as compared to a standard iCBCT reconstruction (top left). As a reference, an iCBCT reconstruction with heuristic metal artifact reduction (MAR) is shown in the bottom panel. W/L = 500/75 HU. Patient data courtesy of Queen's Hospital in Romford

helps to improve the visibility of soft tissue by partially removing shading. However, a certain level of metal artifacts persists. We presume metal artifact reduction with polychromatic reconstruction can be further improved by adapting Acuros® CTS object scatter correction implementation to take RED volumes directly as input and extend its material types to include titanium as well. We also see the combination of polychromatic reconstruction with metal inpainting in projections as a potential investigation direction to enable reliable structure auto segmentation and dose calculation for adaptive radiation therapy.

Acknowledgment

The authors would like to thank Queen's Hospital in Romford for providing the patient data.

References

[1] B. De Man, J. Nuyts, P. Dupont, G. Marchal, and P. Suetens, "An iterative maximum-likelihood polychromatic algorithm for CT," *IEEE transactions on medical imaging*, vol. 20, no. 10, pp. 999–1008, 2001.

[2] I. A. Elbakri and J. A. Fessler, "Statistical image reconstruction for polyenergetic x-ray computed tomography," *IEEE transactions on medical imaging*, vol. 21, no. 2, pp. 89–99, 2002.

[3] ——, "Segmentation-free statistical image reconstruction for polyenergetic x-ray computed tomography with experimental validation," *Physics in Medicine and Biology*, vol. 48, no. 15, pp. 2453–2477, 2003.

[4] J. H. Mason, A. Perelli, W. H. Nailon, and M. E. Davies, "Polyquant CT: direct electron and mass density reconstruction from a single polyenergetic source," *Physics in Medicine & Biology*, vol. 62, no. 22, pp. 8739–8762, 2017.

[5] ——, "Quantitative cone-beam CT reconstruction with polyenergetic scatter model fusion," *Physics in Medicine & Biology*, vol. 63, no. 22, p. 225001, 2018.

[6] L. D. S. Trotta, D. Matenine, M. Martini, K. Stierstorfer, Y. Lemaréchal, P. Francus, and P. Després, "Beam-hardening corrections through a polychromatic projection model integrated to an iterative reconstruction algorithm," *NDT & E International*, p. 102594, 2021.

[7] I. Peterlik, A. Strzelecki, M. Lehmann, P. Messmer, P. Munro, P. Paysan, M. Plamondon, and D. Seghers, "Reducing residual-motion artifacts in iterative 3D CBCT reconstruction in image-guided radiation therapy," *Medical Physics*, vol. 48, no. 10, pp. 6497–6507, 2021.

[8] A. Maslowski, A. Wang, M. Sun, T. Wareing, I. Davis, and J. Star-Lack, "Acuros CTS: A fast, linear Boltzmann transport equation solver for computed tomography scatter–Part I: Core algorithms and validation," *Medical physics*, vol. 45, no. 5, pp. 1899–1913, 2018.

[9] A. Wang, A. Maslowski, P. Messmer, M. Lehmann, A. Strzelecki, E. Yu, P. Paysan, M. Brehm, P. Munro, J. Star-Lack *et al.*, "Acuros CTS: A fast, linear Boltzmann transport equation solver for computed tomography scatter–Part II: System modeling, scatter correction, and optimization," *Medical physics*, vol. 45, no. 5, pp. 1914–1925, 2018.

[10] ICRP, "Adult reference computational phantoms," *Ann ICRP*, vol. 39, pp. 1–166, 2009.

[11] L. I. Rudin, S. Osher, and E. Fatemi, "Nonlinear total variation based noise removal algorithms," *Physica D: nonlinear phenomena*, vol. 60, no. 1-4, pp. 259–268, 1992.

[12] H. Erdogan and J. A. Fessler, "Monotonic algorithms for transmission tomography," in *5th IEEE EMBS International Summer School on Biomedical Imaging, 2002.* IEEE, 2002, pp. 14–pp.

[13] D. Kim, S. Ramani, and J. A. Fessler, "Combining ordered subsets and momentum for accelerated X-ray CT image reconstruction," *IEEE transactions on medical imaging*, vol. 34, no. 1, pp. 167–178, 2014.

[14] Y. Nesterov, "Smooth minimization of non-smooth functions," *Mathematical programming*, vol. 103, no. 1, pp. 127–152, 2005.

[15] M. Sun and J. Star-Lack, "Improved scatter correction using adaptive scatter kernel superposition," *Physics in Medicine & Biology*, vol. 55, no. 22, p. 6695, 2010.

[16] M. Bazalova, C. Coolens, F. Cury, P. Childs, L. Beaulieu, and F. Verhaegen, "Monte Carlo dose calculations for phantoms with hip prostheses," in *Journal of Physics: Conference Series*, vol. 102, no. 1. IOP Publishing, 2008, p. 012001.

Dual-Energy Cone-Beam CT with Three-Material Decomposition for Bone Marrow Edema Imaging

Stephen Z. Liu, Magdalena Herbst, Thomas Weber, Sebastian Vogt, Ludwig Ritchl, Steffen Kappler, Jeffrey H. Siewerdsen, and Wojciech Zbijewski

***Abstract*—We investigate the feasibility of bone marrow edema (BME) detection using a kV-switching Dual-Energy (DE) Cone-Beam CT (CBCT) protocol. This task is challenging due to unmatched x-ray paths in the low-energy (LE) and high-energy (HE) spectral channels, CBCT non-idealities such as x-ray scatter, and narrow spectral separation between fat (bone marrow) and water (BME). We propose a comprehensive DE decomposition framework consisting of projection interpolation onto matching LE and HE view angles, fast Monte Carlo scatter correction with low number of tracked photons and Gaussian denoising, and two-stage three-material decompositions involving two-material (fat-Aluminium) Projection-Domain Decomposition (PDD) followed by image-domain three-material (fat-water-bone) base-change. Performance in BME detection was evaluated in simulations and experiments emulating a kV-switching CBCT wrist imaging protocol on a robotic x-ray system with 60 kV LE beam, 120 kV HE beam, and 0.5° angular shift between the LE and HE views. Cubic B-spline interpolation was found to be adequate to resample HE and LE projections of a wrist onto common view angles required by PDD. The DE decomposition maintained acceptable BME detection specificity (<0.2 mL erroneously detected BME volume compared to 0.85 mL true BME volume) over +/-10% range of scatter magnitude errors, as long as the scatter shape was estimated without major distortions. Physical test bench experiments demonstrated successful discrimination of ~20% change in fat concentrations in trabecular bone-mimicking solutions of varying water and fat content.**

***Index Terms*—Dual-energy CT, cone-beam CT, multi-material decomposition, quantitative imaging, bone marrow edema.**

I. Introduction

BONE trauma is often accompanied by bone marrow edema (BME), which presents as elevated fluid content within the fatty yellow marrow. BME detection is conventionally performed using MRI. However, there has been an increasing interest in BME detection using x-ray CT because of its ubiquity in the emergency department, where the presence of BME might aid the diagnosis of occult fractures [1,2]. Since the x-ray attenuation alone is insufficient to distinguish relative contributions of fat, water and bone to trabecular bone voxels, dual-energy (DE) imaging with three-material decomposition have been proposed for BME identifications, e.g., using a dual-source CT [3]. Here, we investigate the feasibility of DE BME imaging on Flat-Panel Detector (FPD) Cone-Beam CT (CBCT) with the kV-switching protocol.

The primary challenge in DE detection of BME is the narrow energy separation between fat and water-like soft-tissues. Thus, any imaging system biases need to be carefully controlled. This is particularly challenging in CBCT due to its sensitivity to x-ray scatter and detector non-idealities (e.g., glare). We propose a BME imaging framework based on a DE CBCT protocol where the kV changes every 0.5° gantry angle (corresponding to a ~50 ms switching interval). The projections are processed with fast Monte Carlo (MC)-based scatter corrections [4], resampled onto matching view angles, and input into the two-material (fat-Aluminum) Projection-Domain Decomposition (PDD) followed by the three-material (fat-water-cortical bone) image-domain base-change [5,6]. We used both simulations and CBCT benchtop experiments to evaluate impact of scatter estimation errors and other CBCT system non-idealities on the performance of the proposed DE BME imaging framework.

II. Methods

A. kV-Switching DE CBCT of the Wrist

Simulations and benchtop experiments emulated the Siemens Healthineers Multitom Rax twin robotic x-ray system in the wrist CBCT configuration shown in Fig. 1A. The source-axis distance was 750 mm, and the source-detector was 1150 mm. We investigated a realistic kV-switching DE protocol where the x-ray source alternated between the low-energy (LE) beam of 60 kV (+2 mm Al, +0.25 mm Cu) and the high-energy (HE) beam of 120 kV (+2 mm Al, +0.25 mm Cu) every 0.5° views, yielding a total 200 LE and 200 HE frames over 200° rotation.

B. Three-Material DE Decomposition

We applied a two-stage three-material DE decomposition method in Yu *et al* [5] (Fig. 1B). First, PDD was performed by interpolating the precomputed 2D lookup tables to convert DE projections into line integrals of two intermediate bases – fat and Al. This step minimized the effects of beam hardening. Reconstructions of fat and Al line integrals were then obtained and processed with base-change to produce the final fat, water and cortical bone volume fraction maps.

Throughout all studies, fat and Al reconstructions involved the FDK algorithm with 0.5 x 0.5 x 0.5 mm^3 voxels, Parker weighting, and 2D ramp kernel (0.5x Nyquist frequency cutoff and Hann apodization). The base-change was formulated as the

This work was supported by academic-industrial collaboration with Siemens Healthineers, XP Division, and by NIH R01 EB025470. The presented method is not commercially available. Due to regulatory reasons, Siemens Healthinners Multitom Rax in not available in all countries, and its future availability cannot be guaranteed.

S. Z. Liu, J. H. Siewerdsen, and W. Zbijewski are with the Department of Biomedical Engineering, Johns Hopkins University School of Medicine, Baltimore, MD 21205, USA. (e-mail: wzbijewski@jhu.edu).

M. Herbst, T. Weber, S. Vogt, L. Ritchl, and S. Kappler are with Siemens Healthineers, Forchhelm 91301, Germany.

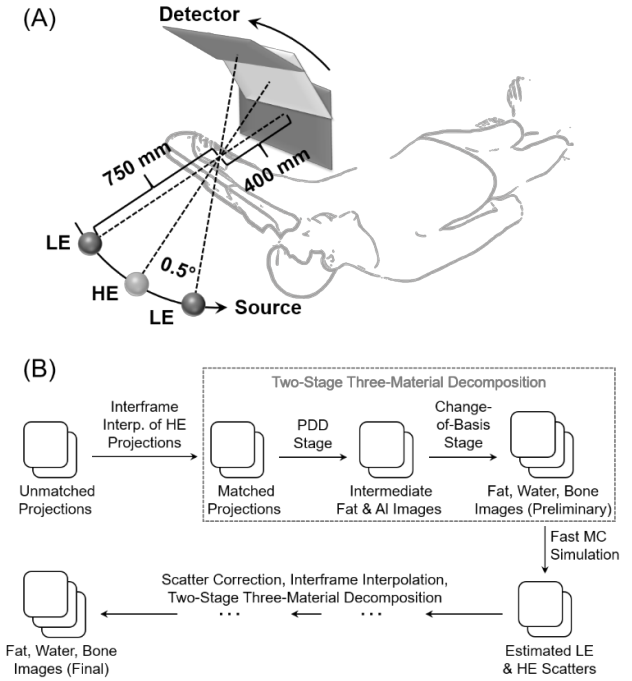
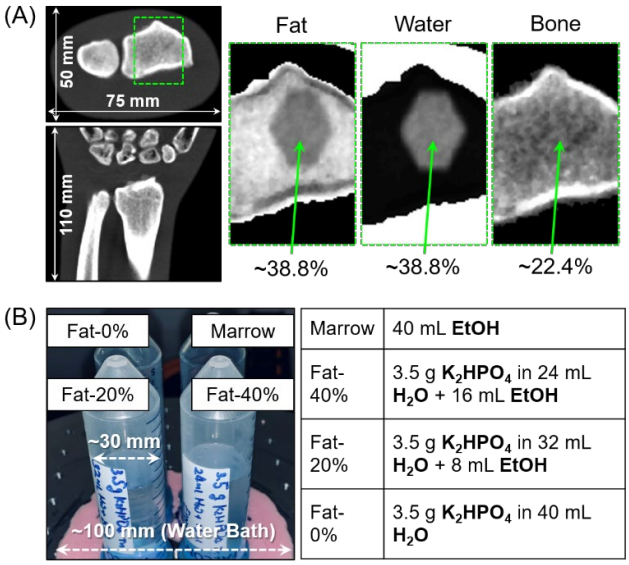


Figure 1. (A) The wrist CBCT scan geometry and the kV-switching protocol. (B) Thee-material (water-fat-bone) DE decomposition framework for the kV-switching CBCT.



Marrow	40 mL **EtOH**
Fat-40%	3.5 g **K_2HPO_4** in 24 mL **H_2O** + 16 mL **EtOH**
Fat-20%	3.5 g **K_2HPO_4** in 32 mL **H_2O** + 8 mL **EtOH**
Fat-0%	3.5 g **K_2HPO_4** in 40 mL **H_2O**

Figure 2. (A) FDK reconstruction of digital normal wrist phantom. Material fraction maps in the BME region (green box) are shown on the right. Volume fractions are provided for each material. (B) Physical phantom consisting of a water cylinder (not shown) with four vials containing the H_2O-EtOH-K_2HPO_4 mixtures specified in the table.

least-square minimization problem constrained by the volume conservation principle [7,8], and optimized using the active-set algorithm. 10 iterations of 3D total-variation denoising with 0.8 penalty strength were applied to the resulting material images.

The kV switching protocol did not provide coinciding LE and HE ray paths required by PDD, due to 0.5° angular separation between successive LE and HE projection views. Therefore, HE projections at the gantry angles matching the LE projections were synthesized using the sinogram interframe interpolation; various interpolation approaches were compared (see Sec. III).

Scatter correction was achieved by a previously developed fast MC-based approach [4]. A rapid, but relatively noisy MC simulation with a low number of tracked photons was first performed; the scatter estimate was then denoised using a 2D Gaussian kernel and subtracted from the projections. The object model for the MC simulation was given by the material maps obtained from a preliminary three-material decomposition of the uncorrected data. A second, final pass of the decomposition was then applied to the corrected projections (see Fig. 1B).

C. *Simulation Studies*

The sensitivity of proposed framework in Fig. 1B to scatter estimation errors was investigated using a previously reported realistic x-ray system model [9]. We assumed a 400 x 400 mm^2 FPD with 0.3 x 0.3 mm^2 pixel size and a 0.6 mm thick CsI scintillator. The x-ray source was operated at 4 mAs/frame for the 60 kV beam, and 1 mAs/frame for the 120 kV beam. X-ray spectra and linear attenuation coefficients of all materials were obtained from *Spektr3.0* [10]; quantum and detector electronic noise with correlations were included in the model. Source axial collimation was set to 150 mm (on the detector). No anti-scatter grid was installed. Ground truth scatter was simulated using the MC algorithm with a relatively large number of photon tracks (10^7) and minimal smoothing (kernel FWHM = 4 mm).

We used two digital anthropomorphic phantoms: a normal wrist of ~50 x 75 mm^2 cross section shown in Fig. 2A, and a large wrist obtained by dilating the soft tissues of the normal wrist to ~80 x 105 mm^2 cross section. The wrists consisted of an outer fat layer (skin), an inner water region (soft tissues), and bones made of a mixture of cortical bone and fat (yellow marrow). A ~0.85 mL BME stimulus containing cortical bone and equal fractions of fat and water was placed in the radius (Fig. 2A). All mixtures fulfilled the volume conservations.

D. *Sensitivity to Scatter Correction Errors*

Simulated kV-switching data were processed following the pipeline in Fig. 1B. Fast MC scatter estimation was performed using 10^6 photons (Fig. 3) – an order of magnitude less than the ground truth in data generation (10^7). We studied the sensitivity of decomposition to inaccuracies in i) shape and ii) magnitude of the estimated scatter. The shape error was introduced by adjusting FWHM of the Gaussian smoothing kernel (ranging 10 - 290 mm), which introduced an increasing distortion of the scatter as illustrated in Fig. 3. The magnitude error was generated by scaling the denoised scatter distribution (ranging 80% - 120%). A total of 9x9 combinations of errors were tested.

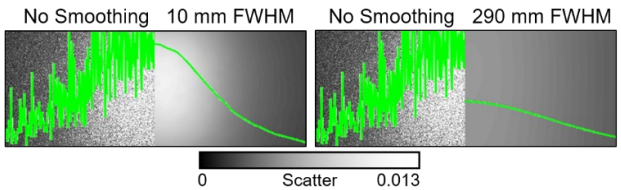


Figure 3. MC-estimated LE scatter profiles using 10^6 photon tracks without smoothing, and with 10 mm and 290 mm FWHM Gaussian kernels. Green curves show the scatter profile along the central detector row.

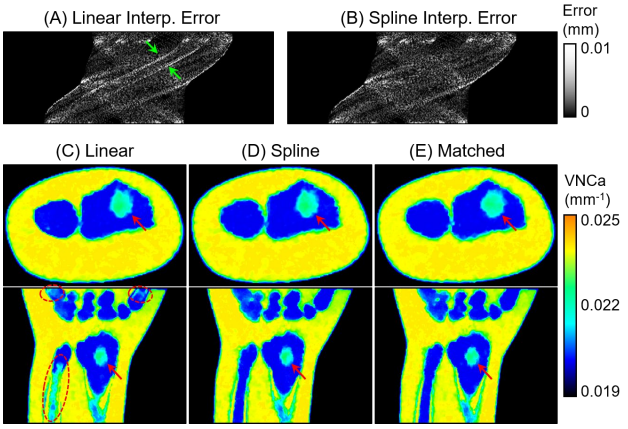


Figure 4. Absolute sinogram errors for linear (A) and spline (B) interframe interpolations of HE data. The errors are plotted for the detector row approx. 200 slices inferior from the central slice. Green arrows indicate the error on the ulna. Decompositions (VNCa images) for DE acquisitions of unmatched LE/HE projections with linear (C) and spline (D) interframe interpolations, and of matched LE/HE projections (idealized reference) (E). Red arrows indicate the true BME region; red dashed circles mark the false BME enhancement.

The performance of decompositions under scatter inaccuracy was evaluated in terms of the specificity of BME detection. We measured whether any connected water voxel clusters >1.5 mm diameter were present in the bones where there should be no fluid – i.e., outside of the true BME stimulus. To this end, the non-zero water voxels within bone boundaries were segmented; the segmentation was then processed with erosion followed by dilation using a 1.5 mm diameter sphere to remove any clusters <1.5 mm diameter. Detection specificity was computed as: $V_{detected} - V_{detected \cap true}$, where $V_{detected}$ is the total volume of water detected (segmented) within bones, and $V_{detected \cap true}$ is the volume of water detected within the true BME region.

E. Physical CBCT Test Bench Study

CBCT test bench equipped with a 300 x 300 mm^2 FPD with 0.388 x 0.388 mm^2 pixels was configured to emulate Multitom Rax wrist scan geometry as in the simulations, with 60 mm axial collimation on the detector. The same DE acquisition protocol as in the simulations was implemented.

A three-material physical phantom was developed to validate BME detection, as shown in Fig. 2B. Ethyl alcohol (EtOH) was used to represent yellow marrow, H_2O for soft-tissue and BME, and K_2HPO_4 for cortical bone. The phantom consisted of a ~100 mm diameter water bath with 4 Falcon® tubes of ~30 mm diameter. The composition in each tube is listed in Fig. 2B.

Detector glare deconvolution [4] was performed on the DE projections prior to applying the two-stage DE decomposition framework of Sec. II.B. Fast MC scatter estimation used 10^6 photons and a 10 mm FWHM smoothing kernel. The PDD stage used polyethylene (PE) and Al as bases; in the base-change stage, they were converted to EtOH, H_2O and K_2HPO_4 bases. Spectra for PDD lookup tables were calibrated from attenuation measurements of Al filters of varying thickness. In the base-change stage, the span of PE-Al that formed each of the final material was obtained from the PDD of a calibration phantom different from the phantom used in the study.

III. Results

Fig. 4 provides a comparison of two different interframe interpolation kernels for synthesizing HE projections at view angles matching the LE data: 2D linear and cubic B-spline. (Note that in this test, the projection data was scatter-free.) The interpolations were applied onto the sinogram of each detector row separately. The differences from a reference HE sinogram simulated directly at the view angles of the LE scan are shown in Fig. 4A for the linear interpolation, and in Fig. 4B for the B-

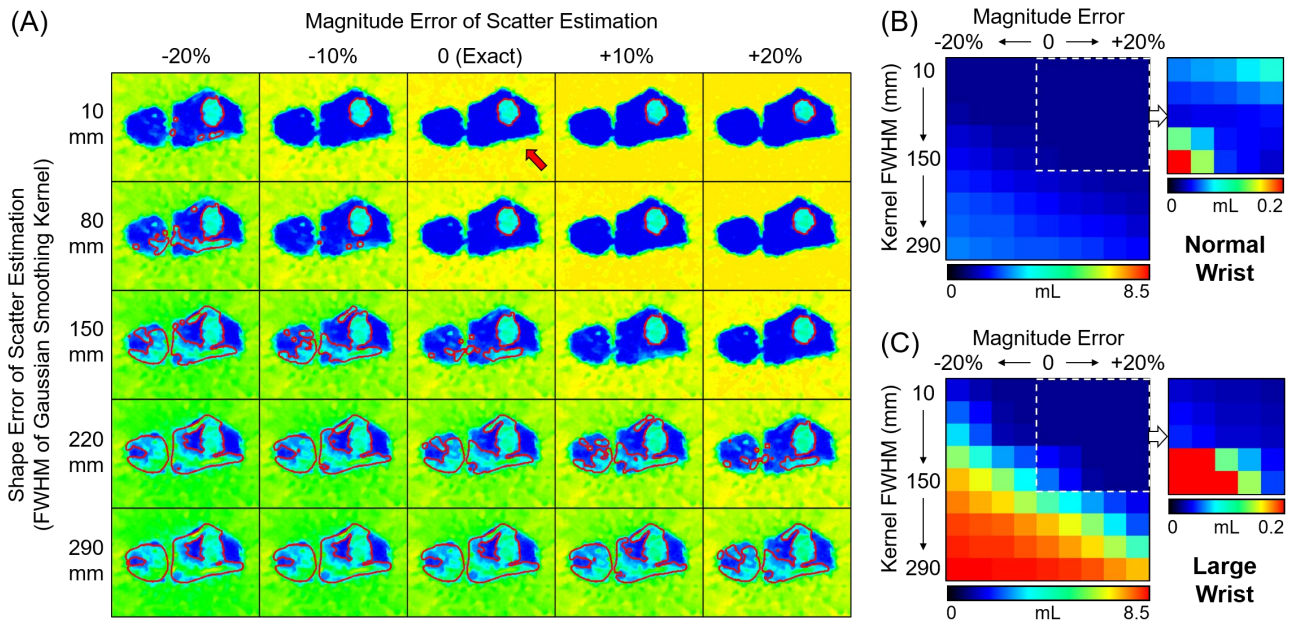


Figure 5. (A) Axial VNCa images (cropped around the bone region) for the scatter error study of the large wrist. Red contours mark the detected BME regions (both true and false positives). (B) BME detection specificity across all scatter error cases for the normal wrist. The region in the dashed box is also shown using a narrower error range. (C) Same as (B) for the large wrist.

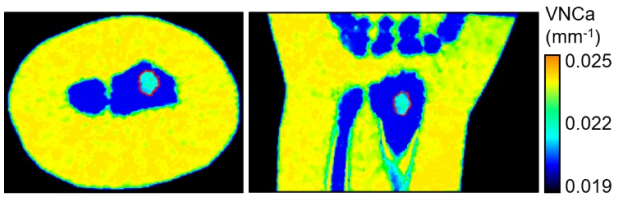


Figure 6. Decomposition of the large wrist scan for MC scatter correction with 0% magnitude error and the 10 mm FWHM Gaussian smoothing kernel. Red contours mark the detected BME areas.

spline interpolation. The decompositions for the two synthesis techniques are displayed as the Virtual Non-Calcium images (VNCa – the attenuation-weighted sum of decomposed fat and water images [11,12]) as in Fig. 4C and 4D. Linear interpolation introduced false BME signals in the ulna and metatarsals, likely due to the sharp transitions in the sinogram of these regions. B-spline interpolation, on the other hand, was free of such errors, indicating that it is an adequate choice for angular resampling of the sinogram. A reference reconstruction of the simulation with matched LE and HE frames is provided in Fig. 4E. Spline projection view synthesis was applied in the subsequent studies.

Fig. 5A shows the VNCa images of the normal wrist for selected combinations of scatter magnitude and shape errors. BME (water) >1.5 mm that were detected in the bones (i.e., true and false positives) are marked. Fig. 5B and C present the BME detection specificity across investigated scatter estimation error settings for the normal and large wrist. Clear visualizations of BME and high detection specificity (<0.2 mL false detection volume compared to 0.85 mL true BME volume) was achieved for both normal and large wrists when the magnitude error was within +/-10% range. This, however, required that the scatter shape was estimated without any major distortions (smoothing FWHM <80 mm). The large wrist was more sensitive to scatter correction inaccuracies, consistent with its elevated scatter-to-primary ratio (SPRs for the normal wrist at the detector center were 6.5% and 4.6% for LE and HE frames, respectively, compared to 14.4% and 10.1% for the large wrist). Fig. 6 shows the axial and coronal VNCa images for the large wrist using scatter simulation with 10 mm FWHM smoothing and 0% introduced magnitude error. There are almost no false BME clusters outside of the true stimulus, illustrating the feasibility of DE CBCT of BME providing adequate scatter correction.

Fig. 7A shows the FDK reconstruction of test-bench LE acquisitions of the physical phantom. The variation in fat (i.e., EtOH) fractions among inserts (less fat – more BME) cannot be appreciated on the attenuation image, but is readily apparent in the DE decompositions with proper artifact corrections. The DE decomposition without glare or scatter correction (Fig. 7B) presents erroneous fat content in the fat-free insert (Fat-0%) and underestimated water concentrations in the water bath. These errors are largely removed with scatter correction (Fig. 7C). Fig. 7D underscores the need for the correction of FPD non-idealities – the VNCa image with glare and scatter correction achieves improved uniformity in the water bath than the image without glare correction.

IV. Discussion and Conclusion

We investigated the feasibility of BME detection using the kV-switching DE CBCT protocol paired with an analytical two-

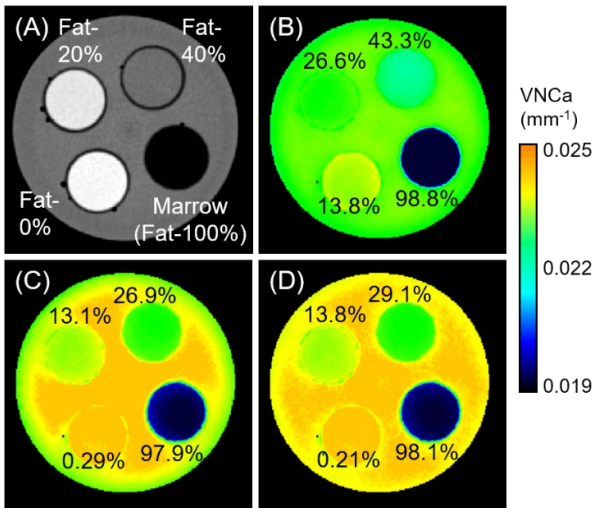


Figure 7. (A) FDK reconstruction of the test-bench LE projections of the physical BME phantom. The ground truth fat (EtOH) fraction in each insert is marked. VNCa images of decomposition with no corrections (B), without glare correction but with scatter correction (C), and with both glare and scatter corrections (D). Estimated fat fractions for each case are indicated.

stage three-material decomposition pipeline. The problem is inherently challenging due to unmatched x-ray paths of the LE and HE channels and similar attenuation properties of fat-like yellow marrow and water-like BME.

Using high-fidelity simulations, we found that the B-spline interpolation is sufficient to resample HE and LE frames onto common view angles required by PDD from an acquisition with a 0.5° shift between the two spectral channels. This result was achieved for the fairly complex anatomy of the wrist, indicating that spline projection synthesis might be useful in a broad range of kV-switching CBCT applications. Considering inaccuracies of scatter estimation, the decomposition maintains acceptable BME detection specificities over +/-10% magnitude errors as long as scatter shape is reasonably accurate. This finding suggests that MC or deep learning-based scatter corrections are likely adequate for BME detection using DE CBCT.

The feasibility of BME detections using proposed DE CBCT framework was validated in physical test bench experiments, demonstrating successful estimation of the difference in fluid and fat content of trabecular bone-mimicking solutions.

Ongoing works include experimental reproducibility study in complex objects under variable imaging conditions [13], and the application of a recently reported model-based one-step DE three-material decomposition algorithm (CMBMD) [8] to BME quantification to avoid interframe interpolations.

Reference

[1] I. T. Ali, D. W. Wong, T. Liang, F. Khosa, M. Mian, S. Jalal, and S. Nicolaou, "Clinical utility of dual-energy CT analysis of bone marrow edema in acute wrist fractures," *American Journal of Roentgenology*, vol. 210, no. 4, pp. 842-847, 2018.

[2] B. Gosangi, J. C. Mandell, M. J. Weaver, J. W. Uyeda, S. E. Smith, A. D. Sodickson, and B. Khurana, "Bone marrow edema at dual-energy CT: a game changer in the emergency department," *RadioGraphics*, vol. 40, no. 3, pp. 859–874, 2020.

[3] B. Petritsch, A. Kosmala, A. M. Weng, B. Krauss, A. Heidemeier, R. Wagner, T. M. Heintel, T. Gassenmaier, and T. A. Bley, "Vertebral compression fractures: third-generation dual-energy CT for detection of

bone marrow edema at visual and quantitative analyses," *Radiology*, vol. 284, no. 1, pp. 161–168, 2017.

[4] A. Sisniega, W. Zbijewski, J. Xu, H. Dang, J. W. Stayman, J. Yorkston, N. Aygun, V. Koliatsos, and J. H. Siewerdsen, "High-fidelity artifact correction for cone-beam CT imaging of the brain," *Physics in Medicine and Biology*, vol. 60, no. 4, pp. 1415–1439, 2015.

[5] L. Yu, X. Liu, and C. H. McCollough, "Pre-reconstruction three-material decomposition in dual-energy CT," In *Medical Imaging 2009: Physics of Medical Imaging*, vol. 7258, p. 72583V, 2009.

[6] S. Z. Liu, C. Zhao, M. Herbst, T. Weber, S. Vogt, L. Ritschl, S. Kappler, J. H. Siewerdsen, and W. Zbijewski, "Feasibility of dual-energy cone-beam CT of bone marrow edema using dual-layer flat panel detectors," In *Medical Imaging 2022: Physics of Medical Imaging*, vol. 12031, pp. 379-386, 2022.

[7] R. F. Barber, E. Y. Sidky, T. G. Schmidt, and X. Pan, "An algorithm for constrained one-step inversion of spectral CT data," *Physics in Medicine & Biology*, vol. 61, no. 10, p. 3784, 2016.

[8] S. Z. Liu, J. H. Siewerdsen, J. W. Stayman, and W. Zbijewski, "Quantitative dual-energy imaging in the presence of metal implants using locally constrained model-based decomposition," In *Medical Imaging 2021: Physics of Medical Imaging*, vol. 11595, p. 115951C, 2021.

[9] C. Zhao, S. Z. Liu, W. Wang, M. Herbst, T. Weber, S. Vogt, L. Ritschl, S. Kappler, J. W. Stayman, J. H. Siewerdsen, and W. Zbijewski, "Effects of x-ray scatter in quantitative dual-energy imaging using dual-layer flat panel detectors," In *Medical Imaging 2021: Physics of Medical Imaging*, vol. 11595, p. 115952A, 2021.

[10] J. Punnoose, J. Xu, A. Sisniega, W. Zbijewski, and J. H. Siewerdsen, "Spektr 3.0—A computational tool for x-ray spectrum modeling and analysis," *Medical Physics*, vol. 43, no. 8, pp. 4711–4717, 2016.

[11] T. T. Kellock, S. Nicolaou, S. S. Y. Kim, S. Al-Busaidi, L. J. Louis, T. W. O'Connell, H. A. Ouellette, and P. D. McLaughlin, "Detection of bone marrow edema in nondisplaced hip fractures: utility of a virtual noncalcium dual-energy CT application," *Radiology*, vol. 284, no. 3, pp. 798-805, 2017.

[12] F. C. Müller., K. K. Gosvig, H. Børgesen, J. S. Gade, M. Brejnebøl, A. Rodell, M. Nèmery, and M. Boesen, "Dual-energy CT for suspected radiographically negative wrist fractures: a prospective diagnostic test accuracy study," *Radiology*, vol. 296, no. 3, pp. 596–602, 2020.

[13] S. Z. Liu, Q. Cao, G. M. Osgood, J. H. Siewerdsen, J. W. Stayman, and W. Zbijewski, "Quantitative assessment of weight-bearing fracture biomechanics using extremity cone-beam CT," In *Medical Imaging 2020: Biomedical Applications in Molecular, Structural, and Functional Imaging*, vol. 11317, p. 113170I, 2020.

Thursday, June 16

Invited Talk on interventional CT

Interventional Imaging

Cardiac CT and Motion Compensation

Real-time Liver Tumor Localization via a Single X-ray Projection Using Deep Graph Network-assisted Biomechanical Modeling

Hua-Chieh Shao, Jing Wang, and You Zhang

***Abstract*—Real-time imaging is highly desirable in image-guided radiotherapy, as it provides instantaneous knowledge of patient's anatomy and motion during the treatment and enables online treatment adaptation to achieve the highest tumor targeting accuracy. Due to extremely limited acquisition time, only one or several X-ray projections can be acquired for real-time imaging, which poses a substantial challenge to localize the tumor from the scarce projections. For liver radiotherapy, such a challenge is further exacerbated by the diminished contrast between the tumor and the normal liver tissues. Here, we propose a framework combining graph neural network-based deep learning and biomechanical modeling to track liver tumor in real time from a single on-board X-ray projection. The liver tumor tracking is achieved in two steps. First, a deep learning network is developed to predict the liver surface deformation, using image features learned from the X-ray projection. Second, the intra-liver deformation is estimated through biomechanical modeling, using the liver surface deformation as the boundary condition to solve intra-liver tumor motion by finite element analysis. The accuracy of the proposed framework was evaluated using a dataset of 10 patients with liver cancer. The results show accurate liver surface registration from the graph-based neural network, which translates into accurate real-time, fiducial-less liver tumor localization (<1.3 mm localization error).**

***Index Terms*—Liver, real-time tumor localization, X-ray, deep learning, graph convolutional network, biomechanical modeling**

I. Introduction

The introduction of conformal radiotherapy enables high-precision dose delivery to the tumor and spares surrounding normal tissues, enabling treatment margin reduction, dose escalation, and improved tumor control [1]. However, internal anatomical motion such as respiratory or cardiac motion leads to tumor location uncertainties, and may cause the radiation beams to miss the tumor and damage normal tissues. Image-guided radiation therapy widely uses X-ray based imaging to localize the tumor before and during the treatment to maintain the delivery accuracy [2]. Real-time imaging, in particular, is highly desired as it can localize the tumor instantly and allow the treatment to adapt to such real-time changes to achieve ultimate treatment accuracy. Due to the stringent temporal resolution requirement (hundreds of milliseconds) of real-time imaging, the volumetric information will be severely under-sampled via current mainstream imaging modalities including cone-beam computed tomography (CBCT). Such a degree of under-sampling makes it impossible to reconstruct high-quality CBCTs using conventional methods for tumor localization.

Due to the recent successes of deep learning (DL), several groups have proposed DL-based methods for real-time imaging. A few network architectures were proposed to reconstruct three-dimensional (3D) CBCT images from single-view or orthogonal-view X-ray projections [3-5]. Such networks, however, were built on an ill-conditioned problem, trying to estimate high-dimensional volumetric data from a single X-ray projection. Considerable reconstruction errors remain, albeit much smaller than those of the conventional reconstruction algorithms. Moreover, to track tumors in real-time, additional steps of image registration or segmentation are necessary to further localize tumors from the reconstructed CBCT images. This is particularly challenging for liver tumors due to the low contrast of liver tumors against surrounding normal liver parenchyma.

To address the above challenges toward real-time imaging, especially for liver tumor localization, we propose a mesh registration-based method combining deep neural networks with biomechanical modeling. The method directly solves the liver tumor motion between a prior CT/CBCT image and a single X-ray projection to localize liver tumors in real-time, and effectively eliminates the need to reconstruct a high-quality, intermediate CBCT image prior to localization. Specifically, a deep graph neural network-based architecture was trained to model the correlation between patient-specific liver boundary motion and features learned on individual X-ray projections. The trained network can then predict liver boundary motion from a single real-time X-ray projection. Using the predicted liver boundary motion as the boundary condition, we further performed finite element analysis-based biomechanical modeling of liver to solve intra-liver tumor motion. The method adopts a deformation-driven approach that incorporates prior information to tackle the extreme under-sampling issue. The two-step-based registration scheme simplifies the complexity of the deep graph neural network with introduced domain knowledge (biomechanical modeling). Biomechanical modeling uses information including structure geometry, material composition and elasticity to derive physiologically and physically meaningful deformation, and complements the intensity information provided in the X-ray projection to further improve the registration and tumor localization accuracy [6].

The accuracy of the proposed technique was evaluated using

The study was supported by funding from the National Institutes of Health (R01CA240808) and a seed grant from the Department of Radiation Oncology at the University of Texas Southwestern Medical Center.

H.-C. Shao, J. Wang, and Y. Zhang are with the Department of Radiation Oncology, University of Texas Southwestern Medical Center, Dallas, TX 75390 USA.

10 patients with liver cancer, and compared with the accuracy of two other techniques. The first technique uses the diaphragm as an anatomic landmark, and tracks the diaphragm motion directly from the on-board projections via template matching to represent liver tumor motion. The second technique is a principal component analysis (PCA) based method which models 3D motion into a few motion eigenvectors for dimension reduction and tumor tracking [7].

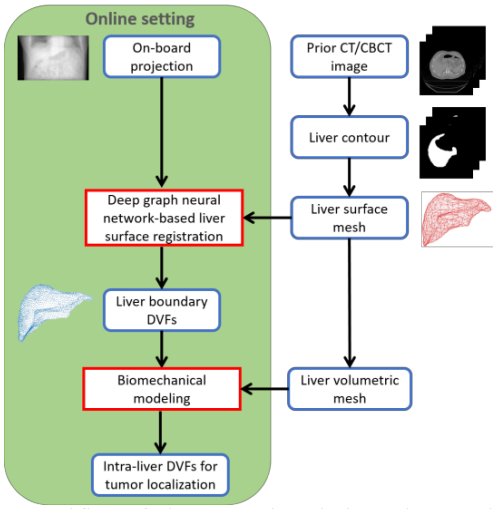


Fig. 1. Workflow of the proposed method. A deep graph neural network-based model was trained to predict liver surface deformation vector field (DVF) from a single X-ray projection. Then a biomechanical model solves the intra-liver DVF using the liver surface DVF as the boundary condition for tumor localization.

II. Materials And Methods

II.A. Method overview

In this study, liver motion and liver tumor localization were solved via deformable registration between a liver mesh (extracted offline from a prior CT/CBCT image available before the treatment) and the liver features projected on a single X-ray projection (Fig. 1). The registration was achieved via two steps: (a) liver surface motion estimation via a deep graph neural network-based structure (Fig. 2); and (b) intra-liver motion estimation via biomechanical modeling. Specifically, in step (a) a patient-specific DL model was trained to predict a liver boundary deformation vector field (DVF) that deforms the prior liver surface mesh to match with the liver shape variations encoded in the X-ray projection. In step (b), a biomechanical model of the liver was built, and an intra-liver DVF was solved through finite element analysis using the liver boundary DVF as the boundary condition.

II.B. The deep-learning network architecture

The network was trained to learn image features from on-board X-ray projections to predict boundary movement from the prior liver mesh to each on-board projection. The DL network architecture is illustrated in Fig. 2, which contains two subnetworks. The first subnetwork extracts image features from each on-board X-ray projection, and the extracted feature maps are pooled and fed into the second subnetwork for liver

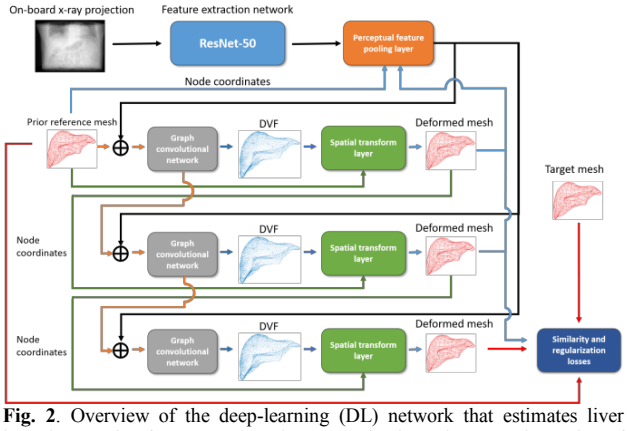


Fig. 2. Overview of the deep-learning (DL) network that estimates liver boundary motion from an on-board X-ray projection. The network consists of two subnetworks performing feature extraction and liver boundary DVF prediction separately. The first subnetwork uses ResNet-50 to extract image features from an X-ray projection. The extracted feature maps were pooled for each node of a liver surface mesh by the perceptual feature pooling layer, based on the projected node coordinates on the X-ray projection. The second subnetwork, consisting of three deformation blocks, progressively estimates liver boundary DVFs. A deformation block comprises of a graph convolutional network (GCN) and a spatial transform layer. The GCN was learned to predict a liver boundary DVF based on the features extracted from the ResNet-50 subnetwork. A spatial transform layer deforms the prior reference mesh or the deformed liver surface mesh from the previous deformation block, using a GCN-predicted DVF.

boundary DVF prediction. Here we used ResNet-50 [8] as the feature extraction network. Consisting of a series of convolutional layers stacked in a residual learning architecture, ResNet-50 extracts encoded liver shape variations, via local and global image features contained in the X-ray projection, and learns short- and long-range dependencies among these extracted features. These learned dependencies are helpful for the deformation estimation because they are shown common in respiration-induced liver motion. The perceptual feature pooling layer pools the ResNet-50 extracted feature maps by associating each 3D node of a liver surface mesh with a 2D point in the feature maps, based on the same geometry of the cone-beam projection.

The second subnetwork comprises a series of deformation blocks that progressively deforms the liver surface mesh nodes based on the extracted feature maps from the first subnetwork. Each deformation block involves a graph convolutional network (GCN, Fig. 3) and a spatial transform layer that deforms the liver surface mesh using the GCN-predicted DVF [9]. GCN performs graph-based convolutions that generalize the standard convolution operations to data structures lack of underlying Euclidean structures, such as functional networks in brain imaging. A non-Euclidean data structure can be represented by a weighted graph comprised of a set of vertices, edges connecting the vertices, and weights associated with each vertex (e.g., vertex features, DVFs, vertex-associated image features). The use of GCN is indicated for our problem, as the liver surface mesh nodes, the geometrical connectivity (edges) between the nodes, and the learned image, DVF, and vertex features associated with each node make a standard non-Euclidean data structure for inputs into the GCN. Using extracted image features that encode the liver shape, preceding DVFs, and learned vertex features from the previous block, the

GCN learns to predict a liver surface DVF to further deform the surface mesh deformed by the previous block. The inputs of the GCN in the first deformation block contain ResNet50-extracted image features and an initial DVF which was set to be zero. For each subsequent GCN, the image features were re-pooled based on the new node coordinates (Fig. 2), deformed via the spatial transform layer and the DVF predicted by the preceding block. The image features were then input into the subsequent GCN, along with the predicted surface DVF and the learned vertex features from the preceding block. We used a GCN of the same architecture as the G-ResNet [9], which is illustrated in Fig. 3. The corresponding network was modified and adapted from the Pixel2Mesh library [10]. The model training was driven by a loss function involving a mesh similarity loss and regularization losses that regularize the deformation and enforce smoothness of the boundary DVFs.

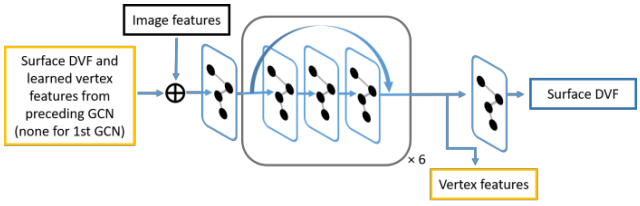


Fig. 3. Graph convolutional network (GCN). The inputs contain pooled image features from the feature extraction ResNet-50 subnetwork (Fig. 2), a surface DVF, and vertex features yielded from the GCN in the previous deformation block (if any). The GCN consists of 20 graph convolution layers that, except for the entrance and exiting layers, were organized in a residual learning architecture. The GCN yields a surface DVF and vertex features to feed into the subsequent deformation block. The inputs of the first GCN in the second subnetwork contains only image features and an initial surface DVF which was set to be zero. The image features were re-pooled for each GCN based on deformed node coordinates. The rounded box in the middle represents a residual learning module containing three graph convolution layers with a shortcut connection, which iterates 6 times.

II.C. Biomechanical modeling

After the deep neural network solves a liver boundary DVF to match with the liver shape features on the X-ray projection, the intra-liver DVF was subsequently derived using a biomechanical model. Here we used the Mooney-Rivlin material model, which describes a hyperelastic (i.e., nonlinear elasticity) material that fits biological tissues well. The details of implementing the biomechanical model can be found in Ref. [6].

II.D. Dataset curation and augmentation

A dataset of 10 patients with liver cancer from our institute was used to evaluate the proposed method. The study was approved under an institutional review boards protocol. Each patient had a contrast-enhanced four-dimensional CT set from treatment planning, and the CT images were binned into 10 respiratory phases (from 0% to 90%), with 0% being the end-of-inhale phase. The CT images were resampled to a uniform size of 256×256×128 with an isotropic resolution of 2 mm×2 mm×2 mm. On-board X-ray cone-beam projections were simulated from the CT images using a ray-tracing algorithm. We simulated projections from three angles: 0, 45, and 90 degrees. The 0- and 90-degree are for anterior-posterior and left-right directions, respectively.

Since each patient had only a 10-phase 4D-CT set, to generate sufficient motion variation scenarios to train the patient-specific network and avoid overfitting, we augmented the dataset of each patient by simulating realistic respiratory deformations encountered in on-board liver imaging. The augmentation was based on a PCA-based motion model of each patient [6, 11]. We first performed deformable registrations between the reference 0% phase and the other phases to attain DVFs, using the open-source software package Elastix. To improve the intra-liver DVF accuracy, we applied biomechanical modeling to derive intra-liver DVFs, using the liver surface DVFs solved by Elastix as boundary conditions. We then replaced the Elastix intra-liver DVFs with the biomechanical modeling-derived intra-liver DVFs. PCA was subsequently performed on these high-quality DVFs of each patient to obtain patient-specific principal motion components. For augmentation, the coefficients of the first three principal motion components were randomly scaled to re-generate DVFs of various magnitudes and patterns [11]. In total, for each patient we generated 1,728 augmented samples which were partitioned into training, validation, and testing sets. The partitioning was assigned according to the original respiratory phases of the PCA coefficients prior to the random scaling. The training set includes the samples of which the original PCA coefficients were from the 10% to 40% phases; the validation set includes the samples whose original PCA coefficients were from the 60% and 70% phases; and the testing set includes the samples whose original PCA coefficients were from the 50%, 80%, and 90% phases. 50% is the end-of-exhale phase that has the largest deformation from the 0% phase.

II.E. Evaluation schemes

The deformation accuracy of liver surface meshes was evaluated using the Hausdorff distance (HD) between the deformed and the 'ground-truth' target liver surface meshes extracted from the augmented dataset [12]. To evaluate the performance of liver tumor tracking, we manually contoured the tumors from the prior CT images at phase 0%. The tumor contours at 0% phase were then propagated using the augmentation DVFs (II.D.) to other augmented motion states, which were used as the 'ground-truth' to evaluate the ones deformed by our method. The accuracy of liver tumor tracking was evaluated by the Dice similarity score (DSC), center-of-mass error (COME), and HD.

III. Results

III.A. Liver deformation accuracy

Figure 4 presents a qualitative comparison of liver surface meshes and projected nodes on X-ray projections at three projection angles (0, 45, and 90 degrees). The first row shows the surface mesh overlays between the prior and 'ground-truth' target meshes (left panel), and between the graph network-deformed and 'ground-truth' target meshes (right panel). The prior and target meshes correspond to the end-of-inhale phase and the end-of-exhale phase (with motion augmentation), respectively. The other rows show the overlay of the pre- (left panel) and post-registration (right panel) surface mesh nodes onto the corresponding X-ray projections

of the end-of-exhale phase, at three different angles. Both the surface mesh overlay and the node-projection overlay demonstrate high registration accuracy.

Table I summarizes the mean (±s.d.) liver HDs of the proposed method and the PCA-based 2D-3D registration method [7]. The DL-based method results in much smaller HDs than the PCA-based 2D-3D method.

III.B. Liver tumor tracking accuracy

Table II summarizes the mean (±s.d.) liver tumor DSCs, COMEs, and HDs of the proposed method at three projection angles. In addition, the results of PCA-based 2D-3D registration and diaphragm tracking are also presented in the Table for comparison. The diaphragm tracking is only able to localize the diaphragm in 2D from a single X-ray projection, thus we only used it to represent liver tumor motion along the superior-inferior (SI) direction. The COME of the diaphragm-based method is thus only for SI direction, with 3D COME potentially being much larger. Table II clearly shows that the proposed method has the best liver tumor localization accuracy, and the performance is consistent among different angles.

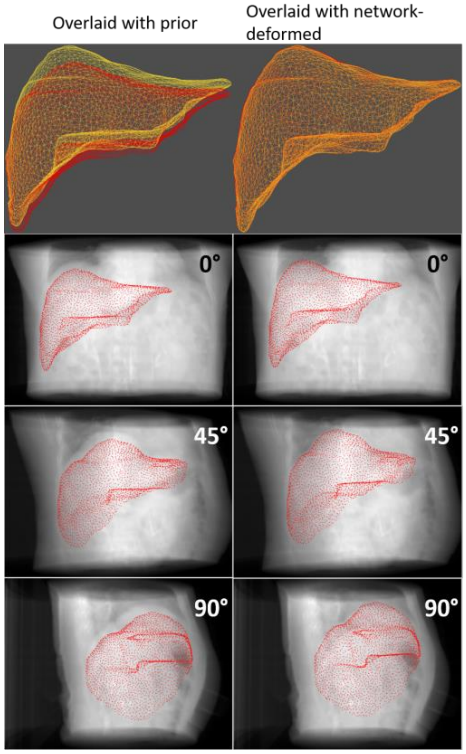


Fig. 4. (*First row*) Liver surface overlays between the prior and 'ground-truth' target meshes (left) and between the graph network-deformed and target meshes (right). The yellow meshes are the target meshes corresponding to the end-of-exhale phase after augmentation, and the red meshes correspond to the prior (left panel) and deformed (right panel) meshes. (*Other rows*) Liver surface nodes projected on X-ray projections at three projection angles. Left and right columns show the projected nodes corresponding to the prior and deformed surface meshes, respectively.

IV. Conclusion

We proposed a method to track liver tumor motion in real-time from a single on-board X-ray projection, using combined DL-based liver surface registration and biomechanical modeling. The results show highly-accurate liver surface registrations, which translate into mean liver tumor localization errors of less than 1.3 mm. Our method can potentially be applied towards intra-treatment tumor monitoring and real-time plan adaptation.

TABLE I. Mean (±s.d.) liver Hausdorff distances.

Projection angle (degree)	Prior (mm)	Method	
		DL prediction (mm)	PCA-based 2D-3D registration (mm)
0	11.77±6.11	2.99±2.42	7.27±4.18
45		3.03±2.39	6.55±3.32
90		3.09±2.55	6.09±2.47

TABLE II. Mean (±s.d.) liver tumor DSC, COME, and HD. The COME for the diaphragm-based method is for the superior-inferior direction only (*).

Project-ion angle (deg.)	Metric	Prior	Method		
			DL prediction	PCA-based 2D-3D registration	Diaphragm tracking
0	DSC	0.547± 0.269	0.895±0.112	0.789±0.205	--
45			0.893±0.110	0.822±0.155	
90			0.886±0.118	0.835±0.134	
0	COME (mm)	6.08± 4.40	1.13±1.33	2.53±4.32	1.68±2.22*
45			1.15±1.32	1.84±1.64	2.69±2.73*
90			1.25±1.41	1.73±1.37	3.08±3.26*
0	HD (mm)	7.24± 4.92	2.81±1.77	3.95±5.00	--
45			2.86±1.77	3.17±1.81	
90			2.93±1.85	3.01±1.36	

References

1. Verellen, D., et al., *Innovations in image-guided radiotherapy.* Nature Reviews Cancer, 2007. **7**(12): p. 949-960.
2. Dhont, J., et al., *Image-guided Radiotherapy to Manage Respiratory Motion: Lung and Liver.* Clinical Oncology, 2020. **32**(12): p. 792-804.
3. Shen, L.Y., W. Zhao, and L. Xing, *Patient-specific reconstruction of volumetric computed tomography images from a single projection view via deep learning.* Nature Biomedical Engineering, 2019. **3**(11): p. 880-888.
4. Lei, Y., et al., *Deep learning-based real-time volumetric imaging for lung stereotactic body radiation therapy: a proof of concept study.* Physics in Medicine and Biology, 2020. **65**(23).
5. Wei, R., et al., *Real-time tumor localization with single x-ray projection at arbitrary gantry angles using a convolutional neural network (CNN).* Physics in Medicine and Biology, 2020. **65**(6).
6. Zhang, Y., et al., *4D liver tumor localization using cone-beam projections and a biomechanical model.* Radiother Oncol, 2019. **133**: p. 183-192.
7. Zhang, Y., et al., *A technique for estimating 4D-CBCT using prior knowledge and limited-angle projections.* Medical Physics, 2013. **40**(12).
8. Paszke, A., et al., *PyTorch: An Imperative Style, High-Performance Deep Learning Library.* Advances in Neural Information Processing Systems 32 (Nips 2019), 2019. **32**.
9. Wang, N.Y., et al., *Pixel2Mesh: Generating 3D Mesh Models from Single RGB Images.* Computer Vision - Eccv 2018, Pt Xi, 2018. **11215**: p. 55-71.
10. Cao, J. *Pixel2Mesh*. 2021 [cited 2022 January 10]; Available from: https://github.com/noahcao/Pixel2Mesh.
11. Jiang, Z., et al., *Enhancing digital tomosynthesis (DTS) for lung radiotherapy guidance using patient-specific deep learning model.* Phys Med Biol, 2021. **66**(3): p. 035009.
12. Taha, A.A. and A. Hanbury, *Metrics for evaluating 3D medical image segmentation: analysis, selection, and tool.* Bmc Medical Imaging, 2015. **15**.

3D Reconstruction of Stents and Guidewires in an Anthropomorphic Phantom From Three X-Ray Projections

Tim Vöth, Thomas König, Elias Eulig, Michael Knaup, Veit Wiesmann, Klaus Hörndler, and Marc Kachelrieß

***Abstract*—Today, the subcutaneous, minimally invasive procedures performed in interventional radiology are usually guided by 2D X-ray fluoroscopy. In 2D X-ray fluoroscopy a series of 2D X-ray images is displayed. For challenging procedures however, 3D X-ray fluoroscopy would be advantageous. In 3D X-ray fluoroscopy, a series of 3D images, which is reconstructed from a series of 2D X-ray images, is displayed. Because the number of images used for guiding an intervention is very high, little dose can be spent per 3D reconstruction of a 3D fluoroscopy. To save dose and to minimize motion artifacts, a reconstruction algorithm that requires very few X-ray projections is desirable. Earlier work showed that guidewires, stents and coils, which are commonly used in interventions, can be reconstructed using only four synthetic X-ray projections. The reconstruction from two or three X-ray projections was only studied briefly. In this work, we improve the method by using a more suitable neural network architecture and by using a multi-channel backprojection instead of a single-channel backprojection. We then apply the improved method to more realistic data measured in an anthropomorphic phantom. The results show that the method produces 3D reconstructions of stents and guidewires with submillimeter accuracy using only three measured X-ray projections.**

***Index Terms*—3D fluoroscopy, sparse view, few view, X-ray, CT, deep learning, stent reconstruction, guidewire reconstruction, 4D, interventional guidance, interventional imaging.**

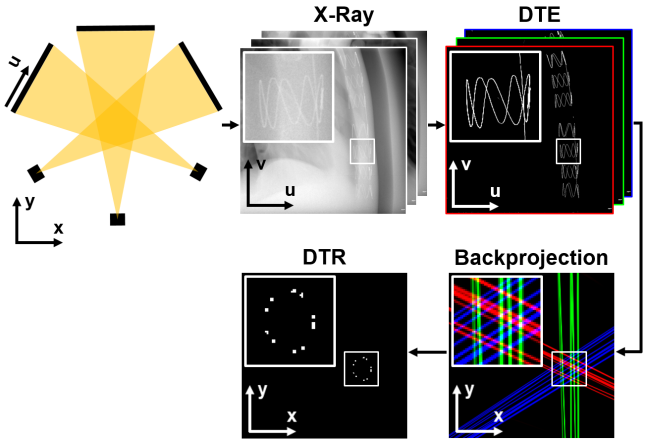


Fig. 1. Sketch of the tool reconstruction pipeline. First, three X-ray projections are acquired with an angular increment of 60°. Then, DTE is applied to each X-ray projection. The DTE output images are backprojected into separate channels (illustrated by different colors) of a backprojection volume. The backprojection is then fed into DTR, which outputs a segmentation of the tools (only one transversal slice of the backprojection and the DTR output are shown).

I. Introduction

APPROACHES for 3D reconstruction of interventional tools from few X-ray projections can be divided into two categories: algorithms specialized in the reconstruction of a single type of interventional tool and more general algorithms capable of reconstructing different types of tools. Belonging to the first category, many algorithms for the 3D reconstruction of curvilinear structures, like guidewires or catheters, have been proposed. A single guidewire [1], [2] or catheter [2] can be reconstructed from one X-ray projection, if a prior 3D dataset showing the patient's vasculature is available. Without a prior 3D dataset, a single guidewire [3], [4] or catheter [3], [5] can be reconstructed from two X-ray projections. An approach for computing a 3D representation of a stent from a single X-ray projection has been proposed [6]. Since it requires a 3D model of the stent as prior knowledge, it is not a 3D reconstruction algorithm, but rather a registration algorithm. To the best of our knowledge, no algorithm specializing in the 3D reconstruction of stents from few X-ray projections has been proposed. In the second category of more general tool reconstruction algorithms, the reconstruction of guidewires and stents from about 16 X-ray projections has been demonstrated [7], [8], [9], [10]. Recently, the reconstruction of guidewires, stents and coils from only four X-ray projections has been demonstrated on synthetic data [11], [12]. In this work, we improve this algorithm and demonstrate on measured data of an anthropomorphic phantom, that guidewires and stents can be reconstructed from only three X-ray projections.

T. Vöth is with Ziehm Imaging GmbH, Lina-Ammon-Straße 10, 90471 Nürnberg, Germany and with the German Cancer Research Center (DKFZ), Im Neuenheimer Feld 280, 69120 Heidelberg, Germany. E-mail: tim.voeth@dkfz.de.

Dr. T. König, V. Wiesmann and K. Hörndler are with Ziehm Imaging GmbH, Lina-Ammon-Straße 10, 90471 Nürnberg, Germany.

E. Eulig, Dr. M. Knaup and Prof. Dr. M. Kachelrieß are with the German Cancer Research Center (DKFZ), Im Neuenheimer Feld 280, 69120 Heidelberg, Germany.

II. Methods

The tool reconstruction pipeline, which is similar to the one proposed by Eulig et al. [12], is outlined in Figure 1. Three X-ray projections, offset by 60°, are acquired simultaneously. The deep tool extraction (DTE) algorithm extracts the tools from each projection. The output images of DTE are then backprojected into a volume. Finally, the deep tool reconstruction (DTR) algorithm transforms the backprojection into a 3D reconstruction of the tools. In this work, we improved

DTR in two ways. First, we replaced the 2.5D convolutional neural network (CNN), which performed DTR in reference [12], with a more suitable 3D CNN. Secondly, we backproject the three DTE output images into separate channels of the backprojection volume, which improves the reconstruction quality significantly.

A. Deep Tool Extraction

The task of DTE is to output the pixel-wise line integral of the X-ray attenuation coefficient of the guidewires and stents in the input projection. Training and validation data were generated online by combining forward projections of simulated guidewires and stents with clinical cone-beam computed tomography (CBCT) projections containing patients without interventional tools. In total, 12,000 guidewire projections, 12,000 stent projections and 2751 clinical projections from nine patients were used. The data were augmented by combining random tool patches and random patient patches, and by simulating varying levels of blur, noise and scatter. DTE was implemented in TensorFlow [13] as a 2D CNN similar to the U-Net [14]. In each resolution stage of the encoder and decoder, two (3×3 convolution + batch normalization [15] + ReLU)-blocks are performed. The number of feature maps of the convolution layers in the nth resolution stage is 32×2^n where n ranges from 1 (highest resolution, 1024×1024) to 7 (lowest resolution, 16×16). Downsampling is performed by 2×2 max pooling, upsampling by nearest-neighbor interpolation followed by a 3×3 convolution layer. The mean absolute error was used as the loss function. Training took 200 epochs, where each epoch consisted of 16,000 training pairs and 4000 validation pairs. Each pair consisted of an input patch and an output patch of size 384×384. The Adam optimizer [16] (learning rate = 1×10^{-5}, $\beta_1 = 0.9$, $\beta_2 = 0.999$) and a mini-batch size of 24 were used.

B. Deep Tool Reconstruction

The task of DTR is to transform the input backprojection volume into a 3D segmentation of the tools. Training and validation data were generated by simulating forward projections of simulated 3D models of guidewires and stents. The backprojection of these projections was used as input volume, a binary voxelization of the 3D models was used as target volume. To make DTR insensitive to errors of DTE, i.e. false-positives and false-negatives, such errors were simulated into the forward projections prior to backprojection. The 3D guidewire models were simulated as curved cylinders around center lines represented by splines with random-walk-generated control points. The 3D stent models were simulated by stacking cylindrical strut segments along their central axis and subsequently bending the stent along a spline. Segment diameter, segment height, number of strut oscillations per segment, strut thickness, strut pattern within the segments (e.g. sinusoidal, zigzag, ...), number of stent segments, axial offset between segments, and bending spline were varied randomly. We simulated aortic stents with diameters between $10\,\mathrm{mm}$ and $30\,\mathrm{mm}$.

The simulated X-ray system has three imaging threads (60° inter-thread angle) with a CBCT-like projection geometry: each thread consists of a point-source and a flat detector (1024^2 pixels, $0.3\,\mathrm{mm}$ pixel pitch, source-detector distance $R_{\mathrm{FD}} = 1100\,\mathrm{mm}$, source-isocenter distance $R_{\mathrm{F}} = 600\,\mathrm{mm}$). For the simulation of the training and validation data and for the application to measured data, a grid of 256^3 voxels of size $(0.6\,\mathrm{mm})^3$ was chosen. To save disk space, since training was performed on patches of size 128^3 anyway, only a 128^3-patch of each simulated 256^3-volume was stored (the full dataset would require $840\,\mathrm{GB}$). The patches were augmented online by z-axis-flips and rotations around the z-axis by integer multiples of 90°. The dataset consisted of 40,000 scenes, each featuring one stent and one or two guidewires. The dataset was split 70/30 between training and validation.

DTR was implemented in TensorFlow as a 3D CNN similar to the 3D U-Net [17]. In each resolution stage of the encoder and decoder, two ($3 \times 3 \times 3$ convolution + batch normalization + ReLU)-blocks are performed. The number of feature maps of the convolution layers in the nth resolution stage is 8×2^n where n ranges from 1 (highest resolution) to 5 (lowest resolution). Downsampling is performed by $2 \times 2 \times 2$ max pooling, upsampling by nearest-neighbor interpolation followed by a $3 \times 3 \times 3$ convolution layer. The soft Dice loss with Laplace smoothing was used for training. Training was performed for 150 epochs, where each epoch consisted of 16,000 training pairs and 4000 validation pairs. Each pair consisted of an input patch and an output patch of size 128^3. The Adam optimizer (learning rate = 1×10^{-4}, $\beta_1 = 0.9$, $\beta_2 = 0.999$) and a mini-batch size of 8 were used.

C. Phantom Measurements

Results will be shown on measured data of an anthropomorphic X-ray phantom (PBU-50, Kyoto Kagaku Co. Ltd. Japan) with a soft-tissue-equivalent extension on the anterior side. The interventional tools were placed between phantom and extension. The X-ray projection geometry was the same as described in Section II-B. The three required projections were selected retrospectively from a 3D scan with fine angular sampling. Results will be shown on three different scenes. The first scene contains two guidewires. The second scene contains one stent. The third scene contains one stent and one guidewire, which was placed inside the stent. All projections were acquired at $80\,\mathrm{kV}$ and $0.80\,\mathrm{mA\,s}$ per projection (scene 1) or $1.15\,\mathrm{mA\,s}$ per projection (scenes 2 and 3).

D. Evaluation on Measured Data

To assess the quality of the output Y of our 3D tool reconstruction pipeline on measured data, a ground truth 3D reconstruction GT is needed. For each of the aforementioned scenes, this ground truth was generated by thresholding a 3D reconstruction that was computed from the projections of the respective 3D scan using the algorithm of Feldkamp, Davis and Kress [18]. To quantify the deviation of Y from GT, one could use the popular Sørensen-Dice coefficient (DSC). However, since the guidewires and stent struts are only a few (1–3) voxels in diameter, the DSC is very sensitive to whether

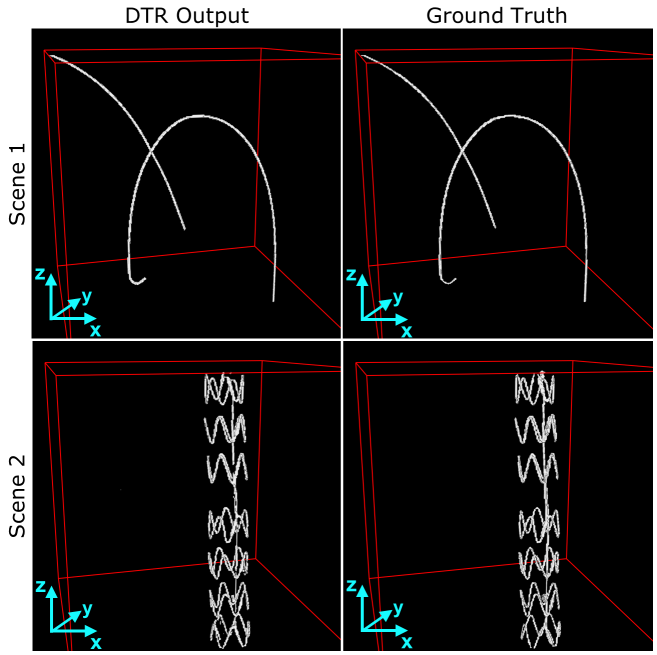


Fig. 2. Volume-rendered DTR output and ground truth of the first and second scene.

a voxel near the surface of such a structure is classified as foreground or background. Because the classification of the surface-voxels in the ground truth is uncertain (it is very sensitive to the threshold used to generate the ground truth), another metric is needed. We therefore decided to use the average Euclidean distance between the voxels of a skeleton of Y and the voxels of a skeleton of GT, $\overline{D}_{S(\mathrm{Y}),S(\mathrm{GT})}$ and the average Euclidean distance between the voxels of a skeleton of GT and the voxels of a skeleton of Y, $\overline{D}_{S(\mathrm{GT}),S(\mathrm{Y})}$. Skeletonization was used to make the metrics less sensitive to the diameter of the guidewires and stent struts. This is desirable, because the diameter in the ground truth is uncertain (as explained above) and because the exact diameter would likely not matter for interventional guidance. Skeletonization was performed by the `skeletonize_3d` function from scikit-image 0.17.2 [19], which implements the algorithm proposed by Lee et al. [20].

E. Real-Time Capability

Inference of DTE on a 1024^2 projection takes about $40\,\mathrm{ms}$, inference of DTR on a 3×256^3 volume takes about $260\,\mathrm{ms}$ on an NVIDIA RTX 3090 GPU. These times were measured in TensorFlow 2.5.0 using mixed precision, graph execution and with the input data already in GPU memory when starting the timer.

III. RESULTS

The training of DTR was performed twice to investigate whether passing the backprojections of the three DTE output images separately (3-channel backprojection) to DTR is advantageous compared to passing a single volume, into which all three DTE output images were backprojected

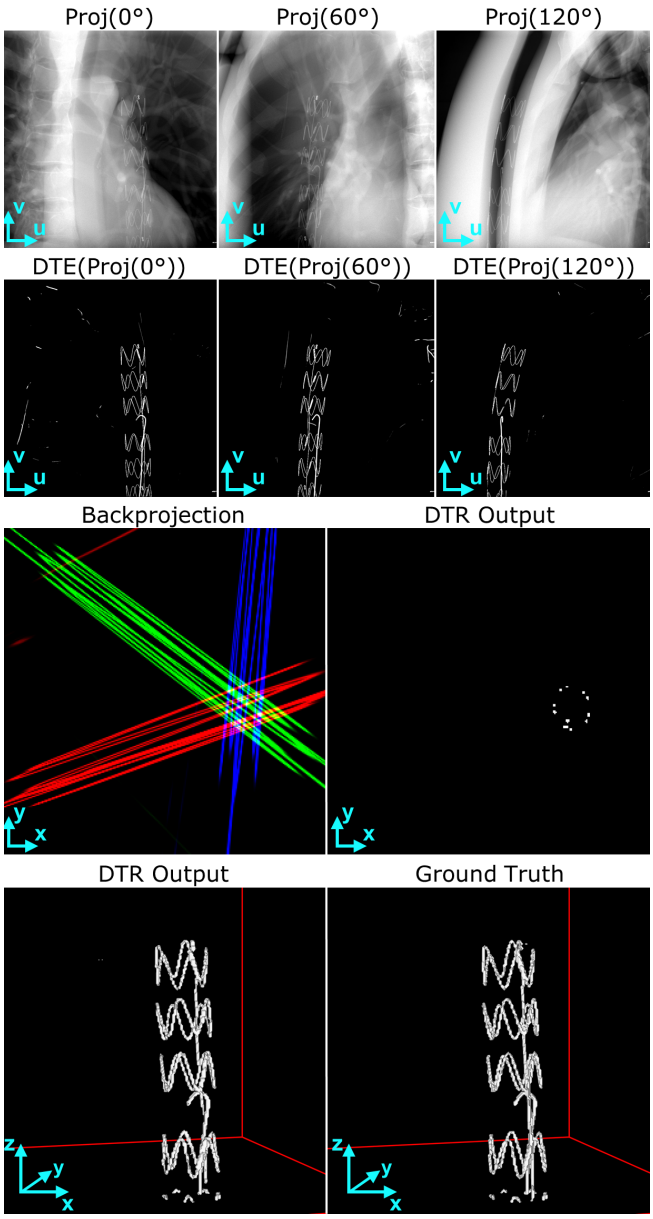


Fig. 3. Inputs, intermediate results and outputs of the tool reconstruction pipeline on the third scene (see Section II-C). First row: the input of the tool reconstruction pipeline are three X-ray projections separated by 60°. Second row: the outputs of DTE, when applied to the images of the first row. Third row: exemplary transversal slices of the backprojection of the DTE outputs and the output of DTR when applied to said backprojection. Fourth row: volume-rendered DTR output and ground truth.

(1-channel backprojection). For the first training, the 3-channel backprojection was used. For the second training, the 1-channel backprojection, which was generated by summing the channels of the 3-channel backprojection, was used. After 150 epochs, training 1 resulted in a DSC of $78.2\,\%$ / $77.5\,\%$ on the training / validation dataset, while training 2 resulted in $73.5\,\%$ / $72.3\,\%$.

The tool reconstruction pipeline was applied to the mea-

	Scene	$\overline{D}_{S(\mathrm{Y}),S(\mathrm{GT})}$	$\overline{D}_{S(\mathrm{GT}),S(\mathrm{Y})}$
3-channel backprojection	1	$0.27\,\mathrm{mm}$	$0.26\,\mathrm{mm}$
	2	$0.36\,\mathrm{mm}$	$0.28\,\mathrm{mm}$
	3	$0.31\,\mathrm{mm}$	$0.33\,\mathrm{mm}$
1-channel backprojection	1	$0.32\,\mathrm{mm}$	$0.31\,\mathrm{mm}$
	2	$0.50\,\mathrm{mm}$	$0.35\,\mathrm{mm}$
	3	$6.20\,\mathrm{mm}$	$0.44\,\mathrm{mm}$

TABLE I
Quantitative evaluation of the deviation of the tool reconstruction from the ground truth using the metrics defined in Section II-D.

sured projections of the three scenes described in Section II-C. The volume-rendered tool reconstructions look very similar to the ground truths, as can be seen in Figure 2 (scenes 1 & 2) and Figure 3 (scene 3). The red box is a cube of size $256 \times 0.6\,\mathrm{mm} = 153.6\,\mathrm{mm}$, indicating the extent of the reconstructed volume. For scene 3, inputs and intermediate results are shown in Figure 3. The deviation of the tool reconstruction from the ground truth was quantified using the metrics defined in II-D. The results are shown in Table I. For all scenes, the tool reconstruction using the 3-channel backprojection outperformed the one using the 1-channel backprojection. The value of $\overline{D}_{S(\mathrm{Y}),S(\mathrm{GT})} = 6.20\,\mathrm{mm}$ for the reconstruction using the 1-channel backprojection on scene 3 is due to outliers in the reconstruction.

IV. NEW WORK TO BE PRESENTED

In this work, we improved an earlier method for computing 3D reconstructions of interventional tools from very few X-ray projections and demonstrate its performance on more realistic data, which were measured in an anthropomorphic phantom.

V. CONCLUSIONS

We demonstrated that our improved pipeline can produce 3D reconstructions of stents and guidewires with submillimeter accuracy from only three X-ray projections measured in an anthropomorphic phantom. Furthermore, we demonstrated that passing the backprojections of the three DTE output images as three separate input channels to DTR, significantly improves the reconstruction quality. The low number of X-ray projections required per 3D reconstruction and the straightforward adaptability to different types of tools makes this algorithm a promising candidate for implementing 3D fluoroscopic interventional guidance.

ACKNOWLEDGMENT

Parts of the reconstruction software were provided by RayConStruct® GmbH, Nürnberg, Germany.

REFERENCES

[1] T. van Walsum, S. A. M. Baert, and W. J. Niessen, "Guide Wire Reconstruction and Visualization in 3DRA Using Monoplane Fluoroscopic Imaging," *IEEE Transactions on Medical Imaging*, vol. 24, no. 5, pp. 612–623, 2005.

[2] M. Brückner, F. Deinzer, and J. Denzler, "Temporal Estimation of the 3D Guide-Wire Position Using 2D X-Ray Images," *Medical Image Computing and Computer-Assisted Intervention – MICCAI 2009*, pp. 386–393, 2009.

[3] M. Wagner, S. Schafer, C. Strother, and C. Mistretta, "4D Interventional Device Reconstruction from Biplane Fluoroscopy," *Medical Physics*, vol. 43, no. 3, pp. 1324–1334, 2016.

[4] S. Baert, E. van de Kraats, T. van Walsum, M. Viergever, and W. Niessen, "Three-Dimensional Guide-Wire Reconstruction from Biplane Image Sequences for Integrated Display in 3d Vasculature," *IEEE Transactions on Medical Imaging*, vol. 22, no. 10, pp. 1252–1258, 2003.

[5] M. Hoffmann, A. Brost, C. Jakob, M. Koch, F. Bourier, K. Kurzidim, J. Hornegger, and N. Strobel, "Reconstruction Method for Curvilinear Structures from Two Views," *SPIE Medical Imaging*, p. 86712F, 2013.

[6] S. Demirci, A. Bigdelou, L. Wang, C. Wachinger, M. Baust, R. Tibrewal, R. Ghotbi, H.-H. Eckstein, and N. Navab, "3D Stent Recovery from One X-Ray Projection," in *Medical Image Computing and Computer-Assisted Intervention – MICCAI 2011*, G. Fichtinger, A. Martel, and T. Peters, Eds. Berlin, Heidelberg: Springer Berlin Heidelberg, 2011, vol. 6891, pp. 178–185, series Title: Lecture Notes in Computer Science.

[7] J. Kuntz, B. Flach, R. Kueres, W. Semmler, M. Kachelrieß, and S. Bartling, "Constrained Reconstructions for 4D Intervention Guidance," *Physics in Medicine and Biology*, vol. 58, no. 10, pp. 3283–3300, 2013.

[8] J. Kuntz, R. Gupta, S. O. Schönberg, W. Semmler, M. Kachelrieß, and S. Bartling, "Real-Time X-Ray-Based 4D Image Guidance of Minimally Invasive Interventions," *European Radiology*, vol. 23, no. 6, pp. 1669–1677, 2013.

[9] B. Flach, M. Brehm, S. Sawall, and M. Kachelrieß, "Deformable 3D–2D Registration for CT and Its Application to Low Dose Tomographic Fluoroscopy," *Physics in Medicine and Biology*, vol. 59, no. 24, pp. 7865–7887, 2014.

[10] B. Flach, J. Kuntz, M. Brehm, R. Kueres, S. Bartling, and M. Kachelrieß, "Low Dose Tomographic Fluoroscopy: 4D Intervention Guidance with Running Prior," *Medical Physics*, vol. 40, no. 10, p. 101909, 2013.

[11] E. Eulig, J. Maier, N. R. Bennett, M. Knaup, K. Hörndler, A. S. Wang, and M. Kachelrieß, "Deep Learning-Aided CBCT Image Reconstruction of Interventional Material from Four X-Ray Projections," in *SPIE Medical Imaging Conference Record*, vol. 113121L:1-7, Mar. 2020.

[12] E. Eulig, J. Maier, M. Knaup, N. R. Bennett, K. Hörndler, A. S. Wang, and M. Kachelrieß, "Deep Learning-Based Reconstruction of Interventional Tools from Four X-Ray Projections for Tomographic Interventional Guidance," *Medical Physics*, vol. 48, no. 10, pp. 5837–5850, Oct. 2021.

[13] Martín Abadi, Ashish Agarwal, Paul Barham, Eugene Brevdo, Zhifeng Chen, Craig Citro, Greg S. Corrado, Andy Davis, Jeffrey Dean, Matthieu Devin, Sanjay Ghemawat, Ian Goodfellow, Andrew Harp, Geoffrey Irving, Michael Isard, Y. Jia, Rafal Jozefowicz, Lukasz Kaiser, Manjunath Kudlur, Josh Levenberg, Dandelion Mané, Rajat Monga, Sherry Moore, Derek Murray, Chris Olah, Mike Schuster, Jonathon Shlens, Benoit Steiner, Ilya Sutskever, Kunal Talwar, Paul Tucker, Vincent Vanhoucke, Vijay Vasudevan, Fernanda Viégas, Oriol Vinyals, Pete Warden, Martin Wattenberg, Martin Wicke, Yuan Yu, and Xiaoqiang Zheng, "Tensorflow: Large-Scale Machine Learning on Heterogeneous Systems," 2015.

[14] O. Ronneberger, P. Fischer, and T. Brox, "U-Net: Convolutional Networks for Biomedical Image Segmentation," *Medical Image Computing and Computer-Assisted Intervention – MICCAI 2015*, pp. 234–241, 2015.

[15] S. Ioffe and C. Szegedy, "Batch Normalization: Accelerating Deep Network Training by Reducing Internal Covariate Shift," *Proceedings of the 32nd International Conference on Machine Learning*, vol. 37, pp. 448–456, 2015.

[16] D. P. Kingma and J. L. Ba, "Adam: A Method for Stochastic Optimization," *3rd International Conference for Learning Representations*, 2015.

[17] Ö. Çiçek, A. Abdulkadir, S. S. Lienkamp, T. Brox, and O. Ronneberger, "3D U-Net: Learning Dense Volumetric Segmentation from Sparse Annotation," *Medical Image Computing and Computer-Assisted Intervention – MICCAI 2016*, pp. 424–432, 2016.

[18] L. A. Feldkamp, L. C. Davis, and J. W. Kress, "Practical Cone-Beam Algorithm," *J. Opt. Soc. Am. A*, vol. 1, no. 6, pp. 612–619, 1984, publisher: OSA.

[19] S. van der Walt, J. L. Schönberger, J. Nunez-Iglesias, F. Boulogne, J. D. Warner, N. Yager, E. Gouillart, and T. Yu, "Scikit-Image: Image Processing in Python," *PeerJ*, vol. 2, p. e453, Jun. 2014.

[20] T. C. Lee, R. L. Kashyap, and C. N. Chu, "Building Skeleton Models Via 3-D Medial Surface Axis Thinning Algorithms," *CVGIP: Graphical Models and Image Processing*, vol. 56, no. 6, pp. 462–478, 1994.

Context-Aware, Reference-Free Local Motion Metric for CBCT Deformable Motion Compensation

H. Huang, J.H. Siewerdsen, W. Zbijewski, C.R. Weiss, M. Unberath, and A. Sisniega

Abstract— Deformable motion is one of the main challenges to image quality in interventional cone beam CT (CBCT). Autofocus methods have been successfully applied for deformable motion compensation in CBCT, using multi-region joint optimization approaches that leverage the moderately smooth spatial variation motion of the deformable motion field with a local neighborhood. However, conventional autofocus metrics enforce images featuring sharp image-appearance, but do not guarantee the preservation of anatomical structures. Our previous work (DL-VIF) showed that deep convolutional neural networks (CNNs) can reproduce metrics of structural similarity (visual information fidelity - VIF), removing the need for a matched motion-free reference, and providing quantification of motion degradation and structural integrity. Application of DL-VIF within local neighborhoods is challenged by the large variability of local image content across a CBCT volume, and requires global context information for successful evaluation of motion effects. In this work, we propose a novel deep autofocus metric, based on a context-aware, multi-resolution, deep CNN design. In addition to the inclusion of contextual information, the resulting metric generates a voxel-wise distribution of reference-free VIF values. The new metric, denoted CADL-VIF, was trained on simulated CBCT abdomen scans with deformable motion at random locations and with amplitude up to 30 mm. The CADL-VIF achieved good correlation with the ground truth VIF map across all test cases with $R^2 = 0.843$ and slope = 0.941. When integrated into a multi-ROI deformable motion compensation method, CADL-VIF consistently reduced motion artifacts, yielding an average increase in SSIM of 0.129 in regions with severe motion and 0.113 in regions with mild motion. This work demonstrated the capability of CADL-VIF to recognize anatomical structures and penalize unrealistic images, which is a key step in developing reliable autofocus for complex deformable motion compensation in CBCT.

***Index Terms*—Interventional CBCT, Motion Compensation, Deformable Motion, Convolutional Neural Network**

I. Introduction

Cone-beam CT (CBCT) provides 3D guidance and intraprocedural imaging in interventional radiology for abdominal procedures but relatively long acquisition time makes it susceptible to patient motion from a complex combination of various periodic and aperiodic sources.

Previous work showed successful application of autofocus optimization for rigid motion compensation [1] using only the acquired CBCT data, with extension to complex deformable motion in abdominal CBCT [2]. Autofocus methods estimate a motion trajectory by minimizing an image-based metric that encourages properties associated to motion-free images (e.g, sharpness or piece-wise constancy). However, such metrics are agnostic to the underlying anatomy and might enforce solutions that satisfy the metric but feature unrealistic structural content.

Our previous work addressed such limitation via a reference-free image similarity metric (DL-VIF) with application to rigid motion compensation in neuro CBCT [3,4]. DL-VIF leveraged the potential of deep convolutional neural networks (CNNs) to extract features specific to motion image degradation and reproduce the capability of Visual Information Fidelity (VIF) [5] to quantify image degradation and structural similarity to a matched motion-free reference, but removing the need for such reference, which is usually not available in clinical settings.

DL-VIF was trained to act on images encompassing the complete head anatomy, with a moderately coarse pixel size, and to provide a global DL-VIF score aggregating contributions to VIF from all structures in the volume into a single scalar. While those assumptions are appropriate for global, rigid, motion compensation, they present various limitations in deformable motion scenarios: i) autofocus deformable motion compensation requires estimation of local VIF, to guide the compensation algorithm towards regions of large motion while ignoring static anatomy; ii) the global nature of the metric makes it susceptible to be dominated by high-contrast structures; and, iii) the coarse voxel size (~2 mm) required for training of the 3D DL-VIF CNN might be not be sufficient for capturing subtle deformation of low-contrast structures.

A patch-based DL-VIF could provide such metric locality, at moderate volume size, but the lack of global context and the inconsistency of image contrast and structure between soft-tissue and bone regions challenges the extraction of meaningful features, resulting in degraded performance, observed in previous attempts to CNN-based deformable autofocus [6].

In this work we propose a novel, context-aware, reference-free autofocus metric, denoted CADL-VIF, that employs a context-aware deep CNN and a voxel-based local VIF definition to provide local estimations of artifacts and structural integrity. CADL-VIF was integrated into a deformable motion compensation framework and was evaluated for soft-tissue deformable motion compensation in simulated cases.

II. Material And Methods

A. CADL-VIF: A Context-Aware Local DL-VIF Design

VIF provides an estimation of the similarity between a test image (motion-corrupted in our case) and a reference image (motion-free) by quantifying the information preserved after a distortion process (motion, in this work), weighted by a

This work was supported by the National Institute of Health, Grant R01-EB-030547.

H. Huang, J. H. Siewerdsen, W. Zbijewski, and A. Sisniega are with the Department of Biomedical Engineering, Johns Hopkins University, Baltimore, MD. J. H. Siewerdsen and C. R. Weiss are with the Department of Radiology, Johns Hopkins University, Baltimore, MD. M. Unberath is with the Department of Computer Science, Johns Hopkins University, Baltimore, MD.

Corresponding Author: A. Sisniega (e-mail: asisniega@jhu.edu).

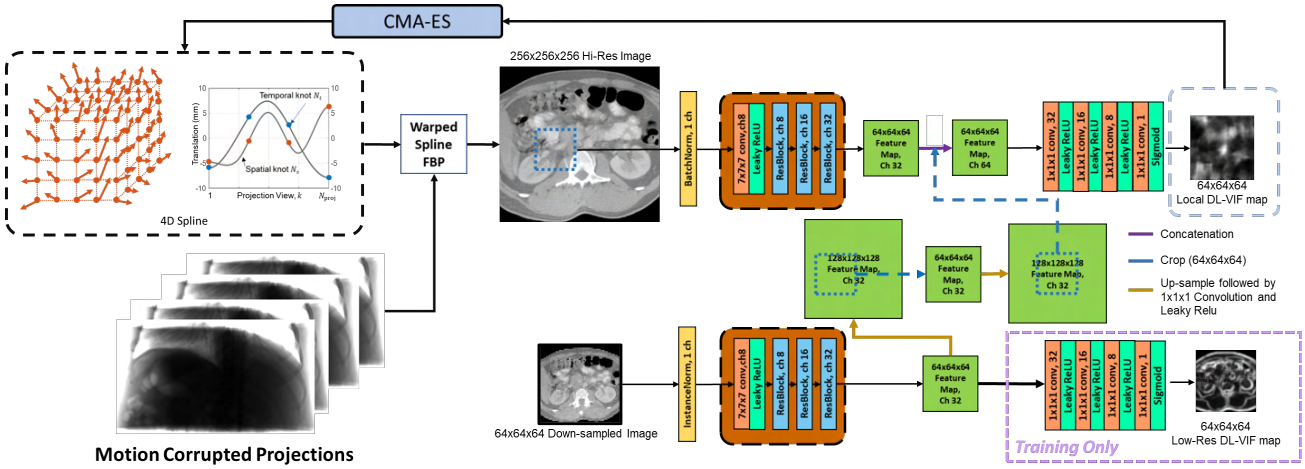


Figure 1. The network architecture of CADL-VIF Map and its integration into the deformable motion compensation framework.

convolution channel, designed as a surrogate of the human visual system response. VIF acts on the integral of the information across the complete image, providing a scalar output. In this work, we extend the definition of VIF to incorporate localized 3D estimation of preserved information.

The resulting VIF_L, was based on the approach in Ref [7] and adapted to 3D volumes. The key difference between VIF_L and original VIF is that when calculating the information contained in motion-corrupted (I_{MC}) and motion-free(I_{MF}) images, VIF_L preserves spatial information by omitting integration across the image dimensions, summing only across different filter channels, as shown in Eqs. 1 and 2.

$$I_{MC} = \sum_{\substack{Filter \\ Channels}} \log(1 + \frac{g^2 \cdot \sigma_{MF}^2}{\sigma_v^2 + \sigma_n^2}) \log\left(1 + \frac{\sigma_{MF}^2}{\sigma_n^2}\right) \quad (1)$$

$$I_{MF} = \sum_{\substack{Filter \\ Channels}} \log\left(1 + \frac{\sigma_{MF}^2}{\sigma_n^2}\right) \quad (2)$$

Where the σ_{MF}^2 is the variance in motion-free image after convolution with kernels designed to mimic the frequency response of the human visual system (HVS), σ_v^2 quantifies the variance introduced by motion artifacts, and σ_n^2 represents the noise in the HVS channel. The term g provides an estimate of the degradation in image information due to patient motion, which depends on the covariance between the motion-corrupted and motion-free images. All terms in Eq. 1 and Eq. 2 were set and calculated as described in Ref [3].

The new VIF_L was reproduced with a novel deep CNN based on our previous DL-VIF design and illustrated in Fig. 1. The network acts on small regions of interest (ROIs) of size 64x64x64 voxels (1 mm isotropic voxel size). Context information is incorporated via a second branch acting on the entire motion-corrupted volume reconstructed at a very coarse voxel size (4 mm isotropically), resulting in a multi-resolution, context aware architecture, inspired by previous work on CT to MR image synthesis [8]. The two branches featured identical layer configurations, with an input 7x7x7 convolution layer and a leaky-ReLU activation, followed by 3 ResBlock layers [3]. Thus, both branches contained an equal number of learnable parameters, yet independently learned. Both branches output 32-channel feature maps, one incorporating local, high-frequency features, and the other providing contextual feature information.

The high-resolution branch incorporated an input batch normalization layer, while the low-resolution branch featured instance normalization. The different normalization responds to different variability presented by the high and low-resolution data. The high-resolution ROIs featured variable soft-tissue and bone regions that present a much larger variability and benefit from batch normalization, compared to the relatively consistent appearance of the low resolution, full abdomen, context.

The context feature map from low-resolution branch was then up-sampled twice and cropped accordingly to match the position and size of the high-resolution ROI. The local and context feature maps were concatenated and input to cascade of 1x1x1 convolution and leaky-ReLU layers to generate the output of high-resolution branch. To facilitate contextual feature learning of the network during training, another series of 1x1x1 convolution with leaky ReLU layers were added to the low-resolution branch after the feature maps, generating low resolution output for the entire volume.

B. Deformable Motion Compensation Framework

CADL-VIF was incorporated as the autofocus metric in a multi-region-based motion compensation framework (see Fig. 1). The time-varying motion vector field (MVF) was estimated with a 4D spline model, integrated into the backprojector algorithm. Deformable motion was estimated as the set of 4D spline coefficients P minimizing the multi-ROI autofocus function:

$$P = argmin_P \sum_{r \in ROIs} \sum_{pixels} -\ln\left[S(\mu_{LR}(P), \mu_{HR}(P, r)) \right] \quad (3)$$

Where S is the autofocus metric for high-resolution ROI, i.e., the high-resolution CADL-VIF Map, calculated from $\mu_{LR}(P)$, the low-resolution image reconstructed with P, and $\mu_{HR}(P, r)$, the high-resolution image reconstructed with P at ROI position r. The negative natural logarithm served as a basic conditioning of the values for optimization. The final autofocus metric was integrated across ROI voxels and across all ROIs. The cost-function was minimized with the Covariance Matrix Adaptation Evolutionary Strategy (CMA-ES) [9].

C. Data Generation and Training

The CADL-VIF network was trained and validated on

simulated data generated using 75 cases from the TCIA lymph node abdomen multi-detector CT (MDCT) database. 61 cases were used for training, 7 cases reserved for validation, and the rest 7 cases for testing. For each simulation instance, a MDCT volume was randomly selected and a 260 mm long sub-volume at a random longitudinal position was extracted from the original volume. The sub-volume was then forward projected using a high-fidelity CBCT projector with geometry pertinent to interventional robotic C-arm systems (source-to-detector distance of 1200 mm, and source-to-axis distance of 785mm). The detector was modeled as a flat-panel with 864 x 660 pixels and 0.64 mm isotropic pixel size. Deformable motion was induced during forward projection, using a MVF with random maximum amplitude ranging from 10 mm to 30 mm at random directions. The MVF featured maximum amplitude at a randomly placed position and decayed smoothly following an elliptical Gaussian kernel with lateral and antero-posterior width randomly chosen from 200 mm to 300 mm and 100 mm to 150 mm, respectively. The MVF followed a cosine temporal pattern with random phase and random frequency ranging from 0.75-1.25 cycles per scan. An additional motion-free scan was simulated to obtain a reference for ground truth VIF_L. Both the motion-corrupted and motion-free images were reconstructed on a 256x256x256 voxels volume with 1x1x1 mm^3 voxel size.

Ground truth VIF_L maps of motion-corrupted images for training were computed using the motion-free images as reference. To emphasize motion artifacts in soft tissue regions, a [0.01, 0.025] mm^{-1} window, followed by CLAHE contrast-enhancement were applied to the motion-corrupted and motion-free volumes before calculation of VIF_L. A total of 610 motion instances were generated for training, 70 for validation, and 70 for testing.

During training, the input data was normalized to [0,1], and data augmentation was achieved via addition of zero mean Gaussian noise with $\sigma = 0.01$. For each training instance pair, a randomly placed 64x64x64 voxels sub-volume was extracted from the motion corrupted image for input to the high-resolution branch, while the full volume, downsampled by a 4x factor, provided the low-resolution contextual input. Training was achieved with a loss function based on the mean square error between the network output and the ground truth VIF_L. The network was trained with the Adam optimizer ($5x10^{-4}$ learning rate), with a batch size of 30, and for 2000 epochs, on 3 Nvidia Quadro RTX A6000 GPUs.

D. Validation Experiments

To test the generalizability of CADL-VIF, a separate dataset with 256 cases was created from the test TCIA volumes, using a larger range of motion frequency: 0.5 to 3 periods per scan. CADL-VIF was estimated on contiguous ROIs of 64 x 64 x 64 voxels, covering the complete 256 mm x 256 mm x 256 mm volume. The sum of CADL-VIF within each ROI was then compared with the sum of ground truth VIF_L in the same region. Results were aggregated from all 64 ROIs in each of 256 cases.

Motion compensation with CADL-VIF was evaluated on 7 simulated cases created analogously to the test dataset, with motion amplitude ranging from 8 to 15 mm, and frequency ranging from 0.8 to 1.2 periods per scan. Motion compensation was performed with a total of 4 local ROIs of 64 mm x 64 mm x 64 mm size, with a common contextual reference, and using a 9x9x9x5 4-dimensional b-spline grid. Results were assessed with SSIM and blurriness estimated using the formula proposed in Ref [7], adapted to 3D, which is calculated as follows: i) each voxel in motion-free and motion-corrupted images is compared with all its neighboring 26 voxels, and the maximum intensity variation is stored in two new volumes, V_{MF} and V_{MC} for the motion-free and motion-corrupted images, respectively; ii) The average value of V_{MF} and V_{MC} is calculated (Z_{MF} and Z_{MC}, respectively); iii) blurriness is defined as $B = \frac{|Z_{MF} - Z_{MC}|}{Z_{MF}}$.

SIM was computed on 6mm x 6mm x 6mm ROIs, that underwent rigid registration to avoid phase mismatches between the motion-corrupted volume, compensated volume, and the static reference. Two ROIs were used per volume, one placed at the center of the motion field and a second one at a quasi-static region.

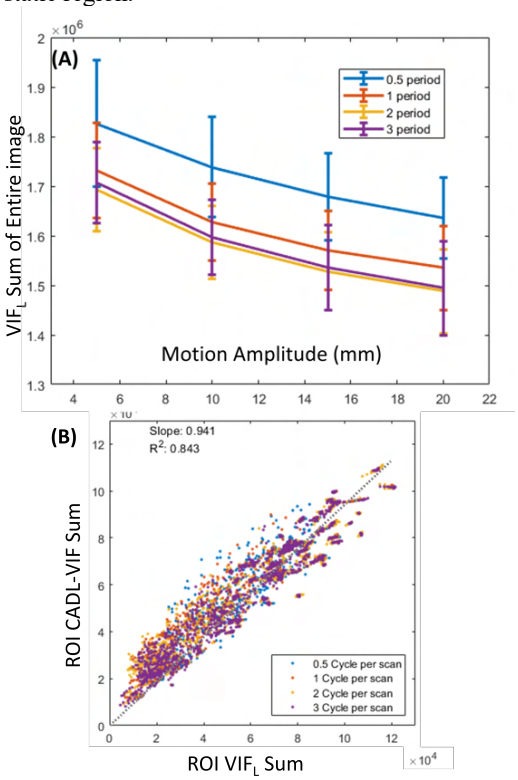


Figure 2. (A) Sum of VIF_L across the entire volume as a function of nominal motion amplitude, for a set of motion frequency values. (B) Agreement between the reference-free CADL-VIF inferences and conventional, reference-based VIFL, both integrated over 64x64x64-voxel ROIs placed in the volume.

III. Results

A. DL-VIF Map and Deformable Motion Severity

Figure 2 (A) shows the variation in VIF_L as a function of motion amplitude and motion frequency, to validate is capability to accurately quantify the effects of motion on image patches. VIF_L values show decreasing trend with increased amplitude (larger motion) and motion frequency (faster motion), making it a suitable metric to quantify motion induced image quality degradation. Validation of the capability of VIF_L to quantify motion is accompanied by assessment of the agreement between reference-based VIF_L maps and CADL-VIF inferences, illustrated in Fig. 2 (B). CADL-VIF showed good

agreement with reference VIF_L across volumes throughout the entire extended dataset, achieving a slope of 0.941 and a linearity of 0.843. Combining these two results: i) VIF_L can accurately reflect the severity of local motion-induced image quality degradation for an extended range of deformable MVF; ii) the combination of contextual and local features learned with CADL-VIF are representative of motion artifacts decreasing the local value of VIF_L, making CADL-VIF a suitable metric for reference-free motion quantification.

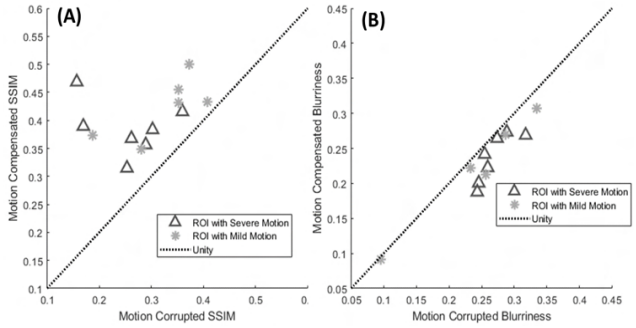


Figure 3. SSIM values (A) and blurriness (B) before and after motion compensation for regions with severe (black triangles) and mild motion (grey stars). The dashed lines mark the unity line, in which motion compensated and motion-corrupted images are equivalent. Values above the unity line in (A) indicate a net increase in SSIM after motion compensation. Values below the unity line in (B) are associated to sharper images.

B. Motion Compensation with CADL-VIF Map

Figure 3 quantitatively illustrate the performance of deformable autofocus based on CADL-VIF for compensation of motion using a multi-ROI approach. CADL-VIF autofocus resulted net improvement in SSIM for all 7 motion cases, yielding an average increase in SSIM of 0.129 for regions with severe motion and 0.113 for regions with mild motion. It is worth noting that the seemingly low SSIM is likely due to the low contrast in soft tissue and the presence of noise, and it is the increase in SSIM value that reflects improved image quality, as illustrated in Figure 4, which shows an example compensation of severely motion-distorted anatomy. Figure 3 (B) show the reduction of image blurriness after motion compensation with CADL-VIF autofocus, with an average reduction of 0.032 in severe motion regions and 0.052 in mild motion regions, which is expected in successful motion compensation.

Consistent improvement in SSIM and reduction in blurriness can be qualitatively appreciated in Figure 4. Motion artifacts severely distorted anatomical structures, with noticeable impact at the center of the volume (region of larger motion amplitude). The distortion and blurriness resulted in a sizable reduction of SSIM, to a value of 0.157. Autofocus motion compensation with CADL-VIF successfully restored the appearance of anatomical structures, mitigated shape distortion, and reduced image blurring, yielding a 3-fold increase in SSIM to 0.469.

IV. Discussion And Conclusion

This work presents a new learning-based image quality metric (CADL-VIF) to quantify the effect of CBCT deformable motion within a local region of interest. The proposed network builds on our previous approach for rigid autofocus with learned metrics providing simultaneous quantification of image quality and structural integrity of the underlying anatomy, by integrating contextual information and extending the reference similarity metrics from scalar values to spatially varying distributions for generation of voxel-wise maps of motion artifacts and distortion on high-resolution ROIs. Such locality is crucial for integration into multi-ROI deformable autofocus compensation methods that would otherwise be unfeasible due to large computational and memory requirements of performing the compensation in the complete volume, while the integration of contextual information allows robust estimation of motion effects by mitigating the effect of confusing factors associated to local variations of image content.

CADL-VIF was able to accurately reproduce the motion quantification capability of the reference similarity metric. When integrated into the CBCT deformable autofocus framework, CADL-VIF proved capable of recovering fine details in soft tissue structures that challenge conventional metrics. Ongoing work targets application of CADL-VIF to clinical data scenarios via training with simulated datasets including complete models of the CBCT imaging chain.

References

[1] A. Sisniega, et al, Phys Med Biol. 2017 ;62(9)
[2] S. Capostagno, et al, Phys Med Biol. 2021 ;66(5)
[3] H. Huang, et al, Phys Med Biol. 2022; in submission
[4] H. Huang, et al, arXiv, Fully3D 2021: 67-71
[5] H. R. Sheikh, et al, IEEE, 2006; 15(2)
[6] A. Sisniega, et al, SPIE. 2021; 241
[7] M. Elsayed, et al, Int J Simul Sci Technol. 2018; 19(4)
[8] K. Klaser, et al, Applied Science, 2021; 11(4)
[9] N. Hansen, CMA-ES, 2014

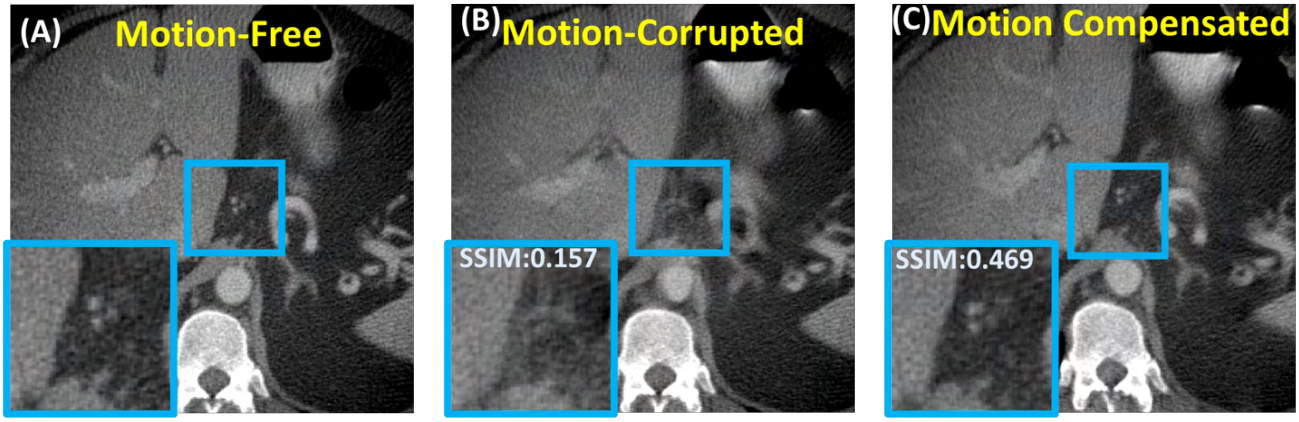


Figure 4. Example of an instance of motion compensation using the CADL-VIF, on one of the test anatomies and motion trajectories.

Simulation of Random Deformable Motion in Soft-Tissue Cone-Beam CT with Learned Models

Y. Hu, H. Huang, J. H. Siewerdsen, W. Zbijewski, M. Unberath, C. R. Weiss, and A. Sisniega

***Abstract*—Cone-beam CT (CBCT) is widely used for guidance in interventional radiology but it is susceptible to motion artifacts. Motion in interventional CBCT features a complex combination of diverse sources including quasi-periodic, consistent motion patterns such as respiratory motion, and aperiodic, quasi-random, motion such as peristalsis. Recent developments in image-based motion compensation methods include approaches that combine autofocus techniques with deep learning models for extraction of image features pertinent to CBCT motion. Training of such deep autofocus models requires the generation of large amounts of realistic, motion-corrupted CBCT. Previous works on motion simulation were mostly focused on quasi-periodic motion patterns, and reliable simulation of complex combined motion with quasi-random components remains an unaddressed challenge.**

This work presents a framework aimed at synthesis of realistic motion trajectories for simulation of deformable motion in soft-tissue CBCT. The approach leveraged the capability of conditional generative adversarial network (GAN) models to learn the complex underlying motion present in unlabeled, motion-corrupted, CBCT volumes. The approach is designed for training with unpaired clinical CBCT in an unsupervised fashion. This work presents a first feasibility study, in which the model was trained with simulated data featuring known motion, providing a controlled scenario for validation of the proposed approach prior to extension to clinical data. Our proof-of-concept study illustrated the potential of the model to generate realistic, variable simulation of CBCT deformable motion fields, consistent with three trends underlying the designed training data: i) the synthetic motion induced only diffeomorphic deformations – with Jacobian Determinant larger than zero; ii) the synthetic motion showed median displacement of 0.5 mm in regions predominantly static in the training (e.g., the posterior aspect of the patient laying supine), compared to a median displacement of 3.8 mm in regions more prone to motion in the training; and iii) the synthetic motion exhibited predominant directionality consistent with the training set, resulting in larger motion in the superior-inferior direction (median and maximum amplitude of 4.58 mm and 20 mm, > 2x larger than the two remaining direction). Together, the proposed framework shows the feasibility for realistic motion simulation and synthesis of variable CBCT data.

***Index Terms* — Interventional CBCT, Motion Simulation, Motion Compensation, Deep Learning.**

I. Introduction

Cone-beam CT (CBCT) is becoming widespread for guidance and intraprocedural imaging in interventional radiology, but it suffers from relatively long image acquisition time that makes it prone to degradation from patient motion. Motion in interventional CBCT displays a complex nature and a wide variety, spanning from rigid aperiodic motion (as in brain CBCT) to multi-source deformable motion in abdominal imaging, mixing quasi-periodic motion components (e.g., respiratory) with aperiodic, quasi-random motion (e.g., peristalsis).

Motion compensation for interventional CBCT has gained significant attention, with image-based approaches including autofocus methods based on handcrafted metrics [1-3], and methods leveraging deep convolutional neural networks (CNNs) to directly learn motion trajectories from distortion patterns [4], or to learn features associated to motion effects that are aggregated into deep autofocus metrics [5, 6]. Common to those approaches is the need for simulation methods that allow the generation of large amounts of realistic, motion-corrupted, CBCT data to enable training and evaluation. The fidelity of simulated datasets to experimental CBCT data is of dual nature: i) the data should show a realistic image appearance, attenuation pertinent to CBCT, and realistic noise and artifacts patterns; and, ii) the synthetic motion should be true to motion observed in clinical CBCT. Recent work showed the capability of fulfilling the first condition via accurate models of the CBCT imaging chain and biological tissues [7]. However, the generation of realistic motion patterns remains an open question in interventional CBCT.

Previous efforts to motion simulation yielded highly accurate models of temporal motion patterns and tissue deformation for quasi-periodic (respiratory and cardiac) motion simulation [8]. However, those models did not provide mechanisms to integrate the remaining sources of motion present in interventional CBCT, some of which feature a highly unpredictable nature (e.g., head involuntary motion or peristaltic motion).

Recent advances in deep learning-based data synthesis architectures and conditional generative adversarial network (GAN) models, have shown the capability of such approaches to learn features associated with complex underlying characteristics of the training data that, when combined with random perturbation models, allowed the synthesis of highly realistic, variable, datasets. Such methods were recently proposed for simulation of non-periodic respiratory motion in 4D CT data synthesis for image-guided radiotherapy applications [9].

In this work we hypothesize that conditional GAN models can be used to learn the underlying motion characteristics in unpaired, motion-corrupted, clinical datasets, with no prior knowledge or prior assumptions on motion nature. A GAN model is proposed, and a proof-of-concept study is presented. This proof-of-concept study used simulated data with known motion fields. The training was completely agnostic to the known motion, analogous to training with clinical datasets, but knowledge of the true motion pattern allowed validation of the characteristics of the random synthetic motion generated by the trained model.

II. Material And Methods

A. Learning Complex Deformable Motion with a GAN model

The proposed GAN architecture is illustrated in Fig. 1. The proposed approach leveraged the concept of Partial Angle

This work was supported by the National Institute of Health under Grant R01-EB-030547.
Y. Hu, and M. Unberath are with the Dept. of Computer Science, Johns Hopkins University, Baltimore, MD, USA.
H. Huang, J. H. Siewerdsen, W. Zbijewski, and A. Sisniega, are with the Dept. of Biomedical Engineering, Johns Hopkins University, Baltimore, MD, USA.
C. R. Weiss is with Russel H. Morgan Department of Radiology, Johns Hopkins University, Baltimore, MD, USA.

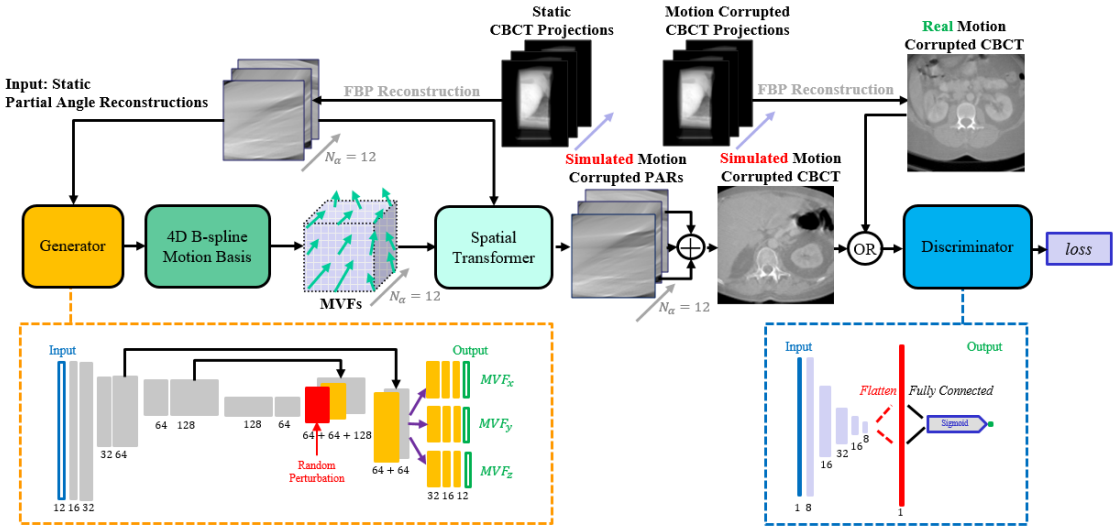


Figure 1. Schematic depiction of the realistic motion simulation framework, based on a deep generative adversarial network model. Motion-free simulated CBCT projections are divided into $N_\alpha = 12$ groups of consecutive projections that generate a set of N_α. The generator network, based on the U-NET architecture, receives at input the set of static PARs and a random perturbation of the latent space. The sparse 4D MVF output by the generator induces the deformation into the individual PARs that were added together into the final motion-corrupted volume. For GAN training, a discriminator receives as input the generated simulated motion volumes and samples of the real motion-corrupted volumes from the training dataset.

Reconstruction (PAR) combined with spatial transformer modules, previously used in CBCT motion compensation [4]. A complete, motion-free, CBCT projection dataset, with a total of 360 projections, is reconstructed into $N_\alpha = 12$ PAR volumes of 512 × 512 × 128 voxels, each containing the backprojection from 30 consecutive projection views. Those PARs are the input to the generator network, which outputs N_α sets of 36 × 36 × 12 B-spline coefficients that serve as a lower dimensionality representation for each of the N_α motion vector fields (MVFs) representing the simulated 4D deformable motion of the volume. Dense MVFs, with size equal to that of the PAR volumes, are then generated via B-spline interpolation. The set of dense MVFs and the original motion-free PARs are then input to a spatial transformer module that applies the simulated deformation to each of the PARs and add them together to obtain a final motion-corrupted volume. In the resulting architecture PARs are considered static, effectively assuming a piecewise constant temporal motion trajectory. A discriminator network was used to provide a GAN loss discriminating between real and simulated motion-corrupted volumes. The architecture of each of the components is discussed below.

Generator: The generator featured a non-symmetrical 3D UNet-like [10] structure with a 3-stage encoding branch, and 2-stage decoding branch. The encoding branch received the motion-free PARs as the input and extracted into the latent space features associated to structural content of the image associated with motion characteristics. This way, the input, motion-free PARs act as the condition variable of the conditional GAN architecture. Each stage on the encoder branch included a set of two 3×3×3 convolution layers, batch normalization, leaky ReLU activation, and a final 2x max pooling layer.

The set of latent space features were combined with a Gaussian random perturbation field to generate random, distinct, motion patterns for a given input condition during both training and inference time. The set of latent features and random perturbation entered the decoder branch, with 2 stages implementing a 3×3×3 convolution layer, a batch normalization layer, and leaky ReLU activation, followed by a 3×3×3 transposed convolution for up-sampling of the feature maps. Skip connections were placed between equivalent levels of the encoder and decoder branches. The output of the decoder is input to three branches implementing a cascade of two 3×3×3 convolution layers, with leaky ReLU activations, that generate the B-spline coefficients for the directional components of the MVFs in the antero-posterior (AP), lateral (LAT), and superior-inferior (SI) directions.

Discriminator: The discriminator acts on motion-corrupted CBCT volumes to predict whether the input comes from a real or simulated instance. During training, the Binary Cross Entropy (BCE) loss was calculated for the simulated and real datasets, and the total loss was defined as the average of both.

In the proposed model, the discriminator featured a cascade of 5 convolution layers (4×4×4 kernel), followed by batch normalization, leaky ReLU activation, and a dropout layer (0.2 dropout). The final fully connected layer (with sigmoid activation) acted on the flattened set of features.

B. Data Generation and Motion Model

For this proof-of-concept study, training and validation data were generated from 70 Multi-detector CT (MDCT) abdominal datasets from the TCIA Lymph Node Abdomen collection. 60 distinct MDCT instances were used for training, 5 for validation, and 5 for testing. For each source MDCT volume, we randomly selected a subvolume of 128 mm length at a random longitudinal position within the abdomen. The

subvolume was then forward projected using a high-fidelity CBCT model with a geometry with source-to-detector distance of 1200 mm, and source-to-axis distance of 785 mm. The detector was modeled as a flat panel with 576 × 440 pixels (0.616 mm isotropic pixel size).

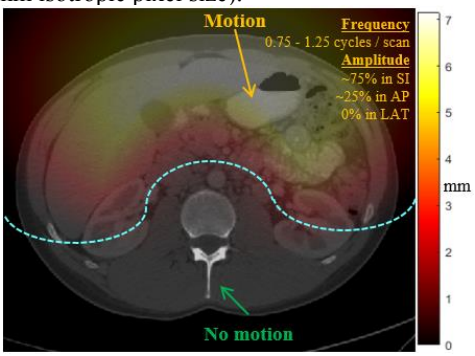


Figure 2. Example training case. Deformable motion was induced following with a motion field with randomly selected central location and amplitude. In this proof-of-concept study, the spinal region was automatically detected and an exclusion mask with smooth boundaries was defined (see cyan delineation). The motion field was attenuated to preserve the spine region quasi-static.

Motion corrupted datasets were obtained by inducing deformable motion during forward projection. The simulated motion field followed a cosine temporal trajectory with random frequency between 0.75 – 1.25 cycles per scan and random phase. Spatial distribution of motion amplitude was modelled as an elliptical field with maximum amplitude (randomly set between 10 and 25 mm) at the center, and randomly placed at a soft-tissue region of the volume. The amplitude faded following a Gaussian decay curve that reached zero at the ellipse axis length, randomly chosen from 200 to 300 mm in the medial-lateral (LAT) direction and between 100 and 150 mm in the antero-posterior (AP) direction. Motion amplitude was kept constant across slices. Motion direction was randomly chosen, allocating between 60% and 80% of motion to the SI direction and the rest to the AP direction, with no lateral motion.

To avoid unrealistic large motion of the spine region, the center of the spine was detected in the volume and a cylindrical motion-exclusion mask with 100 mm radius was defined. The mask performed a smooth transition from one to zero and multiplied the motion field, to minimize the motion in the spine, as illustrated in Fig 2.

The motion-corrupted datasets featured 3 distinct properties that were used for validation of the GAN capability for inference of consistent motion instances: i) the induced motion was composed of diffeomorphic deformations; ii) the spine region remained nearly static for all training instances; and, iii) the majority of the motion was allocated to the SI direction with the rest in the AP direction.

Motion-corrupted datasets were reconstructed into volumetric grids of 512 × 512 × 128 voxels with 0.5 × 0.5 × 0.5 mm^3 voxel size, and motion-free cases were reconstructed into 12 PARs with equivalent parameters. The PARs were downsampled to 128 × 128 × 32 voxels for input to the generator.

C. Network training and validation studies

A total of 720 motion-corrupted volumes based on 60 anatomical instances were included in the training set, and 15 instances from 5 separate anatomical structures were used for validation. Each dataset contained a motion-free collection and a motion-corrupted collection. During training, a sample with N_α = 12 PARs from the motion-free collection was randomly selected as input to the generator, while one sample from motion-corrupted collection was input to the discriminator. Training was performed with an unbalanced scheme in which the generator is updated every 1 batch while the discriminator is updated every 2 batches for 100 epochs, with a batch size of 12. We used the ADAM optimizer for both generator and discriminator with learning rates of 10^{-5} and 10^{-4}, respectively. BCE Loss was selected as objective function to be maximized by generator while minimized by discriminator.

To validate the results, we used a test set of 15 samples based on 5 anatomical structures not seen by the network. For testing, static PARs were input to the generator together with the Gaussian random perturbation. Validity of the generated motion fields was validated via measurements of diffeomorphism, based on the determinants of the Jacobian of the deformation, and metrics of average displacement at regions of maximum motion and regions static in the training set. Furthermore, directional components of the synthetic motion were evaluated in comparison with underlying trends in the training dataset.

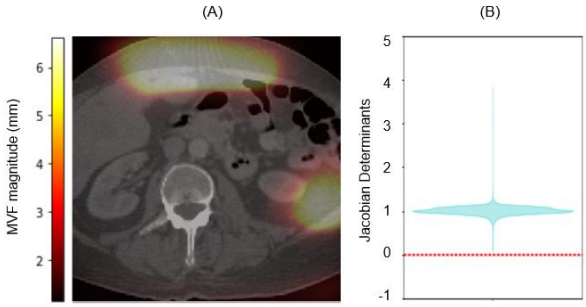


Figure 3. (A) Motion vector field on an example inference instance, with motion predominantly present in soft-tissue structures. (B) Validation of the diffeomorphism of the generated deformable motion fields for the aggregated test set.

III. RESULTS

Figure 3 illustrates the predominant soft-tissue nature of the simulated motion, as well as its diffeomorphic nature. An example simulated MVF is shown if Fig. 3A, demonstrating the majority of the deformation induced to anterior soft-tissue regions with minimal deformation towards the central posterior area, where the spine is located. Fig. 3B shows the accumulated distribution of Jacobian determinant values across the ensemble of test datasets. The induced deformable motion vector fields consistently show Jacobian determinant values larger than zero, consistent with diffeomorphic motion.

Fig. 4A shows the average displacement for the aggregated motion synthesis dataset, obtained by adding the absolute value of the motion amplitude for each time point (viz. PAR) and normalizing the result by the total number of PARs (N_α = 12). Average displacement was evaluated in a soft-tissue region in the anterior area of the abdomen and in a region inside the spine. Results show displacement values in anterior soft-tissue areas of 3.9 ± 2.5 mm, while spine regions showed minimal motion, with average displacement of 0.5 ± 0.0 mm. Fig. 4B shows the directional properties of the random motion instances

synthesized by the GAN model. Consistent with the trends underlying the training data, the synthetic motion fields exhibited larger motion in the SI direction with median amplitude of 4.58 mm and ranging upwards of 20 mm consistent with voluntary or involuntary respiratory motion.

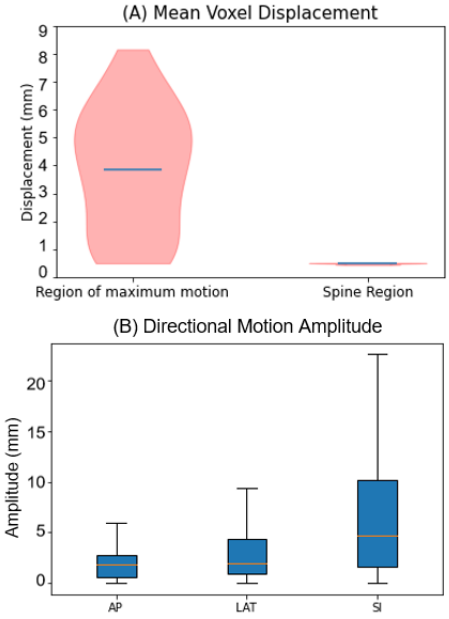


Figure 4. Quantitative evaluation of the synthetic random motion. (A) Average displacement of voxels in the maximum amplitude and in the spine regions, showing preservation of the quasi-static nature of the spine. (B) Motion amplitude in the AP, LAT, and SI directions for the generated random motion vector fields.

Motion amplitude in the AP and LAT directions was lower, with median values of 2.02 mm, and 2.18 mm, respectively. The comparable amplitude observed for the AP and LAT directions illustrate the challenge in differentiation of motion patterns that can result in similar artifacts, as lateral motion was minimal in the training data.

Note that while the current design did not implement any control mechanism on the output motion amplitude, several options for such controlled simulation can be easily integrated, including coarse stratification of the training data into mild, moderate, or severe cases within a semi-supervised training strategy; normalization of the output motion fields; or controlled scaling of the latent space random perturbation.

Validation of the realism of the synthetic motion-corrupted datasets and of capability to generate distinct motion for a given input is shown in Fig. 5. Image results in Fig. 5 show distinct, realistic motion artifacts in soft-tissue regions, with minimal distortion of the (static) spine. Quantitative evaluation of motion amplitude in Figs. 5D and 5H illustrates the generation of variable motion patterns for single input conditions.

IV. Discussion and Conclusion

This work presented an adversarial model for simulation of realistic, random, deformable motion in CBCT using motion-corrupted datasets with no prior assumptions on the motion characteristics. The framework was evaluated in a controlled study in which the properties of the random synthetic motion fields were compared with known motion trends underlying in the training data cohort. The model was able to generate distinct motion instances, while replicating principal properties of the training dataset, such as diffeomorphism, proper spatial distribution of motion amplitude (maximized anteriorly and minimized posteriorly), and predominantly SI motion direction in agreement with learned patterns. The results enable generation of large training datasets for development of deep learning autofocus methods.

V. References

[1] A. Sisniega, et al. 2017. Phys. Med. Biol. 62 3172.
[2] J. Hahn, et al. 2017. Phys. Med. Biol. 44 5795.
[3] S. Capostagno, et al. 2021. Phys. Med. Biol. 66 055010.
[4] J. Maier, et al. 2021. Med Phys. 48 3559.
[5] A. Preuhs, et al. 2020. Arvix.Org. 2006.10390
[6] H. Huang, et al. 2021. Arvix.Org. 10.1117.
[7] P. Wu, et al. 2021. Arvix.Org. 2108.09229.
[8] S. WP, et al. 2008. Med Phys. 8 3800-8.
[9] Y.Chang, et al. 2021. Phys. Med. Biol. 66 11
[10] O. Ronneberger, et al. 2015. Arvix.Org. 1505 04597.

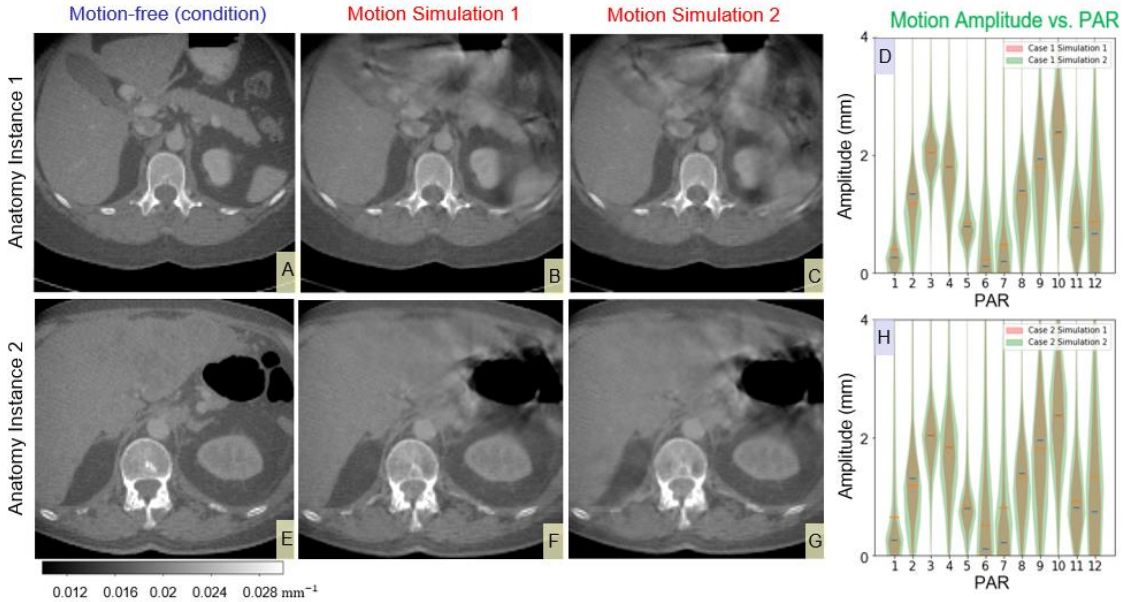


Figure 5. Example motion simulation cases for two motion-free source anatomical instances (A, E), and two instances of the latent space random perturbation (B, C, and F, G), showing distinct motion artifacts but respecting the learned properties in terms of motion distribution and main direction. (D, H) Distribution of motion amplitude as a function of time (viz. PAR index) for the random realization in (B, F) in red and the motion realization in (C, G) in green.

A five-dimensional cardiac CT model for generating virtual CT projections for user-defined bolus dynamics and ECG profiles

Eri Haneda, Bernhard Claus, Jed Pack, Darin Okerlund,
Albert Hsiao, Elliot McVeigh, and Bruno De Man

***Abstract*— Cardiac CT exams are some of the most complex CT exams due to need to carefully time the scan to capture the heart during the quiescent cardiac phase and when the contrast bolus is at its peak concentration. We are interested in developing a robust and autonomous cardiac CT protocol, using deep learning approaches to extract contrast timing and cardiac phase timing directly from pulsed projections. In this paper, we present a new approach to generate large amounts of clinically realistic virtual data for training deep learning networks.**

We propose a five-dimensional cardiac model generated from 4D cardiac coronary CT angiography (CTA) data for synthetic contrast bolus dynamics and patient ECG profiles. We apply deep learning to segment 7 heart compartments and simulate intravenous contrast propagation through each compartment to insert contrast bolus. Additional augmentation techniques by randomizing a bolus curve, patient ECG profile, acquisition timing, and patient motion are applied to increase the amount of data that can be generated. We demonstrate good performance of the deep learning segmentation network, examples of simulated bolus curves using a realistic protocol, and good correspondence between virtually generated projections and real projections from patient scans.

***Index Terms*— Cardiac CTA, CT simulation, deep learning, contrast bolus, cardiac phase**

I. Introduction

Cardiac CT exams such as Coronary CT Angiography (CTA) are some of the most complex CT exams due to need to carefully time the scan to capture the heart during the quiescent cardiac phase (when the heart is moving least) and when the contrast bolus in the heart chambers is at its peak concentration to achieve good contrast enhancement.

Timing the CT scan to coincide with the peak contrast concentration can be done using a 'timing bolus' or with 'bolus tracking'. With a timing bolus, a small volume of contrast is injected to a patient in a pre-session and repeated single slice axial collimated low-dose scans are performed to establish the timing delay between the start of injection and peak enhancement. Then the diagnostic cardiac CTA exam is performed with the full contrast bolus and the CT scan start after this timing delay. With bolus tracking, there is no pre-session: the full contrast bolus volume is injected, and single slice axial collimated scans are performed until the CT number in a region-of-interest reaches a predefined threshold. During the following 'diagnostic delay' of several seconds, the scan table is repositioned, breath hold instructions are delivered, and the scanner is reconfigured, after which the diagnostic CTA scan is performed. Both approaches have pros and cons, and require highly trained operators to achieve consistent bolus enhancement.

The quiescent phase of the cardiac cycle is typically estimated based on recordings from an ECG monitor and evaluated by visually assessing the degree of motion artifact on reconstructed cardiac CTA images. Based on the ECG R-peaks, the time of the next end-diastolic phase (or end-systolic phase for higher heart rates) is estimated [1]. Application of the ECG may take several minutes per exam and in some patients may lead to reliability problems, such as due improper lead positioning.

Our overall project goal is to develop a smart cardiac CT scanner that autonomously determines the optimal scan time interval without ECG, traditional bolus tracking or timing bolus, but based on real-time deep learning analytics of sparsely pulsed projections [2, 3]. However, one of the challenges to developing deep learning algorithms is collecting enough data to train without exposing volunteers to ionizing radiation. To address this challenge, we here present a new approach to generate virtual CT projection data at any view angle, bolus dynamics, and cardiac phase, based on a five-dimensional model of the cardiac CT volume, derived from retrospectively collected clinical images and using a series of data augmentation steps.

II. Methods

Figure 1 shows an overview of our virtual data generation scheme. We first derive five-dimensional cardiac CT models from multi-phase clinical cardiac (cine) CT scans by segmenting the heart compartments and identifying a blood (or contrast) flow propagation map in each compartment. To model contrast concentration at multiple bolus time points from datasets that were acquired at (approximately) a single bolus time point, we segment the cardiac compartments, we parametrize the voxels inside those compartments based on

Eri Haneda (haneda@ge.com), Bernhard Claus, Jed Pack, and Bruno De Man are with GE Research, Niskayuna, NY. Darin Okerlund is with GE Healthcare, Waukesha, WI. Albert Hsiao and Elliot McVeigh are with University of California San Diego, La Jolla, CA.

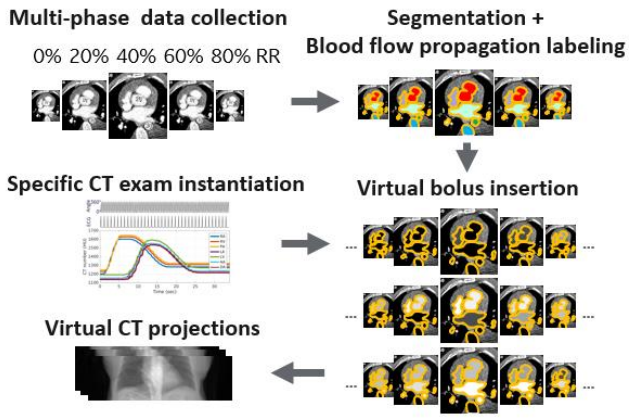


Fig. 1: Overview of our virtual data generation scheme

their location along the flow direction, and then increment the voxel values to model different bolus distributions based on location and time point. We then define multiple instantiations of CT exams based on specific timing of cardiac cycle, contrast bolus, and CT acquisition to generate virtual CT projection data. Each of the steps is described in detail in the next paragraphs.

Clinical datasets: We retrospectively collected multiple phase (cine) cardiac CT data from 40 Transcatheter Aortic Valve Replacement (TAVR) patients. The data were acquired under IRB approval (IRB #191797X) at University of California San Diego using a GE Revolution CT scanner with a 16 cm z-coverage to image the whole heart in one rotation. ECG data was acquired simultaneously, and iodine contrast agent (bolus) was administered to each patient. Cine scan mode was used to image all cardiac phases, resulting in 1.1-1.4 sec scan times and covering at least one full R-R cycle. Phase-specific cardiac volumes were retrospectively reconstructed every 70ms. The image volume size was 512x512x256 voxels covering the full 50-cm-diameter field-of-view and 16 cm in the z direction. Approximately 15-20 volumes (phases) were reconstructed for each patient and the associated R-R% was extracted from the ECG. By interpolation, a cardiac CT image dataset can be extracted at any R-R%, which represents the fourth dimension in the five-dimensional model.

Segmentation: A subset of 149 datasets representing multiple phases were selected from 28 different patients. For each dataset, 7 compartments were manually segmented: right atrium (RA), right ventricle (RV), pulmonary artery (PA), left atrium (LA), left ventricle (LV), ascending aorta (AA), and descending aorta (DA), resulting in 149 3D masks. The PA region was further segmented into three regions: main pulmonary artery (MPA), left pulmonary artery (LPA), and right pulmonary artery (RPA). Any overlap between two adjacent compartments is removed by assigning the overlap voxels to one of the compartments. Any gaps between two adjacent compartments are avoided by inserting a 3D cylinder padding between them. Figure 2 shows an example of the resulting compartment segmentations with color coding.

Blood flow propagation labeling: We defined blood flow propagation labels by parametrizing the voxels inside the compartments based on their position along the blood flow (or bolus) propagation direction. The label assigned to each voxel are integer numbers that approximately represent how the blood propagates from RA to RV, and PA, then from LA to LV, to AA, and to DA. The propagation direction is determined by multiple reference points: the center-of-mass locations of each compartment and a set of touch points where each pair of compartments touch each other. From the manual segmentation masks, we first find touch points between RA-RV (RAV touch), RV-PA (ROV touch), LA-LV (LAV touch), and LV-AA (LOV touch), where O refers to an outgoing vessel from a ventricle.

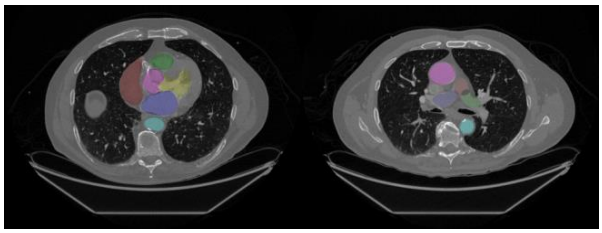


Fig. 2: An example of manual segmentation. The left image contains right atrium (red), right ventricle (green), ascending aorta (pink), left atrium (purple), left ventricle (yellow), and descending aorta (light blue). The right image contains ascending aorta (pink), main pulmonary artery (light red), left pulmonary artery (light green), and right pulmonary artery (light purple), and descending aorta (light blue).

We developed two blood propagation models depending on the compartment. For RA, LA, PA, AA, and DA, a 'linear propagation model' is used. For **RA**, we assign linearly increasing values along the line which passes from the center-of-mass of the atrium (RA_{ctr}) to the RAV touchpoint. Then, each voxel is labeled based on its projected location on the line, normalized from 100 to 199, such that the blood flow mask value increases linearly in the direction from 100 at the entrance to 199 at the exit of the atrium. Similarly, the **LA** voxels are labeled from 400 to 499. For **PA**, we find the end of the LPA branch and the end of RPA branch by identifying maximum distance from ROV touch. Then, voxels are labeled linearly from 300 at ROV touch to 399 at the end of the LPA and the RPA. For **AA**, the direction for linear propagation is defined by the line from LOV touch through the center of the AA (AA_{ctr}) and voxels are labeled from 600 to 699. For **DA**, assuming that DA is a straight vessel running in z direction, the voxels are labeled from 700 at the maximum z to the pixel to 799 at the minimum z position. For **RV** and **LV**, an 'angular propagation model' was developed. First a rotation center point was defined outside of the ventricle, triangulated from the center-of-mass of the ventricle (V_{ctr}) and the two adjacent touch points. Then an angle is computed for each voxel relative to that line, encoding a circular path. Voxels are labeled from 200 to 299 in the RV and from 500 to 599 in the LV.

Blood flow propagation labels were computed for all 149 segmented volumes and will be used for deep learning network training and validation, as described in the next section.

Deep learning segmentation and blood flow propagation labeling - We trained a 3D Unet [4] (Figure 3) to perform both segmentation and voxel labeling to model the blood propagation direction. The input to Unet are patient images, truncated to a 30cm field-of-view and then downsampled to (152x152x64). The network has two output of the same size. The first output is the probability prediction for a voxel to be inside the bloodpool (i.e. one of the 7 compartments). The

second output volume is the voxel label prediction. Finally, both output volumes are combined by setting zero to all voxel coordinates that have less than 50% probability, and the resulting volume is upsampled to the original voxel size.

A two-part loss function was used with one part being a mean squared error on the voxel labels. Outside the ground truth mask, this error was set to zero. The second part of the loss function was a piecewise linear function on the absolute error of the mask output. This function has three sections, with the steep portion residing between 0.4 and 0.6.

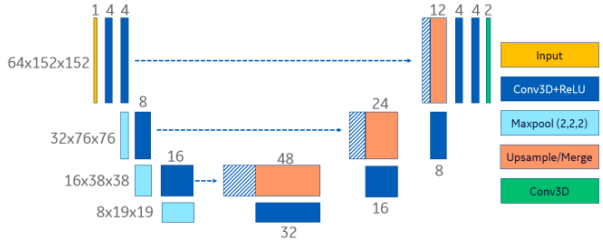


Fig. 3: Neural network architecture to generate a blood flow mask

We finally used this network for inference on additional 488 volumes. We obtained a total of 637 blood flow propagation masks (149 manually and 488 using Unet) from 40 patient data.

Bolus dynamics definition: The obtained volumes are preprocessed to remove the estimated real average bolus level from all compartments and then adding in the simulated bolus level. A simulated bolus time sequence B(t) is defined by the parameters in Figure 4. The bolus curve remains 0 HU for a given *Bolus Delay*, rises from 0 HU to *Bolus Height* during the *Bolus Rise Time* (with a cosine shape), remains at *Bolus Height* for the *Bolus Peak Width*, decays from *Bolus Height* to *Converged Bolus Level* during the *Bolus Decay Time* (with a cosine shape), and remains at *Converged Bolus Level* for the remainder of the time.

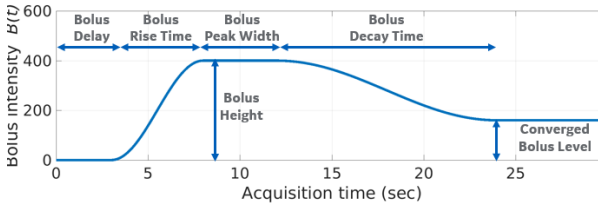


Fig. 4: Bolus curve definition by parameters

The new bolus level is assigned to voxels based on the current simulated scan time and the delay associated with each voxel location. The time delays at the compartment boundaries are set to 0RR, 0.5RR, 1.0 RR, and 1.5RR for the right side of the heart, where RR refers to the R-to-R interval. The pulmonary circulation delay is then added before re-entering the left heart side, after which the following set of additional delays are applied 0RR, 0.5RR, 1.0 RR, 1.5RR and 1.8RR for the compartment boundaries on the left side of the heart. To implement dispersion of bolus in the left cardiac region relative to the right cardiac region, a trapezoid filter is applied. That means the bolus levels for RA, RV, and PA follow the original bolus curve, and the bolus levels for LA, LV, AA, and DA follow the dispersed bolus curve. We also scaled all delays by a random number between [0.3 0.55]. For this particular study we also averaged the bolus curve over each period between two time points at neighboring compartment boundaries and assigned the average value uniformly to each compartment. In future work, this model will be refined to better model realistic behavior in discussion with experts in cardiac blood flow.

Protocol instantiation: We simulate series of acquisitions by randomly selecting a patient, acquisition time points, cardiac conditions, bolus curve parameters, and patient rigid motion parameters. In this study, only anterior-posterior (AP) and posterior-anterior (PA) view acquisitions are simulated with 125 rotations. The rotation time is set to 0.28 sec, therefore data is acquired every 0.14 sec and the total scan time is 35 sec. For each instantiation, we randomly select a bolus curve and a cardiac condition. Table 1 and 2 summarize the value range of bolus curve parameters and cardiac motion parameters. For each acquisition time point t_{aq}, we find two closest R-peaks time (t_{R1}, t_{R2}) from the patient's ECG to compute $RR\% = (t_{aq} - t_{R1})/(t_{R2} - t_{R1})$. We then interpolate a new dataset from the original patient dataset based on the respective RR%. In this study, we used nearest neighbor interpolation.

Table 1: Bolus curve parameters

Parameter	Value range
Bolus Delay	[0.5, 3.5] sec
Bolus Rise Time	[2, 12.5] sec
Bolus Peak Width	[1.0, 7.0] sec
Bolus Decay Time	[7.0, 15.0] sec
Converged Bolus Level	[15, 65]% of peak height
Bolus Height	[120 180]%
Trapezoid filter full-width	[60, 95]% of pulmonary circ delay

Table 2: Cardiac motion parameters

Parameter	Value range
Heart rate	[40, 75] bpm
RR% offset at scan start	[0, 1]
Pulmonary circulation delay	[4.0, 9.0] sec

Extrapolation, Augmentation, and Reprojection:

The voxels outside cone beam are extrapolated from the neighboring slice if available. Extra z slices are also padded on the top and bottom by repeating the boundary slice to prevent boundary artifacts during forward-projection. Then, a random rigid motion augmentation was performed per patient by rotation, scaling and shifting to add more anatomy variation. Finally, we forward-projected the volumes using a distance-driven projector [5] and modeling a GE Revolution CT cone beam geometry.

III. Results

Figure 5 shows a specific CT exam instance as a function of time. The top row shows the CT gantry rotation angle. The second row shows the patient ECG signal. The 7 colored curves show CT number averaged over each of the 7 cardiac compartments. For each curve, we can clearly observe a rising edge, a plateau, and a more gradual decay. The large delay between compartments in the right left sides of the heart is due to the pulmonary circulation. For the same reason there is some additional dispersion in the curves for the left side of the heart. The bottom row shows the time of the injection, the start of the pulsed-mode projections (PMPs), the breath-hold command, and the actual CTA scan. Note that the gantry rotation angle, ECG signal, and PMPs are shown for illustration purpose. The spacing is not exact.

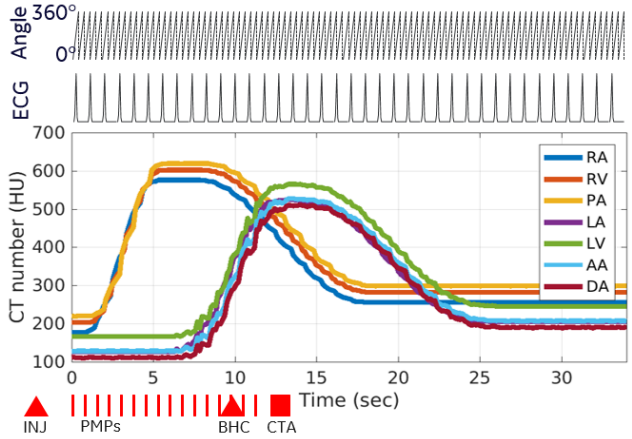


Fig. 5: Example of a CT exam instance as a function of time. From top to bottom: gantry rotation angle, ECG signal, average CT number in the 7 cardiac compartments, and the timing of injection (INJ), pulsed-mode projections, (PMPs), breath-hold command (BHC), and CTA scan.

Figure 6 shows example CT images and the corresponding compartment segmentation and blood flow propagation labels. The color coding reflects the integer labels from 100 to 799. The top row shows two examples of CT images and the corresponding the labels, which were analytically generated. The bottom row shows two test CT images with the inferred segmentation and integer labels. The cardiac regions were quite accurately segmented, and the blood flow propagation labels in seven compartments were visually similar to the training examples from the top row.

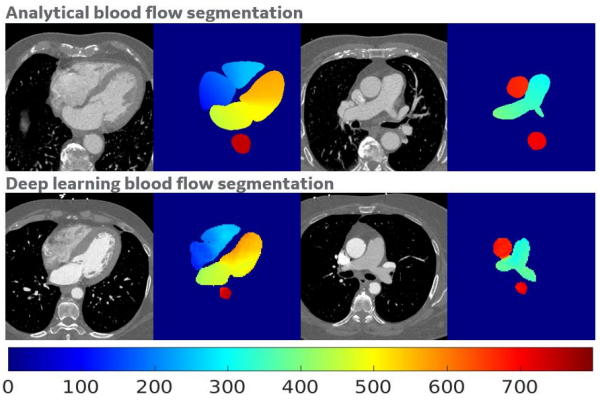


Fig. 6: Four example CT images and corresponding segmentation and blood flow propagation labels: the two training examples show the analytically computed labels (top row) and the two test examples show the inferred values (bottom row).

Figure 7 shows a sequence of AP projections, comparing real measurements (left) and virtual projections (right). The top row shows the first actual projections in the sequence. The next 4 rows show the difference images of the subsequent projections, i.e.: subtracting the first projection in the respective sequence. In the real data, the heart rate was 58.5 bpm and the cardiac phase was 0.89, 0.17, 0.44, 0.71, and 0.99 %RR from the top to bottom. The simulated data was generated using used these same parameters. Bolus insertion and motion augmentation were skipped. The simulated projections visually match the real projections quite well. More work is needed to verify the accuracy of the virtual bolus insertion.

IV. Conclusion

In this paper, we proposed an innovative method to generate large amounts of virtual but clinically realistic cardiac CT projection data. Our approach used a five-dimensional model of the cardiac CT volume derived from multi-phase 3D cardiac CT images with programmable bolus dynamics and cardiac phases. This dataset could be used to train a real-time deep learning network which determines the optimal scan time from raw data without ECG and traditional bolus tracing or timing bolus.

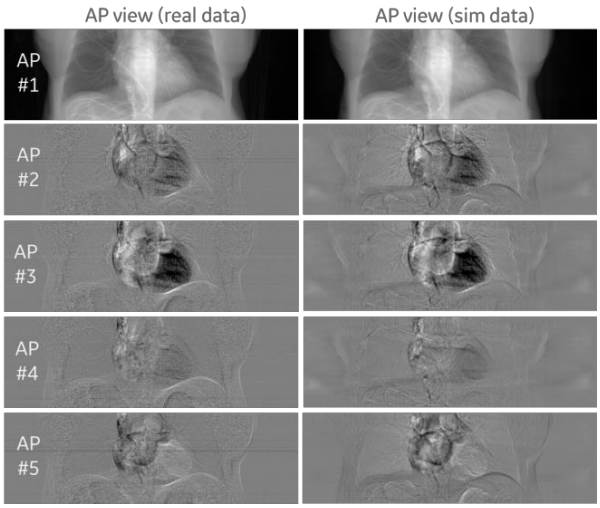


Fig. 7: Comparison between real projections (left) and simulated projections (right) for a sequence of AP projections. The first row in the sequence shows the first AP projection. The next rows show subsequent AP projections after subtracting the first projection in the sequence.

Acknowledgement

Research reported in this publication was supported by the NIH/NHLBI grant R01HL153250. The content is solely the responsibility of the authors and does not necessarily represent the official views of the NIH.

References

[1] B. Ohnesorge and T. Flohr, "Principles of Multi-slice Cardiac CT Imaging," in Multi-slice and Dual-source CT in Cardiac Imaging, 2nd ed., Springer, Berlin/Heidelberg, 2007, pp. 71–126.

[2] B. De Man, E. Haneda, J. Pack, and B. Claus, "Data-based scan gating," U.S. Patent 10736594, Aug. 11, 2020

[3] J. Hsieh, S. Sirohey, R. Nilsen, S. Narayanan, and G. Cao, "Systems and methods for imaging phase selection for computed tomography imaging," U S Patent 9517042, Dec. 13, 2016

[4] O. Ronneberger, P. Fischer, and T. Brox, "U-Net: Convolutional Networks for Biomedical Image Segmentation," arXiv:1505.04597

[5] B. De Man and S. Basu, "Distance-driven projection and backprojection in three dimensions," Phys. Med. Biol. 49(11), 2463–2475 (2004).

A Virtual Imaging Trial Framework to Study Cardiac CT Blooming Artifacts

Ying Fan, Jed Pack, and Bruno De Man

Abstract— **Cardiac CT is a safe, accurate, non-invasive method widely employed for diagnosis of coronary artery disease (CAD) and planning therapeutic interventions. Even with state-of-the-art CT technology, calcium blooming artifacts may limit the accuracy of coronary stenosis assessment. A variety of solutions to reduce blooming artifacts have been proposed, including hardware improvements, protocol optimizations, and software deblooming techniques [1-6]. Hardware developments and clinical studies (for protocol optimization or training data generation) can be expensive, time-consuming, and impractical. Hence, there is an opportunity for a Virtual Clinical Trial (VCT) framework [7-8] to help researchers to evaluate the impact of various solutions on calcium blooming and to create training datasets for developing deep learning solutions for deblooming.**

In this paper, we present a new VCT framework for generating cardiac CT images with calcium blooming with a variety of CT hardware parameters, CT scan protocols and CT reconstruction kernels. As an example, we use the VCT framework to investigate the impact of three common scan and reconstruction parameters (X-ray tube voltage, focal spot size, and reconstruction kernel) on calcium blooming artifacts. We conclude that tube voltage and reconstruction kernel have the most direct impact on calcium blooming, which is consistent with earlier clinical reports [9-12].

Index Terms—**Computed tomography, Cardiac CT, Calcium blooming**

I. Introduction

Cardiac CT Angiography (CCTA) is used for the identification of significant coronary stenoses in patients and is guideline-recommended as a valuable noninvasive alternative in the diagnostic evaluation of CAD [13-14]. One of the main limitations of CCTA is inaccuracy in evaluating calcified lesions. The presence of calcified lesions leads to blooming artifacts in CT images. This may obscure the lumen and cause a false-positive CCTA. Diffuse or extensive calcifications often lead to overestimation or paradoxically, underestimation of coronary stenosis severity.

A variety of solutions to reduce blooming artifacts have been proposed. Hardware improvements have focused on smaller detector cell size to improve spatial resolution [1]. Protocol optimizations have been proposed, either optimizing the X-ray tube voltage to minimize blooming artifacts or subtracting two scans, one with and one without contrast agent [2-3]. Multiple processing techniques have been proposed, including analytical and more recently data-driven deblooming techniques [4-6].

The goal of the research presented in this paper is two-fold. First, we aim to develop a virtual clinical trial (VCT) (or more specifically virtual imaging trial) framework [7-8] to enable inexpensive and relatively fast evaluation of various solutions for calcium blooming. Second, we want to use the VCT framework to evaluate the effect of three example scan/reconstruction parameters: X-ray tube voltage, focal spot size, and reconstruction kernel. In future work, we will use the VCT framework to train and test DL-based deblooming methods.

II. Methods

A. VCT framework

Figure 1 shows an overview of the VCT framework. The framework uses two types of data: clinical images and virtual calcifications, which are combined to produce images with and without blooming artifacts, including the following steps.

Clinical images - When using existing clinical images as software phantom in CT simulations and reconstructions, some blur is introduced in the process, and the simulated-reconstructed image does not look identical to the original clinical image. To minimize that discrepancy, we developed and optimized a frequency-based blur compensation algorithm to minimize the pixelwise error before and after CT simulation and reconstruction. Figure 2 shows the error image without (left) and with (right) frequency-based blur compensation. The latter shows a very small error indicating that the resulting images are representative for real clinical images in terms of noise and resolution.

Virtual calcifications– We used an in-house semiautomatic image segmentation tool to develop a three-dimensional mesh models of virtual calcifications, matching the geometry of the surrounding *calcification-free* coronary arteries (Figure 3). The three-dimensional mesh representation is converted to a voxelized representation with a resolution of 0.1 mm, so it can conveniently be used as input to the CT simulation tool CatSim. (CatSim is also compatible with NURBS-based representations and polygonal phantoms but these were not used here). For this study, the calcifications were all defined as pure calcium.

Ying Fan (fan@research.ge.com), Jed D. Pack, and Bruno De Man are with GE Research, Niskayuna, NY.

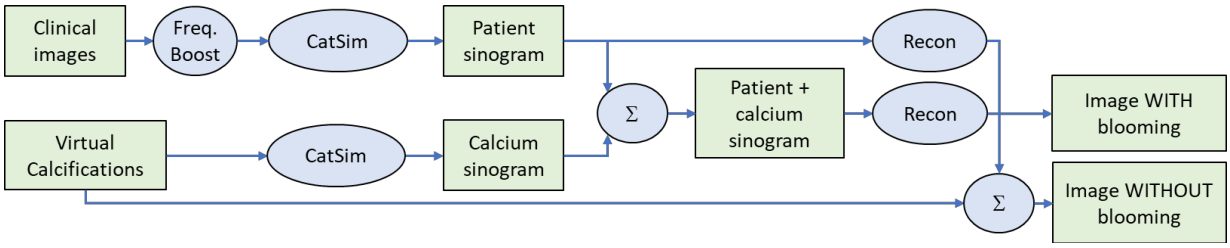


Figure 1. The VCT framework produces images with and without blooming artifacts from patient images and virtual calcifications.

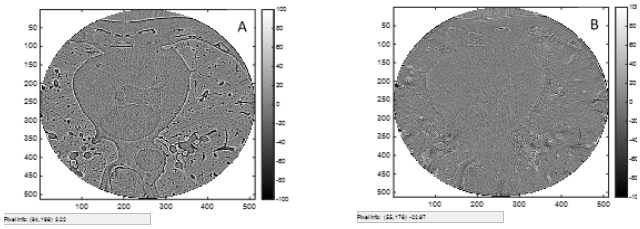


Figure 2. Difference between real clinical images and virtual clinical images without (left) and with (right) blur compensation.

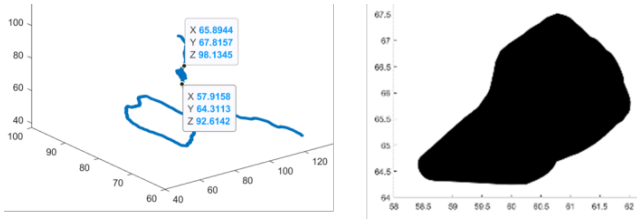


Figure 3. A virtual calcification in the context of a centerline of a coronary artery (left) and a magnified view of the virtual calcification (right).

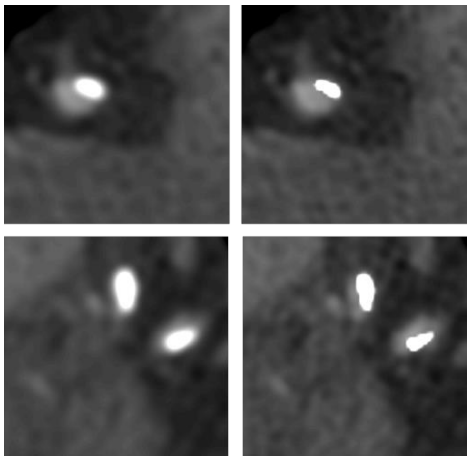


Figure 4. Examples of virtual calcified coronary artery images with (left) and without (right) blooming artifacts.

CT scanner simulation – A Lightspeed VCT scanner model in our CT simulation toolkit CatSim [15] is used to perform two sets of virtual CT scans: one set of the patient images and one set of the virtual calcifications. Unlike the clinical image (which already has all the physics effects of the real CT scan), the calcification needs to be simulated with realistic geometric blur effects related to finite focal spot, finite detector size, and rotation (azimuthal) blur. The distance-driven projector was used, which inherently accounts for finite detector size. We performed an over-sampling study to optimize the number focal spot samples, and the number of view samples for computational efficiency.

The patient sinograms are combined with the calcium sinograms and reconstructed, resulting in virtual clinical images *with* blooming artifacts. The patient sinograms are also reconstructed separately and subsequently combined with a voxelized version of the calcification, resulting in clinical images *without* blooming artifacts.

One limitation of this study is that summing patient and calcium sinograms ignores any non-linear spectral effects. This was done for simplicity and we concluded it is justified since the calcifications are small perturbations, which we have previously shown to cause minimal beam hardening impact. In future work, we will incorporate joint simulation of patient and calcifications. Figure 4 shows the example results of virtual images with and without blooming artifacts.

B. Experiments and analysis

The VCT framework was used to investigate the impact on calcium blooming of three parameters: X-ray tube voltage (kVp), reconstruction kernel, and focal spot size. The following table shows the parameters that were used in the VCT study:

Tube voltage/spectrum	70 keV (mono) 80 kVp 140 kVp
Focal spot size	1.2 mm 0.6 mm
Reconstruction kernel	Standard Detail Bone Edge
Source-to-iso	541 mm
Source-to-detector	949 mm
Detector cell size	1.02 mm x 1.09 mm
Focal spot sub-sampling	3 x 9
View sub-sampling	3

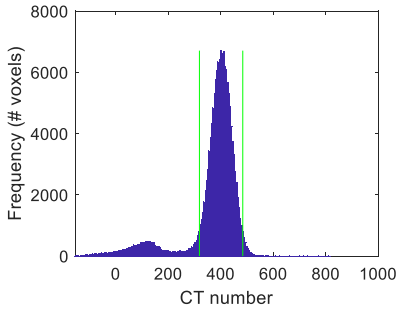


Figure 5. Histogram of CT numbers in a region around the aorta. The aorta CT number is estimated as the average of the peak around 400.

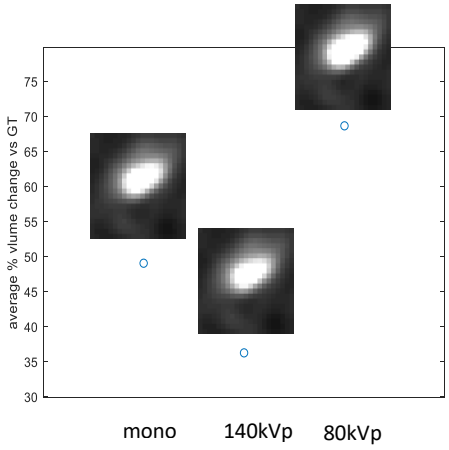


Figure 6. Impact of X-ray tube voltage/spectrum on calcium volume increase due to blooming (in %).

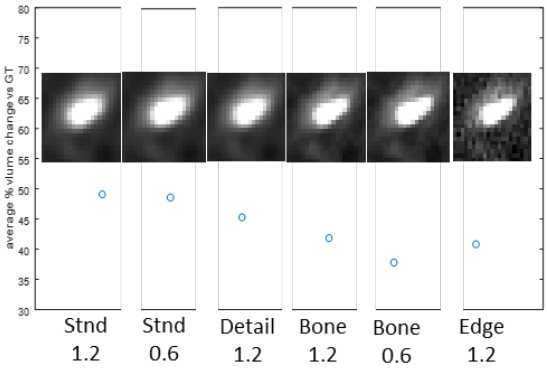


Figure 7. Impact of reconstruction kernel and focal spot size on calcium volume increase due to blooming (in %).

A total of 14 CT exams were used and combined with separate simulations of 130 embedded virtual calcifications. The CT images were acquired by Dr Pontone and the Centro Cardiologico Monzino (University of Milan, Italy) using a Revolution CT scanner (GE Healthcare, Waukesha, WI). Calcifications were detected in both the recon images and the ground truth images using the following automated approach. First a coronary artery mask is computed using a 3D segmentation tool. Second, calcifications are found based on a CT number threshold and a connected-volume threshold (optimized to be robust for image noise).

The calcium segmentation threshold was derived from the average iodine value in the aorta:

$$calcium\ threshold = avg.iodine\ at\ aorta - 25$$

The average iodine value at aorta region is obtained by 1) identifying a volume of interest around aorta; 2) computing the CT number (in HU) histogram for the volume of interest; 3) computing the average CT number of the histogram peak. Figure 5 shows a typical histogram of an aorta region. The green lines are the standard deviation of the iodine values around aorta.

III. Results

Figures 6 and 7 summarize the experimental results. They show the average (over 130 calcifications) percent increase of calcium volume due to blooming for different scan parameters:

$$avg.\%\ volume\ change = \frac{(calcium\ volume\ after\ recon - GT\ calcium\ volume)}{ground\ truth\ (GT)\ calcium\ volume}$$

The insets show corresponding images for one example calcification. Figure 6 shows the impact of tube voltage or standard recon kernel and a 1.2 mm focal spot size. Figure 7 shows the impact of reconstruction kernel (standard, detail, bone, edge) and focal spot size (1.2 mm and 0.6 mm) for a monochromatic 70 keV spectrum.

Figure 6 shows that higher kVp helps reduce the calcium blooming effect. For example, the average percent volume change is approximately 36% when using 140kVp for a cardiac CT scan and the calcium volume change increases to 68% when using 80 kVp.

Figure 7 shows that sharper reconstruction kernels (bone, edge) substantially reduce blooming artifacts, although at the expense of increase noise or aliasing artifacts. Reducing the focal spot size has minimal effect when combined with the lower-resolution standard kernel, and results in a small additional improvement when combined with the higher-resolution bone kernel.

While our results show that both kVp and reconstruction kernel have a significant impact on the degree of calcium blooming, it is important to note that in this study we did not adjust the threshold based on the inherent intensity change for calcium at the different kVps. As a result, it is possible that these results have exaggerated the impact of kVp on calcium blooming. In other words, the reduced apparent blooming one achieves by using a higher kVp might just as well be accomplished by using a wider HU window setting in reviewing the images. Further study is needed to assess this in more detail.

IV. Conclusion

A VCT framework is developed to evaluate calcium blooming as a function of CT scan and reconstruction parameters. An overview of the VCT framework is presented, as well as a study to explore the relative impact of X-ray tube voltage, reconstruction kernel, and focal spot size on calcium blooming severity. We conclude that:

- X-ray tube voltage/spectrum and reconstruction kernel have significant impact on calcium blooming
- Focal spot size is not the limiting factor and only has an impact when combined with a high-resolution reconstruction kernel.
- Calcification segmentation strongly depends on the threshold and could be greatly improved by tuning threshold settings as a function of scan and reconstruction parameters. This is a complex dependency and requires a more dedicated investigation.

In future work, the current study can be refined by using a more realistic range of calcification compositions and by jointly simulating patient and calcifications. Another next step is to work with clinical experts to evaluate the images under various parameter settings and the practical impact on stenosis assessment.

Acknowledgements

We would like to thank Dr Pontone and the Centro Cardiologico Monzino (University of Milan, Italy) for supplying cardiac CT image datasets used in this study.

Research reported in this publication was supported by the NIH/NHLBI grant R01HL151561. The content is solely the responsibility of the authors and does not necessarily represent the official views of the NIH.

References

1. Latina, J., Shabani, M., Kapoor, K., Whelton, S.P., Trost, J.C., Sesso, J., Demehri, S., Mahesh, M., Lima, J.A. and Arbab-Zadeh, A., 2021. Ultra-high-resolution coronary CT angiography for assessment of patients with severe coronary artery calcification: initial experience. *Radiology: Cardiothoracic Imaging*, *3*(4), p.e210053.
2. Fuchs A, Kühl JT, Chen MY, Viladés Medel D, Alomar X, Shanbhag SM, Helqvist S, Kofoed KF. Subtraction CT angiography improves evaluation of significant coronary artery disease in patients with severe calcifications or stents—the C-Sub 320 multicenter trial. Eur Radiol. 2018;28:4077–4085.
3. Steckmann, S., and M. Kachelries. "Blooming Artifact Reduction for Cardiac CT." IEEE Nuclear Science Symposium & Medical Imaging Conference, 2010. doi:10.1109/nssmic.2010.5874133.
4. Li, Ping, Lei Xu, Lin Yang, Rui Wang, Jiang Hsieh, Zhonghua Sun, Zhanming Fan, and Jonathon A. Leipsic. "Blooming Artifact Reduction in Coronary Artery Calcification by A New Deblooming Algorithm: Initial Study." Scientific Reports 8, no. 1 (2018). doi:10.1038/s41598-018-25352-5.
5. Matthias Renker, John W. Nance, Jr, U. Joseph Schoepf, Terrence X. O'Brien, Peter L. Zwerner, Mathias Meyer, J. Matthias Kerl, Ralf W. Bauer, Christian Fink, Thomas J. Vogl, and Thomas Henzler, "Evaluation of Heavily Calcified Vessels with Coronary CT Angiography: Comparison of Iterative and Filtered Back Projection Image Reconstruction" Radiology 2011 260:2, 390-399
6. Yan, Siming, Feng Shi, Yuhua Chen, Damini Dey, Sang-Eun Lee, Hyuk-Jae Chang, Debiao Li, and Yibin Xie. "Calcium Removal from Cardiac Ct Images Using Deep Convolutional Neural Network." 2018 IEEE 15th International Symposium on Biomedical Imaging (ISBI 2018), 2018. doi:10.1109/isbi.2018.8363617.
7. Samei, E., Abadi, E., Kapadia, A., Lo, J., Mazurowski, M. and Segars, P., 2020, March. Virtual imaging trials: an emerging experimental paradigm in imaging research and practice. In *Medical Imaging 2020: Physics of Medical Imaging* (Vol. 11312, p. 113121T). International Society for Optics and Photonics.
8. Abadi, E., Segars, W.P., Tsui, B.M., Kinahan, P.E., Bottenus, N., Frangi, A.F., Maidment, A., Lo, J. and Samei, E., 2020. Virtual clinical trials in medical imaging: a review. *Journal of Medical Imaging*, *7*(4), p.042805.
9. Hoffmann, Udo, Maros Ferencik, Ricardo C. Cury, and Antonio J. Pena. "Coronary CT Angiography." The Journal of Nuclear Medicine, March 13, 2006. Accessed April 23, 2019.
10. Kalisz, Kevin, Ji Buethe, Sachin S. Saboo, and Suhny Abbara. "Artifacts at Cardiac CT: Physics and Solutions." RadioGraphics36, no. 7 (October 21, 2016). Accessed April 23, 2019
11. Zhuangli Liang, Karl W. Clem, Do Synho, Brady Thomas, and Pien Homer. "Analysis and Mitigation of Calcium Artifacts in Cardiac Multidetector CT", IEEE ISBI, 2008.
12. Barrett, Julia F., and Nicholas Keat. "Artifacts in CT: Recognition and Avoidance." RadioGraphics24, no. 6 (2004): 1679-691. doi:10.1148/rg.246045065.
13. Fihn SD, Gardin JM, Abrams J, Berra K, Blankenship JC, Dallas AP, Douglas PS, Foody JM, Gerber TC, Hinderliter AL, King SB, Kligfield PD, Krumholz HM, Kwong RYK, Lim MJ, Linderbaum JA, Mack MJ, Munger MA, Prager RL, Sabik JF, Shaw LJ, Sikkema JD, Smith CR, Smith SC, Spertus JA, Williams SV, American College of Cardiology Foundation, American Heart Association Task Force on Practice Guidelines, American College of Physicians, American Association for Thoracic Surgery, Preventive Cardiovascular Nurses Association, Society for Cardiovascular Angiography and Interventions, Society of Thoracic Surgeons. 2012 ACCF/AHA/ACP/AATS/PCNA/SCAI/STS Guideline for the diagnosis and management of patients with stable ischemic heart disease: a report of the American College of Cardiology Foundation/American Heart Association Task Force on Practice Guidelines, and the American College of Physicians, American Association for Thoracic Surgery, Preventive Cardiovascular Nurses Association, Society for Cardiovascular Angiography and Interventions, and Society of Thoracic Surgeons. J Am Coll Cardiol. 2012;60:e44–e164.
14. Skinner JS, Smeeth L, Kendall JM, Adams PC, Timmis A. NICE guidance. Chest pain of recent onset: assessment and diagnosis of recent onset chest pain or discomfort of suspected cardiac origin. Heart. 2010;96:974–978.
15. B. De Man, S. Basu, N. Chandra, B. Dunham, P. Edic, M. Iatrou, S. McOlash, P. Sainath, C. Shaughnessy, and B. Tower, "CatSim: a new computer assisted tomography simulation environment," *Proc. SPIE 6510, Medical Imaging 2007: Physics of Medical Imaging,* vol. 6510, pp. 65102G, 2007.

X-ray phase contrast computed tomography with an amplitude modulated beam: moving towards clinical use

Grammatiki Lioliou, Lorenzo Massimi, Oriol Roche i Morgo, Tom Partridge, Paul Wolfson, Marco Endrizzi, Peter R. T. Munro, Silvia Cipiccia, Alberto Astolfo, Jinxing Jiang, Chris Thornton, Charlotte Maughan Jones, Michela Esposito, Amir Reza Zekavat, Jure Aleksejev, Ian Buchanan, Dana Shoukroun, Glafkos Havariyoun, Savvas Savvidis, Adam Doherty, Rimcy Palakkappilly Alikunju, Carlos Navarrete Leon, Carlo Peiffer, Sumera Rehman, Amy Ha, Muntzer Mughal, Marco Novelli, Tamara Suaris, J. Louise Jones, Laurence Lovat, Richard M. Waltham, Ian G. Haig, David Bate, Alessandro Olivo, Charlotte K. Hagen

***Abstract*—X-ray phase contrast (XPC) computed tomography (CT) gives access to an increased contrast for weakly attenuating samples such as soft biological tissue, opening new application pathways. We have developed an XPC-CT approach based on amplitude modulation of the beam, achieved by using a mask with alternating absorbing and transmitting septa placed immediately upstream of the scanned sample. Our approach can be implemented with two distinct sensing mechanisms: beam tracking and edge illumination, applicable to different imaging scenarios and experimental constraints. Crucially, both are compatible with "single-frame" retrieval retrieval approaches, which allows fast scans in which the sample is continuously rotated (flyscans).**

While our group has dedicated the best part of the last decade to developing the technique(s), more recently we have explored their potential clinical use. Here, we discuss two applications currently under investigation: the real-time intra-operative scanning of excised breast and oesophageal tissue. We also provide an outlook on how recent technological advances could further improve performance, especially in terms of meeting clinical constraints on scan time.

***Index Terms*—x-ray imaging, phase contrast, computed tomography**

I. Introduction

COMPUTED tomography (CT) is a versatile, non-invasive, imaging modality with applications ranging from biomedical research to clinical practice and material science [1]. It can provide 3D information of the internal structure of materials with a spatial resolution in a wide range of scales, from millimeters down to nanometers. Conventional x-ray CT, however, suffers from insufficient contrast when imaging materials with a low atomic number. Those materials benefit from an alternative imaging method, x-ray phase contrast (XPC) CT.

XPC-CT is based on phase changes that the x-rays undergo as they traverse the sample, while conventional CT is based on attenuation. Both effects are characterised by the complex refractive index,

$$n = 1 - \delta + i\beta \,, \qquad (1)$$

where β and δ describe attenuation and phase shifts, respectively. XPC-CT exploits the fact that δ can be up to three orders of magnitude larger than β, which can result in substantial increases in contrast, and therefore Contrast to Noise Ratio (CNR) at given statistics. Phase effects can manifest on the pixel- and sub-pixel scale; while the former give rise to x-ray refraction, the latter are responsible for the so-called dark-field contrast channel (ultra-small angle x-ray scattering, USAXS), which is related to microscopic features in the sample [2].

II. Amplitude Modulation XPC-CT

Although initially XPC-CT had stringent requirements on the x-ray beam's coherence, imposing the use of either synchrotron or microfocal sources, technological developments enabled its use with non-microfocal laboratory sources. Here, we discuss an amplitude modulation-based XPC-CT method by which the x-ray beam is structured into an array of narrow, physically separated beamlets created by a mask placed immediately upstream of the sample (see Fig. 1). In this configuration, attenuation, phase (refraction), and scattering effects manifest as an intensity reduction, a shift, and a broadening of the individual beamlets, respectively. In the following, we present two approaches to separately retrieve and quantify these three effects under different experimental conditions and constraints.

This work was supported by the Wellcome/EPSRC Centre for Interventional and Surgical Sciences (203145/Z/16/Z) and by EPSRC/UKRI (EP/T005408/1 and EP/P023231/1). Additional funding was obtained from the Wellcome Trust (200137/Z/15/Z). The authors wish to acknowledge the roles of the Breast Cancer Now Tissue Bank in collecting and making available the samples, and the patients who have generously donated their tissues. CKH was supported by the Royal Academy of Engineering, under the Research Fellowship scheme. PRTM was supported by the Royal Society under the University Research Fellowships scheme. AO was supported by the Royal Academy of Engineering under the Chairs in Emerging Technologies scheme.

GL, LM, ORiM, ME, PRTM, SC, AA, PW, JJ, CT, CMJ, TP, ME, ARZ, JA, IB, DS, GH, SS, AD, RPA, CNL, CP, SR, AH, AO, and CKH are with the Department of Medical Physics and Biomedical Engineering, University College London, Malet Place, London WC1E 6BT , UK.

MM is with the Department of Upper Gastro-Intestinal Surgery, UCLH, London NW1 2BU, UK.

MN is with the Research Department of Pathology, Cancer Institute, UCLH, London NW1 2BU, UK.

TS and JLJ are with the St Bartholomew's Hospital, Barts Health NHS Trust, West Smithfields, London EC1A 7BE, UK.

JLJ is also with the Barts and the London School of Medicine and Dentistry, Queen Mary University of London, Newark St, London E1 2AT, UK.

LL is with the Division of Surgery and Interventional Science, UCL, London, WC1E 6BT, UK.

RMW, IGH, and DB are with Nikon X-Tek Systems, Tring Business Centre, Icknield Way, Tring, Hertfordshire HP23 4JX, UK.

GL is the corresponding author (e-mail: g.lioliou@ucl.ac.uk).

A. Sensing Mechanism 1: Beam tracking

A relatively straightforward approach is to resolve the individual beamlets by using a detector with sufficiently high resolution, which allows observing their intensity reduction, shift and broadening directly [3]. A Gaussian fit of the beamlet profiles and subsequent extraction of their integrated area, mean, and standard deviation enables the quantitative separation of the three effects and the consequent retrieval of three contrast channels (attenuation, refraction, dark-field) [4]. Hence, beam-tracking enables multi-contrast imaging from a single exposure of the sample and is compatible with flyscans. A CT scan can be conducted by taking many such exposures as the sample rotates over at least 180 degrees, and reconstructing tomograms (axial slices) from the three separate sinograms corresponding to each of the contrast channels. The beam-tracking mechanism is efficient, in the sense that the entire radiation that passes through the sample reaches the detector (by contrast, this is not the case for XPC techniques relying on an analyzer to sense phase effects). On the downside, since a small pixel is needed and detectors typically have 2,000 x 2,000 pixels, the field of view (FOV) tends to be limited, which excludes applications with a certain (larger) sample size. Although a few exceptions exist, small pixel detectors also tend to have a limited detective quantum efficiency at high x-ray energies, which can increase dose and exposure time.

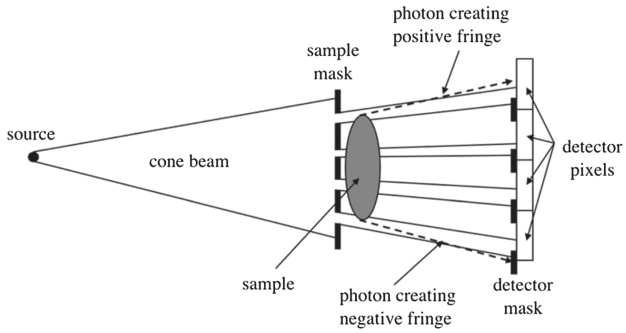


Figure 1. Schematic diagram depicting the implementation of the edge illumination method with laboratory sources and a 2D detector [5]. In the beam tracking case, the detector mask is removed and a detector with ~5-6 pixels per inter-beamlet distance is used.

B. Sensing Mechanism 2: Edge Illumination

The need for small pixels can be circumvented by implementing edge illumination (EI), which works with much larger pixels (~100 μm or more possible). Each pixel receives at most one beamlet, hence the beamlets are not resolved and it is no longer possible to observe their amplitude reduction, shift and broadening directly. Instead, EI requires that a second mask is used, placed immediately upstream of the detector (see Fig. 1), which acts as an analyzer. With that in place, the system's response to attenuation, refraction, and small-angle scattering is encoded via a measurement of the illumination curve (IC). The IC is obtained by step-scanning the sample mask laterally (vertically with respect to Fig. 1) through the setup and recording the intensity per pixel and per scanning step. It is typically well-approximated by a Gaussian, owing to the x-ray tube's Gaussian source profile. When a sample is placed into the setup, refraction causes the IC to shift, small-angle scattering causes it to broaden, and attenuation causes an amplitude reduction. To isolate these effects and thereby retrieve the three contrast channels, we need to measure the IC (with the sample in place) in at least three points (corresponding to three exposures), to obtain three independent equations which allow for the extraction of the three unknowns. To conduct a CT scan, those three exposures must be obtained at each rotational position of the sample. If the sample is known to be non- or weakly scattering, the retrieval problem reduces to two unknowns (attenuation and refraction), which can be solved from two IC measurements.

As the need for three (or two) exposures per angle implies long CT scan times, we have developed an additional, approximation-based retrieval method for which a single exposure per angle is sufficient. In brief, by assuming a constant scaling factor between attenuation and phase (more precisely, δ/β = constant) across the sample, the number of unknowns in the retrieval problem reduces to one, which can be recovered from a single measurement on the IC. This means that the pre-sample mask is kept at a fixed position during the scan (typically at the maximum slope of the IC, which corresponds to maximum phase sensitivity), and the sample can be continuously rotated thus maximizing the acquisition speed.

In practice, the retrieval is performed by applying a dedicated, sample-specific low-pass filter to each acquired frame, yielding the sample's projected thickness, which can be given as input to standard CT reconstruction methods [6]. Although the assumption of constant δ/β ratio appears to be highly restrictive, the "single-frame" retrieval has been found to perform extremely well on biological samples consisting of different soft tissues. In comparison to beam tracking, EI has an approximately 50% reduction in the efficiency with which the beam is used, owing to the second mask which prevents some radiation from reaching the detector. However, since it can be used with larger pixels it offers access to larger FOVs and generally greater flexibility in terms of practical use.

III. Towards Clinical Use

The ability to perform scans in reasonable acquisition times (a few minutes compared to the several hours required by previous approaches) has opened the way to previously inaccessible applications. A key one, which has enabled early experimentation on human patients, is intra-operative specimen imaging. While application areas are broad, this has so far been trialed in breast and oesophageal operations. Both cases require the imaging of relatively large (several cm) specimens, so the EI implementation was used; however, future applications requiring the imaging of smaller specimens would equally allow the use of beam tracking.

In breast operations, we targeted Wide Local Excisions (WLEs), where the issue is making sure that the entirety of the tumour has been removed, i.e. that the margins of the resected WLE are clear. This is a well-known problem in breast

conserving surgery, as tumour involvement at the margins is often only detected at post-operative histopathology, the results of which only become available several weeks after the operation, resulting in the patient having to be re-called for a second operation.

This study was structured in two parts: we first scanned approximately 120 specimens resulting from tissue surplus to the pathology evaluation, as this enabled us to assess the increased sensitivity and specificity of our method compared to standard practice (in the UK, specimen radiography) without disrupting the clinical workflow. This resulted in comparable specificity and a >2.5 fold increase in sensitivity [7]. In the second part of the study, we used the system in real time on entire WLEs as they were extracted from patients undergoing breast conservation surgery, to demonstrate that the system could be used in an operational situation without disrupting the clinical workflow. This showed that full-size (up to 8 cm) specimens could be scanned, with the "single-frame" retrieval approach, in 10-15' without compromising in image quality. One example from the "real time" procedure was reported in [7], the analysis of the additional cases, e.g. see Fig. 2, is underway and results will be presented at the conference.

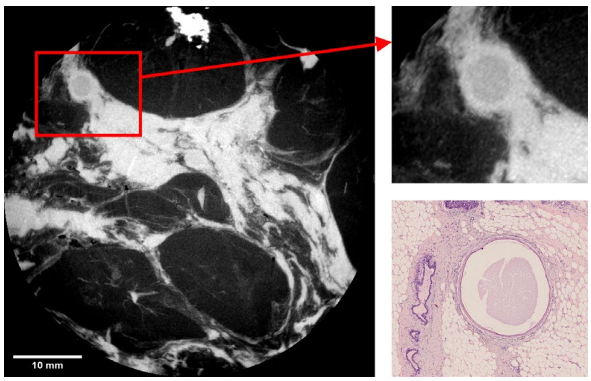


Figure 2. Maximum intensity projection of a XPC CT of a WLE. The zoomed-up region on the top right shows a cyst, with the matching histology shown immediately below.

An alternative use of the same machine was also investigated, based on the observation that the ultimate resolution in our images is determined by the size of the apertures in the pre-sample mask [8]. In practical terms this means that, through longer scans in which the sample is laterally displaced at every projection angle in steps equal to the aperture size over a mask period, a process termed as dithering, we were able to achieve a resolution of the order of 10 μm over the entire specimens. Analysis of the resulting higher-resolution images, combined with the enhanced contrast resulting from XPC, revealed a wealth of previously undetected features: alongside the detection of thinner tumour strands, which could to some extent be expected, it was also possible to detect the tissue response to chemotherapy [9]. This suggest an alternative use of the same machine as a "virtual histology" device, for example to be used during the down-time of the intra-operative use. While this currently requires significantly longer acquisitions, approaches to reduce these are discussed in the next section.

In parallel with the last stages of the breast project, we started a new project on the real-time imaging of oesophageal tissue, resulting from the resection of entire oesophagi during oesophagectomy operations. The aims were similar to those of the breast tissue, with the assessment of the number and infiltration status of the surrounding lymphnodes as an additional target alongside the detection of involved margins. This was a more challenging project compared to the breast one, because the oesophagus consists mostly of muscle tissue which is more absorbing than breast; moreover, tumours typically appear as a distortion of the natural tissue architecture, rather than a clearly identifiable mass.

This required some degree of sample preparation which was not needed in the breast case (namely immersion in ethanol), and a longer exposure time compared to the breast case. While these procedures still need to be optimised, we are reasonably confident of their compatibility with intra-operative use, since oesophagectomies last much longer than breast conserving surgery interventions, thus allowing longer timeframes for intra-operative scans. Also in this case, the early analysis of the obtained image quality reveals a wealth of unexpected details, which seems to suggest an ability to stage tumours directly from the x-ray images (and therefore in real time), which has the potential to change patient pathways (for example by allowing the implementation of additional in-room interventions) which goes beyond the anticipated detection of margin involvement and lymph node infiltration status. Also in this case, data analysis is currently underway, e.g. see Fig. 3, and the early results will be presented at the conference.

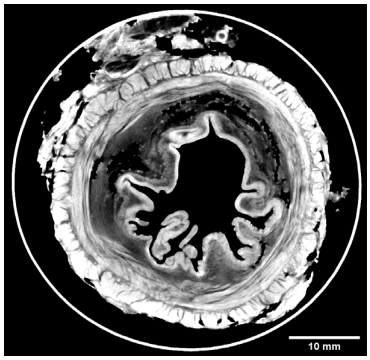


Figure 3. XPC-CT of an entire human oesophagus. While this is a tumour-free example, the exquisite detail with which all the different soft tissue layers are resolved is apparent.

All studies on human tissue were conducted in full compliance with the local ethics regulations and with informed consent from all patients involved.

IV. Recent Advances

The process that, above, has been referred to as dithering (i.e., overcoming the resolution limit otherwise imposed by the period of the sample mask by laterally displacing the sample in steps equal to the mask apertures) has significant advantages when it comes to detecting minute details like tumour margins, strands etc. However, as a substantial downside, dithering prolongs scans due to the multiple exposures involved and, crucially, it is not compatible with flyscans (the fact that a lateral scan of the sample is required for each rotation angle

means that dithered scans are bound to be step-and-shoot scans). Consequently, dithered scans are suitable for digital histology type applications that can run during down-times but may not be performed intra-operatively where a short scanning time is very much of the essence. To overcome this issue, we have been investigating solutions that allow performing EI (or beam tracking) XPC-CT scans with sub-period resolution as flyscans.

So far, we have developed an approach termed "cycloidal" CT [10]. This is an acquisition scheme by which the sample is translated laterally through the setup simultaneously with being rotated. Note the difference to dithering, which involves multiple exposures per angle; in cycloidal CT, only a single exposure is required. This makes it compatible with flyscans, given that it is implemented along with either beam-tracking, or EI with the "single-frame" retrieval method. The name cycloidal stems from the trajectory outlined by the sample as it undergoes the "roto-translation" motion. Its effect is a spreading out of the acquired data in the sinogram. Although this does not *solve* the under-sampling problem created by the amplitude modulating mask (the reason for which resolution is typically limited to its period), a more favorable basis for recovering the missing sinogram entries through a mathematical method (e.g., bivariate interpolation) is achieved. As a result, cycloidal CT can improve resolution compared to a scan in which the sample is only rotated. Early cycloidal CT experiments [10] have resulted in an in-slice resolution of ~30 μm. For comparison, the resolution of a much longer, dithered scan of the same sample was estimated at ~20 μm, and the corresponding rotation only scan provided ~100 μm. Development of the approach is ongoing, but the observation that resolution in an amplitude modulated XPC-CT scan can be increased by translating the sample while it rotates suggests that it could become a way of improving the detection performance of intra-operative scans without exceeding clinical constraints on scan time. Indeed, preliminary experimental results support this [11].

Despite this potential, a limitation of cycloidal CT is that the resolution improvement is confined to the in-slice plane, while the slice thickness remains to be defined by the pixel size. This is because it is implemented with a 1D amplitude modulation (achieved with a sample mask with slit-shaped apertures, Fig. 1) and lateral (that is, horizontal) translation of the sample. Consequently, the reconstructed images have a non-isotropic spatial resolution. To enable a resolution increase isotropically, we are investigating the extension of the cycloidal concept to a cycloidal-spiral approach. The 1D amplitude modulation is replaced by a 2D one (achieved with a sample mask that shapes the beam into an 2D array of pencil beams), and the acquisition involves translating the sample along *both* the vertical and horizontal direction, simultaneously with being rotated. Again, this results in one exposure per rotation angle. Development of this approach, including the interpretation of simulated and experimental data, is currently ongoing. In our talk, we will present results for both the cycloidal- and cycloidal spiral scanning schemes, and discuss their potential for clinical applications of XPCI-CT.

V. Conclusion

In summary, we have developed an amplitude modulation approach to XPC-CT, compatible with laboratory x-ray sources, that has shown significant potential for clinical use in the intra-operative imaging of excised breast or oesophageal tissue. Two possible sensing mechanisms exist (beam tracking, EI), and both can be implemented as flyscans thanks to phase retrieval methods applicable to a single exposure of the sample. While this implies that clinical constraints for scan times can be met, resolution was limited to the period of the sample mask. To increase this, stepping of the sample at each rotation angle was required, inevitably prolonging scans. Our investigation into cycloidal (and, more recently, cycloidal-spiral) acquisition schemes has led to options for increasing resolution that do not prolong scans. We expect those schemes to bear great potential for advancing the clinical utility of amplitude modulation XPC-CT.

References

[1] S. D. Rawson, J. Maksimcuka, P. J. Withers, and S. H. Cartmel, "X-ray computed tomography in life sciences", *BMC Biology*, vol. 18, no. 21, 2020.

[2] N. Matsunaga, K. Yano, M. Endrizzi, and A. Olivo, "Detection of individual sub-pixel features in edge-illumination x-ray phase contrast imaging by means of the dark-field channel", *J. Phys. D: Appl. Phys*, vol. 53, no. 9, art. no. 095401, 2020.

[3] F. A. Vittoria, G. K. N. Kallon, D. Basta, P. C. Diemoz, I. K. Robinson, A. Olivo, and M. Endrizzi, "Beam tracking approach for single–shot retrieval of absorption, refraction, and dark–field signals with laboratory x–ray sources", *Appl. Phys. Lett.*, vol. 106, art. no. 224102, 2015.

[4] F. A. Vittoria, M. Endrizzi, G. K. Kallon, C. K. Hagen, F. Iacoviello, P. De Coppi, and A. Olivo "Multimodal Phase-Based X-Ray Microtomography with Nonmicrofocal Laboratory Sources", *Phys. Rev. Applied*, vol. 8, art. no. 064009, 2017.

[5] A. Olivo, "Edge-illumination x-ray phase-contrast imaging", *J. Phys. Condens. Matter*, vol. 33, art. no. 363002, 2021.

[6] P. C. Diemoz, C. K. Hagen, M. Endrizzi, M. Minuti, R. Bellazzini, L. Urbani, P. De Coppi, and A. Olivo, "Single-Shot X-Ray Phase-Contrast Computed Tomography with Nonmicrofocal Laboratory Sources", Phys. Rev. Applied, vol. 7, art. no. 044029, 2017.

[7] L. Massimi, T. Suaris, C. K. Hagen, M. Endrizzi, P. R. T. Munro, G. Havariyoun, P. M. S. Hawker, B. Smit, A. Astolfo, O. J. Larkin, R. M. Waltham, Z. Shah, S. W. Duffy, R. L. Nelan, A. Peel, J. L. Jones, I. G. Haig, D. Bate, and A. Olivo, "Detection of involved margins in breast specimens with X-ray phase-contrast computed tomography", *Sci. Rep.*, vol. 11, art. no. 3663, 2021.

[8] P. C. Diemoz, F. A. Vittoria, and A. Olivo, "Spatial resolution of edge illumination X-ray phase-contrast imaging", *Opt. Express*, vol. 22, pp. 15514-15529, 2014.

[9] L. Massimi, T. Suaris, C. K. Hagen, M. Endrizzi, P. R. T. Munro, G. Havariyoun, P.M. S. Hawker, B. CSmit, A. Astolfo, O. J. Larkin, R. M. Waltham, Z. Shah, S. W. Duffy, R. L. Nelan, A. Peel, J. L. Jones, I. G. Haig, D. Bate, A. Olivo, "Volumetric high-resolution X-ray phase-contrast virtual histology of breast specimens with a compact laboratory system", *IEEE Trans. Med. Imaging*, in press.

[10] C. K. Hagen, F. A. Vittoria, O. Roche i Morgó, M. Endrizzi, and A. Olivo, "Cycloidal Computed Tomography", *Phys. Rev. Appl.*, vol. 14, art. no. 014069, 2020.

[11] O. Roche i Morgó, L. Massimi, T. Suaris, M. Endrizzi, P. R. T. Munro, S. Savvidis, G. Havariyoun, P. M. S. Hawker, A. Astolfo, O. J. Larkin, R. L. Nelan, J. L. Jones, D. M. Pelt, D. Bate, A. Olivo, and C. K. Hagen, "Exploring the potential of cycloidal computed tomography for advancing intraoperative specimen imaging", *Proc. SPIE*, vol. 11840, Developments in X-Ray Tomography XIII, art. no. 118400R, 2021.

Author List

K

L

M