NDT net e-Journal of Nondestructive Testing - ISSN 1435-4934 - www.ndt.net

14th Conference on Industrial Computed Tomography, Antwerp, Belgium (iCT 2025), www.ict2025.org

Dual Domain Swin Transformer-based Reconstruction method for Sparse-View Computed Tomography

J. Van der Rauwelaert, C. Bossuyt, J. Sijbers

University of Antwerp, Universiteitsplein 1, 2610 Antwerp, Belgium

E-mail: Jonas.VanderRauwelaert@uantwerpen.be

Abstract

A new sparse-view parallel beam computed tomography reconstruction method is proposed that exploits the restoration capabilities of Transformer networks, in particular the Swin Transformer-based image reconstruction network SwinIR. Our method comprises three key blocks: sinogram upsampling via linear interpolation, initial reconstruction using deep learning in both domains, and residual refinement. Two architectures are tested: a long one using neural networks in both domains of the residual refinement block and a short one using a network exclusively in the sinogram domain. Each method is tested with SwinIR and U-Net, resulting in four variants, all of which outperform traditional methods like FBP and SIRT in terms of PSNR and SSIM. The short architecture using SwinIR achieves the best results, with a training and computation time smaller than the SwinIR-based long architecture but larger than both U-Net-based variants.

Keywords: Computed Tomography, Sparse-View, Deep Learning, Swin Transformer, Dual Domain

1 Introduction

Computed Tomography (CT) is a noninvasive imaging method that utilizes X-ray projections to create cross-sectional images of a patient. Because X-rays are ionizing and can damage human cells, it is essential to minimize the radiation dose received by the patient. Therefore, sparse-view X-ray CT is a highly active research area in both clinical and industrial applications as it allows to reduce radiation dose and/or acquisition time by lowering the number of projection angles to less than 100. However, the limited projection data often leads to artifacts in traditional reconstruction methods like Filtered Back Projection (FBP). While iterative reconstruction methods, such as the Simultaneous Iterative Reconstruction Technique (SIRT), offer improved reconstruction quality compared to FBP, they are computationally demanding, leading to prolonged processing times that reduce their practicality.

To address this issue, deep learning (DL) can be implemented using an Artificial Neural Network (ANN) within the sinogram domain to estimate missing projection data [1] or in the image domain to suppress artifacts in the reconstructed images [2]. Dual-domain reconstruction methods consider both the sinogram and image domains simultaneously, with Convolutional Neural Networks (CNNs) in this dual-domain framework demonstrating improved performance compared to single-domain approaches [3]. However, due to the localized nature of convolution operations, CNNs face challenges in capturing long-range dependencies. To address this limitation, Transformer-based methods, which are better suited for modelling such dependencies, have been developed. Transformers [4], widely used in natural language processing, are DL models that utilize self-attention to identify relationships among various input components, enabling the modelling of long-range dependencies. The Vision Transformer (ViT) [5], an adaptation designed for image data, extends this capability to vision tasks, including applications like image restoration [6]. However, the self-attention is computed globally across the entire image, leading to a significant increase in computational complexity as the image size grows. The Shifted Window (Swin) Transformer [7] reduces this computational cost by restricting self-attention calculations to local windows and shifting these windows to enable cross-window self-attention. SwinIR [8], a promising Swin Transformer-based network designed for image restoration, demonstrates superior restorations compared to competitive CNNs, even on a small training dataset of 800 images.

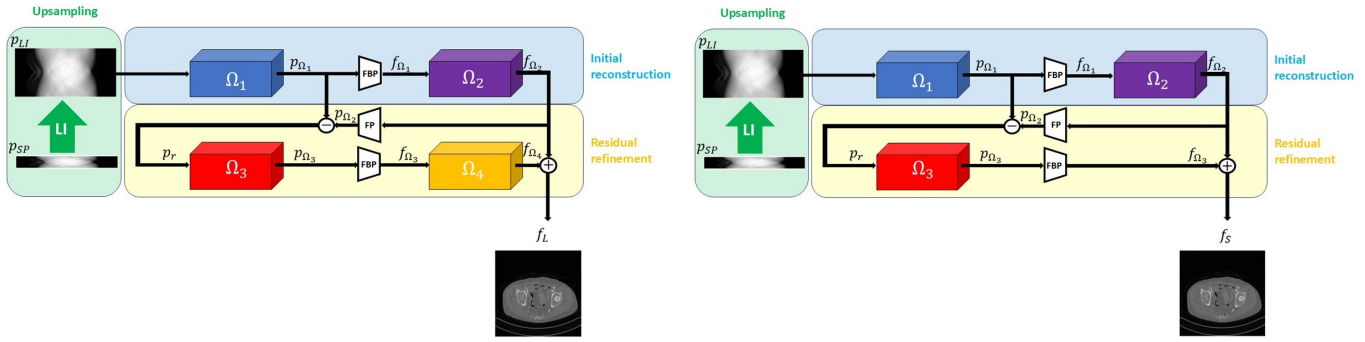
Using DL in the image domain can create inconsistencies with the sinogram domain, potentially resulting in reconstructions that do not accurately represent the corresponding sinogram data. To address this, one approach is to include the sparse data as input to the ANNs [9, 10], but at high levels of sparsity the information contained in these sparse sinograms becomes significantly limited. In [11] an architecture is proposed that employs an edge enhancement CNN and a U-Net [12], a widely used type of CNN, in an initial recovery block, along with an additional U-Nets in each domain of a data consistency block, that aims to reduce the discrepancies between the sinogram and image domains. Additionally, a Swin Transformer-based network is then used for further enhancement, resulting in a total of five ANNs that need to be trained. Furthermore, the inclusion of an ANN in the image domain of the data consistency block could potentially reintroduce inconsistencies between the two domains, countering the goal of this block.

By incorporating the SwinIR image restoration networks in an initial reconstruction block and residual refinement block, the proposed method aims to fully exploit the capabilities of SwinIR for reconstructing images from sparse sinograms ($\sim$ 90 projections) and ultra-sparse sinograms ($\sim$ 30 projections). To the authors' knowledge, this work is the first to implement SwinIR in this architecture for few-view tomographic image reconstruction and to quantify the effect of the residual refinement block in the sinogram domain.

https://doi.org/10.58286/30751

2 Methods

Our proposed reconstruction method is composed of three blocks, shown in Fig. 1: sinogram upsampling (green), initial reconstruction (blue), and residual refinement (yellow). While an ANN could perform sinogram upsampling, it would require retraining for every level of sparsity. Therefore, linear interpolation (LI) was applied to the sparse sinogram p_{SP} resulting in a sinogram p_{LI} with the same dimensions as the full-view sinogram p. The artifacts induced by this interpolation can then be minimalized during the initial reconstruction without the need of retraining, enhancing the method's flexibility. To achieve a more accurate estimation of the full-view sinogram p, the interpolated sinogram p_{LI} is refined using an ANN, denoted as Ω_1, resulting in p_{Ω_1}. The objective is to estimate the ground truth reconstruction f, thus this improved sinogram p_{Ω_1} is then used to obtain the reconstruction f_{Ω_1} through FBP, which is subsequently further enhanced by an ANN, Ω_2, to produce f_{Ω_2}. Interpolation in the sinogram domain introduces errors as the original data cannot be precisely reconstructed. Additionally, Ω_2 may apply corrections in the image domain that do not align with the original full-view sinogram data. To address this issue, a residual refinement step is introduced and evaluated using two different architectures, exploring the impact of incorporating an ANN in the image domain of this block.



(a) Long architecture

(b) Short architecture

Figure 1: Proposed method with (a) the long architecture used in DDSwinL and DDNetL, (b) the short architecture used in DDSwinS, and DDNetS. The ANNs (SwinIR or U-Net) are denoted as Ω_1, Ω_2, Ω_3, and Ω_4.

Long architecture. Fig. 1a shows the long architecture of the proposed method, which incorporates an ANN in both the sinogram domain (Ω_3) as the image domain (Ω_4) of the residual refinement block. In this setup, DDSwinL and DDNetL use SwinIR and U-Net as ANNs, respectively. A forward projection (FP) is applied to reconstruction f_{Ω_2} to generate the corresponding sinogram data p_{Ω_2}, and the residue $p_r = p_{\Omega_1} - p_{\Omega_2}$ is computed. The ANN, Ω_3, then estimates the residue with the full view sinogram $p_r^t = p - p_{\Omega_2}$, denoted as p_{Ω_3}. Subsequently, FBP is applied to the estimated residue in the sinogram domain, p_{Ω_3}, producing the reconstructed residue f_{Ω_3}, which is then used by the ANN Ω_4 to approximate the target residue $f_r^t = f - f_{\Omega_2}$, denoted as f_{Ω_4}. The final output is then given by $f_L = f_{\Omega_4} + f_{\Omega_2}$

Short architecture. Fig. 1b displays the short architecture of the proposed method, featuring an ANN only in the sinogram domain of the residual refinement block, where DDSwinS and DDNetS use SwinIR and U-Net, respectively. Analogous to the long architecture, the residue f_{Ω_2} is computed, to obtain the final reconstruction $f_S = f_{\Omega_3} + f_{\Omega_2}$.

3 Experimental results

3.1 Experimental setup

Dataset. The DL architectures DDNetS, DDNetL, DDSwinS and DDSwinL were trained and tested on the "*2016 NIH-AAPM dataset NIH-AAPM-Mayo Clinic Low Dose CT Grand Challenge*" [13] dataset, which contains CT images (resolution 512×512) of the torso from various patients. For the training dataset, 800 slices were selected from seven patients and the method was tested on 200 slices from three different patients.

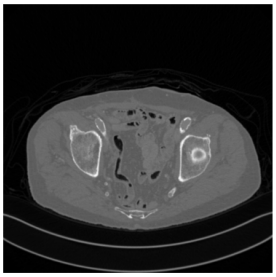
Implementation details and training settings. Guided by the sinogram restoration quality and computation time of network Ω_1 in the sinogram domain, the hyperparameters for all SwinIR networks were set as follows: window size to 8, patch size to 1, number of Residual Swin Transformer Blocks (RSTBs) to 2, number of Swin Transformer Layers (STLs) to 4, number of attention heads to 4, and embedding dimension to 60. The learning rate was initialized at 0.0001, with the Adam optimizer [14] used for training. A batch size of 1 was employed, and the model was trained over 10 epochs.

Geometric setup. The forward and backward projections were performed with the ASTRA-Toolbox [15]. A parallel beam geometry was used, with the rotation point positioned at the centre of the reconstruction grid, equidistant from the source and detector composed of 800 detector pixels. The full-view sinograms, simulated with 400 projections evenly distributed over angles ranging from 0° to 180°, were treated as the ground truth sinograms. Different levels of sparsity were obtained by simulating sinograms composed of 20, 30, 40, 80, 90 or 100 projections.

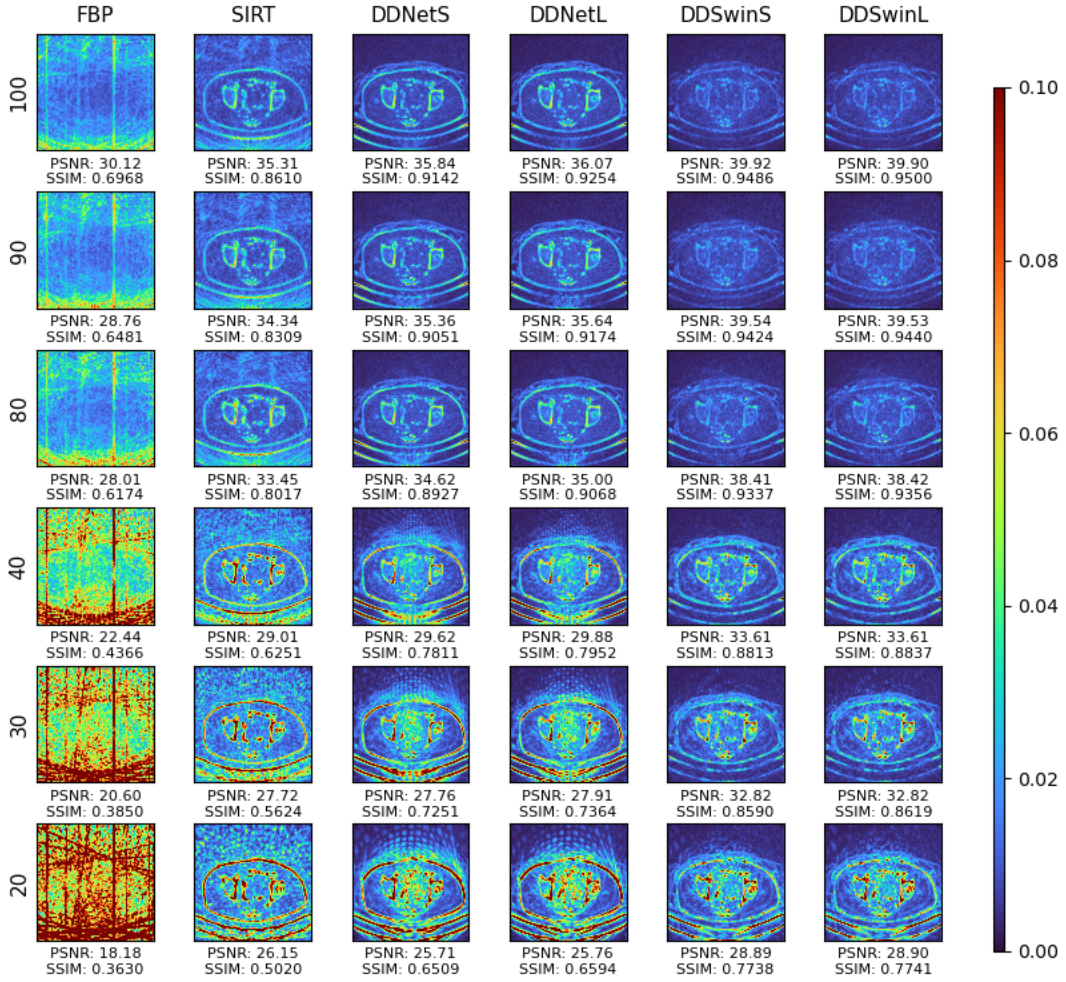
Evaluation metrics. The quality of the reconstructions were evaluated by computing the Structural Similarity Index Metric (SSIM) and the Peak Signal to Noise Ratio (PSNR) of the obtained sinograms and reconstructions.

3.2 Image domain

Fig. 2 shows the absolute differences between the ground truth and the reconstructions using (from left to right) FBP, SIRT, and the four variants of our proposed method: DDNetS, DDNetL, DDSwinS and DDSwinL. The rows indicate the number of input projections. The DL methods evaluated with 100, 90, and 80 projections as input, were trained on sinograms containing 90 projections, while those evaluated on 40, 30, and 20 projections were trained on sinograms consisting of 30 projections. Considering the computational cost and the quality of the reconstruction, the number of iterations for SIRT was set to 500. Results show that the Swin Transformer-based methods achieve the highest PSNR and SSIM, with DDSwinS and DDSwinL producing comparable results to one another with differences of less than 0.03 dB in PSNR and less than 0.003 in SSIM.



(a) Ground truth



(b) Absolute difference images

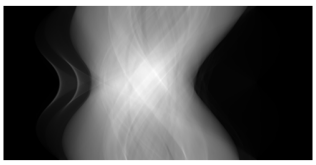
Figure 2: (a) Ground truth reconstruction, (b) Absolute difference between the ground truth and reconstructions obtained with different methods (columns) for different numbers of input projections (rows).

3.3 Sinogram domain

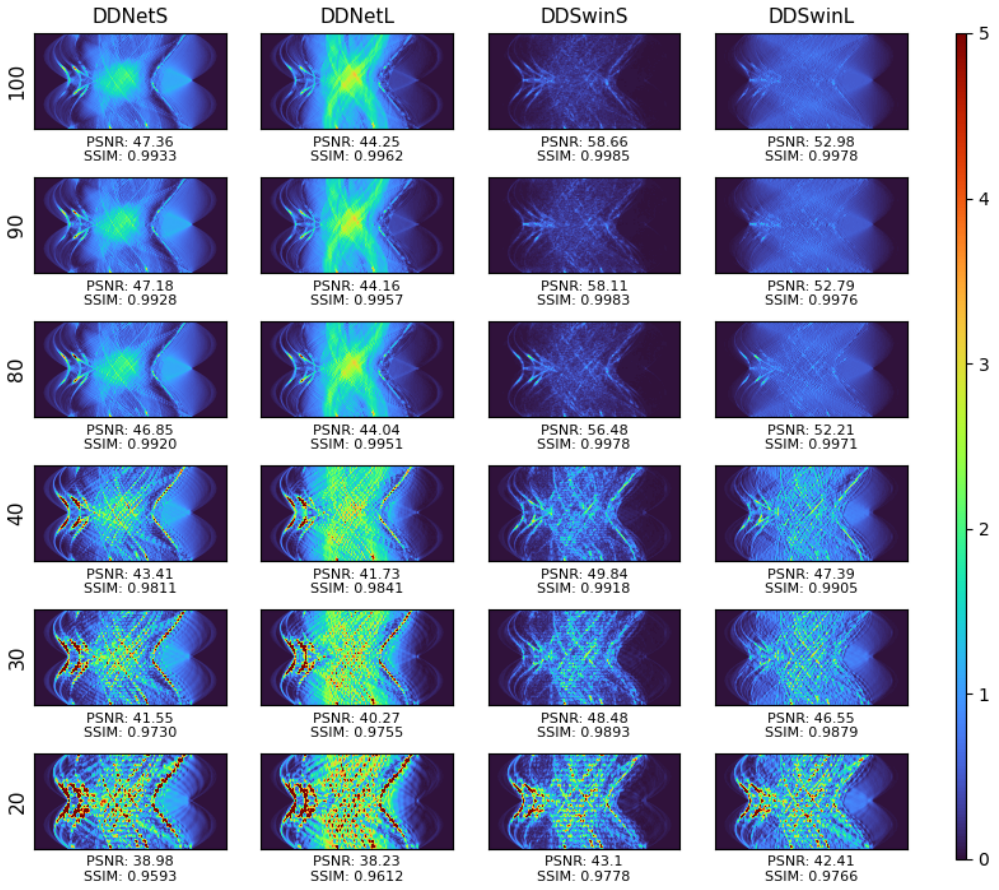
The corresponding sinograms of the reconstructions are simulated and Fig. 3 shows the absolute difference with the simulated ground truth sinogram. The computation time averaged over 50 simulated sinograms with 90 projections on an Intel(R) Core(TM) i7-7820HQ CPU @ 2.90GHz processor, along with the total number of trainable parameters, is presented in Table 1. Combined with Fig. 3 this demonstrates that the short architecture (Fig. 1b) more effectively restores the sinogram and leads to a smaller computation time than the long architecture (Fig. 1a). Additionally, DDSwinS exhibits a 10.93dB improvement in PSNR over DDNetS, at the cost of a computation time that is 13 times larger. When compared to SIRT (500 iterations), DDSwinS achieves a computation time that is 6 times smaller, while still being 144 times larger than that of FBP.

Method	Computation time (s)	Parameters (million)
FBP	0.21	/
SIRT	200.28	/
DDNetS	2.39	1.4
DDNetL	2.54	1.9
DDSwinS	30.29	1.0
DDSwinL	38.82	1.4

Table 1: Mean computation time and the total number of trainable parameters for different methods



(a) Ground truth



(b) Absolute difference images

Figure 3: (a) Ground truth sinogram, (b) Absolute difference between the ground truth sinogram and the corresponding simulated sinograms of the deep learning reconstructions in Fig. 2 obtained with different methods (columns) for different numbers of input projections (rows).

4 Discussion

As shown in Fig. 2 and Fig. 3, despite being trained on only 800 images, our proposed DL methods outperform traditional methods like FBP and SIRT, even when the number of input projections differs by 10 from the number of projections the methods were trained on, highlighting the flexibility of having a sinogram interpolation module. While the Swin Transformer-based methods achieve higher PSNR and SSIM and require fewer trainable parameters than their CNN-based counterparts, they have a longer computation time due to their computationally intensive self-attention mechanism. Nonetheless, the computation time remains shorter than that of SIRT. For the methods built with SwinIR, DDSwinS and DDSwinL, adding the ANN Ω_4 to the architecture has a limited effect on the PSNR and SSIM in the image domain, but it significantly impacts the training process and computation time. Additionally, the inclusion of Ω_4 in the architecture leads to a decrease in PSNR and SSIM in the sinogram domain, resulting in greater discrepancies with the projection data. However, the setup of the SwinIR networks is based on the performance of Ω_1, so further optimization of the settings for the SwinIR networks Ω_2, Ω_3, and Ω_4 could lead to improved results.

5 Conclusion

A dual-domain CT reconstruction method is proposed that employs SwinIR to reduce undersampling artefacts in few-view X-ray CT. By accounting for long-range dependencies, SwinIR demonstrates improved performance over U-Net. Experiments indicate that incorporating an ANN only into the sinogram domain of the residual refinement block rather than in both domains, reduces the absolute error between the simulated full-view sinogram and the simulated sinogram of the resulting reconstruction. Considering these factors, along with the reduced training and computation time of the shorter architecture, it can be concluded that DDSwinS demonstrates the best overall performance.

Acknowledgements
C. Bossuyt and J. Sijbers acknowledge support from the European Union's Horizon 2020 research and innovation programme under grant agreement No 101020100.

References

[1] X. Dong, S. Vekhande and G. Cao. "Sinogram interpolation for sparse-view micro-CT with deep learning neural network". In: *Medical Imaging 2019: Physics of Medical Imaging*. Ed. by Taly Gilat Schmidt, Guang-Hong Chen and Hilde Bosmans. Vol. 10948. International Society for Optics and Photonics. SPIE, 2019, 109482O. DOI: 10.1117/12.2512979. URL: https://doi.org/10.1117/12.2512979.

[2] Z. Zhang et al. "A Sparse-View CT Reconstruction Method Based on Combination of DenseNet and Deconvolution". In: *IEEE Transactions on Medical Imaging* 37.6 (2018), pp. 1407–1417. DOI: 10.1109/TMI.2018.2823338.

[3] D. Lee, S. Choi and H. Kim. "High quality imaging from sparsely sampled computed tomography data with deep learning and wavelet transform in various domains". In: *Medical Physics* 46.1 (Nov. 2018), pp. 104–115. DOI: 10.1002/mp.13258. URL: https://doi.org/10.1002/mp.13258.

[4] A. Vaswani et al. "Attention is all you need". In: *Proceedings of the 31st International Conference on Neural Information Processing Systems*. NIPS'17. Long Beach, California, USA: Curran Associates Inc., 2017, pp. 6000–6010. ISBN: 9781510860964.

[5] A. Dosovitskiy et al. "An Image is Worth 16x16 Words: Transformers for Image Recognition at Scale". In: *9th International Conference on Learning Representations, ICLR 2021, Virtual Event, Austria, May 3-7, 2021*. OpenReview.net, 2021. URL: https://openreview.net/forum?id=YicbFdNTTy.

[6] A. M. Ali et al. "Vision Transformers in Image Restoration: A Survey". In: *Sensors* 23.5 (2023). ISSN: 1424-8220. DOI: 10.3390/s23052385. URL: https://www.mdpi.com/1424-8220/23/5/2385.

[7] Z. Liu et al. "Swin Transformer: Hierarchical Vision Transformer using Shifted Windows". In: *2021 IEEE/CVF International Conference on Computer Vision (ICCV)*. 2021, pp. 9992–10002. DOI: 10.1109/ICCV48922.2021.00986.

[8] J. Liang et al. "SwinIR: Image Restoration Using Swin Transformer". In: *2021 IEEE/CVF International Conference on Computer Vision Workshops (ICCVW)*. 2021, pp. 1833–1844. DOI: 10.1109/ICCVW54120.2021.00210.

[9] C. Wang et al. "DuDoTrans: Dual-Domain Transformer for Sparse-View CT Reconstruction". In: *Machine Learning for Medical Image Reconstruction*. Ed. by Nandinee Haq et al. Cham: Springer International Publishing, 2022, pp. 84–94. ISBN: 978-3-031-17247-2.

[10] H. Yuan, J. Jia and Z. Zhu. "SIPID: A deep learning framework for sinogram interpolation and image denoising in low-dose CT reconstruction". In: *2018 IEEE 15th International Symposium on Biomedical Imaging (ISBI 2018)*. 2018, pp. 1521–1524. DOI: 10.1109/ISBI.2018.8363862.

[11] J. Pan et al. "Multi-domain integrative Swin transformer network for sparse-view tomographic reconstruction". In: *Patterns* 3.6 (2022), p. 100498. ISSN: 2666-3899. DOI: https://doi.org/10.1016/j.patter.2022.100498. URL: https://www.sciencedirect.com/science/article/pii/S2666389922000836.

[12] 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*. Ed. by Nassir Navab et al. Cham: Springer International Publishing, 2015, pp. 234–241. ISBN: 978-3-319-24574-4.

[13] C. McCollough. *2016 Low-Dose CT Grand Challenge*. 2022. DOI: 10.21227/4yqw-2364.

[14] 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*. Ed. by Yoshua Bengio and Yann LeCun. 2015. URL: http://arxiv.org/abs/1412.6980.

[15] W. van Aarle et al. "Fast and flexible X-ray tomography using the ASTRA toolbox". In: *Opt. Express* 24.22 (Oct. 2016), pp. 25129–25147. DOI: 10.1364/OE.24.025129. URL: https://opg.optica.org/oe/abstract.cfm?URI=oe-24-22-25129.