The Deep Steerable Convolutional Framelet Network for Suppressing Directional Artifacts in X-ray Tomosynthesis

Luis F. Alves Pereira
Dept. de Ciência da Computação
Universidade Federal do Agreste de Pernambuco
Garanhuns, Brazil
luis-filipe.pereira@ufape.edu.br

Jan De Beenhouwer
imec-Vision Lab
University of Antwerp
Antwerp, Belgium
jan.debeenhouwer@uantwerpen.be

Jan Sijbers
imec-Vision Lab
University of Antwerp
Antwerp, Belgium
jan.sijbers@uantwerpen.be

***Abstract*—Tomographic artifacts are unwanted distortions and/or structures not present in the scanned body that may appear in the reconstructed images. Recent deep learning-based methods for suppressing artifacts in tomographic images are currently not informed by the nature of these artifacts in the design of their network architectures. In this work, we present the Deep Steerable Convolutional Framelet Network (DSCFN), inspired by the theory of deep convolutional framelets, which exploits the regular pattern of ripple artifacts presented in sparse-view X-ray tomosynthesis images. Experiments with simulated data show that the DSCFN outperforms the regular U-net and its deep convolutional framelet, the tight U-net, in terms of PSNR.**

***Index Terms*—Medical Imaging, Digital Tomosynthesis Imaging, Deep Learning, Deep Convolutional Framelet Networks**

I. Introduction

Both Computed Tomography (CT) and X-ray tomosynthesis allow reconstructing cross sections of a patient's body. In contrast to CT, X-ray Tomosynthesis acquires only a limited number of X-ray projections. As a result, while tomosynthesis reconstructions display a high-resolution parallel to the X-ray detector, the in-depth resolution is much lower compared to that of CT. On the positive side, due to the reduced number of X-ray projections, the radiation dose can be $30\times$ lower than CT [1], and the total examination time much shorter [2]. Moreover, the tomosynthesis average cost per patient is about $2.5\times$ lower compared to CT [3].

To ensure a high quality of tomosynthesis reconstruction images, there is a minimum number of X-ray projections to be acquired for a given angular range [4]. Nevertheless, there is a strive towards reducing the number of projections as much as possible without significantly compromising image quality. Indeed, reducing the number of acquired projections lowers the scanning time as well as the radiation dose. Unfortunately, with decreasing the number of projections, unwanted ripple artifacts start to appear. The method proposed in this work aims at suppressing those ripple artifacts in images obtained from sparse-view X-ray tomosynthesis by post-processing the tomographic data in the reconstruction space. In contrast to recently proposed deep learning solutions, our approach is *artifact-informed* as it leverages the directional nature of the target ripple artifacts to compose the network's architecture.

Many Deep Learning (DL) methods were proposed in the literature for suppressing tomographic artifacts [5]–[9]. However, there is a gap of understanding on the reasoning behind most of those methods. At the same time, the recent theory on Deep Convolutional Framelet Networks (DCFNs) [10] describes a DL architecture that can be viewed as an implementation of convolutional framelets [11] for signal representation using fixed non-local bases combined with data-driven local bases. Therefore, DCFNs can be viewed as an extension of the traditional signal processing theory.

This work proposes a DCFN that incorporates the steerable wavelet as its fixed non-local basis: the Deep Steerable Convolutional Framelet Network (DSCFN). On a steerable basis, the data is decomposed into scale and orientation subbands. Steerable wavelets are more flexible than orthogonal separable wavelets (e.g., Haar, Daubechies, etc) since no orthogonality constraints are applied to the filters. Furthermore, the filters are only constrained to be rotated copies of each other and linear combinations of the basis filters [12], and can hence be selected according to the image characteristics observed in the dataset.

In the sparse-view tomosynthesis problem, the artifacts that emerge in the reconstructed images are mainly ripple effects that are oriented along multiple directions. To reduce the ripple artifacts, we propose the use of DSCFN with band-pass oriented filters, which decompose the image details into four directional components. Furthermore, our DSCFN ensures that all steerable wavelet components from the input image are denoised by independent convolutional layers.

II. Background

A. The Deep Convolutional Framelet Network

Ye *et al.* [13] described the DCFNS - a new class of deep neural networks - that can be viewed as an implementation of Convolutional Framelets [14] for representing signals by convolving with local and non-local basis functions. The architecture of a regular DCFN based on two levels of

decomposition is illustrated in Fig. 1 where $\Phi_{high}^{(l)}$ and $\Phi_{low}^{(l)}$ refer to high-band and low-band non-local bases, respectively, and $\Psi^{(l)}$ and $\tilde{\Psi}^{(l)}$ are learnable local bases.

Mathematically, Ye *et al.* summarized the dataflow in the DCFN using the Eqs. (1) and (2). Eqs. (1) refers to the encoder stage of the DCFN, where $f \in \mathbb{R}^m$ is the input image, $\overline{\Psi}^{(l)}$ is the l-th matrix of flipped convolutional filters $[\overline{\psi_1} \cdots \overline{\psi_q}] \in \mathbb{R}^{d\times q}$ composed of q convolutional filters of length d, $\circledast$ is the convolutional operator that states that $f \circledast \overline{\Psi} = \mathbb{H}_d(f)\Psi$ where $\mathbb{H}_d$ is the Hankel operator such that $\mathbb{H}_d(f) \in \mathbb{R}^{m\times d}$, $\Phi_{low/high}^{(l)} = [\phi_1 \cdots \phi_m] \in \mathbb{R}^{n\times m}$ is the l-th low-band or high-band non-local basis matrix for $\mathbb{R}^n$, ρ is the ReLU activation function, $C_{low}^{(l)}$ is the l-th approximate signal, and $C_{high}^{(l)}$ is the l-th detail signal.

$$\begin{cases} C_{low}^{(1)} = \rho\left(\Phi_{low}^{(1)T}\left(f \circledast \overline{\Psi}^{(1)}\right)\right) \\ C_{low}^{(2)} = \rho\left(\Phi_{low}^{(2)T}\left(C_{low}^{(1)} \circledast \overline{\Psi}^{(2)}\right)\right) \\ C_{high}^{(1)} = \Phi_{high}^{(1)T}\left(f \circledast \overline{\Psi}^{(1)}\right) \\ C_{high}^{(2)} = \Phi_{high}^{(2)T}\left(C_{low}^{(1)} \circledast \overline{\Psi}^{(2)}\right) \end{cases} \tag{1}$$

Additionally, the decoder stage of the DCFN is given by:

$$\begin{cases} \hat{C}_{low}^{(2)} = C_{low}^{(2)} \\ \hat{C}_{low}^{(1)} = \rho\left(\left(\Phi^{(2)}\hat{C}^{(2)}\right) \circledast v(\tilde{\Psi}^{(2)T})\right) \\ \hat{f} = \rho\left(\left(\Phi^{(1)}\hat{C}^{(1)}\right) \circledast v(\tilde{\Psi}^{(1)T})\right) \end{cases} \tag{2}$$

where $\hat{C}_{low}^{(l)}$ is the l-th approximate signal estimation, $v(\tilde{\Psi})$ is given by

$$v(\tilde{\Psi}) := \frac{1}{d}\begin{bmatrix} \tilde{\psi}_1 \\ \vdots \\ \tilde{\psi}_q \end{bmatrix} \in \mathbb{R}^{dq} \tag{3}$$

and $\hat{C}^{(l)}$ is the l-th signal estimation presented in the following equation

$$\Phi^{(l)}\hat{C}^{(l)}\tilde{\Psi}^{(l)T} = \Phi_{low}^{(l)}\hat{C}_{low}^{(l)}\tilde{\Psi}^{(l)T} + \Phi_{high}^{(l)}\hat{C}_{high}^{(l)}\tilde{\Psi}^{(l)T} \tag{4}$$

where $\hat{C}_{high}^{(l)}$ is the l-th detail signal estimation, $\tilde{\Psi}^{(l)} = [\tilde{\psi}_1 \cdots \tilde{\psi}_q] \in \mathbb{R}^{d\times q}$ is the matrix of convolutional filters dual to Ψ such that $\Psi\tilde{\Psi}^T = I_{d\times d}$.

B. The Steerable Wavelet

Steerable Wavelets allow an image to be decomposed into scale and orientation subbands [12]. A set of filters $\{\upsilon_1, \upsilon_2, \cdots, \upsilon_z\}$ composes a steerable basis $\boldsymbol{\Upsilon}$ if the filters are *(i)* rotated copies of each other, and *(ii)* computed as a linear combination of basis filters.

The idealized spectral decomposition promoted by the steerable wavelet is shown in Fig. 2 for $K = 4$ orientations, where $\{\upsilon_{B_0}, \upsilon_{B_1}, \upsilon_{B_2}, \upsilon_{B_3}\}$ are band-pass oriented filters and, υ_{H_0} and υ_{L_1} are non-oriented high-pass and narrowband low-pass filters, respectively. Higher levels of decomposition are obtained recursively by applying the complete steerable decomposition to the downsampled image resulting from L_1.

Since the Fourier spectrum is translation-invariant, and the directional derivatives $\{\upsilon_{B_0}, \cdots, \upsilon_{B_K}\}$ span a rotation equivariant subspace, the final representation obtained is translation-invariant and rotation-equivariant. Furthermore, the Steerable Wavelet is invertible and it forms a tight frame, *i.e.* there is a dual basis $\tilde{\boldsymbol{\Upsilon}}$ that satisfies $\tilde{\boldsymbol{\Upsilon}}\boldsymbol{\Upsilon}^T = I$.

The advantages of the Steerable Wavelet over other separable orthogonal wavelets (such as Haar, and Daubechies) are at least threefold: *(i)* there are no orthogonal constraints applied to the filters; *(ii)* any number K of directional derivates can be used; *(iii)* new steerable filters with specific characteristics can be designed [15], [16].

III. Proposed Method

The Deep Steerable Convolutional Framelet Network (DSCFN) proposed in this work relies on applying the steerable wavelet bases as the non-local basis Φ of the regular DCFN architecture. Furthermore, for learning tomosynthesis artifact suppression rules according to the artifact directional, we introduced new learnable local bases $\{\Psi_{H_0}^{(l)}, \Psi_{B_0}^{(l)}, \cdots, \Psi_{B_3}^{(l)}\}$ for processing high-frequency signals as follows:

$$\begin{cases} \hat{C}_{H_0}^{(l)} = C_{H_0}^{(l)} \circledast \overline{\Psi}_{H_0}^{(l+1)} \\ \hat{C}_{B_1}^{(l)} = C_{B_1}^{(l)} \circledast \overline{\Psi}_{B_1}^{(l+1)} \\ \vdots \\ \hat{C}_{B_3}^{(l)} = C_{B_3}^{(l)} \circledast \overline{\Psi}_{B_3}^{(l+1)} \end{cases} \tag{5}$$

The architecture of the proposed DSCFN is illustrated in Fig.3.

IV. Experiments and Results

Using the Astra Toolbox [17], we simulated a tomosynthesis scanning setup in which the X-ray source moves linearly while the patient and the X-ray detector remain static. Then, sparse-view projections were acquired at equidistant positions, and finally, images were reconstructed using SIRT [18].

Experiments were conducted using The Clinical Proteomic Tumor Analysis Consortium Pancreatic Ductal Adenocarcinoma (CPTAC-PDA) collection[1] that comprises CT, and Magnetic Resonance Imaging (MRI) scans from 116 patients. From these original $512 \times 512 \times slc,\ slc > 120$ volumes, we generated tomographic volumes $512 \times 100 \times slc,\ slc > 120$ by resizing the axial slices. Based on these data, we simulated sparse-sampled tomosynthesis images using 20 X-ray projections and full-sampled tomosynthesis images using 100 X-ray projections.

Then, we trained deep models for mapping sparse-sampled into full-sampled images in order to suppress ripple artifacts. For this, slices 512×100 were fed into the networks, and due to this input image resolution, 2-level decomposition networks were used.

[1] https://wiki.cancerimagingarchive.net/display/Public/CPTAC-PDA

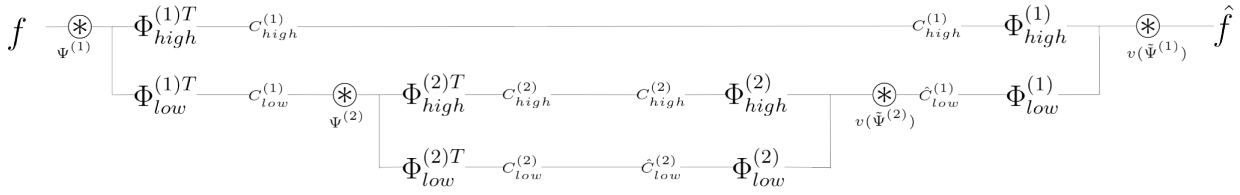

Fig. 1. Architecture of the DCFN proposed by Ye *et al.* [13].

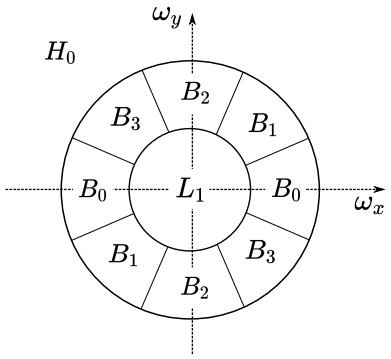

Fig. 2. Idealized spectral decomposition using steerable wavelets within $K = 4$ orientations, where ω_x and ω_y denote the axis in the Fourier space.

The performance of our DSCFN was compared with the U-net [19], and tight U-net [20] (which is a DCFN) in terms of Peak Signal-to-Noise Ratio (PSNR) [21] that is a commonly used metric used to quantify the quality of images subject to noise.

The PSNR metrics in Table I show the strength of our method in relation to both U-net and tight U-net in the vast majority of the scans tested in the CPTAC dataset. In fact, paired samples t-tests resulted in p-values 7×10^{-10} and 4×10^{-6} when comparing the results of the DSCFN with those from U-net and tight U-net. Such p-values rejects the hypothesis that the mean of PSNR distribution are the same for DSCFN and U-net or tight U-net.

To evaluate the generalization capabilities of the models trained in the CPTAC dataset, we performed tests using sparse-view tomosynthesis simulated from CT images of the Visible Human (VH) Project dataset[2]. Table II shows the PSNR metrics related to the images generated from the VH dataset, and Fig. 4 shows the error maps (*i.e.*, the difference between the reconstructions and the ground truths) of SIRT tomosynthesis processed by U-net, tight U-net, and the proposed DSCFN. A closer look at the DSCFN error map reveals thinner horizontal strips in red.

[2]https://mri.radiology.uiowa.edu/visible_human_datasets.html

TABLE I

AVERAGE AND STANDARD ERROR OF THE PSNR VALUES FOR ALL THE SLICES FROM TWENTY SCANS WITHIN CPTAC DATASET. THE HIGHEST AVERAGE VALUES ARE HIGHLIGHTED IN BOLD.

	U-net	**tight U-net**	**DSCFN**
scan#1	51.38(±0.27)	51.56(±0.27)	**51.95(±0.25)**
scan#2	48.44(±0.26)	48.58(±0.27)	**49.02(±0.24)**
scan#3	48.55(±0.26)	48.70(±0.27)	**49.14(±0.24)**
scan#4	48.63(±0.25)	48.93(±0.27)	**49.24(±0.25)**
scan#5	48.64(±0.34)	49.15(±0.37)	**49.22(±0.37)**
scan#6	43.36(±0.42)	43.79(±0.45)	**44.14(±0.48)**
scan#7	43.28(±0.41)	43.69(±0.45)	**44.04(±0.48)**
scan#8	45.59(±0.36)	46.14(±0.41)	**46.38(±0.44)**
scan#9	45.71(±0.35)	46.30(±0.40)	**46.48(±0.42)**
scan#10	38.03(±0.71)	38.47(±0.76)	**38.72(±0.79)**
scan#11	44.88(±0.45)	45.33(±0.49)	**45.57(±0.52)**
scan#12	42.90(±0.25)	**43.28(±0.21)**	43.15(±0.23)
scan#13	44.80(±0.95)	44.76(±0.08)	**45.09(±0.88)**
scan#14	42.49(±0.20)	42.81(±0.19)	**42.82(±0.20)**
scan#15	44.19(±0.23)	**44.81(±0.20)**	**44.81(±0.23)**
scan#16	48.06(±0.34)	48.18(±0.35)	**48.45(±0.34)**
scan#17	47.10(±0.11)	47.16(±0.11)	**47.49(±0.11)**
scan#18	51.42(±0.30)	51.94(±0.31)	**52.29(±0.31)**
scan#19	43.90(±0.13)	43.96(±0.15)	**44.00(±0.17)**
scan#20	50.72(±0.28)	51.17(±0.29)	**51.75(±0.32)**

TABLE II

AVERAGE AND STANDARD ERROR OF THE PSNR VALUES FOR ALL THE SLICES FROM FOUR SCANS WITHIN VH DATASET. THE HIGHEST AVERAGE VALUES ARE HIGHLIGHTED IN BOLD.

	U-net	**tight U-net**	**DSCFN**
scan#1	53.60(±0.35)	54.05(±0.41)	**54.54(±0.41)**
scan#2	57.43(±0.30)	57.59(±0.31)	**57.91(±0.33)**
scan#3	53.14(±0.20)	53.61(±0.22)	**54.02(±0.21)**
scan#4	54.08(±0.25)	54.36(±0.25)	**54.71(±0.25)**

V. CONCLUSION

We proposed the DSCFN for suppressing directional artifacts from tomographic data. The proposed network is a DCFN-like architecture with steerable wavelet bases as non-local bases Φ for decomposing the signal into directional components. Furthermore, to ensure specialized artifact suppression rules for each directional component, we introduced convolutional local bases Ψ for individually processing all the high-frequency signals.

The DSCFN results were superior to those obtained with state-of-the-art networks in terms of PSNR as indicated by

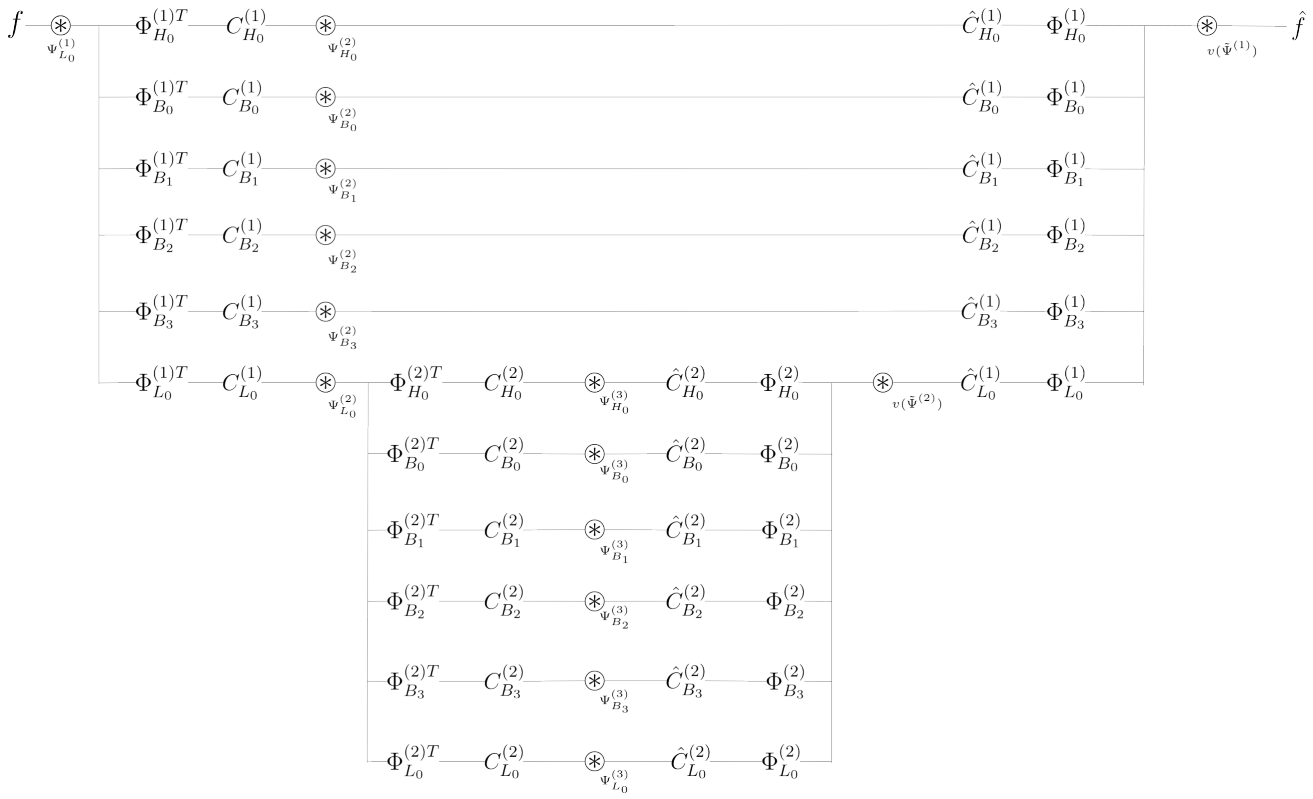

Fig. 3. Architecture of the proposed DSCFN. Notice that each steerable wavelet component $C_K^{(n)}$ is individually processed by its respective convolutional basis $\Psi_K^{(n+1)}$.

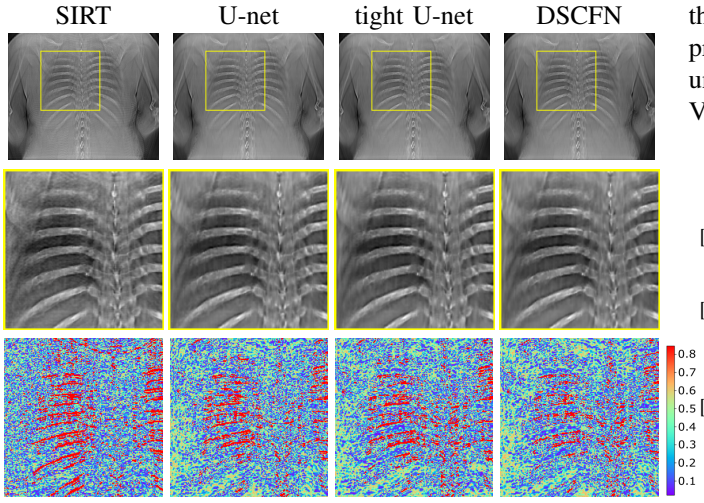

Fig. 4. Top row: SIRT tomosynthesis slice without postprocessing, and processed by U-net, tight U-net, and DSCFN. Second row: Zoom in the ROIs indicated in the first row. Bottom row: error maps of the ROIs.

paired samples t-tests. Furthermore, by assessing reconstruction error maps, we could also confirm the reduction of directional artifacts along a reconstruction.

VI. Acknowledgments

This research is funded by the FWO, Belgium SBO project MetroFlex (S004217N), the European Commission through the INTERREG Vlaanderen Nederland program project Smart*Light (0386), and the Flemish Government under the "Onderzoeksprogramma Artificiele Intelligentie (AI) Vlaanderen" programme.

References

[1] H. Machida, T. Yuhara, M. Tamura, T. Ishikawa, E. Tate, E. Ueno, K. Nye, and J. M. Sabol, "Whole-body Clinical Applications of Digital Tomosynthesis," *Radiographics*, vol. 36, no. 3, pp. 735–750, 2016.

[2] S. Sone, T. Kasuga, F. Sakai, J. Aoki, I. Izuno, Y. Tanizaki, H. Shigeta, and K. Shibata, "Development of a high-resolution digital tomosynthesis system and its clinical application." *Radiographics*, vol. 11, no. 5, pp. 807–822, 1991.

[3] E. Quaia, G. Grisi, E. Baratella, R. Cuttin, G. Poillucci, S. Kus, and M. A. Cova, "Diagnostic Imaging Costs Before and After Digital Tomosynthesis Implementation in Patient Management After Detection of Suspected Thoracic Lesions on Chest Radiography," *Insights into Imaging*, vol. 5, no. 1, pp. 147–155, 2014.

[4] A. R. García and D. M. Fernández, "Breast Tomosynthesis: state of the art," *Radiología (English Edition)*, vol. 61, no. 4, pp. 274–285, 2019.

[5] C. You, Q. Yang, H. Shan, L. Gjesteby, G. Li, S. Ju, Z. Zhang, Z. Zhao, Y. Zhang, W. Cong *et al.*, "Structurally-sensitive Multi-scale Deep Neural Network for Low-dose CT Denoising," *IEEE Access*, vol. 6, pp. 41 839–41 855, 2018.

[6] J. M. Wolterink, T. Leiner, M. A. Viergever, and I. Išgum, "Generative Adversarial Networks for Noise Reduction in Low-dose CT," *IEEE Transactions on Medical Imaging*, vol. 36, no. 12, pp. 2536–2545, 2017.

[7] 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.

[8] Y. Ma, B. Wei, P. Feng, P. He, X. Guo, and G. Wang, "Low-dose CT Image Denoising using a Generative Adversarial Network with a Hybrid Loss Function for Noise Learning," *IEEE Access*, vol. 8, pp. 67 519–67 529, 2020.

[9] 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*, vol. 40, no. 12, pp. 3663–3673, 2021.

[10] J. C. Ye, Y. Han, and E. Cha, "Deep Convolutional Framelets: a general deep learning framework for inverse problems," *SIAM Journal on Imaging Sciences*, vol. 11, no. 2, pp. 991–1048, 2018.

[11] R. Yin, T. Gao, Y. M. Lu, and I. Daubechies, "A Tale of Two Bases: local-nonlocal regularization on image patches with convolution framelets," *SIAM Journal on Imaging Sciences*, vol. 10, no. 2, pp. 711–750, 2017.

[12] E. P. Simoncelli and W. T. Freeman, "The Steerable Pyramid: a flexible architecture for multi-scale derivative computation," in *Proceedings of International Conference on Image Processing*, vol. 3. IEEE, 1995, pp. 444–447.

[13] J. C. Ye, Y. Han, and E. Cha, "Deep convolutional framelets: A general deep learning framework for inverse problems," *SIAM Journal on Imaging Sciences*, vol. 11, no. 2, pp. 991–1048, 2018.

[14] R. Yin, T. Gao, Y. M. Lu, and I. Daubechies, "A tale of two bases: Local-nonlocal regularization on image patches with convolution framelets," *SIAM Journal on Imaging Sciences*, vol. 10, no. 2, pp. 711–750, 2017.

[15] W. T. Freeman, E. H. Adelson *et al.*, "The Design and Use of Steerable Filters," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 13, no. 9, pp. 891–906, 1991.

[16] A. Karasaridis and E. Simoncelli, "A filter Design Technique for Steerable Pyramid Image Transforms," in *Proceedings of IEEE International Conference on Acoustics, Speech, and Signal Processing*, vol. 4. IEEE, 1996, pp. 2387–2390.

[17] W. Van Aarle, W. J. Palenstijn, J. Cant, E. Janssens, F. Bleichrodt, A. Dabravolski, J. De Beenhouwer, K. J. Batenburg, and J. Sijbers, "Fast and Flexible X-ray Tomography Using the ASTRA Toolbox," *Optics Express*, vol. 24, no. 22, pp. 25 129–25 147, 2016.

[18] P. Gilbert, "Iterative Methods for the Three-dimensional Reconstruction of an Object from Projections," *Journal of Theoretical Biology*, vol. 36, no. 1, pp. 105–117, 1972.

[19] 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.

[20] 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.

[21] A. Hore and D. Ziou, "Image quality metrics: Psnr vs. ssim," in *2010 20th international conference on pattern recognition*. IEEE, 2010, pp. 2366–2369.