WIA-LD2ND: Wavelet-based Image Alignment for Self-supervised Low-Dose CT Denoising

Haoyu Zhao[1], Yuliang Gu[1], Zhou Zhao[2], Bo Du[1], Yongchao Xu[1(✉)], and Rui Yu[3]

[1] National Engineering Research Center for Multimedia Software, School of Computer Science, Medical Artificial Intelligence Research Institute of Renmin Hospital, Wuhan University, Wuhan, China
yongchao.xu@whu.edu.cn

[2] School of Computer Science, Central China Normal University, Hubei, China

[3] University of Louisville, Louisville, USA

Abstract. In clinical examinations and diagnoses, low-dose computed tomography (LDCT) is crucial for minimizing health risks compared with normal-dose computed tomography (NDCT). However, reducing the radiation dose compromises the signal-to-noise ratio, leading to degraded quality of CT images. To address this, we analyze LDCT denoising task based on experimental results from the frequency perspective, and then introduce a novel self-supervised CT image denoising method called WIA-LD2ND, only using NDCT data. The proposed WIA-LD2ND comprises two modules: Wavelet-based Image Alignment (WIA) and Frequency-Aware Multi-scale Loss (FAM). First, WIA is introduced to align NDCT with LDCT by mainly adding noise to the high-frequency components, which is the main difference between LDCT and NDCT. Second, to better capture high-frequency components and detailed information, Frequency-Aware Multi-scale Loss (FAM) is proposed by effectively utilizing multi-scale feature space. Extensive experiments on two public LDCT denoising datasets demonstrate that our WIA-LD2ND, only uses NDCT, outperforms existing several state-of-the-art weakly-supervised and self-supervised methods. Source code is available at https://github.com/zhaohaoyu376/WI-LD2ND.

Keywords: Low-dose computed tomography · self-supervised learning · image denoising.

1 Introduction

Computed tomography (CT) has become a widely utilized tool in medical diagnosis. However, increased usage has raised concerns regarding potential risks associated with excessive radiation exposure [11]. The widely recognized principle of ALARA (as low as reasonably achievable) [25] is extensively embraced to minimize exposure through strategies such as employing sparse sampling and reducing tube flux. Reducing the X-ray radiation dose, however, leads to poor-quality images with noticeable noise, which poses challenges for accurate diagnosis [4]. Therefore, the development of image denoising techniques [31] that can effectively handle CT modalities emerges as a critical and urgent need in clinical practice, to ensure both patient safety and diagnostic precision.

In recent years, advanced deep learning networks have proven to be highly effective in reducing noise in low-dose computed tomography (LDCT) than traditional denoising methods [3,7]. Supervised denoising methods [16,15], such as CTformer [27] and ASCON [2], learn the end-to-end mapping from low-dose to normal-dose CT images. Generative adversarial networks (GANs) are also utilized in LDCT denoising task, which do not need paired data, but lots of unpaired data for training[24,1,14,12].

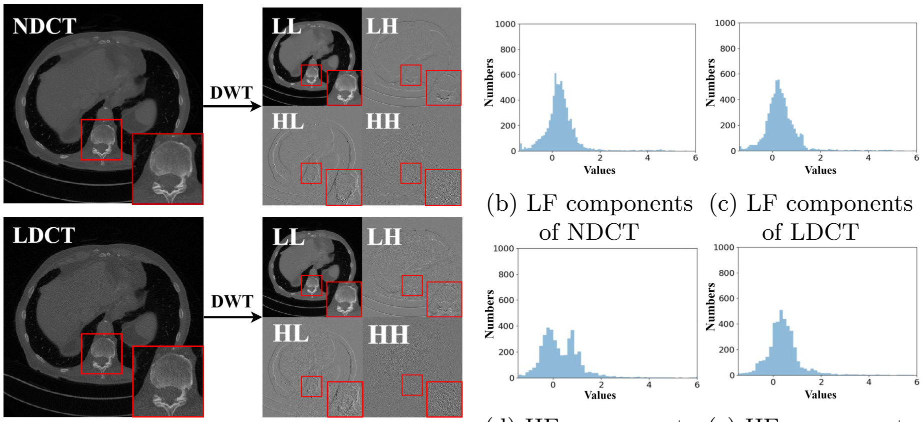
Despite impressive results, these methods encounter challenges as they require both LDCT and NDCT images [18], either paired images or a large amount of unpaired images, which are often unavailable in practice due to high costs, privacy, and ethical concerns. Therefore, it is essential to develop self-supervised methods that harness the potent capabilities of deep neural networks while minimizing the need for extensive labeled data. Several self-supervised methods have been proposed for LDCT denoising, including but not limited to Blin2Unblind [28], Noise2Sim [23], Neighbor2Neighbor [10] and FIRE [19] among others [26,13,28]. However, these methods primarily concentrate on spatial domain information, overlooking the critical importance of frequency domain details. The crucial distinction between low-dose CT and normal-dose CT in high-frequency components (see Fig. 1) is not well explored.

In this paper, we design a novel self-supervised LDCT denoising method, only using NDCT data, called WIA-LD2ND. We first analyze LDCT denoising task from the frequency perspective and then propose a module called Wavelet-based Image Alignment (WIA), which aligns LDCT with NDCT by mainly adding noise to the high-frequency components of both LDCT and NDCT. We also propose a module called Frequency-Aware Multi-scale Loss (FAM) to capture high-frequency components in multi-scale feature space.

Our WIA-LD2ND offers three major contributions as follows: 1) We analyze the LDCT denoising task from a frequency perspective, offering novel insight into its optimization. 2) We introduce a simple and efficient module to align NDCT and LDCT, facilitating self-supervised learning. 3) We propose a frequency-aware multi-scale loss, enabling the reconstruction network to effectively handle high-frequency components.

2 Method

Figure 2 presents the overview of the proposed WIA-LD2ND, comprising of two novel modules: Wavelet-based Image Alignment (WIA) and Frequency-Aware Multi-scale Loss (FAM). NDCT image x is passed through the WIA to destroy some high-frequency components and then fed into the reconstruction network. The high-frequency components of the reconstructed CT y and input x, [y_{LH},



(b) LF components of NDCT (c) LF components of LDCT

(d) HF components of NDCT (e) HF components of LDCT

(a) Visualization of the results of NDCT and LDCT after Wavelet transform results

Fig. 1. (a) Visualization of results of NDCT and LDCT after Discrete Wavelet Transform (DWT). The primary differences between NDCT and LDCT are at the high frequency components $[LH, HL, HH]$. (b-c) Visualize the normalized low-frequency (LF) component LL features of NDCT and LDCT, while (d-e) display the normalized high-frequency (HF) component $[LH, HL, HH]$ features. We adopt the first residual block of pre-trained ResNet-18 [9] to extract image features.

y_{HL}, y_{HH}] and [x_{LH}, x_{HL}, x_{HH}], are relatively flat and hard to capture [5], as shown in Fig. 1 (a). To address this challenge, we propose integrating these high-frequency components into encoders to compute a loss $\mathcal{L}_{FAM}$ within the feature space to enhance the capability of the reconstruction network in capturing high-frequency components and detailed information more effectively. During training, we employ an alternating learning strategy to optimize the reconstruction network and FAM to improve learning efficiency and accuracy of results, which is similar to GAN-based methods [32]. We begin by analyzing the LDCT denoising task from a novel perspective, followed by detailed introduction of the two proposed modules.

2.1 Analysis of LDCT Denoising From Frequency Perspective

Images contain different frequency ranges and spatial locations information. The Discrete Wavelet Transform (DWT), using the Haar wavelet as in [17], is selected for frequency analysis for its simplicity and efficiency. DWT is commonly employed in the field of computer vision and offers a straightforward and computationally effective technique for dividing the input image into low-frequency sub-band and high-frequency sub-bands. It has four filters, LL^T, LH^T, HL^T and HH^T, demonstrating the texture, horizontal details, vertical details, and diagonal information respectively [29], in which low and high pass filters are:

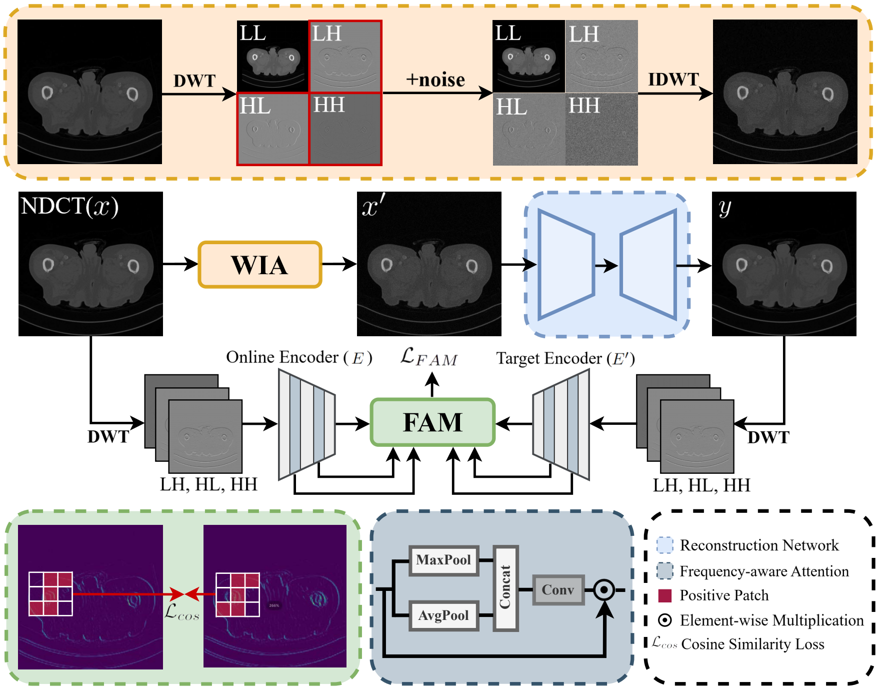
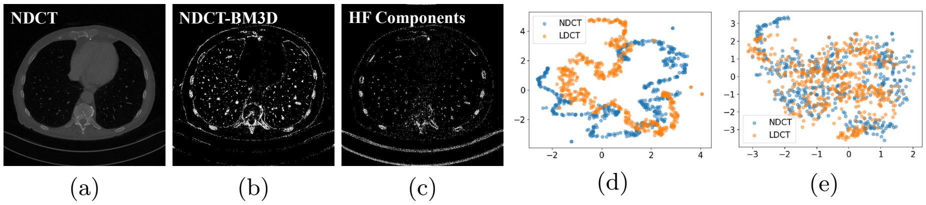


Fig. 2. Overview of our proposed WIA-LD2ND. NDCT image x is passed through the WIA and then fed into the reconstruction network. The high-frequency components of the denoised CT y and the input image x are both fed into FAM to compute the loss capturing high-frequency components in multi-scale feature space.

$$L^T = \frac{1}{\sqrt{2}}[1,1], H^T = \frac{1}{\sqrt{2}}[-1,1] \tag{1}$$

As shown in Fig. 1 (a), after DWT, the main differences between normal-dose CT (NDCT) and low-dose CT (LDCT) are observed in the high-frequency sub-images $[LH, HL, HH]$, with little difference in the LL. Fig. 1 (b-e) further support our conclusion, demonstrating that LDCT and NDCT have significant differences in the high-frequency components at the feature space, while differences in the low-frequency components are comparatively minor. In Fig. 3 (a-c) and Fig. 4, we find that previous LDCT denoising methods such as BM3D [3] performs poorly at reconstructing high-frequency components.

Based on these observations, we conclude that high-frequency components should be the main focus for LDCT denoising, where previous methods falter the most. This highlights the critical necessity for improved techniques in handling high-frequency components during the denoising process.



(a) (b) (c) (d) (e)

Fig. 3. (a-c) Visualization of NDCT, residual between NDCT and result of BM3D [3] (a classical denoising method), and high-frequency components in spatial domain. The residual is converted into a clean binary image for clarity. We filter high-frequency band from image and then convert the result into a binary image. (d-e) Visualization of the tSNE images of feature distribution on the NDCT, LDCT, and their respective transformations after applying WIA on Mayo-2016 dataset. We adopt the first residual block of pre-trained ResNet-18 to extract image features.

2.2 Wavelet-based Image Alignment

According to previous analysis, we employ the Discrete Wavelet Transform (DWT) to decompose the input image x into two sets of components: low-frequency component denoted as x_{LL}, which captures and preserves the smooth surface and texture information, and high-frequency components denoted by $[x_{LH}, x_{HL}, x_{HH}]$. These high-frequency components are essential for capturing intricate texture details, representing the primary distinctions between low-dose CT (LDCT) and normal-dose CT (NDCT) images. Therefore, we make them similar in a simple and effective way, by mainly adding Gaussian noise into the high-frequency components of NDCT and LDCT images.

$$\begin{aligned} x'_{LL} = x_{LL} + noise_{LL}, \quad x'_{LH} = x_{LH} + noise_{LH}, \\ x'_{HL} = x_{HL} + noise_{HL}, \quad x'_{HH} = x_{HH} + noise_{HH}, \end{aligned} \tag{2}$$

where $noise_{LL}, noise_{LH}, noise_{HL}, noise_{HH}$ follow Gaussian distributions with mean 0 and variance $\sigma^2_{LL}, \sigma^2_{LH}, \sigma^2_{HL}, \sigma^2_{HH}$, respectively. Notably, σ_{LH}, σ_{HL}, and σ_{HH} are larger than σ_{LL}. We then conduct inverse DWT (IDWT) on the modified components $[x'_{LL}, x'_{LH}, x'_{HL}, x'_{HH}]$ to reconstruct x'. As shown in Fig. 3 (d-e), after applying WIA module, NDCT and LDCT images share the same feature space, indicating successful alignment.

WIA eliminates the need for paired data. Instead, we only require NDCT images for training a model to denoise from x' (NDCT after WIA) to x (original NDCT), thereby facilitating self-supervised learning.

2.3 Frequency-Aware Multi-scale Loss

High-frequency components play a crucial role in low-dose CT images denoising, as analyzed in Sec. 2.1. However, the CNN-based and Transformer-based models tend to focus primarily on low-frequency representations, making it difficult for models to capture the high-frequency components [5]. Therefore, we design

Frequency-Aware Multi-scale Loss (FAM) which is to focus the network more on the high-frequency components and detail information of the images.

The high-frequency components of denoised CT y and NDCT x, [y_{LH}, y_{HL}, y_{HH}] and [x_{LH}, x_{HL}, x_{HH}], are fed into Online Encoder E and Target Encoder E', respectively, both using the same lightweight architecture. Following previous studies [6,8], the parameters of the Target Encoder E' are an exponential moving average of the parameters in the Online Encoder E. The process is as follows:

$$f_1, f_2, f_3 = E(x_{LH}, x_{HL}, x_{HH}), f'_1, f'_2, f'_3 = E'(y_{LH}, y_{HL}, y_{HH}). \tag{3}$$

We introduce a Frequency-aware Attention mechanism in the encoders, designed to selectively emphasize or de-emphasize areas within the input feature map based on their frequency content, as illustrated in Fig. 2 Given the input feature map $f_{n\{n=1,2,3\}}$, we apply max and average pooling to extract prominent features. We then concatenate them and pass through a convolutional layer with Sigmoid activation to generate spatial attention weights.

Unlike previous studies [2,30], our approach segments multi-scale features extracted from specific layers into patches $f_n^{(i)}$. We then select patches that are most similar to their adjacent counterparts, focusing on those that exhibit shared structural characteristics closely associated with high-frequency components. To this end, we use the cosine similarity s on the feature space:

$$\mathrm{s}(i,j) = {f^{(i)}}^\top f^{(j)} / \|f^{(i)}\|_2 \|f^{(j)}\|_2. \tag{4}$$

We then select the similar feature patches $\{f_n^{(j)}\}_{j\in P^{(i)}}$, where $P^{(i)}$ is a set of feature patch indices of top-4 patches. For the $f'^{(i)}$ from the Online Network, we select the same positive feature patches $P^{(i)}$. We then aggregate positive patches $\{f_n^{(j)}\}_{j\in P^{(i)}}$ and $\{f_n'^{(j)}\}_{j\in P^{(i)}}$ using global average pooling (GAP) and multi-layer perceptron (MLP) yielding g and g', respectively. Finally, the Frequency-Aware Multi-scale Loss $\mathcal{L}_{FAM}$ is given by:

$$\mathcal{L}_{FAM} = \|g - g'\|_2^2. \tag{5}$$

The final loss is defined as $\mathcal{L} = \mathcal{L}_{pixel}(x, y) + \lambda \mathcal{L}_{FAM}$, where $\mathcal{L}_{pixel}$ consists of two common supervised losses: MSE and SSIM, defined as $\mathcal{L}_{pixel}(x, y) = \mathcal{L}_{MSE}(x, y) + \mathcal{L}_{SSIM}(x, y)$. λ is set to 0.01 in this paper.

3 Experiments

3.1 Dataset and Training Details

We conduct experiments on two public LDCT denoising datasets, Mayo-2016[4] and Mayo-2020[5], from the NIH AAPM-Mayo Clinic Low-Dose CT Grand Challenge [21,22]. We select 5410 image pairs (512×512) from 9 patients in Mayo-2016

[4] https://ctcicblog.mayo.edu/2016-low-dose-ct-grand-challenge/

[5] https://wiki.cancerimagingarchive.net/pages/viewpage.action?pageId=52758026

Table 1. Performance comparison on the Mayo-2016 [21] and Mayo-2020 [22] datasets. The best result is in **bold**, and the second best is underlined.

Methods	Mayo-2016		Mayo-2020		Avg	
	PSNR	SSIM	PSNR	SSIM	PSNR	SSIM
BM3D (TIP2007) [3]	35.50	86.95	37.67	89.53	36.59	88.24
DIP (CVPR2018) [26]	37.25	85.94	40.16	95.89	38.71	90.92
Noise2Sim (TMI2022) [23]	**38.51**	**90.15**	39.16	90.90	38.84	90.53
Blind2Unblind (CVPR2022) [28]	35.84	81.15	40.96	94.84	38.40	88.00
Neighbor2Neighbor (TIP2022) [10]	35.27	86.79	36.78	94.06	36.03	90.43
ZS-N2N (CVPR2023) [20]	38.10	87.64	44.00	97.06	41.05	92.35
CycleGAN (ICCV2017) [32]	37.68	89.22	40.18	97.93	38.93	93.58
CUT (ECCV2020) [24]	37.96	89.93	41.11	97.61	39.54	93.77
WIA-LD2ND (ours)	38.15	90.00	**44.64**	**98.31**	**41.40**	**94.16**

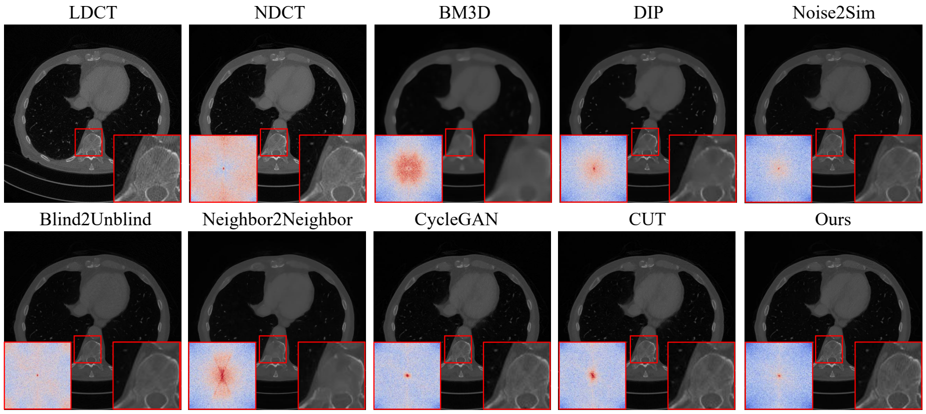


Fig. 4. Qualitative comparison of different methods on the Mayo-2020 dataset [22].

for training and 526 for testing. We select the reconstruction parameter combination of {1mm, D45}. From Mayo-2020, 2082 pairs (512×512) from 12 patients are used for training, with 672 pairs from 4 patients for testing.

In all our experiments, we only use NDCT and choose a backbone identical to that used in [32,24], along with employing the same data augmentation strategy as in [2]. For Mayo-2016, we set $\sigma_{LL} = 100$, $\sigma_{LH} = 200$, $\sigma_{HL} = 200$, and $\sigma_{HH} = 150$ in Eq. (2). For Mayo-2020, the noise variances are $\sigma_{LL} = 25$, $\sigma_{LH} = 50$, $\sigma_{HL} = 50$, and $\sigma_{HH} = 50$. We employ the Adam optimizer with the momentum parameters as $\beta_1 = 0.9$, $\beta_2 = 0.99$ and initial learning rate 1.0×10^{-4}. Our network is trained over 200 epochs using a single NVIDIA GeForce RTX 3090.

3.2 Experiments Results

To evaluate the denoising efficacy of our WIA-LD2ND, we conduct comparative experiments against various denoising methods. The comparisons include traditional methods like BM3D [3], self-supervised methods including DIP [26], Noise2Sim [23], Blind2Unblind [28], Neighbor2Neighbor [10] and ZS-N2N [20] as well as weakly-supervised methods such as CycleGAN [32] and CUT [24]. We

Table 2. Ablation studies are conducted to validate the effectiveness of each module on the Mayo-2016 [21] and Mayo-2020 [22] datasets. WIA* represents directly adding Gaussian noise to NDCT, while FAM* denotes the computation of high-frequency components of y and x at the feature level by directly employing the MSE loss.

Methods	**#Params**	**Mayo-2016**		**Mayo-2020**		**Avg**	
		PSNR	SSIM	PSNR	SSIM	PSNR	SSIM
Baseline	11.37M	34.26	78.03	40.32	96.99	37.29	87.51
Baseline + WIA*	11.37M	35.49	87.73	42.00	98.26	38.75	93.00
Baseline + WIA	11.37M	37.85	89.77	42.73	98.26	40.29	94.02
Baseline + FAM*	13.83M	34.00	79.12	41.97	98.02	37.99	88.69
Baseline + FAM	13.83M	34.78	80.27	42.46	98.25	38.62	89.15
WIA-LD2ND	13.83M	**38.15**	**90.00**	**44.64**	**98.31**	**41.40**	**94.16**

use two widely-adopted metrics, namely peak signal-to-noise ratio (PSNR) and structural similarity index measure (SSIM) to evaluate the performance.

Table 1 shows that our WIA-LD2ND, only using NDCT images, can achieve good performance on both the Mayo-2016 and Mayo-2020 datasets. Compared to the latest state-of-the-art self-supervised and weakly-supervised methods, our WIA-LD2ND achieves significant performance improvements.

Figure 4 presents the reconstruction results, with the subplots in the bottom left corner of the pictures showing the noise power spectrum (NPS), where blue indicates it is closer to the normal-dose CT. As illustrated, our WIA-LD2ND achieves the best results, reconstructing the most detailed information and exhibiting the bluest NPS. Conversely, the results of BM3D and DIP are over-smoothed and compromised with structured artifacts. Additionally, other deep-learning-based methods tend to remove noise aggressively. Our WIA-LD2ND prioritizes the preservation of informative details.

Ablation Studies. To evaluate the effectiveness of our proposed modules, including WIA and FAM, we conduct ablation experiments on Mayo-2016 [21] and Mayo-2020 [22]. The results are shown in Table 2. WIA* represents adding noise directly to NDCT, while FAM* involves the computation of MSE loss directly for the feature of high-frequency components. Our designs achieve better performance than their variants. It reveals that both WIA and FAM are well-designed and contribute to performance gains. More ablation studies on additional hyper-parameters and noise parameters are available in the supplementary materials. WIA-LD2ND incurs an increase of 2.46M parameters compared to the baseline. Considering the significant performance improvement over the baseline model **without any extra inference time**, this slight increase in training cost is acceptable.

4 Discussion and Conclusion

In this paper, we analyze the LDCT denoising task from a novel perspective and propose a self-supervised method called WIA-LD2ND. This method only utilizes NDCT images and incorporates two novel modules: Wavelet-based Image Alignment (WIA), which aligns NDCT and LDCT by destroying some high-frequency components, and Frequency-Aware Multi-scale Loss (FAM), which enhances the reconstruction network's ability to capture high-frequency components and detailed information, thus improving denoising performance. Extensive experimental results demonstrate the superior performance and the effectiveness of our designs. It is noteworthy that WIA-LD2ND increases the number of parameters by 2.46M compared to the baseline during training, without requiring extra inference time. Exploring LDCT denoising from a frequency perspective presents a promising direction for future research.

Acknowledgement. This work was supported in part by the National Key Research and Development Program of China (2023YFC2705700), NSFC 62222112, 62225113, and 62176186, the Postdoctoral Fellowship Program of CPSF (GZC20230924), the Innovative Research Group Project of Hubei Province under Grants (2024AFA017), the NSF of Hubei Province of China (2024AFB245), and CAAI Huawei MindSpore Open Fund.

Disclosure of Interests. The authors have no competing interests to declare that are relevant to the content of this article.

References

1. Bera, S., Biswas, P.K.: Axial Consistent Memory GAN with Interslice Consistency Loss for Low Dose Computed Tomography Image Denoising. IEEE Transactions on Radiation and Plasma Medical Sciences (2023)
2. Chen, Z., Gao, Q., Zhang, Y., Shan, H.: ASCON: Anatomy-aware supervised contrastive learning framework for low-dose CT denoising. In: Proc. of Intl. Conf. on Medical Image Computing and Computer Assisted Intervention. pp. 355–365 (2023)
3. Dabov, K., Foi, A., Katkovnik, V., Egiazarian, K.: Image denoising by sparse 3-D transform-domain collaborative filtering. IEEE Trans. on Image Processing **16**(8), 2080–2095 (2007)
4. Diwakar, M., Kumar, M.: A review on CT image noise and its denoising. Biomedical Signal Processing and Control **42**, 73–88 (2018)
5. Gou, Y., Hu, P., Lv, J., Zhu, H., Peng, X.: Rethinking image super resolution from long-tailed distribution learning perspective. In: Proc. of IEEE Conf. on Computer Vision and Pattern Recognition. pp. 14327–14336 (2023)
6. Grill, J.B., Strub, F., Altché, F., Tallec, C., Richemond, P., Buchatskaya, E., Doersch, C., Avila Pires, B., Guo, Z., Gheshlaghi Azar, M., et al.: Bootstrap your own latent-a new approach to self-supervised learning. Advances in neural information processing systems **33**, 21271–21284 (2020)
7. Gu, S., Zhang, L., Zuo, W., Feng, X.: Weighted nuclear norm minimization with application to image denoising. In: Proc. of IEEE Conf. on Computer Vision and Pattern Recognition. pp. 2862–2869 (2014)
8. He, K., Fan, H., Wu, Y., Xie, S., Girshick, R.: Momentum contrast for unsupervised visual representation learning. In: Proc. of IEEE Conf. on Computer Vision and Pattern Recognition. pp. 9729–9738 (2020)

9. He, K., Zhang, X., Ren, S., Sun, J.: Deep residual learning for image recognition. In: Proc. of IEEE Conf. on Computer Vision and Pattern Recognition. pp. 770–778 (2016)
10. Huang, T., Li, S., Jia, X., Lu, H., Liu, J.: Neighbor2Neighbor: a self-supervised framework for deep image denoising. IEEE Trans. on Image Processing **31**, 4023–4038 (2022)
11. Immonen, E., Wong, J., Nieminen, M., Kekkonen, L., Roine, S., Törnroos, S., Lanca, L., Guan, F., Metsälä, E.: The use of deep learning towards dose optimization in low-dose computed tomography: A scoping review. Radiography **28**(1), 208–214 (2022)
12. Jing, J., Wang, T., Yu, H., Lu, Z., Zhang, Y.: Inter-slice consistency for unpaired low-dose CT denoising using boosted contrastive learning. In: Proc. of Intl. Conf. on Medical Image Computing and Computer Assisted Intervention. pp. 238–247 (2023)
13. Jing, J., Xia, W., Hou, M., Chen, H., Liu, Y., Zhou, J., Zhang, Y.: Training low dose CT denoising network without high quality reference data. Physics in Medicine & Biology **67**(8), 084002 (2022)
14. Kwon, T., Ye, J.C.: Cycle-free cyclegan using invertible generator for unsupervised low-dose CT denoising. IEEE Transactions on Computational Imaging **7**, 1354–1368 (2021)
15. Li, H., Yang, X., Yang, S., Wang, D., Jeon, G.: Transformer with double enhancement for low-dose CT denoising. IEEE journal of biomedical and health informatics (2022)
16. Lin, D.J., Johnson, P.M., Knoll, F., Lui, Y.W.: Artificial intelligence for MR image reconstruction: an overview for clinicians. Journal of Magnetic Resonance Imaging **53**(4), 1015–1028 (2021)
17. Liu, L., Liu, J., Yuan, S., Slabaugh, G., Leonardis, A., Zhou, W., Tian, Q.: Wavelet-based dual-branch network for image demoiréing. In: Proc. of European Conf. on Computer Vision. pp. 86–102 (2020)
18. Liu, X., Liang, X., Deng, L., Tan, S., Xie, Y.: Learning low-dose CT degradation from unpaired data with flow-based model. Medical Physics **49**(12), 7516–7530 (2022)
19. Long, Y., Pan, J., Xi, Y., Zhang, J., Wu, W.: Full image-index remainder based single low-dose DR/CT self-supervised denoising. In: Proc. of Intl. Conf. on Medical Image Computing and Computer Assisted Intervention. pp. 466–475 (2023)
20. Mansour, Y., Heckel, R.: Zero-shot: Efficient image denoising without any data. In: Proc. of IEEE Conf. on Computer Vision and Pattern Recognition. pp. 14018–14027 (2023)
21. McCollough, C.H., Bartley, A.C., Carter, R.E., Chen, B., Drees, T.A., Edwards, P., Holmes III, D.R., Huang, A.E., Khan, F., Leng, S., et al.: Low-dose CT for the detection and classification of metastatic liver lesions: results of the 2016 low dose CT grand challenge. Medical physics **44**(10), e339–e352 (2017)
22. Moen, T.R., Chen, B., Holmes III, D.R., Duan, X., Yu, Z., Yu, L., Leng, S., Fletcher, J.G., McCollough, C.H.: Low-dose CT image and projection dataset. Medical physics **48**(2), 902–911 (2021)
23. Niu, C., Li, M., Fan, F., Wu, W., Guo, X., Lyu, Q., Wang, G.: Noise suppression with similarity-based self-supervised deep learning. IEEE Trans. on Medical Imaging (2022)
24. Park, T., Efros, A.A., Zhang, R., Zhu, J.Y.: Contrastive learning for unpaired image-to-image translation. In: Proc. of European Conf. on Computer Vision. pp. 319–345 (2020)

25. Smith-Bindman, R., Lipson, J., Marcus, R., Kim, K.P., Mahesh, M., Gould, R., De González, A.B., Miglioretti, D.L.: Radiation dose associated with common computed tomography examinations and the associated lifetime attributable risk of cancer. Archives of internal medicine **169**(22), 2078–2086 (2009)
26. Ulyanov, D., Vedaldi, A., Lempitsky, V.: Deep image prior. In: Proc. of IEEE Conf. on Computer Vision and Pattern Recognition. pp. 9446–9454 (2018)
27. Wang, D., Fan, F., Wu, Z., Liu, R., Wang, F., Yu, H.: CTformer: convolution-free Token2Token dilated vision transformer for low-dose CT denoising. Physics in Medicine & Biology **68**(6), 065012 (2023)
28. Wang, Z., Liu, J., Li, G., Han, H.: Blind2unblind: Self-supervised image denoising with visible blind spots. In: Proc. of IEEE Conf. on Computer Vision and Pattern Recognition. pp. 2027–2036 (2022)
29. Yao, T., Pan, Y., Li, Y., Ngo, C.W., Mei, T.: Wave-vit: Unifying wavelet and transformers for visual representation learning. In: Proc. of European Conf. on Computer Vision. pp. 328–345 (2022)
30. Yun, S., Lee, H., Kim, J., Shin, J.: Patch-level representation learning for self-supervised vision transformers. In: Proc. of IEEE Conf. on Computer Vision and Pattern Recognition. pp. 8354–8363 (2022)
31. Zhao, H., Gallo, O., Frosio, I., Kautz, J.: Loss functions for image restoration with neural networks. IEEE Transactions on computational imaging **3**(1), 47–57 (2016)
32. Zhu, J.Y., Park, T., Isola, P., Efros, A.A.: Unpaired image-to-image translation using cycle-consistent adversarial networks. In: Proc. of IEEE Intl. Conf. on Computer Vision. pp. 2223–2232 (2017)