



Full Image-Index Remainder Based Single Low-Dose DR/CT Self-supervised Denoising

Yifei Long¹, Jiayi Pan¹, Yan Xi², Jianjia Zhang^{1(✉)}, and Weiwen Wu^{1(✉)}

¹ School of Biomedical Engineering, Shenzhen Campus of Sun Yat-sen University, Shenzhen 518107, China

{zhangjj225,wuweiw7}@mail.sysu.edu.cn

² Shanghai First-Imaging Information Technology Co., LTD., Shanghai, China

Abstract. Low-dose digital radiography (DR) and computed tomography (CT) play a crucial role in minimizing health risks during clinical examinations and diagnoses. However, reducing the radiation dose often leads to lower signal-to-noise ratio measurements, resulting in degraded image quality. Existing supervised and self-supervised reconstruction techniques have been developed with noisy and clean image pairs or noisy and noisy image pairs, implying they cannot be adapted to single DR and CT image denoising. In this study, we introduce the Full Image-Index Remainder (FIRE) method. Our method begins by dividing the entire high-dimensional image space into multiple low-dimensional sub-image spaces using a full image-index remainder technique. By leveraging the data redundancy present within these sub-image spaces, we identify similar groups of noisy sub-images for training a self-supervised denoising network. Additionally, we establish a sub-space sampling theory specifically designed for self-supervised denoising networks. Finally, we propose a novel regularization optimization function that effectively reduces the disparity between self-supervised and supervised denoising networks, thereby enhancing denoising training. Through comprehensive quantitative and qualitative experiments conducted on both clinical low-dose CT and DR datasets, we demonstrate the remarkable effectiveness and advantages of our FIRE method compared to other state-of-the-art approaches.

Keywords: Digital radiography · Computed tomography · Self-supervised · Remainder · Image denoising

1 Introduction

Digital Radiography (DR) and Computed Tomography (CT) techniques are extensively utilized in the diagnosis of various clinical conditions [15, 24]. However, one major concern associated with these imaging methods is the exposure of

Supplementary Information The online version contains supplementary material available at https://doi.org/10.1007/978-3-031-43990-2_44.

patients to X-ray radiation [12]. Reducing the X-ray radiation dose unavoidably leads to a decline in the number of photons detected, resulting in measurements with a low signal-to-noise ratio and consequent regression in image quality. Consequently, accurately diagnosing clinical conditions based on degraded DR/CT images poses significant challenges. Hence, the development of sophisticated and efficient image-denoising techniques that can effectively address both DR and CT modalities becomes imperative and urgent in clinical applications.

Traditional image denoising methods have relied on exploring spatial pixel features and properties in the transform domain [2, 5, 8]. These methods include approaches like non-local mean (NLM) [23] and BM3D [5]. However, the need for complex parameter adjustment and their relatively slow speed have limited the practical applications of these traditional denoising methods.

With the advancement of neural networks, deep learning-based denoising techniques have shown superior performance compared to traditional methods [1, 3, 9]. Supervised denoising methods, such as U-Net [6], DnCNN [25], and RED-CNN [4], have demonstrated promising results. However, when it comes to digital radiography (DR) and computed tomography (CT), these supervised methods face challenges since they rely on paired noisy and clean image data, which are not readily available in DR and CT imaging [7].

To overcome the limitations of supervised techniques, self-supervised or unsupervised denoising methods have been proposed and developed, leveraging the similarity between noisy images [18, 26]. Several unsupervised learning-based methods have been introduced for image denoising, including Noise2Noise [10, 13, 16], Noise2Sim [20], Noise2Void [14], and Neighbor2Neighbor [11] among others [17]. However, these methods have defects in the DR and CT denoising tasks. They either require noisy and noisy image pairs in training, or sampling strategies resulting in information missing and the same noise level of the sub-image pairs.

Therefore, we designed a new self-supervised image denoising technique, i.e., the full image-index remainder (FIRE), adapting to low-dose DR and CT to address the challenge. Specifically, our FIRE method first divided the whole high-dimensional image space into a series of low-dimensional sub-image spaces with the image-index remainder technique. Based on the remainder of the full image index, a specific image sampler is designed to sample the sub-images. With this strategy, a set of sampled noisy sub-images from a single noisy image is obtained for self-supervised training of the denoising neural network without clean images. We further proved that our proposed sampling strategy is effective in supplementary materials. In addition, we proposed a new loss function to train the unsupervised image-denoising network with dedicated parameter tuning. For further optimizing the feasible domain, a regularization strategy is introduced to reduce the gap between the self-supervised and the supervised denoising network. We evaluated the FIRE on several large-scale clinical DR and CT datasets without clean data. The experimental results show that our FIRE method achieves the best results in noise suppression and visual perception. We also compared our FIRE and other methods using public clinical data with ground truth, and the quantitative and qualitative results demonstrate the out-performance of our method.

The remaining sections are organized as follows. Sections 2 and 3, introduce our proposed FIRE method in terms of the network architectures, sampler design, and loss functions. In Sect. 4, we will test our model on large-scale low-dose DR and CT image datasets and compare it with other well-performing methods. Finally, a summary will be given.

2 Methodology

2.1 Related Theories

Given a clean image x and its corresponding noisy image y , the supervised image denoising methods try to train the denoising network f parameterized by θ minimizing the loss function below:

$$E_{\{x,y\}} = \{f_\theta(y), x\}. \quad (1)$$

Implementing the supervised denoising network on noisy and clean image pairs, the usual loss function can be formulated as:

$$\theta = \arg \min_\theta E_{x,y} \|f_\theta(y) - x\|_2^2. \quad (2)$$

However, in actual situations, it is difficult to obtain the corresponding clean and noisy image pairs, especially in medical imaging scans. To address this issue, Noise2Noise proposed a self-denoising method that does not require real clean images, but it depends on pairs of independent noisy images of the same scene. The method proves that the results obtained by minimizing the following equation are the same as the results obtained by the supervised case above:

$$\theta' = \arg \min_{\theta'} E_{x,y,z} \|f'_\theta(y) - z\|_2^2 \quad (3)$$

where y and z are two independent noisy images conditioned on x . In previous studies [16], it has been proved that the results of training using only noisy and noisy image pairs can be approximately equal to supervised cases. Neighbor2Neighbor [11] focuses on sampling a single noisy image with independent noisy and noisy image pairs from the same scene for network training. However, there are still challenges in applying this method to DR and CT image denoising tasks. Specifically, DR and CT images have a larger range of pixel values and more details in the image, while the current method uses only the part image information in each iteration. In the following section, we propose our FIRE method for DR and CT image denoising tasks.

2.2 Framework

In this section, we propose a self-supervised framework for training a single noisy image-denoising network. First, we design a new subspace sampling technique for generating subspace image groups to train the network. Next, for recovering finer image details and features, a specialized loss function consisting of reconstruction terms and regularization terms is proposed and used for training the network. The overall of our FIRE framework is shown in Fig. 1(a) and (b).

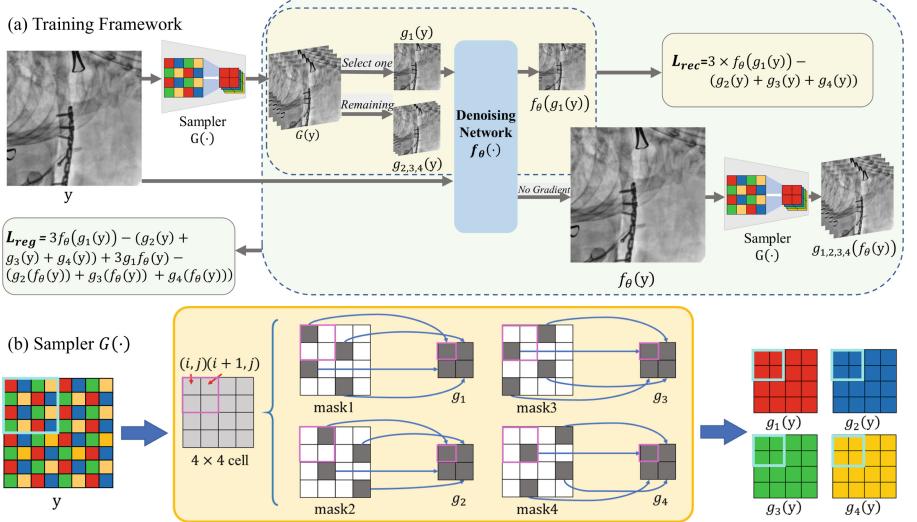


Fig. 1. (a) An overview of our proposed FIRE framework. The regularization loss L is calculated as following two stages. First, the subsampled image is calculated as the image reconstruction term L_{rec} between the network output and the target. Second, we added the regularization term L_{reg} to correct the difference of implying ground truth between the subsampled noisy image. (b) Details of the remainder sampler based on pixel coordinates. The different color boxes, including blue and pink, respectively represent the transformation from one part of the image to this part after sampling.

Subspace Sampling: The remainder with respect to a positive integer (i.e., N) of image pixel coordination is excellent to be used to develop a sampler. With it, a raw noisy image can be divided into a series of sub-image spaces, and such design can fully use the advantages of the image pixels index resulting in satisfying the condition of independence. Let $y \in R^{(W \times H)}$ be one noisy image sample, where W and H represent the width and height of the raw image y . Considering the size of the medical image is usually a multiple of 256, the N is chosen as 4 in this study. In fact, the N can be adjusted depending on the image size. In this case, the sampler G consists of four sub-sampler $G = (g_1, g_2, g_3, g_4)$. The details of the remainder sampler G can be developed as follows:

- Making the sampled pictures contain all the pixels within the original noisy image, it is necessary to ensure that the length and width of the image y can be divisible by N , which is chosen as 4;
- Encoding pixel index of the image y from the left-upper corner, all image pixels can be accessed by the abscissa i and ordinate j . The image indexes can range from $(0, 0)$ to $(W - 1, H - 1)$.
- Calculating the remainder of $(2 \times i + j) \% 4$ and defined it as k . k is an integer in $[0, 3]$. All pixel coordinations satisfying $k = 0$ are retained, and the remaining pixel values are set to 0 to generate the first mask. The second mask is

obtained with pixel coordinations satisfying $k = 1$. The third and fourth masks are obtained by satisfying $k = 2$ and $k = 3$.

- d) Selecting the reserved pixel value from the four pixels in each 2×2 area as the pixel value for the corresponding position of the subsampled image, and put it into $(g_1, g_2, g_3$ and $g_4)$. By doing this, four sampled pictures are obtained, i.e., $(g_1(y), g_2(y), g_3(y)$ and $g_4(y))$. The width and length of all sampled subspace images are $W/2$ and $H/2$ respectively.

To clearly demonstrate the generation diagram of the subspace image group, Fig. 1(b) summarizes the idea of our proposed remainder subspace subsampler. Since the corresponding pixels of $(g_1(y), g_2(y), g_3(y), g_4(y))$ generated by the sampler are adjacent but different in the original image. Each sampled subspace image is independent of the others, and the completeness and difference of image content in the four sampled sub-images can be guaranteed. Here we can divide the four sub-images into two groups of $g_1(y)$ and $(g_2(y), g_3(y), g_4(y))$. Please refer to the supplementary materials for detailed explanations of this grouping.

Regularization Optimization: For self-supervised training, the following minimization optimization problem is formulated by taking advantage of the constraint

$$\min_{\theta} E_{y|x} \|3f_{\theta}(g_1(y)) - (g_2(y) + g_3(y) + g_4(y))\|_2^2 + \lambda E_{y|x} \|3f_{\theta}(g_1(y)) - (g_2(y) + g_3(y) + g_4(y)) + (3g_1(f_{\theta}(y)) - (g_2(f_{\theta}(y)) + g_3(f_{\theta}(y)) + g_4(f_{\theta}(y)))\|_2^2. \quad (4)$$

According to $E_{x,y} = E_x E_{y|x}$, it can be converted into the following regularization optimization problem

$$\min_{\theta} E_{x,y} \|3f_{\theta}(g_1(y)) - (g_2(y) + g_3(y) + g_4(y))\|_2^2 + \lambda E_{x,y} \|3f_{\theta}(g_1(y)) - (g_2(y) + g_3(y) + g_4(y)) + (3g_1(f_{\theta}(y)) - (g_2(f_{\theta}(y)) + g_3(f_{\theta}(y)) + g_4(f_{\theta}(y)))\|_2^2. \quad (5)$$

Finally, the loss function incorporating the regularization term is proposed to train the denoising network with

$$\begin{aligned} L &= L_{rec} + \lambda \times L_{reg} \\ &= \|3f_{\theta}(g_1(y)) - (g_2(y) + g_3(y) + g_4(y))\|_2^2 + \lambda \|3f_{\theta}(g_1(y)) - (g_2(y) + g_3(y) + g_4(y)) + (3g_1(f_{\theta}(y)) - (g_2(f_{\theta}(y)) + g_3(f_{\theta}(y)) + g_4(f_{\theta}(y)))\|_2^2 \end{aligned} \quad (6)$$

where f_{θ} is the denoising network and λ is a hyperparameter to balance the regularization term.

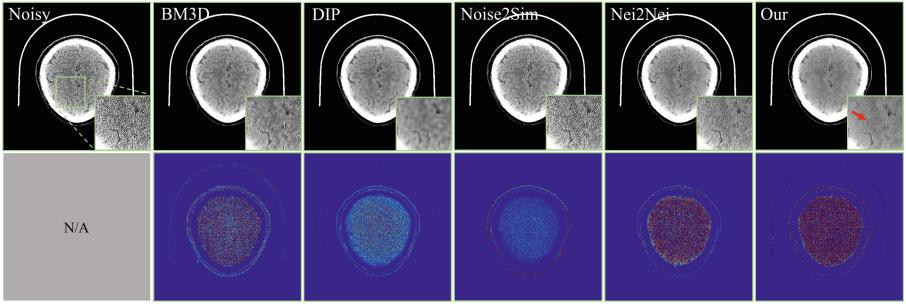


Fig. 2. Visualization comparison of different methods in the clinical brain CT results. The visualization of the results after denoising by different methods is shown above, and the noise gap level between the output image and the original noisy image is shown below.

3 Experiments

In this section, we first introduce the details and configuration of clinical experiments. To evaluate the effectiveness of this method, several advanced image-denoising methods for CT and DR were involved in the comparison. The experiments were conducted on several large-scale CT and DR data sets.

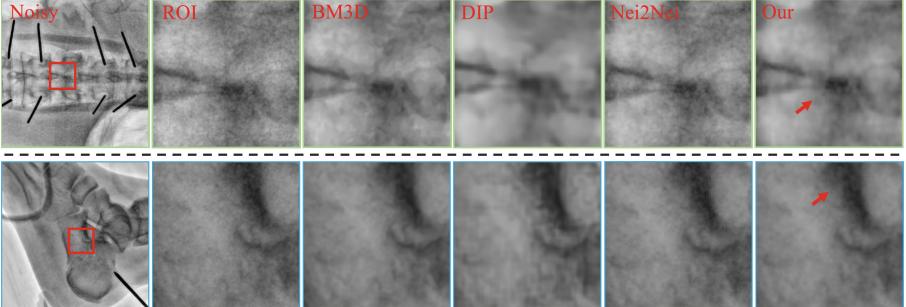


Fig. 3. Visual comparison of different methods in real DR denoising results. The boxes of different colors represent images of different parts. The green image is represented as the spine and the blue as the hand. (Color figure online)

Experimental Configuration: The network architecture of our proposed FIRE framework was a modified U-Net [21]. All experiments were conducted on a PC server equipped with Python3.9.7, PyTorch1.8, and NVIDIA TITAN RTX graphics processors.

In terms of the comparisons, we include BM3D [5], DIP [17], Noise2Sim [20], Noise2Void [14], Blind2Unblind [22], Neighbor2Neighbor(Nei2Nei) [11] and two

supervised denoising algorithms DnCNN [25] and DD-net [27]. According to the applicability of denoising methods, we chose different comparison methods in different experiments. More details can be found in Figs. 2 to 4.

The clinical DR and brain CT image data are collected from the hospital. The public CT image data came from the 2016 NIH-AAPM Mayo Clinical low-dose CT competition [19] (Moen, Chen, Holmes III, Duan, Yu, Yu, Leng, Fletcher & McCollough 2021). In the three experiments of brain CT, clinical DR, and public clinical CT, the image size was 512×512 pixels. The initial learning rate of the ADAM optimizer was set as 0.0003, 0.0003, and 0.0004 respectively. The batch size was set as 1, 16, and 1 respectively. The epoch was set as 40, 200, and 10 respectively to ensure the model was convergent. The regularization term parameter λ was set to $\lambda = 2, 10$, and 1 respectively.

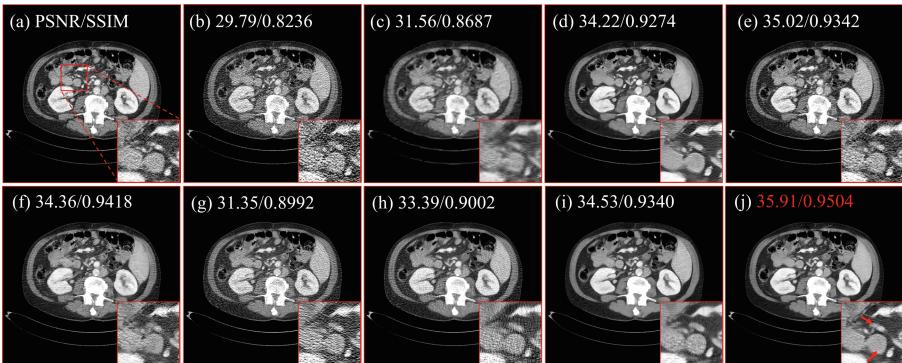


Fig. 4. Visual comparison of different methods on the public CT dataset. (a) is the reference obtained with the high dose, (b) is the noisy image, (c)-(j) are results of BM3D, DnCNN, DDnet, Noise2Sim, Noise2Void, Blind2Unblind, Neighbor2Neighbor, and our proposed FIRE. PSNR and SSIM values are marked at the top of the picture.

Experimental Results: In three different denoising tasks, our FIRE method showed the best results. For the clinical brain CT image denoising, the excellent denoising results of FIRE can be seen in Fig. 2. Due to a certain gap in ground truth between adjacent frames images used in network training, Noise2Sim cannot achieve good enough results. DIP produces a lot of artifacts after removing the noise and misses a lot of details and features. Besides, BM3D and DIP consume long time, and the parameters need to be manually adjusted on different denoising conditions. For the real DR image dataset, Fig. 3 demonstrates the DR image denoising results of different body parts for different patients. It can be seen that the denoising performance of Neighbor2Neighbor is limited. It may be because the Neighbor2Neighbor only uses information of half pixels for training, and the entire structure and pixel information within the DR image is partially

missed. On the public CT image data set, we can intuitively compare the advantages of our method over other methods from the PSNR and SSIM measures. Figure 4 shows the typical slice denoising results. Figure 5(a)-(j) presents the noise power spectrum(NPS) of all methods, where blue indicates it is closer to the reference. As seen, our FIRE obtains the most blue and the least red results. Figure 5(k) reflects the comparison of pixel values on the profile line of a slice of the test set. Among them, our FIRE is closest to the reference.

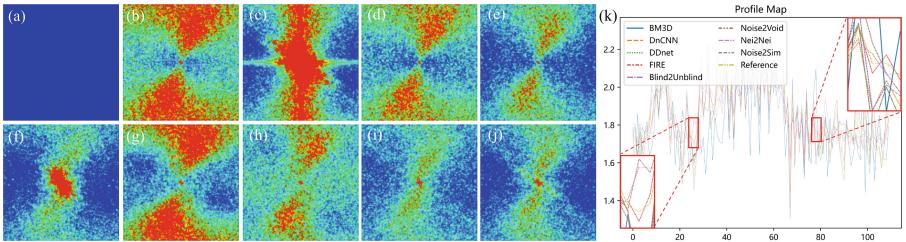


Fig. 5. (a)-(j) correspond to the noise power spectrum of Fig. 4. (k) is the profile map of different methods.

Influence of Regularization Term: We proposed a regularization term in Sect. 3, and the hyperparameter λ is used to adjust the regularization term. Figure 6 shows the visual results of FIRE with different λ values. When $\lambda = 0$, the regularization term is removed. When $\lambda = 1$, the denoising effect of the network is the best. As λ increases, the residual degree of noise increases. This shows that the regularization term can adjust the smoothness and noisiness of denoising. Therefore, choosing an appropriate λ value helps to obtain better denoising results.



Fig. 6. The ablation study of the influence of regularization terms. The following results are the results of different λ values. (a) $\lambda = 0$, (b) $\lambda = 1$, (c) $\lambda = 8$, and (d) $\lambda = 16$.

4 Conclusion

We proposed a full image-index remainder method (FIRE) using only a single noisy image. The proposed FIRE first divided the whole high-dimensional image space into a series of low-dimensional sub-image spaces with the full image-index remainder technique. Our FIRE retains the complete information within the original noisy image. In addition, we proposed a new regularization optimization function to regularize sub-space image training by reducing the gap between the self-supervised and supervised denoising networks. Quantitative and qualitative experiment results indicate that the proposed FIRE is effective in both DR and CT image denoising.

Acknowledgements. This work was supported in part by National Natural Science Foundation of China (grant numbers 62101611 and 62201628), National Key Research and Development Program of China (2022YFA1204200), Guangdong Basic and Applied Basic Research Foundation (grant number 2022A1515011375, 2023A1515012278, 2023A1515011780) and Shenzhen Science and Technology Program (grant number JCYJ20220530145411027, JCYJ20220818102414031).

References

1. Anwar, S., Barnes, N.: Real image denoising with feature attention. In: 2019 IEEE/CVF International Conference on Computer Vision (ICCV), pp. 3155–3164 (2019)
2. Buades, A., Coll, B., Morel, J.M.: A non-local algorithm for image denoising. In: 2005 IEEE Computer Society Conference on Computer Vision and Pattern Recognition (CVPR'05), vol. 2, pp. 60–65 (2005)
3. Chang, M., Li, Q., Feng, H., Xu, Z.: Spatial-adaptive network for single image denoising (2020)
4. Chen, H., et al.: Low-dose CT with a residual encoder-decoder convolutional neural network. *IEEE Trans. Med. Imaging* **36**(12), 2524–2535 (2017)
5. Dabov, K., Foi, A., Katkovnik, V., Egiazarian, K.: Image denoising by sparse 3-D transform-domain collaborative filtering. *IEEE Trans. Image Process.* **16**(8), 2080–2095 (2007)
6. Falk, T., et al.: U-Net: deep learning for cell counting, detection, and morphometry. *Nat. Methods* **16**, 67–70 (2019)
7. Freedman, M.T., Artz, D.S.: Image processing in digital radiography. *Semin. Roentgenol.* **32**(1), 25–37 (1997), digital Radiography Using Storage Phosphor Technology
8. Gu, S., Zhang, L., Zuo, W., Feng, X.: Weighted nuclear norm minimization with application to image denoising, pp. 2862–2869 (2014)
9. Guo, S., Yan, Z., Zhang, K., Zuo, W., Zhang, L.: Toward convolutional blind denoising of real photographs. In: 2019 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), pp. 1712–1722 (2019)
10. Hasan, A.M., Mohebbian, M.R., Wahid, K.A., Babyn, P.: Hybrid-collaborative Noise2Noise denoiser for low-dose CT images. *IEEE Trans. Radiat. Plasma Med. Sci.* **5**(2), 235–244 (2021)

11. Huang, T., Li, S., Jia, X., Lu, H., Liu, J.: Neighbor2Neighbor: a self-supervised framework for deep image denoising. *IEEE Trans. Image Process.* **31**, 4023–4038 (2022)
12. Immonen, E., et al.: The use of deep learning towards dose optimization in low-dose computed tomography: a scoping review. *Radiography* **28**(1), 208–214 (2022)
13. Kashyap, M., Tambwekar, A., Manohara, K., Subramanyam, N.: Speech denoising without clean training data: a Noise2Noise approach (2021)
14. Krull, A., Buchholz, T.O., Jug, F.: Noise2Void - learning denoising from single noisy images. In: 2019 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), pp. 2124–2132 (2019)
15. Krupinski, E.A., et al.: Digital radiography image quality: image processing and display. *J. Am. Coll. Radiol.* **4**(6), 389–400 (2007)
16. Lehtinen, J., et al.: Noise2Noise: learning image restoration without clean data (2018)
17. Lempitsky, V., Vedaldi, A., Ulyanov, D.: Deep image prior. In: 2018 IEEE/CVF Conference on Computer Vision and Pattern Recognition, pp. 9446–9454 (2018)
18. Li, X., Fan, C., Zhao, C., Zou, L., Tian, S.: NIRN: self-supervised noisy image reconstruction network for real-world image denoising. *Appl. Intell.* **52**, 1–18 (2022)
19. McCollough, C., et al.: Low dose CT image and projection data, LDCT and projection data, version 5, data set, the cancer imaging archive (2020). <https://doi.org/10.7937/9NPB-2637>
20. Niu, C., et al.: Suppression of correlated noise with similarity-based unsupervised deep learning (2020)
21. Ronneberger, O., Fischer, P., Brox, T.: U-Net: convolutional networks for biomedical image segmentation (2015)
22. Wang, Z., Liu, J., Li, G., Han, H.: Blind2Unblind: self-supervised image denoising with visible blind spots. In: 2022 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), pp. 2017–2026 (2022)
23. Wiest-Daesslé, N., Prima, S., Coupé, P., Morrissey, S.P., Barillot, C.: Non-local means variants for denoising of diffusion-weighted and diffusion tensor MRI. In: Ayache, N., Ourselin, S., Maeder, A. (eds.) Medical Image Computing and Computer-Assisted Intervention - MICCAI 2007, pp. 344–351. Springer, Berlin, Heidelberg (2007). https://doi.org/10.1007/978-3-540-75759-7_42
24. Williams, M.B., et al.: Digital radiography image quality: image acquisition. *J. Am. Coll. Radiol.* **4**(6), 371–388 (2007)
25. Yin, Z., Wan, B., Yuan, F., Xia, X., Shi, J.: A deep normalization and convolutional neural network for image smoke detection. *IEEE Access* **5**, 18429–18438 (2017)
26. Zainulina, E., Chernyavskiy, A., Dylov, D.V.: Self-supervised physics-based denoising for computed tomography. ArXiv abs/2211.00745 (2022)
27. Zhang, Z., Liang, X., Dong, X., Xie, Y., Cao, G.: A sparse-view CT reconstruction method based on combination of DenseNet and deconvolution. *IEEE Trans. Med. Imaging* **37**(6), 1407–1417 (2018)