VisNet: A Human Visual System Inspired Lightweight Dual-Path Network for Medical Images Denoising

Hailin Yue[1], Hulin Kuang[1⋆], Lei Ma[1], Jin Liu[1,2], Junjian Li[1], Jianhong Cheng[3], and Jianxin Wang[1,2]

[1] Hunan Provincial Key Lab on Bioinformatics, School of Computer Science and Engineering, Central South University, Changsha, China
hulinkuang@csu.edu.cn
[2] Xinjiang Engineering Research Center of Big Data and Intelligent Software, School of Software, Xinjiang University, Urumqi, China
[3] Institute of Guizhou Aerospace Measuring and Testing Technology, Guiyang, China

Abstract. Efficiently and accurately removing noise from medical images is crucial for clinical diagnosis. Nevertheless, most deep learning-based medical images denoising methods are highly complex and inaccurate in preserving the edge and shape of different organs, resulting in suboptimal denoising performance. In our study, we propose a Human **Vis**ual System Inspired Lightweight Dual-Path **Net**work for medical images denoising (**VisNet**), which can efficiently and accurately remove noise from different types of medical images. Specifically, to simulate the mechanism in the visual system where magnocellular and parvocellular pathways capture significant and subtle noise, respectively, we design a dual-path multi-scale perception module. Then, to simulate the function of the primary visual cortex, we propose an edge detection and shape adaptation module to preserve the structural information of the medical images. Finally, inspired by dorsal and ventral pathways, a spatial-semantic information extraction module is designed to enhance the main semantic information in the image through the interactive fusion between the spatial and semantic pathways. Experimental results demonstrate that VisNet achieves superior performance across three medical datasets compared to nine existing baselines, while maintaining minimal computational complexity (Params=0.15, FLOPs=16.41). In addition, for brain tumor classification, using denoised images of VisNet as input significantly improves accuracy (87.5% vs 96.7%) and achieves performance comparable to noise-free images. Code of VisNet is available at https://github.com/yuehailin/VisNet.

Keywords: Medical images denoising · Visual system inspired · Lightweight.

⋆ Corresponding author

1 Introduction

Nowadays, medical images have become an indispensable means in clinical disease diagnosis. Common medical images include Computed Tomography (CT) [4], X-ray Radiation (X-ray) [1], and Magnetic Resonance Imaging (MRI) [20]. These medical images can provide a visual display of a patient's internal structures, aiding doctors in making more accurate diagnoses and treatment decisions [31,2]. However, various factors may introduce noise, including artifacts, spots, and peaks. Such noise can compromise image clarity and quality, potentially leading to incorrect diagnoses or treatment decisions, which in turn may pose unnecessary risks and harm to patients [16]. Therefore, it is necessary to design a method that can efficiently remove noise and improve the quality of medical images.

Existing medical images denoising methods can be broadly classified into filtering-based and deep learning-based methods. Filtering methods, such as median, average, and Gaussian filters [5,23,17], effectively reduce noise but often lead to the blurring of image details and structure edges, resulting in distortion. With the advancement of deep learning, convolutional neural networks (CNNs) have been applied to image denoising, yielding significant results [14,11]. Subsequently, generative adversarial networks (GANs) have also emerged as a powerful tool in this domain [29,9,21]. As attention mechanisms have progressed, transformer models have become an important tool for image denoising [19,13,30,27]. More recently, with the rise of selective state space, Mamba-based image denoising techniques have made notable progress [12]. However, most methods are typically tailored to specific types of medical images, limiting their adaptability to organs with varying shapes, and leading to suboptimal performance. The human visual system shows powerful visual perception capabilities [22] (as shown in Fig. 1 (A)), and simulating its process of processing visual signals may be able to more effectively filter noise [28]. Moreover, these methods are usually complex and require significant computing resources, making it challenging to achieve a double breakthrough in denoising efficiency and denoising performance (Fig. 1 (B)). Thus, there is a growing need for lightweight, adaptable methods that can efficiently denoise diverse types of medical images.

Based on these considerations, we propose a Human **Vis**ual System Inspired Lightweight Dual-path **Net**work (**VisNet**) for medical images denoising. As we can see from Fig. 1 (A), to simulate the functions of the magnocellular and parvocellular pathways in the human visual system, we design a dual-path multi-scale perception module that effectively captures multi-scale noise in different types of medical images and suppresses this noise through two distinct scale pathways. Then, inspired by visual signal processing in the primary visual cortex (V1), we propose an edge detection and shape adaptation module that uses dense blocks and deformable convolutions to extract the edges and shapes of organs in different types of medical images. Finally, we design a spatial-semantic information extraction module to simulate the processes of the dorsal and ventral pathways in processing spatial and semantic information, respectively.

The contributions of this work can be summarized as follows: 1) We take the human visual system as the prototype and propose a lightweight medical images

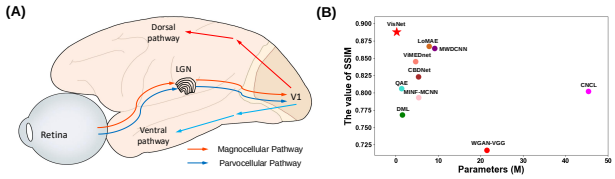

Fig. 1. (A): The visual pathway of the human visual system. (B): Comparison of the SSIM and Parameters for different denoising methods.

denoising method, which achieves superior denoising performance across multiple datasets, while maintaining minimal computational complexity. 2) Inspired by the visual system's magnocellular and parvocellular pathways, primary visual cortex, and dorsal and ventral pathways, we design a dual-path multi-scale perception module, an edge detection and shape adaptation module, and a spatial semantic information extraction module, respectively. 3) Using denoised images as input, the performance in common brain tumor classification is significantly higher than that with noisy images, and closely the performance with noise-free images, demonstrates its potential in clinical diagnosis.

2 Methods

2.1 Architecture Overview

In our study, we design a lightweight network inspired by a visual system for medical images denoising. As shown in Fig. 2, the dual-path multi-scale perception module can better perceive organs of various sizes in multiple types of medical images through large and small pathways, enabling more accurate modeling of noise at different scales. The edge detection and shape adaptation module enhances the model's ability to preserve the edge features of organs by emphasizing image edge structures and incorporating deformable convolution, which facilitates better adaptation to the diverse shapes of organs, thereby improving the model's capacity to effectively process complex organ structures. The spatial-semantic information extraction module includes spatial and semantic pathways that capture respective spatial and semantic information. The interactions between these two pathways enrich features across different depths, thereby enhancing the model's understanding of spatial-semantic information.

2.2 Dual-path Multi-scale Perception Module

In the human early visual pathways, the magnocellular pathway (orange line in Fig. 1 (A)) captures large-scale information with larger cell bodies and receptive fields, while the parvocellular pathway (blue line in Fig. 1 (A)) focuses on image details with smaller cell bodies and receptive fields. Inspired by these pathways, we propose a dual-path multi-scale perception module to capture both significant and subtle visual signals at different scales.

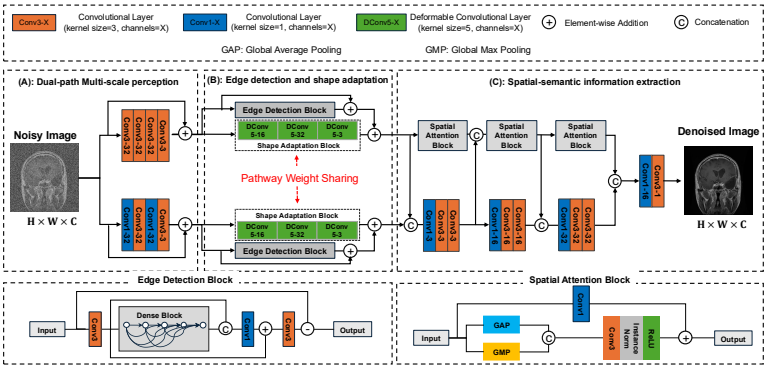

Fig. 2. Overview of Human Visual System Inspired Lightweight Dual-Path Network.

As shown in Fig. 2 (A), the dual-path multi-scale perception module (DMP) is consistent with the early visual pathways and also contains two pathways. The large-scale pathway consists of four 3×3 convolutions to obtain a large receptive field. By replacing a large convolution with multiple smaller ones, this approach not only reduces the number of parameters but also enhances the model's ability to fit non-linearities [24], thereby improving its capacity to capture global image features. In contrast, the small-scale pathway adopts an alternating concatenation strategy, combining two 1×1 convolutions and two 3×3 convolutions. Among them, the role of 1×1 convolution is used to control the number of channels, thereby reducing computational complexity while retaining information, while 3×3 convolution can effectively capture smaller-scale detailed information in the image. Through the parallel processing of large and small-scale pathways, the module effectively removes multi-scale noise in a targeted manner.

2.3 Edge Detection and Shape Adaptation Module

As shown in Fig. 1 (A), the magnocellular and parvocellular pathways in the human visual system transmit visual signals of varying scales to V1 for processing, including the extraction of features such as edges and shapes. Inspired by this, we propose an edge detection and shape adaptation module, following the dual-path multi-scale perception module (as shown in Fig. 2 (B)), to extract edge and shape features from visual signals at multiple scales.

Edge Detection Block. The edge detection (ED) module consists of three convolutional layers and a dense block. Among them, the role of the convolutional layer is to adjust the spatial dimension. The role of the dense block is to capture image features. To maintain a balance between complexity and feature extraction capability, we use 3 layers of convolutions within the dense block for dense connections, followed by a 1×1 convolution. We obtain edge information through a subtraction operation. The p-th layer in the dense block receives the feature maps from all previous layers as input and produces the output X_p as follows:

$$X_p = H_p([X_0, X_1, ..., X_{p-1}]), \tag{1}$$

where $[X_0, X_1, ..., X_{p-1}]$ represents the concatenation of the feature maps generated in the $0, 1, ..., p-1$ layer. At the same time, H_p represents the current convolutional layer, and X_p represents the output of p-th layer.

Shape Adaptation Block. In medical images, noise can present irregular features on organs of different shapes. Traditional convolutions with fixed-size kernels may not be able to effectively adapt to the changes in the shape of different types of medical images [3]. Therefore, this section adopts the shape adaptation (SA) block, which consists of deformable convolution. Specifically, the shape adaptation block consists of three layers of deformable convolutions, with feature map channels of 16, 32, and 3, respectively. Each deformable convolution layer adaptively adjusts the kernel's shape and position based on local input features, enabling better adaptation to organs of different shapes. Stacking multiple layers allows the model to progressively learn more abstract feature representations, enhancing the performance in image denoising task. The formula of deformable convolution can be expressed as follows:

$$Y(p) = \sum_{i=1}^{H \times K} W(i) \cdot X(p + \Delta(p, i)) + b, \tag{2}$$

where p represents the position in the output feature map, $H \times K$ is the size of the convolution kernel, $W(i)$ is the weight of the convolution kernel, $X(p+\Delta(p,i))+b$ represents the value of the input feature map at position $p + \Delta(p, i)$, Y is the output feature map, b is the bias.

2.4 Spatial-Semantic Extraction Module

In the human visual system, the V1 module transmits visual signals to the dorsal pathway and ventral pathway (as shown in Fig. 1 (A)), where they process spatial information (such as location, motion, and depth) and semantic information (such as object recognition and categorization), respectively. Inspired by this, we propose a spatial-semantic extraction module.

As shown in Fig. 2 (C), the Spatial-Semantic Extraction (SSE) module consists of two pathways. The spatial pathway contains three spatial attention blocks. Each spatial attention block first processes the input through global average pooling (GAP) and global max pooling (GMP) to obtain a feature map. The feature map is then pass through a 3×3 convolution operation followed by InstanceNorm, adding residual pixels to the input feature map. The semantic pathway comprises three multi-channel stacked convolutions. As the model deepens, the number of channels in three stacked convolutions increases to 3, 16, and 32, respectively. Additionally, there are three channel interactions between the spatial and semantic pathways to facilitate feature complementarity. The data flow in the second spatial attention block and the third stacked convolution in Fig. 2 (C) is summarized as follows:

$$X_{d2} = \text{SpatialAttention}\left([X_{d1}, X_{v1}]\right) \tag{3}$$

$$X_{v3} = C^{k=3}\left(C^{k=3}\left(C^{k=1}\left([X_{v2}, X_{d2}]\right)\right)\right) \tag{4}$$

where X_{d1} and X_{d2} are the outputs of the first and second spatial attention blocks. X_{v1}, X_{v2} and X_{v3} are the outputs of three stacked convolutions respectively. $C^{k=n}$ is the convolution with kernel size of n.

3 Experiments

3.1 Experimental Settings

Datasets. We evaluate VisNet on three types of public medical datasets. **CT**: The 2016 NIH-AAPM-Mayo Clinic Low Dose CT Grand Challenge includes 2378 CT images [27]. We use 1709 images for training and 669 images for testing. **X-ray**: The Japanese Society of Radiological Technology (JSRT) contains 247 chest X-ray images [18]. We use 197 images for training and 50 images for testing. **MRI**: We use a public Kaggle dataset of brain tumor MRI images [26], adding Gaussian and Poisson noise with levels of 15 and 50 ($\sigma = 15$ and $\sigma = 50$). The training set contains 5712 images, and the test set contains 1311 images.
Implementation Details. We use mean square error (MSE) as the optimization objective with Adam as the optimizer and a learning rate of 0.001. The results of all comparative experiments are reproduced based on the original settings and presented in terms of mean and standard deviation.
Evaluation Metric. In our study, we use Peak Signal-to-Noise Ratio (PSNR) and Structural Similarity (SSIM) to evaluate the performance of different methods. PSNR is a widely used metric for evaluating image quality, particularly in assessing the performance of image denoising. SSIM is used to evaluate the similarity between a denoised image and its corresponding ground truth. SSIM takes into account factors such as luminance, contrast, and structural information.

3.2 Compared with State-of-the-Art Methods

Results on CT and X-ray Denoising. Table 1 shows the quantitative denoising results for various methods on the CT and X-ray datasets. VisNet outperforms all other methods in both CT and X-ray denoising tasks, achieving the highest PSNR and SSIM values, demonstrating its superior denoising capabilities across these modalities.
Results on MRI Denoising. Table 2 displays the results for the MRI denoising task at different noise levels ($\sigma = 15$ and $\sigma = 50$). VisNet again shows the best performance in terms of both PSNR and SSIM, surpassing other methods at both noise levels. These results validate the superiority of VisNet in addressing MRI denoising with varying levels of noise.
Visualization. In this section, we use X-ray as an example to visualize the denoising performance of different methods. As shown in Fig. 3, compared to other methods, the denoised image produced by VisNet is visually closest to the ground truth, which visually highlights the superiority of VisNet.

Computational Complexity. Despite achieving superior denoising performance across CT, X-ray, and MRI datasets, VisNet also exhibits the lowest computational complexity against all compared methods. Specifically, as shown in Tables 1 and 2, the parameters (Params) of VisNet is 0.15 and the floating point operations (FLOPs) of VisNet are 16.41, both of which are the lowest among all the methods, making it an efficient and practical method for medical images denoising tasks.

3.3 Ablation Study

To evaluate each component's effectiveness (DMP, ED, SA and SSE), we sequentially remove one while keeping the others unchanged. As can be seen from Table 3, removing any component (DMP, ED, SA, or SSE) from the model leads to a decrease in the performance of the model. Specifically, omitting the DMP module reduces the model's ability to capture both subset and global features at multiple scales, while removing the ED block degrades edge preservation, crucial for object delineation. Excluding the SA block weakens shape adaptability,

Table 1. Quantitative results of all methods in CT and X-ray denoising task.

Methods	CT		X-ray		Params	FLOPs
	PSNR	SSIM	PSNR	SSIM		
Noisy	27.563 2.038	0.812 0.071	9.001 0.955	0.354 0.085	-	-
WGAN-VGG [29]	30.851 1.787	0.848 0.018	17.517 1.205	0.717 0.023	20.49	1082.77
CBDNet [10]	28.765 2.391	0.815 0.043	27.379 1.030	0.832 0.023	4.36	40.28
QAE [7]	31.364 1.243	0.859 0.041	26.424 1.201	0.806 0.038	0.38	153.22
DML [15]	31.535 1.612	0.863 0.020	26.305 1.571	0.768 0.043	0.56	62.12
CNCL [8]	31.679 1.574	0.866 0.047	26.334 0.943	0.802 0.040	44.43	277.77
MWDCNN [25]	31.056 1.722	0.845 0.028	29.412 1.205	0.864 0.051	8.18	834.06
MINF-MCNN [6]	32.062 1.321	0.872 0.027	26.822 0.677	0.793 0.052	4.40	576.86
ViMEDnet [12]	31.625 1.263	0.851 0.025	28.964 0.877	0.845 0.049	3.68	34.26
LoMAE [27]	31.835 1.364	0.870 0.025	26.833 0.593	0.867 0.025	6.87	107.52
VisNet (ours)	**32.231** 1.395	**0.880** 0.027	**31.338** 0.619	**0.888** 0.014	**0.15**	**16.41**

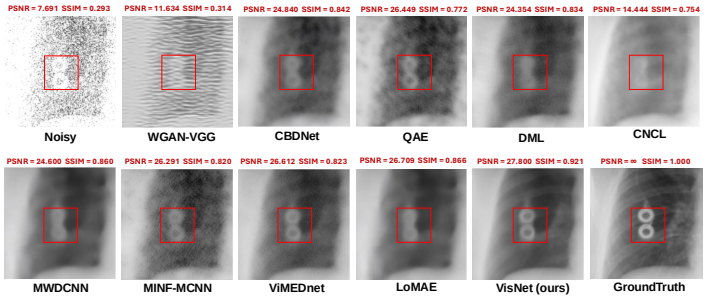

Fig. 3. Visualization results of all methods in X-ray image denoising task.

Table 2. Quantitative results of all methods in MRI denoising task.

Methods	MRI ($\sigma = 15$)		MRI ($\sigma = 50$)		Params	FLOPs
	PSNR	SSIM	PSNR	SSIM		
Noisy	12.738 2.110	0.291 0.085	8.724 0.602	0.050 0.021	-	-
WGAN-VGG [29]	27.494 1.896	0.766 0.151	25.961 1.307	0.739 0.145	20.49	1082.77
CBDNet [10]	33.917 3.385	0.846 0.057	28.332 3.788	0.746 0.074	4.36	40.28
QAE [7]	31.453 1.765	0.835 0.047	27.332 1.458	0.715 0.054	0.38	153.22
DML [15]	30.813 1.816	0.801 0.045	26.907 1.321	0.723 0.071	0.56	63.12
CNCL [8]	28.910 0.533	0.628 0.154	27.268 0.623	0.445 0.158	44.43	277.77
MWDCNN [25]	33.524 1.736	0.854 0.066	28.227 1.403	0.736 0.062	8.18	834.06
MINF-MCNN [6]	34.518 3.419	0.889 0.031	28.254 3.613	0.741 0.073	4.40	576.86
ViMEDnet [12]	34.224 2.817	0.872 0.046	28.319 2.375	0.737 0.058	3.68	34.26
LoMAE [27]	34.337 3.601	0.891 0.029	28.216 3.578	0.736 0.069	6.87	107.52
VisNet (ours)	**35.024** 3.359	**0.924** 0.020	**28.907** 4.021	**0.753** 0.071	**0.15**	**16.41**

Table 3. Ablation experimental results of key modules on three datasets.

Methods	CT		X-ray		MRI ($\sigma = 15$)	
	PSNR	SSIM	PSNR	SSIM	PSNR	SSIM
w/o DMP	26.736 1.318	0.824 0.036	29.848 0.562	0.864 0.013	33.732 3.775	0.901 0.017
w/o ED	32.013 1.665	0.874 0.028	26.419 1.253	0.807 0.013	34.450 3.817	0.917 0.019
w/o SA	32.028 1.747	0.875 0.029	30.490 0.505	0.868 0.014	34.506 3.662	0.912 0.018
w/o SSE	31.945 1.707	0.873 0.029	30.880 0.570	0.879 0.015	30.837 3.447	0.815 0.046
VisNet	**32.231** 1.395	**0.880** 0.027	**31.338** 0.619	**0.888** 0.014	**35.024** 3.359	**0.924** 0.020

impairing generalization. Lastly, removing the SSE module hampers the model's ability to capture semantic relationships, adversely affecting contextual understanding. These results highlight the importance of each module, emphasizing their contribution to VisNet's performance.

In addition, we also analyze the impact of different learning rates and different normalization methods in the spatial-semantic extraction module on model performance. The experimental results show that a learning rate of 0.001 yields the best model performance. Additionally, InstanceNorm proves to be more suitable for this study.

3.4 Clinical Application of VisNet

To further validate the clinical application value of the denoised images produced by VisNet, we conduct brain tumor classification experiments using Res-BRNet [32] as the baseline. When noisy MRI images are used as input, the accuracy of Res-BRNet is 87.5%. However, when MRI images denoised by VisNet are used as input, the accuracy of Res-BRNet improved to 96.7%, which is close to the 97.5% accuracy achieved with noise-free images as input. These results highlight the clinical value of VisNet in clinical diagnosis.

4 Conclusion

We propose a lightweight dual-path network inspired by the human visual system for medical images denoising. Overall, VisNet presents an accurate and efficient method for medical images denoising. In future work, we will further investigate the interpretability of different modules in VisNet. Additionally, we aim to explore the clinical application value of VisNet across various types of medical images, to assess its effectiveness in real-world medical scenarios.

Acknowledgement. The work was supported in part by the National Natural Science Foundation of China (No. U24A20256), Science and Technology Major Project of Changsha (No. kh2502004), Xinjiang Uygur Autonomous Region Key R&D program (No. 2024B03039-3), central government guides local funds for scientific and technological development of China (No. ZYYD2025QY25), Scientific Research Fund of Hunan Provincial Education Department (No.23A0020), the Natural Science Foundation of Hunan Province (No. 2025JJ50374), and the Guiyang City Science and Technology Plan Project (No. [2024]2-18). This work was carried out in part using computing resources at the High-Performance Computing Center of Central South University.

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

References

1. Al-Mallah, M.H., Aljizeeri, A., Alharthi, M., Alsaileek, A.: Routine low-radiation-dose coronary computed tomography angiography. European Heart Journal Supplements **16**, B12–B16 (2014)
2. Bluethgen, C., Chambon, P., Delbrouck, J.B., van der Sluijs, R., Połacin, M., Zambrano Chaves, J.M., Abraham, T.M., Purohit, S., Langlotz, C.P., Chaudhari, A.S.: A vision–language foundation model for the generation of realistic chest x-ray images. Nature Biomedical Engineering pp. 1–13 (2024)
3. Chang, M., Li, Q., Feng, H., Xu, Z.: Spatial-adaptive network for single image denoising. In: Computer Vision–ECCV 2020: 16th European Conference, Glasgow, UK, August 23–28, 2020, Proceedings, Part XXX 16. pp. 171–187. Springer (2020)
4. Chen, H., Zhang, Y., Kalra, M.K., Lin, F., Chen, Y., Liao, P., Zhou, J., Wang, G.: Low-dose ct with a residual encoder-decoder convolutional neural network. IEEE transactions on medical imaging **36**(12), 2524–2535 (2017)
5. Chen, T., Ma, K.K., Chen, L.H.: Tri-state median filter for image denoising. IEEE Transactions on Image processing **8**(12), 1834–1838 (1999)
6. Du, Q., Tang, Y., Wang, J., Hou, X., Wu, Z., Li, M., Yang, X., Zheng, J.: X-ray ct image denoising with minf: A modularized iterative network framework for data from multiple dose levels. Computers in Biology and Medicine **152**, 106419 (2023)
7. Fan, F., Shan, H., Kalra, M.K., Singh, R., Qian, G., Getzin, M., Teng, Y., Hahn, J., Wang, G.: Quadratic autoencoder (q-ae) for low-dose ct denoising. IEEE transactions on medical imaging **39**(6), 2035–2050 (2019)
8. Geng, M., Meng, X., Yu, J., Zhu, L., Jin, L., Jiang, Z., Qiu, B., Li, H., Kong, H., Yuan, J., Yang, K., Shan, H., Han, H., Yang, Z., Ren, Q., Lu, Y.: Content-noise complementary learning for medical image denoising. IEEE Transactions on Medical Imaging **41**(2), 407–419 (2022)

9. Geng, M., Meng, X., Yu, J., Zhu, L., Jin, L., Jiang, Z., Qiu, B., Li, H., Kong, H., Yuan, J., et al.: Content-noise complementary learning for medical image denoising. IEEE transactions on medical imaging **41**(2), 407–419 (2021)
10. Guo, S., Yan, Z., Zhang, K., Zuo, W., Zhang, L.: Toward convolutional blind denoising of real photographs. In: Proceedings of the IEEE/CVF conference on computer vision and pattern recognition. pp. 1712–1722 (2019)
11. Huang, J., Chen, K., Ren, Y., Sun, J., Wang, Y., Tao, T., Pu, X.: Cddnet: Cross-domain denoising network for low-dose ct image via local and global information alignment. Computers in Biology and Medicine p. 107219 (2023)
12. Huang, J., Zhong, A., Wei, Y.: A new visual state space model for low-dose ct denoising. Medical Physics (2024)
13. Jang, S.I., Pan, T., Li, Y., Heidari, P., Chen, J., Li, Q., Gong, K.: Spach transformer: spatial and channel-wise transformer based on local and global self-attentions for pet image denoising. IEEE Transactions on Medical Imaging (2023)
14. Jifara, W., Jiang, F., Rho, S., Cheng, M., Liu, S.: Medical image denoising using convolutional neural network: a residual learning approach. The Journal of Supercomputing **75**, 704–718 (2019)
15. Jung, C., Lee, J., You, S., Ye, J.C.: Patch-wise deep metric learning for unsupervised low-dose ct denoising. In: International Conference on Medical Image Computing and Computer-Assisted Intervention. pp. 634–643. Springer (2022)
16. Karimi, D., Dou, H., Warfield, S.K., Gholipour, A.: Deep learning with noisy labels: Exploring techniques and remedies in medical image analysis. Medical image analysis **65**, 101759 (2020)
17. Kumar, B.S.: Image denoising based on gaussian/bilateral filter and its method noise thresholding. Signal Image Video Process. **7**(6), 1159–1172 (2013)
18. Li, Y., Zhang, K., Shi, W., Miao, Y., Jiang, Z.: A novel medical image denoising method based on conditional generative adversarial network. Computational and Mathematical Methods in Medicine **2021**(1), 9974017 (2021)
19. Ma, Y., Yan, Q., Liu, Y., Liu, J., Zhang, J., Zhao, Y.: Strunet: Perceptual and low-rank regularized transformer for medical image denoising. Medical Physics (2023)
20. Murphy, K.J., Brunberg, J.A.: Adult claustrophobia, anxiety and sedation in mri. Magnetic resonance imaging **15**(1), 51–54 (1997)
21. Özbey, M., Dalmaz, O., Dar, S.U., Bedel, H.A., Özturk, Ş., Güngör, A., Çukur, T.: Unsupervised medical image translation with adversarial diffusion models. IEEE Transactions on Medical Imaging (2023)
22. Pitcher, D., Ungerleider, L.G.: Evidence for a third visual pathway specialized for social perception. Trends in cognitive sciences **25**(2), 100–110 (2021)
23. Song, Q., Ma, L., Cao, J., Han, X.: Image denoising based on mean filter and wavelet transform. In: 2015 4th International Conference on Advanced Information Technology and Sensor Application (AITS). pp. 39–42. IEEE (2015)
24. Szegedy, C., Liu, W., Jia, Y., Sermanet, P., Reed, S., Anguelov, D., Erhan, D., Vanhoucke, V., Rabinovich, A.: Going deeper with convolutions. In: IEEE Conference on Computer Vision and Pattern Recognition (CVPR). pp. 1–9 (2015)
25. Tian, C., Zheng, M., Zuo, W., Zhang, B., Zhang, Y., Zhang, D.: Multi-stage image denoising with the wavelet transform. Pattern Recognition **134**, 109050 (2023)
26. Velayudham, A., Kumar, K.M., Priya, M.K.: Improving brain mri denoising using convolutional autoencoder and sparse representations. Expert Systems with Applications **263**, 125711 (2025)
27. Wang, D., Han, S., Xu, Y., Wu, Z., Zhou, L., Morovati, B., Yu, H.: Lomae: simple streamlined low-level masked autoencoders for robust, generalized, and inter-

pretable low-dose ct denoising. IEEE Journal of Biomedical and Health Informatics (2024)
28. Wang, S., Wang, C.Y., Wang, P., Wang, C., Li, Z.A., Pan, C., Dai, Y., Gao, A., Liu, C., Liu, J., et al.: Networking retinomorphic sensor with memristive crossbar for brain-inspired visual perception. National science review **8**(2), nwaa172 (2021)
29. Yang, Q., Yan, P., Zhang, Y., Yu, H., Shi, Y., Mou, X., Kalra, M.K., Zhang, Y., Sun, L., Wang, G.: Low-dose ct image denoising using a generative adversarial network with wasserstein distance and perceptual loss. IEEE transactions on medical imaging **37**(6), 1348–1357 (2018)
30. Yuan, J., Zhou, F., Guo, Z., Li, X., Yu, H.: Hcformer: hybrid cnn-transformer for ldct image denoising. Journal of Digital Imaging **36**(5), 2290–2305 (2023)
31. Yue, H., Liu, J., Li, J., Kuang, H., Lang, J., Cheng, J., Peng, L., Han, Y., Bai, H., Wang, Y., et al.: Mldrl: Multi-loss disentangled representation learning for predicting esophageal cancer response to neoadjuvant chemoradiotherapy using longitudinal ct images. Medical image analysis **79**, 102423 (2022)
32. Zahoor, M.M., Khan, S.H., Alahmadi, T.J., Alsahfi, T., Mazroa, A.S.A., Sakr, H.A., Alqahtani, S., Albanyan, A., Alshemaimri, B.K.: Brain tumor mri classification using a novel deep residual and regional cnn. Biomedicines **12**(7), 1395 (2024)