



Article

AsymUNet: An Efficient Multi-Layer Perceptron Model Based on Asymmetric U-Net for Medical Image Noise Removal

Yan Cui ¹ , Xiangming Hong ², Haidong Yang ³, Zhili Ge ¹ and Jielin Jiang ^{2,4,5,*} 

- ¹ School of Mathematics and Information Science, Nanjing Normal University of Special Education, Nanjing 210038, China; csyanncui@njts.edu.cn (Y.C.); 250053@njts.edu.cn (Z.G.)
- ² School of Software, Nanjing University of Information Science and Technology, Nanjing 210044, China; 202212490289@nuist.edu.cn
- ³ School of Information Engineering, Inner Mongolia University of Technology, Hohhot 010051, China; imperman@imut.edu.cn
- ⁴ State Key Laboratory for Novel Software Technology, Nanjing University, Nanjing 210023, China
- ⁵ Jiangsu Province Engineering Research Center of Advanced Computing and Intelligent Services, Nanjing University of Information Science and Technology, Nanjing 210044, China
- * Correspondence: jiangjielin2008@nuist.edu.cn

Abstract: With the continuous advancement of deep learning technology, U-Net-based algorithms for image denoising play a crucial role in medical image processing. However, most U-Net-based medical image denoising algorithms typically have large parameter sizes, which poses significant limitations in practical applications where computational resources are limited or large-scale patient data processing are required. In this paper, we propose a medical image denoising algorithm called AsymUNet, developed using an asymmetric U-Net framework and a spatially rearranged multilayer perceptron (MLP). AsymUNet utilizes an asymmetric U-Net to reduce the computational burden, while a multiscale feature fusion module enhances the feature interaction between the encoder and decoder. To better preserve the image details, spatially rearranged MLP blocks serve as the core building blocks of AsymUNet. These blocks effectively extract both the local and global features of the image, reducing the model's reliance on prior knowledge of the image and further accelerating the training and inference processes. Experimental results demonstrate that AsymUNet achieves superior performance metrics and visual results compared with other state-of-the-art methods.

Keywords: medical image denoising; asymmetric U-Net; multi-scale feature fusion module; spatially rearrangement MLP blocks



Citation: Cui, Y.; Hong, X.; Yang, H.; Ge, Z.; Jiang, J. AsymUNet: An Efficient Multi-Layer Perceptron Model Based on Asymmetric U-Net for Medical Image Noise Removal. *Electronics* **2024**, *13*, 3191. <https://doi.org/10.3390/electronics13163191>

Academic Editor: Chiman Kwan

Received: 6 July 2024

Revised: 6 August 2024

Accepted: 11 August 2024

Published: 12 August 2024



Copyright: © 2024 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<https://creativecommons.org/licenses/by/4.0/>).

1. Introduction

With the rapid advancement of medical technology, the field of medical image denoising has made significant progress as a fundamental task in computer vision. Despite the improvements in imaging equipment, medical image denoising continues to present challenges in real-world applications. As a result, there is a growing interest among scholars in developing effective medical image denoising algorithms. Medical image denoising is not only an essential preprocessing step in medical imaging workflows, but it also has a direct impact on the accuracy and reliability of subsequent analyses. Therefore, achieving accurate image denoising is crucial for enhancing the quality of medical images.

Traditional denoising methods often rely on filters to remove noise from images. While these filter-based algorithms are effective, they do have a few drawbacks: (1) each denoising task requires a specific model, (2) iterative denoising methods are time consuming, (3) they lack generality for various data types, and (4) manual or semi-automatic parameter tuning is necessary. These limitations greatly affect the performance of medical image denoising algorithms. However, with significant advancements in deep-learning algorithms for medical image denoising, deep-learning-based methods have achieved state-of-the-art performance by adaptively extracting more complex features from medical image data.

Among these methods, convolutional neural networks (CNNs) [1–3] stand out for their ability to learn local patterns from extensive medical image datasets. For medical image denoising based on deep learning, U-Net-based CNNs [4] are widely used. Meanwhile, transformers [5–7] have shown impressive results on large datasets by relying less on predefined data assumptions during training. However, both CNNs and transformers face challenges: CNNs often struggle with limited receptive fields, which leads to suboptimal denoising outcomes, while transformers require intricate self-attention mechanisms to process input data, resulting in a high computational overhead and prolonged inference times. Therefore, the MLP-mixer architecture, which is based on multilayer perceptrons (MLPs), emerges as a favorable balance between the two. Specifically, MLPs have lower data priors and larger receptive fields compared with CNNs, while also being simpler with smaller computational demands compared with transformer networks. Moreover, MLPs primarily rely on matrix multiplication, which makes them highly efficient on GPUs designed for parallel processing. However, despite the excellent performance of MLPs in many tasks, they do not effectively account for the local information of images. In medical image denoising tasks, the local self-similarity of images is a crucial attribute, and this limitation can hinder the performance of MLPs in handling such tasks.

Initially used in biomedical image segmentation, U-Net has also shown outstanding results in image denoising. Recent research has investigated ways to improve its denoising capabilities by combining U-Net with transformers or adversarial neural networks. Zhang et al. [7] merged U-Net with transformers, effectively utilizing global image features at different scales and employing local attention mechanisms to enhance the local image information. Huang et al. [8] integrated U-Net with adversarial networks to capture both global and local variations between denoised and original images, resulting in superior performance metrics. Despite their effectiveness in extracting multiscale feature maps through their encoder–decoder structure, these U-Net variants become more computationally demanding when additional basic modules are stacked. Additionally, conventional U-Nets primarily rely on skip connections, which restrict feature map interactions between different scales. These approaches fail to consider the structural and texture similarities present in both multiscale and same-scale feature maps, ultimately limiting their denoising capabilities.

To overcome these limitations, we propose the asymmetric multilayer perceptron U-Net (AsymUNet) for medical image denoising. AsymUNet uses an asymmetric U-Net architecture to reduce computational load, along with a multiscale feature fusion module (MSFFM) to enhance the information interaction between the encoder and decoder. Furthermore, spatially rearranged multilayer perceptron blocks function as core building blocks, effectively extracting local features by reorganizing feature map spatial structures. AsymUNet demonstrates superior performance metrics and improved visual effects compared with existing methods. Extensive experiments confirm AsymUNet's exceptional performance across various metrics and its ability to deliver exceptional visual quality. The primary contributions of this study are summarized as follows:

1. We propose AsymUNet, a denoising algorithm for medical images based on an asymmetric U-Net framework and a spatially rearranged MLP. AsymUNet effectively reduces the computational load compared to conventional U-Net structures, while maintaining an excellent denoising performance.
2. We introduce a MSFFM that integrates the feature information from all scales, including both multiscale and same-scale feature maps. This improvement in the decoder greatly enhances the denoising effectiveness and achieves higher performance metrics.
3. We use MLP blocks that are spatially rearranged as core modules in our approach. These blocks extract local information from a spatial perspective and global information from a channel perspective, improving image feature representations and preserving image details. Additionally, MLPs utilize basic matrix multiplication for feature extraction, which results in a superior inference speed in practical applications.

The remainder of this paper is organized as follows: Section 2 reviews related work. Section 3 introduces the AsymUNet model. In Section 4, we present extensive experiments conducted to evaluate the performance of AsymUNet. Finally, Section 5 concludes the paper with a summary of findings and potential future directions.

2. The Related Work

In this section, our focus is primarily on two aspects of medical image denoising: denoising Gaussian grayscale/color medical images and denoising low-dose CT images.

2.1. Grayscale/Color Medical Image Denoising

CNNs have shown a strong performance in denoising Gaussian grayscale/color medical images, particularly in capturing local information and other relevant image priors. In the early stages of denoising grayscale/color medical images, researchers mainly used filters for noise reduction. Xu et al. [9] proposed a parallel NLM algorithm that enhances the noise-weighted kernel function to achieve a higher denoising quality in medical images. They also utilized GPU-accelerated algorithms for faster inference times. Diwakar et al. [10] introduced a method that combines the NLM filter with wavelet transform, effectively suppressing noise in medical images by adjusting the wavelet transform threshold, resulting in improved performance metrics. Fisnandya et al. [11] utilized the block matching and 3D collaborative filtering (BM3D) algorithm [12] for medical image denoising. This algorithm aggregates similar feature maps through block matching to achieve excellent measurement metrics. However, these algorithms often require significant computational resources and are sensitive to parameter settings, which limit their practical application. Consequently, researchers have increasingly turned their attention to deep learning approaches. Liu et al. [13] proposed a model for denoising medical images based on a genetic algorithm. This model enhances performance by using a greedy approach to retain high-performance genes and suppress defective ones. Rawat et al. [14] developed CVMIDNet, a model based on complex domain CNNs that combines phase and amplitude information to extract the feature maps, demonstrating an excellent denoising performance. Dong et al. [15] introduced the feature-guided denoising convolutional neural network, which incorporates a feature masking layer and hierarchical framework to effectively preserve key feature information. Geng et al. [16] proposed the content-noise complementary learning method, which accurately removes noise by separately predicting image content and noise components. Sharif et al. [17] introduced a deep dynamic residual attention network that integrates attention mechanisms into residual CNNs for superior medical image restoration results. Finally, Ghahremani et al. [18] proposed ADL, leveraging residual blocks and a feature pyramid to enhance image feature extraction and improve the visual effects in medical image denoising tasks.

2.2. Low-Dose CT Image Denoising

Low-dose CT (LDCT) denoising of medical images has become a prominent area of research in deep learning. Traditional methods, such as NLM algorithms and BM3D, have been widely used for LDCT denoising. For example, Zhang et al. [19] proposed a region-adaptive NLM algorithm that divides LDCT image pixels based on edge information, adjusting search windows, block sizes, and filtering parameters to improve visual quality. Chen et al. [20] introduced an improved BM3D method for reducing noise in LDCT, integrating a visual attention technique to highlight lesions and enhance the imaging contrast of diseased tissues. However, these traditional approaches rely on manual prior knowledge and parameter tuning, often resulting in over-smoothing during denoising, which can be particularly problematic in LDCT due to non-uniform noise that can degrade texture and structural details. In response, deep learning has emerged as a promising approach by integrating LDCT denoising algorithms with neural networks to automatically extract image feature maps, achieving a robust performance and better generalization compared with traditional methods. Chen et al. [21] developed the residual encoder–decoder CNN (RED-

CNN), which combines auto-encoders, deconvolution networks, and skip connections into residual networks to effectively preserve more detailed image information. Liang et al. [22] introduced the edge-enhanced CNN (EDCNN), which incorporates dense connections to integrate extracted edge information throughout the network, facilitating end-to-end image denoising. Luthra et al. [23] introduced the transformer architecture into medical image denoising, enhancing edge information using learnable Sobel–Feldman operators (EFormer), thereby improving denoising performance. Additionally, Wolterink et al. [24] combined GANs with CNNs to effectively suppress noise in LDCT images while maintaining visual quality. Ma et al. [25] introduced StruNet, an encoder–decoder architecture that integrates Swin transformer modules and residual blocks, leveraging perceptual loss and low-rank regularization to reduce noise and artifacts in CT images. Yang et al. [26] proposed a GAN based on Wasserstein distance and perceptual similarity (WGAN), optimizing network performance by minimizing differences in feature space and preserving image texture information. These advancements underscore how deep learning has enabled the development of more sophisticated techniques, offering enhanced generalization and better preservation of intricate image details.

3. AsymUNet

In this section, we provide a detailed explanation of the proposed AsymUNet for denoising medical images. This explanation covers the overall model structure, the design of multiscale feature fusion module, the design of basic blocks, and the loss function used during training.

3.1. Main Framework

As we know, the U-Net architecture has gained widespread popularity in image denoising due to its ability to preserve image details through skip connections that transfer feature maps from the encoder to the decoder. However, this method primarily focuses on feature maps at specific scales, which can limit its denoising performance. Moreover, integrating computationally intensive modules such as MLPs or transformers can further burden the neural network, impacting its efficiency. To address these challenges, we propose AsymUNet, a medical image denoising algorithm based on an asymmetric U-Net framework and spatially rearranged multilayer perceptrons. AsymUNet adopts an asymmetric U-Net structure to reduce the computational load and processing time while effectively capturing both global and local information. Additionally, it includes an MSFFM to enhance the integration of feature maps. Figure 1 illustrates the overall structure of AsymUNet.

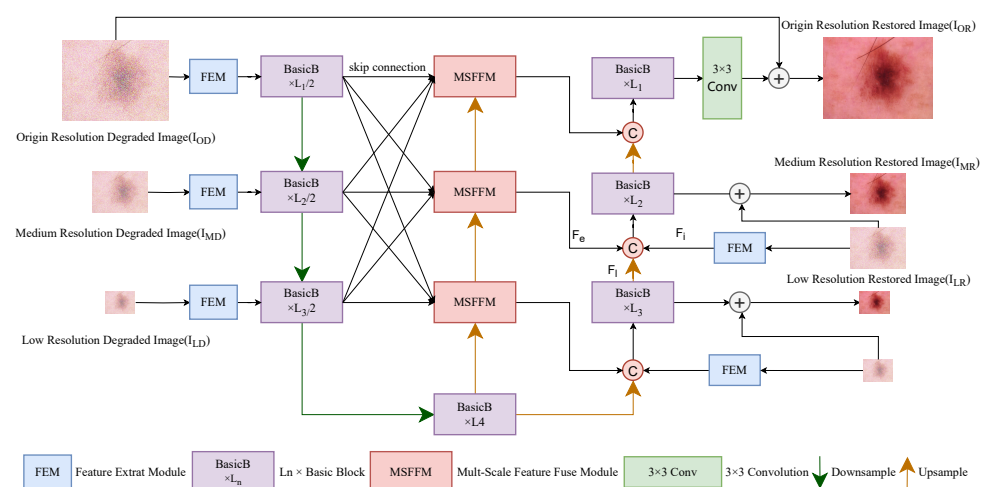


Figure 1. The overall structure of AsymUNet.

To reduce the computational load, the asymmetric U-Net reduces the number of basic blocks in the encoder while enhancing the encoder structure to maintain the denoising performance despite the smaller model size. At each downsampling layer, the encoder incorporates feature maps from the degraded image as additional inputs, ensuring that the texture information at different scales is preserved during the downsampling process. Effective feature map transfer between the encoder and decoder is crucial for the denoising performance. To optimize the utilization of multiscale feature maps and enhance the denoising efficacy, we introduce an MSFFM. This module aims to combine features from various scales and transfer them effectively to the corresponding decoder layers. The detailed mechanism of this module will be explained in Section 3.2. As for the decoder design, we propose an architecture that integrates inputs from three different sources: the merged feature maps from the MSFFM (F_I), the restored features from the previous decoder layer (F_I), and the texture features extracted from the degraded image (F_I). By integrating these three feature maps, the decoder becomes better at reconstructing both the overall structure and fine details of the image. This decoder architecture functions similarly to a compact neural network. By inputting comprehensive image features into the decoder, it significantly reduces the encoder's workload for feature extraction. To further improve the model's computational efficiency, we introduce spatially rearranged MLP blocks that rely on matrix multiplication as fundamental components. These blocks restructure the spatial arrangement of features to extract local features and derive global features from the image from a channel perspective. This approach helps achieve superior performance metrics. The detailed design of these basic blocks is presented in Section 3.3. Additionally, AsymUNet uses a three-layer encoder–decoder architecture; at the bottleneck, the feature maps extracted by the encoders are sent to spatially rearranged MLP blocks, which are then ready to be decoded through the same number of stages as they were encoded.

3.2. Multiscale Feature Fusion Module

Skip connections are essential in U-Net as they enable crucial interaction between the encoder and decoder, significantly enhancing feature fusion. However, despite being efficient and effective, skip connections do not fully exploit the correlations between feature maps at different scales. This limitation can restrict the denoising performance. Taking inspiration from research by Li et al. [27], it becomes evident that image feature maps exhibit commonalities on two levels. Firstly, within images at the same scale, feature maps at the corresponding positions display significant texture similarity. Secondly, across images of varying scales, structural feature maps exhibit clear similarities. Building on these insights, this paper argues that preserving this intrinsic coherence is vital for achieving comprehensive and effective information integration in multiscale feature fusion.

To effectively reduce noise in medical images, we propose an MSFFM that merges image feature maps across different scales to preserve as much detail and structure as possible. Taking a mid-scale MSFFM as an example (Figure 2), this module integrates four inputs: the feature map F_N from the decoder output at the current resolution, the feature map F_{HD} from the decoder output at a higher resolution, the feature map F_{LD} from the decoder output at a lower resolution, and the output from the preceding MSFFM. This fusion process not only maintains the unique structural characteristics of each scale, but also effectively preserves the texture information at the current scale. The operations of the MSFFM are represented by Equation (1):

$$F_O = \phi(F_N, F_{HD}, F_{LD}, F_{BO}), \quad (1)$$

where ϕ denotes the operation of integrating these four feature maps, involving a spatial rearrangement MLP (SRCMLP) block and a channel attention module; F_O represents the output of the MSFFM at the current resolution; and F_{BO} is the output from the previous MSFFM.

With the SRCMLP block, AsymUNet efficiently extracts detailed image feature maps. By introducing the channel attention mechanism, it is able to capture global information

within these feature maps. The resulting output from this module not only contributes to the next higher resolution fusion feature module, but also significantly enhances the denoising capability of the decoder within our proposed framework.

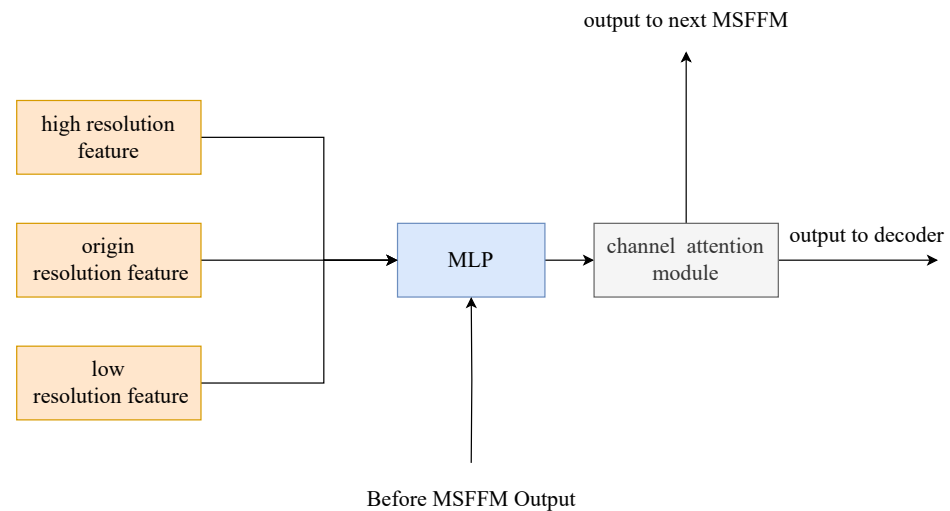


Figure 2. Diagram illustrating the structure of MSFFM.

3.3. Spatial Rearrangement Multilayer Perceptron Block

In MLP-Mixer [28], the Mix-Token MLP block accepts tokens that have been linearly expanded as the input feature maps, and it uses fully connected layers to model the relationships between these tokens. However, the fixed input dimensions of the fully connected layers in the Mix-Token MLP block make it incompatible with the variable image sizes that are encountered in medical image processing. Furthermore, as the Mix-Token MLP block extracts global information directly from the sequence on each pass, there is a risk of overlooking crucial local information.

To address this issue, we propose a SRCMLP block as the foundational block of the model. Similar to MLP-Mixer, each SRCMLP consists of two sub-blocks: the rearrangement module (RC) and the channel MLP module (CMLP). These sub-blocks are designed to extract spatial and channel information from the input feature maps, respectively. Given an input feature X with dimensions $R^{H \times W \times C}$, where H , W , and C represent the height, width, and original channel count of the image, respectively, the operations of the SRCMLP can be expressed as shown in Equations (2) and (3):

$$Y = RC(X) + X \quad (2)$$

$$Z = CMLP(Y) + Y \quad (3)$$

where Y and Z represent the intermediate feature maps and the final output feature maps, respectively. The specific operations of the RC module are shown in Figure 3. First, the spatial dimension of the input feature maps is divided into multiple fixed-size blocks according to a specified square region size. Then, the feature maps in these fixed-size blocks are reorganized. Suppose the specified side length for division is Len . The input feature map X with shape $H \times W \times C$ is divided into T regions, each with a shape of $Len \times Len \times C$. Next, the feature maps of each region are reorganized, transforming the shape of the input feature map X into $\frac{H}{Len} \times \frac{W}{Len} \times (C \times Len \times Len)$. By utilizing two fully connected layers and a batch normalization (BN) layer, the local information within the input feature maps are effectively extracted. Finally, the CMLP module is employed to derive global information from these inputs, ultimately enhancing the network's denoising performance.

As shown in Figure 4, to capture more detailed image features, the CMLP module uses a fully connected layer to increase the feature map dimension to $4 \times C$. Furthermore,

the inclusion of BN, GELU activation, and dropout improves the stability and performance of the network, enabling the model to converge quickly. Lastly, a fully connected layer is employed to restore the original number of channels.

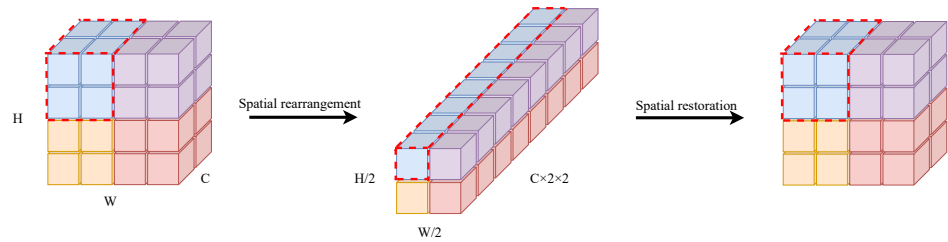


Figure 3. The proposed rearrangement module.

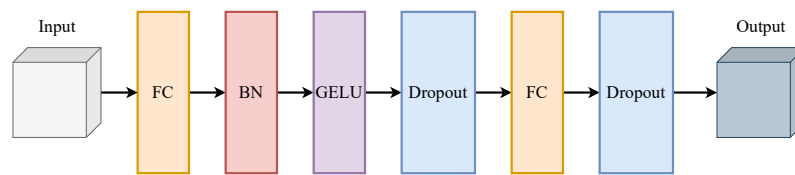


Figure 4. The proposed channel multilayer perceptron module.

In the SRCMLP block, the fully connected layer operation is the primary computational and parameter burden. Assume the shape of the input feature map is $H \times W \times C$. In the RCMLP module, suppose the region’s side length after division is Len . After rearrangement, the shape of the feature map becomes $\frac{H}{Len} \times \frac{W}{Len} \times (C \times Len \times Len)$. Assume the feature map dimension in the bottleneck is reduced to $C/2$. Therefore, the parameter count for the rearrangement module is $(C \times Len \times Len) \times \frac{C}{2} \times 2 = C^2Len^2$. The floating point operations per second (FLOPs) are calculated as $\frac{H}{Len} \times \frac{W}{Len} \times (C \times Len \times Len) \times \frac{C}{2} \times 2 = HWC^2$. Next, in the CMLP module, assume the feature map dimension in the bottleneck is modified to $4 \times C$. The parameters and FLOPs for the CMLP are $C \times 4 \times C \times 2 = 8C^2$ and $H \times W \times C \times 4 \times C \times 2 = 8HWC^2$, respectively. The total parameter count and FLOPs are $(8 + Len^2)C^2$ and $9HWC^2$, respectively. If the fully connected layers are replaced by conventional 3×3 convolution layers, the total parameter count and FLOPs would be $(72 + 9Len^2)C^2$ and $81HWC^2$, respectively, which are significantly greater than the former. Therefore, SRCMLP has fewer parameters and FLOPs compared with using convolution operations. MLP uses matrix multiplication instead of convolutional kernels for feature map processing, significantly improving the computational efficiency through parallel computation on GPUs.

3.4. The Loss Function

Most image denoising algorithms primarily focus on calculating the loss between the output image and the clean image. However, in the U-Net architecture, the image is progressively restored through multiple hierarchical levels. If we only compute the loss between the final output layer and the clean image, we will overlook the denoising processes at various stages within the network. To address this issue, AsymUNet introduces a multiscale loss function, which consists of multiple L1 loss functions. The L1 loss function can be represented by Equation (4) as follows:

$$L_1(I_O, I_R) = 1/T \sum_{t=1}^T \|I_O^t - I_R^t\| \tag{4}$$

where I_O is the set of original images, I_R is the set of images restored by AsymUNet, and I_O^t and I_R^t represent the t -th original image and denoised image, respectively. T is the number of samples.

Meanwhile, the multiscale loss function consists of three levels of loss functions, which can be represented by Equation (5):

$$L_{cout} = \sum_{e=1}^E L_1^e(I_O, I_R) \quad (5)$$

where E is the number of layers. The $L1$ loss will be accumulated between the input and output images at each layer of the network to account for the overall loss.

To improve the performance of denoising, auxiliary loss terms can be incorporated into the loss function. For AsymUNet, this involves integrating a multiscale frequency reconstruction (MSFR) loss function with the existing multiscale loss function. The MSFR loss measures the $L1$ distance in the frequency domain at different scales between the original image and the denoised image. This relationship is represented by Equation (6):

$$L_{MSFR}(I_O, I_R) = 1/T \sum_{t=1}^T \|F(I_O^t) - F(I_R^t)\| \quad (6)$$

where F represents the fast Fourier transform (FFT), which converts the image signal into the frequency domain information.

The overall loss function for training AsymUNet consists of two components: L_{cout} and L_{MSFR} . To control the influence of the auxiliary term L_{MSFR} , parameter μ is introduced as a coefficient before L_{MSFR} . Therefore, the overall loss function can be represented as shown in Equation (7):

$$L_{total} = L_{cout} + \mu L_{MSFR} \quad (7)$$

4. Experiments

Section 4.1 describes the experimental setup, while Section 4.2 presents the experimental results. Additionally, Section 4.3 presents the results of the ablation experiments.

4.1. Experimental Settings

We conducted relevant experiments on the proposed AsymUNet network for various medical image denoising tasks, including: (1) Gaussian grayscale/color medical image denoising and (2) LDCT image denoising.

(1) Gaussian Grayscale/Color Medical Image Denoising: Images from the HAM10000 dataset [29] and the Chest X-ray dataset [30] are used as the training and testing datasets. During training, the white Gaussian noise is added to the training set images. The standard deviation (σ) is randomly chosen from the interval [0, 55]. The HAM10000 dataset consists of 10,000 RGB images of pigmented skin lesions from various individuals, each sized $450 \times 600 \times 3$. For training, 9400 images are randomly selected, leaving 300 images for testing. The chest X-ray dataset was collected from pediatric patients aged one to five at the Guangzhou Women and Children's Medical Center. It includes 5216 images for training and 624 images for testing.

(2) Low-Dose CT Medical Image Denoising: To evaluate the proposed AsymUNet on the actual medical images, we used the Mayo Clinic CT dataset [31] in our experiments. This dataset was created for the 2016 Low Dose CT Grand Challenge, which was jointly organized by the National Institutes of Health, the American Association of Physicists in Medicine (AAPM), and the Mayo Clinic. Its purpose is to compare different LDCT image denoising algorithms. The dataset consists of 150 sets of projection data and corresponding image data obtained from clinical CT examinations using SOMATOM definition CT systems. For our experiments, we chose 2167 pairs of quarter-dose and full-dose CT images for training, and 100 pairs for testing.

The AsymUNet network is trained using the PyTorch framework on a workstation equipped with an Intel(R) Core(TM) i5-13600KF CPU @3.50 GHz processor and an NVIDIA RTX 3090 GPU. For all of the experiments, the training parameters are configured as

follows: we employed the AdamW optimizer with parameters $\beta_1 = 0.9$, $\beta_2 = 0.999$, and a weight decay of 1×10^{-4} . The loss function used is L1 loss, and training proceeds for 300,000 iterations. The initial learning rate is set to 3×10^{-4} and is decreased gradually to 1×10^{-6} using a cosine annealing strategy. To facilitate smooth training, a progressive learning strategy is adopted. Initially, the patch size and batch size are set to 128×128 and 8, respectively. As training progresses and the number of iterations reaches 92,000, 156,000, 204,000, 240,000, and 276,000, the patch size and batch size pairs are adjusted as follows: $[(160^2, 4), (192^2, 3), (256^2, 2), (320^2, 2), (384^2, 1)]$, respectively.

4.2. Results

In this section, we demonstrate the denoising capabilities of the proposed AsymUNet on different types of medical images. This includes Gaussian grayscale and color images from dermatological and pulmonary datasets. Additionally, we demonstrate the network's effectiveness in denoising LDCT images.

(1) Denoising of Gaussian Grayscale/Color Medical Images: We evaluated the performance of AsymUNet against four competitive algorithms: the traditional BM3D [12] algorithm, as well as three deep learning methods including DNCNN [1], ADL [18], and DRANet [2]. To assess the effectiveness of AsymUNet, we utilized the peak signal-to-noise ratio (PSNR) and structural similarity index (SSIM) as evaluation metrics.

To improve the generalizability of the proposed AsymUNet, Gaussian white noise at various random levels was added to the training images. During the testing phase, the network's ability to generalize was assessed across five different noise levels: $\sigma = 10, 15, 25, 35, 50$. The experimental results are detailed in Table 1.

Table 1. Comparison of denoising results (AVERAGE \pm STANDARD DEVIATION) on Gaussian Color/Grayscale Medical Images.

Dataset	Noise Levels	PSNR					SSIM				
		BM3D	DNCNN	ADL	DRANet	AsymUNet	BM3D	DNCNN	ADL	DRANet	AsymUNet
Human	10	37.47 \pm 1.3	38.07 \pm 1.4	38.82 \pm 1.3	39.08 \pm 1.4	39.30 \pm 1.2	0.91 \pm 0.02	0.91 \pm 0.02	0.93 \pm 0.02	0.93 \pm 0.02	0.93 \pm 0.01
	15	35.92 \pm 1.3	35.85 \pm 1.7	37.51 \pm 1.3	38.01 \pm 1.4	38.12 \pm 1.3	0.87 \pm 0.03	0.86 \pm 0.03	0.92 \pm 0.02	0.92 \pm 0.02	0.92 \pm 0.02
	25	34.44 \pm 1.5	34.94 \pm 1.6	36.17 \pm 1.4	36.48 \pm 1.4	36.71 \pm 1.3	0.85 \pm 0.03	0.86 \pm 0.04	0.87 \pm 0.02	0.87 \pm 0.03	0.88 \pm 0.02
	35	33.45 \pm 1.6	33.97 \pm 1.7	35.24 \pm 1.5	35.63 \pm 1.6	35.82 \pm 1.4	0.84 \pm 0.04	0.84 \pm 0.05	0.86 \pm 0.02	0.86 \pm 0.03	0.87 \pm 0.02
	50	32.36 \pm 1.6	31.28 \pm 1.8	34.23 \pm 1.6	34.44 \pm 1.6	34.87 \pm 1.5	0.83 \pm 0.04	0.82 \pm 0.05	0.84 \pm 0.03	0.84 \pm 0.03	0.86 \pm 0.02
Chest	10	38.67 \pm 1.1	38.82 \pm 1.1	39.61 \pm 1.1	39.23 \pm 1.1	39.75 \pm 1	0.94 \pm 0.02	0.94 \pm 0.02	0.96 \pm 0.02	0.95 \pm 0.02	0.96 \pm 0.02
	15	37.62 \pm 1.1	37.94 \pm 1.2	38.62 \pm 1.2	38.04 \pm 1.1	38.68 \pm 1.1	0.91 \pm 0.02	0.91 \pm 0.02	0.94 \pm 0.02	0.92 \pm 0.02	0.95 \pm 0.02
	25	35.48 \pm 0.9	35.86 \pm 1	37.11 \pm 1.2	36.57 \pm 1.2	37.16 \pm 1.1	0.86 \pm 0.02	0.87 \pm 0.02	0.89 \pm 0.03	0.88 \pm 0.02	0.90 \pm 0.02
X-ray	35	33.84 \pm 0.9	34.54 \pm 0.9	35.74 \pm 1.1	35.61 \pm 1.3	36.15 \pm 1.2	0.84 \pm 0.02	0.84 \pm 0.02	0.86 \pm 0.03	0.86 \pm 0.03	0.88 \pm 0.02
	50	32.16 \pm 0.7	32.97 \pm 1	34.45 \pm 1.2	34.69 \pm 1	34.95 \pm 1	0.83 \pm 0.03	0.83 \pm 0.03	0.84 \pm 0.03	0.84 \pm 0.03	0.86 \pm 0.03

From the data presented in Table 1, it is evident that AsymUNet shows significant enhancements in both PSNR and SSIM measurements when compared with the classical BM3D algorithm. This indicates that AsymUNet is able to make effective use of extensive medical image datasets to obtain various denoising capabilities, and can adaptively reduce noise through the application of these learned functions. Additionally, when compared with other deep-learning algorithms, AsymUNet consistently exhibits a superior performance across the evaluation metrics, thus underscoring its robust ability to effectively denoise medical images across various levels of noise.

To showcase the visual capabilities of AsymUNet, Figures 5 and 6 depict the results of AsymUNet alongside those of other comparable algorithms at a noise level of $\sigma = 25$. The figures illustrate that AsymUNet outperforms the other three methods in terms of clarity and preservation of image details. Whether applied to the HAM10000 color dataset or the Chest X-ray grayscale dataset, AsymUNet effectively restores image edges and enhances contrast, thereby preserving important image details.

(2) Low-Dose CT Image Denoising: This section highlights the effectiveness of AsymUNet in reducing noise in LDCT medical images. The model's performance was evaluated using the Mayo Clinic LDCT Grand Challenge dataset from the AAPM. Table 2

presents the test results comparing AsymUNet to WGAN [26], EDCNN [22], REDCNN [21], and Eformer [23] on the AAPM dataset. The results indicate that AsymUNet outperforms all comparison algorithms, underscoring its superior denoising capability for medical images. Additionally, Figure 7 visually demonstrates AsymUNet’s performance on the AAPM dataset, showing its ability to effectively reduce noise while preserving crucial image details, thereby significantly mitigating the risk of misdiagnosis in medical applications.

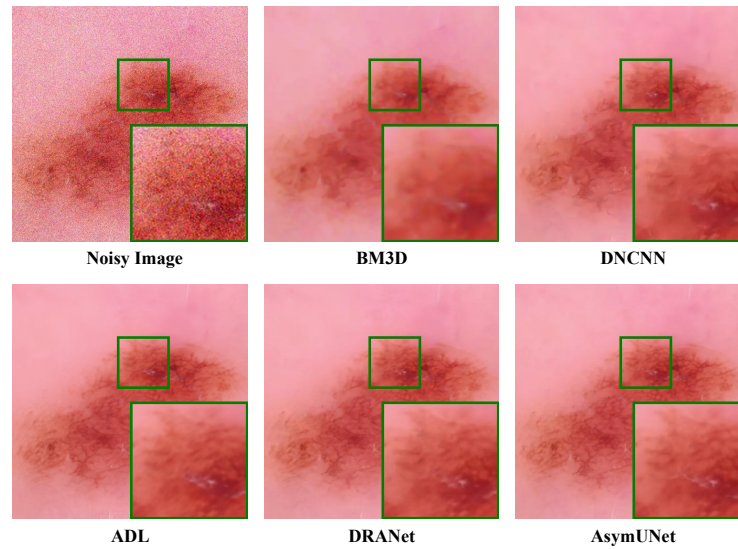


Figure 5. Denoising results on the HAM10000.

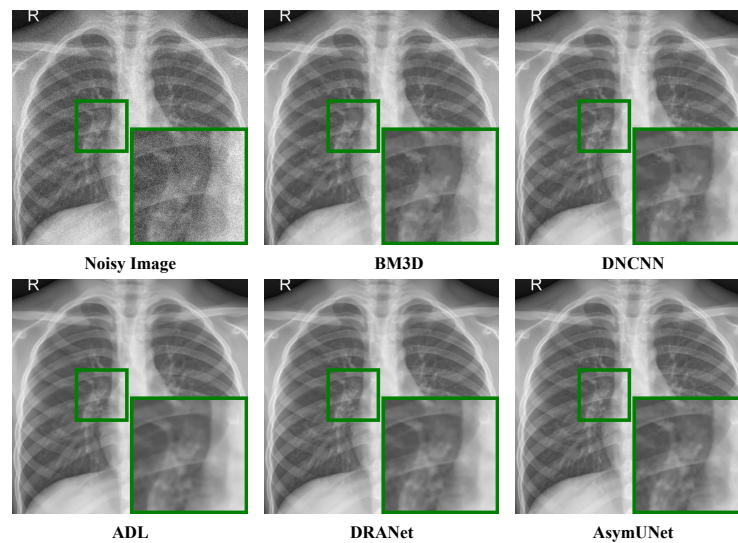


Figure 6. Denoising results on the Chest X-ray.

Table 2. Comparison of Denoising Results(AVERAGE ± STANDARD DEVIATION) on Low-Dose CT Images.

Dataset	PSNR					SSIM				
	WGAN	EDCNN	REDCNN	EFormer	AsymUNet	WGAN	EDCNN	REDCNN	EFormer	AsymUNet
AAPM	38.60 ± 1.7	42.08 ± 1.6	42.38 ± 1.6	43.48 ± 1.4	44.67 ± 1.4	0.9647 ± 0.0316	0.9866 ± 0.0287	0.9856 ± 0.0224	0.9852 ± 0.0292	0.9864 ± 0.0213

In real-world applications, the processing speed of neural networks is crucial. Therefore, we also compared the inference times of different models. Figure 8 shows the total inference time for each model on the HAM10000 test set. Among these models, AsymUNet

demonstrates the fastest inference speed. This advantage is attributed to its asymmetric U-Net architecture and MLP-based structure, which collectively reduce the computational time significantly.

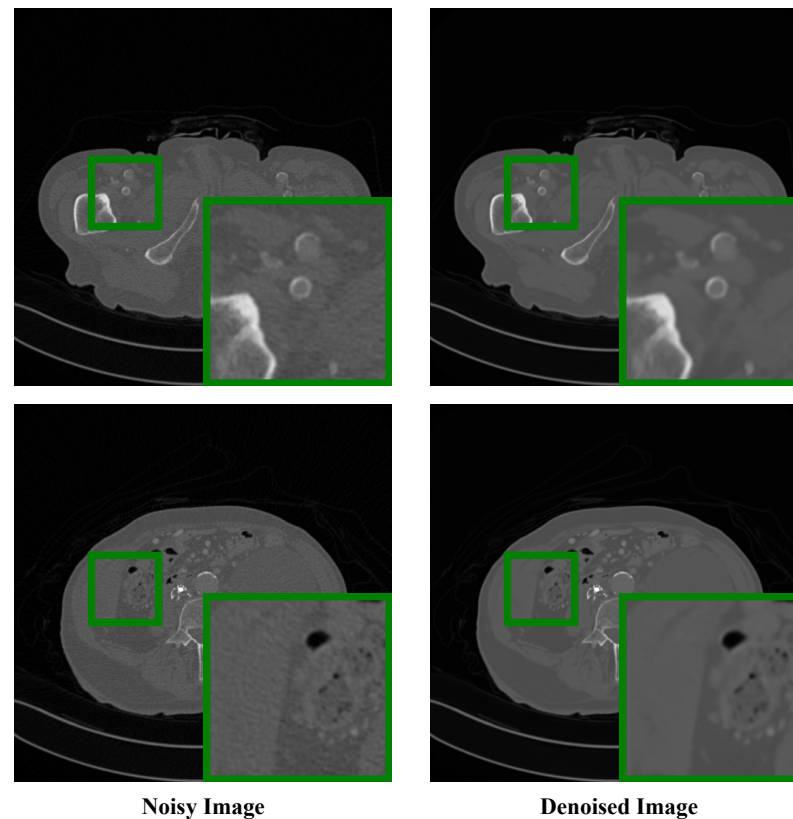


Figure 7. Denoising results on the AAPM.

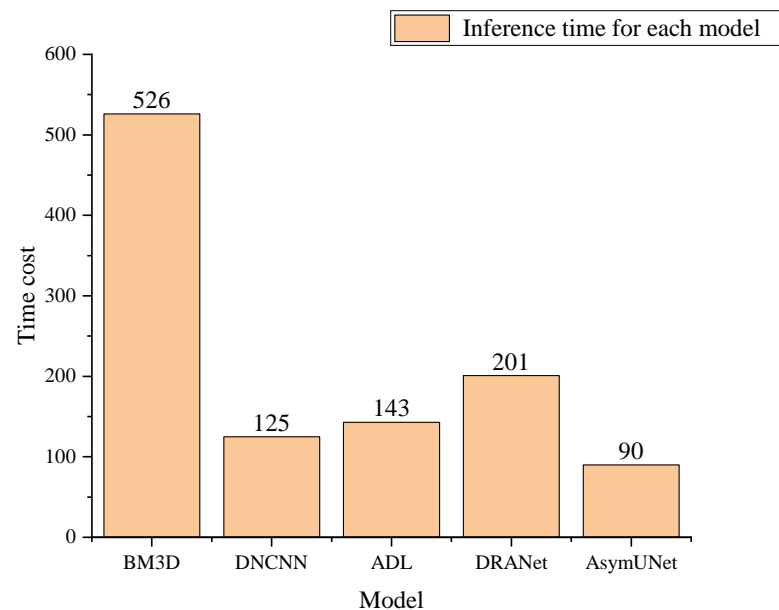


Figure 8. Comparison of the inference time for each model.

4.3. Ablation Experiments

To further investigate the effectiveness of various components, this section carried out several ablation experiments. All of the experiments were performed using the Chest X-ray dataset, utilizing a consistent progressive training strategy. For testing, the test set from

the Chest X-ray dataset was utilized, and all of the results were evaluated in the presence of Gaussian noise with a noise level of $\sigma = 25$. The influence of each component will be discussed in the subsequent sections.

(1) Multiscale Feature Fusion Module: In this subsection, we conducted experiments where the skip connection part of the MSFFM in AsymUNet was removed, resulting in a model referred to as Baseline. We compared the experimental results of Baseline, Baseline + Skip Connection (SC), and Baseline + MSFFM. Table 3 illustrates that incorporating MSFFM leads to better metrics compared with using only skip connections. This highlights that integrating feature maps across multiple scales and transmitting them to the decoder allows the model to effectively utilize image features, thereby achieving a superior denoising performance.

Table 3. Ablation experiments for the MSFFM.

NetWork	PSNR (dB)	SSIM
Baseline (a)	36.14	88.62
Baseline+SC (b)	36.85	90.34
Baseline+MSFFM (c)	37.16	90.73

(2) Spatial Rearrangement Multilayer Perceptron Block: To assess the effectiveness of SRCMLP, we conducted an experiment where SRCMLP was removed from AsymUNet and replaced with a standard MLP, denoted as W/O SRCMLP. As indicated in Table 4, the denoising performance of the neural network notably decreased after the removal of SRCMLP.

Table 4. Ablation experiments for the SRCMLP.

NetWork	PSNR (dB)	SSIM
AsymUNet	37.16	90.73
W/O SRCMLP	37.19	89.65

(3) Asymmetric U-Net: Table 5 presents the computational load and denoising performance results comparing the Asymmetric U-Net with the Symmetric U-Net. It is evident that the asymmetric variant reduced the computational load by nearly a quarter while maintaining an equivalent denoising performance compared with the symmetric counterpart.

Table 5. Ablation experiments for the asymmetric U-Net.

NetWork	PSNR (dB)	SSIM	FLOPs (G)	Params (M)
SymUNet	37.18	90.74	220.93	20.52
AsymUNet	37.16	90.73	177.44	18.20

5. Conclusions

This paper presents AsymUNet, an asymmetric multilayer perceptron U-Net designed to improve the denoising performance while reducing the inference time. AsymUNet utilizes an asymmetric U-Net architecture, which contributes to its efficiency gains. Additionally, the model incorporates SRCMLP as a foundational block, enhancing performance metrics and accelerating inference by utilizing RCMLP for local information extraction and CMLP for global information extraction. Furthermore, taking inspiration from the similarities in structure and texture across multiscale and same-scale feature maps, the paper introduces an MSFFM. This module enhances information flow between the encoder and decoder, thereby improving the overall denoising effectiveness. By combining these advancements, AsymUNet achieves state-of-the-art performance in both color/gray image denoising and LDCT image denoising tasks. Specifically, on the AAPM dataset, the proposed AsymUNet demonstrated a superior performance, achieving PSNR and SSIM scores

of 44.67 dB and 0.9864, respectively. Notably, it exhibited an average PSNR improvement of 1.19 dB over its closest rival (EFormer).

Author Contributions: Conceptualization, Y.C. and J.J.; methodology, X.H. and Y.C.; software, Z.G. and H.Y.; validation, X.H., Z.G. and H.Y.; formal analysis, Y.C., Z.G. and J.J.; investigation, Y.C. and J.J.; resources, Y.C.; data curation, X.H., Z.G. and H.Y.; writing—original draft preparation, Y.C. and X.H.; writing—review and editing, Y.C., X.H. and J.J.; visualization, Z.G. and H.Y.; supervision, Y.C. and J.J.; project administration, J.J.; funding acquisition, Y.C. All authors have read and agreed to the published version of the manuscript.

Funding: This work was supported in part by the Universities Directly Under the Inner Mongolia Autonomous Region Funded by the Fundamental Research Fund Project under Grant JY20220157, in part by the National Natural Science Foundation of China under Grant 62001236, in part by the Natural Science Foundation of the Jiangsu Higher Education Institutions of China under Grant 20KJA520003, in part by the Six Talent Peaks Project of Jiangsu Province under Grant JY-051.

Data Availability Statement: Publicly available datasets were analyzed in this study. Chest X-ray datasets can be found here: <https://www.kaggle.com/datasets/paultimothymooney/chest-xray-pneumonia> (accessed on 6 December 2023). HAM10000 datasets can be found here: <https://dataverse.harvard.edu/dataset.xhtml?persistentId=doi:10.7910/DVN/DBW86T> (accessed on 16 December 2023). AAPMT datasets can be found here: <https://www.aapm.org/GrandChallenge/LowDoseCT/> (accessed on 26 December 2023).

Conflicts of Interest: The authors declare no conflicts of interest.

References

- Zhang, K.; Zuo, W.; Chen, Y.; Meng, D.; Zhang, L. Beyond a gaussian denoiser: Residual learning of deep cnn for image denoising. *IEEE Trans. Image Process.* **2017**, *26*, 3142–3155. [CrossRef] [PubMed]
- Wu, W.; Liu, S.; Xia, Y.; Zhang, Y. Dual residual attention network for image denoising. *Pattern Recognit.* **2024**, *149*, 110291. [CrossRef]
- Jifara, W.; Jiang, F.; Rho, S.; Cheng, M.; Liu, S. Medical image denoising using convolutional neural network: A residual learning approach. *J. Supercomput.* **2019**, *75*, 704–718. [CrossRef]
- Zhang, Y.; Dou, Y.; Zhao, L.; Jiao, Y.; Guo, D. EDUNet++: An Enhanced Denoising Unet++ for Ice-Covered Transmission Line Images. *Electronics* **2024**, *13*, 2085. [CrossRef]
- Yang, L.; Li, Z.; Ge, R.; Zhao, J.; Si, H.; Zhang, D. Low-dose CT denoising via sinogram inner-structure transformer. *IEEE Trans. Med. Imaging* **2022**, *42*, 910–921. [CrossRef] [PubMed]
- Song, M.; Wang, W.; Zhao, Y. A Dynamic Network with Transformer for Image Denoising. *Electronics* **2024**, *13*, 1676. [CrossRef]
- Zhang, J.; Shangguan, Z.; Gong, W.; Cheng, Y. A novel denoising method for low-dose CT images based on transformer and CNN. *Comput. Biol. Med.* **2023**, *163*, 107162. [CrossRef]
- Huang, Z.; Zhang, J.; Zhang, Y.; Shan, H. DU-GAN: Generative adversarial networks with dual-domain U-Net-based discriminators for low-dose CT denoising. *IEEE Trans. Instrum. Meas.* **2021**, *71*, 1–12. [CrossRef]
- Xu, M.; Lv, P.; Li, M.; Fang, H.; Zhao, H.; Zhou, B.; Lin, Y.; Zhou, L. Medical image denoising by parallel non-local means. *Neurocomputing* **2016**, *195*, 117–122.
- Diwakar, M.; Kumar, M. CT image denoising using NLM and correlation-based wavelet packet thresholding. *IET Image Process.* **2018**, *12*, 708–715. [CrossRef]
- Astari, F.M.; Mulyantoro, D.K.; Indrati, R. Analysis of BM3D Denoising Techniques to Improvement of Thoracal MRI Image Quality; Study on Low Field MRI. *J. Med. Imaging Radiat. Sci.* **2022**, *53*, S24. [CrossRef]
- Dabov, K.; Foi, A.; Katkovnik, V.; Egiazarian, K. Image denoising with block-matching and 3D filtering. In *Image Processing: Algorithms and Systems, Neural Networks, and Machine Learning*; SPIE: Bellingham, WA, USA, 2006; Volume 6064, pp. 354–365.
- Liu, P.; El Basha, M.D.; Li, Y.; Xiao, Y.; Sanelli, P.C.; Fang, R. Deep evolutionary networks with expedited genetic algorithms for medical image denoising. *Med. Image Anal.* **2019**, *54*, 306–315. [CrossRef]
- Rawat, S.; Rana, K.; Kumar, V. A novel complex-valued convolutional neural network for medical image denoising. *Biomed. Signal Process. Control* **2021**, *69*, 102859. [CrossRef]
- Dong, G.; Ma, Y.; Basu, A. Feature-guided CNN for denoising images from portable ultrasound devices. *IEEE Access* **2021**, *9*, 28272–28281. [CrossRef]
- 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 Trans. Med. Imaging* **2021**, *41*, 407–419. [CrossRef]
- Sharif, S.M.A.; Naqvi, R.A.; Biswas, M. Learning medical image denoising with deep dynamic residual attention network. *Mathematics* **2020**, *8*, 2192. [CrossRef]

18. Ghahremani, M.; Khateri, M.; Sierra, A.; Tohka, J. Adversarial distortion learning for medical image denoising. *arXiv* **2022**, arXiv:2204.14100.
19. Zhang, P.; Liu, Y.; Gui, Z.; Chen, Y.; Jia, L. A region-adaptive non-local denoising algorithm for low-dose computed tomography images. *Math. Biosci. Eng. MBE* **2023**, *20*, 2831–2846. [[CrossRef](#)]
20. Chen, L.L.; Gou, S.P.; Yao, Y.; Bai, J.; Jiao, L.; Sheng, K. Denoising of low dose CT image with context-based BM3D. In Proceedings of the 2016 IEEE Region 10 Conference (TENCON), Singapore, 22–25 November 2016; pp. 682–685. [[CrossRef](#)]
21. 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 Trans. Med. Imaging* **2017**, *36*, 2524–2535. [[CrossRef](#)] [[PubMed](#)]
22. Liang, T.; Jin, Y.; Li, Y.; Wang, T. Edcnn: Edge enhancement-based densely connected network with compound loss for low-dose ct denoising. In Proceedings of the 15th IEEE International Conference on Signal Processing (ICSP), Beijing, China, 6–9 December 2020; pp. 193–198.
23. Luthra, A.; Sulakhe, H.; Mittal, T.; Iyer, A.; Yadav, S. Eformer: Edge enhancement based transformer for medical image denoising. *arXiv* **2021**, arXiv:2109.08044.
24. Wolterink, J.M.; Leiner, T.; Viergever, M.A.; Išgum, I. Generative adversarial networks for noise reduction in low-dose CT. *IEEE Trans. Med. Imaging* **2017**, *36*, 2536–2545. [[CrossRef](#)]
25. Ma, Y.; Yan, Q.; Liu, Y.; Liu, J.; Zhang, J.; Zhao, Y. StruNet: Perceptual and low-rank regularized transformer for medical image denoising. *Med. Phys.* **2023**, *50*, 7654–7669. [[CrossRef](#)]
26. 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 Trans. Med. Imaging* **2018**, *37*, 1348–1357. [[CrossRef](#)]
27. Li, Y.; Fan, Y.; Xiang, X.; Demandolx, D.; Ranjan, R.; Timofte, R.; Van Gool, L. Efficient and explicit modelling of image hierarchies for image restoration. In Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, Vancouver, BC, Canada, 17–24 June 2023; pp. 18278–18289.
28. Tolstikhin, I.O.; Houlsby, N.; Kolesnikov, A.; Beyer, L.; Zhai, X.; Unterthiner, T.; Yung, J.; Steiner, A.; Keysers, D.; Uszkoreit, J.; et al. Mlp-mixer: An all-mlp architecture for vision. *Adv. Neural Inf. Process. Syst.* **2021**, *34*, 24261–24272.
29. Tschandl, P.; Rosendahl, C.; Kittler, H. The HAM10000 dataset, a large collection of multi-source dermatoscopic images of common pigmented skin lesions. *Sci. Data* **2018**, *5*, 180161. [[CrossRef](#)]
30. Kermany, D.S.; Goldbaum, M.; Cai, W.; Valentim, C.C.; Liang, H.; Baxter, S.L.; McKeown, A.; Yang, G.; Wu, X.; Yan, F.; et al. Identifying medical diagnoses and treatable diseases by image-based deep learning. *Cell* **2018**, *172*, 1122–1131. [[CrossRef](#)]
31. 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. *Med. Phys.* **2017**, *44*, e339–e352. [[CrossRef](#)]

Disclaimer/Publisher’s Note: The statements, opinions and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions or products referred to in the content.