


## Research Article

# Denoising Chest LDCT Images Using a Variant Unet Based on Hybrid Attention Mechanism and Diffusion Model

Xiaoyu Chen<sup>1</sup>, Yiming Chen<sup>2</sup>, Xi Liu<sup>2</sup>, Longling Fan<sup>1\*</sup> , Dechuan Zhang<sup>2</sup>

<sup>1</sup>Key Laboratory of Applied Statistics and Data Analysis of Department of Education of Yunnan Province, Research Center for Mathematics and Interdisciplinary Sciences, Faculty of Science, Kunming University of Science and Technology, Kunming, 650500, China

<sup>2</sup>Department of Radiology, Chongqing Hospital of Traditional Chinese Medicine, Chongqing, 400021, China  
E-mail: fan\_longling2008@126.com

**Received:** 26 June 2025; **Revised:** 18 August 2025; **Accepted:** 18 September 2025

**Abstract:** Background: Computed Tomography (CT) may increase cancer risk due to high-dose radiation exposure. Although Low-Dose CT (LDCT) reduces radiation, its noise compromises diagnostic accuracy. This study deeply optimized LDCT image denoising techniques to balance image quality and noise reduction effects. Methods: This study investigates three LDCT image denoising methods combining Unet and diffusion models. The UDiff model enhances image quality by adding Gaussian noise and employing Unet for denoising. The UAdiff model omits the noise-adding step and introduces a hybrid attention mechanism to strengthen feature extraction. The DUPAnet model utilizes LDCT images as prior information, integrating a hybrid attention mechanism for efficient denoising. Experiments were conducted using chest phantom data acquired from multi-vendor CT scanners, optimizing the models with a hybrid loss function, and evaluating denoising performance through various metrics during 2024. Results: The UDiff model exhibits denoising blurring and poor stability when processing similar phantom data. The UAdiff model weakens attention feature capture, leading to detail loss and performance fluctuations. The DUPAnet model shows the best performance in both denoising and detail preservation, with the best performance in both objective metrics Peak Signal-to-Noise Ratio (PSNR) and SSIM (30.37 and 0.87) and subjective images. It enhances the denoising ability of LDCT images, has strong adaptability to various scanning parameters, and has excellent generalization performance. Conclusion: The improved Unet model integrates LDCT priors and a hybrid attention mechanism, effectively denoising while preserving chest LDCT image details, thereby enhancing image quality and generalization capability. Future work will focus on optimizing the algorithm's generalization performance, exploring multimodal fusion and real-time processing to advance clinical applications.

**Keywords:** Low-Dose Computed Tomography (LDCT), denoise, unet, diffusion model, attention module

**MSC:** 68T07, 68U10, 92C55

## 1. Introduction

Computed Tomography (CT) is an advanced medical imaging technology that merges X-ray imaging with sophisticated computer algorithm [1, 2]. Its significant technical advantages such as a wide detection range, rapid imaging speed

and high-density resolution. It is adept at clearly displaying pathological information, including tumors, fractures and vascular lesions, and has become an indispensable tool in modern medical diagnosis. However, although CT technology is of great value in diagnosis, the radiation dose generated in the scanning process has always been the focus of the public and the medical community. A number of studies have pointed out that the radiation dose of a single CT scan can be as high as 100 to 1,000 times that of conventional X-ray films. Excessive exposure will increase the risk of various types of cancers by 0.4%, and there is a linear relationship between the radiation dose and cancer risk. Frequent CT scans significantly increase the potential cancer risk [3–6]. In order to reduce the negative impact of radiation dose on the subjects, Low-Dose CT (LDCT) technology has been developed. In clinic, the radiation is usually reduced by reducing the tube current, voltage and other scanning parameters [7], but the noise generated in LDCT images can interfere with lesion recognition, including inaccurate assessment of lesion size, blurring of important anatomical features, and masking of small tumors, thereby affecting the accuracy of clinical diagnosis. In view of this, the current research faces two main challenges: balancing radiation dose and image quality, and optimizing denoising strategies. A large amount of research has attempted to introduce advanced reconstruction and denoising techniques [8, 9], model based iterative reconstruction methods have shown significant advantages in noise suppression, while deep learning algorithms have also achieved certain results in learning complex noise distributions and reconstructing fine structural details. However, these advances have not fully addressed the core challenge of denoising low-dose CT images, details and structures are still susceptible to damage even with reduced doses. Deep learning methods suffer from overfitting when applied across devices and populations. The sensitivity of detection varies among different types of lesions, such as small pulmonary nodules. Therefore, this research direction is worth further exploration, especially in terms of interpretability, generalization ability, and multi center clinical validation of methods, in order to gradually improve their clinical generalizability and long-term safety.

Low dose CT reconstruction techniques are mainly divided into three categories: Filtered Back Projection (FBP), Iterative Reconstruction (IR) and reconstruction based on Deep Learning (DLR). FBP is one of the most traditional and classic methods in LDCT image reconstruction [10]. However, when the dose is low, the FBP method falls short in noise suppression. IR method progressively converges towards the real image through iterative optimization, thus to more effectively suppress noise and reduce artifacts, but its calculation cost in iterative projection/back projection step remains high [11, 12]. DLR automatically identifies noise features in images and efficiently retrieves image details via massive learning from high-dose and LDCT image pairs. In particular, Convolutional Neural Network (CNN) performs well in denoising and has become the mainstream technology [13–16]. In 2015, the Unet network structure was proposed. Because of its unique symmetrical encoder decoder structure, combined with jump connection, it effectively retains high-resolution features and enhances spatial information transmission, and its simple and stable structure and high training efficiency, it has been widely used in the field of image processing, including tasks such as image segmentation and denoising, primarily for image detail processing [17, 18]. Heinrich et al. Discussed two CNN architectures, ResFCN and ResUNet, for denoising LDCT images. The results showed that ResUNet, a multi-scale U-shaped network architecture with residual connection, had better performance and performance than ResFCN [19]. However, Unet's performance is constrained by the local receptive field of convolution, making it challenging to capture the long-range dependencies. Additionally, high-frequency details are prone to loss during the down sampling process, leading to an overly smooth denoising effect. The strong assumption of noise distribution also limits the generalization ability of the model.

The diffusion model has received widespread attention due to its outstanding performance in advanced image generation. Compared with traditional CNNs, it boasts a superior global generation capability [20, 21]. Gupta et al. proposed an innovative anisotropic diffusion model, effectively solved the problem of step effect and edge blur in the conventional model through optimizing the diffusion coefficient and image dependent threshold parameters [22]. However, the diffusion model relies on multi-step iteration, which has low computational efficiency and complex training process, and may also produce artifacts or distorted details under extreme noise conditions. These methods have improved image quality to some extent in various scenarios, yet issues persist with insufficient or excessive denoising and insufficient preservation of detail structure. Both of them face the challenges of high data dependence, poor adaptability to mixed noise, and lack of explicability. In addition, the existing research mainly focuses on the denoising technology. Although this is the key to improve the image quality, the differences between different LDCT images and their impact on denoising methods are also important issues worthy of further discussion.

This study is based on the Unet architecture and diffusion model, integrating a mixed attention mechanism and prior information to construct a new image denoising model, and the effectiveness of the mixed attention mechanism, prior information guidance, and model structure optimization in improving the denoising effect of LDCT images were evaluated. Through multiple sets of chest phantom CT experiments with different scanning parameters and noise levels, the quantitative indicators of denoising results (such as signal-to-noise ratio, structural similarity) and subjective visual quality were systematically analyzed; and the performance differences of three techniques in processing different noise patterns and anatomical structures were explored. At the same time, the influence of key parameters such as scanning interlayer spacing and voltage on the denoising effect was revealed. This study aims to clarify the robustness and applicability of these improved methods in LDCT image processing, demonstrate their advantages in suppressing noise, improving detail preservation, and diagnostic usability, objectively analyze their limitations in computational complexity and generalization ability, and provide theoretical basis and practical guidance for LDCT image quality optimization technology.

## 2. Materials and methods

### 2.1 Low dose CT image denoising problem model

It is difficult to obtain the original projection data scanned by CT instrument. For the denoising of LDCT image, post-processing is usually carried out after image reconstruction, so as to transform the CT image denoising problem into image domain processing. Different from natural image restoration, the difficulty of LDCT image denoising is that its statistical characteristics are difficult to determine directly in the image domain [23], thereby constraining the effectiveness of traditional denoising methods. The methods based on Deep Learning (DL) are not limited by this type of noise dependence, because they rely on training samples rather than the statistical distribution of noise. Therefore, the modeling of LDCT denoising problem is as follows [16].

Assume  $X \in \mathbb{R}^{m \times n}$  is LDCT image,  $Y \in \mathbb{R}^{m \times n}$  is the corresponding normal dose CT image, the relationship between them will be defined as

$$X = \sigma(Y), \quad (1)$$

where  $\sigma : \mathbb{R}^{m \times n} \rightarrow \mathbb{R}^{m \times n}$  represents the complex degradation process of normal dose CT images due to noise pollution. Therefore, the denoising problem can be defined as the following function problem

$$\arg \min_f \|f(X) - Y\|_2^2, \quad (2)$$

where  $f$  is considered the optimal approximation and can be estimated based on deep learning methods.

### 2.2 Chest low dose CT image denoising method based on UDiff model

The UDiff model is extended and improved on the basis of the U-shaped structure of the classic Unet, and the diffusion model is introduced to improve the denoising performance. Gaussian noise is usually added in the forward process, and the overall structure is shown in Figure 1. The Unet structure is composed of encoder and decoder. This design can effectively solve the problem of loss of details caused by down sampling operation, so as to help the network locate more accurately [24]. The encoding part gradually extracts image features through a series of convolution and pooling operations, while the decoding part restores the spatial resolution of the image through deconvolution. Each layer of decoder is combined with the corresponding encoder through hopping connection to retain details, which is very important for image denoising task.

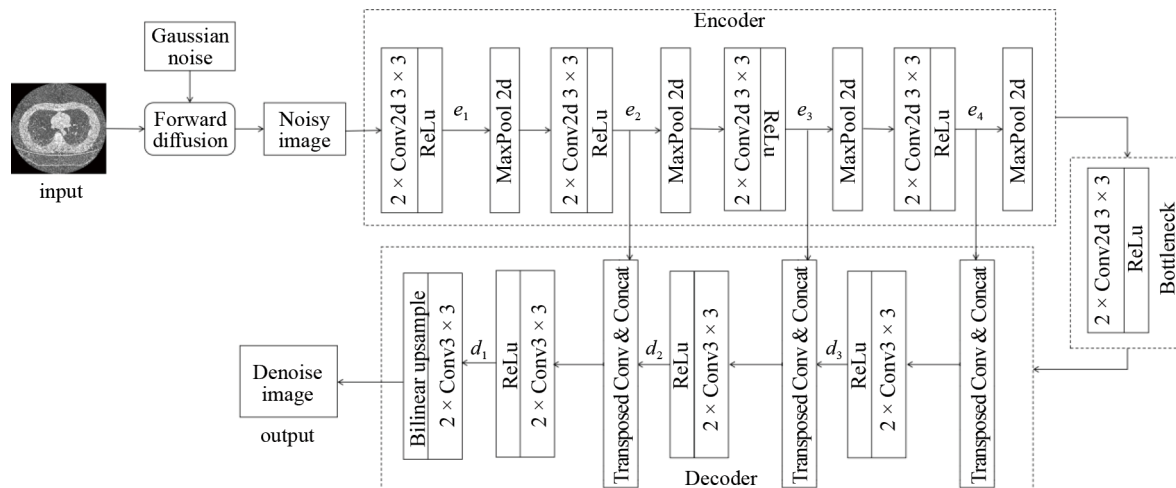


Figure 1. Overall architecture of UDiff model

In recent years, the diffusion model has also played an outstanding role in medical image denoising [25, 26]. It generates new data by adding the diffusion process of noise simulation data and the reverse process denoising. The main core includes the forward process and the reverse process. The former gradually adds noise until the image becomes a complete noise image. The latter relies on the model to learn how to gradually remove the noise, so as to restore the original image. Its basic structure is shown in Figure 2.

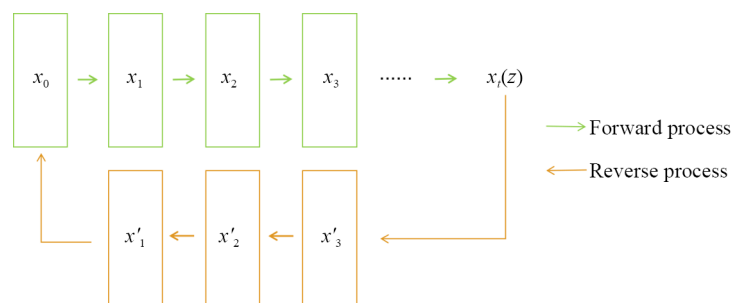


Figure 2. Diffusion model

The forward process gradually adds noise to the data, given an original image  $x_0$  add noise to the image at each time step  $t$ , forming a noisy image  $x_t$  where  $t$  is from 1 to  $T$ , the added noise is determined according to the preset noise schedule. With the increase of time  $t$ , the noise component increases gradually. At the maximum time step  $T$ , it gradually approaches a pure noise state [25]. The forward process (single step) of this method can be expressed as:

$$x_t = \sqrt{1 - \beta_t} \cdot x_{t-1} + \sqrt{\beta_t} \cdot \varepsilon, \quad (3)$$

Where  $\beta_t$  is a predefined noise scheduling parameter, defined as  $\beta_1 = 0.0001$ ,  $\beta_2 = 0.02$ .  $\varepsilon$  is a noise sampled from the standard normal distribution  $N(0, 1)$ .

The inverse process is the key to the diffusion model. Through the model learning denoising method, we can reverse the process of adding noise in the training process, so as to eliminate the noise in stages, and finally restore the original clean image. The target of the reverse process is  $p(x_{t-1}|x_t)$ . Given a noisy image  $x_t$ , the model attempts to predict the

definition of the last time step  $x_{t-1}$ , and finally recover the original image from the noisy image. The inverse process of this model uses the Unet network structure for denoising. The basic Unet network structure is shown in Figure 3 [27].

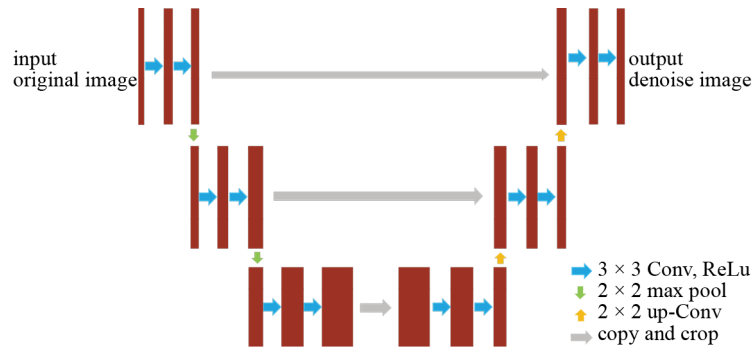


Figure 3. The structure of Unet

The part of the current image  $x_t$  predicted by the model to add noise is recorded as

$$\hat{\epsilon}_t = \text{model}(x_t, t), \quad (4)$$

At each time step  $t$ , the denoised image  $x_{t-1}$  can be calculated by the following formula

$$x_{t-1} = \frac{1}{\sqrt{\alpha_t}} \left( x_t - \frac{1 - \alpha_t}{\sqrt{1 - \alpha_{t-1}}} \cdot \hat{\epsilon}_t \right), \quad (5)$$

where  $\alpha_t = 1 - \beta_t$ , and  $\beta_t$  is the noise scheduling in the forward process. The final reverse process can be expressed as

$$p_{\theta}(x_{t-1} | x_t) = N \left( x_{t-1}; \mu_{\theta}(x_t, t), \sum_{\theta}(x_t, t) \right), \quad (6)$$

where  $\mu_{\theta}(x_t, t)$  is the denoising mean value predicted by the Unet network, and  $\sum_{\theta}(x_t, t)$  is the denoising variance.

### 2.3 Chest low dose CT image denoising method based on UAdiff model

Because the forward noise adding process of the diffusion model is controllable, that is, the added noise is known. In the UDiff model, the Gaussian noise introduced by the forward process has certain pertinence according to the given qualitative noise, but this has limitations in the research and discussion. However, in order to control the radiation dose, the LDCT image has complex and uncertain noise by adjusting the scanning voltage, current and other factors. We want to denoise directly without additional noise, so on the basis of the above method, we directly omit the forward noise adding process of the diffusion model. Although the U-shaped structure network performs well in feature extraction, it usually lacks a direct mechanism to focus the key information of the image. In LDCT images, noise interference often masks important structures and details. Since the initial proposal of attention mechanism in 2014, it has been widely used in various fields with in-depth research. Attention mechanism can adaptively focus on important areas of the image. In image denoising, a large number of studies have confirmed that it helps to distinguish between noise and detail features,

improve the denoising effect, and retain the image detail structure [28, 29]. After removing the noise adding process, we consider introducing a hybrid attention mechanism in the decoding stage to assist image denoising, mainly including the channel attention module and the spatial attention module, through which we can comprehensively improve the ability to focus on the feature map. The following shows the UAdiff model that omits the forward noise adding process and adds the mixed attention mechanism in the coding phase. Its overall network structure and the mixed attention module are shown in Figure 4. Then, we further introduce the main components of mixed attention.

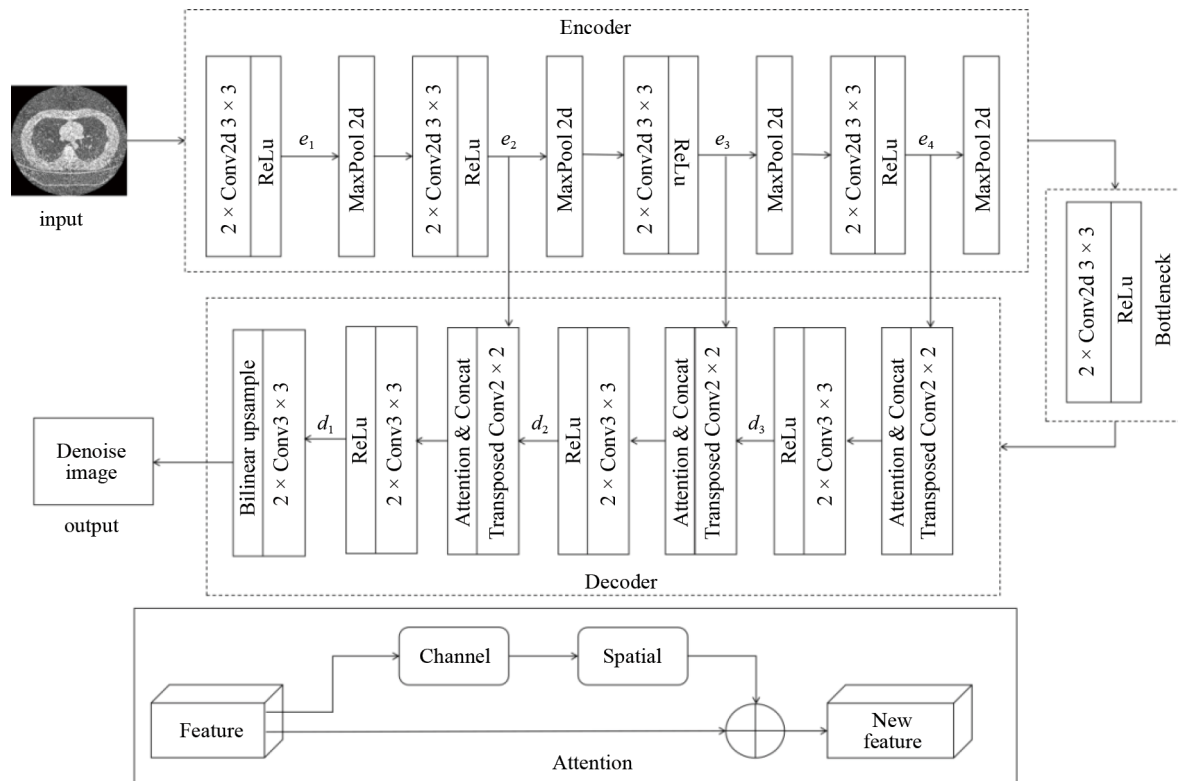


Figure 4. Overall architecture of UAdiff model and mix-attention module

### 2.3.1 Channel attention module

For any input feature map  $F \in R^{L \times W \times H}$ , first use the compression operation to obtain the overall information of the feature image, and then use the aggregation operation to obtain the correlation between channels [30]. The compression operation includes global maximum pooling and global average pooling to generate the characteristic information of two different spaces. The aggregation operation uses the feature information of two different spaces to obtain the overall information of the feature layer using the multi hidden neural network with self coding structure and shared parameters.

In the hybrid attention mechanism adopted in this section, the channel attention model is specifically described as follows. Given the input characteristic graph  $X$ , first, two one-dimensional statistical features  $F_{\text{avg}}$  and  $F_{\text{max}}$  are extracted respectively through the global average pooling and maximum pooling operations. Then, the two features are processed by two shared full connection layers (including ReLU and Sigmoid activation function) to generate channel weight vectors. Finally, each channel of the feature map is multiplied by the corresponding attention weight to obtain a weighted feature map. The structure of the channel attention module is shown in Figure 5. Its mathematical expression can be expressed as

$$CA(X) = X \cdot [\sigma(f(F_{\text{avg}}(X))) + \sigma(f(F_{\text{max}}(X)))], \quad (7)$$

where  $\sigma$  is the Sigmoid activation function,  $f$  is the full connection layer,  $F_{\text{avg}}(X)$  and  $F_{\text{max}}(X)$  is the statistical feature obtained through global average pooling and maximum pooling.

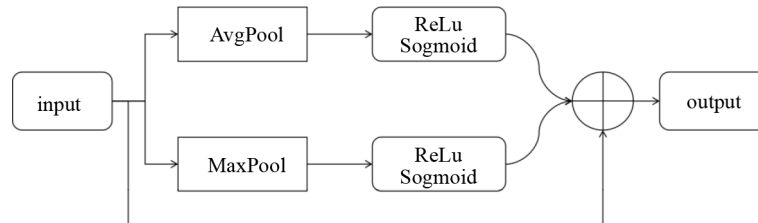


Figure 5. Structure of channel attention module

### 2.3.2 Spatial attention module

Different from the channel attention module, the spatial attention module enhances attention to spatial regions by calculating the weight of each position in the spatial dimension [30]. In the hybrid attention mechanism described in this section, the spatial attention module first calculates the global average pooling and maximum pooling of the feature map in the spatial dimension, then merges these two into a single channel feature map, and finally generates spatial attention weights through convolution operation. The specific structure of the spatial attention module is shown in Figure 6. The mathematical expression is as follows:

$$SA(X) = X \cdot \sigma(\text{Conv}([\text{AvgPool}(X), \text{MaxPool}(X)])), \quad (8)$$

where  $\text{AvgPool}(X)$  and  $\text{MaxPool}(X)$  are pooling operations in the spatial dimension,  $\text{Conv}$  is the convolution layer, and  $\sigma$  uses the Sigmoid activation function.

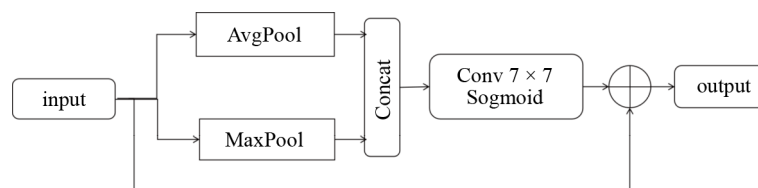


Figure 6. Structure of spatial attention module

## 2.4 Chest low dose CT image denoising method based on DUPAnet model

After a series of preliminary experimental analyses, we found that the main function of adding attention mechanism in the model is to effectively distinguish between noise and detail features. If the forward process of adding noise is directly omitted, it will result in a lack of guidance for noise feature information, which will significantly affect the denoising effect. Drawing on the idea of using Magnetic Resonance (MR) prior information to improve the denoising performance of Positron Emission Computed Tomography (PET) images in reference [31], we incorporate LDCT images as prior information into the forward process to guide subsequent denoising tasks. On the basis of the UAdiff model, chest LDCT



images are introduced as noise prior information and input in dual channel form with corresponding normal dose images, while retaining other structures, to construct the DUPAnet model. The overall network structure is shown in Figure 7.

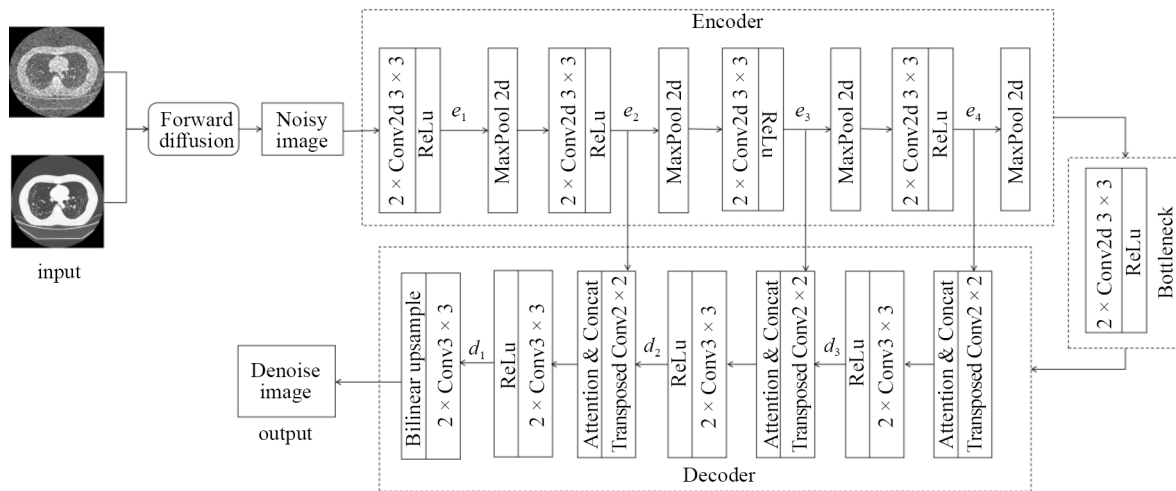


Figure 7. Overall architecture of DUPAnet model

## 2.5 Loss function

Generally, the denoising task of LDCT images is mainly to construct a loss function by minimizing the distance between Normal-Dose CT (NDCT) images and processed LDCT images, which can be referred to as the Mean Squared Error (MSE) loss function,

$$L_{\text{MSE}} = \arg \min \frac{1}{N} \sum_{i=1}^N \|y_{\text{pred}, i} - y_{\text{true}, i}\|_2^2, \quad (9)$$

where  $y_{\text{pred}, i}$  represents the denoised image,  $y_{\text{true}, i}$  represents the original normal dose image, and  $N$  is the number of images. MSE is sensitive to large errors, but may be overly sensitive to outliers.

In addition, the  $L_1$  loss function (absolute error) is widely used in many tasks, especially in image reconstruction, denoising, image compression, and model generation tasks. In many denoising algorithms [32], the  $L_1$  loss function is used to measure the difference between the denoised image and the real image, and its mathematical expression is

$$L_1 = \frac{1}{n} \sum_{i=1}^n \|y_i - \hat{y}_i\|_2^2, \quad (10)$$

where  $y_i$  is the pixel value in the real image,  $\hat{y}_i$  is the pixel value of the denoised image, and  $n$  is the total number of pixels in the image. Compared to the MSE loss function,  $L_1$  loss is less sensitive to significant errors (outliers) in the image and can compensate for some of the shortcomings of the MSE loss function. Consider combining the two.

Structural similarity is a standard for measuring the similarity between two images, which comprehensively considers the brightness, contrast, and structural information of the images, and can better reflect the structural differences of the images [33]. The Structural Similarity Index Measure (SSIM) loss used here is achieved by subtracting the SSIM value from 1, with the aim of minimizing structural differences in the image. Typically, SSIM is defined as



$$\text{SSIM}(x, y) = \frac{(2\mu_x\mu_y + c_1)(2\sigma_{xy} + c_2)}{(\mu_x^2 + \mu_y^2 + c_1)(\sigma_x^2 + \sigma_y^2 + c_2)} \quad (11)$$

where  $\mu_x$  and  $\mu_y$  are the mean of two images,  $\sigma_x^2$  and  $\sigma_y^2$  are the variance of the images,  $\sigma_{xy}$  is the covariance of the two images,  $c_1$  and  $c_2$  are a constant used to avoid zero division errors. The range of this value  $[-1, 1]$  is usually normalized to  $[0, 1]$ . The SSIM loss function is ultimately expressed as follows:

$$L_{\text{SSIM}} = 1 - \text{SSIM}. \quad (12)$$

The above three methods all adopt a mixed loss function optimization model in terms of loss function, combining MSE loss,  $L_1$  loss, and SSIM loss. MSE loss penalizes pixel level errors,  $L_1$  loss emphasizes sensitivity to sparse errors, while SSIM loss helps maintain the structural information of the image, especially ensuring the integrity of details and textures during denoising. The final loss function is defined as:

$$L_{\text{LOSS}} = \omega_1 L_{\text{MSE}} + \omega_2 L_1 + \omega_3 L_{\text{SSIM}} \quad (13)$$

where  $\omega_1$ ,  $\omega_2$  and  $\omega_3$  are the weighting coefficients of  $L_{\text{MSE}}$ ,  $L_1$  and  $L_{\text{SSIM}}$ , with weights of 0.5, 0.3, and 0.2 respectively [34].

## 2.6 Experimental phantom datas

The simulated human chest phantom comes from the “Lungman N1” model produced by Kyoto Kagaku in Japan, with a size of approximately 430 mm × 400 mm × 480 mm, a chest circumference of approximately 940 mm, and a mass of approximately 18 kg. The phantom includes the simulated chest wall, diaphragm, mediastinum, heart, lungs, and pulmonary blood vessels. The space between the pulmonary blood vessels in the chest contains air, which is similar to that of a normal human body and has the same tissue radiation attenuation properties. The phantom has both arms raised upwards, in the same position as the patient during the actual scan.

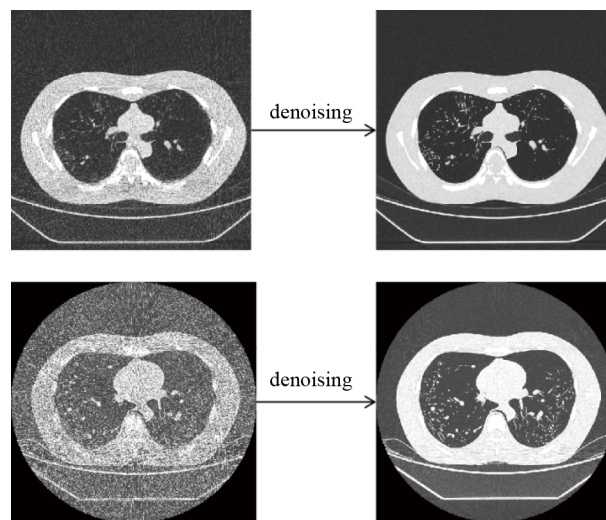
Select 6 CT scanners from different manufacturers and models (GE 660, GE 256, GE VCT produced by GE in the United States, United Imaging 710 and United Imaging 860 produced by China United Imaging Corporation, and Somatom Force produced by Siemens in Germany) to scan the simulated phantom from the lung apex to the lung floor. Adjustable scanning parameters encompass: Kv, mA/mAs, pitch, iteration parameters, speed Field of View (FOV), Layer number, layer thickness, detector width, and scanning time. Based on the principle of decreasing radiation dose from high to low, starting from the default parameters of the scanner for routine chest scanning, the technical parameters are adjusted step by step to achieve the goal of reducing dose. All parameter combinations are unique. The image reconstruction is completed on the machine host, mainly using bone algorithms to reconstruct lung window images and standard algorithms to reconstruct mediastinal window images. The specific situation of the phantom dataset is shown in Table 1.

**Table 1.** Specific situation of the phantom dataset used in the experiment

Model	Equipment	Scan sequence	Dose type	Kv	mA	mAs	Pitch	Recon option	Rotational time (s)	FOV (mm)	Number of images	Thickness/Interval (mm/mm)	Detector coverage (mm)	Total exposure time (s)	DLP (mGy·cm)
1	GE optima 660	$n = 1$	Normal	120	250-350	-	1.375	AsiR 50%	0.5	38.7	59	5/5	40	3.22	317.5
		$n = 15$	Low dose	80-120	10-50	-	1.375	AsiR 10%-100%	0.5/0.6/0.7	38.7	59	5/5; 10/10	40	3.22-5.97	3.55-56.12
2	GE revolution 160	$n = 1$	Normal	120	50-500	-	0.992	AsiR 20%	0.5	37	59	5/5	80	2.22	297.00
		$n = 12$	Low dose	70-120	14-90	-	0.508-1.531	AsiR 20%-80%	0.5	37	59	5/5	80	1.6-4.41	2.37-107.74
3	Philips iCT	$n = 1$	Normal	120	-	151	0.8	iDos 2	0.5	39	59	5/5	80	3.5	406.4
		$n = 4$	Low dose	80/120	10-30	-	0.8/1	iDos 2	0.33/0.5	39	59	5/5	40/80	2.9-3.9	7.7-84.5
4	United $\mu$ CT860	$n = 1$	Normal	120	110-248	188	1.093	3D.2	0.5	37	66	5/5	80	3	332.65
		$n = 12$	Low dose	60-120	21-50	19-23	1.093-1.231	3D.1	0.5	37	66	5/5	80	2.1-2.4	2.8-76.57
5	Siemens new force	$n = 1$	Normal	120	-	56	1	Does4	0.5	37	59	5/5	115	3.82	131.90
		$n = 10$	Low dose	70/80/Sn100	-	188	0.75-1.5	Does4	0.5	37	59	5/5	115	2.62-4.86	1.5-23.4
6	United $\mu$ CT860	$n = 1$	Normal	120	188	86	1.094	3D.7	0.5	37	59	5/5	80	2.3	320.88
		$n = 6$	Low dose	60/70/120	20-41	10-19	1.094-1.169	3D.7/3D.8	0.5	37	59	5/5	80	2.1-2.3	2.85-73.03
7	United $\mu$ CT710	$n = 1$	Normal	120	-	249	1.088	3D.5	0.5	37	59	5/5	40	4.2	388.30
		$n = 8$	Low dose	70/120	17-42	6-20	1.088-1.388	3D.5-3Dk.9	0.5	37	59	5/5	40	3.3-4.2	2.48-60.19
8	GE VCT	$n = 1$	Normal	120	180-400	-	0.984	AsiR 20%	0.4	37	59	5/5	40	3.42	330.77
		$n = 5$	Low dose	120	10	-	0.984/1.375	AsiR 30%-90%	0.4/0.5	37	59	5/5	40	2.57-3.42	2.61-10.4

## 2.7 Experiment situation

This article uses 8 sets of phantom data of chest normal dose CT and LDCT under different current and voltage conditions provided by Chongqing Traditional Chinese Medicine Hospital. Due to different scanning instruments, there are mainly two forms of imaging, as shown in Figure 8. Our training and validation sets randomly selected normal doses and corresponding low-dose sequences from these 8 sets of phantom data, totaling 1,920 images, and divided them in a 9 : 1 ratio. The test set randomly selects low-dose sequences from different models that did not participate in training for detection (each sequence of each model contains 59 to 350 images). The selection of some test sets avoids complete randomness and aims to control specific variables for subsequent comparative experimental discussions on CT images themselves.



**Figure 8.** LDCT images (left) and normal dose CT images (right) of Phantom 1 and Phantom 6

The operating environment for this experiment is Intel (R) Core (TM) i7-13700F CPU, with 64 GB RAM running memory and NVIDIA GeForce RTX 3060 Ti graphics card configuration. All experiments are built on the Pytorch framework. During the training process, the training batch is set to 4, the initial learning rate is set to  $1.0 \times 10^{-4}$ , and the model is optimized using the Adam optimizer. The losses of the training set and validation set are calculated separately and the results of the validation set are saved. The weights of the current model are saved for each epoch during the training process. To avoid overfitting, the model uses an early stopping strategy and saves the optimal model weights at the end of each epoch. During training, we evaluate the performance of the model on the validation set and measure its denoising effect using SSIM and Peak Signal-to-Noise Ratio (PSNR) metrics.

## 2.8 Evaluation metric

Use the three different methods mentioned above to denoise LDCT images generated from different phantoms, scan intervals, and voltage scans. The quality evaluation of denoising effect depends on subjective and objective evaluation indicators. Subjective evaluation includes random visual inspection of denoised images; objective evaluation is mainly based on indicators such as PSNR, SSIM, and Root Mean Square Error (RMSE). Specifically, the higher the PSNR score, the better the image quality; The closer the SSIM value is to 1, the better the image quality; The lower the RMSE value, the more ideal the result.

Peak signal-to-noise ratio is used to evaluate the similarity between reconstructed images and original images, and its formula is:

$$\text{PSNR} = 10 \cdot \log_{10} \left( \frac{\text{MAX}^2}{\text{MSE}} \right), \quad (14)$$

where MAX represents the maximum value of image pixels (255 for 8-bit images), which is the mean square error, the smaller the MSE, the smaller the difference between the denoised image and the original image, and the larger the PSNR value, indicating better denoising effect.

Structural similarity index measure as a supplement to PSNR, aims to address the issue of inconsistency between PSNR evaluation results and the image quality observed by the naked eye in certain scenarios. Specifically, sometimes the visual quality of the image is high, but the PSNR value is low, and conversely, some images with high PSNR values may not look as good to the eye. This difference is due to the limited sensitivity of the human eye to overall pixel errors in the image, especially to differences in edge details, while being relatively tolerant to differences in smooth areas. In addition, the human eye is more sensitive to changes in brightness than changes in color, which explains why many image and video compression techniques tend to sacrifice color information to preserve brightness details. SSIM comprehensively compares the similarity of brightness, contrast, and structural information, with a value range of 0-1. The closer the SSIM value of the denoised image is to 1, the higher its structural similarity with the original image and the better the denoising effect.

Root mean square error is a commonly used indicator to measure the difference between predicted values and true values. It is widely used in fields such as image quality assessment, image reconstruction, denoising, and compression, serving as a key indicators for evaluating the effectiveness of image processing algorithms [35]. In image processing and computer vision, RMSE can be used to measure the differences or similarities between two images. By calculating the error of each pixel value and squaring it, RMSE can provide a quantitative measure of image differences. Assuming there are two images  $I_1$  and  $I_2$ , one representing the original image and the other representing the processed image, each with a size of  $M \times N$ , where  $M$  and  $N$  are the width and height of the images, and the pixels are  $p_1(i, j)$  and  $p_2(i, j)$ , the calculation formula for RMSE is as follows

$$\text{RMSE} = \sqrt{\frac{1}{M \times N} \sum_{i=1}^M \sum_{j=1}^N (p_1(i, j) - p_2(i, j))^2}. \quad (15)$$

### 3. Results

#### 3.1 Experimental results and analysis

##### 3.1.1 Analysis of denoising results of chest low dose CT images with different phantoms

The quantitative results of the performance evaluation of three different methods for denoising different phantom chest LDCT images are shown in Table 2. We conduct a preliminary analysis of the results of the three different methods in the table.

(1) Analysis of Denoising Results by the UDiff Model.

The denoising results of the UDiff model show that the PSNR values of 8 phantoms fluctuate greatly, while the SSIM values are relatively stable, indicating that the method performs well in maintaining image structure. From a data perspective, the indicators of Phantom 1, Phantom 6, and Phantom 8 are relatively high, indicating that their denoising performance is good. However, the indicators of Phantom 2, Phantom 3, and Phantom 5 are relatively low, indicating that their performance in noise removal and structural preservation is average. Among them, Phantom 1 has the highest PSNR value (29.4709) and good denoising effect; The SSIM value of Phantom 2 (0.8580) is stable, indicating good similarity in brightness, contrast, and structure. However, the RMSE value is relatively high (0.0868), indicating that the denoising effect is not ideal. The SSIM value of Phantom 7 (0.8631) is optimal, but the PSNR value (27.4323) is not optimal.

(2) Analysis of Denoising Results by the UAdiff Model.

From the denoising results of the UAdiff model, there is a certain fluctuation in its overall performance among the 8 models. In Phantom 2, Phantom 7, and Phantom 8, the SSIM values were increased to 0.8720, 0.8722, and 0.8542, respectively, demonstrating good structural retention ability. However, the generalization of the model is insufficient. The PSNR and SSIM of Phantom 3 significantly decreased (19.7348 and 0.6868), while the RMSE increased to 0.1012, showing significant oversmoothing and loss of details. The three indicators of Phantom 5 are also poor, further indicating that this method has limited performance in handling complex textures. In contrast, the indicators of Phantom 1, Phantom 4, and Phantom 6 are basically the same as those of Method 1. The extreme fluctuations of Phantom 3 and Phantom 5 reveal the insufficient robustness of this method, indicating that its denoising model or prior assumptions have not fully covered multiple image features.

**Table 2.** Results of three denoising methods

		PSNR	SSIM	RMSE
UDiff model	Phantom 1	29.4709	0.8623	0.0328
	Phantom 2	21.7473	0.8580	0.0868
	Phantom 3	22.7212	0.8461	0.0625
	Phantom 4	27.4334	0.8492	0.0427
	Phantom 5	23.2288	0.7962	0.0681
	Phantom 6	28.8470	0.8272	0.0345
	Phantom 7	27.4323	0.8631	0.0419
	Phantom 8	29.3183	0.8494	0.0337
UAdiff model	Phantom 1	29.5040	0.8579	0.0327
	Phantom 2	21.9627	0.8720	0.0846
	Phantom 3	19.7348	0.6868	0.1012
	Phantom 4	27.1627	0.8413	0.0450
	Phantom 5	21.8145	0.6992	0.0797
	Phantom 6	28.8914	0.8257	0.0343
	Phantom 7	27.4413	0.8722	0.0419
	Phantom 8	29.1469	0.8542	0.0344
DUPAnet model	Phantom 1	30.3087	0.8734	0.0296
	Phantom 2	27.4702	0.8867	0.0438
	Phantom 3	22.7528	0.7341	0.0611
	Phantom 4	27.8633	0.8519	0.0399
	Phantom 5	24.2458	0.7551	0.0607
	Phantom 6	29.0352	0.8299	0.0337
	Phantom 7	27.4542	0.8688	0.0419
	Phantom 8	30.3702	0.8691	0.0297

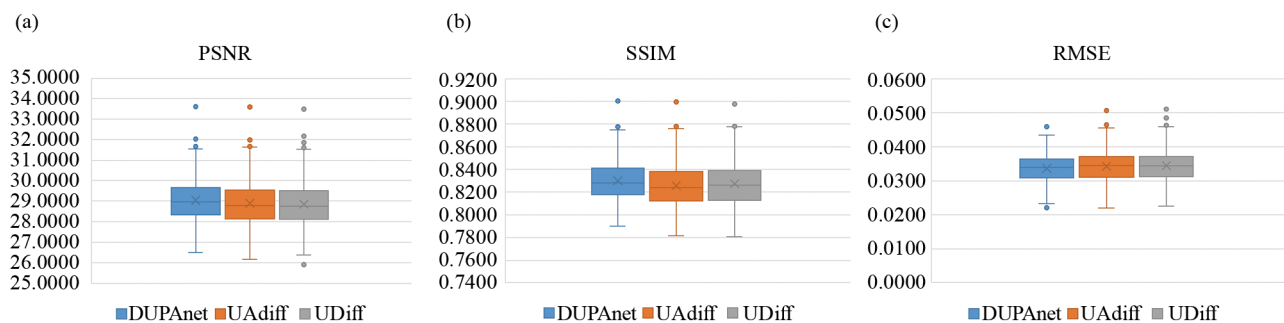
### (3) Analysis of Denoising Results by the DUPAnet Model.

The denoising results of the DUPAnet model show that the overall performance of this method is better than the first two methods on 8 phantoms. The PSNR values of all phantoms have been improved, indicating that they have stronger noise suppression ability and signal-to-noise ratio improvement effect, especially on Phantoms 2, 3, and 5, which perform poorly in the first two methods, with Phantom 2 showing the most significant improvement. The PSNR value of Phantom 8 reaches 30.3702, demonstrating outstanding performance. In terms of SSIM, the overall improvement is not significant, but it exhibits stronger robustness and can better maintain image structural information. At the same time, the RMSE values of all phantoms have decreased, indicating that this method performs better in pixel level denoising

accuracy and global error control, and the improvement is particularly evident in Phantoms 2 and 3. The DUPAnet model improves PSNR and RMSE metrics, significantly reduces pixel errors, and exhibits smaller fluctuation ranges between different phantoms, indicating its stronger adaptability to diverse image features. The stable performance of SSIM further validates its advantage in maintaining structural information, and the reduction of performance fluctuations highlights the robustness of this method.

#### (4) Analysis of the performance distribution across all models.

In order to compare and analyze the performance differences of DUPAnet, UAdiff, and UDiff methods on SSIM, PSNR, and RMSE indicators more clearly, we selected the results of Phantom 8 to draw box-and-whisker plot of the performance distribution of each indicator for the three methods, as shown in Figure 9. The results showed that the SSIM of the three methods were concentrated in the range of 0.80-0.85, with the median of DUPAnet and UDiff slightly higher than UAdiff, indicating their advantages in structural preservation. The overall distribution of PSNR is concentrated between 28-30, and the overall level of DUPAnet is relatively higher, indicating that it performs better in terms of signal-to-noise ratio. The distribution of RMSE is concentrated in the range of 0.03-0.04, and the median of UDiff and DUPAnet is lower than that of UAdiff, indicating that their reconstruction errors are smaller. Overall, DUPAnet performs the best in SSIM, PSNR, and RMSE, with high overall reconstruction quality. The UDiff results are relatively stable, while the performance of UAdiff is slightly inferior.

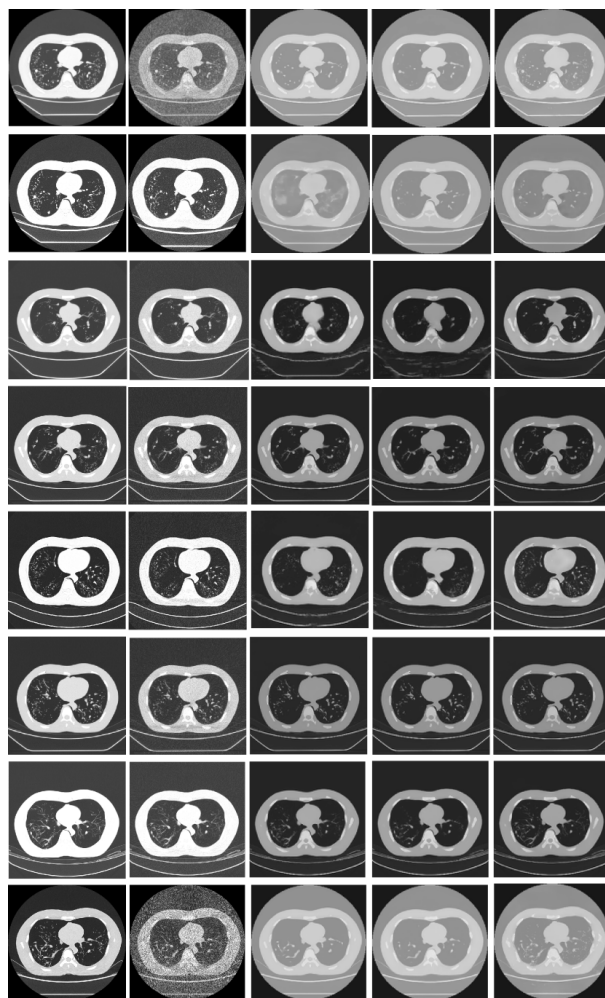


**Figure 9.** Box-and-whisker plot of performance distribution of indicators for three models using Phantom 8 as an example. (a) PSNR for the three models of Phantom 8, (b) SSIM for the three models of Phantom 8, (c) RMSE for the three models of Phantom 8

### 3.1.2 Image denoising analysis of chest low dose CT images with different phantoms

The results of denoising LDCT images of 8 phantoms using three methods are shown in Figure 10. It can be seen that the denoising results of the UDiff model denoising method are mostly acceptable. The results of Phantom 1 and Phantom 8 show good denoising effects, but there are many missing details. For Phantom 4, Phantom 5, and Phantom 6, the denoising results are good. The image reveals that some details are preserved during the denoising process. However, for Phantom 2, the denoising results are unsatisfactory, with details becoming excessively blurred after denoising. For Phantom 3, although the details are restored, the display is not clear and there is also blurring. The overall denoising results of the UAdiff model are generally good, with good results for Phantoms 4, 6, and 7. Nevertheless, for other phantoms, more details are lost, indicating that incorporating attention mechanisms can improve feature focus and boost noise removal ability. However, the early feature extraction lacks guidance from noise features and may not be able to directly distinguish between noise and detail features, resulting in insufficient performance in detail recovery. The overall denoising results of the DUPAnet model are the best among the three methods, as evidenced by the comparative analysis of various denoising techniques. While the details restored for Phantom 2 and Phantom 3 are not complete enough, they avoid blurring. Compared to the UAdiff model, the DUPAnet model restores more details while denoising other phantoms, and the overall denoising results are visually improved.





**Figure 10.** Eight types of phantom denoising results (from left to right are normal dose CT, LDCT, UDiff model denoising results, UAdiff model denoising results, DUPAnet model denoising results)

### 3.1.3 Analysis of denoising results of chest low dose CT images with different scan layer thicknesses

In order to control variables for better comparison of different scan intervals, we observed the data scan parameters of different phantoms and only used other sequences from Phantom 4 that were not used for training and validation for this comparative experiment. Table 3 presents a comparative analysis of the efficacy of the three different denoising methods for chest LDCT images at different scan slice intervals. From Table 3, it can be seen that the PSNR results of the UDiff model and UAdiff model with a slice spacing of 1.0 mm are not as good as those with a slice spacing of 5.0 mm, and the SSIM results are relatively poor. However, the DUPAnet model shows good results in handling various indicators of the two slice spacings, which indicates that the DUPAnet model can avoid the difference between slice spacings and maintain more stable denoising performance. Even though the PSNR results for a layer spacing of 1.0 mm are better, the SSIM results of the layer spacing of 5.0 mm are more favorable, indicating that the sequence images with a 1.0 mm spacing have more details in their images, indicating that the DUPAnet model has good denoising ability, but there is still a certain lack of ability to restore details. However, the overall denoising performance is better than the three methods. The best among them.



**Table 3.** Results of three methods for LDCT images with different scan slice intervals

Denoising method	Slice spacing (mm)	PSNR	SSIM	RMSE
UDiff model	$1.0 \times 1.0$	24.4295	0.7439	0.0598
	$5.0 \times 5.0$	27.4310	0.8492	0.0428
UAdiff model	$1.0 \times 1.0$	24.5159	0.7498	0.0592
	$5.0 \times 5.0$	27.2932	0.8425	0.0446
DUPAnet model	$1.0 \times 1.0$	28.2375	0.8096	0.0379
	$5.0 \times 5.0$	27.8672	0.8519	0.0398

### 3.1.4 Analysis of denoising results of chest low dose CT images generated by different voltages

In order to control the variables and better compare the chest LDCT cases with different voltages separately, we only selected partial sequences from Phantom 4 and Phantom 7 that were almost identical to other variables for this comparative experiment. Table 4 shows the performance evaluation of chest LDCT image denoising tasks, generated by different voltages using the three different methods as previously mentioned. From the data in Table 4, it can be seen that the UDiff model and UAdiff model have poor results in PSNR, SSIM, and RMSE metrics regardless of voltage magnitude. A voltage change of 10 units does not seem to have much impact on denoising. The DUPAnet model performs better than the first two methods in all aspects. On the PSNR, the DUPAnet model achieved 27.3584 at 80 kV, 27.3556 at 70 kV, and 27.3617 at 60 kV, showing better results compared to the previous two methods. The PSNR values were almost consistent at different voltages, indicating that its denoising ability is very stable. The SSIM value is higher than the results of the other two methods, and it remains almost consistent at different voltages, indicating that its image detail restoration ability is also relatively stable. RMSE remained consistently at 0.0419 in the DUPAnet model, significantly lower than the other two methods, indicating that the difference between the denoised image and the normal dose image was smaller, and also demonstrating its advantage in denoising performance.

Overall, although all three methods are not affected by voltage, the overall results of the UDiff and UAdiff models are not as good as the DUPAnet model in denoising tasks, especially in terms of noise removal ability, which is significantly better than the UDiff and UAdiff models.

**Table 4.** Results of three methods for LDCT images with different voltages

Denoising method	Kv	PSNR	SSIM	RMSE
UDiff model	80	24.8033	0.7559	0.0565
	70	24.7084	0.7552	0.0574
	60	24.3438	0.7493	0.0600
UAdiff model	80	24.4294	0.7522	0.0592
	70	24.4671	0.7525	0.0590
	60	24.3275	0.7502	0.0601
DUPAnet model	80	27.3584	0.7900	0.0419
	70	27.3556	0.7901	0.0419
	60	27.3617	0.7903	0.0419

## 4. Discussion

### 4.1 Research results

This study compared the denoising effects of three denoising models, namely UDiff, UAdiff and DUPAnet, on chest LDCT images. The main results show that the DUPAnet model outperforms the other two models in terms of

metrics such as PSNR, SSIM and RMSE, significantly improving the visual quality of denoised images. The UDiff model demonstrates superior denoising performance when a notable difference between LDCT and normal dose CT, but when the difference between the two is not significant (such as Phantom 2 and 3), the denoising results tend to exhibit blurring. The UAdiff model omits the process of adding noise, which to some extent avoids the limitations of qualitative denoising, and introduces a mixed attention mechanism to maintain denoising ability. However, the results showed a strong dependence on input data and a significant loss of image detail structure. Due to the lack of guidance from noise feature information, the attention ability of attention features is weakened, which makes it difficult to remove complex noise in LDCT, especially for phantoms with non prominent textures (such as Phantoms 3 and 5), where excessive denoising occurs, affecting the denoising effect. The DUPAnet model introduces LDCT images as prior information of noise, provides noise feature guidance, and adds a mixed attention mechanism in the denoising Unet decoding part to effectively distinguish noise and detail features, enhance attention to different features, and thus improve denoising ability while preserving detail structures. This method shows better results and visual effects for denoising different phantoms.

For the difference in detail texture prominence between normal dose CT and LDCT, the phantom exhibits stable denoising ability, which to some extent improves the model's generalization ability. For LDCT images with a scan spacing of 1.0 mm, due to the presentation of more detailed structures, the denoising effects of UDiff and UAdiff models are slightly inferior compared to images with a scan spacing of 5.0 mm. However, the DUPAnet model completely overcomes the impact of the difference in scan spacing and even exhibits better denoising effects at a scan spacing of 1.0 mm. In addition, different scanning voltages have little effect on the denoising performance of the three methods, and the overall results show that the DUPAnet model outperforms the other two models in terms of denoising performance. Therefore, the DUPAnet model proposed in this article can more effectively remove noise from chest LDCT images while preserving some detailed structures, significantly improving image quality and visual effects, and effectively improving the impact of different spacing LDCT images on denoising, although there are still some limitations in denoising generalization ability on different phantoms.

## 4.2 The advantages and limits of different denoising model

Liu et al. proposed the residual diffusion model [36], which gradually refines the image through iterative denoising and denoising processes, effectively handling noise distribution. At the same time, convolution operations are good at capturing local noise patterns and are easy to expand. Wu et al. proposed a denoising diffusion probability model [37], which uses U-Net to predict noise and iteratively refines it through diffusion process control to achieve high-quality denoising. The UDiff model combines the diffusion model and the classic Unet structure for denoising chest LDCT images. In the forward process, the model selectively adds Gaussian noise and then uses the Unet structure with efficient feature expression ability for denoising. The experimental results show that the UDiff model has poor denoising performance on Phantom 2, 3, and 5, failing to clearly differentiate between normal dose and low dose. The denoised image results show blurriness and significant loss of details. In contrast, the numerical outcomes of Phantoms 1 and 8 are slightly better than Phantoms 4 and 6, however, when examining the images, Phantoms 4 and 6 have better visual effects, making their final results more aligned with practical requirements. The denoising ability of this method lacks stability, and there are certain limitations in adding specified noise in experiments, resulting in poor generalization ability to complex or unknown noise.

To overcome the limitations of qualitative noise, we initially eliminate the forward noise process utilizing the aforementioned mode. According to the research results of Tian et al. [38], embedding attention mechanisms into CNN network architecture can significantly enhance image denoising performance. The UAdiff model abandons the forward denoising step and introduces a channel space hybrid attention mechanism in the decoding stage of the Unet network during the reverse denoising process. This mechanism enhances the model's ability to distinguish between noise and detail features by dynamically adjusting feature weights and constructing long-range dependencies, strengthens the extraction of key features while effectively eliminating noise, ensuring the preservation of image details. The experimental results reveal that the UAdiff model exhibits significant fluctuations in denoising performance across different phantoms and has a strong dependence on input data features. From the results, it can be seen that the denoising effect of the model on Phantom 3 and Phantom 5 has decreased, while there is a slight improvement on Phantom 2, 6, and 7, and the overall pixel level error has been reduced. Although the UAdiff model incorporates attention mechanism in the denoising method,

it removes the forward process of the diffusion model, resulting in a lack of guidance from noise features in the overall denoising model and weakening the reinforcement ability of attention features. This fails to fully preserve details in the image denoising process and performs poorly in terms of generalization ability for denoising different models.

Given the guiding role of noise features in the denoising process and the need to avoid qualitative noise, this study draws on the PET image denoising method based on the denoising diffusion probability model proposed by Gong et al., combined with the prior information of MR images, in order to further improve the denoising effect and reduce uncertainty [31]. Based on the concept of multimodal image fusion, this study proposes a denoising method using the DUPAnet model. In the forward process of the diffusion model, LDCT images are introduced as prior information for denoising, that is, noise is added to normal dose CT images and input into the network in a dual channel form. In the encoding stage of the Unet network for reverse denoising, the mixed attention mechanism is continued to be used to denoise the chest LDCT images. The experimental results show that the DUPAnet model denoising method exhibits more stable and excellent denoising effects on the three evaluation indicators of PSNR, SSIM, and RMSE, especially in error control and denoising ability, demonstrating high stability. In experiments with different phantoms, there has been a certain degree of performance improvement, especially evident in Phantom 2. From the image results, the overall denoising visual effect is good, and the preservation of detailed structures has also been enhanced. This indicates that using LDCT images as prior information for denoising is effective, which can better handle the denoising problem of different chest LDCT images and improve the generalization ability for handling complex noise. This method has shown significant potential in the field of denoising technology.

### 4.3 Optimize of loss function

In previous studies on LDCT image denoising, the use of the MSE loss function alone often resulted in excessively smooth denoised images [39]. The three methods used in this study all employed mixed loss functions, mainly including MSE loss function,  $L_1$  loss function, and SSIM loss function. In the field of image denoising, the MSE loss function is widely used due to its mathematical differentiability and optimization stability, which effectively constrains pixel-level errors. but it can easily cause image smoothing and loss of details. The  $L_1$  loss function has stronger robustness to noise and outliers, and can better preserve edge details, but it is slightly inferior to MSE in terms of convergence speed and stability [32]. The SSIM loss function evaluates image quality from multiple dimensions such as structure, brightness, and contrast, which is more in line with human perception and can effectively improve visual quality. However, optimization is difficult and computationally complex [40]. By mixing these three loss functions, it is possible to comprehensively consider pixel accuracy, edge preservation, and structure perception, thereby improving the stability and perceptual quality of denoising effects and compensating for the shortcomings of a single loss function. This study also adopted a hybrid attention mechanism, which replaces the serial independent structure of the basic Convolutional Block Attention Module (CBAM) hybrid attention mechanism through bidirectional information exchange of channel spatial attention [41]. Introducing spatial dimension dynamic weight guidance in channel attention, while integrating channel importance information in spatial attention to form bidirectional feedback and enhance the synergy of attention modules. This hybrid attention mechanism effectively integrates spatial and channel information, enhances feature expression ability, accurately locates noisy areas, and significantly improves image denoising and detail restoration capabilities.

### 4.4 Impacts of low dose CT scanning parameters

The imaging quality of LDCT scanning is affected by factors such as exposure parameters, tube current, tube voltage, radiation dose, slice thickness, and reconstruction algorithm [42]. For example, low tube current can increase noise and reduce imaging quality. If the tube voltage is too high, it will reduce the image contrast. Therefore, it is necessary to adjust the tube current and tube voltage according to the actual situation to achieve the best imaging effect [43]. Røhme et al. [44] evaluated the image quality and radiation dose differences of acute abdominal CT protocols using 40 different CT scanners from 33 radiology departments in Norwegian hospitals and research institutes, covering scanner models from four different suppliers. The study found significant differences in radiation dose and image quality among different CT scanners. Due to differences in hardware and software solutions between CT manufacturers and scanners, achieving unified standards

for image quality and radiation dose across scanners and suppliers has become challenging [45]. This study compares and analyzes different parameters that affect LDCT images, mainly focusing on different instruments, scanning voltages, and slice spacing. The experimental results showed that the UDiff model and UAdiff model performed relatively poorly on Phantom 2, 3 and 5, while the DUPAnet model showed some improvement overall, especially outstanding performance on Phantom 2. However, there are still shortcomings in Phantom 3 and Phantom 5. There are certain differences in the LDCT images generated by different instruments, and the generalization ability between instruments needs to be improved. For LDCT images generated at different voltages, the results showed that the differences in Phantom 4 under different voltages did not significantly affect the denoising ability of the three methods mentioned in the article, with the DUPAnet model performing the best. The number of CT images generated by scanning with different slice spacing varies, and the specific differences are reflected in the image details. The experimental results show that the UDiff model and UAdiff model have different results at different distances. The result with a distance of 5.0 mm is better, while the DUPAnet model has similar results at the two distances. On the contrary, the result with a distance of 1.0 mm is higher. This indicates that the method can still maintain good denoising ability in the case of increased image details, further verifying the effectiveness of the method. In clinical applications, by combining denoising models and optimizing scanning parameters, such as selecting appropriate voltage and slice spacing settings, optimizing denoising algorithms to achieve optimal image quality, it is possible to reduce image noise while maintaining diagnostic quality of the image, thereby providing more accurate imaging data for clinical use.

#### 4.5 Limits and future prospects

This study experimentally validated three different denoising methods using a large amount of diverse phantom data, and the results showed that all three methods have certain denoising capabilities. Among them, our proposed variant diffusion model, incorporating LDCT prior information and mixed attention mechanism, exhibits better denoising effect under multiple body model data and different scanning voltages and spacing conditions, confirming its ability to effectively remove noise from chest LDCT images while preserving detailed structures and significantly improving image quality.

However, the model still struggles with generalizing to different phantoms, and its performance may degrade in clinical settings with complex, real-world noise patterns. While phantom data was used extensively in this study, real human LDCT data has not been incorporated, which limits the model's practical applicability. Collecting real human data is challenging due to variability in scan conditions, patient anatomy, and noise. This poses a significant obstacle to evaluating the model's true clinical effectiveness. Moreover, the model's ability to generalize to other anatomical regions beyond chest LDCT images remains uncertain, as denoising methods are typically region-specific.

Future research should focus on improving the model's generalization by including real human data, exploring multimodal denoising approaches with other imaging techniques like Magnetic Resonance Imaging (MRI) and PET, and addressing its applicability across different anatomical regions. Additionally, integrating real-time denoising algorithms into CT scanning systems could significantly enhance diagnostic workflows, enabling seamless image processing and improving clinical efficiency as medical imaging technologies evolve.

## 5. Conclusions

Computed tomography technology plays an important role in the diagnosis of diseases such as cancer, but its radiation can cause harm to the human body and even further catalyze the impact of diseases on the human body. The images of low radiation LDCT technology have obvious noise and artifacts, which affect clinical diagnosis. The research on denoising technology is very important to solve the quality problem of LDCT images. Research experiments have shown that the DUPAnet denoising method proposed in this paper performs the best in terms of PSNR, SSIM, and RMSE when processing LDCT images. It can effectively remove noise and preserve details, and performs the most stably under different scan spacing and voltage conditions, thereby improving the quality of chest LDCT images. According to the discussion, there is still a lot of room for exploration in this research question in the future. In addition to further optimizing denoising algorithms and improving their generalization ability to different LDCT conditions and CT scans of different parts, it is

also necessary to enhance real-time processing capabilities and adaptability to meet the demand for fast, efficient, and accurate medical images in clinical applications.

## Acknowledgement

Authors are grateful to the reviewers and editors for their suggestions and comments to improve the manuscript. Authors also thank Yunnan Fundamental Research Projects (No. 202301AU070184), Chongqing Science and Health Joint Medical Research key Project (No. 2024ZDXM001), and Chongqing Natural Science Foundation project (No. cstc2021jcyj-msxm0727, CSTB2023NSCQ-MSX0295), and Xinglin Scholar Research Promotion Project of Chengdu University of Traditional Chinese Medicine (No. YYZX2022136) and Natural Science Research Foundation of Kunming University of Science and Technology (No. KKZ3202307033).

## Conflict of interest

The authors declare no competing financial interest.

## References

- [1] Goldman LW. Principles of CT and CT technology. *Journal of Nuclear Medicine Technology*. 2007; 35(3): 115-128. Available from: <https://doi.org/10.2967/jnmt.107.042978>.
- [2] Brenner DJ, Hall EJ. Computed tomography-An increasing source of radiation exposure. *New England Journal of Medicine*. 2007; 357(22): 2277-2284. Available from: <https://doi.org/10.1056/NEJMra072149>.
- [3] Demaria S, Golden EB, Formenti SC. Role of local radiation therapy in cancer immunotherapy. *JAMA Oncology*. 2015; 1(9): 1325-1332. Available from: <https://doi.org/10.1001/jamaoncol.2015.2756>.
- [4] Hauptmann M, Byrnes G, Cardis E, Bernier MO, Blettner M, Dabin J, et al. Brain cancer after radiation exposure from CT examinations of children and young adults: results from the EPI-CT cohort study. *The Lancet Oncology*. 2023; 24(1): 45-53. Available from: [https://doi.org/10.1016/s1470-2045\(22\)00655-6](https://doi.org/10.1016/s1470-2045(22)00655-6).
- [5] Bosch de Basea Gomez M, Thierry-Chef I, Harbron R, Hauptmann M, Byrnes G, Bernier MO, et al. Risk of hematological malignancies from CT radiation exposure in children, adolescents and young adults. *Nature Medicine*. 2023; 29(12): 3111-3119. Available from: <https://doi.org/10.1038/s41591-023-02620-0>.
- [6] Smith-Bindman R, Chu PW, Firdaus HA, Stewar C, Malekheadayat M, Alber S, et al. Projected lifetime cancer risks from current computed tomography imaging. *JAMA Internal Medicine*. 2025; 185(6): 710-719. Available from: <https://doi.org/10.1001/jamainternmed.2025.0505>.
- [7] Moen TR, Chen B, Holmes III DR, Duan X, Yu Z, Yu L, et al. Low-dose CT image and projection dataset. *Medical Physics*. 2021; 48(2): 902-911. Available from: <https://doi.org/10.1002/mp.14594>.
- [8] Shin YJ, Chang W, Ye JC, Kang E, Oh DY, Lee YJ, et al. Low-dose abdominal CT using a deep learning-based denoising algorithm: A comparison with CT reconstructed with filtered back projection or iterative reconstruction algorithm. *Korean Journal of Radiology*. 2020; 21(3): 356-364. Available from: <https://doi.org/10.3348/kjr.2019.0413>.
- [9] Yang C, Wang W, Cui D, Zhang J, Liu L, Wang Y, et al. Deep learning image reconstruction algorithms in low-dose radiation abdominal computed tomography: Assessment of image quality and lesion diagnostic confidence. *Quantitative Imaging in Medicine and Surgery*. 2023; 13(5): 3161. Available from: <https://doi.org/10.21037/qims-22-1227>.
- [10] Singh S, Kalra MK, Hsieh J, Licato PE, Do S, Pien HH, et al. Abdominal CT: Comparison of adaptive statistical iterative and filtered back projection reconstruction techniques. *Radiology*. 2010; 257(2): 373-383. Available from: <https://doi.org/10.1148/radiol.10092212>.
- [11] Beister M, Kolditz D, Kalender WA. Iterative reconstruction methods in X-ray CT. *Physica Medica*. 2012; 28(2): 94-108. Available from: <https://doi.org/10.1016/j.ejmp.2012.01.003>.

- [12] Deák Z, Grimm JM, Treitl M, Geyer LL, Linsenmaier U, Körner M, et al. Filtered back projection, adaptive statistical iterative reconstruction, and a model-based iterative reconstruction in abdominal CT: An experimental clinical study. *Radiology*. 2013; 266(1): 197-206. Available from: <https://doi.org/10.1148/radiol.12112707>.
- [13] Shen D, Wu G, Suk HI. Deep learning in medical image analysis. *Annual Review of Biomedical Engineering*. 2017; 19(1): 221-248. Available from: <https://doi.org/10.1146/annurev-bioeng-071516-044442>.
- [14] Immonen E, Wong J, Nieminen M, Lanca L, Guan F, Metsälä E. The use of deep learning towards dose optimization in low-dose computed tomography: A scoping review. *Radiography*. 2022; 28(1): 208-214. Available from: <https://doi.org/10.1016/j.radi.2021.07.010>.
- [15] Zhang J, Shangguan Z, Gong W, Cheng Y. A novel denoising method for low-dose CT images based on transformer and CNN. *Computers in Biology and Medicine*. 2023; 163: 107162. Available from: <https://doi.org/10.1016/j.compbimed.2023.107162>.
- [16] Chen H, Zhang Y, Kalra MK, Lin F, Chen Y, Liao P, et al. Low-dose CT with a residual encoder-decoder convolutional neural network. *IEEE Transactions on Medical Imaging*. 2017; 36(12): 2524-2535. Available from: <https://doi.org/10.1109/TMI.2017.2715284>.
- [17] Gunawan R, Tran Y, Zheng J, Nguyen H, Chai R. Image recovery from synthetic noise artifacts in CT scans using modified U-Net. *Sensors*. 2022; 22(18): 7031. Available from: <https://doi.org/10.3390/s22187031>.
- [18] 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 Transactions on Instrumentation and Measurement*. 2021; 71: 1-12. Available from: <https://doi.org/10.1109/TIM.2021.3128703>.
- [19] Heinrich MP, Stille M, Buzug TM. Residual U-net convolutional neural network architecture for low-dose CT denoising. *Current Directions in Biomedical Engineering*. 2018; 4(1): 297-300. Available from: <https://doi.org/10.1515/cdbme-2018-0072>.
- [20] Liu X, Xie Y, Liu C, Diao S, Tan S, Liang X. Diffusion probabilistic priors for zero-shot low-dose CT image denoising. *Medical Physics*. 2023; 52(1): 329-345. Available from: <https://doi.org/10.1002/mp.17431>.
- [21] Gao Q, Li Z, Zhang J, Shan H. CoreDiff: Contextual error-modulated generalized diffusion model for low-dose CT denoising and generalization. *IEEE Transactions on Medical Imaging*. 2023; 43(2): 745-759. Available from: <https://doi.org/10.1109/TMI.2023.3320812>.
- [22] Gupta B, Lamba SS. An efficient anisotropic diffusion model for image denoising with edge preservation. *Computers and Mathematics with Applications*. 2021; 93: 106-119. Available from: <https://doi.org/10.1016/j.camwa.2021.03.029>.
- [23] 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 2020 15th IEEE International Conference on Signal Processing (ICSP)*. Beijing, China: IEEE; 2020. p.193-198.
- [24] Jeon SY, Kim W, Choi JH. MM-Net: Multiframe and multimask-based unsupervised deep denoising for low-dose computed tomography. *IEEE Transactions on Radiation and Plasma Medical Sciences*. 2022; 7(3): 296-306. Available from: <https://doi.org/10.1109/TRPMS.2022.3224553>.
- [25] Kazerouni A, Aghdam EK, Heidari M, Azad R, Fayyaz M, Merhof IHD. Diffusion models in medical imaging: A comprehensive survey. *Medical Image Analysis*. 2023; 88: 102846. Available from: <https://doi.org/10.1016/j.media.2023.102846>.
- [26] Huang B, Lu S, Zhang La. One-sample diffusion modeling in projection domain for low-dose CT imaging. *IEEE Transactions on Radiation and Plasma Medical Sciences*. 2024; 8(8): 902-915.
- [27] Ronneberger O, Fischer P, Brox T. U-net: Convolutional networks for biomedical image segmentation. In: *Medical Image Computing and Computer-Assisted Intervention-MICCAI 2015*. Cham; 2015. p.234-241.
- [28] Li M, Hsu W, Xie X, Cong J, Gao W. SACNN: Self-attention convolutional neural network for low-dose CT denoising with self-supervised perceptual loss network. *IEEE Transactions on Medical Imaging*. 2020; 39(7): 2289-2301. Available from: <https://doi.org/10.1109/tmi.2020.2968472>.
- [29] Zhang X, Han Z, Shangguan H, Han X, Cui X, Wang A. Artifact and detail attention generative adversarial networks for low-dose CT denoising. *IEEE Transactions on Medical Imaging*. 2021; 40(12): 3901-3918. Available from: <https://doi.org/10.1109/tmi.2021.3101616>.
- [30] Wang Y, Song X, Chen K. Channel and space attention neural network for image denoising. *IEEE Signal Processing Letters*. 2021; 28: 424-428. Available from: <https://doi.org/10.1109/LSP.2021.3057544>.



- [31] Gong K, Johnson K, El Fakhri G, Li Q, Pan T. PET image denoising based on denoising diffusion probabilistic model. *European Journal of Nuclear Medicine and Molecular Imaging*. 2024; 51(2): 358-368. Available from: <https://doi.org/10.1007/s00259-023-06417-8>.
- [32] Ma Y, Wei B, Feng P, He P, Guo X, Wang G. Low-dose CT image denoising using a generative adversarial network with a hybrid loss function for noise learning. *IEEE Access*. 2020; 8: 67519-67529.
- [33] Kaur R, Juneja M, Mandal AK. A comprehensive review of denoising techniques for abdominal CT images. *Multimedia Tools and Applications*. 2018; 77: 22735-22770. Available from: <https://doi.org/10.1007/s11042-017-5500-5>.
- [34] Zhang K, Zuo W, Gu S, Zhang L. Learning deep CNN denoiser prior for image restoration *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*. Honolulu, HI, USA: IEEE; 2017. p.3929-3938.
- [35] Zubair M, Md Rais HB, Ullah F, Al-Tashi Q, Faheem M, Ahmad Khan A. Enabling predication of the deep learning algorithms for low-dose CT scan image denoising models: A systematic literature review. *IEEE Access*. 2024; 12: 79025-79050. Available from: <https://doi.org/10.1109/access.2024.3407774>.
- [36] Liu J, Wang Q, Fan H, Wang Y, Tang Y, Qu L. Residual denoising diffusion models. In: *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*. Seattle, WA, USA: IEEE; 2024. p.2773-2783.
- [37] Wu Z, Chen X, Xie S, Shen J, Zeng Y. Super-resolution of brain MRI images based on denoising diffusion probabilistic model. *Biomedical Signal Processing and Control*. 2023; 85: 104901. Available from: <https://doi.org/10.1016/j.bspc.2023.104901>.
- [38] Tian C, Xu Y, Li Z, Zuo W, Fei L, Liu H. Attention-guided CNN for image denoising. *Neural Networks*. 2020; 124: 117-129. Available from: <https://doi.org/10.1016/j.neunet.2019.12.024>.
- [39] Zhang J, Niu Y, Shangguan Z, Gong W, Cheng Y. A novel denoising method for CT images based on U-net and multi-attention. *Computers in Biology and Medicine*. 2023; 152: 106387. Available from: <https://doi.org/10.1016/j.combiomed.2022.106387>.
- [40] Wang XW, Chen CY. Real-world image denoising based on U-shaped multi-scale attention method. *Computer Technology and Development*. 2024; 34(4): 48-54.
- [41] Cai G. CBAM-DnCNN: An improved method for image denoising. In: *2023 3rd International Conference on Electronic Information Engineering and Computer Science (EIECS)*. Changchun, China: IEEE; 2023. p.580-584.
- [42] Varghese BA, Cen SY, Jensen K, Levy J, Andersen HK, Schulz A, et al. Investigating the role of imaging factors in the variability of CT-based texture analysis metrics. *Journal of Applied Clinical Medical Physics*. 2024; 25(4): e14192. Available from: <https://doi.org/10.1002/acm2.14192>.
- [43] Tao S, Sheedy E, Bruesewitz M, Weber N, Williams K, Halaweish A, et al. Technical note: kV-independent coronary calcium scoring: A phantom evaluation of score accuracy and potential radiation dose reduction. *Medical Physics*. 2021; 48(3): 1307-1314. Available from: <https://doi.org/10.1002/mp.14663>.
- [44] Røhme LAG, Homme THF, Johansen ECK, Schulz A, Aaløkken TM, Johansson E, et al. Image quality and radiation doses in abdominal CT: A multicenter study. *European Journal of Radiology*. 2024; 178: 111642. Available from: <https://doi.org/10.1016/j.ejrad.2024.111642>.
- [45] Kahraman G, Haberal KM, Ağildere AM. Establishment of local diagnostic reference levels for computed tomography with cloud-based automated dose-tracking software in Türkiye. *Diagnostic and Interventional Radiology*. 2024; 30(3): 205-211. Available from: <https://doi.org/10.4274/dir.2023.232265>.