Computers in Biology and Medicine 152 (2023) 106387

Contents lists available at ScienceDirect

Computers in Biology and Medicine

journal homepage: www.elsevier.com/locate/compbiomed

A novel denoising method for CT images based on U-net and multi-attention

Ju Zhang [a], Yan Niu [b], Zhibo Shangguan [c], Weiwei Gong [c], Yun Cheng [d,*]

[a] College of Information Science and Technology, Hangzhou Normal University, Hangzhou, 311121, China
[b] College of Information Engineering, Zhejiang University of Technology, Hangzhou, 310023, China
[c] College of Computer Science and Technology, Zhejiang University of Technology, Hangzhou, 310023, China
[d] Department of Ultrasound, Zhejiang Hospital, Hangzhou, 310013, China

ARTICLE INFO

Keywords:
Medical image denoising
Attention mechanism
Convolutional neural network
U-net
Residual structure

ABSTRACT

Reducing the radiation dose may lead to increased noise in medical computed tomography (CT), which can adversely affect the radiologists' judgment. Many efforts have been devoted to the denoising of low-dose CT (LDCT) images. However, it is often observed that denoised medical images usually lose some important clinical lesion edge information and may affect doctors' clinical diagnosis. For a denoising neural network, it is expected that the neural network can well retain the detailed features and make the network more anthropomorphic, and to simulate the attention mechanism of observation, being a valuable feature of the thinking process of human brain. Based on U-network (U-Net) and multi-attention mechanism, a novel denoising method for medical CT images is proposed in this study. To obtain different feature information in CT images, three attention modules are proposed in our method. The local attention module is developed to localize the surrounding information of the feature map and calculate each pixel from the context extracted from the feature map. The multi-feature channel attention module can automatically learn and extract features, suppress some invalid information and add different weights to each channel in the feature map according to different tasks. The hierarchical attention module allows the deep neural network to extract a large amount of feature information. This study also introduces an enhanced learning module to learn and retain the detail information in the image by stacking multi-layer convolution layer, batch normalization (BN) layer and activation function layer to increase the network depth. Experimental studies are conducted, and comparisons with the state-of-the-art networks are made, and the results demonstrate that the developed method can effectively remove the noise in CT images and improve the image quality in the evaluation metrics of peak signal to noise ratio (PSNR) and structural similarity (SSIM). Our method achieved 34.7329 of PSNR and 0.9293 of SSIM for $\sigma = 10$ on the QIN_LUNG_CT dataset, and achieved 28.9163 of PSNR and 0.8602 of SSIM on the Mayo Clinic LDCT Grand Challenge dataset.

1. Introduction

The extensive use of computed tomography (CT) in medical practice has raised a public concern over the associated radiation dose to the patient. Reducing the radiation dose may lead to increased noise, which can adversely affect the radiologists' judgment. Many efforts have been devoted to the denoising of low-dose CT (LDCT) images [1].

Spatial pixel image denoising algorithm, such as non-local mean (NLM) image denoising method, uses the whole image to denoising. Looking for similar areas in the image as a block of images, and averaging the areas, can remove well the noise in the image. In the transformation domain (Discrete Cosine Transform domain, Discrete Fourier Transform domain, Shearlet Transform domain) [5], the threshold denoising method is used to denoise the stacked image block, obtaining an estimate of similar blocks, and then according to the mean weight, achieving the aggregation of image denoising. For Wavelet based image denoising, the Wavelet coefficient of image signal is relatively large and the wavelet coefficient of the noise is small. By selecting a suitable threshold value and keeping those coefficients larger than the threshold, wavelet coefficients smaller than the threshold are set to 0, and the denoised image is obtained by inverse transformation. Deep learning methods provide new ideas for the denoising of medical images. For deep convolutional neural networks (CNN), the denoising process is an end-to-end method. Generally, increasing the number of layers in the

* Corresponding author.
E-mail address: 962110508@qq.com (Y. Cheng).

https://doi.org/10.1016/j.compbiomed.2022.106387
Received 29 May 2022; Received in revised form 20 November 2022; Accepted 28 November 2022
Available online 1 December 2022

network improves model performance. Compared with the traditional feature extraction algorithm, CNN can take the image as the network input data, reduce the complex preprocessing of the image and improve the effectiveness and accuracy of the process [2,3]. In traditional image processing, with the improvement of pixels, the number of parameters required increases, which has an effect on image processing speed. CNN can effectively reduce the number of parameters. In CNN, the initial image can be divided into multiple block units, and CNN uses the weight sharing method. The weight sharing method is used to select a random block after determining the weight, and the size of the corresponding weight can be applied to other image blocks.

Most existing CNN image denoising models have many layers. In this case, the model requires many parameters, and the computational cost of training is very high. Research has shown that expanding the receptive field is effective in improving the performance of CNN [4]. In general training tasks, the larger the receptive field of the convolution layer, the better the performances. In image classification, the receptive field from the input layer to the output layer is generally larger than the image. The deeper the depth of the neural network, the larger the receptive field, and the better the performance. There are many ways to expand the receptive field. One option is to simply stack more convolution layers to increase depth [4]. However, it inevitably brings more parameters and increases the computational burden. Another option is to use pooling. The main task of the pooling operation is to reduce the dimension of data to reduce the network parameters, to improve the calculation efficiency of the network and to avoid over fitting of training parameters caused by too many parameters. However, a separate pooling operation cannot be directly applied to image denoising. Using hole convolution is a common method to increase the receptive field. Compared with ordinary convolution, hole convolution increases the receptive field, but it does not increase the total parameters of the entire neural network because hole convolution operates selectively in a specific range of pixels during convolution operation. The residual network [6], can guarantee performance while increasing the number of network layers and improves overfitting problems.

The U-net network gets smaller features than the original image after the down-sampling process in the encode stage. After a decoding process, the output image can be restored to a higher quality than the input image. This idea can also be used for image denoising. In the training stage, the noisy image is put into the encode-decode as input, and the goal is to restore the original image after denoising. With the progress of down-sampling, the receptive field will gradually expand, the area that can be perceived becomes larger, and the low-frequency information of the image is perceived more. Up-sampling is a recovery process. At the same time, the network architecture carries out skip connection, which is conducive to the integration of the information of each stage of down-sampling in the up-sampling process. However, in the past experiments, it is found that this U-net network also has some shortcomings. For example, compared with many networks, U-net has too shallow layers and too few parameters, so it is easy to overfit during training. Adding attention mechanism to the original U-net structure can increase the hyperparameters in the network, improve the network performance, prevent falling into local optimum and speed up the network convergence. At the same time, the essence of attention mechanism is a set of weight coefficients learned autonomously through the network, and in the way of dynamic weighting to emphasize the region we are interested in, while suppressing the irrelevant background region, this feature plays a great role in suppressing image noise.

It is often observed that denoised medical images usually lose some important clinical lesion edge information and may affect doctors' clinical diagnosis [16,17]. For a denoising neural network, it is expected that the neural network can well retain the detailed features and make the network more anthropomorphic, and to simulate the attention mechanism of observation, being a valuable feature of the thinking process of human brain.

A novel denoising method for CT images based on U-Net and multi-attention is proposed in this study. The local attention module is developed to localize the surrounding information of the feature map. The multi-feature channel attention module can learn and extract features, suppress some invalid information and add different weights to each channel in the feature map according to different tasks. The hierarchical attention module allows the deep neural network to extract a large amount of feature information and ensure the lightweight of the network. Experimental studies and comparisons with the state-of-the-art networks show that the method developed in this study can effectively remove the noise in CT images and improve the image quality.

The following are the main contributions of this study:

1) Based on U-network (U-Net) and multi-attention mechanism, a novel denoising method for medical CT images is proposed in this study. To obtain different feature information in CT images, three attention modules are proposed in our method.
2) The local attention module is developed to localize the surrounding information of the feature map, and calculate each pixel from the context extracted from the feature map. The multi-feature channel attention module can automatically learn and extract features, and suppress some invalid information. The hierarchical attention module allows the deep neural network to extract a large amount of feature information and ensure the network's lightweight.
3) After each attention module, an enhancement learning module is introduced to merge the feature maps, to learn and retain the detail information in the image by stacking multi-layer convolution layer, BN layer, and activation function layer to increase the network depth.
4) The proposed method is verified on two datasets, i.e. the QIN_LUNG_CT dataset, and the Mayo Clinic LDCT Grand Challenge dataset. The developed denoising method is compared with the-state-of-the-art methods, demonstrating the performances and advantages of the proposed denoising method.
5) To show the role of each attention module proposed in this study, several groups of ablation experiments were conducted to observe the effect of the attention module on network performance, and the effectiveness of each attention module in the network is verified.

Our paper is organized as follows: In Section 2, we introduce backgrounds and related work. Section 3 describes U-network and multi-attention based denoising of CT Images. In Section 4, we present the experimental and comparison studies of the proposed approach. Finally, we conclude our work in Section 5. The code of this study is publicly available at https://github.com/jundao520/model/blob/main/train.py.

2. Backgrounds and related work

2.1. Application of deep neural networks in image denoising

Extensive efforts have been made to design better image reconstruction or image processing methods to reduce LDCT noise and suppress artifacts. These methods generally fall into three categories: (a) sinogram filtration before reconstruction, (b) iterative reconstruction, and (c) image post-processing after reconstruction. Image post-processing directly operates on an image. Many efforts were made in the image domain to reduce LDCT noise and suppress artifacts.

Jain et al. [7] used deep neural networks (DNNs) to significantly improve the Gaussian denoising performance. Most early depth models cannot achieve satisfactory denoising performance. Schmidt et al. [8] and Trainable nonlinear reaction diffusion (TNRD) [9] launched the optimization algorithm in the model field to learn the phased reasoning process. It is essentially a deep neural network denoising model, which can perform noise removal for medical CT images. A CNN and multi-feature extraction based denoising method for CT images is proposed to solve noise removal in medical CT images [21]. By combining residual learning [6] and batch normalization, Zhang et al. proposed a

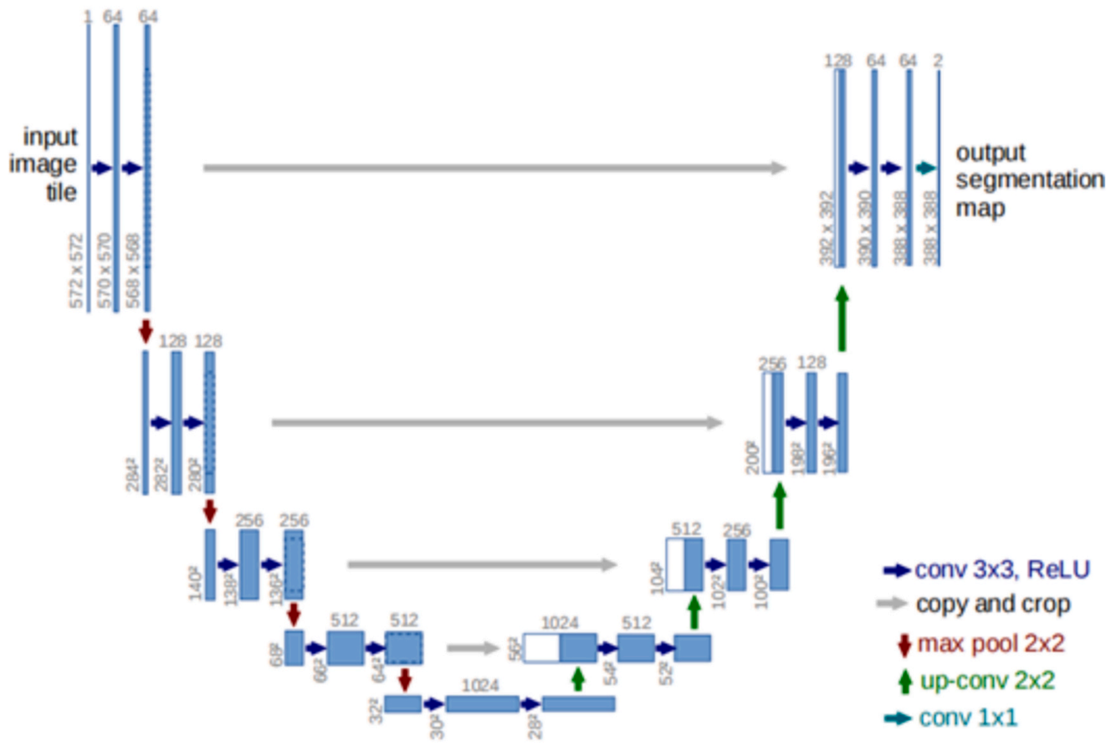

Fig. 1. U-Net structure.

denoising CNN (DnCNN) [10], which has better performance than the traditional non-CNN method. Noise2noise has also reached the advanced level without using clean data. Recently, other CNN methods such as RED30, MemNet [31], BM3D-Net [30], MWCNN [32], and FFDNet [11] have also been developed with good denoising performance. Research shows that learning a single model is feasible for Gaussian denoising. However, these blind models may overfit AWGN and cannot deal with real noise. Conversely, unblinded CNN denoisers, such as FFDNet [11], can obtain satisfactory results on most real noisy images by manually setting appropriate or relatively high noise levels.

Chen et al. [33] developed an attention augmented U-net structure. The spatial attention is developed to capture the long-range dependencies in different resolution feature maps. A self-attention mechanism is also proposed to refine the final feature maps according to the feature interaction in the neighboring positions. Wang et al. [34] proposed a normalized attention network (NAN) to learn the relationships between channels, which smooths the optimization landscape and speeds up the convergence process for training an attention model. The NAN is introduced to convolutional network denoising, in which each channel gets gain and channels can play different roles in the subsequent convolution. Dou et al. [35] adopted a 3D-based SSCA block neural network of U-Net architecture for remote sensing HSI denoising which is mainly constructed by a SSCA block. The SSCA block consists of a spatial attention (SA) block and a spectral-channel attention (SCA) block, in which the SA block extract spatial information and enhance spatial representation ability, as well as the SCA block explore the band-wise relationship within HSIs for preserving spectral information. Zhang et al. [36] proposed an improved deep convolutional U-Net framework (RatUNet) for image denoising. In order to better process the edge information of the image, RatUNet uses depth-wise and polarized self-attention mechanism to guide a CNN for image denoising. Zhu et al. [37] proposed a multi-scale residual dense attention network, which takes advantage of the context and attention information of images. Dilated convolution, dense connection and attention strategies are combined with the high-resolution network.

2.2. U-net

The U-Net structure [12] is composed of several convolution layers, which are arranged in a top-down and bottom-up manner to form a U-shaped network. Therefore, this type of convolution layer with two paths from top to bottom and then from bottom to top is called U-Net. Fig. 1 shows the U-Net structure [12]. The top-down path is called the contraction path, while the bottom-up path is called the expansion path. The contraction path is used to capture the context of the image, and the expansion path is used to effectively locate the region of interest. The shrink path consists of three blocks. Each block on the shrink path consists of two 3×3 convolutional layers, a 2×2 maximum pooling layer (max pooling layer with step size of two), and the rectified linear unit (ReLU) activation function, which is used in each convolution layer to down-sample the original picture. The extension path consists of three blocks. In the deconvolution of the extended path, there is a single 2×2 convolution layer (the activation function is also ReLU) and two 3×3 convolution layers in each step. Simultaneously, the up-sampling of each step adds the characteristic graph from the corresponding contraction path (trimmed to maintain the same shape). In the last layer of the network, there is a single 1×1 convolution layer. The feature vector of the channel can be converted into the required classification results using this operation. Finally, the entire U-Net has 23 convolution layers. The ability of U-Net to convolute images of any shape and size, particularly images of any size, is a key feature. The up-sampling operation is used to increase the image size. The extended path aids in the identification, location, and segmentation of objects of interest.

2.3. Attention mechanism

Recently, attention mechanism has been widely used in the fields of computer vision and image processing, and many practices have proved that attention mechanism is important in improving the performance of the model. In the field of computer vision or image processing, some information in the input data may play a decisive role in the correct judgment of neural network. Because only a few trainable parameters

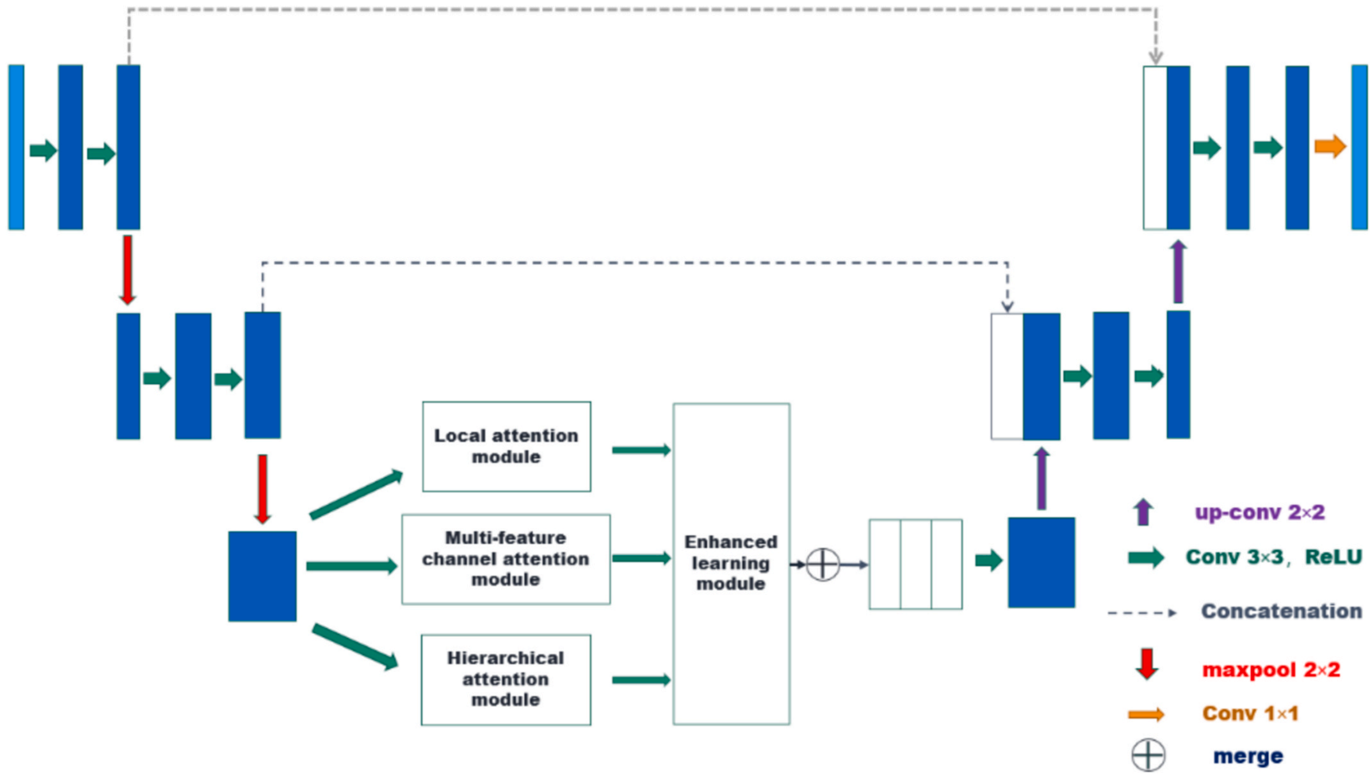

Fig. 2. The overall network framework of the proposed method.

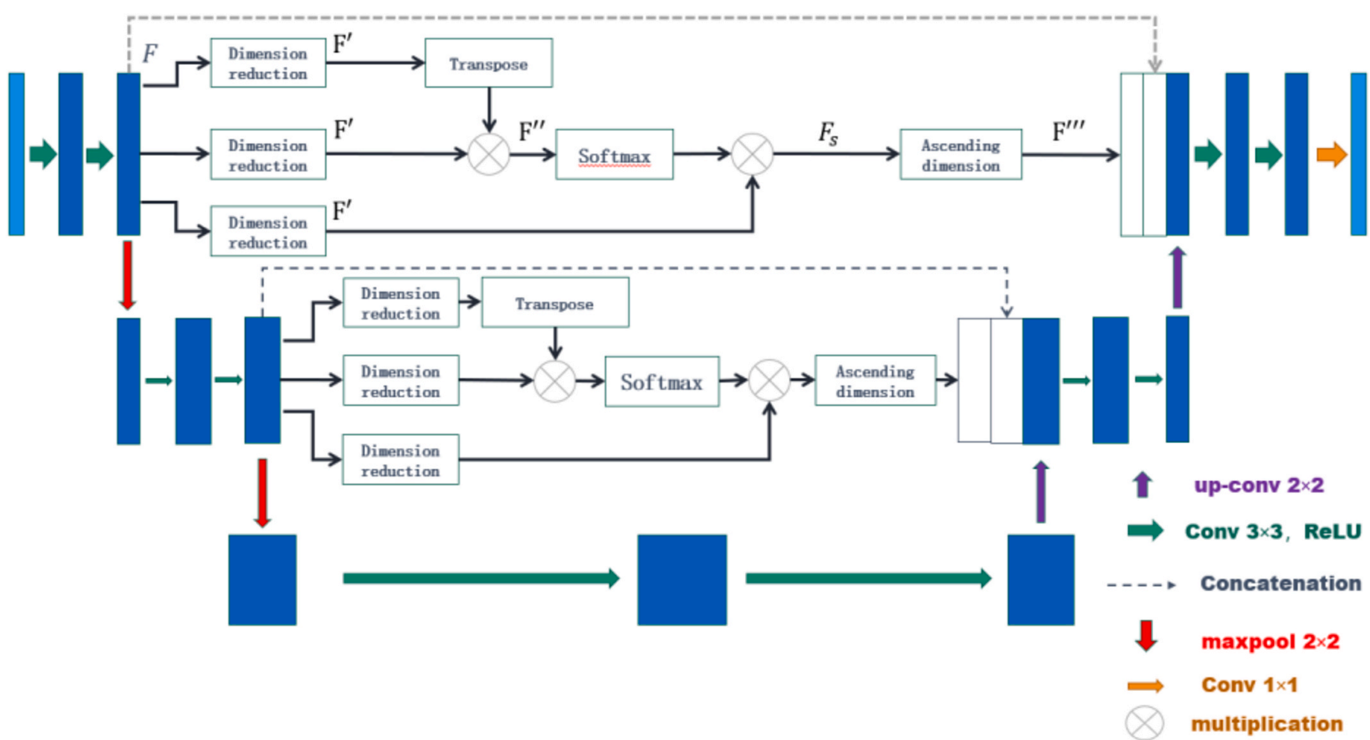

Fig. 3. Structure of the local module.

are required to learn high-performance network architecture, technology must be used to assist in focusing on the most relevant parts of the data. One way to achieve this is by fusing multimodal data features, such as channels (potential) and spatial relationships in the image, in which the information transmitted from different modes helps enhance and weaken its individual influence.

An attention function can be described as mapping a Query and a set of Key-Value pairs to an output, where the Query, Keys, Values, and output are all vectors. The output is computed as a weighted sum of the values, where the weight assigned to each value is computed by a

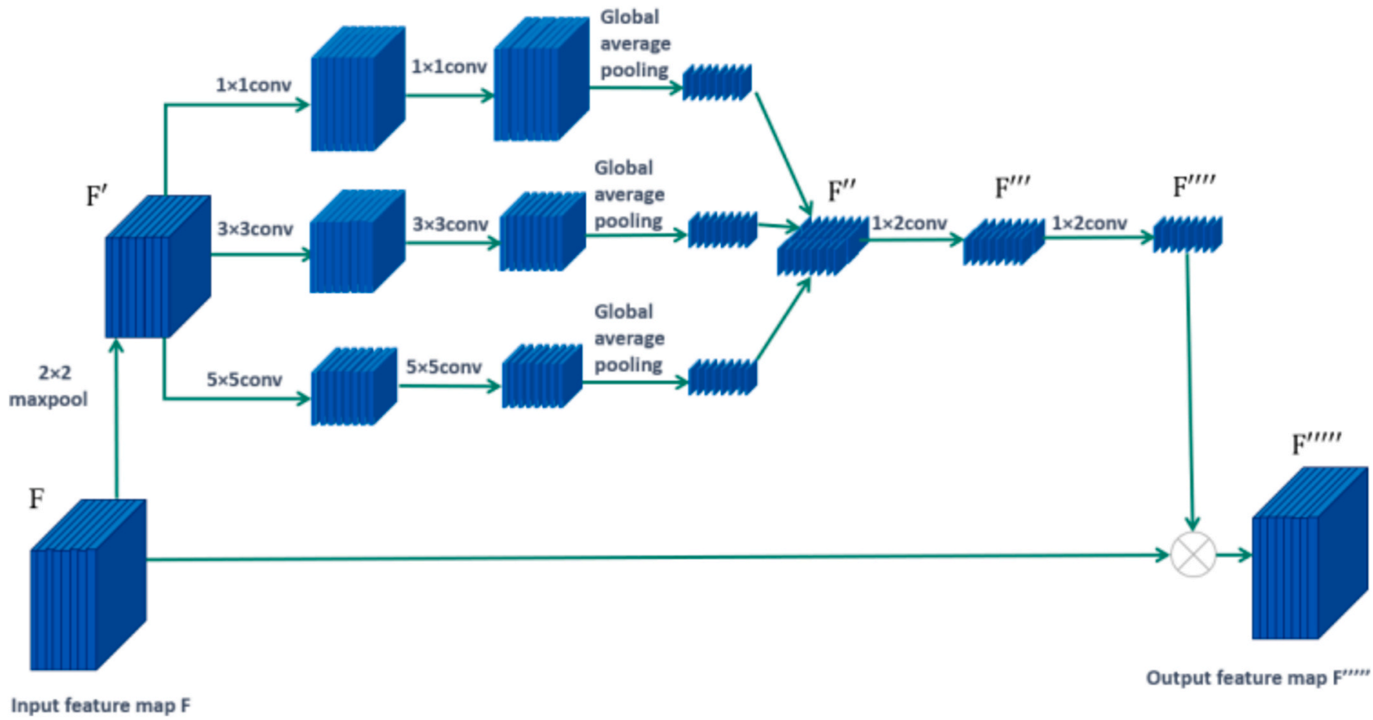

Fig. 4. Structure diagram of multi-feature channel attention module.

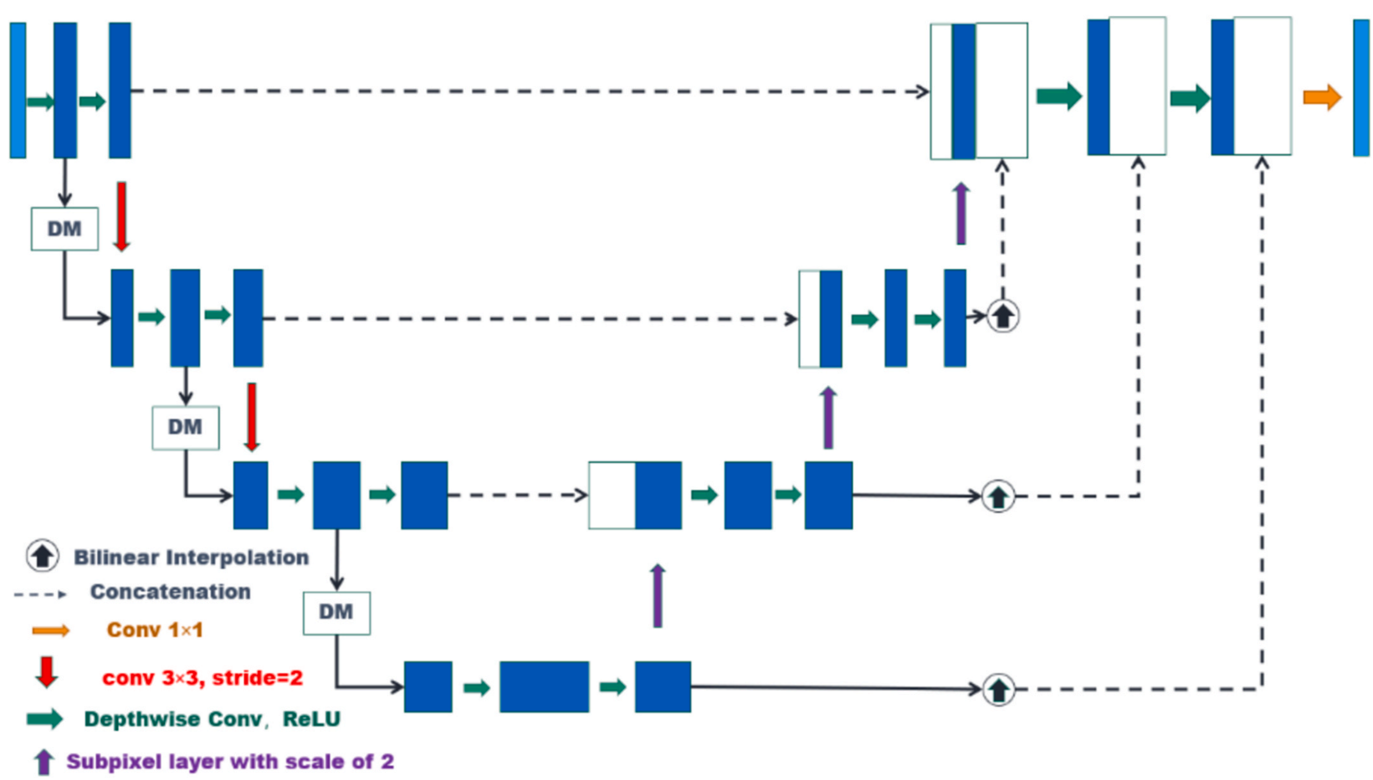

Fig. 5. Structure diagram of 4-level modules (DM:Dimension Matching).

compatibility function of the Query with the corresponding Key [22]. Specifically, we can think of a Source as a set of data pairs with Keys and Values $\{\mathrm{Key}_i, \mathrm{Value}_i \mid i = 1, 2, \ldots, m\}$. At this time, given the Query of an element in the target, the weight coefficient of the Value corresponding to each Key is obtained by calculating the similarity or correlation between the Query and each Key. Then, the weighted sum of the Values is carried out to obtain the final value of attention. Therefore, essentially, the attention mechanism is to weighted sum the Values of the elements in the Source, while Query and Key are used to calculate the weight coefficients of the corresponding Values. The calculation process can be roughly divided into three steps:

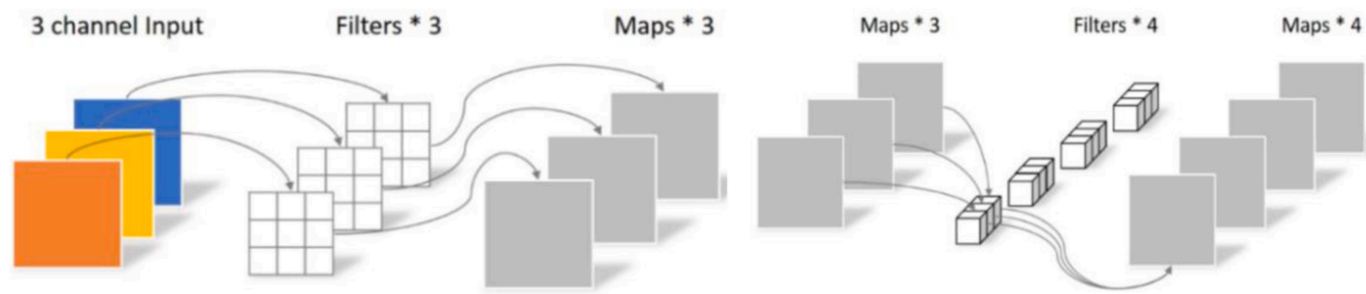

Fig. 6. Channel by channel convolution (left) and Point by point convolution(right).

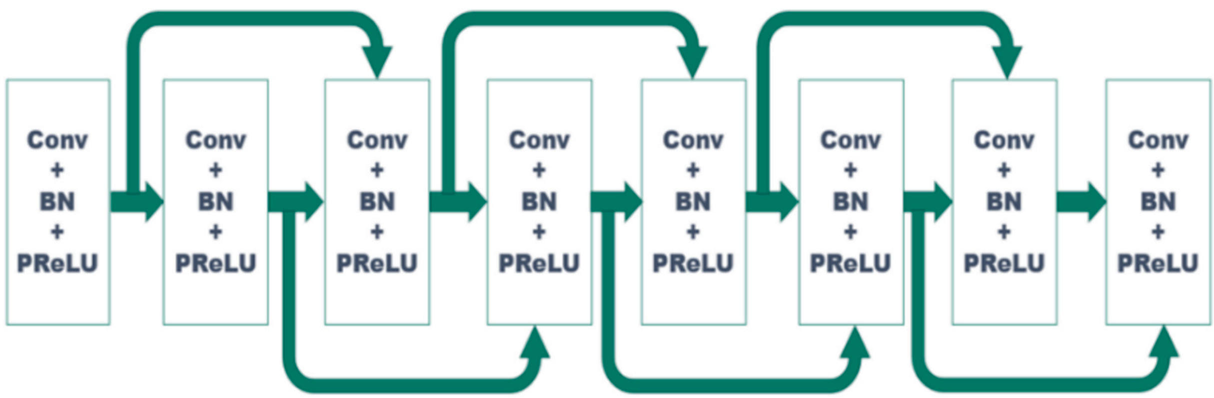

Fig. 7. Structure of enhanced learning module.

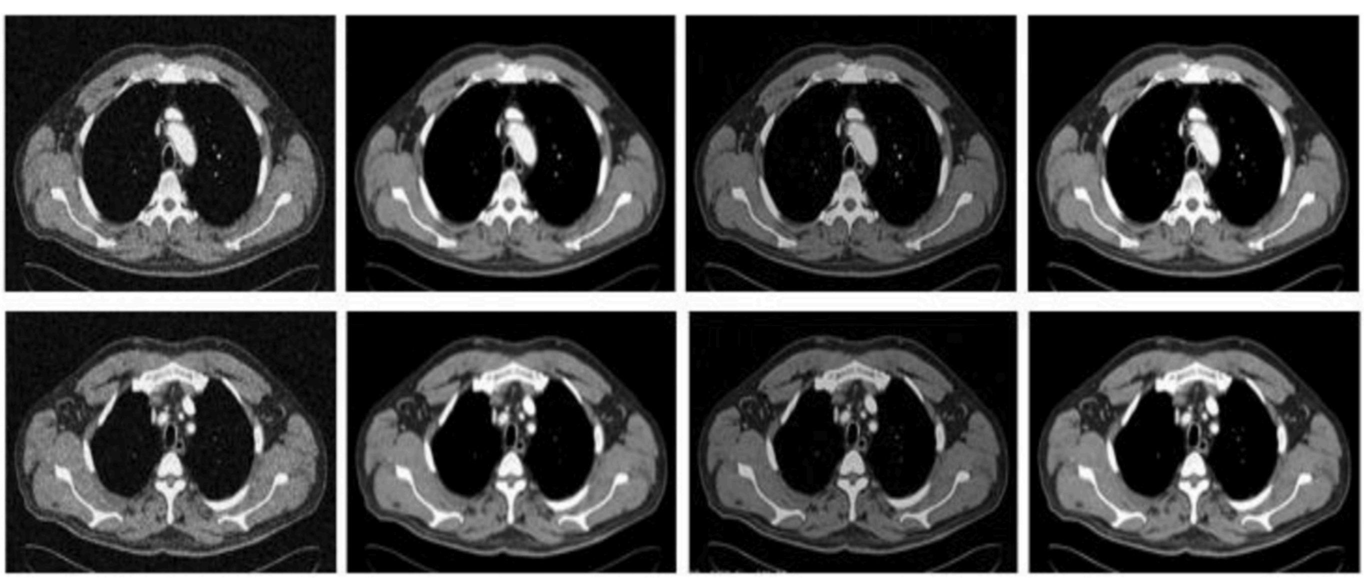

Fig. 8. QIN_LUNG_CT dataset (partial).

- Calculate the similarity or correlation between Query and Key.

Different functions and evaluation mechanisms can be introduced to calculate the similarity or correlation between Query and a certain Key_i. The most common ways to do this include taking the dot product of their vectors, finding their Cosine similarity, or applying an extra neural network.

- Carry out Softmax operation on the obtained similarity to normalize the weight.
- According to the calculated weight, the weighted sum of all Values is calculated to get the Attention vector.

In 2018, squeeze and exception network (SE-Net) [26] showed excellent performance in the ImageNet classification competition and won the champion of the classification competition in that year. Their goal is to improve the representation ability of the network using a lightweight network to learn the relationship between different channels in the feature. To achieve this, an SE block module is proposed in SE-Net. Through the SE block, we can learn to use global information, highlight the important information in the feature channel in the figure

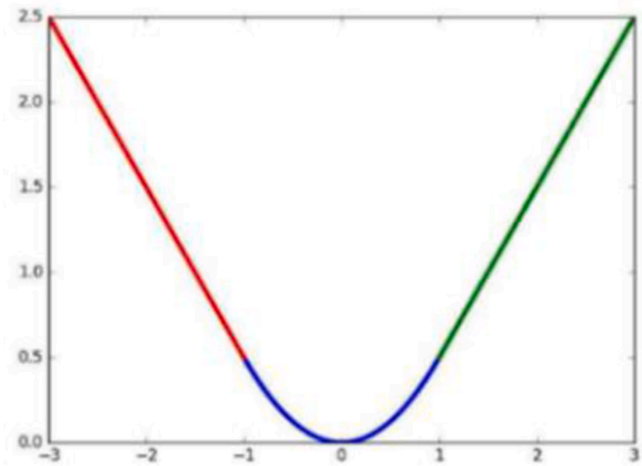

Fig. 9. Smooth L1 loss function.

and reduce the impact of unimportant information. In 2016, Jaderberg et al. proposed a powerful new module called spatial transformer network (STN) [29]. SangHyun woo et al. proposed a lightweight attention module, convolutional block attention module (CBAM) [28], which is based on channel attention mechanism and spatial attention mechanism, and effectively integrates the two mechanisms into one module. The author adds CBAM module to the structures of ResNet and MobileNet [27]. This module is an effective attention module, which can be embedded into any CNN to improve its representation ability.

Accordingly, work on integrating the attention mechanism in U-net is also continuing to develop. For example, in the Attention U-Net [23] proposed by Ozan Oktay et al., in 2018, the attention mechanism was introduced into the U-Net. Before concatenating the features of each resolution in the encoder with the corresponding features in the decoder, an attention module was used to readjust the output features of the encoder. In our study, we also selected some of the latest developments in this field to conduct comparative experiments. For example, NBNet [24], authors proposed subspace attention (SSA), a Non-Local attention module, to explicitly learn basis generation and subspace projection. Z.D.Wang et al. proposed Uformer [25], a transformer based U-Net structure denoising network.

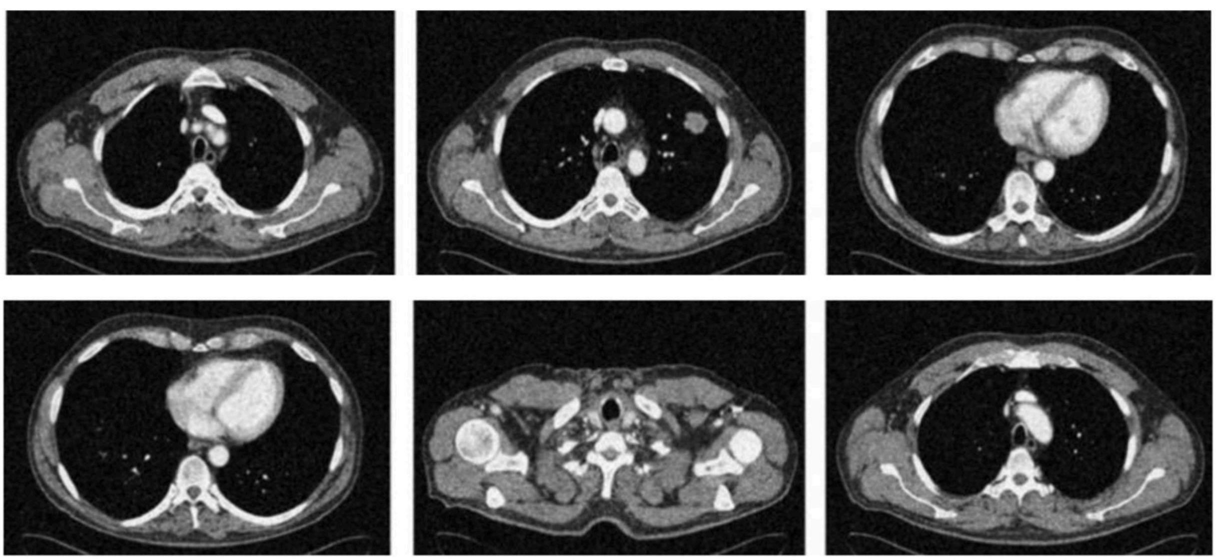

Fig. 10. CT images (before denoising).

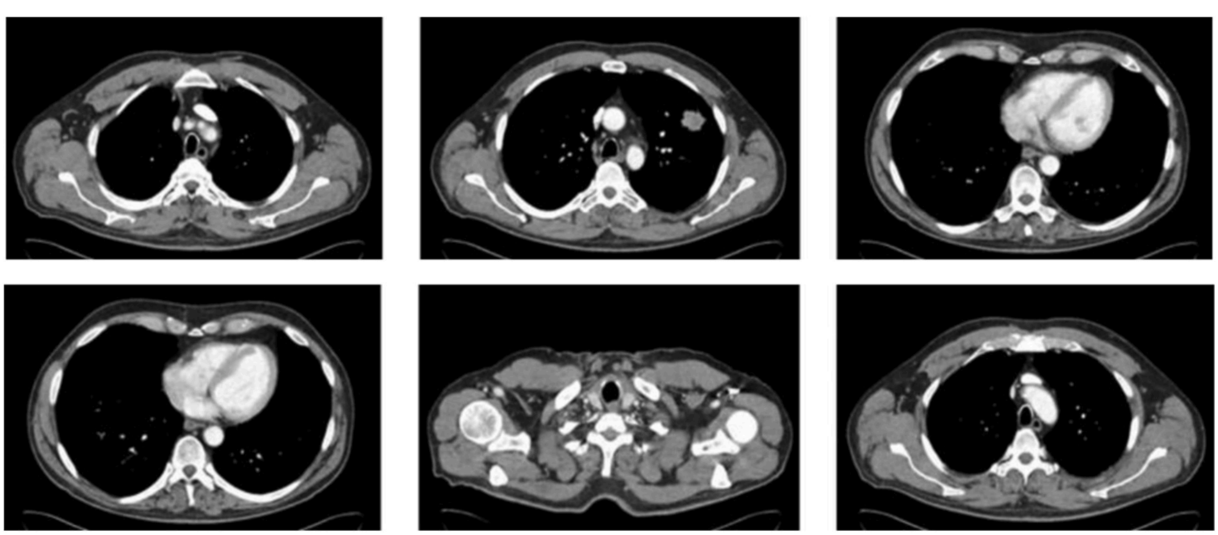

Fig. 11. Denoising results of CT images (after denoising).

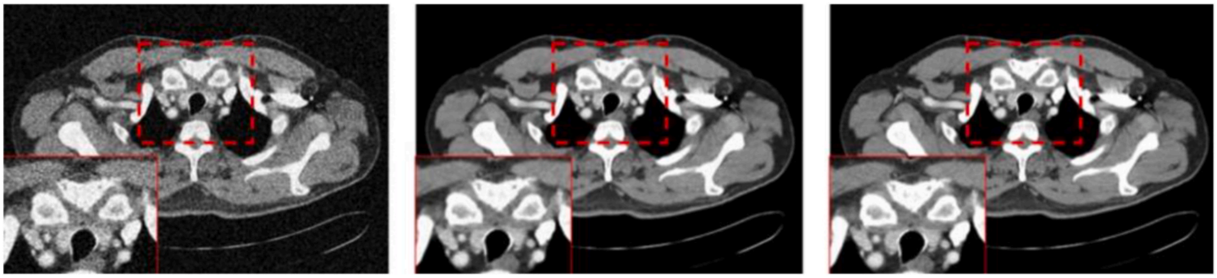

Fig. 12. Detailed features of CT images (noisy, denoised, and original images respectively, from left to right).

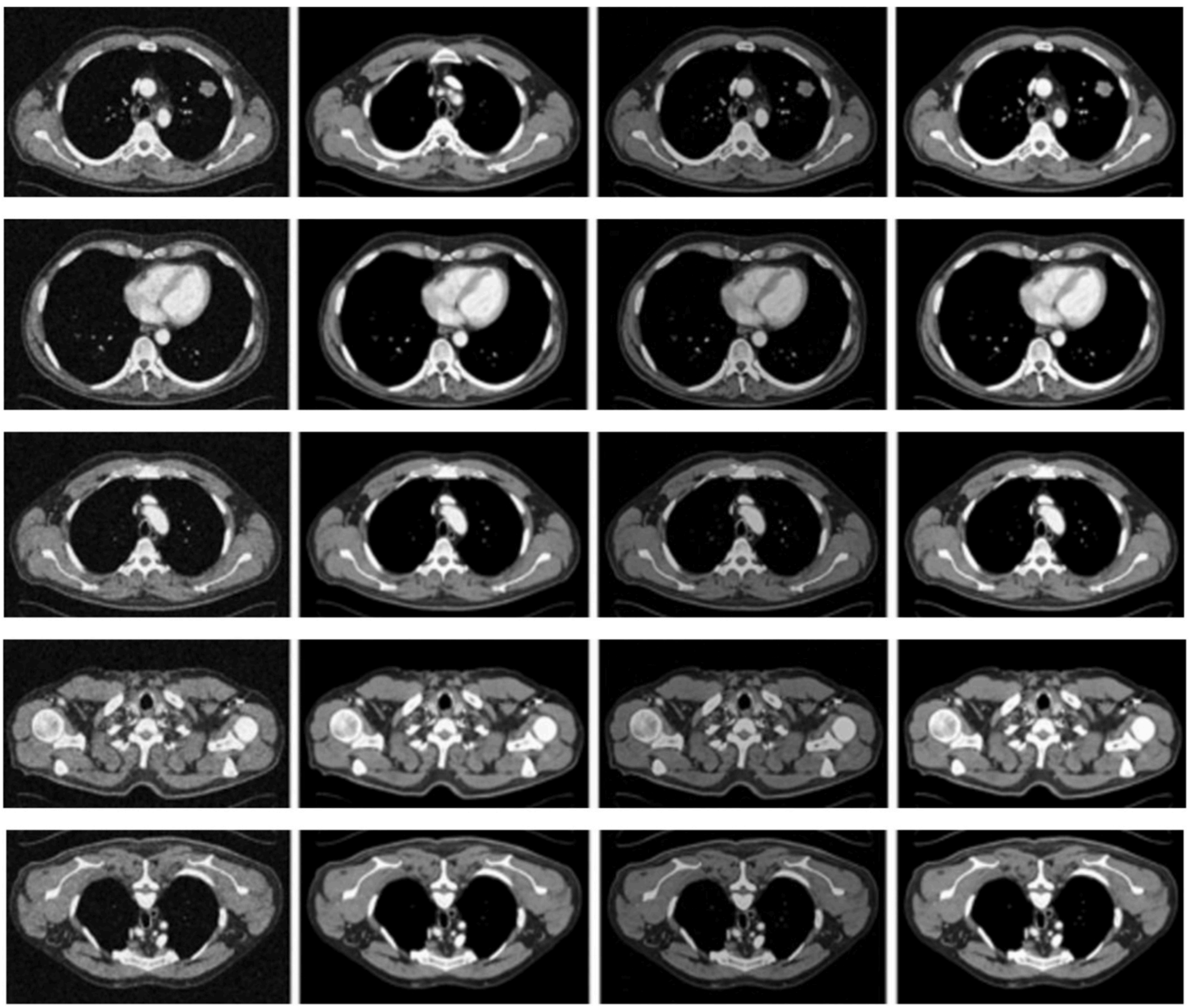

Fig. 13. Image changing process during the training stage. (From left to right are noisy image and resulting images with epoch = 5, 10, and 40, respectively).

Another common method is to learn multi-scale features to obtain rich data representation. The network based on attention mechanism has a strong function of information extraction. Usually, the network with attention module has two branches, that is, characteristic branch and attention branch. Feature branching is similar to traditional networks, in which neural structures extract features from data. In the attention branch, the network quantifies the importance of the input features concerned. However, the attention mechanism enables the network to focus on specific and context-related feature subsets. Following the introduction of the attention mechanism, several studies have been conducted to develop the most advanced network, which will enrich the attention mechanism and outperform the previous attention network. Many experimental results demonstrated that the attention mechanism can reduce over adaptation and improve accuracy.

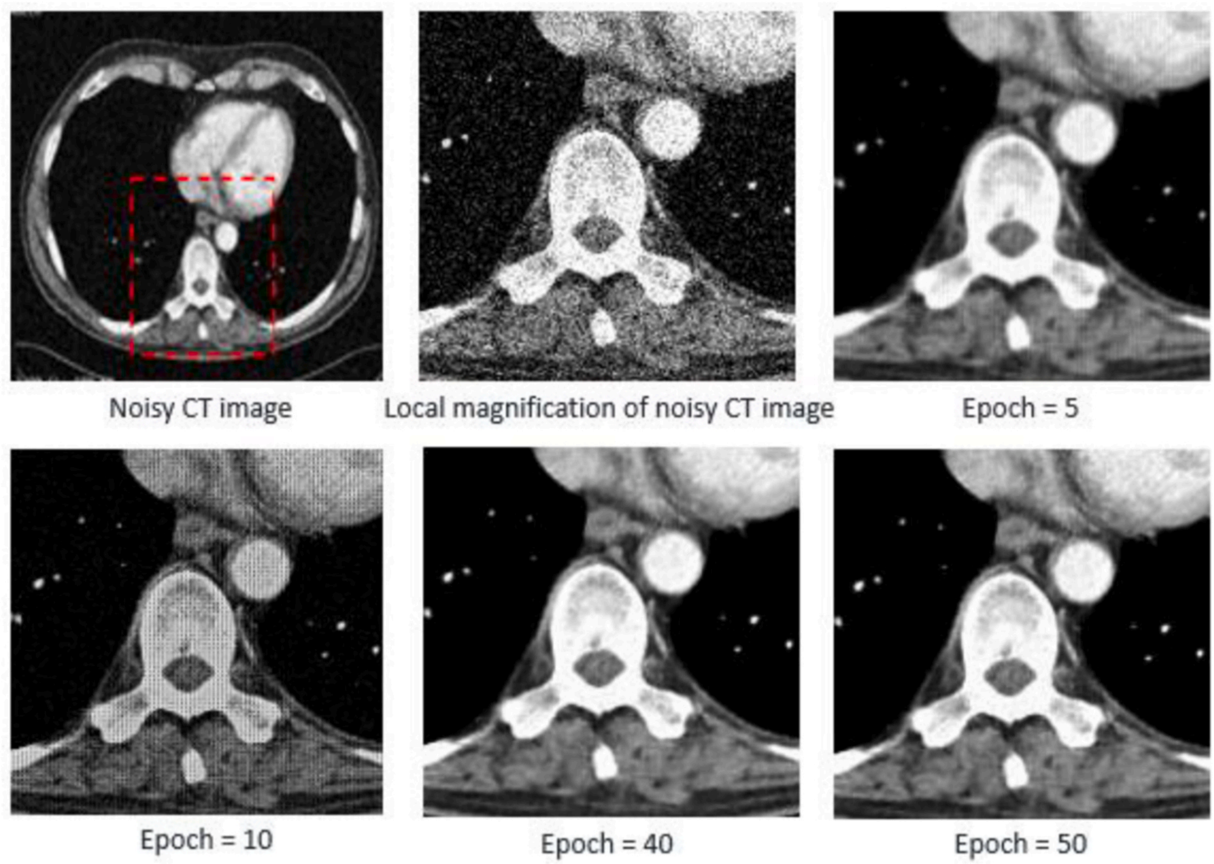

Fig. 14. Changes of image-detailed features during training.

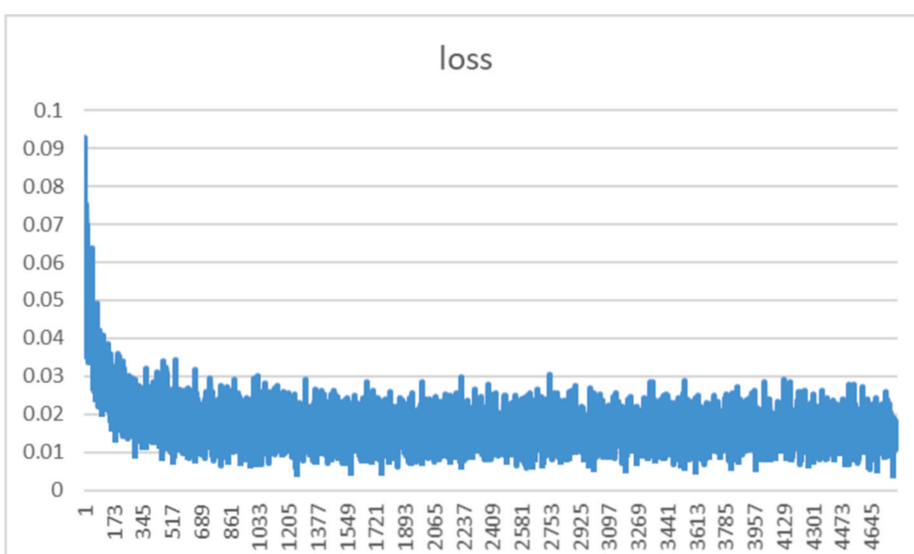

Fig. 15. Loss function during training.

3. U-network and multi-attention-based denoising of CT images

3.1. U-net with multi-channel and multi-attention mechanism

Although numerous experiments have demonstrated that U-Net is an excellent model for image information extraction, the results obtained using U-Net to denoise CT images directly are frequently insufficient to meet the medical requirements. Based on the U-Net model [12] and its encoding and decoding idea, this study proposes a local attention module, a multi-feature channel attention module, and a hierarchical attention module. Adding a convolution attention mechanism to the U-Net can effectively improve the ability of the network to identify key features. In this study, each attention model is set between the compression and expansion paths of the U-Net. After each attention module, an enhancement learning module is added to merge the feature maps. Fig. 2 shows the overall framework of the network structure of

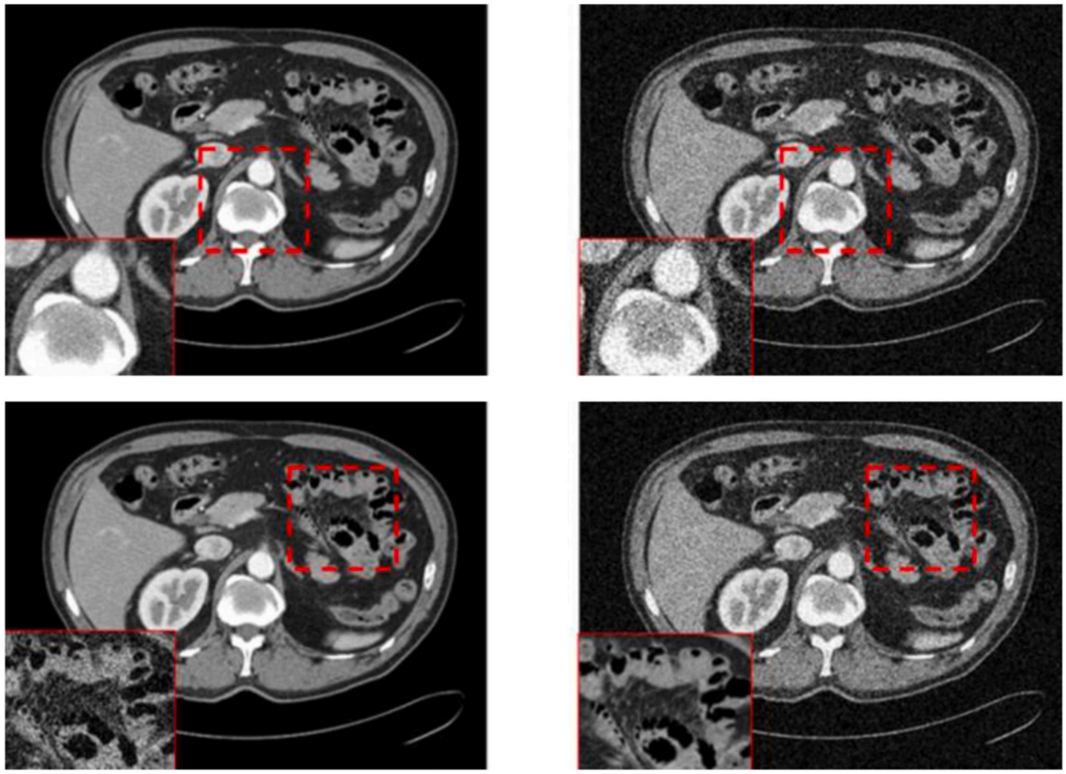

Fig. 16. Detailed features of original CT image and noisy CT image.

this study.

The contracting path of the de-noising network consists of two blocks. Each block on the contracting path is an unpadded convolutional layer of every two 3 × 3 followed by a maximum pooling layer of 2 × 2 (max-pooling layer: step size 2), and the original image is down-sampled by ReLU activation function behind each convolution layer. The extended path consists of 2 blocks. In the deconvolution of the extended path, each step will have one convolution layer of 2 × 2 (the activation function is also ReLU) and two convolution layers of 3 × 3. At the same time, the up-sampling of each step will add the feature map from the corresponding contracted path. The last layer of the denoising network is a convolution layer of 1 × 1. In this paper, each attention model is set between the contracting path and the expansion path of U-Net network, and local attention module, multi-feature channel attention module and hierarchical attention module are introduced. After each attention module, the enhanced learning module is provided, and the feature maps obtained by the enhanced learning module are combined, and then passed to the expansion path after up-sampling. The basic idea of each module in this proposed de-noising network will be introduced and details of local attention module, multi-feature channel attention module and hierarchical attention module will be presented in detail in the following sections.

3.2. Local attention module

In the clinical diagnosis and analysis of CT images, some local structures of the image can effectively help doctors make accurate medical diagnoses. For example, the cavity sign in lung CT images is one of the important bases for doctors to diagnose lung cancer. We believe that the CT image to be denoised must focus on the detailed features of the lung along with local information. Therefore, this study proposes a local attention module, as shown in Figs. 3–2. Inspired by the U-Net structure, biomedical image processing generally must localize the surrounding information, which means that feature pixel extraction alone is not enough to solve this task. The context extracted from the image must then be applied to each pixel. Because the required output resolution is sometimes significantly large, and considering that the model based on attention mechanism is a lightweight and embeddable neural network structure, this study designs the main architecture of the local attention module based on the encoding and decoding idea of U-Net. On the left side of the structure is a series of down-sampling operations composed of convolution and maximum pooling. The U-Net makes this part a compressed path. The compression path in this study is composed of two blocks. Each block uses three 3 × 3 convolutions and one 2 × 2 maximum pool down-sampling. After each convolution, the ReLU activation function is used for nonlinear processing. Because the convolution kernel of 2 × 2 is selected for down-sampling, the size of the feature map obtained after each down-sampling is half of the original one. The purpose of down-sampling is to fuse the peripheral position information of the feature map to the feature channel. The number of channels of the feature map after down-sampling each time is doubled. Therefore, it can be seen that the size of the feature map changes after each down-sampling. The structure on the right side of the local attention module is similar to that on the left. In U-Net, this part is called the extension path. It is also composed of two blocks. At the beginning of each block, the size of the feature map of the upper layer is expanded twice by deconvolution, and the number of channels of the feature map is doubled. Then, the output feature maps of the blocks with symmetrical position compression paths in the local attention module are combined.

In the local attention module, to keep more local information in the channel of the pixel, the structure shown in Fig. 3 is added at the symmetrical position between the compression and expansion paths, the size of the feature map output from the first block in the compression path is compressed, and the two-dimensional data on each channel in the feature map are compressed into one-dimensional data. That is, the characteristic graph $F \in R^{H\times W\times C}$ of the first block in the compression path compresses the size of $N\times C$, i.e., $F' \in R^{N\times C}$, where $N = H\times W$. Transpose the first F' and perform matrix cross matching with the second

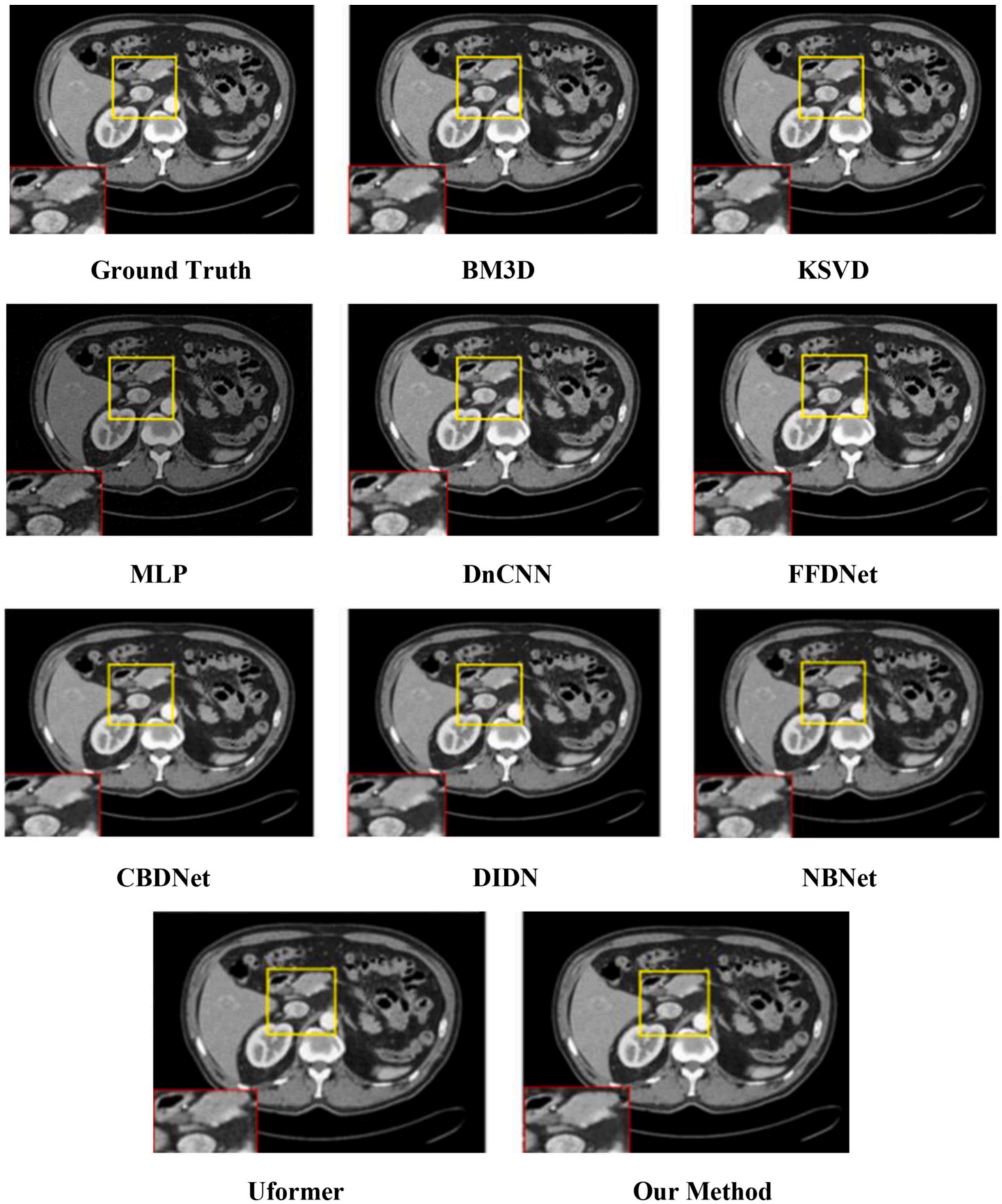

Fig. 17. Comparisons of the denoising effects of different denoising methods on Qin_LUNG_CT.

F' to obtain F'', where $F'' \in R^{C\times C}$, and multiply the obtained F'' to obtain F_s of the Softmax function, expressed as Equation (1).

$$F_{s_{ij}} = \frac{\exp\left(F''_{ij}\right)}{\sum_{i=1}^{C} \exp\left(F''_{ij}\right)} \tag{1}$$

where $F_{S_{ij}}$ represents the value of row i and column j in the characteristic diagram F_s. The third F' and F_s are matrix cross multiplied, and the characteristic graph obtained after multiplication is dimension transformed to obtain F''', where $F''' \in R^{H\times W\times C}$. Finally, the feature graph F''' is spliced to the block of symmetrical position compression path in the local attention module.

3.3. Multi-feature channel attention module

CT images in the early stages of diseases often show that the characteristics of lesions are not obvious. For example, identifying pulmonary nodules at an early stage requires greater attention from radiologists because nodule density may have anatomical features

Table 1
Comparison of PSNR between the proposed method and other denoising networks on Qin_LUNG_CT.

Methods	BM3D	KSVD	MLP	DnCNN	FFDNet	CBDNet	DIDN	NBNet	Uformer	Our Method
σ = 10	31.8345	31.1347	29.8981	31.1344	32.6127	33.3614	34.4321	34.6621	34.7037	34.7329
σ = 25	27.7673	27.8478	27.1363	28.2367	30.0813	30.9787	31.1461	31.5753	31.6977	31.7835
σ = 35	26.3414	26.1298	26.1592	26.0311	28.6761	29.4727	29.6094	30.5784	30.6146	30.6234
σ = 45	25.5625	24.8931	25.8633	25.2426	27.4172	28.3051	28.9334	29.3372	29.2376	29.6134
σ = 50	24.2324	24.0981	24.4566	24.8914	27.0351	27.6135	27.5851	27.5895	27.6583	27.8951
σ = 55	23.1441	23.5366	23.8912	24.4166	26.8581	26.7452	26.8581	26.9349	27.0057	27.1089
σ = 60	23.3413	23.1363	23.5131	24.1358	25.4912	25.8681	26.1767	26.0715	26.1036	26.0872
σ = 70	21.3144	23.1029	22.2328	23.0154	25.1234	25.8924	25.6264	25.8433	25.8779	25.9334

Table 2
SSIM comparison between the proposed method and other denoising networks on Qin_LUNG_CT.

Methods	BM3D	KSVD	MLP	DnCNN	FFDNet	CBDNet	DIDN	NBNet	Uformer	Our Method
σ = 10	0.8935	0.8512	0.8221	0.8501	0.9031	0.9149	0.9211	0.9268	0.9271	0.9293
σ = 25	0.7881	0.7923	0.7501	0.7985	0.8233	0.8445	0.8566	0.8594	0.8631	0.8654
σ = 35	0.7201	0.7121	0.7162	0.7031	0.7792	0.7994	0.8081	0.8105	0.8238	0.8205
σ = 45	0.6631	0.6013	0.6881	0.6401	0.7331	0.7664	0.7796	0.7874	0.7895	0.8077
σ = 50	0.6102	0.6031	0.6331	0.6513	0.7544	0.7407	0.7379	0.7382	0.7447	0.7502
σ = 55	0.5364	0.5621	0.5703	0.6294	0.7432	0.7318	0.7491	0.7488	0.7569	0.7584
σ = 60	0.4912	0.4831	0.5004	0.5991	0.6549	0.6893	0.7264	0.7042	0.7181	0.7132
σ = 70	0.3891	0.4771	0.4495	0.4752	0.6031	0.6972	0.6863	0.6994	0.7065	0.7103

Table 3
Comparisons between the proposed method and other denoising networks on Mayo.

Methods	LDCT	DnCNN	CBDNet	DIDN	NBNet	Uformer	Our Method
PSNR	24.4688	28.2813	28.4354	28.8761	28.7401	28.7354	28.9163
SSIM	0.8246	0.8513	0.8541	0.8590	0.8562	0.8547	0.8602

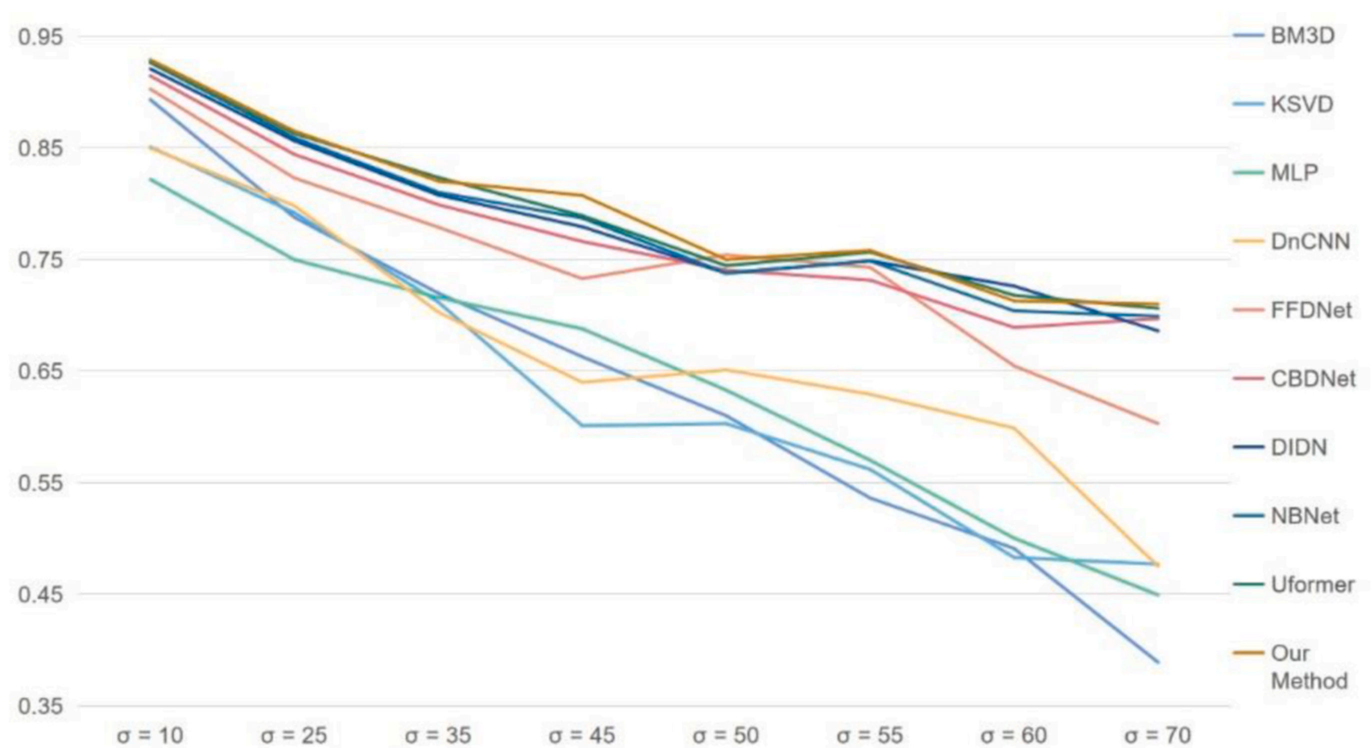

Fig. 18. Comparison of PSNR values.

similar to other lung structures. Computer aided detection (CAD) systems are an alternative to automatic lung lesion detection to help radiologists overcome the problem of routine readings. Because the candidate features and their surrounding environment have similar visual features, the key task of CAD is to accurately identify the candidate feature boundary in CT images. Additionally, extracting the salient

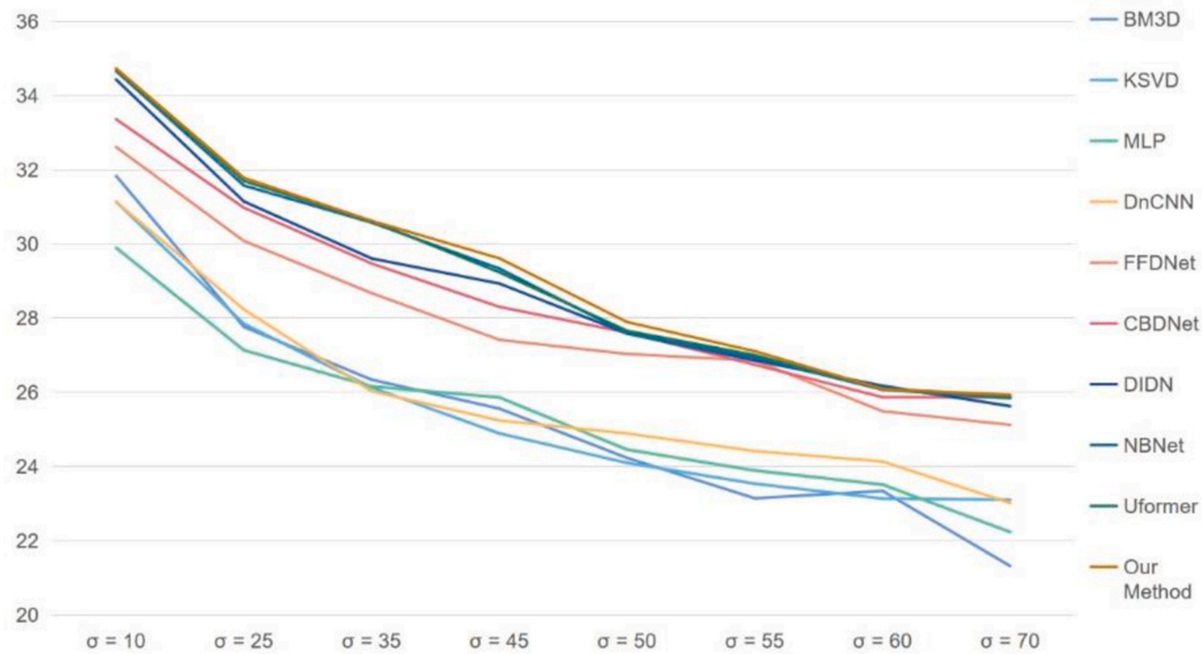

Fig. 19. Comparison of SSIM values.

features of candidate regions also plays an important role. For lung feature extraction, traditional manual features are usually used to extract features. However, this process is time-consuming and complex. We believe that in the CT image to be denoised, we must use deep learning to build a lightweight module to automatically learn and extract features and suppress some invalid information. Therefore, this study proposes a multi-feature channel attention module. Inspired by the SE-Net structure, the multi-feature channel attention module is based on the feature map after weighting the features extracted by different convolution cores, and adds different weights to each channel in the feature map according to different tasks. Fig. 4 shows the multi-feature channel attention module structure.

The network structure of the model inserts a bypass branch on a forward channel. On this branch, 2×2 maxpool is used, and the input characteristic graph F, ($F \in R^{H \times W \times C}$), is down-sampled to F', ($F' \in R^{(H/2) \times (W/2) \times C}$). In the image processing, convolution filters of different sizes will extract the feature information in different ranges. This study uses three different sizes of filters to extract features for F', which are the convolution filters of 1×1, 3×3, and 5×5. Two Conv + ReLU layers are used on each channel. Considering that the attention module must be lightweight and embeddable, zero filling operation (padding of zero) is not performed during the convolution operation. To reduce the impact of the limited neighborhood size, the average pooling is used after two Conv + ReLU layers. After pooling, three feature maps with a size of $1 \times 1 \times C$ are obtained. The three feature maps are spliced in the w dimension to obtain F'' ($F'' \in R^{1 \times 3 \times C}$). Each channel of the combined feature map F'' contains the feature information extracted using different convolution filters with different sizes to fuse the feature map information.

The convolution kernel of 1×2 is used for convolution, and through ReLU, the size of the characteristic graph is $1 \times 3 \times C$, and then the convolution kernel of 1×2 and ReLU are used to obtain the characteristic graph F''', ($F''' \in R^{1 \times 1 \times C}$). The feature map F'''' and the input feature map F have the same number of channels. Multiplying the two-dimensional data of each channel of the input feature map by the number of corresponding channels in F'''', produces feature map F''''.

3.4. Hierarchical attention module

The module based on the attention mechanism is a lightweight and embeddable module, which leads to the shallow depth of the previously proposed attention module. It is easy to ignore the importance of network depth. Although a deep network may lead to gradient disappearance or gradient explosion of the network, it will also make the attention module heavyweight. However, more valuable information in the input feature map can be extracted using the depth CNN [13]. Therefore, to make the network retain more effective information and ensure the lightweight of the network, this study proposes a hierarchical attention model as shown in Fig. 5.

To make the network lightweight and embeddable, instead of using the traditional convolution method, the deep separable convolution [14] (Depthwise Conv. in Fig. 6) is used. The basic idea of deep separable convolution is to divide the traditional convolution layer into two network layers, namely, channel-by-channel convolution layer and point-by-point convolution layer. The first layer is used to fuse the spatial information on each channel in the feature map. The second layer fuses different information on each channel in the feature map. They are shown in Fig. 6.

Depth convolution first calculates the convolution of each channel in the feature map, that is, channel-by-channel convolution. For feature channel processing, a single-layer convolution kernel is used. The number of convolution kernels must be equal to the number of channels in the previous layer. For example, after calculating the characteristic diagram of the three channels, the characteristic diagram of the three channels is obtained. Following channel-by-channel convolution is a point-by-point convolution, and the convolution kernel used in point-by-point convolution is $1 \times 1 \times M$. Therefore, the convolution operation in this stage will weigh and combine the feature map obtained in the previous stage in the dimension of feature channel. Therefore, the number of convolution kernels and output characteristic subgraphs in the pointwise convolution stage is equal. Using deep separable convolution can greatly reduce the network training parameters and speed up the parameter training without losing the network performance. Assume that the size of the input feature is $D_F \times D_F \times M$, the size of the output feature graph is $D_F \times D_F \times N$, where D_F is the width and height of the feature graph, N and M refer to the number of channels. For an ordinary

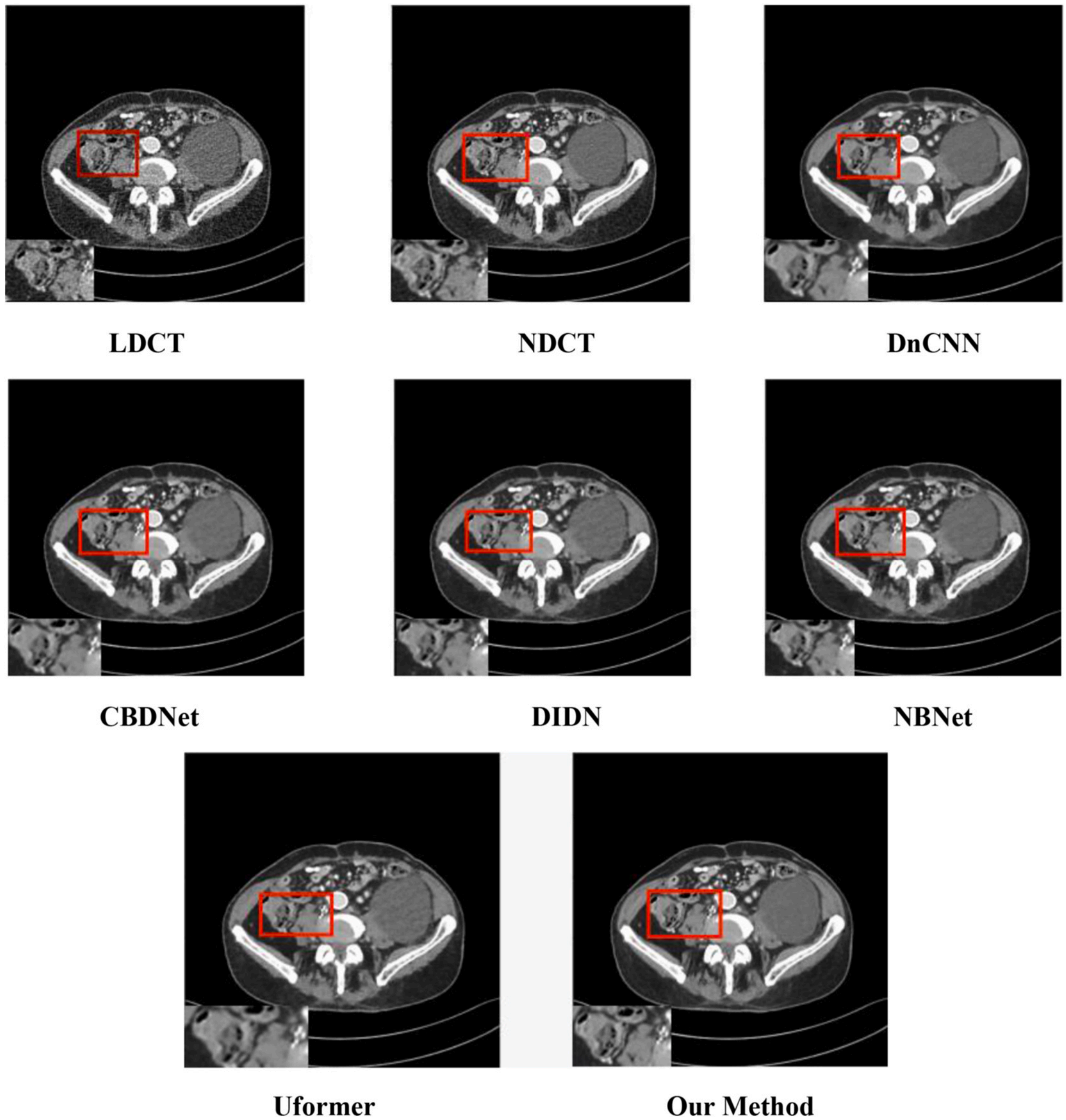

Fig. 20. Comparison of the denoising effects of different denoising methods on Mayo.

Table 4
Comparisons of ablation evaluation indexes of the local attention module.

Methods	①	②	③	SSIM	PSNR
Our method-①②③	×	×	×	0.8585	31.1082
Our method-②③	✓	×	×	0.8741	31.6314
Our method-①	×	✓	✓	0.9013	34.1983
Our method	✓	✓	✓	0.9293	34.7329

Table 5
Comparison of the ablation evaluation indexes of multi-feature channel modules.

methods	①	②	③	SSIM	PSNR
Our method-①②③	×	×	×	0.8585	31.1082
Our method-①③	×	✓	×	0.8901	31.7526
Our method-②	✓	×	✓	0.9032	33.4615
Our method	✓	✓	✓	0.9293	34.7329

convolution, the number of parameters is $D_k \times D_k \times M \times N \times D_F \times D_F$. The number of parameters of channel-by-channel convolution is $D_k \times D_k \times M \times D_F \times D_F$, and the number of network parameters of point-by-point convolution is $N \times M \times D_F \times D_F$, and thus the number of parameters of deep separable convolution is expressed as Equation (2):

$$D_k \times D_k \times M \times D_F \times D_F + N \times M \times D_F \times D_F \tag{2}$$

Table 6
Comparison of ablation evaluation indexes of 5-level attention modules.

method	①	②	③	SSIM	PSNR
Our method-①②③	×	×	×	0.8585	31.1082
Our method-①②	×	×	✓	0.9012	33.7812
Our method-③	✓	✓	×	0.9203	33.5968
Our method	✓	✓	✓	0.9293	34.7329

Through comparison, it can be found that the ratio relationship between deep separable convolution and ordinary convolution is expressed as Equation (3):

$$\frac{D_k \times D_k \times M \times D_F \times D_F + M \times N \times D_F \times D_F}{D_k \times D_k \times M \times N \times D_F \times D_F} = \frac{1}{N} + \frac{1}{D_K^2} \tag{3}$$

To make the hierarchical attention model learn and retain more feature information in the compressed path and extended path, this study splices the feature map output by each block with the top-level feature map in the extended path. Bilinear interpolation is used to expand the size of the output feature graph in the expansion path to completely match the size of the feature graph at the top level.

3.5. Enhanced learning module

The enhanced learning module is composed of eight convolutions, batch normalization and PReLU, and the size of the convolution kernel in each layer is 3 × 3. The number of filters is 256, that is, 256 feature maps are generated after the convolution operation in each layer. The enhanced learning module enables the network to learn and retain more feature information by increasing the depth. The jump connection mode is adopted between the Conv + BN + PReLU layers in the enhanced learning module, which can effectively learn the image feature information between the network layers. Simultaneously, it can also effectively alleviate the gradient explosion and gradient disappearance. Jumping feature information transmission can enhance the feature learning ability of the network. The enhanced learning network module is shown in Fig. 7.

4. Experimental studies of the proposed CT denoising method

The pseudocode of the proposed method is shown in algorithm 1, and the code is publicly available at https://github.com/jundao520/model/blob/main/train.py.

Algorithm 1. The pseudocode for denoising network-

Algorithm 1 The pseudocode for denoising network

- Establish the backbone of denoising model with U-net
- Construct local attention module, multi-feature attention module and hierarchical module
- Introduce enhanced learning module
- Data preprocessing, the images participating in the training are cut into image block with size of 64*64

for number of training epochs do (spend about 10 min per epoch)

  for each batch do

   · Obtain noisy images and noiseless images from the datasets

   · Calculation of loss values with Smooth L1 loss function

   · Update network parameters by minimizing the error of loss function

   · The peak signal-to-noise ratio (PSNR) averaged over a batch was calculated

  End for

End for

4.1. Datasets

In this paper, we train and verify the proposed method on two different datasets. One lung CT image dataset used in the experimental studies comes from the medical image data provided in the cancer imaging Archive (TCIA, https://wiki.nci.nih.gov/pages/viewpage.action?pageId=39291472). TCIA is a service library that will identify and host many cancers' medical image archives available for researchers to download. Data are organized in the form of collection. The images of patients are usually related to common diseases (such as lung cancer), image morphology, type (such as MRI, CT, and digital histopathology) or research focus. It also provides image-related data, such as patient results, treatment details, genomics and expert analysis. Various medical image resources provided in the TCIA library are widely used in the research and diagnosis of medical diseases. The data storage format in the TCIA library is DICOM, which is converted into JPG format. To facilitate network training, the format converted image is gray transformed. In this work, the medical CT image dataset obtained from the TCIA database is Qin_LUNG_CT, and some dataset images are shown in Fig. 8.

In the image preprocessing stage, the initial dataset is divided into training, verification, and test datasets, of which the proportion of three datasets are 90%, 5%, and 5%, respectively. The size of the training and verification datasets are cut to 64 × 64, the number of segmented training set image blocks is 134907, and the number of verification dataset image blocks is 7494.

Another publicly released datasets comes from 2016 NIH-AAPM Mayo Clinic LDCT Grand Challenge (https://www.aapm.org/GrandChallenge/LowDoseCT/). This dataset includes low-dose CT images with 1.0 mm slice thickness and normal-dose CT images from 10 anonymous patients (about 500 slices per patient) and is often used for model training and testing. The data of patient L506 were selected for model effect verification, and the data of the remaining 9 patients were used for model training.

4.2. Neural network training

The training of deep learning network is carried out through an optimization process by minimizing the error obtained using a cost function or loss function, which represents the deviation between the network prediction result and the real result. To correct the deviation, the network uses the back-propagation algorithm to update the network weight according to their contribution to the calculation loss [5]. This process is repeated several times until a stop criterion is reached. The smooth L1 loss function was used in this study. The loss function combines the advantages of L2 and L1 and avoids their disadvantages [15].

Smooth L1 loss function combines the advantages that the L2 loss function can be derived everywhere and has the advantages of stability and robustness of the L1 loss function [18]. Mathematically, smooth L1 loss function can be expressed as Equation (4).

$$smooth_{L_1}(x) = \begin{cases} 0.5x^2, if|x| < 1 \\ |x| - 0.5, otherwise \end{cases} \tag{4}$$

Taking $\mathrm{f}(x) - y$ as the horizontal axis and smooth L1 as the vertical axis, the following smooth L1 loss function image can be obtained, as shown in Fig. 9.

4.3. Experimental environment

The training and experiment of the proposed method use the following hardware configuration, CPU: Intel i9-9900k@3.60 GHz, memory: 64 GB and graphics card: NVIDIA rtx2080 11 GB. In the Windows 10 system environment, the Pytorch framework is used to build and write the network framework. The NVIDIA GPU computing toolkit is version 10.1.

4.4. Experimental results

In this study, peak signal-to-noise ratio (PSNR) and structural similarity (SSIM) are used to evaluate the denoising results of CT images. The dataset used in this training is from Qin in TCIA_ LUNG_ Medical CT image dataset of CT. In this study, 90% of the dataset is used as the training data. The image data in the test set are used for the denoising network performance test. The data in the test set does not participate in the training of the denoising network. The trained denoising network is used to denoise the data in the test set, and the denoising results are shown in Figs. 10 and 11, respectively. The detailed features of noisy images, denoised images and original lung CT images are shown in Fig. 12.

In the denoising network training stage, the Adam optimizer with an initial learning rate of 0.0001 is used for training, and the learning rate is halved every three cycles. Epoch is set to 50. In the process of denoising network training, changes of the images in the training process are recorded in the form of resulting images, as shown in Fig. 13. The changes in image-detailed features during training are shown in Fig. 14. During the process of training iteration of the denoising network, the parameters in the network are updated and optimized to improve the denoising performance of the network.

In this study, epoch is set to 50 when training the denoising network. During the process of denoising network training, 96 iterative updates were set in each epoch. The loss results calculated by each iteration of the lung CT image are shown in Fig. 15.

4.5. Comparisons with the state-of-the-art denoising networks

To show the performances and advantages of the proposed method, we will compare it with the most commonly used and popular denoising methods, including BM3D, KSVD, and MLP, and state-of-the-art methods, including DNCNN [10], FFDNet [11], CBDNet [19], DIDN [20], NBNet [24], Uformer [25]. Fig. 16 shows the detailed features of the original CT image and noisy CT image. Figs. 17 and 20 show the denoising effects with different denoising methods on the two different datasets respectively. The comparisons of PSNR and SSIM evaluation indexes on the dataset Qin_LUNG_CT of different methods are shown in Table 1 and Table 2, respectively. Comparison results on the dataset Mayo Clinic LDCT Grand Challenge are shown in Table 3. The algorithm proposed in this study improves both PSNR and SSIM values. From the comparison results of these indicators, the proposed method played a certain role in improving the performance compared with the most popular image denoising methods. It can be concluded that the performance of other non-machine learning denoising methods or ordinary non-U-Net variant networks has been improved. Figs. 18 and 19 graphically show the comparisons of network evaluation index using line charts and the improvements of PSNR and SSIM values of the algorithm proposed in this study. Fig. 17 shows that the denoising effect of KSVD is the worst, the denoising result is smooth with many texture details, and its boundary is particularly blurred. The denoising result of this proposed method can clearly restore valuable information from the noisy image, which is the closest to the ground truth.

4.6. Ablation experiments and analysis of attention module

To verify the role of each attention module proposed in this study, several groups of ablation experiments were conducted to observe the effect of the attention module on network performance. Through the analysis of data, the effectiveness of the functions of each attention module in the network is verified.

4.6.1. Local attention module

To observe the influence of the local attention module on the denoising network, we compared the effects of removing attention modules, removing attention modules other than the local attention modules, removing only the local attention module and removing the entire denoising network. The results of the comparative experiment are shown in Table 4. The experimental results show that adding the local attention module to the denoising network increases the local feature space correlation and can improve the denoising performance of the network. Among them, ① represents the local attention module, ② represents the multi-feature channel attention module, and ③ represents the hierarchical attention module.

4.6.2. Multi-feature channel attention module

To observe the influence of the multi-feature channel attention module on the denoising network, we compared the effects of removing the attention modules, removing the attention modules other than the multi-feature channel attention module, removing only the multi-feature channel attention module and removing the entire denoising network. Table 5 shows the comparative experiment results. The experimental results show that adding multi-feature attention module to the denoising network increases the extraction ability of feature map information and can improve the denoising performance of the network.

4.6.3. Hierarchical attention module

To observe the influence of the hierarchical attention module on the denoising network, we compared the effects of removing the attention modules, removing the attention modules other than the hierarchical attention module, removing only the hierarchical attention module and removing the entire denoising network. Table 6 shows the comparative experiment results. The experimental results show that adding a hierarchical attention module to the denoising network increases the network depth, retains more effective information, and improves the denoising performance of the network.

5. Conclusions

The CT image to be denoised must focus on the detailed features of the lung along with the local information. Therefore, this study proposes a local attention module to localize the peripheral information of the feature map and calculate each pixel from the context extracted from the feature map. To build a lightweight module to automatically learn and extract features and suppress some invalid information, this study proposes a multi-feature channel attention module. The module can add different weights to each channel in the feature map according to different tasks. Deep CNN can extract more valuable information from the input feature map. To make the network retain more effective information and ensure the lightweight of the network, this study proposes a hierarchical attention model. This study also uses the enhanced learning module to increase the depth of the network. Experimental results show that this method can effectively remove the noise from lung CT images while improving image quality.

Data availability statement

We confirm that the data is available upon reasonable request.

Declaration of competing interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Acknowledgements

This work is partially supported by the National Natural Science Foundation of China (60974042).

References

[1] McLeavy C, Chunara M, Gravell R, Rauf A, Cushnie A, Talbot CS et al. The future of CT: deep learning reconstruction. Clin. Radiol.. 2021(76):407-415.
[2] A novel wavelet thresholding method for adaptive image denoising, in: I. Hussain, H. Yin (Eds.), Proceedings of 3rd International Symposium on Communications, Control and Signal Processing, 2008.
[3] A. Qayyum, S.M. Anwar, M. Awais, M. Majid, Medical image retrieval using deep convolutional neural network, Neurocomputing (266) (2017) 8–20.
[4] G. Habib, S. Qureshi, Optimization and Acceleration of Convolutional Neural Networks: A Survey, Journal of King Saud University-Computer and Information Sciences, 2020.
[5] Ceyhan O. Algorithmic Complexities in Backpropagation and Tropical Neural Networks. arXiv:210100717. 2021.
[6] Deep residual learning for image recognition, in: K. He, X. Zhang, S. Ren, J. Sun (Eds.), Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, CVPR), 2016.
[7] V. Jain, S. Seung, Natural image denoising with convolutional networks, in: Advances in Neural Information Processing Systems, 2008.
[8] U. Schmidt, S. Roth, Shrinkage fields for effective image restoration, in: Proceedings of IEEE Conference on Computer Vision and Pattern Recognition, CVPR), 2014.
[9] Y. Chen, T. Pock, Trainable nonlinear reaction diffusion: a flexible framework for fast and effective image restoration, IEEE Trans. Pattern Anal. Mach. Intell. (39) (2017) 1256–1272.

[10] K. Zhang, W. Zuo, Y. Chen, D. Meng, L. Zhang, Beyond a Gaussian denoiser: residual learning of deep cnn for image denoising, IEEE Trans. Image Process. (26) (2017) 3142–3155.
[11] K. Zhang, W. Zuo, L. Zhang, FFDNet: toward a fast and flexible solution for CNN-based image denoising, IEEE Trans. Image Process. (27) (2018) 4608–4622.
[12] O. Ronneberger, P. Fischer, T. Brox, U-net: convolutional networks for biomedical image segmentation, in: Proceedings of the International Conference on Medical Image Computing and Computer-Assisted Intervention, Springer, 2015.
[13] A. Khan, A. Sohail, U. Zahoora, A.S. Qureshi, A survey of the recent architectures of deep convolutional neural networks, Artif. Intell. Rev. (53) (2020) 5455–5516.
[14] F. Chollet, Xception: deep learning with depthwise separable convolutions, in: Proceedings of IEEE Conference on Computer Vision and Pattern Recognition, CVPR), 2017.
[15] X. Jin, L. Wu, X. Li, X. Zhang, J. Chi, S. Peng, et al., ILGNet: inception modules with connected local and global features for efficient image aesthetic quality classification using domain adaptation, IET Comput. Vis. (13) (2019) 206–212.
[16] A survey on medical image denoising using optimisation technique and classification, in: D.K. Priya, B.B. Sam, S. Lavanya, A.P. Sajin (Eds.), Proceedings of the International Conference on Information Communication and Embedded Systems, ICICES), 2017.
[17] A survey on noise reduction techniques for lung cancer detection, in: J. Madhura, D.R. Babu (Eds.), Proceedings of the International Conference on Innovative Mechanisms for Industry Applications, ICIMIA), 2017.
[18] R. Girshick, Fast R-CNN. Proceedings of the IEEE International Conference on Computer Vision (ICCV), 2015, pp. 1440–1448.
[19] S. Guo, Z. Yan, K. Zhang, W. Zuo, L. Zhang, Toward convolutional blind denoising of real photographs, in: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, CVPR), 2019.
[20] Deep iterative down-up cnn for image denoising, in: S. Yu, B. Park, J. Jeong (Eds.), Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition Workshops, 2019.
[21] Ju Zhang, HaiLin Zhou, Yan Niu, JinCheng Lv, Jian Chen, Yun Cheng, CNN and multi-feature extraction based denoising of CT images, Biomed. Signal Process Control 67 (2021), 102545.
[22] Ashish Vaswani, Noam Shazeer, Parmar, et al., Attention is all you need, in: Proceedings of the 31st International Conference on Neural Information Processing Systems (NIPS), 2017.
[23] Ozan Oktay, Schlemper Jo, Loic Le Folgoc, Matthew Lee, et al., Attention U-net: learning where to look for the pancreas, Medical Imaging with Deep Learning (2018).
[24] S. Cheng, Y. Wang, H. Huang, D. Liu, H. Fan, S. Liu Nbnet, Noise basis learning for image denoising with subspace projection, in: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, CVPR), 2021.
[25] Zhendong Wang, Cun Xiaodong, Jianmin Bao, Wengang Zhou, Jianzhuang Liu, Houqiang Li, Uformer: a general U-shaped transformer for image restoration, in: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, CVPR), 2022.
[26] Jie Hu, Li Shen, Gang Sun, Squeeze-and-Excitation networks, in: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR), 2018, pp. 7132–7141.
[27] Andrew G. Howard, Menglong Zhu, Bo Chen, et al., Mobilenets: Efficient Convolutional Neural Networks for Mobile Vision Applications, 2017, 04861 arXiv: 1704.
[28] Sanghyun Woo, Jongchan Park, Joon-Young Lee, Kweon In So, CBAM: convolutional block attention module, in: Proceedings of the European Conference on Computer Vision, ECCV), 2018.
[29] Max Jaderberg, Karen Simonyan, Andrew Zisserman, Koray Kavukcuoglu, Spatial transformer networks, in: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, CVPR), 2016.
[30] Yang Dong, Jian Sun, BM3D-Net: a convolutional neural network for transform-domain collaborative filtering, IEEE Signal Process. Lett. (25) (2018) 55–59.
[31] Ying Tai, Jian Yang, Xiaoming Liu, Chunyan Xu, MemNet: a persistent memory network for image restoration, in: Proceedings of the IEEE International Conference on Computer Vision (ICCV), 2017.
[32] P. Liu, H. Zhang, K. Zhang, Multi-level wavelet-CNN for image restoration, in: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition Workshops, CVPRW), 2018.
[33] R. Chen, H. Zhang, J. Liu, Multi-attention augmented network for single image super-resolution, Pattern Recogn. 122 (2022), 108349.
[34] Y. Wang, X. Song, G. Gong, N. Li, A multi-scale feature extraction-based normalized attention neural network for image denoising, Electronics 10 (3) (2021) 319.
[35] H.-X. Dou, X.-M. Pan, C. Wang, H.-Z. Shen, L.-J. Deng, Spatial and spectral-channel attention network for denoising on hyperspectral remote sensing image, Rem. Sens. 14 (14) (2022) 3338.
[36] H. Zhang, Q. Lian, J. Zhao, Y. Wang, Y. Yang, S. Feng, RatUNet: residual U-Net based on attention mechanism for image denoising, PeerJ Computer Science 8 (2022), e970.
[37] J. Zhu, C. Yao, Y. Tang, Y. Gao, L. Zhou, H. Hu, MRDA-net: multiscale residual dense attention network for image denoising. International conference on artificial intelligence and security, ICAIS (2022) 213–230.