ORIGINAL ARTICLE

Re-UNet: a novel multi-scale reverse U-shape network architecture for low-dose CT image reconstruction

Lianjin Xiong[1] · Ning Li[1] · Wei Qiu[1] · Yiqian Luo[1] · Yishi Li[2] · Yangsong Zhang[1,3,4]

Abstract

In recent years, the growing awareness of public health has brought attention to low-dose computed tomography (LDCT) scans. However, the CT image generated in this way contains a lot of noise or artifacts, which make increasing researchers to investigate methods to enhance image quality. The advancement of deep learning technology has provided researchers with novel approaches to enhance the quality of LDCT images. In the past, numerous studies based on convolutional neural networks (CNN) have yielded remarkable results in LDCT image reconstruction. Nonetheless, they all tend to continue to design new networks based on the fixed network architecture of UNet shape, which also leads to more and more complex networks. In this paper, we proposed a novel network model with a reverse U-shape architecture for the noise reduction in the LDCT image reconstruction task. In the model, we further designed a novel multi-scale feature extractor and edge enhancement module that yields a positive impact on CT images to exhibit strong structural characteristics. Evaluated on a public dataset, the experimental results demonstrate that the proposed model outperforms the compared algorithms based on traditional U-shaped architecture in terms of preserving texture details and reducing noise, as demonstrated by achieving the highest PSNR, SSIM and RMSE value. This study may shed light on the reverse U-shaped network architecture for CT image reconstruction, and could investigate the potential on other medical image processing.

Keywords Low-dose CT · UNet · Deep learning · Image reconstruction

1 Introduction

Computed tomography (CT) system plays an important role in the field of radiology diagnosis. Due to noninvasive and convenience of CT, doctors can easily see the specific lesions in the patient's body through CT images, such as the size and shape of lesions [1]. The X-ray radiation employed in CT scans poses a potential threat to human health, as it can induce irreversible cellular and DNA damage during the scanning process [2]. Prolonged exposure to radiation increases the probability of cancer, particularly among children, who are more susceptible to radiation-induced damage [3]. The radiation dose received by conventional CT scanning is about 1.5–20 mSv (millisieverts), and the radiation dose received by low-dose CT (LDCT) scanning is about one quarter of that of conventional CT scanning [4]. Since the direct factor of radiation dose is X-rays [5], dose reduction can be achieved by two ways. The first one is to decrease the tube current or reduce the tube voltage. However, reducing the tube current results in lower signal-to-noise ratios of the projection data, which can make it difficult to identify areas with similar densities [6]. The second way for dose reduction is to shorten the scanning time by sparse sampling. However, CT images reconstructed from such projection data often contain streak artifacts [7]. These streak artifacts will greatly affect the subjective judgement of radiologists, and may even lead to misdiagnosis. Therefore, improving the quality of the final

✉ Yangsong Zhang
zhangysacademy@gmail.com

1 School of Computer Science and Technology, Laboratory for Brain Science and Medical Artificial Intelligence, Southwest University of Science and Technology, Mianyang 621010, China

2 Department of Respiratory and Critical Care Medicine, The First Affiliated Hospital of Chongqing Medical University, Chongqing 400016, China

3 NHC Key Laboratory of Nuclear Technology Medical Transformation (MIANYANG CENTRAL HOSPITAL), Mianyang 621000, China

4 Key Laboratory of Testing Technology for Manufacturing Process, Ministry of Education, Southwest University of Science and Technology, Mianyang 621010, China

LDCT image is a very promising and meaningful research topic, and many studies have achieved very good performance. We classify previous algorithms for LDCT image reconstruction into three categories, i.e., sinogram filtration, iterative reconstruction, and post-processing [8].

Sinogram filtration is a processing method based on raw data, also known as analytical reconstruction algorithm. This algorithm is designed to suppress noise in projection data by employing the filtering algorithm that reduces the signal-to-noise ratio of low-dose projection data. Over the past decade, numerous effective projection domain preprocessing methods have been proposed. Sahiner et al. proposed a method to effectively suppress the high-frequency noise in the reconstructed CT image by constructing a filter in the wavelet space [9]. Zhang et al. employed a methodology to segment projection data based on variations in noise intensity, which enabled the use of distinct filtering techniques for different noise intensities [10]. Compared with iterative reconstruction algorithms, these algorithms can significantly shorten the reconstruction time and have greater advantages in computing speed. However, the original projection data is sensitive to the processing method, which often leads to data overcorrection, resulting in the distortion of the reconstructed image and new artifacts.

Iterative reconstruction algorithms are among the most used methods for reducing noise in LDCT images. Compared to analytical reconstruction algorithms, they offer superior performance in terms of noise suppression and artifact elimination. By incorporating prior information in the image space, these algorithms can also enhance edge sharpness [11]. With the introduction of the compressed sensing techonology, Sidky and Pan used total variation (TV) minimization constraint reconstruction to effectively suppress and reduce the noise and artifacts of the LDCT images [12]. Browne et al. proposed the method of row update to realize projection reconstruction, which greatly improved the reconstruction speed [13]. Fessler et al. proposed that by constructing projection, the image alternate subspace accelerates the iterative convergence [14], but this makes each iterative calculation more complex, and the overall reconstruction time of the algorithm has not been significantly reduced. Although these reconstruction algorithms have effectively improved the final imaging quality, they all have a common disadvantage that they are computationally expensive. Even though there are some algorithms to speed up reconstruction, there is still a certain distance between iterative reconstruction algorithms and the need for rapid imaging in clinical needs.

In recent years, deep learning technology has made great achievements in many fields such as image segmentation, classification, detection, noise reduction, etc. Some researchers have also introduced this technology into LDCT image reconstruction, and achieved superior performance. For example, Chen et al. verified its effectiveness in LDCT image reconstruction by building a simple neural network [15]. In addition, there are numerous CNN-based network structures have been proposed, which has significantly contributed to the amelioration of LDCT image fidelity [16–20]. Yang et al. proposed the WGAN-VGG network [21], which uses the generative adversarial network (GAN) to achieve CT image reconstruction. The perception loss used in this network greatly improves the problem of image over-smooth [22]. Zhang et al. achieved good noise reduction effect by using CLEAR designed based on GAN [23]. They used multi-level consistency loss and made great progress in Peak signal-to-noise ratio (PSNR) and Structural Similarity (SSIM). Due to large amount of superparameters, it is difficult to train the model, which hinders the application. Wang et al. utilized the multi-head self-attention mechanism instead of the CNN and proposed a CTformer model, which played a significant role in suppressing noise and artifacts in LDCT images [24]. Liu et al. believed that noise has a strong correlation with deviant features, and construct the deviant feature sensitive noise estimate network, which also achieved very good noise reduction effects [25]. Bera et al. reduced the dependence on pairwise training of LDCT and NDCT by using self supervised methods, and also achieved good mapping from LDCT to NDCT [26]. Lu et al. introduced neural structure search for the first time to conduct LDCT image reconstruction, and achieved better noise reduction effect by searching for a better architecture [27].

While previous deep learning models have demonstrated significant progress in LDCT image reconstruction research, their model architectures are mainly based on the UNet network structure as shown in Fig. 1(a). This type of approach

(a) U-shape Network

(b) Reverse U-shape Network
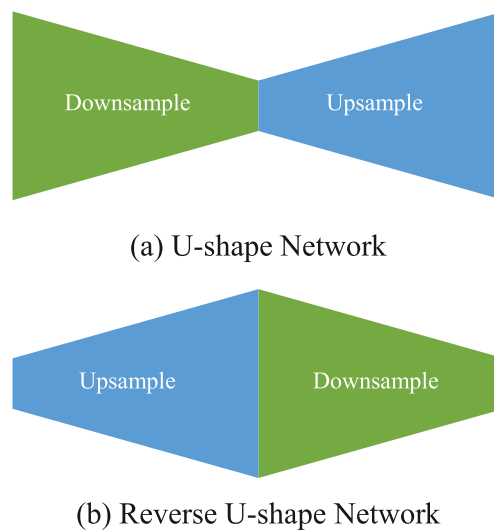
Fig. 1 Comparison of architectures of image restoration models. (a) U-shape architecture, (b) Reverse U-shape architecture. Some details of the model have been omitted, e.g., residual connection, feature fusion modules, the number of convolutional blocks, and etc

first downsamples to reduce the image size, and then upsamples to enlarge the image size. Reducing the image size will inevitably cause a loss of texture details, which is unacceptable for our medical images. As shown in Fig. 2, we conducted a rough simulation experiment. First, we reduced the image size of Fig. 2(a) to half of the original size through the Bilinear interpolation algorithm, and then restored the size through the Bilinear interpolation algorithm to get the result of Fig. 2(b). Compared with the original image, the image has become smoother. After calculating the absolute difference between the two images, as shown in Fig. 2(c), there has been a significant difference between the two images. In addition, we also use the Bilinear interpolation algorithm to enlarge Fig. 2(a) to twice the original size and then restore it to get the result of Fig. 2(d), which is almost the same as the original image. Figure 2(e) also shows that their absolute difference is relatively small. Inspired by this inspiration, in order to achieve the reconstruction of the CT image and preserve a greater amount of image information, we employ an upsampling-downsampling approach using a novel reverse U-shape network architecture as shown Fig. 1(b). The effectiveness of the proposed model is verified by extensive experiments on the public and widely used dataset.

The overall contributions of this paper can be summarized as follows:

1. To best of our knowledge, the reverse U-shape network architecture was introduced for the domain of LDCT image reconstruction for the first time. The enlargement of image feature map size should be taken as the first phase of the reconstruction process.
2. The proposed multi-scale feature extraction module facilitates the reconstruction of LDCT images.

2 Materials and methods

For the LDCT image reconstruction operation, it can be modeled with the following formula:

$$Y = X + n, \tag{1}$$

where n is the noise to be removed, $X \in R^{N \times N}$ denotes a LDCT image, $Y \in R^{N \times N}$ represents the corresponding NDCT image. The complete reconstruction process can be interpreted as the requirement to derive a model to realize the mapping from LDCT image X to NDCT image Y.

2.1 The proposed model

In this study, our goal is to design a network architecture as simple as possible to realize LDCT image reconstruction. Inspired by the structural design of U-Net [28] and RED-CNN [18], we developed a reverse U-shape network architecture, termed as Re-UNet. The model framework is illustrated in Fig. 3.

Fig. 2 The impact of employing the bilinear interpolation algorithm to initially magnify and subsequently restoration the image, as well as initially reduce and subsequently restoration it. (a) Origin image, (b) image magnification followed by restoration, (c) absolute difference between (a) and (b), (d) image reduction followed by restoration, (e) absolute difference between (a) and (d). In order to better see the difference, we enhanced the two images (c) and (e), adding 28 to their red channel value, 74 to their green channel value, and 152 to their blue channel value

(a) Origin Image

(b)

(c)

(d)

(e)
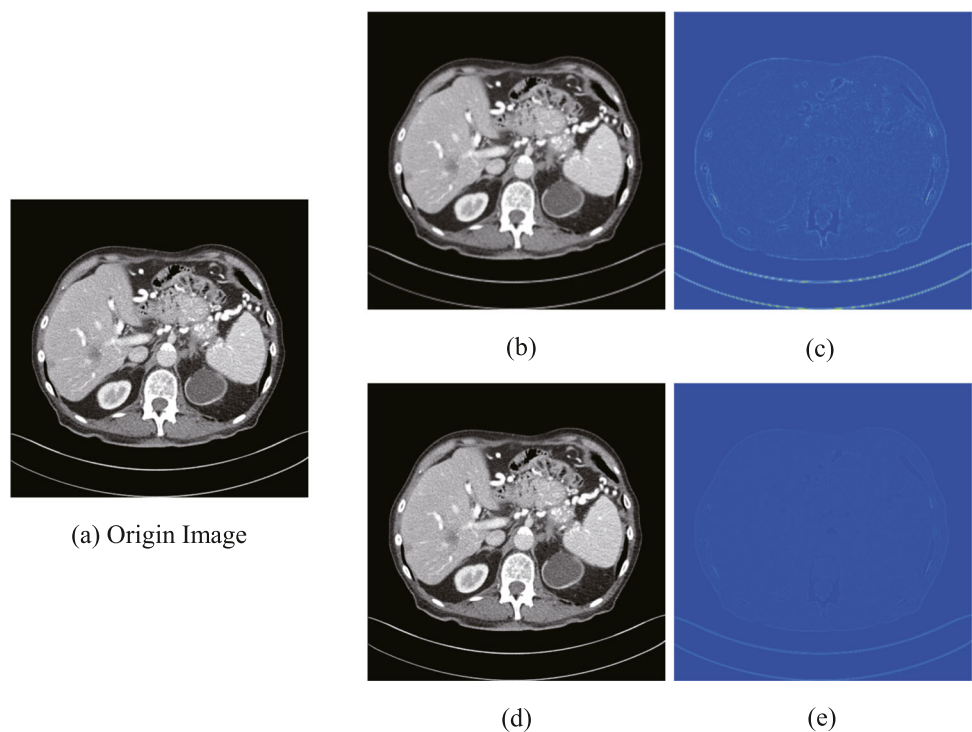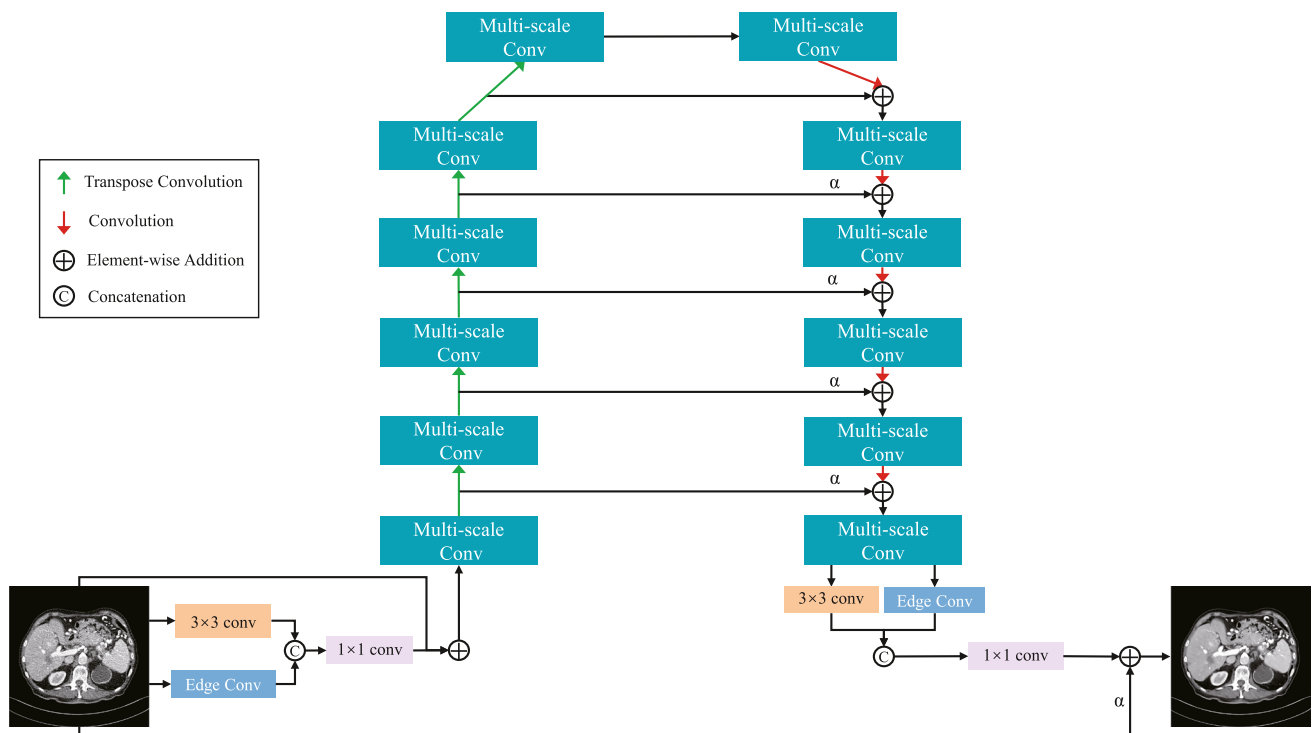
Fig. 3 Architecture of Re-UNet for LDCT images reconstruction. The proposed model features a reversed U-shape architecture, in which image processing is conducted through sequential upsampling and downsampling operations

In the initial stage of the network, two branches were used, one of which performed a 3×3 convolution for feature extraction, and another branch leveraged an edge extraction module to achieve image edge enhancement. Subsequently, the output features of the two branches are integrated via the concatenated operation. To diminish the computational complexity of subsequent convolutions, a 1×1 convolution was introduced, with the purpose of reducing the number of channels, and then integrated with the initial input to form the residual connection. After these preliminary operations, the feature information will be incorporated into the upsampling stage, which was the first key stage of the network. But before each upsampling step, we implemented a multi-scale feature extraction module aimed at refining the extracted features. The transposed convolution was used to achieve the steps of upsampling. Throughout the process of upsampling, we employed six multi-scale feature extraction modules and six transposed convolutions. After passing through the upsampling stage, the feature information will change from 512×512 to 536×536. In the downsampling stage, we used the convolution to achieve the purpose of gradually reducing the feature size. To reduce the risk of vanishing gradients, we used residual connections after each layer of downsampling. When performing feature fusion of residual connections, the lower-level features were multiplied by a learnable parameter α and added to the output of the downsampling process. Similar to the initial phase of the network,

we again extracted the feature edge before applying a 1×1 kernel convolution to reduce the number of channels to 1, and then obtained the final reconstructed NDCT image.

2.1.1 Edge extraction module

Considering the complexity and structure of CT images, clear edges of various organs or lesion areas in CT images are importance for doctors to make accurate diagnoses. Therefore, the edge enhancement operation on the LDCT image is benifitial for the subsequent reconstruction, and the following experimental results verify the feasibility of this idea. In the design of edge extraction module, we used the traditional sobel filter which is widely used in natural images. As shown in Fig. 4 (b), the two sobel convolution kernels represent the horizontal and vertical convolution kernels, respectively. To extract the edge features, the feature maps of the odd-valued channels were convolved a horizontal convolution kernel, and the feature maps of the even-valued channels were convolved with a vertical convolution kernel. Owing to the particularity of CT images, an additional learnable parameter β was introduced on the basis of the original convolution kernel, whose function was to enable the convolution kernel to better learn the edge extraction of CT images. Figure 4(c) shows the output results of CT images after vertical and horizontal edge extraction respectively. The extracted edges enable a clear visualization of the organ outlines.

Fig. 4 The sobel operator of the edge convolution module and the result of edge extraction. (a) A randomly selected LDCT image; (b) Sobel convolution kernel operator. (c) Result of the LDCT image edge extraction

(a) Origin Image

(b) Sobel Edge Convolution

(c) Edge Convolution Image
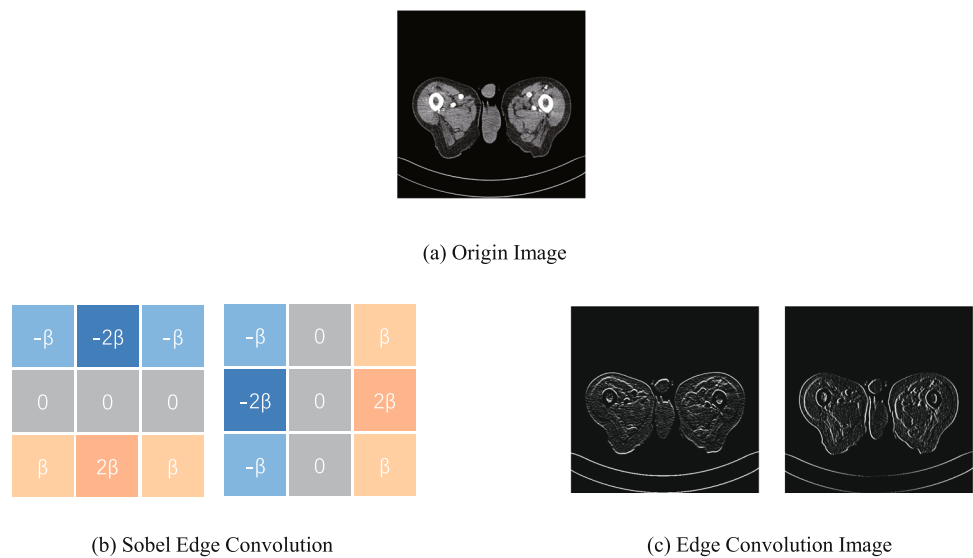
2.1.2 Multi-scale feature extraction modules

Considering the notable disparities in body tissue and organ sizes depicted in CT images, it could be suboptimal to exclusively employ a single kernel size throughout the entire neural network architecture. In view of this, we developed a multi-scale convolution module to enhance the network's ability to effectively address the challenge in the reconstruction of different tissues and organs. Figure 5 shows the specific design of our multi-scale feature extraction module, which contains five different convolution kernels of different sizes, i.e., 1×1, 3×3, 5×5, 7×7 and 9×9, respectively. In order to reduce the parameter amount of the model, we used depthwise convolution for the kernels of sizes 5×5, 7×7 and 9×9, respectively. After each convolution operation, a LeakyRelu activation function was used to realize the nonlinear characteristics of the model. Finally, the features after multi-scale convolutions were concatenated together, and then further fused through a convolution with a kernel size of 3×3. In addition, the number of channels can be kept consistent with that of the input.

2.1.3 Loss function

During model training, we use a hybrid loss function to measure the difference between the output of the model and NDCT images, which consists of two parts. In the first part, we used the mean absolute error (MAE). This loss function can well reflect the pixel-wise differences between the NDCT image and the reconstructed image by the model. By utilizing MAE, the network's reconstruction outcome has the potential to yield a superior PSNR value. The MAE loss (L_{MAE}) can be expressed by the following formula:

$$L_{MAE} = \frac{1}{N} \sum_{i=1}^{N} |x_i - y_i| \tag{2}$$

Fig. 5 The structure of the multi-scale feature extraction module
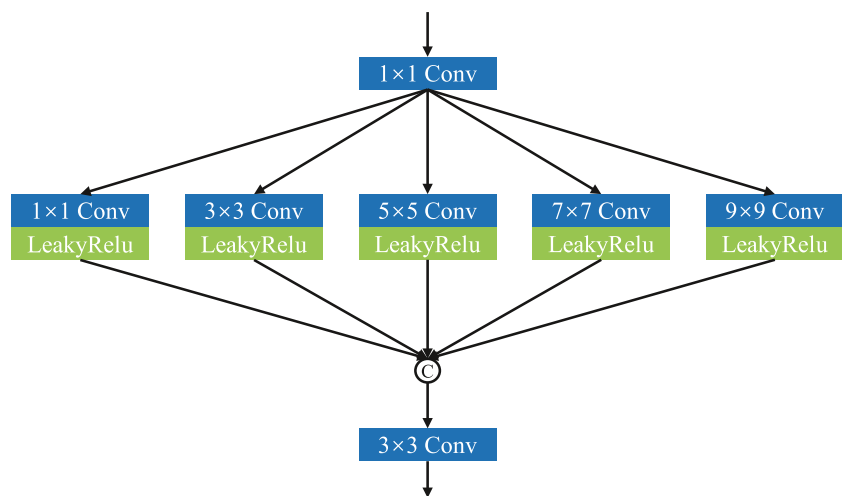
In addition, we also used structural similarity (SSIM) [29] to measure the structural differences between model output and NDCT images, and incorporated it into our mixture loss. The loss based on the SSIM (L_{SSIM}) is calculated as follows:

$$L_{SSIM} = 1 - \frac{\left(2\mu_x\mu_y + C_1\right)\left(2\mu_x\mu_y + C_1\right)}{\left(\mu_x^2 + \mu_y^2 + C_1\right)\left(\sigma_x^2 + \sigma_y^2 + C_2\right)} \quad (3)$$

where μ_x and μ_y are the average values, σ_x and σ_y are the variances, of all the pixels in the NDCT image and the reconstructed image, respectively. The final loss function consists of the above two parts and is expressed by the following formula:

$$L_{loss} = L_{MAE} + \alpha L_{SSIM} \quad (4)$$

where α is the weight of L_{SSIM}. We empirically set the value of α to be 0.5 in all the following experiments.

2.1.4 Baseline methods

We adopted four deep learning models as the baseline methods to compare with our method. They are RED-CNN [18], WGAN-VGG [21], EDCNN [19], and CTformer [24] respectively. Among them, RED-CNN is an outstanding representative of using CNN to achieve LDCT image reconstruction. The EDCNN network has achieved a good combination of edge enhancement and CNN, and achieved good reconstruction results. The WGAN-VGG network effectively introduces the GAN network into the LDCT image reconstruction task and greatly improves the clarity of LDCT images. CTformer is a network structure based on Transformer, which also performs well in denoising LDCT images. Except for our model structure, all other models are similar to the UNet structure, where downsampling is performed first and then upsampling is performed to reconstruct the final CT images.

2.1.5 Datasets

In this study, we used the dataset from 2016 NIH-AAPM-Mayo LDCT Grand Challenge [30]. We selected a total of 10 CT scan sequences of anonymous patients, each containing normal-dose CT images and corresponding simulated quarter-dose CT images. The scan thickness of these CT datasets is 3 mm, and finally there are 2378 CT images for our experiments. We used 9 sequences of CT images as our training set and the remaining one (named L506) to verify the performance of the proposed model. During the training stage, we randomly cropped 8 patches of size 64×64 from CT images of size 512×512.

2.1.6 Implementation details

For the model training, we adopted the Adam [31] algorithm to optimize our proposed network. We set the hyperparameters β_1 and β_2 to 0.9 and 0.999, respectively. For the learning rates, it was set to 1e-4. And the total number of epochs was set to 100. In the multi-scale network module, we used a large number of LeakyReLU, the slope value of it was set to 0.01. Our algorithms was built using the PyTorch 1.11.0 platform. The model was trained and tested on an NVIDIA RTX 3090Ti GPU. Different from dividing patches in training stage, we used LDCT images with a size of 512×512 as input in the testing stage.

2.1.7 Evaluation metrics

We considered the subjectivity of evaluating the reconstruction quality by naked eyes. To evaluate the performance of the model, five objective and widely used metrics, i.e., PSNR, SSIM, root mean square error (RMSE), Contrast-to-Noise Ratio (CNR) and recontruction time were used in this study. These metrics can serve as effective measures of image quality and assess the model's performance. Specifically, PSNR is an indicator used to evaluate image quality. The higher the PSNR value, the higher the reconstruction quality of the image. SSIM is used to measure the structural similarity between two images, the larger the value, the more similar the two images, and the upper limit of this value is 1. In addition, RMSE is also an important indicator to measure the similarity of two images. It reflects the error of the two images by calculating the pixel-wise differences of the reconstructed CT image and the NDCT image. Obviously, the smaller the value, the smaller the differences between the two images, reflecting the higher quality of the reconstructed CT image. CNR is a frequently employed metric within the domain of medical imaging for the purpose of image quality assessment. It quantitatively characterizes the correlation between the contrast exhibited by meaningful signals within an image and the level of noise present therein. The utilization of CT image reconstruction time serves as a means for assessing both the model's intricacy and the efficacy of its inference processes.

## 3 Experimental results and discussion

3.1 Results

Table 1 shows the average values of various evaluation metrics for each model on the Mayo dataset. The proposed Re-UNet outperforms the previous state-of-the-art CTformer by 0.56 dB in terms of PSNR. From the SSIM, the similarity between the CT image generated by Re-UNet and the

Table 1 Performance comparison of different methods on L506 dataset. The best results are highlighted in boldface

	PSNR↑	SSIM↑	RMSE↓	CNR↑	Time↓
LDCT	29.2489	0.8759	14.2416	−3.4204	−
RED-CNN	32.8209	0.9092	9.3144	−3.0951	**0.35 s**
WGAN	31.1007	0.8996	11.3596	−	−
EDCNN	32.3036	0.9034	9.9121	**−3.0711**	0.35 s
CTformer	33.0793	0.9119	9.0726	−3.0960	0.43 s
Re-UNet	**33.7460**	**0.9185**	**8.4101**	−3.0724	0.45 s

NDCT image reaches 0.9185, which is also optimal when compared with other models. In addition, the Re-UNet is 0.6625 lower than CTformer in terms of RMSE. In terms of CNR, the Re-UNet only shows 0.001 worse than the EDCNN model. For the computational complexity, although the proposed Re-UNet needs slightly more time, all the models can finish the reconstruction for single LDCT image with 1 s. Therefore, the Re-UNet shows better performance in most of cases under the four evaluation metrics. In order to conduct a more intuitive comparison of the reconstruction effects, we selected two relatively representative CT images from L506,

as shown in Figs. 6 and 7. The two images clearly show that our model achieves the best reconstruction effect. The sense of boundaries of tissues and organs after reconstruction has been significantly improved, and there is a lesion in the blue rectangle box in Fig. 7, and it becomes more clearly in the reconstructed CT image. We enlarged the area of the blue rectangular box in Fig. 7(f) to separately present in Fig. 8. The position pointed by the red arrow in the image is a lesion area, which is affected by noise in the original LDCT image, and the lesion becomes blurry, which brings some difficulty to the doctor's diagnosis. However, after reconstructing the LDCT image through our model, the noise is significantly reduced, and the size and shape of the lesion are clearly visible. The CT image reconstructed by other models also achieves noise reduction, but EDCNN and CTformer add additional noise and artifacts to the images. Only RED-CNN achieves roughly the same visual effect as our model. We further compared the superiority of the proposed model with baseline models in terms of quantitative metrics in the fixed regions. We selected four regions of interest (ROI) from the NDCT image in Fig. 7, and marked them with red rectangles. We performed PSNR, SSIM and RMSE metric calculations on the four regions

(a) LDCT   (b) RED-CNN   (c) EDCNN

(d) CTformer   (e) Re-UNet   (f) NDCT
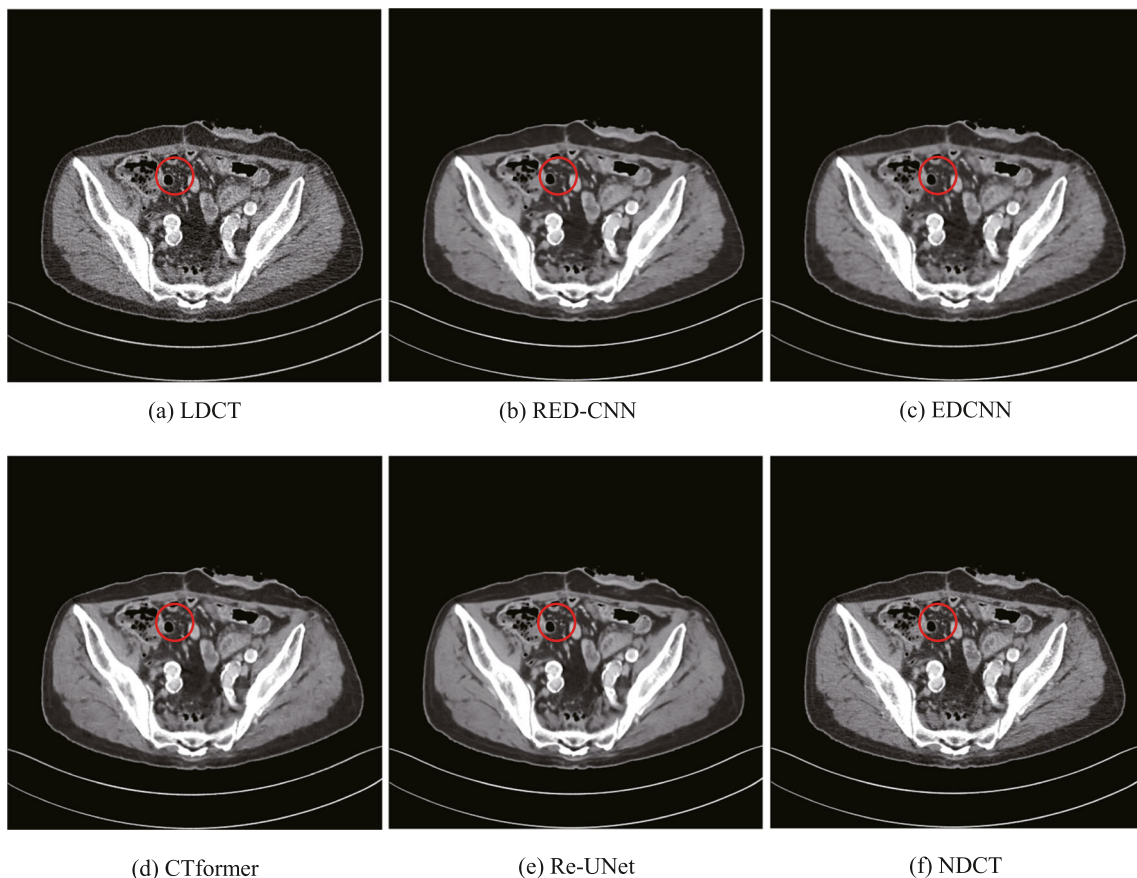
Fig. 6 Results from the different methods for comparison. The display window is [-160, 240] HU. (a) a LDCT image; (b) RED-CNN; (c) EDCNN; (d) CTformer; (e) Re-UNet. (f) a NDCT image
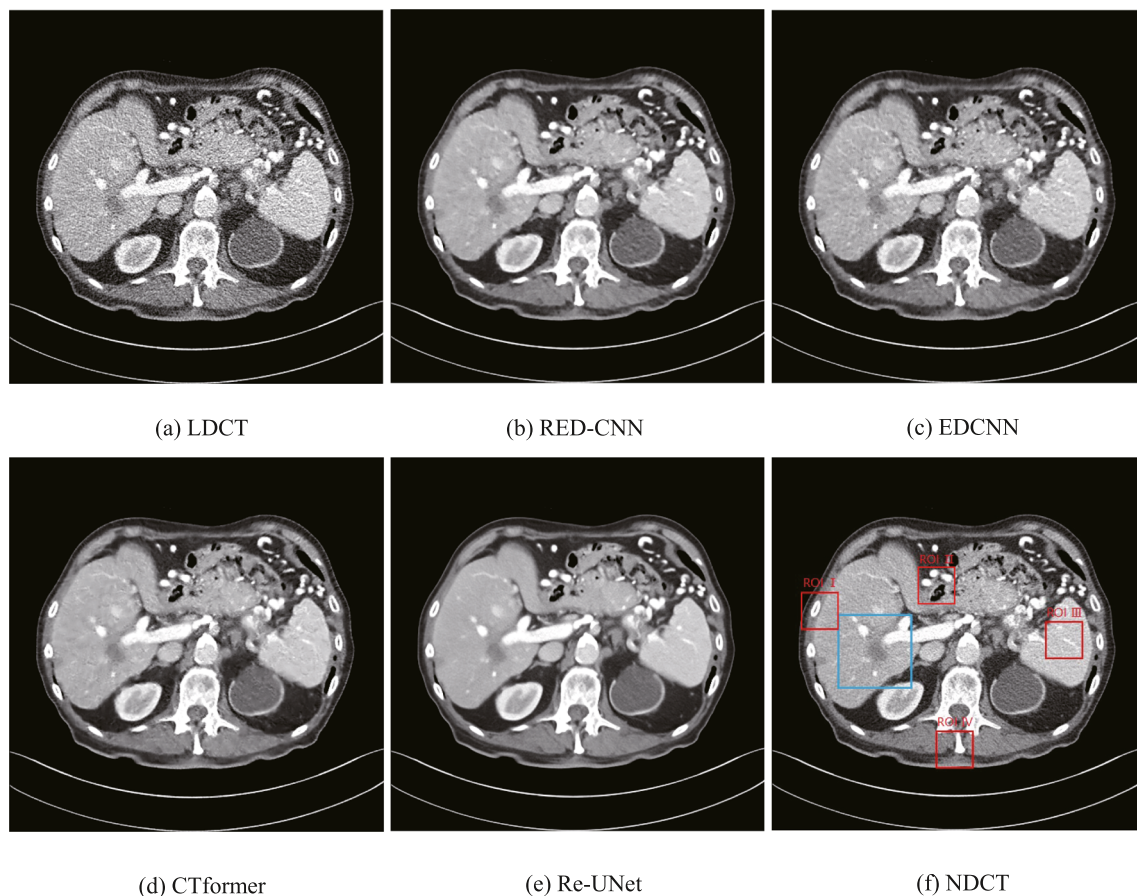
Fig. 7 Results from the different methods for comparison. The display window is [-160, 240] HU. (a) a LDCT image; (b) RED-CNN; (c) EDCNN; (d) CTformer; (e) Re-UNet. (f) a NDCT image. The red rectangle is the several defined ROIs

respectively, which are shown in Fig. 9. It can be found that Re-UNet also achieves the optimal performance in these local regions. Utilizing the delineated ROI, we computed the mean CNR, as presented in Table 1. Notably, although not attaining the optimum, this value differs by a mere 0.001 when compared to the performance of the EDCNN model. We also quantified the average temporal investment for each model during the reconstruction of 211 CT images, revealing a discrepancy of 0.1 s in relation to the optimal model.

3.2 Discussion

In this paper, we present an innovative network architecture Re-UNet for LDCT image reconstruction. Previous network architectures, such as RED-CNN and CTformer, exhibit a U-shaped network structure. In the context of these architectural configurations, an inherent procedure involves an initial employment of downsampling operations in the earlier layers to compress the sizes of intermediate feature maps, which will inevitably lead to the loss of feature information at the higher levels of the neural network, making the reconstruction task difficult. The U-D model upsampling stage

inherently expands the image dimensions, thereby enabling enhanced feature information retention, and the model will also learn how to remove excess noise information during the downsampling stage. Furthermore, Within the context of D-U architectural models, it is evident that an increase in the convolution kernel size can result in the deterioration in model performance. For the proposed Re-UNet, the U-D configuration has consistently exhibited a sustained and stable performance in the domain of noise reduction. It shows that U-D model is more robust to changes in convolution kernel size. Therefore, we believe that the U-D model will be an important structure in the field of LDCT image reconstruction.

Some limitations need to be addressed. Our experiment lacks radiologists to evaluate the reconstruction quality of LDCT images, which is important in actually diagnosing the images. Besides, only the Mayo dataset was adopted to evaluate the performance. This dataset was publicly available, to serve as the benchmark dataset to evaluate various reconstruction methods. In future, extra cohort datasets should be collected to further verify the efficiency and effectiveness in future work. In current stage, the proposed Re-UNet was not

(a) LDCT   (b) RED-CNN   (c) EDCNN

(d) CTformer   (e) Re-UNet   (f) NDCT
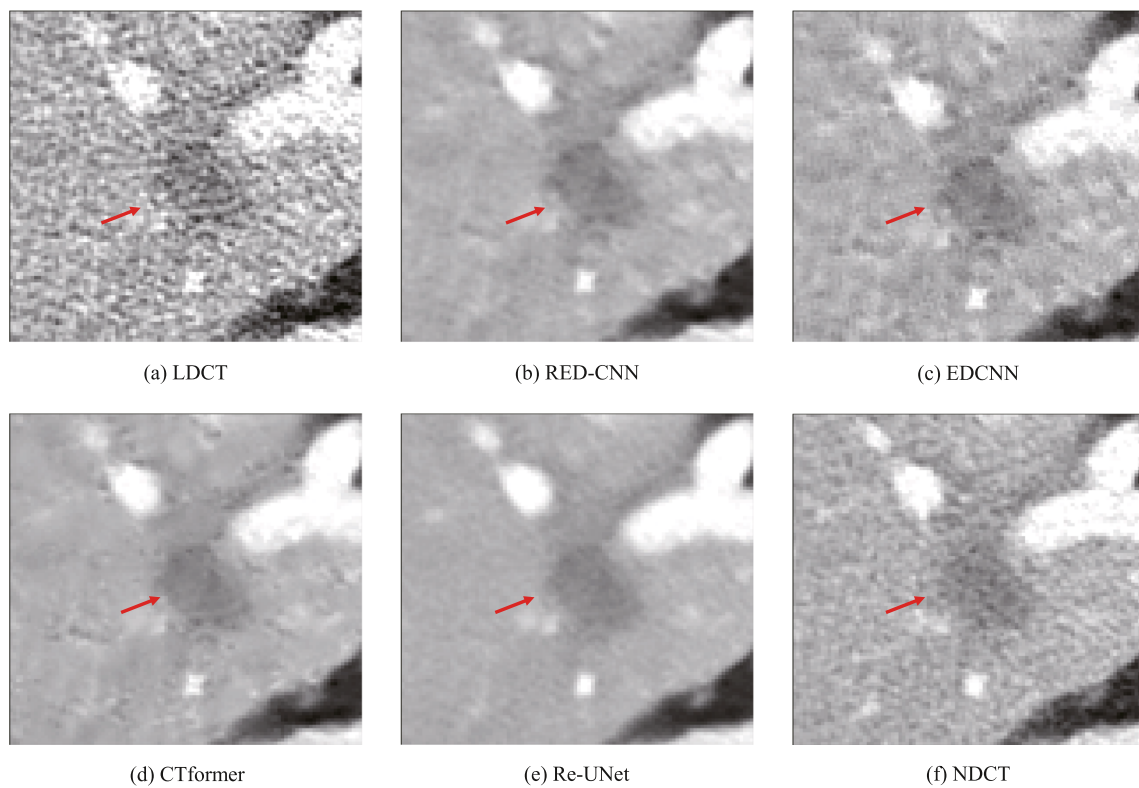
Fig. 8 The amplified ROI images of different methods outputs in the blue rectangle marked from Fig. 7
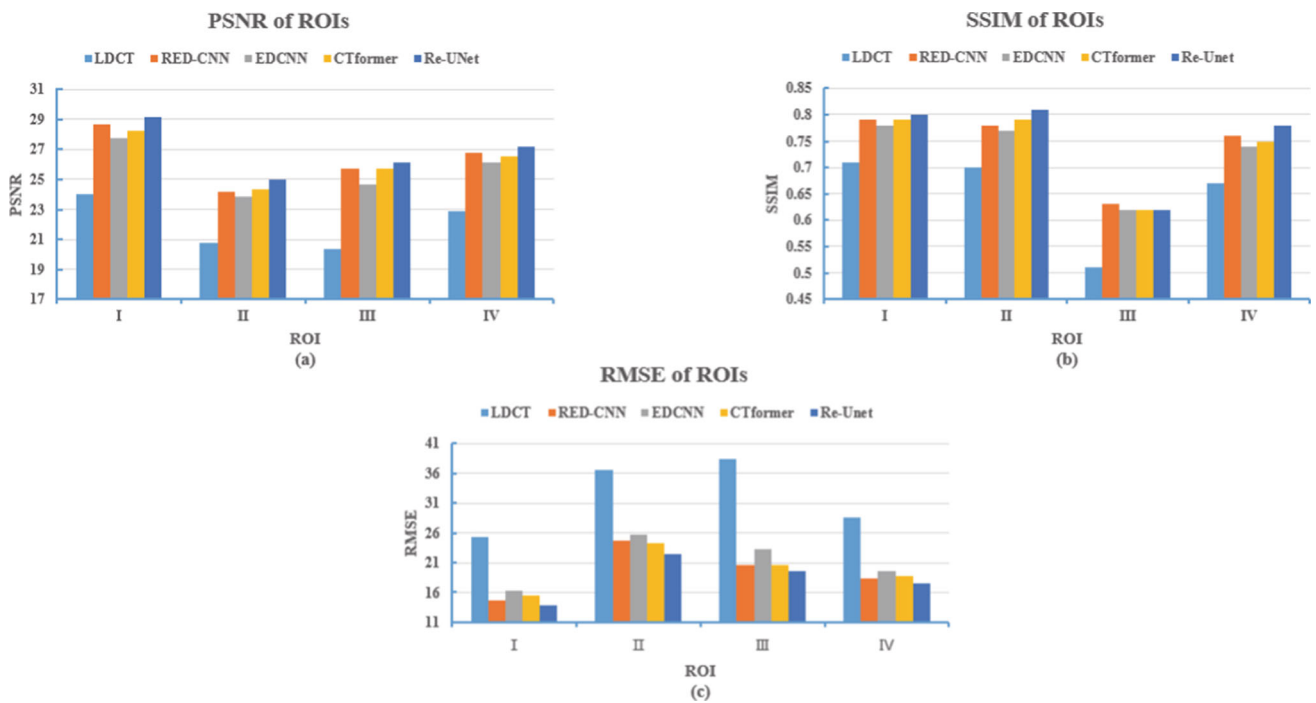
Fig. 9 The PSNR, SSIM and RMSE histogram of ROIs from Fig. 7 under different algorithms. (a) PSNR of ROIs; (b) SSIM of ROIs; (c) RMSE of ROIs

deployed within a practical CT scanning system to investigate the performance; it needs more endeavors for further optimizing the algorithm with CT equipment manufacturers and clinical practitioners to achieve that goal.

3.3 Ablation study

We designed a series of ablation experiments based on the Mayo dataset to verify the effectiveness of our network structure. To investigate the influence of upsampling and downsampling modules on the validation of LDCT image reconstruction, extra experiments with two model architectures were undertaken. Specifically, the first architecture involves a sequential application of upsampling operations followed by downsampling operations, denoted as the "U-D" architecture, and the second employs downsampling operations followed by upsampling operations, referred to as the "D-U" architecture. We want to investigate the effect of changes in the number of network layers and convolution kernel sizes on the reconstruction performance for the two model architectures. These two parameters have been considered due to their respective impacts on the feature map dimensions within the network architecture. As the U-D network structure increases in depth or when a larger convolution kernel is employed, the feature map at the highest layer will become larger. For the number of sampling layers of the network, the results are shown in Table 2, the U-D architecture consistently exhibits a superior performance over the corresponding D-U architecture. For the convolution kernel size, the results are shown in Table 3, the performance of the U-D architecture is also consistently better than that of the D-U architecture. Increasing the sizes of the convolution kernel first improves, and then leads to rapid deterioration in the performance of D-U architecture. For the U-D architecture, it is more robust to the setting of the convolution kernel sizes. These findings indicate the efficiency and robustness of the proposed U-D architecture for LDCT image reconstruction, which could be considered in future studies.

To substantiate the ameliorative impact of the additional modules we introduced for noise reduction, a series

Table 3 Experimental results of D-U model and U-D model with different convolution kernel sizes. The best results are highlighted in boldface

layers	Kernel size	model	PSNR↑	SSIM↑	RMSE↓
6	3×3	D-U	32.4054	0.9041	9.7576
		U-D	**33.5497**	**0.9165**	**8.5936**
	5×5	D-U	33.4509	0.9151	8.6860
		U-D	**33.6015**	**0.9165**	**8.5394**
	7×7	D-U	33.4337	0.9152	8.7055
		U-D	**33.6599**	**0.9182**	**8.4862**
	9×9	D-U	32.8788	0.9100	9.2621
		U-D	**33.5641**	**0.9168**	**8.5795**
	11×11	D-U	26.4269	0.8267	19.1887
		U-D	**33.6095**	**0.9180**	**8.5370**

D and U denote the downsampling and upsampling operations, respectively

of ablation experiments were meticulously conducted. The ensuing results of these investigations are meticulously documented in Table 4 for comprehensive analysis and reference. First of all, we designed two basic models. The first model is downsampling and then upsampling, as shown in Fig. 1(a), termed as UNet. The second model is to exchange the order of sampling operation in UNet model. We first upsample the CT image and then downsample it, termed as R-UNet. In these two basic models, we used convolution for downsampling, and deconvolution for upsampling. In the UNet model, we used a total of five convolutions for downsampling. The size of the CT images was reduced from 512×512 to 488×488. Then, through five consecutive transposed convolutions to restore the size of the image to 512×512. When the hyperparameters of the two models are the same, the basic R-UNet model demonstrates superior performance on the testset in comparison to the basic UNet model. From Table 4, we can see that the test result of the R-UNet is 0.31 dB higher than that of UNet on PSNR. However, there is not much difference between the two in the number of parameters, we only keep two decimal places. It can be seen that the parameter amount of the R-UNet model is 1.03M, but the performance

Table 2 Experimental results of D-U model and U-D model with different sampling layers. The best results are highlighted in boldface

Kernel size	layers	model	PSNR↑	SSIM↑	RMSE↓
5×5	5	D-U	33.2882	0.9118	8.8359
		U-D	**33.3511**	**0.9137**	**8.7809**
	6	D-U	33.4509	0.9151	8.6860
		U-D	**33.6015**	**0.9165**	**8.5394**

D and U denote the downsampling and upsampling operations, respectively

Table 4 The results of ablation experiments on L506 dataset. The best results are highlighted in boldface

	Reverse	Multi	Edge	PSNR↑	Params
UNet			×	33.2882	**1.03**
R-UNet	✓	×	×	33.6015	1.03
MR-UNet	✓	✓	×	33.6910	4.14
Re-UNet	✓	✓	✓	**33.7460**	4.15

of the model has been greatly improved. Next, we tested the effectiveness of our proposed multi-scale module through experiments. Compared with R-UNet, we just replaced the original convolution block with a multi-scale module, the model we named MR-UNet. The PSNR value of the MR-UNet model is 33.6910 dB, which has some improvements compared to R-UNet. Finally, after we added the edge convolution, the test performance of the model on the L506 dataset was further improved. The PSNR of the final model has been improved to 33.7460 dB.

4 Conclusion

In summary, we proposed a novel Re-UNet model architecture for the LDCT images reconstruction. Different from the well-known U-Net architecture, the upsampling operations were first used, followed by the downsampling operations. Besides, the multi-scale feature extraction block was designed to reconstruct the different organs. Extensive experiments indicate that the Re-UNet outperforms the compared methods. This study may provide a new avenue for CT image reconstruction with Re-UNet, and demonstrates a great potential for designing the DL models with reverse U-shape architecture for image processing.

References

1. Seeram E (2015) Computed tomography-e-book: physical principles, clinical applications, and quality control. Elsevier Health Sciences
2. Brenner DJ, Hall EJ (2007) Computed tomography-an increasing source of radiation exposure. N Engl J Med 357(22):2277–2284
3. Shrimpton PC, Hillier MC, Lewis MA, Dunn M (2005) Doses from computed tomography (CT) examinations in the UK-2003 review, vol 67. NRPB Chilton
4. Gartenschläger M, Schweden F, Gast K, Westermeier T, Kauczor HU, Von Zitzewitz H, Thelen M (1998) Pulmonary nodules: detection with low-dose vs conventional-dose spiral CT. Eur Radiol 8:609–614
5. Kalra MK, Maher MM, Toth TL, Hamberg LM, Blake MA, Shepard JA, Saini S (2004) Strategies for CT radiation dose optimization. Radiology 230(3):619–628
6. Wang J, Li T, Liang Z, Xing L (2008) Dose reduction for kilovotage cone-beam computed tomography in radiation therapy. Phys Med Biol 53(11):2897
7. Bian J, Siewerdsen JH, Han X, Sidky EY, Prince JL, Pelizzari CA, Pan X (2010) Evaluation of sparse-view reconstruction from flat-panel-detector cone-beam CT. Phys Med Biol 55(22):6575

8. Xia W, Shan H, Wang G, Zhang Y (2023) Physics-/model-based and data-driven methods for low-dose computed tomography: A survey. IEEE Signal Proc Mag 40(2):89–100
9. Sahiner B, Yagle AE (1993) Image reconstruction from projections under wavelet constraints. IEEE Trans Signal Process 41(12):3579–3584
10. Zhang Y, Zhang J, Lu H (2010) Statistical sinogram smoothing for low-dose CT with segmentation-based adaptive filtering. IEEE Trans Nucl Sci 57(5):2587–2598
11. Sukovic P, Clinthorne NH (2000) Penalized weighted least-squares image reconstruction for dual energy X-ray transmission tomography. IEEE Trans Med Imaging 19(11):1075–1081
12. Sidky EY, Pan X (2008) Image reconstruction in circular cone-beam computed tomography by constrained, total-variation minimization. Phys Med Biol 53(17):4777
13. Browne J, De Pierro AB (1996) A row-action alternative to the EM algorithm for maximizing likelihood in emission tomography. IEEE Trans Med Imaging 15(5):687–699
14. Fessler JA, Hero AO (1994) Space-alternating generalized expectation-maximization algorithm. IEEE Trans Signal Process 42(10):2664–2677
15. Chen H, Zhang Y, Zhang W, Liao P, Li K, Zhou J, Wang G (2017) Low-dose ct via convolutional neural network. Biomed Opt Express 8(2):679–694
16. Gao X, Zhang L, Mou X (2018) Single image super-resolution using dual-branch convolutional neural network. IEEE Access 7:15767–15778
17. He K, Zhang X, Ren S, Sun J (2016) Deep residual learning for image recognition. In: Proceedings of the IEEE conference on computer vision and pattern recognition, pp 770–778
18. Chen H, Zhang Y, Kalra MK, Lin F, Chen Y, Liao P, Zhou J, Wang G (2017) Low-dose CT with a residual encoder-decoder convolutional neural network. IEEE Trans Med Imaging 36(12):2524–2535
19. Liang T, Jin Y, Li Y, Wang T (2020) Edcnn: Edge enhancement-based densely connected network with compound loss for low-dose ct denoising. In: 2020 15th IEEE International conference on signal processing (ICSP), vol 1, pp 193–198. IEEE
20. Li S, Li Q, Li R, Wu W, Zhao J, Qiang Y, Tian Y (2022) An adaptive self-guided wavelet convolutional neural network with compound loss for low-dose CT denoising. Biomed Signal Process Control 75:103543
21. Yang Q, Yan P, Zhang Y, Yu H, Shi Y, Mou X, Kalra MK, Zhang Y, Sun L, Wang G (2018) Low-dose CT image denoising using a generative adversarial network with Wasserstein distance and perceptual loss. IEEE Trans Med Imaging 37(6):1348–1357
22. Johnson J, Alahi A, Fei-Fei L (2016) Perceptual losses for real-time style transfer and super-resolution. In:Computer vision–ECCV 2016: 14th european conference, Amsterdam, The Netherlands, October 11-14, 2016, Proceedings, Part II 14, pp 694–711. Springer
23. Zhang Y, Hu D, Zhao Q, Quan G, Liu J, Liu Q, Zhang Y, Coatrieux G, Chen Y, Yu H (2021) Clear: comprehensive learning enabled adversarial reconstruction for subtle structure enhanced low-dose CT imaging. IEEE Trans Med Imaging 40(11):3089–3101
24. Wang D, Fan F, Wu Z, Liu R, Wang F, Yu H (2022) Ctformer: Convolution-free token2token dilated vision transformer for low-dose ct denoising. arXiv:2202.13517
25. Liu J, Jiang H, Ning F, Li M (2022) Pang W Dfsne-net: Deviant feature sensitive noise estimate network for low-dose CT denoising. Comput Biol Med 149:106061
26. Bera S, Biswas PK (2023) Self supervised low dose computed tomography image denoising using invertible network exploiting inter slice congruence. In: Proceedings of the IEEE/CVF winter conference on applications of computer vision, pp 5614–5623
27. Lu Z, Xia W, Huang Y, Hou M, Chen H, Zhou J, Shan H, Zhang Y (2022) M 3 nas: Multi-scale and multi-level memory-efficient

neural architecture search for low-dose ct denoising. IEEE Trans Med Imaging 42(3):850–863

28. Ronneberger O, Fischer P, Brox T (2015) U-net: Convolutional networks for biomedical image segmentation. In: Medical image computing and computer-assisted intervention–MICCAI 2015: 18th international conference, Munich, Germany, October 5-9, 2015, Proceedings, Part III 18, pp 234–241. Springer

29. Wang Z, Bovik AC, Sheikh HR, Simoncelli EP (2004) Image quality assessment: from error visibility to structural similarity. IEEE Trans Image Process 13(4):600–612

30. McCollough CH, Bartley AC, Carter RE, Chen B, Drees TA, Edwards P, Holmes DR III, Huang AE, Khan F, Leng S et al (2017) Low-dose CT for the detection and classification of metastatic liver lesions: results of the 2016 low dose CT grand challenge. Med Phys 44(10):e339–e352

31. Kingma DP, Ba J (2014) Adam: A method for stochastic optimization. arXiv:1412.6980

Wei Qiu received the B.S. degree from Sichuan Tourism University, Sichuan, China, in 2020. She is currently studying for a master's degree at Southwest University of Science and Technology in Sichuan, China. Her current research focus is on medical images and image registration.

Yiqian Luo received the B.S. degree from the Southwest University of Science and Technology, Sichuan, China, in 2021. He is currently pursuing the master degree in the Southwest University of Science and Technology, Sichuan, China. His current research focus on the medical image and image classification.

Lianjin Xiong received the B.S. degree from Sichuan Tourism University in 2019. He is currently studying for a master's degree in the School of Computer and Technology, Southwest University of Science and Technology. His research interests include medical image processing and medical imaging technology.

Yishi Li MD, Associate Chief Physician. He is an assistant to the director of Respiratory and Critical Care Medicine at The First Affiliated Hospital of Chongqing Medical University. He serves as Vice Chairman of the Respiratory Endoscopic Interventional Committee of the Chinese Society for Tuberculosis. His research focuses on interventional respiratory diseases, medical artificial intelligence, etc.

Ning Li received the B.S. degree from the North University of China, Taiyuan, China, in 2020. He is currently pursuing the master degree in the Southwest University of Science and Technology, Sichuan, China. His current research focus on the medical image and image segmentation.

Yangsong Zhang received the Ph.D. degree in signal and information processing from the School of Life Science and Technology, University of Electronic Science and Technology of China, in 2013. He is currently a Professor with the School of Computer Science and Technology, Southwest University of Science and Technology, China. His research interests include brain-computer interface, deep learning, machine learning, medical imaging processing, etc.