Low-dose CT Image Denoising Using Classification Densely Connected Residual Network

Jun Ming, **Benshun Yi***, **Yungang Zhang and Huixin Li**
School of Electronic Information, Wuhan University,
Wuhan 430072,China
[e-mail: yibs@whu.edu.cn]
*Corresponding author: Benshun Yi

Abstract

Considering that high-dose X-ray radiation during CT scans may bring potential risks to patients, in the medical imaging industry there has been increasing emphasis on low-dose CT. Due to complex statistical characteristics of noise found in low-dose CT images, many traditional methods are difficult to preserve structural details effectively while suppressing noise and artifacts. Inspired by the deep learning techniques, we propose a densely connected residual network (DCRN) for low-dose CT image noise cancelation, which combines the ideas of dense connection with residual learning. On one hand, dense connection maximizes information flow between layers in the network, which is beneficial to maintain structural details when denoising images. On the other hand, residual learning paired with batch normalization would allow for decreased training speed and better noise reduction performance in images. The experiments are performed on the 100 CT images selected from a public medical dataset—TCIA(The Cancer Imaging Archive). Compared with the other three competitive denoising algorithms, both subjective visual effect and objective evaluation indexes which include PSNR, RMSE, MAE and SSIM show that the proposed network can improve LDCT images quality more effectively while maintaining a low computational cost. In the objective evaluation indexes, the highest PSNR 33.67, RMSE 5.659, MAE 1.965 and SSIM 0.9434 are achieved by the proposed method. Especially for RMSE, compare with the best performing algorithm in the comparison algorithms, the proposed network increases it by 7 percentage points.

1. Introduction

With the development of CT technology, use of CT imaging in medical diagnosis has expanded. Considering the potential risks to patients from high-dose radiation during CT scans, reducing radiation dose while maintaining good image quality for medical diagnosis has become an important research field in medical imaging. In 1990, Naidich et al. [1] conceptualized low-dose CT (LDCT); they lowered the radiation dose by reducing X-ray tube current with keeping other scanning parameters unchanged. When the tube current is reduced, the number of photons received by the detector also decreases, causing the projection data to be contaminated by noise. Thus, the CT image reconstructed from such contaminated projection data will have high noise and streaking artifacts [2], resulting in adverse effects to medical diagnosis. To improve upon the issue of low quality low-dose CT images, many methods are proposed. These methods are encompassed by three overarching categories: sinogram filtering, image reconstruction and image domain denoising.

The sinogram filtering performs on the projection data before image reconstruction. Typical algorithms include generalized multi-dimensional adaptive filtering [3], penalized weighted least-squares (PWLS) algorithm [4], bilateral filtering [5], structure adaptive filtering [6] and so on. The advantage of sinogram filtering is that they can make full use of the statistical characteristics of noise in the sinogram domain [7]. However, when edges in the sinogram are not well preserved, sinogram filtering is impaired by spatial resolution loss. This causes new noise and artifacts to be introduced after image reconstruction. The filtered back projection (FBP) algorithm [8] is the most common CT image reconstruction method, which has the advantages of high resolution and fast imaging speed. But at the same time, the FBP algorithm has a high requirement for the completeness of projection data. When the amount of projection data is insufficient, the quality of the reconstructed image will be significantly reduced. Recently, in order to ameliorate the reconstructed image quality, researchers have recommended iterative reconstruction algorithms, such as adaptive statistical iterative reconstruction (ASIR) and model-based iterative reconstruction (MBIR). However, despite the improved image quality, these processes are very time-consuming.

Sinogram filtering and image reconstruction both require access to projection data, but the projection data is difficult to be accessed by normal users since it is the intermediate result of CT scanner. In contrast, image domain denoising algorithms can be performed on the reconstructed CT images directly, which do not rely on the projection data. To take advantage of similar features found within a large neighborhood in an image, an adaptive nonlocal means algorithm was suggested [9]. Proposed by Kang et al. [10], an adaptive block-matching 3D algorithm achieved high efficiency in low-dose CT image noise reduction tasks. Chen et al. [11] proposed an artifact suppression dictionary learning (ASDL) algorithm, which integrates the direction and scale information of artifacts into dictionary training, and then eliminates artifacts by sparse representation. Furthermore, Chen et al. [12] proposed a discriminative feature representation (DFR) algorithm, which uses the feature dictionary to decompose high-dose CT image features from low-dose CT images for noise-free image estimation. Although the statistical characteristics of noise in projection data are well known as Poisson distribution, they become complex in reconstructed image. In [13], the non-local means (NLM) was adapted for CT image denoising. Khan et al. proposed a 2-D Adaptive Trimmed Mean Autoregressive (ATMAR) model to denoise of medical images corrupted with Poisson noise [14]. Meanwhile, a Weighted Gradient Filter for Poisson noise in medical images is proposed in [15]. Through this image domain noise reduction process, the image quality is significantly improved, but smoothing and/or residual errors often occur in the processed image. These

problems are difficult to solve considering the non-uniform distribution of CT image noise. Thus, the traditional methods are difficult to achieve an optimal balance between denoising performance and detail preservation, in other words, it is easy to lose details in image domain while suppressing noise and artifacts. In the recent years with rapid deep learning breakthroughs [16], convolutional neural network (CNN) has shown great advantages and has been applied to many computer visual tasks, ranging from image noise reduction, blur reduction and super-resolution to segmentation, detection and recognition. With strong capabilities of feature learning and mapping, CNN-based methods typically demonstrate promising advantages compared to traditional methods in the process of removing complex noise found in low-dose CT images. Chen et al. proposed a simple CNN model for low-dose CT image denoising, which performs better than the traditional methods in both visual quality and quantitative indexes [17]. To fully demonstrate the superiority of the convolutional neural network when compared with traditional methods, a more detailed experiment was then carried out [18]. Consequently, they further considered a residual encoder-decoder convolutional neural network (RED-CNN), which allowed them to attain avant-garde noise reduction performance [19]. Kang et al. Determined a deep convolutional neural network to adequately subdue CT-specific noise by applying the deep CNN to low-dose CT image wavelet transform coefficients [20]. Accordingly, they imparted the wavelet residual network, which seamlessly integrates the extensive reach of deep learning with the heightened performance standard provided by framelet-based noise reduction algorithms to ascertain a striking framelet-based denoising algorithm [21]. Yang et al. proposed a two-dimensional and three-dimensional deep residual network to remove noise and artifacts while effectively preserving details [22]. Gholizadeh-Ansari et al. introduced dilated convolution into the proposed residual network, which expands the receptive field of the network and enables the network to obtain better denoising results with fewer convolutional layers [23]. Although some studies involved construction of deeper network, most image denoising tasks are considered as "low-level" tasks since limited layers without no intention to extract features. This is in clear contract to high-level task such as recognition or detection, in which deep CNN layers and other operations are widely used to capture deep features of images [16].

Inspired by ResNet [24] and DenseNet [25], we recommend a densely connected residual network (DCRN) for low-dose CT image noise cancelation, which is carefully designed to achieve harmony between denoising performance and computational setbacks. The rest of this paper is organized as follows. The structural characteristics of our proposed DCRN are specified in section 2. In section 3, various qualitative and quantitative experiments performed are provided to support the proposed DCRN model, along with an in-depth discussion and analysis on the impacts of the structural features on the denoising performance. Finally, a conclusion is drawn in Section 4.

2. Method

2.1 Densely connected residual network

When CNN goes deeper, the information about the input and gradient will vanish gradually during the process of propagation, which makes it more difficult to train deep CNN model. Since residual mapping can be easier understood compared to the original unreferenced mapping, residual learning [24] of CNN would provide for an effective solution to the issue of performance degradation from increased network depth. Consequently, very deep CNN are not only easily trained, but also highly accurate. Although residual learning was firstly applied

to image classification and object detection, it has been successfully extended to various computer vison tasks including image denoising. In addition to residual learning, batch normalization (BN) [26] has also been widely used to improve training efficiency; it applies a normalization step in addition to a scale and shift step prior to nonlinear activation layer in order to account for the internal covariate shift. Using BN layer in CNN can bring several benefits, such as faster convergence, improved results, and reduced sensitivity to initialization. As for how batch normalization interacts with residual learning, on one hand, batch normalization alleviates internal covariate shift within the residual block, thus improving the performance of residual learning. On the other hand, residual learning makes the inputs of the hidden layers closer to the Gaussian distribution, which reduces the correlation between them and helps batch normalization to accurately adjust to internal covariate shift. Therefore, the integration of residual learning and batch normalization will allow for an efficient training process and a better denoising performance [27]. Our proposed network adopts this strategy.

Different from ResNet [24] that bypasses signal from one layer to the next via summation, DenseNet [25] maximizes information exchange between layers in the network through a simple connectivity model- layers of the same feature map size to be connected directly to each other. Each layer receives inputs from all previous layers and passes on its own feature maps to all subsequent layers. In image denoising, due to the information loss during the forward propagation process of the network, many CNN-based methods without skip connections are difficult to preserve details effectively in denoised image. Now with this dense connectivity pattern, later layers can make full use of the feature maps of all early layers, thus significantly reducing information loss. Our proposed network has also drawn on the experience of this dense connectivity pattern.

In dense blocks, the input of the subsequent convolution modules consists of the input and output feature maps of the previous module, so that the output feature maps of each convolution module can be directly utilized by all subsequent convolution modules. The advantages of this dense connectivity pattern are as follows: on the one hand, feature map dense connection establishes a direct connection path between the front and rear layers of the network, and the gradient can directly reach any layer in front of the network in the process of back propagation, thus preventing gradient disappearance during network training. On the other hand, contrary to traditional CNN, which only uses the output feature maps of the previous layer, this dense connectivity pattern enables any layer in the network to maximize usage of the output feature maps of all the preceding layers, which encourages the reuse of the feature maps and improves the network's feature learning ability. As the convolution module continuously connects the input and output feature maps, the amount of feature maps inside the network will increase sharply in accordance to the increasing number of convolution modules, augmenting computation significantly. Therefore, the network implemented in this paper is appropriately simplified while referring to this dense connectivity pattern. As shown in **Fig. 1**, the output feature maps of the all convolution modules are directly connected to the rear of the Concat layer, and then the Bottleneck layer compresses the feature maps on the dimension, which not only ensures that the output feature maps of the network of each convolution module figure make full use of it, but also improves the computation efficiency of the network. Based on these theoretical foundations, we proposed a densely connected residual network (DCRN) for low-dose CT image noise, which combines the ideas of dense connection with residual learning.
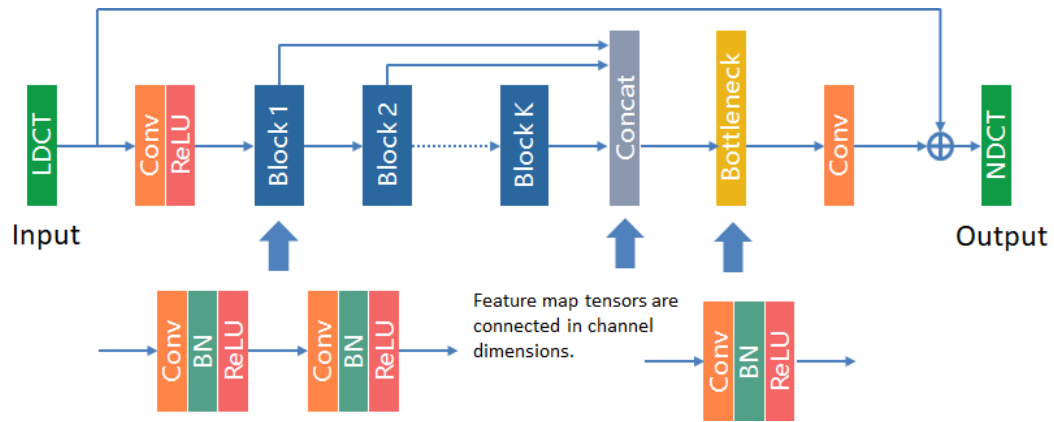
Fig. 1. The architecture of our proposed network.

Fig. 1 demonstrates the overall architecture of our proposed network. The input of DCRN is LDCT image, the output is the corresponding NDCT image; Conv, BN and ReLU represent convolutional layer, batch normalization layer and rectified linear unit [28] respectively; Concat denotes the concatenation operation, which concatenates feature maps from different layers in the channel dimension as proposed in DenseNet; Bottleneck denotes a dimensional reduction operation consisting of a Conv, BN and ReLU, where the convolution kernel size is set to 1×1; Block denotes the combination of two Conv-BN-ReLU subblocks. The algorithm flow chart is shown in **Fig. 2**.
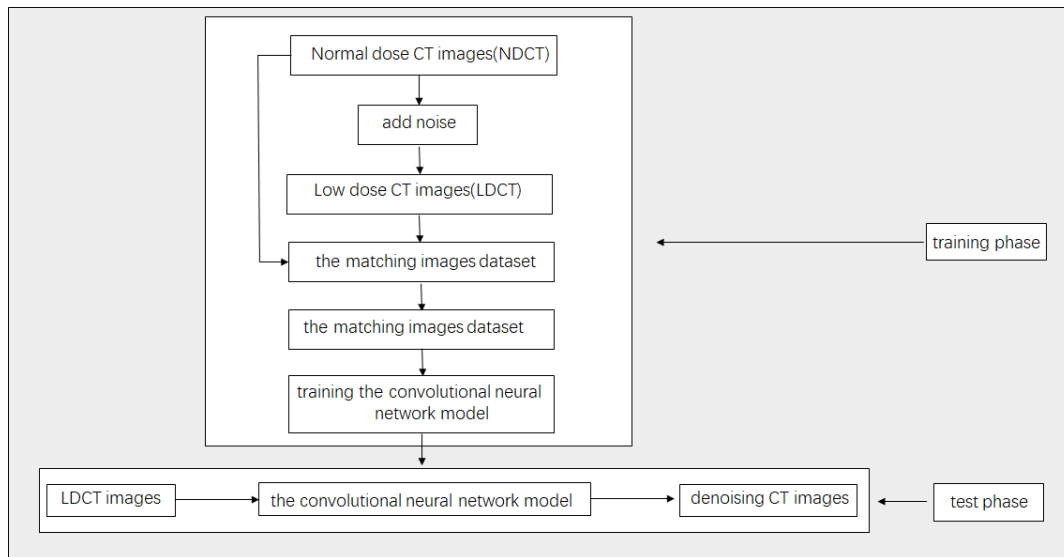

Fig. 2. The algorithm flow chart.

The workflow of our proposed DCRN can be divided into three stages: 1) After the LDCT image is input into the network, a Conv layer and a ReLU layer are first applied to obtain a set of initial feature maps. 2) Several cascaded Blocks are used to gradually extract features of higher-level abstraction, and then the output feature maps of each Block are concatenated in the Concat layer. The number of the feature maps are reduced by the Bottleneck layer which have augmented as a result of the concatenation. Note that the dense connectivity pattern applied here is a little different from that proposed in DenseNet [25], in other words, we have

modified the dense connectivity pattern to reduce its computational cost while maintaining the full utilization of the output feature maps of each Block in the network. 3) The last Conv layer aggregates the feature maps to output a residual map, the estimated NDCT image is obtained by adding the residual map to the LDCT image.

2.2 Implement Details

In convolutional kernel design of convolutional neural network, it is conventional and advantageous to stack several small convolutional kernels instead of using a single large convolutional kernel. Particularly, when the receptive fields are the same, the number of weights can be saved by stacking several small convolution kernels. Additionally, the use of multiple small convolution kernels increases nonlinearity to the network.

The convolution kernel size is set to 3×3 for all convolutional layers except for the first one which is set to 5×5, and for every convolutional layer, we pad zeros to maintain the input and output feature map size. K and F, important hyperparameters that affect the network performance, represent the number of Blocks and the number of feature maps in each convolutional layer respectively. In this paper, K=5 and F=64 by default, and the significance of these values on network performance will be discussed in Section 3.3.

3. Experiments and evaluations

All the experiments were performed on the computer platform with the following configurations: CPU is Intel Core i7-6850K, GPU is Nvidia GeForce GTX 1080 Ti. We used Caffe [29] to train CNN models and then evaluated them with MATLAB R2016b.

3.1 Training

Training convolution neural network needs many of the corresponding input and output images to each other, but in practice to obtain real and mutual corresponding LDCT and NDCT image is very difficult. In the field of low-dose CT image denoising, the LDCT image is mainly simulated from NDCT image by the algorithm.

TCIA (The Cancer Imaging Archive) [30] is a public database containing common tumor medical images and corresponding clinical information, from which we selected 200 NDCT images various human body parts as training data, and the size of the images is 512×512. The simulation algorithm proposed by Zeng et al. [31] is used to add Poisson noise to the NDCT image to generate the corresponding LDCT image. Let S be the simulated sinogram before log-transform, the corresponding contaminated sinogram Sn was obtained by the following formula:

$$S_n = \text{Poisson}(b \cdot e^{-S} + r) \tag{1}$$

where b is the blank scan factor that controls the simulated noise level, b is read-out noise, Possion() denotes the process of adding Poisson noise. In our simulation, b is set to 10^6, the LDCT image is reconstructed from S_n by FBP algorithm [8]. For example, **Fig. 3** is an NDCT picture, and **Fig. 4** is a picture of the corresponding LDCT, which is a picture after adding Poisson noise in **Fig. 3**.

Fig. 3. A NDCT image

Fig. 4. The image of corresponding LDCT

In order to train the network more efficiently, our proposed DCRN was trained based on image patches with a size of 55×55 and a sliding interval of 8 pixels. After extracting the patches, data augmentation including rotation and flipping was applied to expand the training set, which is beneficial to suppress overfitting and improve the robustness of the network.

In the training phase, the hyperparameters were set according to the following: starting base learning rate was 10^{-3} with a decreasing rate of 10^{-5} per step, convolution kernels initialized with random Gaussian distribution of zero mean and standard deviation 0.01, all bias terms set to zero, batch size of one training iteration set to 64, Mean Squared Error (MSE) applied as the loss function and optimized by Adam[32].

3.2 Evaluation

We randomly selected 100 CT images from TCIA to test our proposed DCRN, these images were not included in the training set. Moreover, BM3D [33], D-ResNet [23] and RED-CNN [19] were selected to compare with DCRN. BM3D uses similarity between image blocks for joint filtering; D-ResNet contains 7 convolution layers, and the middle 5 convolution layers use empty convolution to increase the receptive field; RED-CNN adopts the structure of Residual Codec and comprises of 5 symmetric convolutional layers and 5 deconvolution layers.

3.2.1 Subjective Visual Effects

Fig. 5 exhibits the qualitative results of each type of image. As can be seen, the LDCT image not only has a high noise level, but also is accompanied by radial artifacts. Although BM3D effectively removes noise and makes the image visually smoother, it does not suppress artifacts well, and there are still a lot of artifacts remained in the image. D-ResNet removes most of the artifacts, but it does not effectively remove the noise, and the denoised image is visually blurry. In contrast, RED-CNN and our proposed DCRN are better in noise removal and artifact suppression. Similar conclusion can be drawn from **Fig. 6**. which demonstrates the qualitative results of another test image.
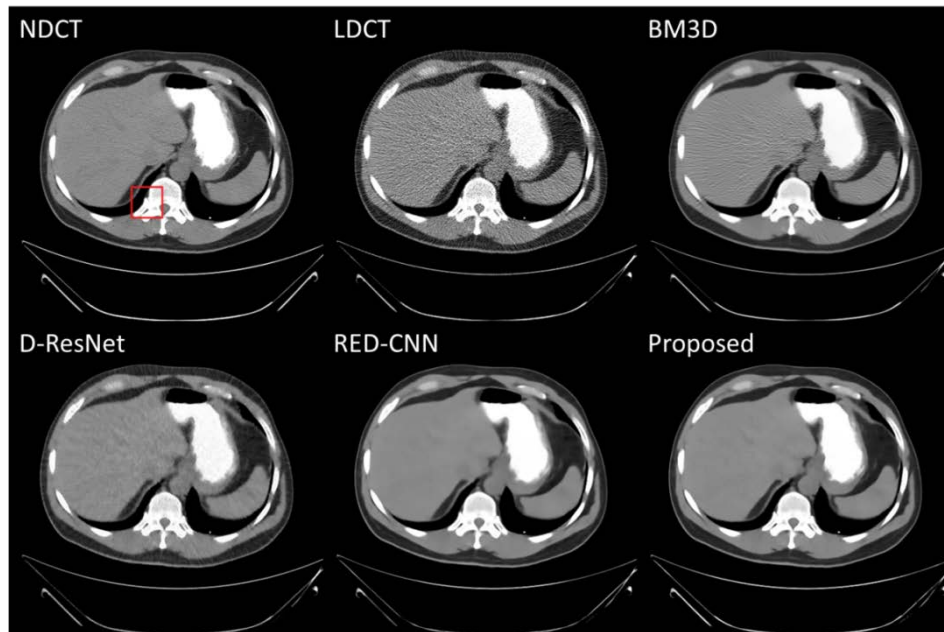
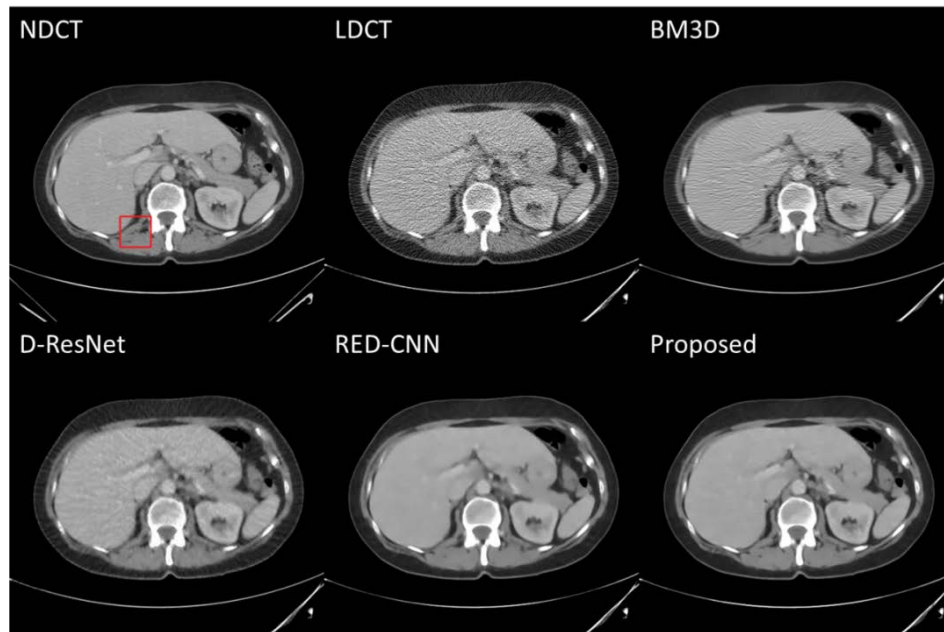**Fig. 5.** Qualitative results of the test image A

Fig. 6. Qualitative results of the test image B.

Although the denoising results of our proposed DCRN is very similar to those of RED-CNN, it can still be found that DCRN is better than RED-CNN in preserving image details **Fig. 7** and **Fig. 8** magnify the area marked by the red box found in **Fig. 5** and **Fig. 6**, respectively. In **Fig. 7**, BM3D and D-ResNet cannot effectively preserve the small gap between the bone indicated by the red arrow, while RED-CNN and DCRN preserve it completely. For the small bone groove indicated by the blue arrow, it becomes much shorter after denoising by BM3D, D-ResNet and RED-CNN, only our proposed DCRN preserve it well. In **Fig. 8**, the red and

blue arrows point to the gap between the tissues. In the images denoised by BM3D, D-ResNet and RED-CNN, the gaps have different degrees of blurring, in contrast, the gaps in the denoised image of DCRN are clearly visible.
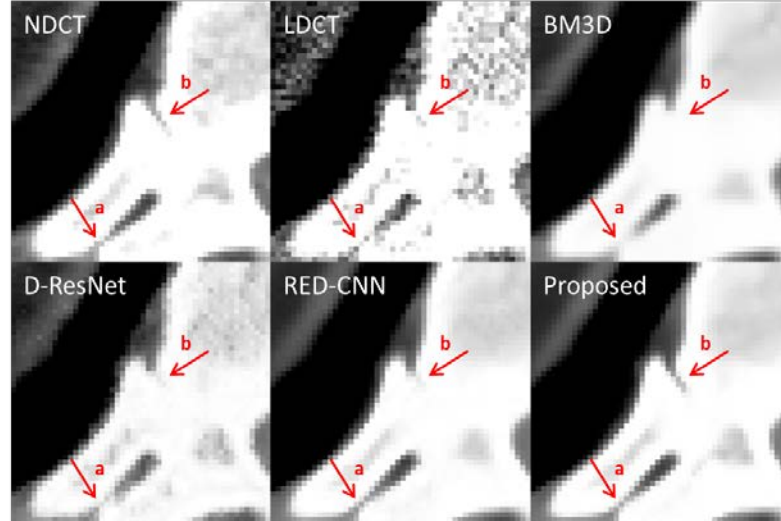

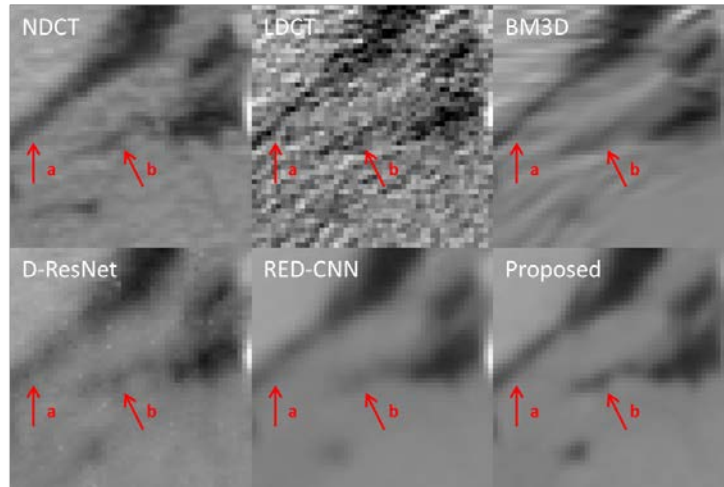**Fig. 7.** magnification of the region marked enclosed by the red box in **Fig. 5**.

Fig. 8. magnification of the region enclosed by the red box in **Fig.6**.

3.2.2 Objective Evaluation Index

Common objective evaluation indexes of image quality include Peak Signal to Noise Ratio (PSNR), Root Mean Square Error (RMSE), Mean Absolute Error (MAE) and Structural Similarity (SSIM) [34]. The calculation of PSNR and RMSE is based on mean square error, so it is sensitive to large pixel value error. The calculation of them are as follows:

$$MSE = \frac{1}{mn}\sum_{i=1}^{m}\sum_{j=1}^{n}\left\|I(i,j) - I_{ref}(i,j)\right\|^2$$

(2)

$$PSNR = 10 \log_{10} \frac{(2^b - 1)^2}{MSE} \tag{3}$$

$$RMSE = \sqrt{MSE} \tag{4}$$

Where, I and I_{ref} are the image to be evaluated and the reference image of size $m \times n$, respectively, and b is the number of bits occupied by the pixel point storage, for example, b=8 for the 8-bit gray image. The larger the value of PSNR and the smaller the value of RMSE, the better the image quality. MAE calculates the mean of the absolute error. The smaller the value, the better. The formula is as follows:

$$MAE = \frac{1}{mn} \sum_{i=1}^{m} \sum_{j=1}^{n} \left| I(i,j) - I_{ref}(i,j) \right| \tag{5}$$

PSNR, RMSE and MAE evaluate the image quality based on the pixel value error between the test image and the reference image, without considering visual characteristics of the human eye, so the evaluation results may be inconsistent with people's subjective feelings. In order to make the evaluation of the image quality more in line with people's subjective feelings, SSIM evaluates the image quality based on brightness, contrast, and structure. A condensed form of the SSIM is given by:

$$SSIM = \frac{(2\mu_x\mu_y + C_1)(2\sigma_{xy} + C_2)}{(\mu_x^2 + \mu_y^2 + C_1)(\sigma_x^2 + \sigma_y^2 + C_2)} \tag{6}$$

Where, x, y represent the image to be evaluated and the reference image, respectively; μ_x, μ_y are the means of x and y; σ_x, σ_y are the standard deviations of x and y; σ_{xy} is the correlation coefficient between x and y. C_1 and C_2 are the two constants to stabilize the division with weak denominator. The value is constrained in [0,1] where the larger value represents a higher image quality. As for quantitative evaluation, Peak Signal to Noise Ratio (PSNR), Root Mean Square Error (RMSE), Mean Absolute Error (MAE) and Structural Similarity (SSIM) were chosen as evaluation indexes of image quality with the results of different methods provided in **Table 1**. We selected 100 test images from TCIA and calculated these objective evaluation indicators of denoised images by different methods on MATLAB R2016b. All the methods can significantly improve the quantitative evaluation indexes after image denoising; however, RED-CNN and our proposed DCRN perform at a higher level compared to BM3D and D-ResNet. Furthermore, when solely examining RED-CNN, which achieves cutting edge denoising results, our DCRN method remains slightly ahead in each quantitative evaluation indexes.

Table 1. Quantitative evaluation indexes of different methods

Method	PSNR	RMSE	MAE	SSIM
LDCT	26.91	12.77	5.417	0.8421
BM3D	29.29	9.346	3.458	0.8940
D-ResNet	31.62	7.081	2.628	0.9231
RED-CNN	33.11	6.110	1.998	0.9427
Proposed	**33.67**	**5.659**	**1.965**	**0.9434**

3.2.3 Computational Complexity

In addition to subjective visual effects and objective evaluation indexes, computational complexity is also an important factor in measuring the overall performance of an algorithmic model. Table 2 shows the model weights, CPU running calculation time, and GPU running calculation time, GPU training time, model size for various denoising methods. The running time and training time are in seconds. The measurement method is to calculate the average operation time of the corresponding method on 100 test images. Among the four methods, the execution time of traditional method BM3D is the shortest, but the denoising effect of it is the worst. Although the training time, running time in test images as well as the model size of D-ResNet are the best among the three networks, the objective evaluation indexes of it are far lower than the other two networks as shown in Table 1. Due to the shortcuts connection mechanism of RED-CNN, the training time of the proposed network is longer than RED-CNN, but the proposed network has fewer parameters and smaller model size, the execution time on test images is much shorter. Once the network is trained off, the trained network model can be used to process the test images, and the shorter running time on test images can make the method have higher engineering value. Considering the trade off between the performance and running time on test images as well as the training time, we suggest that the proposed method is superior to other comparison methods.

Table 2. Computational cost of different methods

Method	Parameters	Running time on CPU (secs)	Running time on GPU (secs)	Training time on GPU (secs)	Model size(KB)
BM3D	/	2.697	/	/	/
D-ResNet	2.25×10^5	2.951	0.0015	20209.0672	2716
RED-CNN	1.85×10^6	11.244	0.1800	20920.7125	21745
Proposed	3.95×10^5	4.437	0.0020	23656.9195	4789

3.3 The impact of network structure

This subsection will explore the impact of different network structural configurations on the denoising performance. In the calculation of CPU time, due to the large amount of computation of deep CNNs, the traditional method BM3D is faster than other methods based on CNNs; but at the same time, due to the computation of the convolutional neural network can be highly parallel, the method based on the convolutional neural network can achieve extremely low computation time on the GPU, while BM3D cannot be compared due to the lack of GPU-based implementation. Among the three methods based on convolutional neural network, D-Resnet has the least number of weights and the fastest CPU/GPU computing speed, but the cost is that its denoising effect is still far from the other two network models. For RED-CNN and the network implemented in this paper, the effect of the network implemented in this paper is slightly ahead of that of RED-CNN, but also reduces the weight by nearly 79%, greatly simplifying the complexity of the model. Meanwhile, the CPU/GPU computing speed of the network implemented in this paper is also significantly faster than that of RED-CNN, and the computing efficiency is significantly improved.

Table 3. The evaluation indexes with different number of Blocks

Number of Blocks	PSNR	RMSE	MAE	SSIM
3	33.45	5.800	2.017	0.9420
4	33.58	5.714	1.978	0.9431
5	**33.67**	5.659	**1.965**	0.9434
6	33.59	5.707	1.974	0.9432
7	33.66	**5.655**	1.966	**0.9436**

Table 4. The evaluation indexes with different number of feature maps

Number of feature maps	PSNR	RMSE	MAE	SSIM
32	33.51	5.756	2.012	0.9421
48	33.49	5.770	2.004	0.9422
64	**33.67**	5.659	**1.965**	**0.9434**
80	33.64	5.661	1.974	0.9431
96	**33.67**	**5.623**	1.982	0.9431

Table 3 shows the evaluation indexes with different number of Blocks (denoted as K). **Fig. 9** shows the polyline graph of the variation of evaluation indexes in the table with the number of convolution modules. As the number of convolution modules increases, the evaluation index trend rises until it reaches a certain threshold and then oscillates within a certain range. The PSNR and MAE indexes peak when $K = 5$, while the RMSE and SSIM indexes peak when $K = 7$. Since the values of the indexes in both cases are very close, it can be considered that both cases achieve the optimal denoising. performance. Therefore, the default setting of $K = 5$ reduces the number of network parameters while consistently producing optimal denoising results. **Table 4** shows the evaluation indexes with different number of feature maps (denoted as F) of each convolutional layer. **Fig. 10** shows that with the increase of the number of feature maps, the evaluation indexes overall also show the characteristics of first getting better and then oscillating. When F rises from 48 to 64, there is a significant improvement in the evaluation indexes. But when we continue to increase F, the indexes do not change much. Setting $F = 64$ can make a good balance between denoising performance and network complexity. From the experiments above, we can see that increasing the depth and width of the network does not necessarily lead to performance improvement, excessive complex network may also cause problems such as overfitting and thus have a negative impact on network performance.
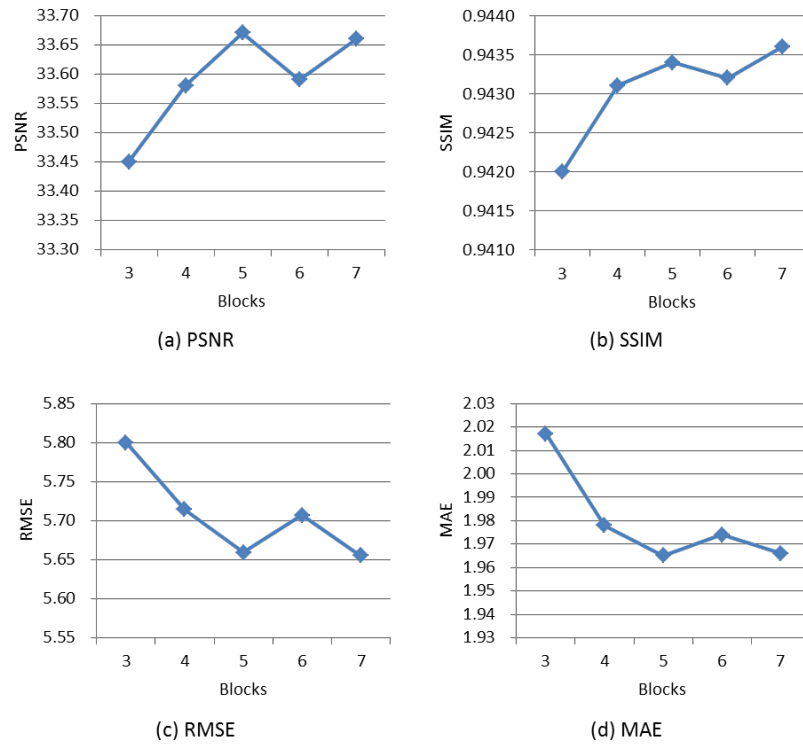
Fig. 9. The graph of the evaluation index with the number of convolution module
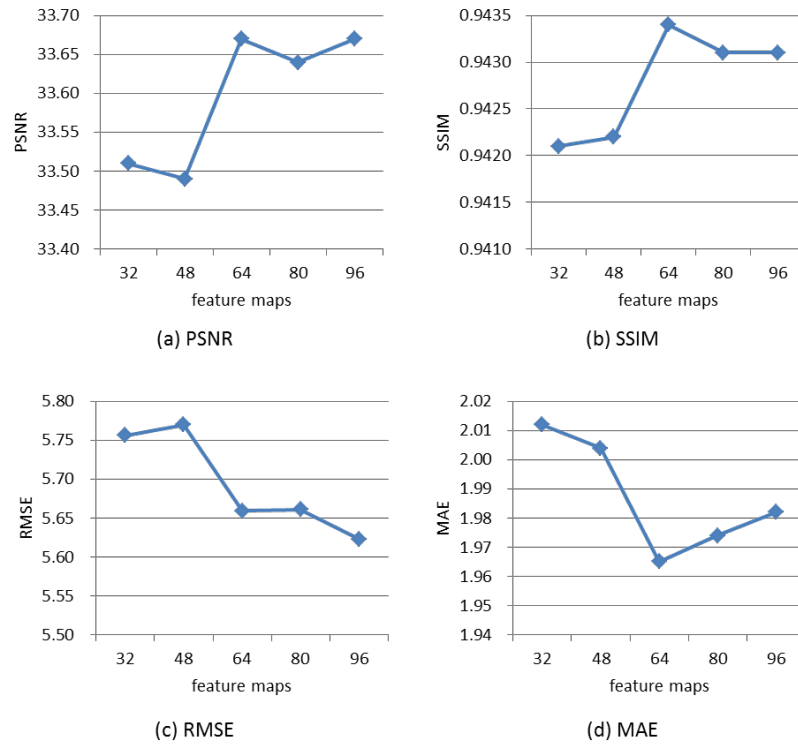

Fig. 10. The graph of the evaluation index changes with the quantity of convolutional layer output feature maps

Besides, we have also explored the impact of batch normalization, residual learning and dense connection on network denoising performance with the results displayed in **Table 5**. The evaluation indexes show a significant decline after removing batch normalization, residual learning or dense connection, which confirms the positive role of these three mechanisms in improving the denoising performance of the network.

Table 5. The evaluation indexes with different structural configurations

Architecture	PSNR	RMSE	MAE	SSIM
Baseline	**33.67**	**5.659**	**1.965**	**0.9434**
Baseline without batch normalization	33.52	5.739	2.049	0.9405
Baseline without residual learning	33.14	6.071	2.052	0.9419
Baseline without dense connection	33.45	5.800	1.983	0.9427

4. Conclusion

Since there are complex noise and artifacts in low-dose CT images, traditional methods often lose structural details when denoising images. Recently, due to the strong feature representation capabilities of deep learning methods, breakthroughs have been made in the field of computer vision. Inspired, we have proposed DCRN for low-dose CT image denoising, which mainly improves the denoising performance through three mechanisms: batch normalization, residual learning and dense connection. From the results of the experiment, our suggested DCRN obtains better performance in both visual quality and quantitative indexes while maintaining a low computational cost. If the proposed network can be trained with real images instead of simulated ones in the future, the reliability of the network model will be greatly improved, which will help translate the research results into actual products.

References

[1] D. P. Naidich, C. H. Marshall, C. Gribbin, R. S. Arams, and D. I. McCauley, "Low-dose of the lungs: preliminary observations," *Radiology*, vol.175, no.3, pp.729-731, Jun. 1990. Article (CrossRef Link)

[2] I. Mori, Y. Machida, M. Osanai, and K. Iinuma, "Photon starvation artifacts of X-ray CT: their true cause and a solution," *Radiological physics and technology*, vol.6, no.1, pp. 130-141, Jan. 2013. Article (CrossRef Link)

[3] Kachelriess M, Watzke O and Kalender W A, "Kalender 2001 Generalized multi-dimensional adaptive filtering for conventional and spiral single-slice, multi-slice, and cone-beam CT," *Medical physics*, vol. 28, no.4, pp. 475-490, Apr. 2001. Article (CrossRef Link)

[4] J. Wang, T. Li, H. Lu, and Z. Liang, "Penalized weighted least-squares approach to sinogram noise reduction and image reconstruction for low-dose X-ray computed tomography," *IEEE transactions on medical imaging*, vol. 25, no.10, pp. 1272-1283, Oct. 2006. Article (CrossRef Link)

[5]   A. Manduca, L. Yu, J. D. Trzasko, N. Khaylova, J. M. Kofler, C. M. McCollough, and J. G. Fletcher, "Projection space denoising with bilateral filtering and CT noise modeling for dose reduction in CT," *Medical physics*, vol. 36, no.11, pp. 4911-4919, Nov. 2009. Article (CrossRef Link)

[6]   M. Balda, J. Hornegger, and B. Heismann, "Ray contribution masks for structure adaptive sinogram filtering," *IEEE transactions on medical imaging*, vol. 31, no.6, pp. 1228-1239, Jun. 2012. Article (CrossRef Link)

[7]   B. R. Whiting, "Signal statistics in x-ray computed tomography," in *Proc. of SPIE Medical Imaging*, vol. 4682, pp. 53–60, May, 2002. Article (CrossRef Link)

[8]   Mengfei Li, Yunsong Zhao, and Peng Zhang, "Accurate Iterative FBP Reconstruction Method for Material Decomposition of Dual Energy CT," *IEEE transactions on medical imaging*, vol. 38, no.3, pp. 802-812, Mar. 2019. Article (CrossRef Link)

[9]   Z. Li, L. Yu, J. D. Trzasko, D. S. Lake, D. J. Blezek, J. G. Fletcher, C. H. McCollough, and A. Manduca, "Adaptive nonlocal means filtering based on local noise level for CT denoising," *Medical physics*, vol.41, no.1, pp. 011908, 2014. Article (CrossRef Link)

[10] D. Kang, P. Slomka, R. Nakazato, J. Woo, D. S. Berman, C.-C. J. Kuo, and D. Dey, "Image denoising of low-radiation dose coronary CT angiography by an adaptive block-matching 3D algorithm," in *Proc. of SPIE Medical Imaging*, vol. 8669, pp. 86692G, Mar. 2013. Article (CrossRef Link)

[11] Y. Chen, X. Yin, L. Shi, H. Shu, L. Luo, J.-L. Coatrieux, and C. Toumoulin, "Improving abdomen tumor low-dose CT images using a fast dictionary learning based processing," *Physics in Medicine & Biology*, vol.58, no.16, pp. 5803-5820, Aug. 2013. Article (CrossRef Link)

[12] Y. Chen, J. Liu, Y. Hu, J. Yang, L. Shi, H. Shu, Z. Gui, G. Coatrieux, and L. Luo, "Discriminative feature representation: an effective postprocessing solution to low dose CT imaging," *Physics in Medicine & Biology*, vol.62, no.6, pp. 2103-2131, 2017. Article (CrossRef Link)

[13] M. Diwakar, M. Kumar, "CT image denoising using NLM and correlation-based wavelet packet thresholding," *IET Image Processing*, vol.12, no.5, pp.708-715, May. 2018.
      Article (CrossRef Link)

[14] K. B. Khan, M. Shahid, H. Ullah, E. Rehman and M. M. Khan, "Adaptive trimmed mean autoregressive model for reduction of Poisson noise in scintigraphic images," *IIUM Engineering Journal*, vol. 19, no. 2, pp. 68-79, Dec. 2018. Article (CrossRef Link)

[15] K. B. Khan, A. A. Khaliq, M. Shahid and J. A. Shah, "A new approach of weighted gradient filter for denoising of medical images in the presence of Poisson noise," *Tehnički vjesnik*, vol.23, no.6, pp. 1755-1762, 2016. Article (CrossRef Link)

[16] Y. LeCun, Y. Bengio, and G. Hinton, "Deep learning," *Nature*, vol.521, no.7553, pp. 436-444, May. 2015. Article (CrossRef Link)

[17] H. Chen, Y. Zhang, W. Zhang, P. Liao, K. Li, J. Zhou, and G. Wang, "Low-dose CT denoising with convolutional neural network," in *Proc. of 14th IEEE International Symposium on Biomedical Imaging*, pp.143-146, Apr. 18-21, 2017. Article (CrossRef Link)

[18] H. Chen, Y. Zhang, W. Zhang, P. Liao, K. Li, J. Zhou, and G. Wang, "Low-dose CT via convolutional neural network," *Biomedical Optics Express*, vol.8, no.2, pp. 679-694, Feb. 2017. Article (CrossRef Link)

[19] H. Chen, Y. Zhang, M. K. Kalra, F. Lin, Y. Chen, P. Liao, J. Zhou, and G. Wang, "Low-dose CT with a residual encoder-decoder convolutional neural network," *IEEE transactions on medical imaging*, vol.36, no.12, pp. 2524-2535, Dec. 2017. Article (CrossRef Link)

[20] E. Kang, J. Min, and J. C. Ye, "A deep convolutional neural network using directional wavelets for low-dose X-ray CT reconstruction," *Medical physics*, vol.44, no.10, pp. e360-e375, Oct. 2017. Article (CrossRef Link)

[21] E. Kang, W. Chang, J. Yoo, and J. C. Ye, "Deep convolutional framelet denosing for low-dose CT via wavelet residual network," *IEEE transactions on medical imaging*, vol.37, no.6, pp. 1358-1369, Jun. 2018. Article (CrossRef Link)

[22] W. Yang, H. Zhang, J. Yang, J. Wu, X. Yin, Y. Chen, H. Shu, L. Luo, G. Coatrieux, Z. Gui, and Q. Feng, "Improving low- dose CT image using residual convolutional network," *IEEE Access*, vol.5, pp. 24698 - 24705, Oct. 2017. Article (CrossRef Link)

[23] M. Gholizadeh-Ansari, J. Alirezaie, and P. Babyn, "Low-dose CT denoising with dilated residual network," in *Proc. of 40th Annual International Conference of the IEEE Engineering in Medicine and Biology Society*, pp.5117-5120, Jul. 18-21,2018. Article (CrossRef Link)

[24] K. He, X. Zhang, S. Ren, and J. Sun, "Deep residual learning for image recognition," in *Proc. of the IEEE Conference on Computer Vision and Pattern Recognition*, pp.770-778, Jun. 26-Jul. 1, 2016. Article (CrossRef Link)

[25] G. Huang, Z. Liu, L. van der Maaten, and K. Q. Weinberger, "Densely connected convolutional networks," in *Proc. of the IEEE Conference on Computer Vision and Pattern Recognition*, pp. 4700-4708, Jul. 21-26, 2017 . Article (CrossRef Link)

[26] S. Ioffe, and C. Szegedy, "Batch normalization: Accelerating deep network training by reducing internal covariate shift," in *Proc. of the 32nd International Conference on Machine Learning*, pp.448-456, Jul. 7-9, 2015. Article (CrossRef Link)

[27] K. Zhang, W. Zuo, Y. Chen, D. Meng, and L. Zhang, "Beyond a Gaussian denoiser: Residual learning of deep CNN for image denoising," *IEEE Transactions on Image Processing*, vol.26, no.7, pp. 3142-3155, Jul. 2017. Article (CrossRef Link)

[28] V. Nair, and G. E. Hinton, "Rectified linear units improve restricted boltzmann machines," in *Proc. of the 27th International Conference on Machine Learning*, pp.807-814, Jun. 21-24, 2010. Article (CrossRef Link)

[29] Y. Jia, E. Shelhamer, J. Donahue, S. Karayev, J. Long, R. Girshick, S. Guadarrama, and T. Darrell "Caffe: Convolutional architecture for fast feature embedding," in *Proc. of the 22nd ACM international conference on Multimedia*, pp.675-678, Nov. 3-7 2014. Article (CrossRef Link)

[30] K. ClarkEmail, B. Vendt, K. Smith, J. Freymann, J. Kirby, P. Koppel, S. Moore, S. Phillips, D. Maffitt, M. Pringle, L. Tarbox, and F. Prior, "The Cancer Imaging Archive (TCIA): maintaining and operating a public information repository," *Journal of digital imaging*, vol.26, no.6, pp. 1045-1057, June , Dec. 2013. Article (CrossRef Link)

[31] D Zeng, J Huang, Z Bian, et al, "A simple low-dose X-ray CT simulation from high-dose scan," *IEEE Transactions on Nuclear Science*, vol.62, no.5, pp.2226-2233, Oct. 2015. Article (CrossRef Link)

[32] D. P. Kingma, and J. Ba, "Adam: A method for stochastic optimization," in *Proc. of the 3rd International Conference for Learning Representations*, pp.1-15, 2014. Article (CrossRef Link)

[33] K. Dabov, A. Foi, V. Katkovnik, and K, Egiazarian, "Image denoising with block-matching and 3D filtering," in *Proc. of. SPIE*, *Image Processing: Algorithms and Systems*, *Neural Networks*, *and Machine Learning*, vol. 6064, pp.606414, Jan. 2006. Article (CrossRef Link)

[34] Z. Wang, A. C. Bovik, H. R. Sheikh, and E. P. Simoncelli, "Image quality assessment: from error visibility to structural similarity," *IEEE transactions on image processing*, vol.13, no.4, pp.600-612, Apr. 2004. Article (CrossRef Link)

Jun Ming, received the B.S degree in communication engineering from Central China Normal University, Wuhan, China in 2018. She is now pursuing the M.S. degree in Wuhan University, Wuhan, China. Her research interests include computer vision.

Benshun Yi, received the B.S, M.S and Ph.D. degrees in electrical engineering from Huazhong University of Science and Technology, Wuhan, China, in 1986, 1989 and 1996, respectively. He is a Professor of Electronic Information school of Wuhan University, Wuhan, China. His research interests fall in the general areas of multimedia network communication, wireless network, and channel coding and he has published extensively in these areas.

YunGang Zhang, received the B.S degree in communication engineering and the M.S degree in information and communication engineering from Wuhan University, Wuhan, China, in 2016 and 2019, respectively. His research interests include computer vision.

Huixin Li, received the B.S degree in communication engineering from Wuhan University of Technology, Wuhan, China in 2019. She is now pursuing the M.S. degree in Wuhan University, Wuhan, China. Her research interests include computer vision.