

A selective kernel-based cycle-consistent generative adversarial network for unpaired low-dose CT denoising

Chaoqun Tan¹, Mingming Yang², Zhisheng You¹, Hu Chen^{1,2,*} and Yi Zhang^{2,*}

¹National Key Laboratory of Fundamental Science on Synthetic Vision, Sichuan University, Chengdu 610065, China

²College of Computer Science, Sichuan University, Chengdu 610065, China

*Correspondence: Hu Chen, huchen@scu.edu.cn; Yi Zhang, yzhang@scu.edu.cn

Abstract

Low-dose computed tomography (LDCT) denoising is an indispensable procedure in the medical imaging field, which not only improves image quality, but can mitigate the potential hazard to patients caused by routine doses. Despite the improvement in performance of the cycle-consistent generative adversarial network (CycleGAN) due to the well-paired CT images shortage, there is still a need to further reduce image noise while retaining detailed features. Inspired by the residual encoder-decoder convolutional neural network (RED-CNN) and U-Net, we propose a novel unsupervised model using CycleGAN for LDCT imaging, which injects a two-sided network into selective kernel networks (SK-NET) to adaptively select features, and uses the patchGAN discriminator to generate CT images with more detail maintenance, aided by added perceptual loss. Based on patch-based training, the experimental results demonstrated that the proposed SKFCycleGAN outperforms competing methods in both a clinical dataset and the Mayo dataset. The main advantages of our method lie in noise suppression and edge preservation.

Keywords: cycle-consistent adversarial network, selective kernel networks, unsupervised low dose CT, image denoising, clinical dataset

Introduction

X-Ray computed tomography (CT) as a common non-invasive radiological diagnostic method is widely known to have potential for the examination of various diseases such as pneumonia, tumor, infarction and bleeding.^{1,2} The important role of CT in following-up on the effects on lung tissue has been widely recognized for clinical diagnosis and monitoring of COVID-19.³ One increasing concern about CT is the threat of excessive radiation dose. Research reducing CT dose under the guiding principle of ALARA (as low as reasonably achievable) has aroused strong attention.⁴ The universal and effective strategy to minimize the risk is to obtain low dose CT (LDCT) by decreasing the tube current of the X-ray tube and shorten the exposure time during shooting.⁵ However, lowering radiation dose inevitably increases artifacts and causes noise in reconstructed images, which could degrade the signal-to-noise ratio and affect the judgment performance. Thus, how to improve the image quality for LDCT has become a significant topic in the field of image denoising.

To date, algorithms generally include sinogram domain filtration,^{6,7} iterative reconstruction (IR),^{8,9} and image processing.^{10,11} The common sinogram filtering methods are difficult to perform on transparent raw data from commercial scanners before image reconstruction, which can lead to resolution loss and edge blurring. Simultaneously, these methods may induce artifacts in the generated image during data processing. In comparison, IR has contributed greatly to the field of LDCT. These algorithms optimize an objective function that incorporates an accurate system model, a statistical noise model, and prior information in the

image domain. The common algorithms include total variation and its variants,¹²⁻¹⁴ dictionary learning,^{15,16} low-rank¹⁷ and so on. These iterative reconstruction algorithms greatly improve image quality but images may still lose some detail and suffer from remaining artifacts. Also, they require a high computational cost, which is a bottleneck in practical applications.

As an effective alternative, image post-processing has advantages with regard to non-essential raw data and efficiency. Due to developments in artificial intelligence and deep learning (DL),¹⁸⁻²¹ DL-based algorithms have attracted extensive attention to learn the mapping pixels-to-pixels for the corresponding routine-dose image by training with pairs of low-dose images and matched high-dose CT data. Kang *et al.*²² have proposed the DL-model combined with wavelet transform to effectively suppress noise and artifacts in LDCT, but a long training time is required. Chen *et al.*²³ used the classical residual encoder-decoder convolutional neural network (RED-CNN) for LDCT denoising, and five-layer to simplify network structure and to outperform the state-of-the-art methods. Noise reduction in the above-mentioned end-to-end network notwithstanding, mean square error (MSE)-based methods have usually oversmoothed the subtle structural details by minimized per-pixel MSE. Therefore, the generative adversarial network (GAN) is used to overcome these limitations.²⁴ Wolterink *et al.*²⁵ have applied GAN to achieve noise supersession in LDCT. Yang *et al.*²⁶ have introduced the Wasserstein distance to design WGAN-GP to better retain feature information in LDCT, simultaneously, perceptual loss is used to optimize the loss function. Du *et al.*²⁷ have brought a visual attention mechanism into GAN to

Received: December 2, 2021. Accepted: April 15, 2022. Published: 20 April 2022

© The Author(s) 2022. Published by Oxford University Press on behalf of the West China School of Medicine & West China Hospital of Sichuan University. This is an Open Access article distributed under the terms of the Creative Commons Attribution License (<https://creativecommons.org/licenses/by/4.0/>), which permits unrestricted reuse, distribution, and reproduction in any medium, provided the original work is properly cited.

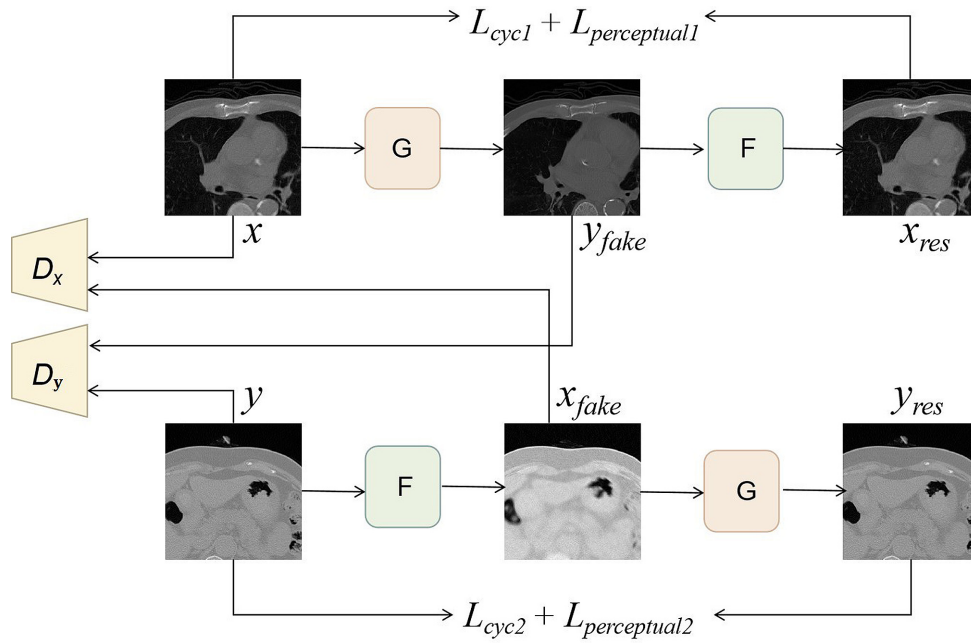


Figure 1. The overall framework of the CT image denoising method based on adaptive feature selection. NDCT (normal-dose CT) and LDCT images are generated by generators G and F, respectively. Discriminators D_x and D_y are used to distinguish which data-domain the input belongs to. Cyclic-consistency loss and perceptual loss are used to constrain the generated CT image corresponding to the input image.

propose VAGAN, which focuses on the information needed to achieve satisfactory performance and effectively suppress noise. You *et al.*²⁸ have proposed a semi-supervised network, which achieves the generation of high-resolution CT from low-resolution CT. In general, GAN has obtained widespread interest by generating different models in LDCT.

Despite the remarkable improvement in performance, well-paired CT images for supervised training are difficult to obtain in clinical practice. Furthermore, due to the potential mode-collapsing characteristics of GAN, redundant features could be generated to affect the accuracy of diagnosis in clinical practice. Thus, unsupervised learning has caused public concern regarding unmatched-pair LDCT images. The cycle-consistent generative adversarial network (CycleGAN) is an image-to-image conversion algorithm, which is composed of two generators and two discriminators to achieve cycle-consistency by performing input-to-target image domain translation without well-paired data.²⁹ However, unsupervised LDCT denoising using CycleGAN is not very effective in noise suppression.³⁰⁻³² In this study, we designed a novel selective feature network using CycleGAN, where the unsupervised learning model can reduce LDCT image noise by adaptively selecting features.

Methods

Overview of LDCT denoising model

The purpose of noise reduction is to make LDCT images as similar as possible to normal dose CT (NDCT) images. This process can be simplified to the following:

$$G = x \rightarrow y$$

Where X images domain is defined as LDCT data and Y images domain is defined as NDCT data. The overview structure of the network is shown in Fig. 1.

When LDCT x ($x \in X$) is input, the corresponding NDCT y_{fake} ($y_{fake} \in Y$) is generated by G. Then, the corresponding LDCT x_{res} ($x_{res} \in X$) is regenerated from y_{fake} ($y_{fake} \in Y$) by F. D_x is used to discriminate x and x_{fake} , and D_y is used to distinguish y and y_{fake} . The cyclic-consistency loss function³³ is introduced to constrain the generated CT image corresponding to the input x . The perceptual loss³³ function is introduced to calculate the pixel-to-pixel distance between x and x_{res} . When NDCT y ($y \in Y$) is input, the corresponding LDCT x_{fake} ($x_{fake} \in X$) and NDCT y_{res} ($y_{res} \in Y$) are generated, respectively. Similarly, the cyclic-consistency loss function and the perceptual loss function are used to compute the distance between y and y_{res} .

Adaptive feature selection generator

RED-CNN applies convolution and deconvolution instead of pooling and up-sampling, and shortcut connections which can reverse the loss of LDCT image edges and structural information. However, RED-CNN is used as the generator for unsupervised training, so noise and artifacts cannot be effectively suppressed in a clinical CT dataset. In addition, pooling and up-sampling are used to effectively remove noise and artifacts in U-NET,³⁴ but result in serious loss of structural detail and cause blurred edges of images. Inspired by this, the combination of RED-CNN, U-NET, and a 1×1 convolutional layer as extractor, which is injected into SK-NET,³⁵ is used to adaptively select features obtained from different networks with different convolution kernels. The overall structure of the proposed generator is shown in Fig. 2A. The improved RED-CNN and U-NET as a bilateral network with 1×1 convolutional block is used to extract the features of LDCT. The three different feature maps of U_1 , U_2 and U_3 are obtained by using the different convolution kernels of SK-NET. Then different feature maps are made by element-wise summation $U = U_1 + U_2 + U_3$, and encode global information through global-average pooling to generate channel-level information $S \in R^c$. The formation is defined

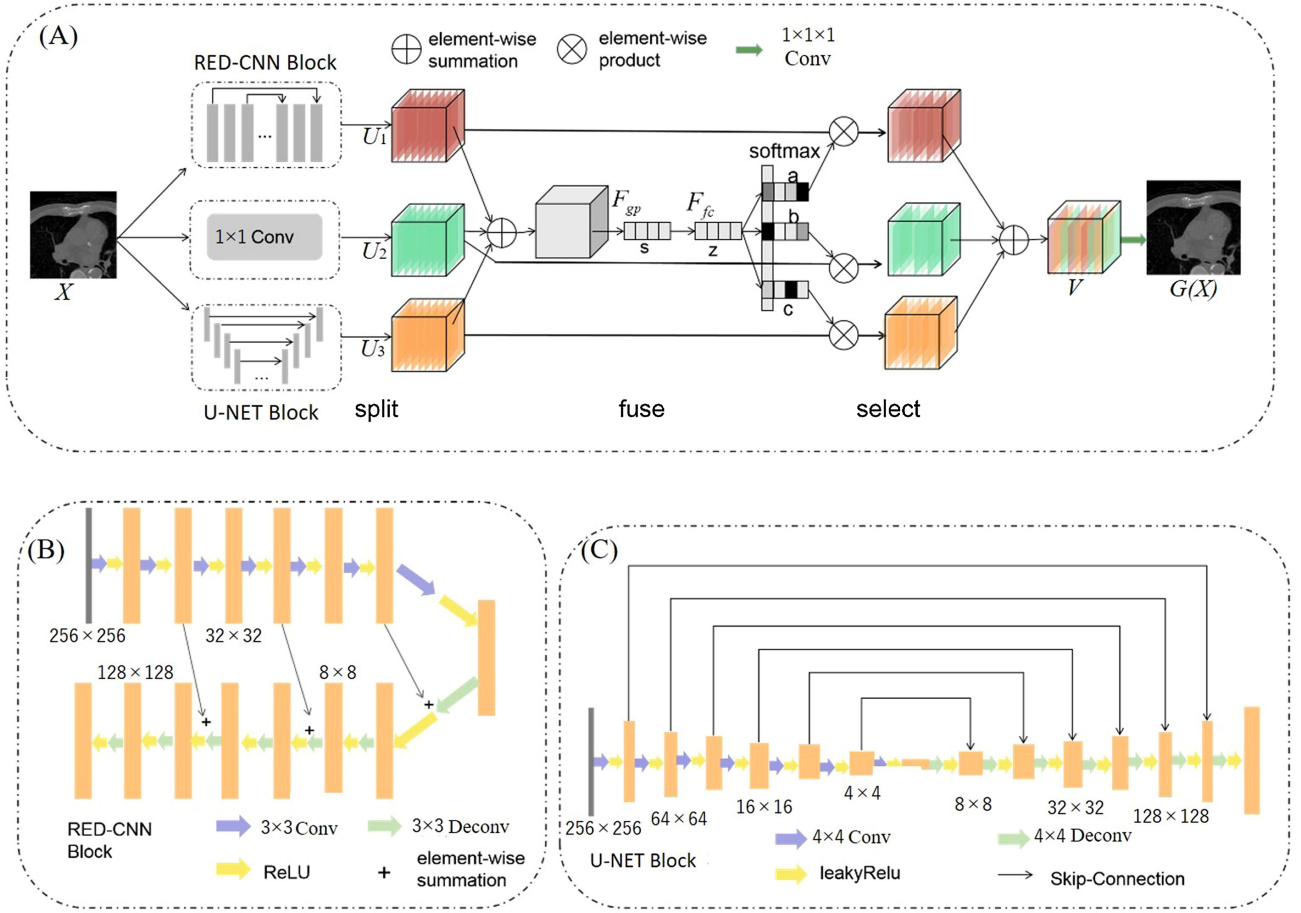


Figure 2. Network structure of the generator: (A) framework of generator, (B) RED-CNN block, and (C) U-NET block.

as:

$$S_c = F_{gp}(U_c) = \frac{1}{H \times W} \sum_{i=1}^H \sum_{j=1}^W U_c(i, j) \quad (1)$$

The dimensionality of S is reduced to produce a compact global feature $Z \in \mathbb{R}^{d \times 1}$ from a fully connected layer.

$$Z = F_{fc}(S) = \delta(B(W_s)) \quad (2)$$

The softmax layer is used to generate three different weights a , b and c , where $a_c + b_c + c_c = 1$.

$$a_c = \frac{e^{A \cdot z}}{e^{A \cdot z} + e^{B \cdot z} + e^{C \cdot z}}, b_c = \frac{e^{B \cdot z}}{e^{A \cdot z} + e^{B \cdot z} + e^{C \cdot z}}, c_c = \frac{e^{C \cdot z}}{e^{A \cdot z} + e^{B \cdot z} + e^{C \cdot z}} \quad (3)$$

Finally, the 1×1 convolutional layer is applied to obtain the output feature V of CT images. The formation is defined as:

$$V_c = a_c \cdot U_1 + b_c \cdot U_2 + c_c \cdot U_3 \quad (4)$$

Both good denoising performance and clear structural details can be gained. The details of the generator network are described as follows.

Abundant structural detail

To reduce the distortion of structural detail caused by up-sampling, a 1×1 convolution layer with 64 channels is added

to better achieve cross-channel interaction and retain the integration of LDCT information. In this study, RED-CNN, U-NET, and a 1×1 convolutional block are used to extract the features of LDCT, and are split into three different feature maps of the same size $256 \times 256 \times 64$.

Adaptive feature selection

Due to the redundancy of features extracted from different networks, the effective features can be selected in this study. Inspired by SK-NET, based on split, fuse, and select, a network that can adaptively select the features obtained by different neural networks is designed. Not only can the size of the receptive field of different convolution kernels be adjusted, but the features extracted by different networks can be effectively fused.

Improved RED-CNN and U-NET

The network structure of RED-CNN is shown in Fig. 2B. It consists of 14 layers, including 7 convolutional and 7 deconvolutional symmetrical layers. The convolutional and deconvolutional layers have the same kernel size of 3×3 and 64 channels. Rectified linear units (ReLU)³⁶ are added for each layer. The shortcuts connection is matched between the convolutional and deconvolutional layers, which improves the convergence speed. U-NET is composed of 7 convolutional and 7 deconvolutional layers with the same kernel size of 4×4 and 64 channels. An ReLU is added for each layer. Feature maps of $256 \times 256 \times 64$ are obtained. The network structure of U-NET is shown in Fig. 2C.

Discriminator design

The discriminator scores the generated images and guides the training of the generator. Inspired by PatchGAN,²² the input images are mapped to a patch of size 4×4 , which is the feature map obtained by the convolutional layer. x_{ij} is the probability of belonging to NDCT, which corresponds to each patch X of the input image. The average of x_{ij} is the output of the discriminator. For training of the generated NDCT, the receptive field of the discriminator can be effectively improved to retain high-definition details of images. The network structure of the discriminator is shown in Supplementary Fig. 1, see online supplementary material.

Loss function

The overall loss function is as follows:

$$\text{Loss} = L_{GAN}(G, D_Y, X, Y) + L_{GAN}(F, D_X, Y, X) + \lambda L_{cyc}(G, F) + L_{perceptual} \quad (5)$$

where L_{GAN} is the adversarial loss, L_{cyc} is the cycle-consistency loss, $L_{perceptual}$ is the perceptual loss function.

Adversarial loss

Adversarial loss³³ is used to calculate the two mapping functions. The loss function is shown as follows.

$$L_{GAN}(G, D_Y, X, Y) = E_{Y \sim P_{data}(y)} [\log D_Y(y)] + E_{X \sim P_{data}(x)} [\log(1 - D_Y(G(x)))] \quad (6)$$

where X is LDCT; Y is NDCT; G is the generator for $X \rightarrow G(X)$. D_Y aims to distinguish $G(x)$ and Y . G aims to minimize the gap and maximize the difference of adversary, i.e. $\min_G \max_{D_Y} L_{GAN}(G, D_Y, X, Y)$. Similarly, the function F is introduced to map $Y \rightarrow F(Y)$, i.e. $\min_F \max_{D_X} L_{GAN}(F, D_X, Y, X)$.

Cycle-consistency loss

Theoretically, adversarial loss can compute which target domain the generated output belongs to. However, with large enough capacity, the mapping of inputs to outputs in the target domain can be randomly arranged, where the mapping cannot be guaranteed to pair the output with the corresponding input. Therefore, to further constrain the matching of the generated image and decrease the data domain space of mapping functions, the cycle consistency can be loaded in the mapping functions G and F . The cycle-consistency loss³³ is defined as

$$L_{cyc}(G, F) = E_{X \sim P_{data}(x)} [\|F(G(x)) - x\|_1] + E_{Y \sim P_{data}(y)} [\|G(F(y)) - y\|_1] \quad (7)$$

For the translation cycle of $X \rightarrow G(X) \rightarrow F(G(X))$, the cycle-consistency function can give the generated images corresponding to X by calculating the L1-norm between X and $F(G(X))$. This is called forward cycle-consistency. Similarly, the mapping functions G and F should also conform to backward cycle-consistency for the translation of $Y \rightarrow F(Y) \rightarrow G(F(Y))$.

Perceptual loss

Cycle-consistency loss can calculate the pixel-level distance to ensure the matching of generated image and input image; however, structural texture and details are lost. That is why the perceptual loss is added to guide the generator to learn more feature details of images. The perceptual loss³⁷ is usually calculated from

the feature maps of VGG-16. The perceptual loss function is defined as follows.

$$\text{LOSS}_{perceptual} = E_{X \sim P_{data}(x)} \|\phi(x) - \phi(F(G(x)))\|_1 + E_{Y \sim P_{data}(x)} \|\phi(y) - \phi(G(F(y)))\|_1 \quad (8)$$

where ϕ is the feature maps. In this section, both the second max-pooling layer with 128 channels and the last max-pooling layer with 512 channels in VGG-16 are used to calculate the perceptual loss.

Evaluation metrics

Root mean square error

Root mean square error (RMSE) is used to measure the deviation between the generated CT image and NDCT. It is computed by using the arithmetic square root of MSE. The definition is shown in Eq. (9).

$$\text{RMSE} = \sqrt{\text{MSE}}, \quad (9)$$

$$\text{MSE} = \frac{1}{mn} \sum_{i=0}^{m-1} \sum_{j=0}^{n-1} [I(i, j) - K(i, j)]^2$$

where $m \times n$ is the size of the clear image I and the noise image K .

Peak signal-to-noise ratio

Peak signal-to-noise ratio (PSNR) reflects the ratio between the maximum signal of the image and the noise of the image. It is an evaluation index for calculating errors between corresponding pixels. The higher the PSNR, the better the image quality. PSNR is defined in Eq. (10).

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

where MAX_I^2 is the maximum of pixels.

Structural SIMilarity

Structural SIMilarity (SSIM) is applied to judge the similarity between X and Y , based on the brightness, contrast, and structure of images. SSIM is defined in Eq. (11).

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

where μ_x, μ_y are the mean of x, y ; σ_x^2, σ_y^2 are the variances of x and y , respectively; σ_{xy} is the covariance of x, y ; $c_1 = (k_1L)^2$, $c_2 = (k_2L)^2$ are constants, avoid division by zero, where L is the range of pixel values. Usually, $k_1 = 0.01$, $k_2 = 0.02$ and $c_3 = c_2/2$.

Qualitative evaluation

In addition to quantitative indicators, a qualitative indicator may also be needed to evaluate the quality of the generated CT image. Since CT images are to help clinicians make pathological diagnoses, the denoising results of LDCT are judged by the subjective senses of two professionals.

Experimental design and results

Experimental datasets

Clinical dataset

A real unmatched-pair clinical database was applied for evaluating the performance of the proposed model. There were two different 512×512 LDCT images, 10 ma and 30 ma respectively. 160ma is the dose of NDCT image in both datasets. This clinical dataset

was the main problem to be solved. The specific parameters of two different LDCT images are shown in Supplementary Table 1, see online supplementary material.

Due to the diversity of the human body, different window widths are required for observing and displaying CT images. Window widths of 0.2–0.28 and 0–0.33 were selected, respectively, where 0.2–0.28 was used for observing the tissue and 0–0.33 was used for observing the lung. Some typical CT images are shown in Supplementary Fig. 2 (see online supplementary material). The 10 ma LDCT dataset was composed of 337 pairs of LDCT and NDCT images from 6 anonymous patients. In our experiments, 218 pairs of LDCT and NDCT images from 5 patients were randomly selected for training and the remaining pairs were used for testing. To effectively train the network, patch-based extraction was performed to obtain CT images with local details required for denoising training and to increase the number of samples. All CT images were cropped into a 256×256 patch every 16 pixels, and 81 209 pairs of 256×256 LDCT and NDCT images were obtained.

The 30 ma LDCT dataset contained 562 pairs of LDCT and NDCT images from 10 anonymous patients. A total of 498 pairs of LDCT and NDCT images from 9 patients were randomly selected for training and the remaining pairs were used for testing. Similarly, 143 922 pairs of 256×256 LDCT and NDCT images were obtained.

Mayo dataset

Mayo,³⁸ a publicly available dataset with paired NDCT and LDCT images, was created in “the 2016 NIH-AAPM-Mayo Clinic Low Dose CT Ground Challenge” to evaluate the performance of LDCT denoising algorithms. The dataset included 5936 NDCT images with 512×512 and quarter-dose simulated LDCT images from 10 patients. In this paper, 135 250 pairs of 256×256 CT image patches from 9 patients were randomly extracted as the training set, and the remaining data from 1 patient was used as a testing set. For the research of unsupervised LDCT image denoising, all pairs were in disorder. This was used to validate the effectiveness of the proposed model.

Training parameter

In this study, the model was optimized by the Adam³⁹ algorithm. The initial learning rate was 10^{-5} which decreased linearly from iterating 100 000 steps. It remained unchanged until it was reduced to 10^{-7} at 700 000 steps. The weight parameter λ of the generator was set to 10 and the batch size was set to 1. The experiment was based on Tensorflow14 and Python 3.6. The model was trained on a PC (Intel i7 processor and 11G video memory) with a graphics processing unit card (Nvidia 2080TI). The final model was obtained when 1 million steps was reached.

The experiments were performed on two different LDCT datasets. Multiple different state-of-the-art algorithms were compared with ours in the clinical dataset, including BM3D,⁴⁰ K-SVD,⁴¹ and unsupervised models CCADN³⁰ and CycleGAN. Moreover, BM3D, K-SVD, supervised methods such as RED-CNN²³ and Q-AE,⁴² and unsupervised algorithm CycleGAN and CCADN were compared with ours on Mayo dataset. Due to the unpaired clinical dataset, it was impossible to calculate quantitative indicators such as PSNR, RMSE, and SSIM. Therefore, qualitative evaluations were mainly carried out by the visual sense of experts.

Experimental results

Clinical dataset

0.2-0.28 window width

Two LDCT images were used to test the performance of the model, with window widths of 0.2–0.28 and 0–0.33, respectively. The experimental results and the magnified images for a region of interest (ROI) with window width of 0.2–0.28 are shown in Fig. 3A and B. The tissue can be observed, i.e. the areas with higher gray values. For the LDCT of Fig. 3A, image noise and artifacts exist near the structure with a high attenuation coefficient. All methods suppressed image noise to different extents. For multiple comparison results, our method has achieved the best experimental results.

However, other methods could not effectively suppress noise or cause blurring of CT images to various degrees. As seen in Fig. 3A, the noise of LDCT could be effectively suppressed for BM3D, but the generated CT images were so blurred that there was serious loss of structural details. K-SVD can also cause blurring of edges and a good quality CT image is not obtained. Compared with traditional algorithms, the unsupervised algorithm CCADN achieved better denoising, but there was still noise and artifacts. It can be seen, particularly from Fig. 3B, that with CCADN and CycleGAN the noise was not effectively suppressed and artifacts in the generated CT images could cover the local details of the images. However, it is clear that compared to the above-mentioned algorithms, the noise was more effectively suppressed and clearer structural features are retained.

0-0.33 window width

CT images of lungs could be observed with a window width of 0–0.33, i.e. the black area marked by the blue rectangle in Fig. 4. The results were similar to the training results with the 0.2–0.28 window width. It can be concluded that ours achieved the best denoising performance compared with other algorithms, and the best quality of CT image was generated. The experimental results and magnified images of an ROI based on the different algorithms are shown in Fig. 4A and B.

Although most image noise and artifacts were eliminated, the structural details were smoothed by BM3D and K-SVD. Meanwhile, for the unsupervised learning algorithms CCADN and CycleGAN, obviously, there was still more noise and artifacts compared with ours, especially for the results based on CycleGAN. SK-FCycleGAN can generate a clear lung CT image.

Usually, the five metrics including noise suppression, artifact reduction, lesion discrimination, contrast retention, and overall quality are used for subjective evaluation by doctors (5 is best, 1 is lowest). On the basis of the different algorithms, two radiologists with 6 years of clinical experience respectively provided their scores. The unpaired NDCT images were references, and the average scores of the two experts were used as final results. The statistical results are shown in Supplementary Table 2, see online supplementary material. From the five indicators, all of the methods could suppress LDCT noise and effectively reduce artifacts, but our algorithm produced better scores than the other methods.

Mayo dataset

To better verify the robustness and generalization of the proposed method, it was tested on the Mayo dataset and compared with the traditional algorithms such as BM3D and K-SVD, the classical supervised learning algorithms RED-CNN and Q-AE, and the unsupervised learning algorithms CycleGAN and CCADN. The experimental results on the Mayo dataset are shown in Fig. 5.

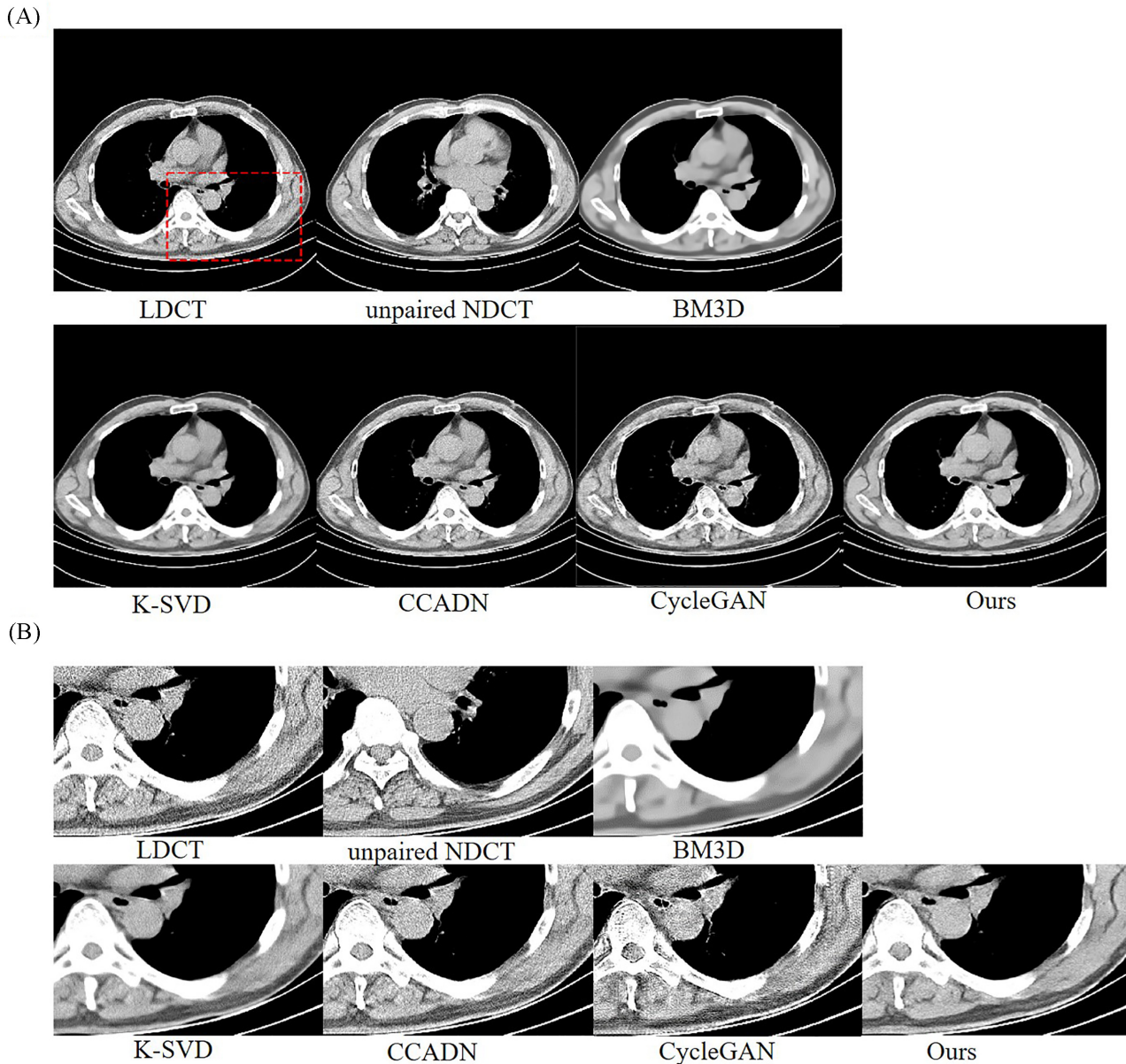


Figure 3. Results and magnified images over a ROI of a 30 ma dataset with 0.2–0.28 window width for comparison. The BM3D, K-SVD, and unsupervised algorithms CCADN and CycleGAN have been compared in this study. The area indicated by a red rectangle is the ROI that was magnified to show the experimental results.

For Fig. 5A, the denoising result of BM3D was over-smoothed, and noise and artifacts were not effectively suppressed in the K-SVD image. Even if RED-CNN gave the best performance, our method had better denoising results compared to traditional algorithms, CCADN and CycleGAN. The results and the magnified ROI for the Mayo dataset are shown in Fig. 5B.

In the magnified ROI of BM3D, the experimental results obtained still presented too blurred CT images. Since MSE-based algorithm is usually trained by minimized per-pixel MSE, the generated CT has higher quantitative value and is closer to NDCT images. For the unsupervised learning algorithm, there was a lack of matched LDCT and NDCT images, thus denoising results could only be obtained and compared visually. A comparison of the quantitative results obtained using the different methods is shown in Table 1.

It can be seen that the MSE-based supervision learning methods had the best evaluation metrics compared with other mod-

els. RMSE, PSNR, and SSIM were also calculated by MSE. Compared with traditional algorithms and unsupervised learning algorithms, the best performance was obtained for RMSE, PSNR, and SSIM with our algorithm, with values of 0.0085, 41.45 and 0.9535, respectively.

Different model and performance trade-offs

In these experiments, to verify the effectiveness of designed generator networks, ours was compared with four different models, including a generator with only RED-CNN, a generator with only U-NET, and a generator with concatenated three-feature vectors and without a 1×1 convolutional block. In addition, a model without a perceptual loss function was also examined to prove the availability of the added perceptual loss function. To ensure fairness, the remaining parameters remain unchanged in the comparison.

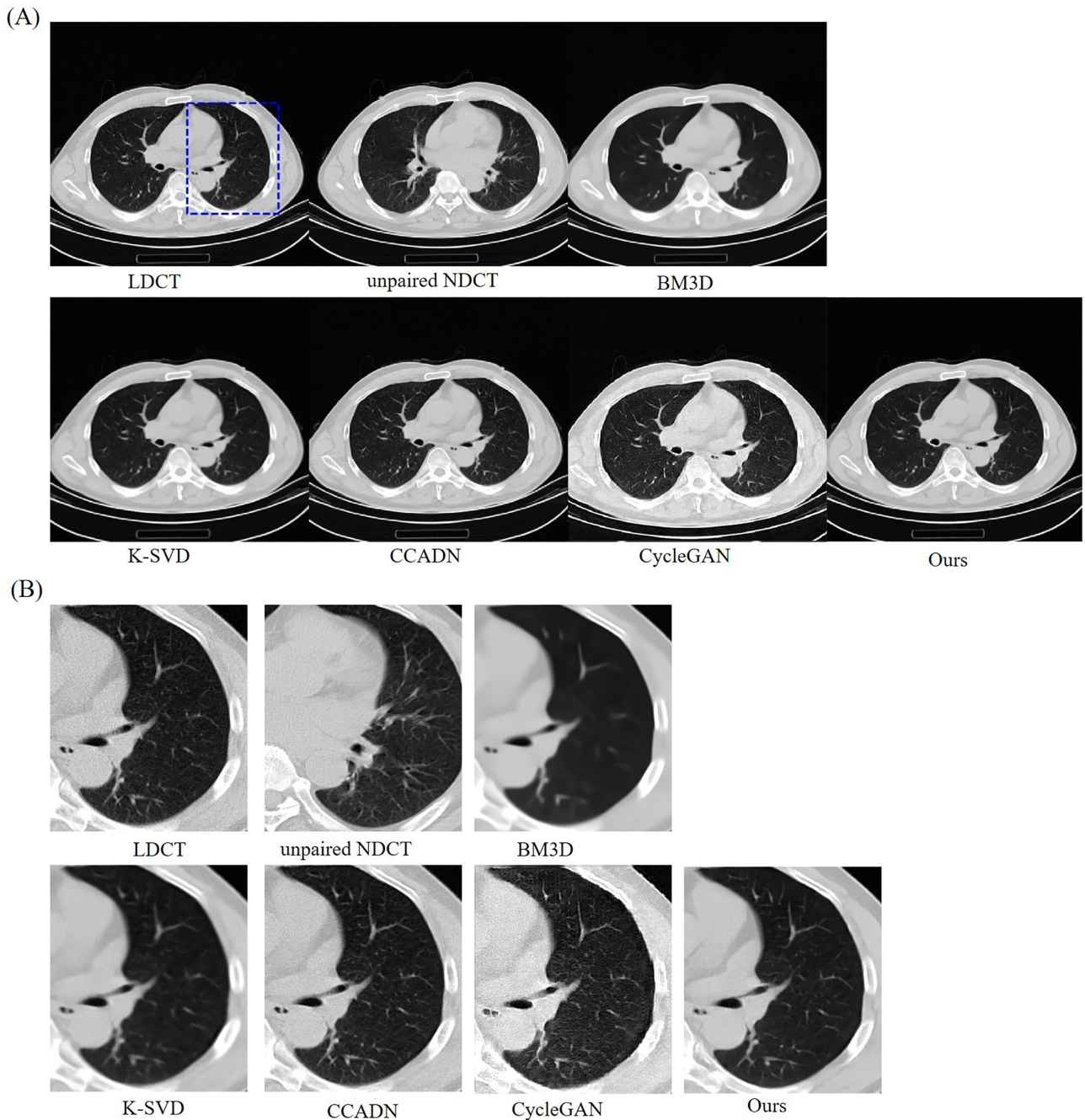


Figure 4. Results and magnified images of a ROI of a 30 ma dataset with 0–0.33 window width for comparison. The BM3D, K-SVD, and unsupervised algorithms CCADN and CycleGAN were compared in this study. The black area indicated by the blue rectangle is the ROI that was magnified to show the experimental results.

The experimental results and the magnified ROI for the different models are shown in Fig. 6.

In Fig. 6B, the generated CT images obtained still had noise, and artifacts were introduced. CT image contrast was reduced only for RED-CNN as the generator. Detailed features of the generated CT image were severely lost when the improved U-NET was used as generator. Some artifacts appeared when the three-feature vectors were directly concatenated. Additionally, using our designed network without perceptual loss, the noise could not completely be removed and a certain amount of blurring was produced. The 1×1 convolutional layer was added to achieve better cross-channel correlation and retain the integration of LDCT

information. It also introduced more nonlinearity and improves generalization ability. It is clear that there was still noise and distortion of structural detail without the 1×1 convolutional block. Finally, our method effectively suppressed noise and artifacts, retained more structure and the edges of detailed features, and outperformed competing models.

Different dose and performance trade-offs

Based on the reduced dose, ensuring the completeness of the image details as much as possible is suitable for clinical practice. In clinical datasets, 10 ma and 30 ma images have differences for the preservation of detailed information and the degree of noise.

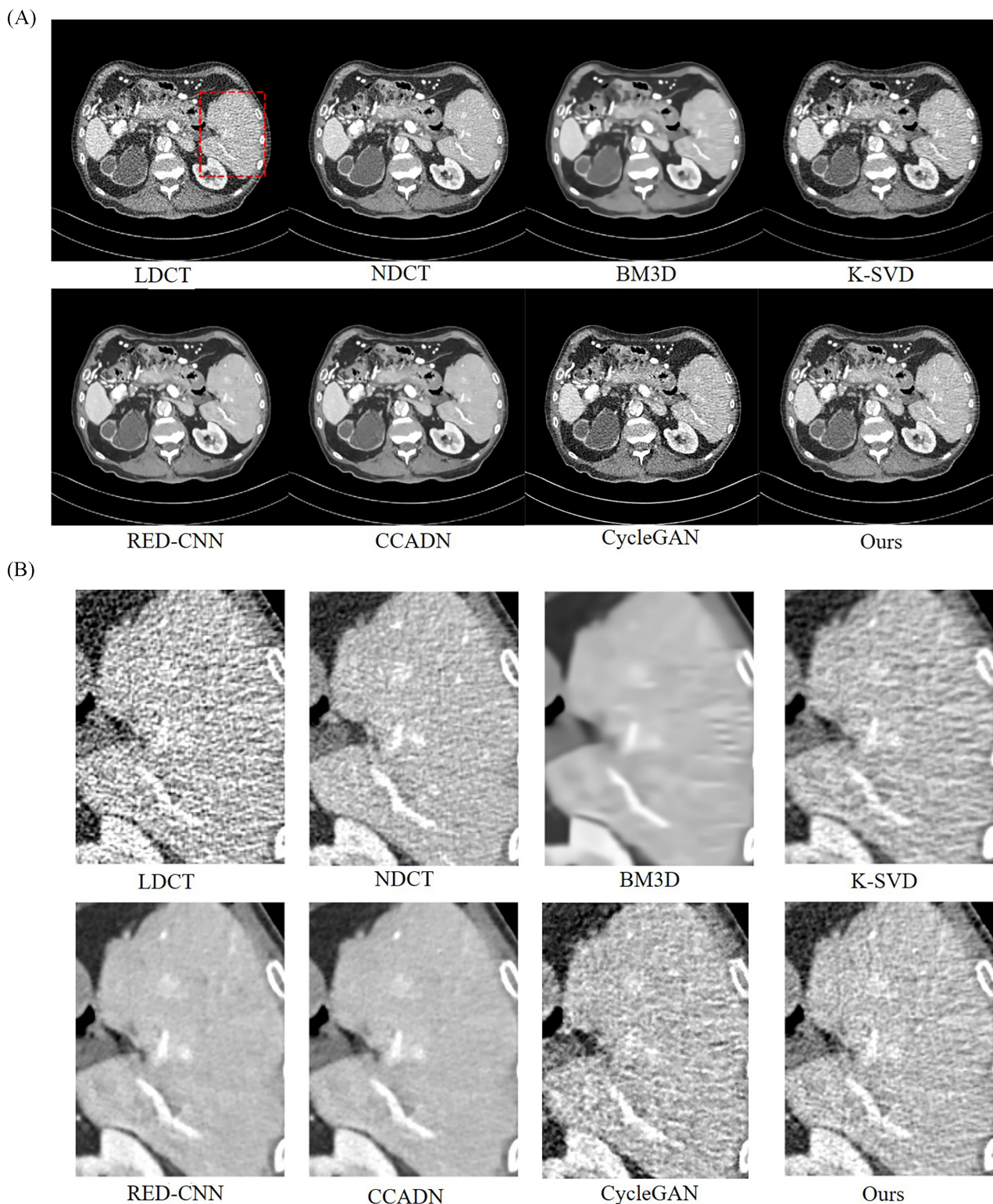


Figure 5. Results of the Mayo dataset and a magnified ROI for comparison of the BM3D, K-SVD, RED-CNN, CCADN and CycleGAN algorithms used in this study. The area indicated by a red rectangle is the ROI that was magnified to show the experimental results.

In this paper, 30 ma LDCT images were used as the main research data, and 10 ma LDCT images were tested to verify the significance of the proposed method. The unsupervised learning algorithm CCADN was compared with ours. Discussion of different doses of CT images has also demonstrated the applicability of our

method. The experimental results and the magnified ROI of different doses are shown in Fig. 7.

From Fig. 7B it can be seen that our method achieved good denoising results on the 10 ma CT dataset, and existing structural details were properly retained, as indicated by yellow and purple

Table 1. Comparison of quantitative results associated with the Mayo dataset.

Model	RMSE	PSNR	SSIM
BM3D	0.0097	40.23	0.9385
K-SVD	0.1284	37.83	0.9455
RED-CNN	0.0065	43.71	0.9686
CCADN	0.0089	41.09	0.9471
CycleGAN	18.8017	22.65	0.8469
Ours	0.0085	41.45	0.9535

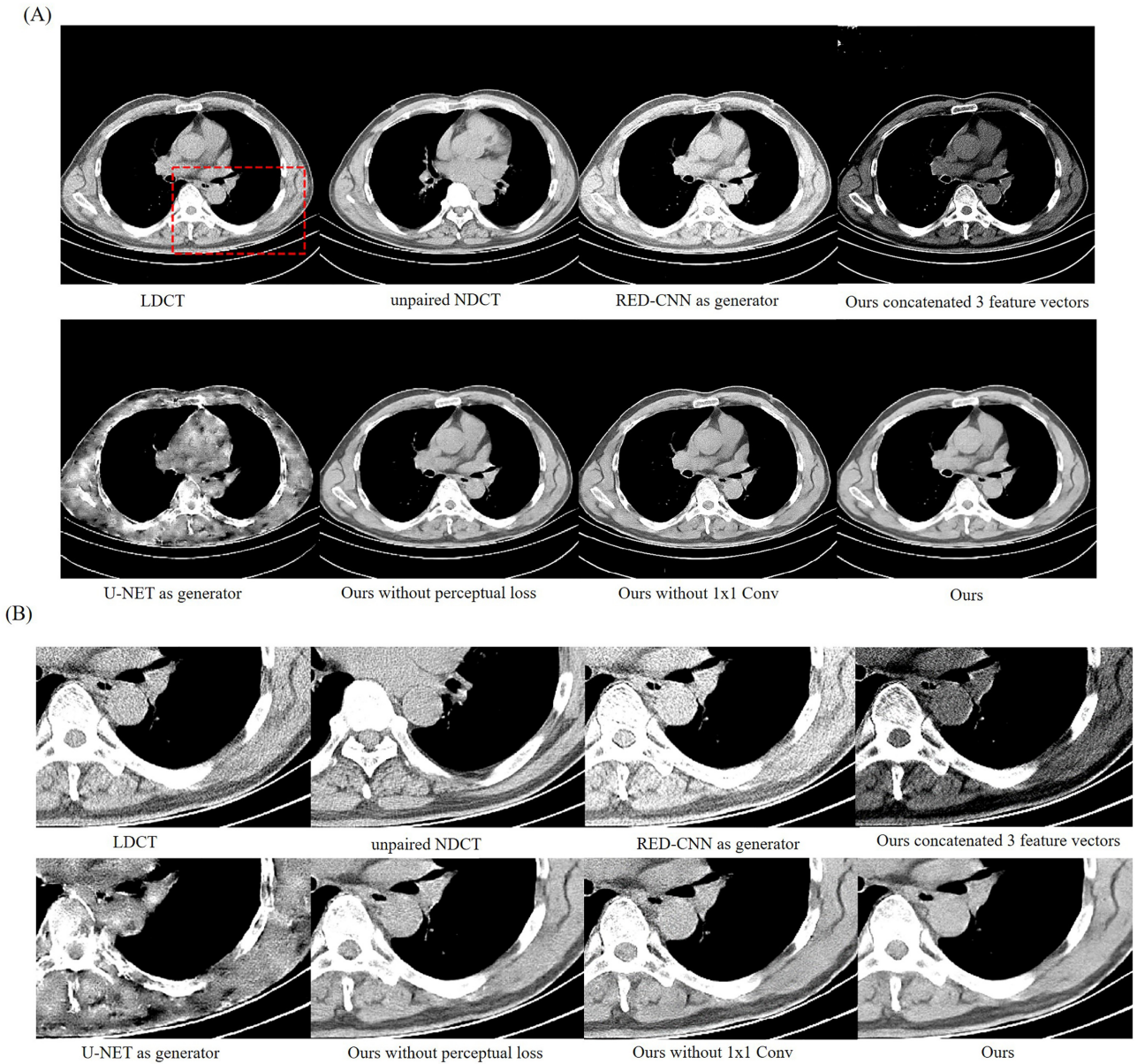


Figure 6. Experimental results and the magnified ROI of the different models. The four different models were compared, including a generator with only RED-CNN, a generator with only U-NET, and a generator with concatenated three-feature vectors and without a 1×1 convolution block. A model without perceptual loss function was also examined to prove the availability of the added perceptual loss function. To ensure fairness, the other parameters remain unchanged in the comparison. The area indicated by a red rectangle is the ROI that was magnified to show the experimental results.

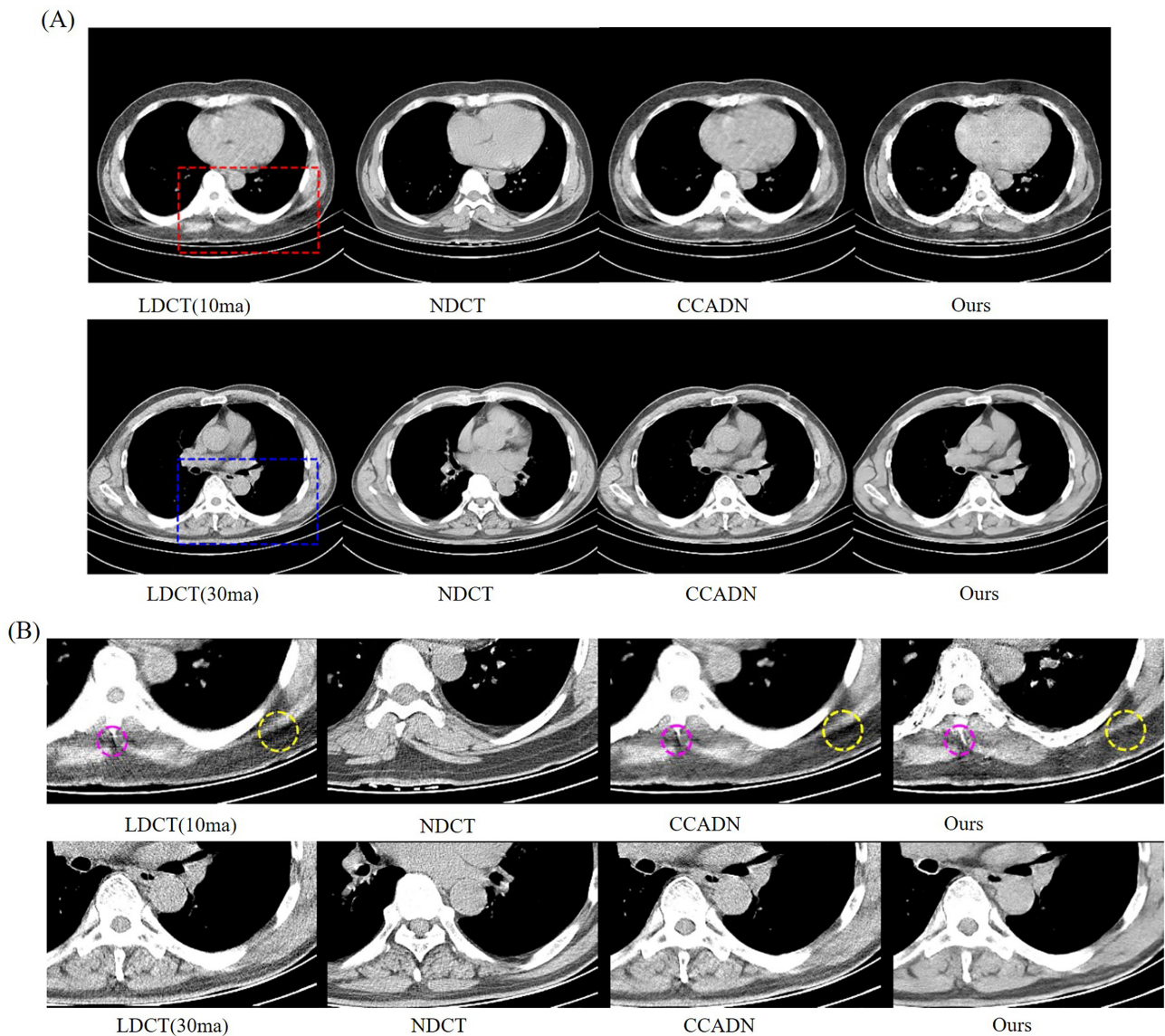


Figure 7. The experimental results and magnified parts of different 10 ma and 30 ma dose datasets for comparison. CCADN have been compared with ours in this paper. The red and blue rectangles indicate the magnified images of an ROI of 10 ma and 30 ma dose datasets, respectively. More structural details can be seen, as highlighted by the yellow and purple circles.

circles. Compared with 30 ma LDCT, 10 ma data was processed with noise reduction, and although lost features were restored to a certain extent, there was still detail loss. For 30 ma LDCT data, CT images with completely detailed information is obtained after denoising processing. In clinical diagnosis, excessive loss of detailed information can lead to changes of results. Hence, 30 ma CT images were selected as the main experimental dataset to evaluate the proposed model. The proposed model could effectively reduce the harm caused by CT detection without affecting the quality of diagnosis.

Conclusion

Considering the shortage of well-paired CT images, inspired by RED-CNN and U-Net, we propose a novel unsupervised learning model based on CycleGAN for LDCT image denoising. An adaptive feature selection generator is designed, and a patchGAN discriminator is used to generate CT images maintaining more detail, that is aided by added perceptual loss. Compared with tradi-

tional methods and other unsupervised learning algorithms, the experimental results have confirmed that our proposed model is superior with both a clinical dataset and the Mayo dataset. The main advantages of our method lie in noise suppression and edge preservation.

Supplementary data

Supplementary codes are available at <https://github.com/tcq1122/SKFCycleGAN.git>. Other supplementary data is available at *PCME DI Journal* online.

Acknowledgements

This study was funded by the National Natural Science Foundation of China (Grants No. 61871277 and 61671312), and in part by the Project of State Administration of Traditional Chinese Medicine of Sichuan (Grant No. 2021MS012).

Conflict of 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.

References

- Jung KJ, Lee KS, Kim SY, et al. Low-dose, volumetric helical CT: image quality, radiation dose, and usefulness for evaluation of bronchiectasis. *Invest Radiol* 2000;**35**:557–63. doi: 10.1097/00004424-200009000-00007.
- Gartenschläger M, Schweden F, Gast K, et al. Pulmonary nodules: detection with low-dose vs conventional-dose spiral CT. *Eur Radiol* 1998;**8**:609–14. doi: 10.1007/s003300050445.
- Ismael AM, Engür AS. Deep learning approaches for COVID-19 detection based on chest X-ray images. *Expert Syst Appl* 2021;**164**:114054. doi: 10.1016/j.eswa.2020.114054.
- Balda M, Hornegger J, Heismann B. Ray contribution masks for structure adaptive sinogram filtering. *IEEE Trans Med Imaging* 2011;**30**:1116–28. doi: 10.1109/TMI.2012.2187213.
- Naidich DP, Marshall CH, Gribbin C, et al. Low-dose CT of the lungs: preliminary observations. *Radiology* 1990;**175**:729–31. doi: 10.1148/radiology.175.3.2343122.
- Manduca A, Yu L, Trzasko JD. Projection space denoising with bilateral filtering and CT noise modeling for dose reduction in CT. *Med Phys* 2009;**36**:4911–9. doi: 10.1118/1.3232004.
- Zhang Y, Zhang J, Lu H. Statistical sinogram smoothing for low-dose CT with segmentation-based adaptive filtering. *IEEE Trans Nucl Sci* 2010;**57**:2587–98. doi: 10.1109/TNS.2010.2060356.
- Wang J, Li T, Xing L. Iterative image reconstruction for CBCT using edge-preserving prior. *Med Phys* 2009;**36**:252–60. doi: 10.1118/1.3036112.
- Sidky EY, Kao CM, Pan X. Accurate image reconstruction from few-views and limited-angle data in divergent-beam CT. *J X-ray Sci Technol* 2006;**14**:119–39. doi: 10.48550/arXiv.0904.4495.
- Chen Y, Yang Z, Hu Y, et al. Thoracic low-dose CT image processing using an artifact suppressed large-scale nonlocal means. *Phys Med Biol* 2012;**57**:2667–88. doi: 10.1088/0031-9155/57/9/2667.
- Li Z, Yu L, Trzasko JD, et al. Adaptive nonlocal means filtering based on local noise level for CT denoising. *Med Phys* 2014;**41**:011908. doi: 10.1118/1.4851635.
- Yu H, Wang G. Compressed sensing based interior tomography. *Phys Med Biol* 2009;**54**:2791–805. doi: 10.1088/0031-9155/54/9/014.
- Chen M, Pu YF, Bai YC. Low-dose CT image denoising using residual convolutional network with fractional TV loss. *Neurocomputing* 2021;**452**:510–20. doi: 10.1016/j.neucom.2020.10.004.
- Sidky EY, Pan X. Image reconstruction in circular cone-beam computed tomography by constrained, total-variation minimization. *Phys Med Biol* 2008;**53**:4777–807. doi: 10.1088/0031-9155/53/17/021.
- Xu Q, Yu H, Mou X, et al. Low-dose x-ray CT reconstruction via dictionary learning. *IEEE Trans Med Imaging* 2012;**31**:1682–97. doi: 10.1109/TMI.2012.2195669.
- Zhang Y, Mou X, Wang G, et al. Tensor-based dictionary learning for spectral CT reconstruction. *IEEE Trans Med Imaging* 2017;**36**:142–54. doi: 10.1109/TMI.2016.2600249.
- Cai JF, Jia X, Gao H, et al. Cine cone beam CT reconstruction using low-rank matrix factorization: algorithm and a proof-of-principle study. *IEEE Trans Med Imaging* 2014;**33**:1581–91. doi: 10.1109/TMI.2014.2319055.
- Goodfellow I, Bengio Y, Courville A. *Deep Learning*. Cambridge, MA: The MIT Press, 2016.
- LeCun Y, Bengio Y, Hinton G. Deep learning. *Nature* 2015;**521**:436–44. doi: 10.1038/nature14539.
- Zoran D, Chrzanowski M, Huang PS et al. Towards robust image classification using sequential attention models. 2019 *IEEE Conference on Computer Vision and Pattern Recognition*. 2019; 9483–92.
- Wang K, Peng X, Yang J, et al. Suppressing uncertainties for large-scale facial expression recognition. 2020 *IEEE Conference on Computer Vision and Pattern Recognition*. 2020; 6897–906.
- Kang E, Min J, Ye JC. A deep convolutional neural network using directional wavelets for low-dose X-ray CT reconstruction. *Med Phys* 2017;**44**:e360–75. doi: 10.1002/mp.12344.
- Chen H, Zhang Y, Kalra MK, et al. Low-dose CT with a residual encoder-decoder convolutional neural network. *IEEE Trans Med Imaging* 2017;**36**:2524–35. doi: 10.1109/TMI.2017.2715284.
- Yi X, Babyn P. Sharpness-aware low-dose CT denoising using conditional generative adversarial network. *J Digit Imaging* 2018;**31**:655–69. doi: 10.1007/s10278-018-0056-0.
- Wolterink JM, Leiner T, Viergever MA. Generative Adversarial Networks for Noise Reduction in Low-Dose CT. *IEEE Trans Med Imaging* 2017;**36**:2536–45. doi: 10.1109/TMI.2017.2708987.
- Yang Q, Yan P, Zhang Y, et al. Low dose CT image denoising using a generative adversarial network with Wasserstein distance and perceptual loss. *IEEE Trans Med Imaging* 2018;**37**:1348–57. doi: 10.1109/TMI.2018.2827462.
- Du W, Chen H, Liao P, et al. Visual attention network for low-dose CT. *IEEE Signal Proc Let* 2019;**26**:1152–6. doi: 10.1109/LSP.2019.2922851.
- You C, Li G, Zhang Y, et al. CT super-resolution GAN constrained by the identical, residual, and cycle learning ensemble (GAN-CIRCLE). *IEEE Trans Med Imaging* 2019;**39**:188–203. doi: 10.1109/TMI.2019.2922960.
- Tang C, Li J, Wang L, et al. Unpaired low-dose CT denoising network based on cycle-consistent generative adversarial network with prior image information. *Comput Math Method M* 2019;**12**:1–11. doi: 10.1155/2019/8639825.
- Kang E, Koo HJ, Yang DH, et al. Cycle-consistent adversarial denoising network for multiphase coronary CT angiography. *Med Phys* 2019;**46**:550–62. doi: 10.1002/mp.13284.
- Li ZH, Zhou SW, Huang JZ, et al. Investigation of low-dose CT image denoising using unpaired deep learning methods. *IEEE Trans Plasma Sci* 2021;**5**:224–34. doi: 10.1109/trpms.2020.3007583.
- Huang Z, Chen Z, Zhang Q, et al. CaGAN: a cycle-consistent generative adversarial network with attention for low-dose CT imaging. *IEEE Trans Comput Imag* 2020;**6**:1203–18. doi: 10.1109/TCI.2020.3012928.
- Zhu JY, Park T, Isola P, et al. Unpaired image-to-image translation using cycle-consistent adversarial networks. 2016 *IEEE Conference on International Conference on Computer Vision*. 2017.
- Ronneberger O, Fischer P, Brox T. U-net: Convolutional networks for biomedical image segmentation. *International Conference on Medical Image Computing and Computer-Assisted Intervention* 2015. 2015; 234–41.
- Li X, Wang W, Hu XL, et al. Selective kernel networks. 2018 *IEEE Conference on Computer Vision and Pattern Recognition*. 2018; 510–9.
- Krizhevsky A, Sutskever I, Hinton GE. ImageNet classification with deep convolutional neural networks. *Comm ACM* 2017;**60**: 84–90. <https://doi.org/10.1145/3065386>
- Gatys LA, Ecker AS, Bethge M. Image style transfer using convolutional neural networks. 2016 *IEEE Conference on Computer Vision and Pattern Recognition* 2016; 2414–23.

38. Chen H, Zhang Y, Zhang W, et al. Low-dose CT via convolutional neural network. *Biomed Opt Express* 2017;**8**:679–94. doi: 10.1364/BOE.8.000679.
39. Kingma DP, Ba J. Adam: A method for stochastic optimization. *arXiv* 2014.
40. Feruglio PF, Vinegoni C, Gros J, et al. Block matching 3D random noise filtering for absorption optical projection tomography. *Phys Med Boil* 2010;**55**:5401–15. doi: 10.1088/0031-9155/55/18/009.
41. Aharon M, Elad M, Bruckstein A. K-SVD: An algorithm for designing over complete dictionaries for sparse representation. *IEEE Trans signal Proces* 2006;**54**:4311–22. doi: 10.1109/TSP.2006.881199.
42. Fan F, Shan H, Kalra MK, et al. Quadratic autoencoder (Q-AE) for low-dose CT denoising. *IEEE Trans Med Imaging* 2019;**39**:2035–50. doi: 10.1109/TMI.2019.2963248.