J. Meas. Sci. Instrum., 2024, 15(4): 490-498.
DOI: 10.62756/jmsi.1674-8042.2024049
http://jmsi.publish.founderss.cn
jmsi@nuc.edu.cn

Low-dose CT image denoising method based on generative adversarial network

JIAO Fengyuan[1,2*], YANG Zhixiu[1], SHI Shaojie[1], CAO Weiguo[1,3]

1.Shanxi Provincial Key Laboratory for Biomedical Imaging and Big Data, North University of China, Taiyuan 030051, China;
2.School of Information and Communication Engineering, North University of China, Taiyuan 030051, China;
3. School of Environment and Safety Engineering, North University of China, Taiyuan 030051, China

***Corresponding author:** JIAO Fengyuan (jfengyuan@163.com).

Received: December 27, 2022 **Revised:** May 20, 2023 **Accepted:** June 11, 2023

Abstract: In order to solve the problems of artifacts and noise in low-dose computed tomography (CT) images in clinical medical diagnosis, an improved image denoising algorithm under the architecture of generative adversarial network (GAN) was proposed. First, a noise model based on style GAN2 was constructed to estimate the real noise distribution, and the noise information similar to the real noise distribution was generated as the experimental noise data set. Then, a network model with encoder-decoder architecture as the core based on GAN idea was constructed, and the network model was trained with the generated noise data set until it reached the optimal value. Finally, the noise and artifacts in low-dose CT images could be removed by inputting low-dose CT images into the denoising network. The experimental results showed that the constructed network model based on GAN architecture improved the utilization rate of noise feature information and the stability of network training, removed image noise and artifacts, and reconstructed image with rich texture and realistic visual effect.

Key words: low-dose CT image; generative adversarial network; noise and artifacts; encoder-decoder; atrous spatial pyramid pooling (ASPP)

0 Introduction

Computed tomography (CT) technology uses X-ray fluoroscopy for human imaging to assist doctors in diagnosis and treatment of diseases. It has the characteristics of fast scanning and clear images. It is widely used in clinical medicine[1]. CT imaging quality is closely related to X-ray radiation dose. Increasing the radiation dose can improve the image quality, but too much radiation dose is harmful to the human body. Therefore, CT technology research has mainly focused on how to obtain high-quality images under low radiation dose conditions. Starting with ray sources, detectors, algorithms, and other aspects[2], researchers carried out active explorations. Compared with the long research period and high cost of ray sources and detectors[3,4], CT imaging algorithms are low in cost and easy to implement, so they have always been one of the hot spots.

The post-processing technique for low-dose CT image is a key means to improve the image quality, and researchers have done a lot of theoretical research and practical exploration. Yu et al.[5-9] proposed several nonlinear noise filters with easy algorithms, nonlinear distribution, and large parameter space. Wang et al.[10-16] proposed iterative denoising algorithms based on statistics, which were suitable for sampling data in different ways, and reconstructed images for incomplete data, but the amount of calculation was large and the reconstruction speed was slow. Elbakri et al.[17] believed that the number of photons detected approximately obeyed the Poisson distribution where the background noise was Gaussian noise, and proposed a penalty maximum likelihood estimation algorithm for denoising. In addition, compressed sensing theory has gradually been applied in CT imaging technology. Chen et al.[18,19] used compressed sensing theory to establish a low-noise dictionary, which effectively improved the signal-to-noise ratio of the image, but the edges of the organs were blurred and the structural features were excessively smooth.

Deep learning technology has been widely used for CT imaging. The super-resolution imaging method based on

convolutional neural network (CNN) greatly improves the image quality. In order to remove additive Gaussian noise, Zhang et al. [20] proposed a network structure, which combined residual network with batch normalization (BN) layer to improve the imaging quality of natural images. Han et al.[21] used residual theory to construct U-shaped network structure to predict the strip artifacts in CT images by sparse-view construction. Ledig et al. [22,23] proposed an image denoising algorithm different from the traditional deep convolution network. Firstly, a residual network based on generative adversarial network (GAN) was constructed, and then a special loss function for network training was constructed. As a result, the algorithm not only enhanced the content and color of the original image, but also removed some unnecessary noises and retained more details.

An optimized GAN-based network structure with encoder-decoder was proposed as the core, and it was composed of convolution layer (Conv), MobileNet V3, and activation layer (ReLU). The lightweight encoder-decoder architecture can effectively extract image features and improve the stability of network training process and the quality of image. The discriminant network consists of multiple convolution layers and deconvolution layers, and the multi-scale feature fusion layer is used to connect them. Thus the abstract information of the image was capatured to determine whether the input image contained a lot of noise. Noisy and noiseless images were input respectively at the output end of the model, and the model used convolutional autoencoder for image denoising deep learning.

1 Proposed methodology

1.1 GAN model

GAN is a dual-loop generative network model[24], which is mainly composed of two parts: generator (G) and discriminator (D). The model structure is shown in Fig.1.

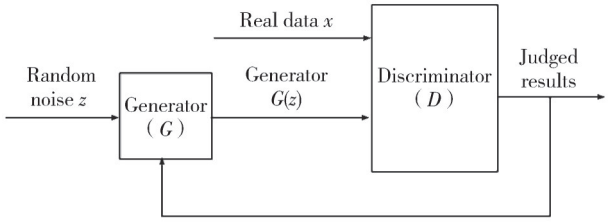

Fig. 1 GAN model

The idea of GAN originates from the Nash equilibrium in game theory[25], that is, the G receives random noise and generates fake sample data $G(z)$, and the D distinguishes between real data x and fake sample data $G(z)$ and feeds the discrimination results back to the generator. The G and D continuously optimize their network capabilities through iterative confrontation training until D cannot distinguish between the fake sample data $G(z)$ and the real data x. The objective function is expressed as

$$\min_G \max_D V(D,G) = E_{x\sim P_{data}(x)}\left[\log D(x)\right] + E_{z\sim P(z)}\left[\log\left(1-D\left(G(z)\right)\right)\right], \quad (1)$$

where $P_{\text{data}}(x)$ represents the distribution of real samples; $P(z)$ represents the distribution of low-dimensional noise; $D(x)$ represents the probability that real data are recognized as true; $D(G(z))$ represents the probability that the generated data are recognized as true; $E_{x\sim P_{\text{data}}}[\log D(x)]$ represents the expectation of true data in D, and $E_{z\sim P(z)}[\log(1\text{-}D(G(z))]$ represents the expectation of generated data in D. When GAN trains an adversarial network, it repairs the other while training one, and alternates training.

1.2 Noise model

The deep learning model can learn the prior knowledge of noise by training, so as to improve the noise processing ability. However, training data are usually limited, while the types of noise are unlimited[26]. Generally, the model has a good processing effect on the noise that has been learned in the training set, but has a poor processing effect on the noise that has not been learned in the training set, which indicates that the generalization ability of the model is inadequate. The existing CNN often simplifies noise model to additive Gaussian noise, but in practical applications, the noise distribution in low-dose CT images is unknown. Therefore, the denoising network trained by artificial additive Gaussian noise as a data set cannot guarantee the denoising effect of real noisy images. In response to this problem, the proposed algorithm used the real noise data set to train the GAN to estimate the real noise distribution, and then generated noise information that was more similar to the real noise distribution.

1.3 MobileNet V3 model

MobileNet network is a lightweight deep neural network model constructed based on deep separable convolution structure with encoder-decoder architecture[27]. It uses the lightweight squeeze-and-excitation (SE) module to perform attention extraction of the spatial information on the corresponding channel of the input feature, so that the network can learn the important information in the feature,

and then aggregate the features by point-by-point addition. In this way, the SE module is used to assign different weights to different parts of the image. So that the edge details in the image noise reduction can be maintained well for the local information of the shallow network. And for the feature map output by the deep network, encoder can capture the information of different scales of the same image to enhance the network's receptive field and spatial details. The MobileNet V3 architecture is shown in Fig.2.

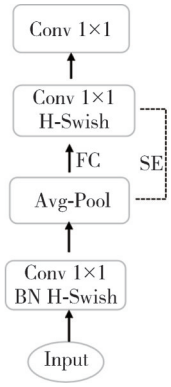

Fig. 2 MobileNet V3 architecture

In order to reduce the time consumption caused by the SE module, the channel of the expansion layer becomes 1/4 of the original in all the structures containing the SE module, thus the accuracy of the model can be effectively improved and the possibility of information loss after training is reduced without increasing the time consumption. Since the calculation of sigmoid function takes a long time, $\frac{ReLU6(x+3)}{6}$ is used to approximately replace the sigmoid function to reduce the requirements on the performance of computing equipment. In the quantization process, replacing *h-Swish* with *Swish* will not bring about a large loss of precision, and the efficiency can be increased by about 15%. The calculation equations are expressed as

$$Swish(x)=x\sigma(x), \tag{2}$$

$$h\text{-}Swish(x)=x\frac{ReLU6(x+3)}{6}. \tag{3}$$

1.4 Atrous spatial pyramid pooling module

The atrous spatial pyramid pooling (ASPP) module uses atrous convolutions with four sampling rates (1, 6, 12, 18) in parallel on the top feature map to extract multi-scale semantic information from the feature map of the backbone network, as shown in Fig.3.

After the network feature of the input image is extracted, its multi-scale context feature is extracted by the ASPP. ASPP fills $r-1$ zeros between sampling points by setting different atrous rates r, which expands receptive field from $K\times K$ to $S\times S$.

$$S=K+(K-1)(r-1), \tag{4}$$

$$F_{out}(i)=\sum_{K}F_{in}(i+rK)W(K), \tag{5}$$

where F_{in} is the input feature; K is the filter size; r is the atrous rate; i is the position on the input feature; and W is the filter; F_{out} is the output feature.

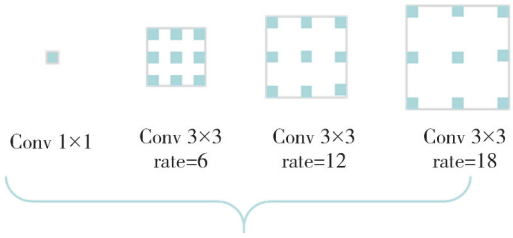

Fig. 3 ASPP module

In the multi-scale feature fusion module, the contact operation is expressed as

$$J=C\left[c(x_1),c(x_2),\cdots,c(x_3),\cdots,c(x_n)\right] \tag{6}$$

where x_i is the feature map of the ith layer; $c(\cdot)$ is the convolution operation, and $C[\cdot]$ is the contact operation.

ASPP performs multi-scale feature extraction and fusion of the image by learning the information to be emphasized or suppressed, which can effectively extract the spatial information of the image, enhance the information flow in the network, expand the receptive field, reduce the computational complexity, and greatly improve decoder performance during feature fusion.

2 Network model and algorithm

2.1 Noise model

2.1.1 Noisy block extraction

Noisy block extraction is one of the important steps to correctly train GAN to simulate unknown noise. Because the noise distribution can be better estimated from the noise-dominated data. To reduce the influence of the original background, a set of approximate noisy blocks are extracted first from the weaker background of the given noisy image. In this way, the network model is more accurate by learning the noise distribution. The noisy point is determined by the difference of the mean and the variance between p_i and q_j^i of each j, and the two constraints are defined as

$$\left|Mean(q_j^i)-Mean(p_i)\right|\leqslant\mu Mean(p_i), \tag{7}$$

$$\left|Var(q_j^i)-Var(p_i)\right|\leqslant\gamma Var(p_i), \tag{8}$$

where $Mean(\cdot)$ and $Var(\cdot)$ represent mean and variance, respectively. When the algorithm is applied to all noisy images, $S=\{S_1, S_2, \cdots, S_t\}$ can be obtained,

and approximate noisy block set $V=\{V_1, V_2, \cdots, V_i\}$ can be derived from $V_i = S_i - Mean(i)$.

2.1.2 Noise modeling

Due to limited pattern and the number of noisy blocks extracted, the training results of deep GAN are not satisfactory. In order to make the noise closer to the noise in low-dose CT images, StyleGAN2 is introduced to model the noise from the extracted noisy block set V, so as to improve the training results of GAN and generate more high-quality and realistic noise data. The model structure is shown in Fig.4.

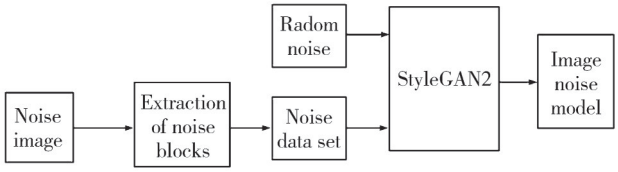

Fig. 4 Noise modeling process

The image noise modeling process based on StyleGAN2 network is described as follows.

Step 1: Obtain a noise map. The image noise is extracted from the noisy image to obtain a noise map.

Step 2: Extract noisy blocks. Extract multiple sets of similar noisy blocks from the weaker background of the noisy image, remove structural noise and single image noise, and form a noise data set composed of high-quality noisy blocks.

Step 3: StyleGAN2 training. Use the noise data set obtained in step 2 to train StyleGAN2 network to obtain the parameters with a good fit of this data set. In this way, the noise model based on StyleGAN2 is more dependent on noise data and is not representative.

Step 4: Establish an image noise model. The noise data and random noise obtained in step 3 are sent to the StyleGAN2 for further training to obtain the required image noise model.

StyleGAN2 controls the overall feature of the image without changing the feature of key information of the image, and random noise provides the local diversity of the image. The network has the ability to learn complex distributions. GAN can be trained through the back propagation algorithm and noise samples can be generated through forward propagation. Furthermore, the weight of StyleGAN2 is adjusted by

$$w'_{ijk} = s_i w_{ijk}, \tag{9}$$

$$\sigma_j = \sqrt{\sum_{i,k} w'^{\,2}_{ijk}}, \tag{10}$$

$$w''_{ijk} = \frac{w'_{ijk}}{\sqrt{\sum w'^{\,2}_{ijk} + \varepsilon}}, \tag{11}$$

where i indicates the ith input feature map; w_{ijk} is convolution weight; s_i is scaling coefficient, and ε means to increase the training volume by 40% to avoid numerical instability. After AdaIN normalization, mottle artifacts like water droplets can be removed. The StyleGAN2 makes the noise distribution easier, and a relatively complex process can be automatically fitted through deep learning process. The trained neural network model can generate a larger set of noise samples $V'=\{V_1', V_2', \cdots, V_w'\}$.

When the noise information of the image is unknown, the noise distribution model of noisy image obtained with this method can reduce the difficulty of constructing training data set and effectively enhance the generalization ability of noise model.

2.2 Generator

The generator network consists of Conv, MobileNet V3, ReLU, and ASPP modules. The size of convolution kernel of Conv layer is 3×3. When a module extracts the feature of the image, the image size is reduced after a downsampling. MobileNet V3 module learns the weight of feature mapping on different channels by channel attention mechanism of the feature, which highlights the important feature mapping and ignores the unimportant feature mapping, so as to extract effective features and reduce the amount of network computing and the number of parameters. Since over downsampling will seriously damage the spatial information of the image, the ASPP module for multi-scale feature extraction can reduce spatial loss.

2.3 Discriminator

The discriminator network is composed of Conv, ReLU, and Deconv, which is connected by multi-scale feature fusion layer. When upsampling, in addition to the output of the previous module, the output feature corresponding to its size in the downsampling module is also received. The Concat function fuses the detail information of the low-level network and the position information of the high-level network, which can not only ensure the network to learn sufficient feature, but also gradually restore the feature to the original image. As a result, the details and spatial information of high-level feature can be efficiently restored.

2.4 GAN-based denoising model with encoder-decoder architecture

To effectively reduce the noise and artifacts in low-dose CT images, a GAN-based denoising model with encoder-decoder architecture is designed, as shown in Fig.5.

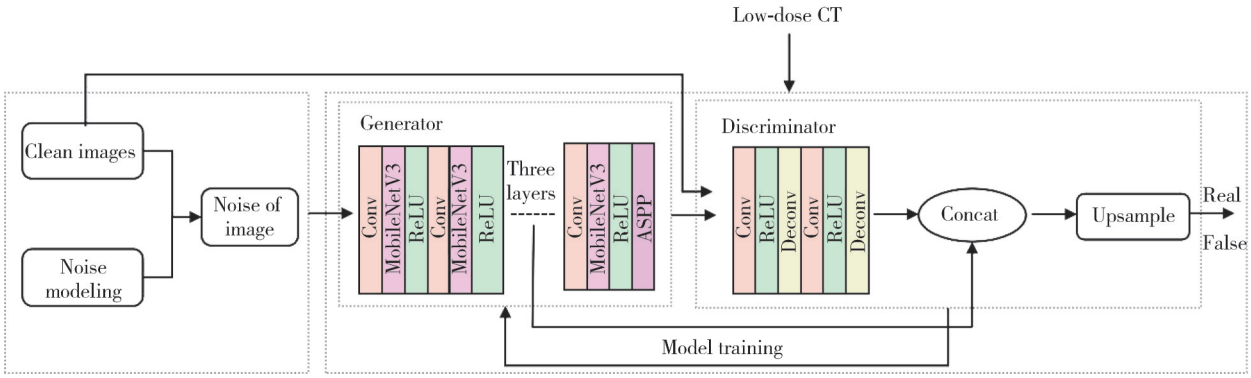

Fig. 5 GAN-based denoising model for low-dose CT image

Firstly, the noise distribution of a noisy image was learned based on StyleGAN2, and the training data set of noisy image was constructed. Thus the noisy image closer to the real noise of low-dose CT image was obtained. Then, a network model with encoder-decoder architecture based on GAN was designed. The generator adopted a lightweight MobileNet V3 and ASPP module with the size of 3×3 and the atrous rates of 1, 6, 12 and 18 to extract multi-scale feature. It ensured the continuous memory function of the network by means of dense connection of convolution layer and activation layer of deep network, thus the previous and next feature information of the network could be reused to improve the utilization rate of feature information and image quality. After the discriminator received feature expression from the generator, the deconvolution layer gradually restored the details and spatial dimensions of the noiy image. The feature map of shallow network is local information, and the feature map of deep network is global information. By extracting the feature maps generated at different levels for feature fusion, the feature map with rich semantic information can be obtained. After that, the noisy image was input to train the model until the optimal network model was obtained. Although the noise was eliminated step by step in case of full convolution following the lost details of the image, the main information of the image could be retained owing to discriminator. Finally, the low-dose CT noisy image was input, and a good denoising effect was achieved.

3 Experiment

The hardware parameters included IntelCorei5-10400F 4.00 GHz and NVIDIA TITAN RTX 24 G. The data set used 35 000 images in the self-built data set. The verification set and the test set included 6 000 images, respectively. The self-built data set was made up of real CT data taken by the hospital, and the images contained at least 2 040 pixels vertically or horizontally.

The proposed algorithm was used to train and verify the self-built dataset, and the experimental results were compared with those of SRCNN under nearest and deep learning.

The weighted sum of mean square error (MSE) loss and confrontation loss was used as the objective function of training generator network. The noise reduction effect was evaluated by structural similarity (SSIM, in the range of 0—1) and peak signal-to-noise ratio (PSNR with unit of dB).

The pixel-based MSE loss is used to preserve the low frequency part of the image, and the anti-loss is used to recover the high frequency information of the image. The MSE loss function is expressed as

$$L_{MSE(\theta)}=\frac{1}{N}\sum_{1}^{N}\left\| I_i^H-G\left(I_i^L,\theta\right)\right\|^2, \tag{12}$$

where I_i^{H} represents the high-resolution image; I_i^{L} represents the low-resolution image, and $G\left(I_i^L,\theta\right)$ represents the mapping function between the low-resolution image and the high-resolution image that the generation network learns.

Based on the confrontation mechanism of the generation network and the discriminator network, the confrontation loss is expressed as

$$L_A(\theta)=\frac{1}{N}\sum_{1}^{N}-\lg D\left(G\left(I_i^L,\theta\right)\right), \tag{13}$$

where $D(G(I_i^{\mathrm{L}},\ \theta))$ is the probability that the discriminator network determines that the generated image is a real image.

SSIM and PSNR reflect the error between the corresponding pixels of the two images. The higher the value, the less the distortion of the output image, and the better the image reconstruction quality. SSIM represents the similarity of two images. The closer the value is to 1, the closer the output image is to the original high resolution image, that is, the better the reconstruction effect. The

larger the values of PSNR and SSIM are, the better the effect is. The image effect is often judged by the values of the two. The PSNR and SSIM are expressed as

$$PSNR = 10\lg\left(\frac{(2^n-1)^2}{MSE}\right), \tag{14}$$

$$SSIM = \frac{(2\mu_x\mu_y + C_1)(2\sigma_{xy} + C_2)}{(\mu_x^2 + \mu_y^2 + C_1)(\sigma_x^2 + \sigma_y^2 + C_1)}, \tag{15}$$

where μ_x and σ_x represent the gray mean and variance of the original image; μ_y and σ_y represent the gray mean and variance of the reconstructed image; σ_{xy} is the covariance between the original image and the reconstructed image, and C_1 and C_2 are constants.

4 Results and discussion

The model is trained, and the training effects are shown in Figs.6 and 7. It can be seen from Fig.6 that before the discriminator has been trained 60 000 times, the loss curve fluctuates relatively largely. After 60 000 times, the curve becomes steady. After the network iteration is about 140 000 times, the loss value floats in a very small range of 0.03, and the model enters steady training state. It can be seen from Fig.7 that the discriminator does not work in the initial stage and its performance is very poor. After 140 000 trainings, the performance is getting better and better. In the end, the generator and the discriminator confront with each other and a good balance is reached.

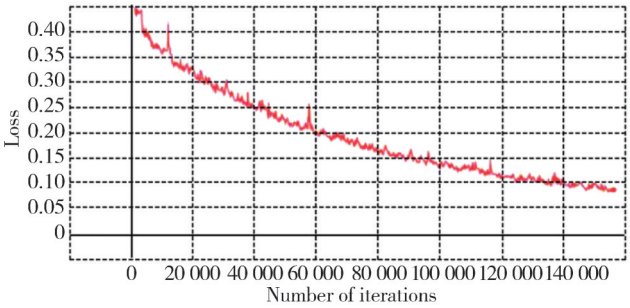

Fig. 6 Generator loss curve after being trained

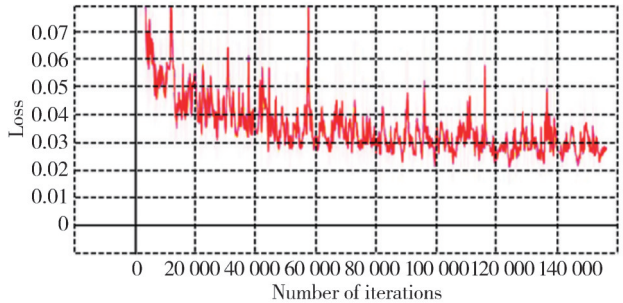

Fig. 7. Discriminator loss curve after being trained

The abdominal, pectoral, and pelvic CT images from the self-built data set are selected, and the denoising results of nearest, SRCNN, and the proposed method are compared, as shown in Figs.8—10. It can be seen that the overall quality of the obtained image using the proposed method is better than that using the other two methods, especially the textures of the image detials are clearer.

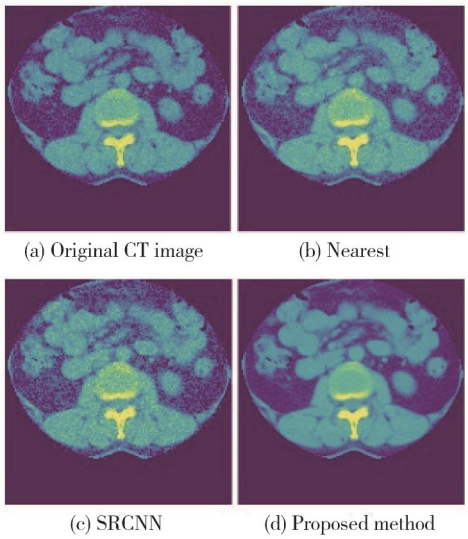

Fig. 8 Results of different denoising methods on abdominal CT images

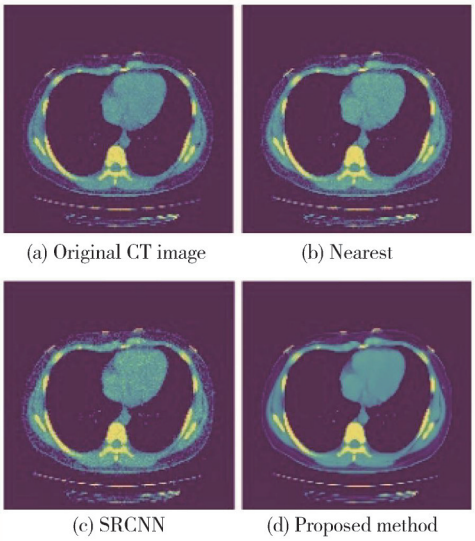

Fig. 9 Results of different denoising methods on pectoral CT images

Figs.8—10 respectively represent the original image in the self-built dataset, the image after nearest noise reduction, the image after SRCNN noise reduction, and the image after noise reduction using the proposed method. It can be seen that the texture of the detail part of the image after noise reduction using the proposed method is clearer, and the overall quality is better than that of the other two methods. Fig. 11 shows enlarged views of reconstucted images using proposed denoising method.

It can be seen that the details of the images are true and the local organs are smooth by the proposed

denoising method. In the marked local area, the contour of the image is relatively even, the splicing effect of the organ surface is good, and no artifacts are formed on the surface with varying depth. In addition, the denoising method can realistically reflect the convex and concave characteristics of the real organ surface, and has good robustness to low-dose CT images. In Fig. 11(b), the denoised edges can well restore the original boundary features of the organ, and the geometry of the organ edge is more obvious.

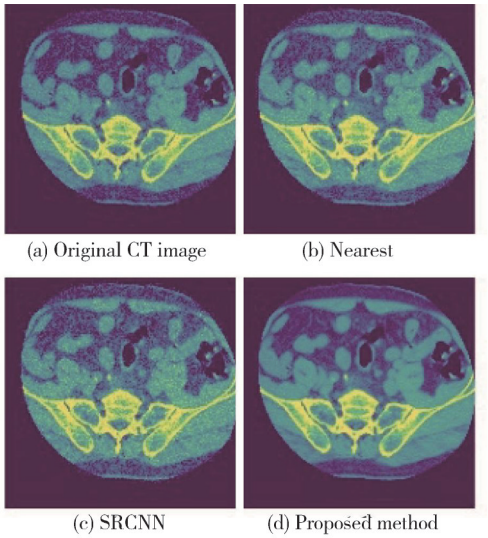

Fig. 10 Results of different denoising methods on pelvic CT images

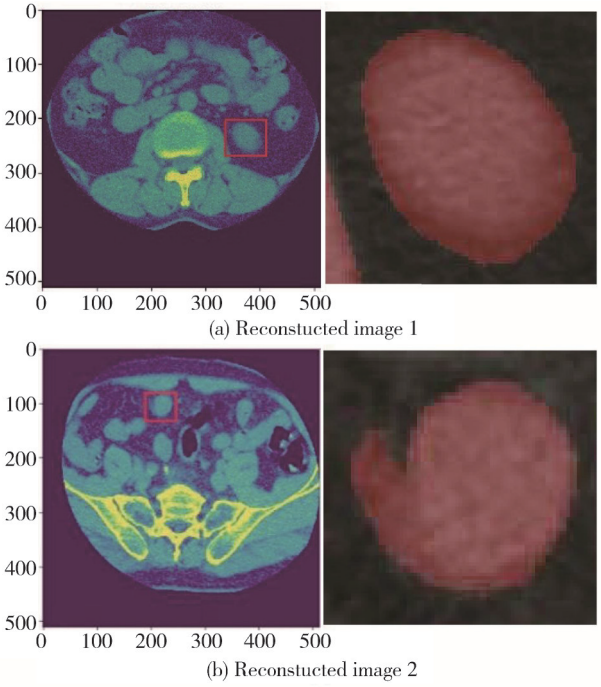

Fig. 11 Enlarged views of reconstucted images using proposed denoising method

The PSNR and SSIM values between the real image and the reconstructed image are calculated by generator under the optimized image reconstruction network, as well as the average values of PSNR and SSIM under each test set, and the objective evaluation results are obtained under different denoising methods, as shown in Table 1.

Table 1 Comparison of evaluation index results of different denoising methods

Method	PSNR/dB	SSIM
Nearest	24.607	0.700
SRCNN	28.089	0.814
Proposed	28.653	0.816

It can be seen that the values of PSNR and SSIM of nearest, SRCNN, and proposed denoising method show an increasing trend, and the values of PSNR and SSIM of the proposed method are the highest, indicating that the network structure designed can make the network training more stable and the algorithm is more superior.

5 Conclusions

A denoising method based on GAN was proposed for strip artifacts and noise in low-dose CT images. The encoder-decoder model based on GAN was constructed to effectively extract deep features, improve the utilization rate of feature information, reduce the possibility of information loss after training, improve the network accuracy, and reduce the calculation amount. The experimental results showed that the proposed method could effectively extract the features of low-dose CT images, better maintain the edge information of organs in CT images during convolution, fully recover the texture information of images, and improve the quality of image reconstruction while accelerating the convergence of the model. Future research mainly focuses on denoising of specific areas or specific scene images by combining deep learning with traditional methods.

Acknowledgement

This work was supported by National Natural Science Foundation of China (No. 11802272), and China Postdoctoral Science Foundation (No.2019M651085).

Declaration of conflicting interests

The authors have no conflict of interests related to this publication.

References

[1] TANG Y, CUI X Y, ZHANG Q, et al. Denoising algorithm for low-dose CT projection based on partial

differential equation. Computer Application & Software, 2014, 31(3): 179-182.

[2] ZHANG H Y, TENG G J, CAO B, et al. Advances in X-ray imaging technology. Scientia Sinica Vitae, 2020, 50(11): 1202-1212.

[3] LI T F, LI X, YANG Y, et al. Simultaneous reduction of radiation dose and scatter for CBCT by using collimators. Medical Physics, 2013, 40(12): 121913.

[4] CHAI N, WANG S J, CHEN L J, et al. Low-dose micro-CT imaging method based on progressive network processing. Computerized Tomography Theory and Applications, 2020, 29(4): 435-446,.

[5] YU L F, MANDUCA A, TRZASKO J D, et al. Sinogram smoothing with bilateral filtering for low-dose CT//Medical Imaging 2008: Physics of Medical Imaging, March 18, 2008, San Diego, CA. Bellingham: SPIE, 2008: 768-775.

[6] DEMIRKAYA O. Reduction of noise and image artifacts in computed tomography by nonlinear filtration of the projection images//Medical Imaging 2001: Physics of Medical Imaging, February 16, 2001, San Diego, CA, USA. Bellingham: SPIE, 2001: 917-923.

[7] KACHELRIEß M, WATZKE O, KALENDER W A. Generalized multi-dimensional adaptive filtering for conventional and spiral single-slice, multi-slice, and cone-beam CT. Medical Physics, 2001, 28(4): 475-490.

[8] WANG J, LU H B, LI T F, et al. Sinogram noise reduction for low-dose CT by statistics-based onlinear filters//Medical Imaging 2005: Image Processing, February 13-15, 2005, San Diego, CA, USA. Bellingham: SPIE, 2005: 2058-2066.

[9] LIU Y, ZHANG Q, GUI Z G. Noise reduction for low-dose CT sinogram based on fuzzy entropy. Journal of Electronics & Information Technology, 2014, 35(6): 1421-1427.

[10] WANG J, LI T F, LU H B, et al. Penalized weighted least-squares approach to sinogram noise reduction and image reconstruction for low-dose X-ray computed tomography. IEEE Transactions on Medical Imaging, 2006, 25(10): 1272-1283.

[11] LI T F, LI X, WANG J. Nonlinear sinogram smoothing for low-dose X-Ray CT. IEEE Transactions on Nuclear Science, 2004, 51(5): 2505-2513.

[12] ZHANG W K, LI J S, SUN J Q, et al. FBP initialized few-view CT reconstruction algorithm using similar prior image constraint//38th Annual International Conference of the IEEE Engineering in Medicine and Biology Society, August 16-20, 2016, Orlando, FL, USA. New York: IEEE, 2016: 3949-3952.

[13] FANG L, LI L. Application progress of adaptive statistical iterative reconstruction in reducing radiation dose. CT Theory and Applications, 2013, 22(2): 207-213.

[14] SHIN H B, KIM M S, LAW M, et al. Application of sigmoidal optimization to reconstruct nuclear medicine image: Comparison with filtered back projection and iterative reconstruction method. Nuclear Engineering and Technology, 2021, 53(1): 258-265.

[15] ZHANG Y K, ZHANG J Y, LU H B. Statistical sinogram smoothing for low-dose CT with segmentation-based adaptive filtering. IEEE Transactions on Nuclear Science, 2010, 57(5): 2587-2598.

[16] ZHANG Q, GUI Z, CHEN Y, et al. Bayesian sinogram smoothing with an anisotropic diffusion weighted prior for low-dose X-ray computed tomography. Optik, 2013, 124(17): 2811-2816.

[17] ELBAKRI I A, FESSLER J A. Efficient and accurate likelihood for iterative image reconstruction in X-ray computed tomography//Medical Imaging 2003, June 20, 2003, San Diego, CA, USA. Bellingham: SPIE, 2003: 1839-1850.

[18] CHEN Y, YANG Z, HU Y, et al., Thoracic low-dose CT image processing using an artifact suppressed large-scale nonlocal means. Physics in Medicine and Biology, 2013, 57(9): 2667-2688.

[19] YIN J, FAN Y L, QIN S H. Research of CT reconstruction algorithms based on compressed sensing. China Medical Devices, 2019, 34(8): 166-169.

[20] ZHANG K, ZUO W M, CHEN Y J, et al. Beyond a Gaussian denoiser: residual learning of deep CNN for image denoising. IEEE Transactions on Image Processing: a Publication of the IEEE Signal Processing Society, 2017, 26(7): 3142-3155.

[21] HAN Y S, YOO J, YE J C. Deep residual learning for compressed sensing CT Reconstruction via persistent homology analysis. arXiv:1611.06391.

[22] LEDIG C, THEIS L, HUSZÁR F, et al. Photo-realistic single image super-resolution using a generative adversarial network//IEEE Conference on Computer Vision and Pattern Recognition, July 21-26, 2017, Honolulu, HI, USA. Piscataway: IEEE, 2017: 4681-4690.

[23] DONG C, LOY C C, HE K M, et al. Image super-resolution using deep convolutional networks. IEEE Transactions on Pattern Analysis and Machine Intelligence, 2016, 38(2): 295-307.

[24] WANG K F, GOU C, DUAN Y J, et al. Generative adversarial networks: the state of the art and beyond. Acta Automatica Sinica, 2017, 43(3): 321-332.

[25] WANG G M, QIAO J F, WANG L. A generative adversarial network based on energy function. Acta Automatica Sinica, 2018, 44(5): 28-38.

[26] CHEN J W, CHEN J W, CHAO H Y, et al. Image blind denoising with generative adversarial network based noise modeling//2018 IEEE/CVF Conference on Computer Vision and Pattern Recognition, June 18-23, 2018, Salt Lake City, UT, USA. Piscataway: IEEE, 2018: 3155-3164.

[27] HOWARD A, SANDLER M, CHEN B, et al. Searching for MobileNetV3//2019 IEEE/CVF International Conference on Computer Vision, October 27-November 2, 2019, Seoul, Korea (South). Piscataway: IEEE, 2019: 1314-1324.

基于生成对抗网络的低剂量CT图像降噪方法

焦枫媛[1,2*], 杨志秀[1], 石韶杰[1], 曹卫国[1,3]

1. 中北大学 生物医学成像与影像大数据山西省重点实验室, 山西 太原 030051;
2. 中北大学 信息与通信工程学院, 山西 太原 030051;
3. 中北大学 环境与安全工程学院, 山西 太原 030051

摘　要: 为解决临床医学诊断中低剂量计算机断层扫描(Computed tomography, CT)图像存在的伪影和噪声问题, 提出了一种改进的生成对抗网络(Generative adversarial network, GAN)架构下的图像降噪算法。首先, 构造了一个基于Style GAN2的噪声模型, 并通过训练该模型来估计真实噪声分布, 从而生成与真实噪声分布相似的噪声信息作为实验噪声数据集。然后, 构造了一个基于GAN思想的以编码器-解码器(Encoder-decoder)架构为核心的网络模型, 用生成的噪声数据集训练该网络模型直至达到最优。最后, 将低剂量CT图像输入该降噪网络, 可去除低剂量CT图像中的噪声和伪影。实验结果表明, 利用所构造的基于GAN架构的网络模型, 可提高噪声特征信息利用率和网络训练稳定性, 去除图像噪声和伪影, 且重建图像纹理丰富, 视觉效果逼真。

关键词: 低剂量CT图像; 生成对抗网络; 噪声和伪影; 编码器-解码器; 空洞金字塔池化

引用格式: JIAO Fengyuan, YANG Zhixiu, SHI Shaojie, et al. Low-dose CT image denoising method based on generative adversarial network. Journal of Measurement Science and Instrumentation, 2024, 15(4): 490-498.