

Low-Dose CT Image Denoising Based on Improved WGAN-gp

Xiaoli Li^{1,*}, Chao Ye¹, Yujia Yan² and Zhenlong Du¹

Abstract: In order to improve the quality of low-dose computational tomography (CT) images, the paper proposes an improved image denoising approach based on WGAN-gp with Wasserstein distance. For improving the training and the convergence efficiency, the given method introduces the gradient penalty term to WGAN network. The novel perceptual loss is introduced to make the texture information of the low-dose images sensitive to the diagnostician eye. The experimental results show that compared with the state-of-art methods, the time complexity is reduced, and the visual quality of low-dose CT images is significantly improved.

Keywords: WGAN-gp, low-dose CT image, image denoising, Wasserstein distance.

1 Introduction

Computational tomography (CT) has been applied extensively attention and development in recent years, but the radiation generated during CT scanning might pose a potential hazard to the human body. However, reducing the radiation dose by reducing the X-ray tube current can introduce noise and artifacts for CT images, which affects the physician's diagnosis. In response to these problems, many methods [Xia, Xiong, Athanasios et al. (2017); Du, Li and Guo (2013)] have been proposed to improve the quality of reconstructed low-dose CT images (LDCT). These methods are mainly divided into three categories, including projection domain method, iterative reconstruction method and post-processing method.

The projection domain method decides the denoising algorithm suitable for the characteristics of the projection domain, and then uses the traditional analytical reconstruction algorithm which is the filtered back projection (FBP) algorithm to reconstruct the image. The iterative reconstruction method uses the likelihood function to connect the image and the projected image according to the statistical characteristics. The image reconstruction [Munteanu, Brisan, Chiroiu et al. (2012); Cui, McIntosh and Sun (2018); Nie, Xu, Feng et al. (2018)] is performed in an iterative manner by incorporating a priori information into the objective function. [Chen, Zhang, Zhang et al. (2017a)] proposed adaptive weighted non-local priors according to the non-local mean Algorithms. These methods improve the quality of CT image reconstruction, but the disadvantage of long iteration time can occur. The post-processing method directly operates on low-dose

¹ School of Computer Science and Technology, Nanjing Tech University, Nanjing, 211816, China.

² Electrical and Computer Engineering, University of Rochester, Rochester, 14620, NY, USA.

* Corresponding Author: Xiaoli Li. Email: lixl@njtech.edu.cn.

CT images [Jayashree and Bhuvaneshwaran (2019)]. Since the features of LDCT images contain significant strip artifacts and noise, the purpose of the post-processing method is to remove artifacts and noise in the image. Chen et al. [Chen, Zhang, Zhang et al. (2017b)] proposed a fast dictionary learning method to improve the quality of the reconstructed CT images based on dictionary learning and sparse representation, but the dictionary learning method may introduce blurring and artifacts. Chen et al. [Chen, Zhang, Zhang et al. (2017a)] applied the non-local mean (NLM) method to CT image denoising. This method can significantly improve image quality, but residual error and excessive smoothing can still be observed in processed images, and because CT image noise is unevenly distributed, these problems have not been solved well.

The recent rapid development of deep neural networks provides a new way to solve the problem of low-dose CT image denoising. Due to the powerful feature learning and mapping capabilities of deep neural networks, deep neural networks show better reconstruction quality and faster speed than traditional methods. So many work is carried out in this respect. Chen et al. [Chen, Zhang, Zhang et al. (2017b)] first applied convolutional neural network (CNN) to low-dose CT image denoising. Compared with traditional methods in visual effects and evaluation indicators, it shows certain superiority. Chen et al. [Chen, Zhang, Zhang et al. (2017b)] proposed the residual encoder convolutional neural network (RED-CNN), and obtained the best results on the objective evaluation index, but the network complexity is high.

These deep neural networks have achieved good results in low-dose CT image denoising. These networks all use the mean square error (MSE) as a loss function, minimizing the mean square error usually results in excessive edge smoothing and loss of details. At the same time, the image texture that is crucial to human perception will be neglected. For the original GAN, there are problems such as difficulty in training, and the gradient is easy to disappear. Yang et al. [Yang, Yan, Zhang et al. (2018)] applied WGAN to low-dose CT image denoising, and achieved good results. However, due to the inherent defects of WGAN design, the convergence speed can be further improved.

2 Denoising based on WGAN-gp

This section firstly introduces the denoising model, and then introduces the proposed WGAN-gp in this paper. WGAN uses Wasserstein distance instead of simple pixel MSE as its loss function. This method can well overcome the problem of excessive edge smoothing and the loss of details. At the same time, WGAN-gp with gradient penalty term improves the deficiency of WGAN and accelerates the convergence speed. The proposed novel perceptual loss is also well preserving the image texture that is vital to human perception.

2.1 Denoising model

Hypothesis $x \in \mathbb{R}^{N \times N}$ represents a normal dose CT image (NDCT), $z \in \mathbb{R}^{N \times N}$ indicating a corresponding low-dose CT image (LDCT). The goal of denoising is to find a function G that maps z to x .

$$G: z \rightarrow x \tag{1}$$

$x \in \mathbb{R}^{N \times N}$ can be regarded as a sample of the NDCT image distribution P_r , and $z \in \mathbb{R}^{N \times N}$ is a sample of the corresponding LDCT image distribution P_L . The function G maps the LDCT image distribution P_L to a specific image distribution P_g . It makes the generated distribution similar to the real sample distribution P_r .

2.2 WGAN-gp

The Generative Adversarial Networks (GAN) is a combined neural network comprising a generator network G and a discriminator network D . The generator G accepts a random vector z to generate an image $G(z)$. The discriminator D receives the real image x and generates an image $G(z)$. To give the real image a higher score, the discriminator gives the generated “false” image a lower score. With multiple iterations of training, the image generated by the generator G is getting closer to the real image. At the same time, the discriminator D cannot judge whether the input image is real or generated. Then the network is trained successfully. The training D and G are solved by the following methods. Regularized Wasserstein distance, which is also called the entropy regularized optimal transport distance, which formula is as the follow.

$$\min_G \max_D L_{GAN}(D, G) = \mathbb{E}_{x \sim P_r} [\log D(x)] + \mathbb{E}_{z \sim P_z} [\log(1 - D(G(z)))] \quad (2)$$

where \mathbb{E} represents the expectation, and P_z is the random vector(noise) sample distribution and P_r is the real data sample distribution, P_g is the sample distribution for the generator transformation. $D(x)$ and $D(G(z))$ are both probabilities with a value of $[0,1]$. The original GAN has a fatal flaw: the better the discriminator training, the more serious the generator gradient disappears. Specifically, for any sample x , the contribution to the discriminator loss function is that.

$$-P_r(x) \log D(x) - P_g(x) \log[1 - D(x)] \quad (3)$$

Deriving for $D(x)$, the optimal discriminator function is as the follow.

$$D^*(x) = \frac{P_r(x)}{P_r(x) + P_g(x)} \quad (4)$$

Substituting the generator loss function $L(G) = -\mathbb{E}_{x \sim P_g} [D(x)]$, the JS divergence is as the follow.

$$E_{x \sim P_r} \log \frac{2P_r(x)}{P_r(x) + P_g(x)} + E_{x \sim P_g} \log \frac{2P_g(x)}{P_r(x) + P_g(x)} - 2 \log 2 \quad (5)$$

In this case, if the intersection between P_r and P_g are zero measures, the generator’s loss function is always zero and the gradient disappears. For this problem, the Wasserstein distance [Arjovsky, Chintala and Bottou (2017); Champion, Pascale and Juutinen (2008)] is used instead of the JS divergence. The advantage of the Wasserstein distance is that even if there is no overlap with the two distributions, and they can still measure their distance. The method of WGAN training D and G as the follow.

$$\min_G \max_D L_{WGAN}(D, G) = -\mathbb{E}_x [D(x)] + \mathbb{E}_z [D(G(z))] + \lambda \mathbb{E}_{\hat{x}} [(\|\nabla_{\hat{x}} [D(\hat{x})]\|_2 - 1)^2] \quad (6)$$

The first two terms perform Wasserstein distance estimation, and the last one is the network regularization gradient penalty term. Compared with the original GAN, WGAN removes the logarithmic function in the loss, and also removes the sigmoid layer in the discriminator D. WGAN is small enough to be simple and simple, but there are still disadvantages of training difficulties and slow convergence.

In the literature [Kingma and Ba (2014)] WGAN was used in the low-dose CT image denoising and achieved good results. By adding gradient penalty terms, the convergence speed is further improved. The specific improvement is:

The loss function of WGAN is defined as the follow:

$$\begin{aligned} L(D) &= -E_{x \sim P_r} [D(x)] + E_{x \sim P_g} [D(x)] \\ L(G) &= -E_{x \sim P_g} [D(x)] \end{aligned} \quad (7)$$

For the discriminator, in order to give scores after the input sample changed, there will be no drastic changes by adding Lipschitz limits.

$$\|\nabla_x D(x)\|_p \leq K \quad (8)$$

The LP norm of the discriminator $D(x)$ gradient is not more than a constant k . The Lipschitz limit is specifically achieved by weight clipping. For each time the discriminator parameters are updated, the absolute value of all the discriminator parameters is checked whether there is more than one threshold. If a threshold exists, a regression threshold range is enforced. This introduces two drawbacks. The first makes the neural network a binary network, and the parameters are easy to take up and down the bounds for the discriminator wants to maximize the score gap between the true and false samples. The second is that the gradient disappears or the gradient explodes, because the discriminator multi-layer network enlarges or reduces the gradient.

The Lipschitz limit can be expressed by additionally setting the loss term. Set k to 1 and get the discriminator loss of WGAN-gp by weighting the original discriminator loss with WGAN.

$$L_{\text{WGAN-gp}}(D) = -\mathbb{E}_{x \sim P_r} [D(x)] + \mathbb{E}_{x \sim P_g} [D(x)] + \lambda \mathbb{E}_{x \sim \mathbb{X}} [(\|\nabla_x D(x)\|_p - 1)^2] \quad (9)$$

The third term \mathbb{X} represents the entire sample space, and the estimation expectation in the high-dimensional sample space is exponential, which is practically difficult to implement. Therefore, it is not necessary to add the Lipschitz constraint to the entire space. As long as the sample is generated, the central region between the generated sample and the real sample are sampled. Therefore, the final discriminator loss is as the following.

$$L_{\text{WGAN-gp}}(D) = -\mathbb{E}_{x \sim P_r} [D(x)] + \mathbb{E}_{x \sim P_g} [D(x)] + \lambda \mathbb{E}_{x \sim \hat{P}_x} [(\|\nabla_x D(x)\|_p - 1)^2] \quad (10)$$

where $x \sim P_x$, in the third term \hat{x} is the random interpolation sample on the line of x_r and x_g , $\hat{x} = \varepsilon x_r + (1 - \varepsilon)x_g$, where $x_r \sim P_r$, $x_g \sim P_g$, $\varepsilon \sim \text{Uniform}[0,1]$.

2.3 Loss function

The purpose of the loss function in the network is generally to preserve image detail, usually using mean square error (MSE) as the loss function.

$$L_{MSE}(G) = E_{(x,z)} \left[\frac{1}{N^2} \|G(z) - x\|_F^2 \right] \quad (11)$$

Although this method minimizes the difference in pixels before generating the denoised image $G(z)$ and the NDCT image x , it may result in image blurring and loss of details. The improved idea is to introduce a perceptual loss based on advanced features to better characterize the image. This paper uses the perceived loss proposed in literature [Johnson, Alahi and Li (2016)].

$$L_{PL} = E_{(x,z)} \left[\frac{1}{whd} \|\phi(G(z)) - \phi(x)\|_F^2 \right] \quad (12)$$

where ϕ show the feature extractor, w , h and d represent the width, height and depth of the feature space. The perceived loss in this paper is defined as the loss of the ReLU active layer of the pre-trained 19-layer visual geometry group (VGG-19) network, thus ϕ representing VGG-19.

The compound loss function proposed in this paper is as the follow.

$$L = L_{MSE} + \lambda_p L_{PL} \quad (13)$$

where λ_p is the Lagrangian parameter which controls the trade-off. So the total loss function is defined as the follow.

$$\min_G \max_D L_{WGAN-gp}(D,G) + L_{MSE} + \lambda_p L_{PL} \quad (14)$$

2.4 Network architecture design

WGAN-gp network structure consists of three parts. Fig. 1. shows generator G , which is a convolutional neural network with 8 convolutional layers. The convolution kernel size of each convolutional layer is 3×3 pixels, the generator's first seven hidden layers have 32 filters. The last layer uses a 3×3 filter to generate a feature map. The modified linear unit (ReLU) is the activation function, where n is the convolution kernel and s is the convolution step. Fig. 3. shows the discriminator D network. The discriminator D has six convolutional layers, the first two convolutional layers have 64 filters, the middle two convolutional layers have 128 filters, and the last two convolutional layers have 256 filters. As in the generator, the convolution kernel size of each convolutional layer is 3×3 pixels. After the convolutional layer, there are two fully connected layers, the first has 1024 outputs, the other has only an output. Fig. 2. shows the VGG-19 network. Through this network, the novel perceptual loss proposed in this paper is obtained.

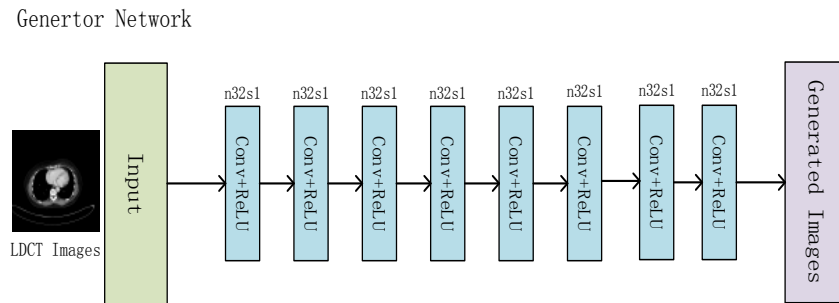


Figure 1: The structure of generator network G

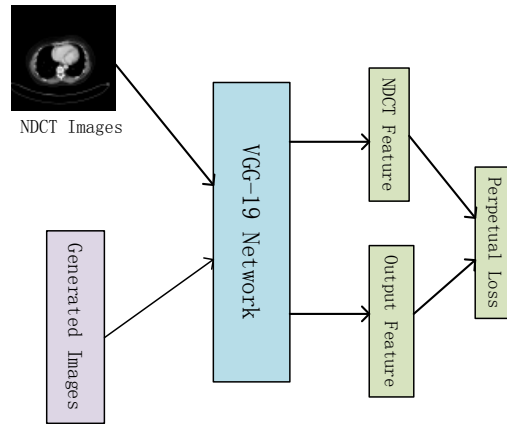


Figure 2: The network structure of perceived loss

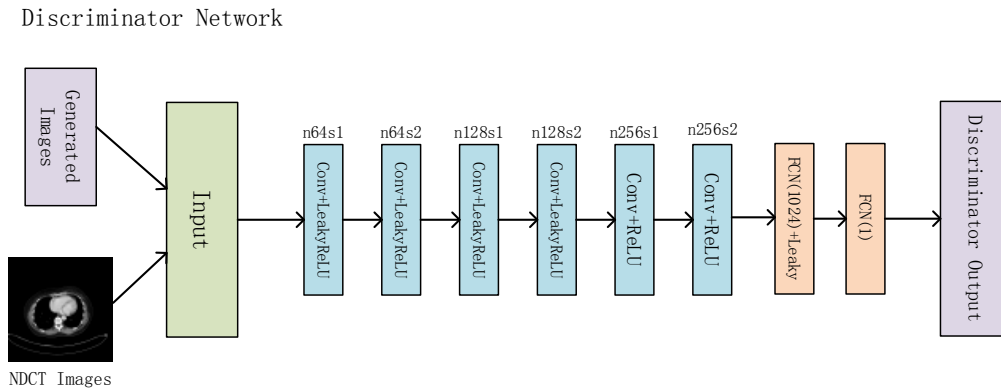


Figure 3: The structure of discriminator network D

3 Experimental results and analysis

The experimental computer hardware is configured for Intel Core i7-6700k and NVIDIA GeForce GTX 1080Ti. The model is tested on Python using the TensorFlow library.

3.1 Network training

300 different CT images were selected as training data with the size of 512×512 pixels from the TCGA-COAD clinical CT dataset. Low-dose CT images can be obtained by simulation: fan beam projection for normal dose CT (NDCT) images transform. The obtained projection matrix S is subjected to exponential operation and Poisson noise is added. Then the logarithm is taken, and transformed into the projection domain by the FBP algorithm to obtain a simulated low-dose CT image. The projection domain is added with noise.

$$\begin{aligned} \mathbf{I}_n &= \text{Poisson}[b \cdot \exp(-S)] \\ S_n &= \ln(b / \mathbf{I}_n) \end{aligned} \quad (15)$$

where b is taken as 10^6 , which is the number of emitted photons; \mathbf{I}_n is the number of photons received by the detector; \mathbf{S}_n is the projection matrix after noise pollution.

The input to the training network is an image block of size 64×64 pixels. The blank image block is pre-excluded when the image block is selected. In the experiment, all networks are optimized by the Adam algorithm [Kingma and Ba (2014)], which can replace the first-order optimization algorithm of the traditional stochastic gradient descent process, and iteratively update the neural network weight based on the training data. In this paper, the super-parameter of the Adam algorithm is set to: $\alpha = e^{-5}$, $\beta_1 = 0.5$, $\beta_2 = 0.9$, $\lambda = 10$ according to experimental experience. At the same time, the batch size of the network is set to 128.

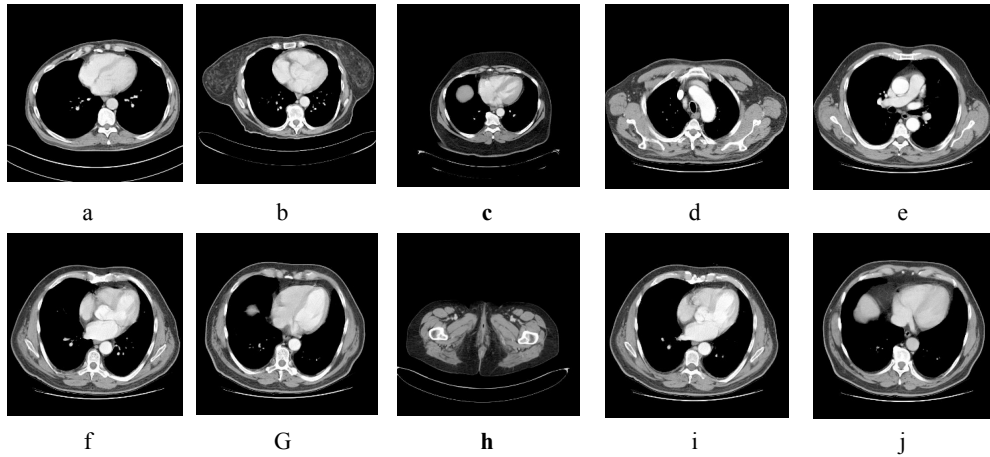
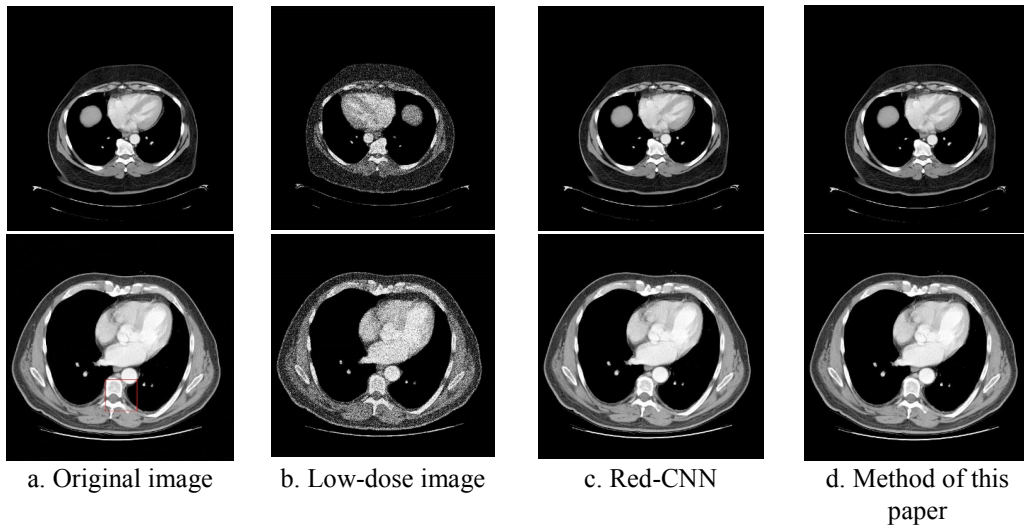
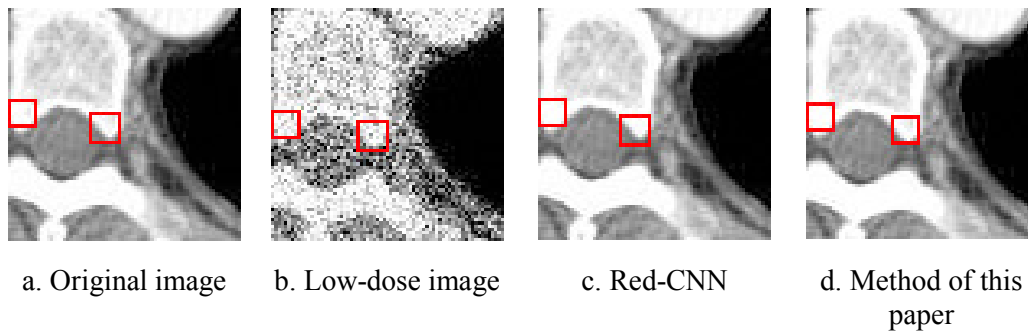


Figure 4: Experimental data

3.2 Subjective evaluation

Ten images of TCGA-COAD clinical dataset were randomly selected as test images, which were not repeated with the 300 images trained. The selected 10 images are shown in Fig. 4. Comparing this method with the literature RED-CNN method, the denoising effect is shown in Fig. 5. It can be seen from the figure that the proposed method and Red-CNN have achieved good results. The difference is that the network mentioned in this paper retains multiple image details. Fig. 6. is an enlarged view of the boxed area of Fig. 5. As shown in the red box, this method retains more image detail than Red-CNN. The gray bone area using Red-CNN in Fig. 6(c). is already difficult to identify, while the method in this paper is better preserved.

**Figure 5:** Denoising results comparison**Figure 6:** Zoom-in denoising results

3.3 Objective comparison

For quantitative analysis, the peak signal-to-noise ratio (PSNR) and structural similarity (SSIM) are used as the evaluation indicators for the denoising effect of low-dose CT images. The detailed data of the test chart are shown in Tab. 1.

As can be seen from Tab. 1, the proposed method has 7 test charts in comparison with Red-CNN in objective indicators. The average PSNR is higher than Red-CNN by 0.8 dB. SSIM is slightly ahead of Red-CNN in terms of indicators.

Table 1: Experimental results from three methods

Number	Index	Low-dose CT	Red-CNN	Method of this paper
a	PSNR/dB	24.8377	28.8700	30.7347
	SSIM	0.8924	0.9302	0.9314
b	PSNR/dB	26.8932	31.1257	32.7612
	SSIM	0.7316	0.7924	0.9104
c	PSNR/dB	24.8719	27.4356	28.5178
	SSIM	0.7893	0.8234	0.9012
d	PSNR/dB	22.5732	25.0124	31.3736
	SSIM	0.7538	0.8257	0.9088
e	PSNR/dB	23.9637	28.7472	31.4743
	SSIM	0.8137	0.8538	0.8837
f	PSNR/dB	31.4883	32.8477	31.7723
	SSIM	0.8473	0.9277	0.9173
g	PSNR/dB	28.3847	30.3648	32.4884
	SSIM	0.8928	0.9034	0.9374
h	PSNR/dB	28.3874	30.7744	34.3748
	SSIM	0.8137	0.8937	0.9134
i	PSNR/dB	25.3774	30.9983	32.1192
	SSIM	0.8353	0.8938	0.9147
j	PSNR/dB	26.8492	32.7482	31.5733
	SSIM	0.7937	0.9145	0.8972
Average	PSNR/dB	26.3627	29.8924	31.7190
	SSIM	0.8164	0.8759	0.9116

3.4 Algorithm complexity comparison

By calculating the average time consumption of 20 forward propagations for each test chart, the average time-consuming data is obtained, as shown in Tab. 2. The calculation formula of the network complexity is: $O\left\{\sum_l n_{l-1} f_l^2 n_l\right\}$, n_l is the number of feature maps outputted by the l th network layer, f_l is the l th convolution kernel size. By calling the Caffe interface in python, the average time is obtained as Tab. 2 showing.

Table 2: Complexity comparison chart

	RED-CNN	WGAN	WGAN-gp
Complexity	3837000	2434000	2247000
Time/s(CPU)	18.981	9.760	2.740
Time/s(GPU)	4.746	1.073	0.220

As it can be seen from Tab. 2, WGAN is more than 50% less time-consuming than RED-CNN. WGAN-gp speeds up the convergence rate, which is 64% less than WGAN time-consuming, and will have a better performance on the GPU.

4 Conclusion

In view of the difficulty of traditional algorithms in suppressing noise in low-dose CT images, this paper uses the generation-fighting network (WGAN-gp) to greatly shorten the convergence process. With the new loss function, the relevant details of the image are preserved. Experimental results show that in the denoising effect of low-dose CT images, the proposed method is better than Red-CNN. Compared with WGAN network, WGAN-gp network with gradient penalty term significantly improves the convergence speed.

Acknowledgements: The paper is supported by National Natural Science Foundation of China (61672279), Project of “Six Talents Peak” in Jiangsu (2012-WLW-023), and Open Foundation of State Key Laboratory of Hydrology-Water Resources and Hydraulic Engineering, Nanjing Hydraulic Research Institute, China (2016491411).

References

- Arjovsky, M.; Chintala, S.; Bottou, L. (2017): Wasserstein GAN. *Stat*, vol. 1050, pp. 26-56.
- Champion, T.; Pascale, L. D.; Juutinen, P. (2008): The Wasserstein distance: local solutions and existence of optimal transport maps. *SIAM Journal on Mathematical Analysis*, vol. 40, no. 1, pp. 1-20.
- Chen, H.; Zhang, Y.; Kalra, M. K.; Lin, F.; Chen, Y. et al. (2017a): Low-dose CT with a residual encoder-decoder convolutional neural network (RED-CNN). *IEEE Transactions on Medical Imaging*, vol. 36, no. 12, pp. 2524-2535.
- Chen, H.; Zhang, Y.; Zhang, W.; Liao, P.; Li, K. et al. (2017b): Low-dose CT denoising with convolutional neural network. *Biomedical Imaging*, pp. 143-146.
- Chen, Y.; Yu, F.; Luo, L.; Christine, T. (2013): Improving abdomen tumor low-dose CT images using dictionary learning based patch processing and unsharp filtering. *Engineering in Medicine and Biology Society*, pp. 4014-4017.
- Chen, Y.; Gao, D.; Nie, C.; Luo, L.; Chen, W. et al. (2009): Bayesian statistical reconstruction for low-dose X-ray computed tomography using an adaptive-weighting nonlocal prior. *Computerized Medical Imaging and Graphics*, vol. 33, no. 7, pp. 495-500.
- Cui, Q.; McIntosh, S.; Sun, H. (2018): Identifying materials of photographic images and photorealistic computer generated graphics based on deep CNNs. *Computers, Materials & Continua*, vol. 55, no. 2, pp. 229-241.
- Dou, Y.; Eeden, C. V. (2009): Comparisons of the performances of estimators of a bounded normal mean under squared-error loss. *Revstat Statistical Journal*, vol. 7 no. 3, pp. 203-226.
- Du, Z.; Li, X.; Guo, Y. (2013): Exposing blur kernel from retouch image. *Proceedings of 13th International Conference on Computer-Aided Design and Computer Graphics*, pp. 407-408.
- Goodfellow, I.; Pouget-Abadie, J.; Mirza, M.; Xu, B.; Warde-Farley, D. et al. (2014): Generative adversarial nets. *Advances in Neural Information Processing Systems*, vol. 3, pp. 2672-2680.

- Johnson, J.; Alahi, A.; Li, F. F.** (2016): Perceptual losses for real-time style transfer and super-resolution. *European Conference on Computer Vision*, pp. 694-711.
- Jayashree, N.; Bhuvaneshwaran, R. S.** (2019): A robust image watermarking scheme using z-transform, discrete wavelet transform and bidiagonal singular value decomposition. *Computers, Materials & Continua*, vol. 58, no. 1, pp. 263-285.
- Kingma, D. P.; Ba, J.** (2014): Adam: a method for stochastic optimization. *Computer Science*, vol. 1412, no. 6980.
- Ma, J.; Huang, J.; Feng, Q.; Zhang, H.; Lu, H. et al.** (2011): Low-dose computed tomography image restoration using previous normal-dose scan. *Medical Physics*, vol. 38, no. 10, pp. 5713-5731.
- Munteanu, L.; Brisan, C.; Chiroiu, V.; Donescu, S.** (2012): On the compression viewed as a geometric transformation. *Computers, Materials & Continua*, vol. 31, no. 2, pp. 127-146.
- Nie, Q.; Xu, X.; Feng, B.; Zhang, L.** (2018): Defining embedding distortion for intra prediction mode-based video steganography. *Computers, Materials & Continua*, vol. 55, no. 1, pp. 59-50.
- Quan, H.; Mohammed, G.** (2008): Scope of validity of PSNR in image/video quality assessment. *Electronics Letters*, vol. 44, no. 13, pp. 800-801.
- Wolterink, J. M.; Leiner, T.; Viergever, M. A.; Isgum, I.** (2017): Generative adversarial networks for noise reduction in low-dose CT. *IEEE Transactions on Medical Imaging*, vol. 36, no. 12, pp. 2536-2545.
- Xia, Z.; Lu, L.; Qiu, T.; Shim, H. J.; Chen, X. et al.** (2019): A privacy-preserving image retrieval based on AC-coefficients and color histograms in cloud environment. *Computers, Materials & Continua*, vol. 58, no. 1, pp. 27-43.
- Xia, Z.; Ma, X.; Shen, Z.; Sun, X.; Xiong, N. N. et al.** (2019): Secure image LBP feature extraction in cloud-based smart campus. *IEEE Access*, vol. 6, no. 1, pp. 30392-30401.
- Xia, Z.; Xiong, N. N.; Vasilakos, V. A.; Sun, X.** (2017): EPCBIR: an efficient and privacy-preserving content-based image retrieval scheme in cloud computing. *Information Sciences*, vol. 387, pp. 195-204.
- Yang, Q.; Yan, P.; Zhang, Y.; Yu, H.; Shi, Y. et al.** (2018): Low dose CT image denoising using a generative adversarial network with Wasserstein distance and perceptual loss. *IEEE Transactions on Medical Imaging*, vol. 37, no. 6, pp. 1348-1357.
- Chen, H.; Zhang, Y.; Zhang, W.; Liao, P.; Li, K. et al.** (2017): Low-dose CT denoising with convolutional neural network. *Medical Physics*, vol. 4, pp. 123-129.