Original Article

Synthetic high-energy computed tomography image via a Wasserstein generative adversarial network with the convolutional block attention module

Hai Kong[1#], Zhidong Yuan[2#], Haojie Zhou[3], Ganglin Liang[3], Zhonghong Yan[1], Guanxun Cheng[2], Zhanli Hu[3]

[1]School of Pharmacy and Bioengineering, Chongqing University of Technology, Chongqing, China; [2]Department of Radiology, Peking University Shenzhen Hospital, Shenzhen, China; [3]Lauterbur Research Center for Biomedical Imaging, Shenzhen Institute of Advanced Technology, Chinese Academy of Sciences, Shenzhen, China

Contributions: (I) Conception and design: H Kong; (II) Administrative support: Z Hu; (III) Provision of study materials or patients: Z Yuan; (IV) Collection and assembly of data: G Cheng; (V) Data analysis and interpretation: H Kong, H Zhou, G Liang, Z Yan; (VI) Manuscript writing: All authors; (VII) Final approval of manuscript: All authors.

[#]These authors contributed equally to this work.

Correspondence to: Zhanli Hu, PhD. Lauterbur Research Center for Biomedical Imaging, Shenzhen Institute of Advanced Technology, Chinese Academy of Sciences, No. 1068, Xueyuan Avenue, Shenzhen, China. Email: zl.hu@siat.ac.cn; Guanxun Cheng, PhD. Department of Radiology, Peking University Shenzhen Hospital, No. 1120, Lianhua Road, Shenzhen, China. Email: 18903015678@189.cn.

Background: Computed tomography (CT) is now universally applied into clinical practice with its non-invasive quality and reliability for lesion detection, which highly improves the diagnostic accuracy of patients with systemic diseases. Although low-dose CT reduces X-ray radiation dose and harm to the human body, it inevitably produces noise and artifacts that are detrimental to information acquisition and medical diagnosis for CT images.

Methods: This paper proposes a Wasserstein generative adversarial network (WGAN) with a convolutional block attention module (CBAM) to realize a method of directly synthesizing high-energy CT (HECT) images through low-energy scanning, which greatly reduces X-ray radiation from high-energy scanning. Specifically, our proposed generator structure in WGAN consists of Visual Geometry Group Network (Vgg16), 9 residual blocks, upsampling and CBAM, a subsequent attention block. The convolutional block attention module is integrated into the generator for improving the denoising ability of the network as verified by our ablation comparison experiments.

Results: Experimental results of the generator attention module ablation comparison indicate an optimization boost to the overall generator model, obtaining the synthesized high-energy CT with the best metric and denoising effect. In different methods comparison experiments, it can be clearly observed that our proposed method is superior in the peak signal-to-noise ratio (PSNR), structural similarity index measure (SSIM) and most of the statistics (average CT value and its standard deviation) compared to other methods. Because $P<0.05$, the samples are significantly different. The data distribution at the pixel level between the images synthesized by the method in this paper and the high-energy CT images is also most similar.

Conclusions: Experimental results indicate that CBAM is able to suppress the noise and artifacts effectively and suggest that the image synthesized by the proposed method is closest to the high-energy CT image in terms of visual perception and objective evaluation metrics.

Keywords: Low-dose computed tomography; convolutional block attention module (CBAM); generative adversarial networks (GAN)

Submitted Sep 09, 2022. Accepted for publication Apr 28, 2023. Published online June 15, 2023.
doi: 10.21037/qims-22-947
View this article at: https://dx.doi.org/10.21037/qims-22-947

Introduction

Over the past decades, an increasing number of researchers and specialists have concentrated on the important subject of how to reduce patient X-ray radiation doses while obtaining satisfying pictures. However, to obtain better imaging results with conventional computed tomography (CT) systems, patients must receive a higher dose of X-ray radiation, which is harmful to their health (1). Dual energy CT (DECT) mainly utilizes the different absorptions produced by substances at different energy X-rays, showing great technical advantages in the detection and imaging of various systems of the human body (2). The DECT imaging system consists of two X-ray sources and two detectors. the acquisition processes for high and low-energy are independent of each other, and their noise generation conditions are almost the same. The radiation dose required is not twice that of a conventional scan, but essentially equivalent or even less (3-7). In addition, DECT is capable of providing more favorable data on the function, morphology, occurrence, development, and prognosis of diseases as well, in addition to obtaining ordinary CT scan images (8). A variety of clinical applications have been successfully confirmed as well, making it possible to resolve problems that come with ordinary CT. For example, DECT was able to distinguish between calcified and noncalcified calculi in all cases, and dual-energy urinary calculus analysis was also effective with a low-dose protocol (9). DECT can also improve the ability to show tiny lesions, and the case (10) gave an example of dual-energy CT making it more effective than single-energy CT to detect bladder cancer.

Related work

Leading industrial CT suppliers have implemented two different systems to obtain distinct energy data. One with a dual source sets two tubes running at different voltages and corresponding detectors mounted orthogonally in one gantry, leading to inaccurate data information caused by a low sampling rate or signal crosstalk. The other is a single source whose main function is quickly switching the voltage of the single-tube emission source to achieve high- and low-energy scanning, which leads to the time interval between two scans affecting the information of the image (11). For the purpose of reducing the radiation dose of DECT, on the one hand we are able to improve the reconstruction algorithm to minimize noise and dose. On the other hand, the tube current modulation technique is used to directly reduce the tube current. The first type is based on some iterative reconstruction algorithms or analytical reconstruction methods such as the total variation (TV)-based methods, the penalized weight least-square (PWLS) method (12,13), as well as the cone-beam computed tomography (CBCT) reconstruction approach in the field of deep learning (14), by means of introducing with compressed sensing theory we are able to perform sparse reconstruction. With the second type, white noise will be generated by sparse X-ray photons in a low current environment, resulting in a negative impact on image quality. Iterative reconstruction techniques are now well established and reliable. However, compared with the traditional filtered back projection (FBP), the reconstruction speed is slightly slower due to the use of nonlinear operations that consume more resources (15-17). Therefore, we hope to identify the mapping relationship between low-energy scans and high-energy scans to reduce high-energy CT scans in the DECT system as much as possible and replace them with low-energy CT scans.

In recent years, data-driven deep learning-based methods relying on powerful nonlinear mapping capabilities have become a new trend and have been widely used in the field of medical imaging, including in segmentation (18), denoising (3,19-21) and reconstruction (22) tasks. Particularly for low-dose CT noise reduction, novel methods have been developed to convert between images presenting similar anatomy but different energy. For example, a residual encoder-decoder convolutional neural network (RED-CNN) (23) was proposed for generating low-dose CT images to replace normal-dose CT images. However, only taking into account the mean square error between the generated CT image and the real image might induce the resultant images to be vulnerable to blurred edges and missing detail information. To tackle this problem, the generative adversarial network (GAN) (24) was introduced for low-dose CT. Then, it was found that the Wasserstein distance is superior to the Jensen-Shannon (JS) divergence even if the two distributions do not overlap, because

the Wasserstein distance still reflects their proximity and overcome the difficulty in training GAN, yielding the Wasserstein distance-based GAN (WGAN) (25). Furthermore, In order to avoid gradient disappearance and gradient explosion during WGAN training, Improved Training of Wasserstein GANs (26) proposes a gradient penalty to solve this problem by setting an additional loss term similar to the L2 regularity. Importantly, Wasserstein generative adversarial network with visual geometry group perceptual loss (WGAN-VGG) (27) enhances the texture details of the generated images to make the recovered images more visually appealing and more in line with human sensory characteristics. The perceptual loss was implemented by visual geometry group (VGG) (28), which was pretrained on natural images.

In addition, Huang *et al.* (29) proposed a cycle-consistent generative adversarial network with attention (CaGAN) for low-dose CT noise reduction, which demonstrated that the attention module has a positive effect on improving the quality of generated images.

Contributions

Although these prior network architectures have achieved significant performance gains, it is still a great challenge in low-dose CT imaging to ensure that image details meet the diagnostic requirements. Consequently, inspired by these methods, we make use of the WGAN and devise a generator network structure in which an attention module (CBAM) and residual block are integrated for low-dose CT noise reduction. Our contributions can be summarized as follows:

(I) We propose a new generator model consisting of Vgg16, 9 residual blocks, upsampling and subsequent CBAM for synthesizing high-energy CT from low-energy CT, which is equivalent to reducing the noise of low-dose CT and the radiation dose of high-energy CT.

(II) We make the parameters of CBAM adjustable, by which we attempt to perform a large number of ablation experiments with training and test data to demonstrate that this model is acceptable.

Methods

Data sources

The study was conducted in accordance with the Declaration of Helsinki (as revised in 2013) and was approved by the ethics committee of Shenzhen Institutes of Advanced Technology, Chinese Academy of Sciences (Shenzhen, China). Individual consent for this retrospective analysis was waived. We analyzed a dataset of 80 patients, including different parts (chest, abdomen and head) of the human body. The dataset was divided into 6:2:2 beforehand for training, testing, and validation. The images were acquired using Siemens Somatom Definition Flash Dual-Source CT, with 80 kV scanning in system A and 140 kV scanning in System B. There was no time difference between different energy scans. The phase angle difference between the two sets of Siemens Somatom Definition Flash dual-source CT systems is 93 degrees in the XY-axis direction. Because it is spiral scanning and the scanning bed moves at a certain speed at the same time, there may be spatial dislocation in the images of different energies collected by the two sets of systems. In this experiment, a thin layer collimation layer thickness (64*0.6 mm), fast rack speed (0.33 s/turn) and small pitch (0.75) were used to reduce the possible spatial dislocation in the acquisition of the same layer. Meanwhile, Siemens' unique AMPR conical wire harness artifact correction image reconstruction algorithm could also greatly reduce the possible spatial dislocation between two sets of images. Finally, all data were acquired using the same machine in the same institution.

All the data in this group were scanned by dual-source CT, and their image sizes were all 512*512. To improve the training speed of the network model, all the datasets must be transformed into the form of tfrecord. It is necessary for the dataset to make a clipping between 0 and 1 by dividing by 3,000. We normalize the input images between 0 and 1 because the model will be increasingly convergent with the backpropagation gradient being under control.

The 80 patients included 55 males and 25 females with ages ranging from 25 to 75 years old. The low and high energies of the scanned images are 80 kV and 140 kV, respectively. We use the reconstructed paired high-low energy images as the training set, which contains 48 patients aggregating approximately 30,000 slices, and choose 16 patients with approximately 10,000 slices for testing. The remaining 16 patients with approximately 10,000 slices were prepared for validation. In addition, the patch size is set to 224*224 by means of a function that randomly crops the original 512*512 pictures.

Wasserstein generative adversarial network (WGAN)

As shown in *Figure 1*, a generative model G and a

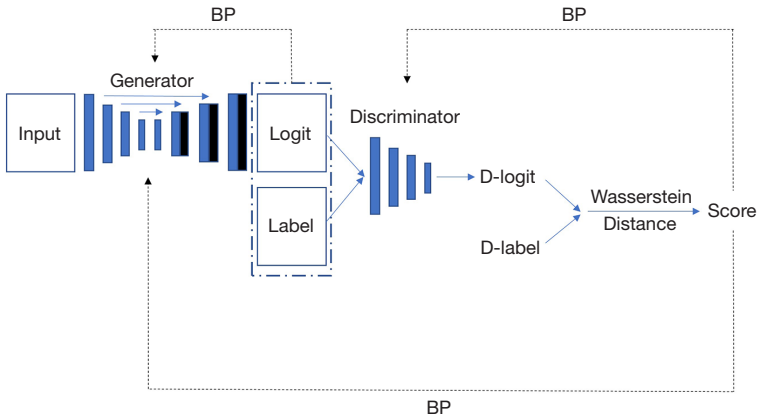

Figure 1 Overall workflow of the WGAN: Input, noise ~ U (–1,1). BP, backpagation; D-logit, formed from logit image going through Discriminator network; D-label, formed from label image going through Discriminator network; Score, measurement of similarity between images; WGAN, Wasserstein generative adversarial network.

discriminative model D constitute the WGAN. The generative model uses a pair of low-energy CT and high-energy CT to learn a function that maps a low-energy CT value distribution to a high-energy CT value distribution. The task of the discriminative model is to determine whether the images are from the synthesized high-energy CT distribution or the real high-energy CT distribution. Here, G and D are trained to solve the following min-max problem:

$$\min_G \max_D L_{GAN}(G,D) = E_{x\sim P_r}\left[\log D(x)\right] + E_{x\sim P_g}\left[\log\left(1-D(x)\right)\right] \quad [1]$$

where P_r is the true sample distribution and P_g is the sample distribution generated by the generator. the superiority of the Wasserstein distance over the JS divergence is that even if the two distributions do not overlap, the Wasserstein distance still reflects their proximity and overcome the difficulty in training GAN, yielding the Wasserstein distance-based GAN (WGAN) (25).

Network architecture

Figure 2: A diagram of the proposed generator structure in WGAN, consisting of Vgg16, 9 residual blocks, upsampling and a subsequent attention block (CBAM). In this model, we make use of the low-energy CT images that are randomly cropped from 512*512 to 224*224 as input for the requirement of Vgg16. We use a pretrained Vgg16 network to replace the encoder part of the generator, which can greatly reduce the training time.

In addition, as shown in *Figure 3*, the residual module is introduced in the generator structure, and nine tandem residual neural network blocks are used for deep connection, which can continuously capture the edge, texture, local features and global features with richer semantic information of the input low-dose CT and complete the transformation of low-dose CT features to those of high-dose CT.

Furthermore, the convolutional block attention module (CBAM) (30) is integrated into the generator for improving the denoising ability of the network (31). As shown in *Figure 4A*, CBAM applies the channel and spatial attention modules in turn to emphasize meaningful features in the two dimensions of space and channel. *Figure 4B* illustrates that the channel dimension is unchanged and the spatial dimension is compressed. The module focuses on meaningful information in the input image. Since the SENet (32) only maps spatial information into the channel through Global Average Pooling (GAP) to get a local optimal solution of the feature, which leads to the problem of lost information, we consider using Global Max Pooling (GMP) to compensate for the problem of lost spatial information when using GAP alone, making the extracted high-level features more comprehensive and richer. Next in the shared two-layer perceptron (MLP), the features are further extracted, finally by feature fusion and sigmoid function, attention vector of channel domain can be obtained. *Figure 4C* describe the process for obtaining attention vector of spatial domain, the main steps are as follows:

(I) Channel-refined feature F_C goes through both GAP and GMP operations along the channel dimension

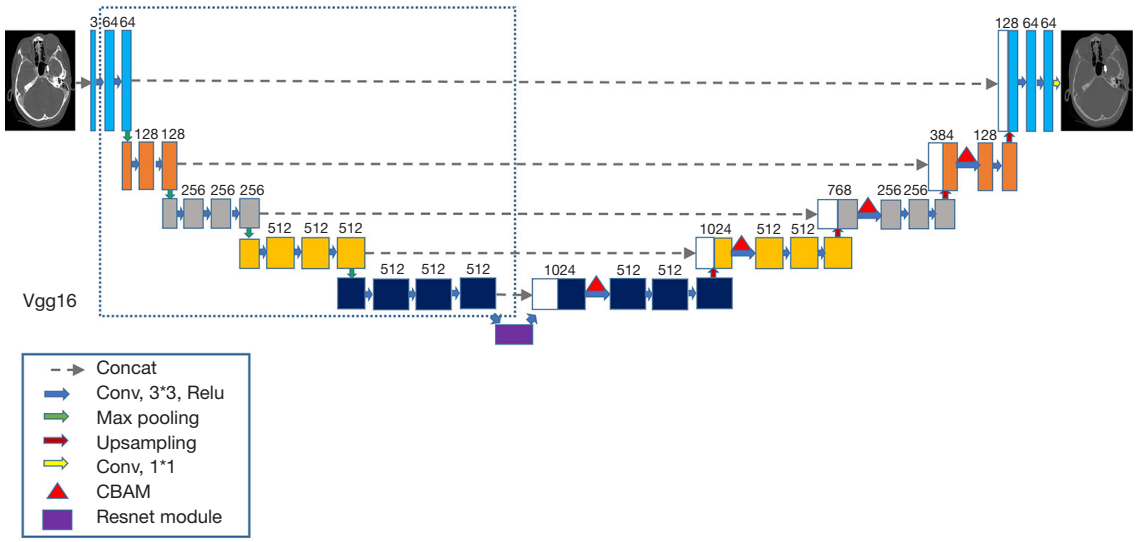

Figure 2 The architecture of the generator: Concat, concatenation for 3 channels; Conv, 3*3, Relu, convolution operation, convolution kernel size, activation function, respectively; CBAM, convolutional block attention module.

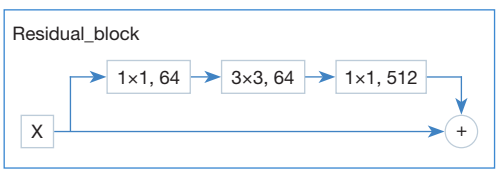

Figure 3 The architecture of the residual block: each box represents the size and number of convolution kernel.

to obtain two different channel feature description operators, respectively.

(II) The above results are stitched together and then subjected to a 7*7 convolution operation for expanding Receptive Field, which is significant to process spatial information.

(III) By sigmoid function, attention vector of spatial domain can be obtained.

Accordingly, we believe that CBAM can guide the network to concentrate on the meaningful features and suppress unimportant ones.

Table 1 reflects the parameters of CBAM adjustable, 'A' represents CBAM, and the subsequent number stands for the number of CBAM counting from bottom to top in *Figure 2*. The ratio represents the channel multiplier of the full connection regardless of ascending and descending. As the number of feature map channels increases or decreases exponentially in our generator, deeper features require a larger ratio to simplify their redundant information. Therefore, the number of fully connected neurons is reduced, which can reduce the complexity of the model parameters and prevent the model from overfitting. That is the reason why we set the attention module with adjustable parameters in the U-shaped generator network structure.

It can be seen from the *Figure 5* that the structure of the discriminator network D has four convolutional blocks that consist of two convolution operations followed by a LeaklyReLu (33) activation function structure. In the last part of this network, we use two fully connected layers, and the final layer removes the LeaklyReLu activation function to obtain a single output, which is a Wasserstein distance between the high-dose CT distribution and low-dose CT distribution.

Hybrid loss

The prior investigation justifies the use of a hybrid loss function for optimal diagnostic quality (34). As a comprehensive metric for the consistency of generated and real images, the hybrid loss enables the generated images to synthesize multi-scale information and improve the generalization performance of the model.

Adversarial loss

For the purpose of accelerating the training process and obtaining better denoised images, gradient penalty is often applied for the WGAN (26) loss.

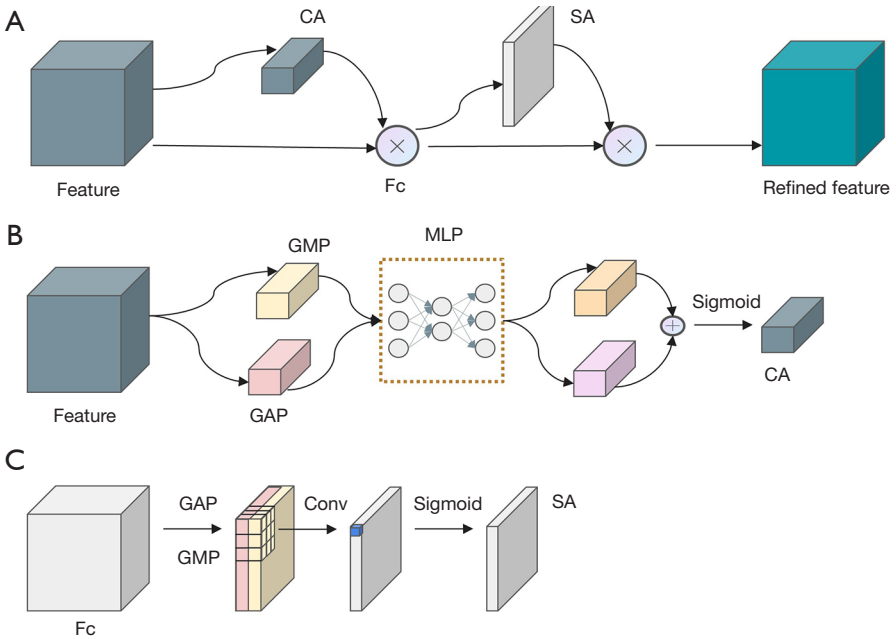

Figure 4 Overview of the convolutional block attention module. (A) Two parts of the convolutional block attention module. (B) Channel attention module. (C) Spatial attention module. CA, channel attention module; Fc, channel refined feature; SA, spatial attention module; GMP, Global Max Pooling; GAP, Global Average Pooling; MLP, two-layer perceptron.

Table 1 The parameters of convolutional block attention module

Attention module	Ratio
A0	–
A1	1/64
A2	1/32
A3	1/16
A4	1/8

Attention module, the number of convolutional block attention module; Ratio, the channel multiplier of the full connection regardless of ascending and descending; A0, original network with no convolutional block attention module; A1, 1 convolutional block attention modules added; A2, 2 convolutional block attention modules added; A3, 3 convolutional block attention modules added; A4, 4 convolutional block attention modules added.

$$L_{WGAN-GP}(G,D) = -E_{x\sim P_r}\left[D(x)\right] + E_{x\sim P_g}\left[D(x)\right] + \lambda E_{\hat{x}}\left[\left(\left\|\nabla\hat{x}D(\hat{x})\right\|_2 - 1\right)^2\right] \quad [2]$$

In order to avoid gradient disappearance and gradient explosion during WGAN training, Improved Training of Wasserstein GANs (26) proposes a gradient penalty to solve this problem by setting an additional loss term similar to the L_2 regularity. The last term restricts Critic's gradient norm to converge to 1. $\hat{x}$ is the linear interpolation between the generated samples and the real samples and λ=10.

Perceptual loss

It is of vital importance for medical images to preserve the significant features and details as a reference for diagnosis of diseases (35). Perceptual loss has been widely investigated in image transformation tasks over the past years (1). Compared with the normal L2 loss, the details of the output characteristics can be enhanced. To enhance the texture details of the generated images and make the recovered images visually better and more consistent with human sensory features, we add perceptual loss to the loss function of the generator. The perceptual loss can be described as follows:

$$L_{Perceptual}(G) = E_{(x,y)}\left[\frac{1}{whd}\left\|\phi(G(y)) - \phi(x)\right\|_F^2\right] \quad [3]$$

where ϕ denotes the pretrained VGG-19 network for the feature extractor. $\|*\|_F$ is the Frobenius norm. w, h, and d represent the width, height, and depth, respectively. The VGG-19 model was trained originally for the ILSVRC Challenge in 2014 and won first place in ILSVRC

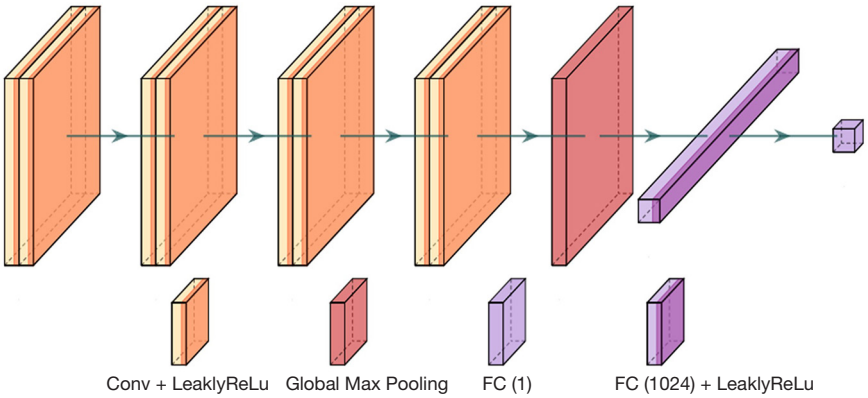

Figure 5 The architecture of the discriminator. LeaklyReLu, activation function; FC, fully connected network.

positioning and second place in classification, which is one of the seminal works laying the foundation for the field of deep learning.

Similarity loss

Natural images are highly structured, which is reflected in strong correlations between the pixels of the image, especially if they are spatially similar to medical CT images of different dose levels. These correlations carry important information about the structure of the object in the visual scene. The SSIM is calculated in three dimensions: luminance, contrast, and structure. When measuring the distance between two figures, more emphasis is placed on the structural similarity of the two figures, rather than using MSE or PSNR, which calculate the difference between the two images on a pixel-by-pixel basis. The SSIM can be described as follows:

$$SSIM\left(G(x),Y\right)=\frac{2\mu_x\mu_y+C_1}{\mu_x^2+\mu_y^2+C_1}*\frac{2\sigma_x\sigma_y+C_2}{\sigma_x^2+\sigma_y^2+C_2}*\frac{\sigma_{xy}+C_3}{\sigma_x\sigma_y+C_3} \quad [4]$$

where μ_x, μ_y, σ_x, σ_y, and σ_{xy} denote the means, standard deviations, and the covariance value of $G(x)$ and Y, respectively. In addition, the value of *SSIM* takes values in the range of 0–1. The *SSIM* loss can be defined as equation 5:

$$L_{SSIM}\left(G\right)=1-SSIM\left(G(x),Y\right) \quad [5]$$

The total loss of proposed generator is summarized as below:

$$L_{Total}=\alpha L_{WGAN}\left(C,D\right)+\beta L_{Perceptual}\left(G\right)+\gamma L_{SSIM}\left(G\right) \quad [6]$$

Where α, β and γ are weight coefficients of the above three terms. By fixing one variable and then optimizing the remaining two variables, these weight coefficients were assigned: $\alpha=10^{-3}$, $\beta=10^{-4}$, and $\gamma=5$.

Results

Contrast enhancement

Ioversol Injection 350 100 mL (350 mg iodine per mL) produced by Jiangsu Hengrui Pharmaceutical Co., Ltd. was used in this study, and 1.2 mL contrast agent was injected into the patient's right elbow vein evenly according to the injection duration of 18 s. After contrast agent injection, normal saline was injected at a slightly faster flow rate for 10 s. The contrast agent tracking technique was used, and the area of interest was set on the abdominal aorta 2 cm below the level of the renal artery. After the computed tomography value reached 120 HU, computed tomography angiogram (CTA) scanning in the cephalic direction was performed with a delay of 12–15 s (36-40). After the injection of contrast agent into the patient, the patient's right upper extremity vein, right subclavian vein, superior vena cava, heart, aorta, arteries of all parts of the body, liver, spleen, pancreas, kidney and other organs were successively enhanced display.

There is different attenuation for the same substance in different energies of X-rays. The same substance decays differently under different X-ray energies. Under low X-ray energy, the more it decays, the higher density it presents, while under high X-ray energy, the less it decays, the lower density it presents. As is shown in *Figure 6*, contrast agents in low-energy CT show high density, while in high-energy CT, they show low density. Under identical display circumstances, high density and low density show different appearances of white and black. Notably, the same ROI

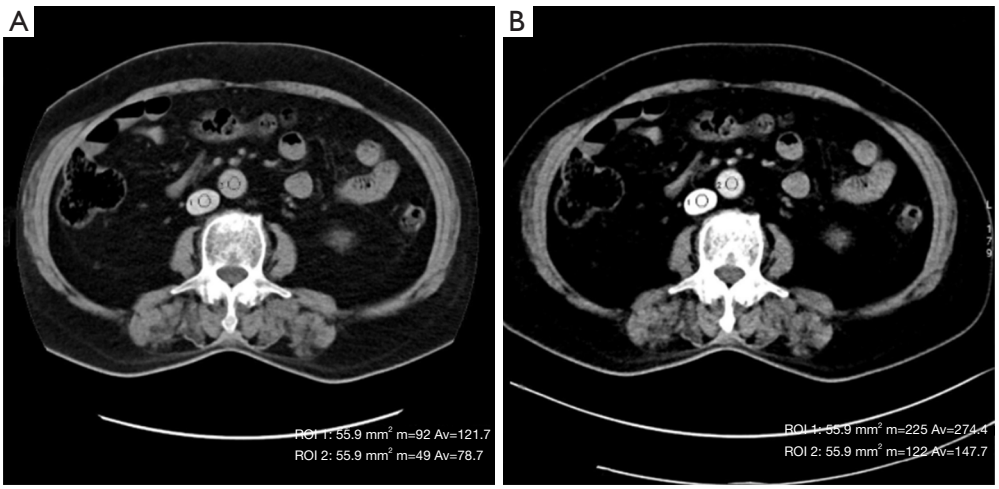

Figure 6 High energy (140 kV) and low energy (80 kV) CT for lower extremity venous case. (A) CT values for the 140 kV abdominal aorta and inferior vena cava, respectively. (B) CT values for the 80 kV abdominal aorta and inferior vena cava, respectively. CT, computed tomography; ROI 1, inferior vena cava; ROI 2, abdominal aorta.

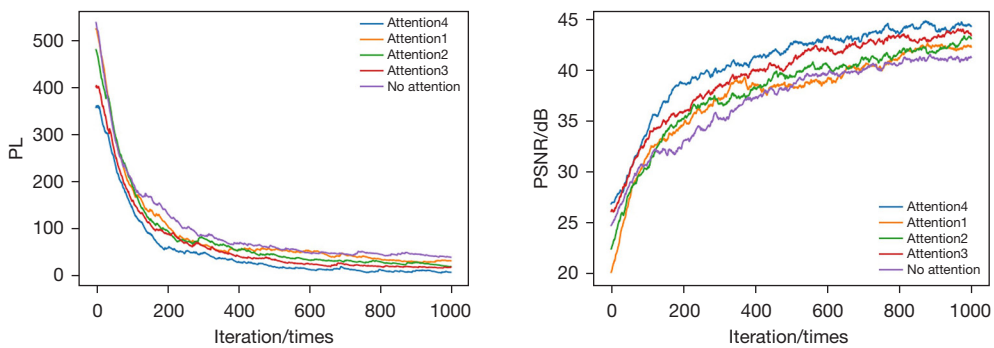

Figure 7 The change in PSNR and perceptual loss during 10,000 iterations. PL, perceptual loss; Iterations, the process of training a batch with 4 samples; PSNR, the peak signal-to-noise ratio.

appears sharper in high-energy images and has smaller CT values.

Ablation comparison experiment of the generator attention module

To verify that the CBAM attention mechanism will optimize the parameters of the original U-shaped network during training, we introduce the attention module variables A1–A4 sequentially to the original generator U-shaped structure A0 and analyze the evaluation metrics of the test images. Specifically, the entire training data are out of order, and the original image size of 512*512 is randomly cropped to 224*224 and then input into the improved WGAN network. The results are as follows after 10,000 iterations.

According to the *Figure* 7, it shows that the perceptual loss decreases as the number of iterations increases and finally converges to the minimum value continuously. In addition, adding the attention module can make the model converge faster and effectively reduce the PL value so that the images generated by the generator are more similar to the label images at the perceptual level; the PSNR increases as the number of iterations increases and finally converges to the maximum value continuously. In addition, adding the attention module can make the model converge faster and effectively increase the PSNR value so that the images generated by the generator are more similar to the label images at the pixel level.

We also selected a representative slice from the results

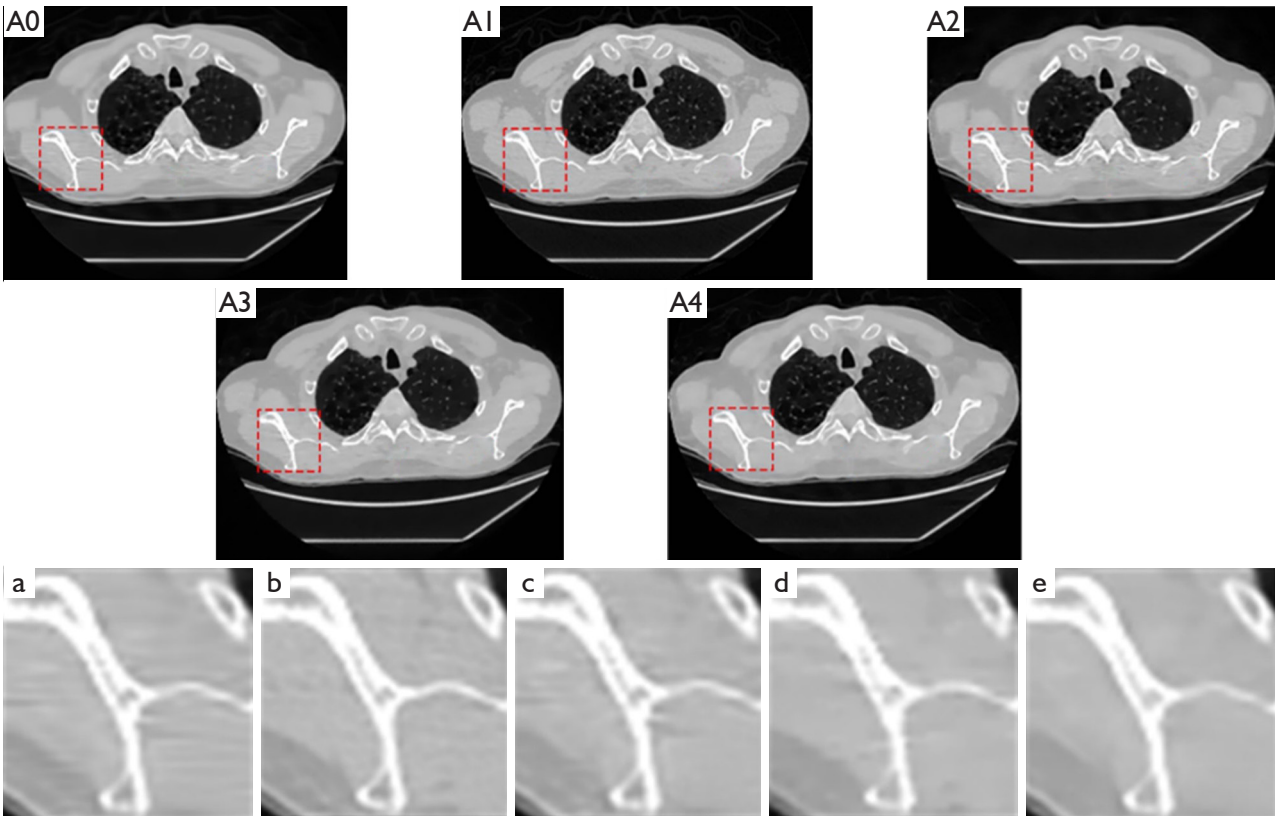

Figure 8 Results with a chest image. (a–e) The zoomed regions within the red box in A0–A4. A0, original network with no convolutional block attention module; A1, 1 convolutional block attention modules added; A2, 2 convolutional block attention modules added; A3, 3 convolutional block attention modules added; A4, 4 convolutional block attention modules added.

of the test set. *Figure 8* contains the results using different CBAM.

In *Figure 8* (a), the noise and artifacts caused by the lack of incident photons severely degrade the image quality. When we insert mutable CBAM in different depths of the generator, the images from (b) to (e) can be increasingly clear and smooth, which demonstrates that CBAM has a positive effect for the original network (A0) to generate better quality images. For the purpose of further comparison, the PSNR, PL and NMI were measured for all the test images in the red box.

Table 2 and *Figure 9* show the quantitative results of *Figure 8* were calculated. With the number of CBAM increasing, the PSNR and NMI rise gradually while PL drops steadily.

In *Figure 9*, the box plot shows a significant increase in PSNR and a slight decrease in PL after sequentially adding attention modules A1-A4, which indicates an optimization boost to the overall generator model. Moreover, the PSNR distribution shows a negative skew, and the PL distribution shows a positive skew, indicating that most of the data are distributed on the side with a larger PSNR and the side with a smaller PL.

Different methods comparison experiments

The above experiments establish the best U-shaped structure of the generator for this experiment, i.e., the addition of four attention modules, and in the field of deep learning, the experiments comparing several synthetic low-dose CT images are conducted on the basis of this optimized WGAN model.

The green box represents ROI 1, and the red box refers to ROI 2. *Figure 10* indicates that the low-dose CT image has the worst performance in the two ROIs. ROI 2 is white with a blurred overall structure, and the organization or detailed information cannot be distinguished at all. Compared to LDCT, clearer edges could be found in RED-CNN and WGAN-VGG, but the black block in the upper

Table 2 Experimental results of the generator attention module ablation comparison

Generator network structure	PSNR	PL	NMI
A0	41.784±2.001	7.336±3.530	1.539±0.064
A1	41.998±1.966	7.262±3.490	1.587±0.060
A2	42.083±1.938	7.128±3.352	1.608±0.061
A3	43.172±1.953	6.999±3.301	1.614±0.070
A4	44.240±1.932	6.918±3.371	1.621±0.074

Data are expressed as mean ± standard deviation. A0, original network with no convolutional block attention module; A1, 1 convolutional block attention modules added; A2, 2 convolutional block attention modules added; A3, 3 convolutional block attention modules added; A4, 4 convolutional block attention modules added; PSNR, the peak signal-to-noise ratio; PL, perceptual loss; NMI, normalized mutual information.

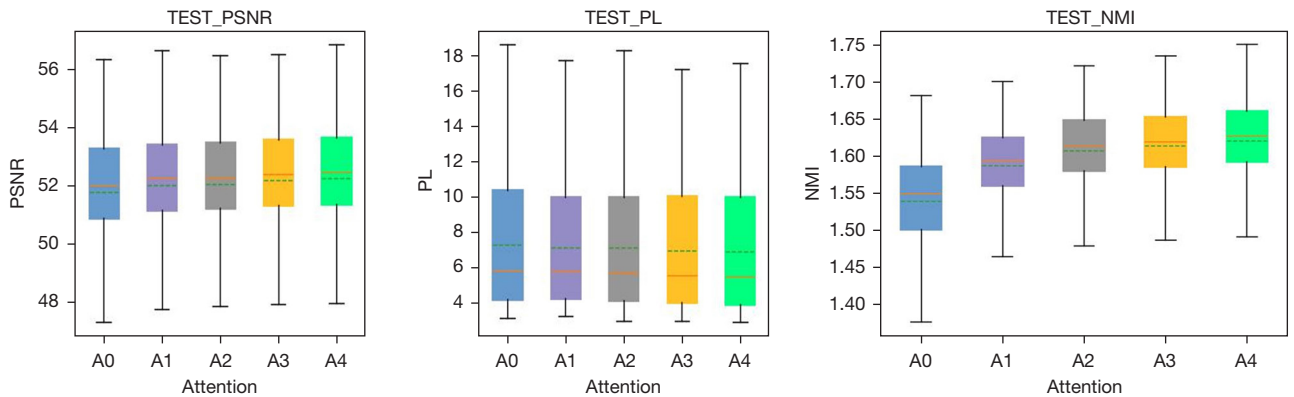

Figure 9 Changes in evaluation indicators with different convolutional block attention modules. PSNR, the peak signal-to-noise ratio; PL, perceptual loss; NMI, normalized mutual information; TEST_PSNR, test results for PSNR; TEST_PL, test results for PL; TEST_NMI, test results for NMI. A0, original network with no convolutional block attention module; A1, 1 convolutional block attention modules added; A2, 2 convolutional block attention modules added; A3, 3 convolutional block attention modules added; A4, 4 convolutional block attention modules added.

left corner of ROI 1 could hardly be found. In contrast, our method shows richer detailed information than the other methods. In addition, the statistics of some image metrics in the two ROIs are listed below.

In view of our output images are of array format whose intensity is equivalent to pixel values, i.e., between 0 and 1. We are able to make a reverse shift in intensity that is multiplied by 3,000 for mapping the output intensity ranges to HU. We only preserved the 0–3,000 range of CT values. All the training images are in the axial plane in the model.

To evaluate whether the sampled images of LECT and the synthetic images of other methods have statistically significant differences, we make use of a t test for validation, and the steps are listed below.

(I) Make a hypothesis and determine the level of bilateral test (α=0.05).

(II) Perform a t-statistic test and query the t value to obtain the corresponding P value.

(III) Comparing the values between P and α, if P≤α, the samples are significantly different.

$$t = \frac{\overline{x}_1 - \overline{x}_2}{S_{x_1 - x_2}} \quad [7]$$

$$S_{x_1 - x_2} = \sqrt{\frac{S_1^2}{n_1} + \frac{S_2^2}{n_2}} \quad [8]$$

Table 3 shows the results of the quantitative analysis of this experiment, and the best data records have been marked. It can be clearly observed that our proposed method is superior in the image evaluation metrics PSNR,

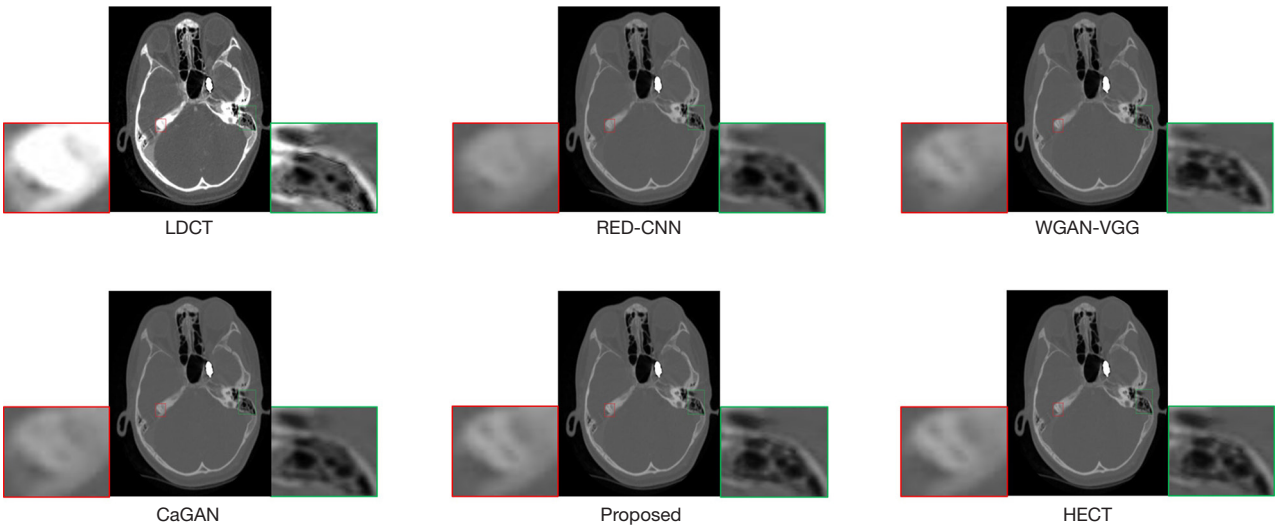

Figure 10 Generated head images from different methods. CT, computed tomography; LDCT, low-dose CT; RED-CNN, residual encoder-decoder convolutional neural network; WGAN-VGG, Wasserstein generative adversarial network with visual geometry group perceptual loss; CaGAN, cycle-consistent generative adversarial network with attention; Proposed, original network with 4 convolutional block attention modules; HECT, high-energy CT.

Table 3 Evaluation metrics of the test set images under different methods

Region	Metric	Methods					
		HECT	LECT	RED-CNN	WGAN-VGG	CaGAN	Proposed
ROI 2	SSIM	–	0.485	0.651	0.725	0.764	0.815
	Average CT value (HU)	415	776	658	542	476	440
	SD	5.06	8.67	8.16	6.25	6.00	5.43
	P value	–	–	0.036	0.004	<0.001	<0.001
ROI 1	SSIM	–	0.445	0.621	0.736	0.796	0.831
	Average CT value (HU)	85	175	152	118	96	91
	SD	4.81	6.69	5.82	5.46	4.73	5.33
	P value	–	–	0.008	0.023	0.016	<0.001

ROI, region of interest; SSIM, structural similarity; HU, Hounsfield unit; SD, standard deviation; P value, the probability of a status quo or worse scenario when the original assumptions are assumed to be correct; HECT, high-energy CT; LECT, low energy CT; RED-CNN, residual encoder-decoder convolutional neural network; WGAN-VGG, Wasserstein generative adversarial network with visual geometry group perceptual loss.

SSIM and most of the statistics (average CT value and its standard deviation) compared to other methods. Because $P<0.05$, the samples are significantly different. The data distribution at the pixel level between the images synthesized by the method in this paper and the real images is also most similar. In *Figure 11*, we calculate the pixel distance between the real images and the images generated by these methods, from which we can visually see the difference at the pixel level.

The pixel distance between the method from left to right compared with the real image gradually decreases, which proves that the image generated by the proposed method is closest to the real image at the pixel level. To demonstrate that the outlines and edges of the synthetic images matched those of the real images, as seen in *Figure 12*, we mark the edges of the synthetic images clearly in the real images. The

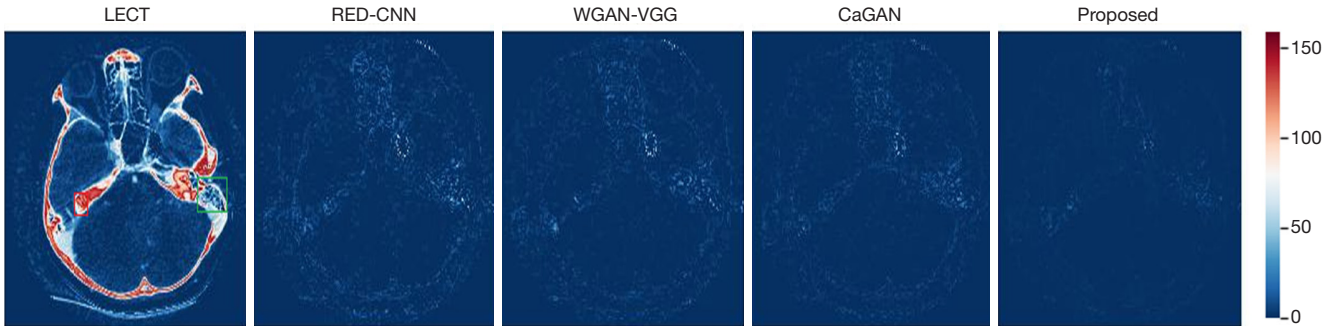

Figure 11 Heat map of pixel differences between images generated by different methods and high-energy CT images. LECT, low energy CT; RED-CNN, residual encoder-decoder convolutional neural network; WGAN-VGG, Wasserstein generative adversarial network with visual geometry group perceptual loss; CaGAN, cycle-consistent generative adversarial network with attention; CT, computed tomography.

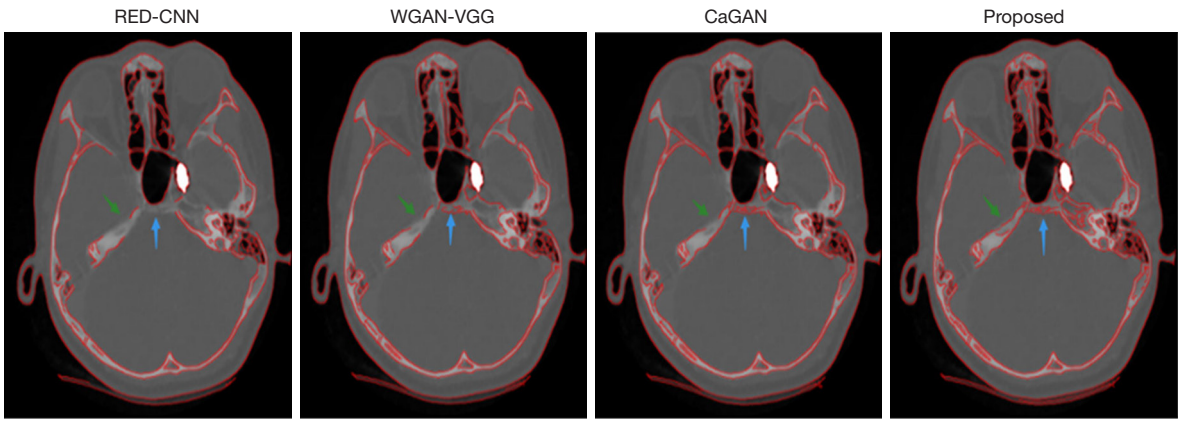

Figure 12 Overlays of synthetic image edges on the real image. The positions marked by the green and blue arrows refer to the area where different methods are compared. RED-CNN, residual encoder-decoder convolutional neural network; WGAN-VGG, Wasserstein generative adversarial network with visual geometry group perceptual loss; CaGAN, cycle-consistent generative adversarial network with attention.

positions marked by the green and blue arrows indicate superiority of the proposed method, which can generate more contour details similar to the real image than other methods. Moreover, the method is based on Canny edge detection with the same parameters.

Discussion

Training details

In the repeated experiments, we verify the effectiveness of the following training parameters. We use the Adam optimizer (41) with β_1=0.5 and β_2=0.9 in our network. At the beginning the learning rate is assigned to 10^{-4} and becomes half of the original after each epoch. The number of epochs is set to 100, 50,000 iterations in total. The slope of the LeaklyReLU activation function is set to 0.2. In addition, the patch size is set to 224*224 by means of a function that randomly crops the original 512*512 pictures. We also normalize the input images between 0 and 1 because the model will be increasingly convergent with the backpropagation gradient being under control.

In terms of avoiding overfitting appearance, after every epoch, we attempt to document the generator loss of both the training and validation parts in turn. If the effect of model training is no longer enhanced, for example, if the training loss is decreasing while the verification loss is no longer decreasing or even increasing, then the model training should be ended.

As shown in *Figure 13*, a comparison between the

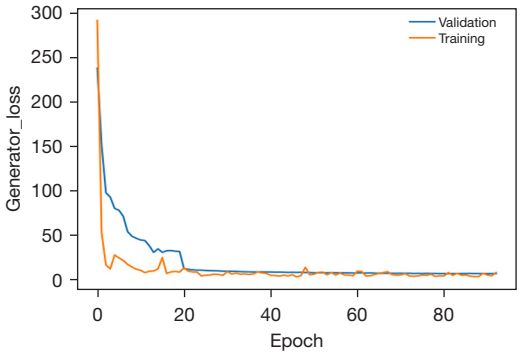

Figure 13 Variation in generator loss in the training and validation sets: Generator_loss, the total loss of generator; Epoch, the process of training once using all the samples in the training set.

validation and training, we can draw the conclusion that validation loss keeps decreasing as well when training loss is reducing. Consequently, overfitting does not occur in the training.

Image quality evaluation

Two out of three evaluation metrics were used in training/optimization. Since the optimized metrics were biased toward optimal values, the performance reported is not informative of the method capability. Other metrics that were not used in training should be used in evaluation.

To quantitatively measure the similarity between generated images and real images, we adopt three evaluation criteria: peak signal-to-noise ratio (PSNR) (42) perceptual loss (PL) (1) for the training, and normalized mutual information (NMI) (43) for the test evaluation. PSNR focuses on comparing the differences between the pixel points of the two images, as an evaluation metric of image quality. PL not only enhances the texture details of the generated images but also makes the recovered images visually better and more consistent with human sensory features. In addition, we introduce the normalized mutual information for the test evaluation, a supplementary evaluation metric that measures the similarity of two images, and its larger value represents the higher similarity of the two images.

$$NMI = \frac{H(A)+H(B)}{H(A,B)} \quad [9]$$

where $H(X)=-\sum_{x\in X} x\log x$ is the entropy. It ranges from 1 (perfectly uncorrelated image values) to 2 (perfectly correlated image values, whether positively or negatively).

Limitations

If the model was trained using axial slices, the other image planes (coronal and sagittal) should be showed for visual inspection of the synthetic outcomes. However, all the training images are in the axial plane in the model, and the other image planes (coronal and sagittal) are not available in our dataset. In the future, we will make use of a 3D dataset. Additionally, we plan to take other attention mechanisms into account as a comparison.

The output image is not in HU, we achieve HU mapping to grayscale pixel size by cropping the image to within 0–3,000 and dividing by 3,000. The whole process is reversible and can be interconverted, but it may result in loss of image structure information.

Conclusions

In our paper, traditional generator network combining with an attention mechanism is proposed for synthesizing low-dose CT images, which makes use of the CBAM in a generator network and enhances a WGAN with a hybrid loss. First, the active role of attention in the structure of the generator U-shaped network is established through attention module ablation comparison experiments. Then, quantitative comparison of low-dose images synthesized by different methods is conducted. Experimental results indicate that our method is able to suppress the noise and artifacts effectively and prove that the image synthesized by the proposed method is closest to the high-energy CT image in terms of visual perception and objective evaluation metrics.

Acknowledgments

Funding: This work was supported by the National Natural Science Foundation of China (Nos. 61672246, 61272068, 12071345, and 62101540); the Shenzhen Excellent Technological Innovation Talent Training Project of China (No. RCJC20200714114436080); the Fundamental Research Funds for the Central Universities, HUST (No. 2016YXMS018); the Programs for Science and Technology Development of Henan Province (No. 212102210511); the Guangdong Innovation Platform of Translational Research for Cerebrovascular Diseases of China.

Footnote

Conflicts of Interest: All authors have completed the ICMJE uniform disclosure form (available at https://qims.amegroups.com/article/view/10.21037/qims-22-947/coif). The authors have no conflicts of interest to declare.

Ethical Statement: The authors are accountable for all aspects of the work in ensuring that questions related to the accuracy or integrity of any part of the work are appropriately investigated and resolved. The study was conducted in accordance with the Declaration of Helsinki (as revised in 2013) and was approved by the ethics committee of Shenzhen Institutes of Advanced Technology, Chinese Academy of Sciences (Shenzhen, China). Individual consent for this retrospective analysis was waived.

References

1. Li Q, Li S, Li R, Wu W, Dong Y, Zhao J, Qiang Y, Aftab R. Low-dose computed tomography image reconstruction via a multistage convolutional neural network with autoencoder perceptual loss network. Quant Imaging Med Surg 2022;12:1929-57.
2. Marin D, Boll DT, Mileto A, Nelson RC. State of the art: dual-energy CT of the abdomen. Radiology 2014;271:327-42.
3. Zhou H, Liu X, Wang H, Chen Q, Wang R, Pang ZF, Zhang Y, Hu Z. The synthesis of high-energy CT images from low-energy CT images using an improved cycle generative adversarial network. Quant Imaging Med Surg 2022;12:28-42.
4. Wortman JR, Shyu JY, Dileo J, Uyeda JW, Sodickson AD. Dual-energy CT for routine imaging of the abdomen and pelvis: radiation dose and image quality. Emerg Radiol 2020;27:45-50.
5. Wang S, Wu W, Cai A, Xu Y, Vardhanabhuti V, Liu F, Yu H. Image-spectral decomposition extended-learning assisted by sparsity for multi-energy computed tomography reconstruction. Quant Imaging Med Surg 2023;13:610-30.
6. Uhrig M, Simons D, Kachelrieß M, Pisana F, Kuchenbecker S, Schlemmer HP. Advanced abdominal imaging with dual energy CT is feasible without increasing radiation dose. Cancer Imaging 2016;16:15.
7. Duan X, Ananthakrishnan L, Guild JB, Xi Y, Rajiah P. Radiation doses and image quality of abdominal CT scans at different patient sizes using spectral detector CT scanner: a phantom and clinical study. Abdom Radiol (NY) 2020;45:3361-8.
8. Sodickson AD, Keraliya A, Czakowski B, Primak A, Wortman J, Uyeda JW. Dual energy CT in clinical routine: how it works and how it adds value. Emerg Radiol 2021;28:103-17.
9. Thomas C, Patschan O, Ketelsen D, Tsiflikas I, Reimann A, Brodoefel H, Buchgeister M, Nagele U, Stenzl A, Claussen C, Kopp A, Heuschmid M, Schlemmer HP. Dual-energy CT for the characterization of urinary calculi: In vitro and in vivo evaluation of a low-dose scanning protocol. Eur Radiol 2009;19:1553-9.
10. Nakagawa M, Naiki T, Naiki-Ito A, Ozawa Y, Shimohira M, Ohnishi M, Shibamoto Y. Usefulness of advanced monoenergetic reconstruction technique in dual-energy computed tomography for detecting bladder cancer. Jpn J Radiol 2022;40:177-83.
11. Obmann MM, Punjabi G, Obmann VC, Boll DT, Heye T, Benz MR, Yeh BM. Dual-energy CT of acute bowel ischemia. Abdom Radiol (NY) 2022;47:1660-83.
12. Huang Z, Chen Z, Quan G, Du Y, Yang Y, Liu X, et al. Deep Cascade Residual Networks (DCRNs): Optimizing an Encoder–Decoder Convolutional Neural Network for Low-Dose CT Imaging. IEEE Transactions on Radiation and Plasma Medical Sciences 2022;6:829-40.
13. Ma J, Zhang H, Gao Y, Huang J, Liang Z, Feng Q, Chen W. Iterative image reconstruction for cerebral perfusion CT using a pre-contrast scan induced edge-preserving prior. Phys Med Biol 2012;57:7519-42.
14. Lee HC, Song B, Kim JS, Jung JJ, Li HH, Mutic S, Park JC. An efficient iterative CBCT reconstruction approach using gradient projection sparse reconstruction algorithm. Oncotarget 2016;7:87342-50.
15. Szczykutowicz TP, Toia GV, Dhanantwari A, Nett B. A review of deep learning CT reconstruction: concepts, limitations, and promise in clinical practice. Current Radiology Reports 2022;10:101-15.
16. Geyer LL, Schoepf UJ, Meinel FG, Nance JW Jr, Bastarrika G, Leipsic JA, Paul NS, Rengo M, Laghi A, De Cecco CN. State of the Art: Iterative CT Reconstruction Techniques. Radiology 2015;276:339-57.
17. Fletcher JG, Yu L, Fidler JL, Levin DL, DeLone DR, Hough DM, Takahashi N, Venkatesh SK, Sykes AG, White D, Lindell RM, Kotsenas AL, Campeau NG, Lehman VT, Bartley AC, Leng S, Holmes DR 3rd, Toledano AY, Carter RE, McCollough CH. Estimation of Observer Performance for Reduced Radiation Dose Levels in CT: Eliminating Reduced Dose Levels That Are Too Low Is the First Step. Acad Radiol 2017;24:876-90.
18. Minaee S, Boykov Y, Porikli F, Plaza A, Kehtarnavaz N, Terzopoulos D. Image Segmentation Using Deep Learning: A Survey. IEEE Trans Pattern Anal Mach Intell 2022;44:3523-42.
19. Lee JY, Kim W, Lee Y, Ko E, Choi JH. Unsupervised Domain Adaptation for Low-Dose Computed Tomography Denoising.

IEEE Access 2022;10:126580-92.
20. Hu Z, Jiang C, Sun F, Zhang Q, Ge Y, Yang Y, Liu X, Zheng H, Liang D. Artifact correction in low-dose dental CT imaging using Wasserstein generative adversarial networks. Med Phys 2019;46:1686-96.
21. Huang Z, Liu Z, He P, Ren Y, Li S, Lei Y, Luo D, Liang D, Shao D, Hu Z, Zhang N. Segmentation-guided Denoising Network for Low-dose CT Imaging. Comput Methods Programs Biomed 2022;227:107199.
22. Cao Q, Mao Y, Qin L, Quan G, Yan F, Yang W. Improving image quality and lung nodule detection for low-dose chest CT by using generative adversarial network reconstruction. Br J Radiol 2022;95:20210125.
23. Chen H, Zhang Y, Kalra MK, Lin F, Chen Y, Liao P, Zhou J, Wang G. Low-Dose CT With a Residual Encoder-Decoder Convolutional Neural Network. IEEE Trans Med Imaging 2017;36:2524-35.
24. Goodfellow I, Pouget-Abadie J, Mirza M, Xu B, Warde-Farley D, Ozair S, Courville A, Bengio Y. Generative adversarial networks. Communications of the ACM 2020;63:139-44.
25. Gulrajani I, Ahmed F, Arjovsky M, Dumoulin V, Courville AC. Improved training of wasserstein GANs. In Proceedings of the 31st International Conference on Neural Information Processing Systems (NIPS'17). Curran Associates Inc., Red Hook, NY, USA, 5769-79.
26. Hein D, Persson M. Spectral CT denoising using a conditional Wasserstein generative adversarial network. In Medical Imaging 2023: Physics of Medical Imaging 2023;12463:700-3.
27. Yang Q, Yan P, Zhang Y, Yu H, Shi Y, Mou X, Kalra MK, Zhang Y, Sun L, Wang G. Low-Dose CT Image Denoising Using a Generative Adversarial Network With Wasserstein Distance and Perceptual Loss. IEEE Trans Med Imaging 2018;37:1348-57.
28. Ge Y, Su T, Zhu J, Deng X, Zhang Q, Chen J, Hu Z, Zheng H, Liang D. ADAPTIVE-NET: deep computed tomography reconstruction network with analytical domain transformation knowledge. Quant Imaging Med Surg 2020;10:415-27.
29. Huang Z, Chen Z, Zhang Q, Quan G, Ji M, Zhang C, Yang Y, Liu X, Liang D, Zheng H, Hu Z. CaGAN: a cycle-consistent generative adversarial network with attention for low-dose CT imaging. IEEE Transactions on Computational Imaging 2020;6:1203-18.
30. Woo S, Park J, Lee JY, Kweon IS. Cbam: Convolutional block attention module. In Proceedings of the European conference on computer vision (ECCV) 2018:3-19.
31. Kim DW, Ryun Chung J, Jung SW. Grdn: Grouped residual dense network for real image denoising and GAN-based real-world noise modeling. In Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition Workshops 2019.
32. Hu J, Shen L, Sun G. Squeeze-and-excitation networks. In Proceedings of the IEEE conference on computer vision and pattern recognition 2018: 7132-7141.
33. Nayef BH, Abdullah SNHS, Sulaiman R, Alyasseri ZAA. Optimized leaky ReLU for handwritten Arabic character recognition using convolution neural networks. Multimedia Tools and Applications 2022:1-30.
34. Sun Y, Pan B, Li Q, Wang J, Wang X, Chen H, Cao Q, Liu H, Feng T, Sun H, Xiao Y, Gong NJ. Clinical ultra-high resolution CT scans enabled by using a generative adversarial network. Med Phys 2022. [Epub ahead of print]. doi: 10.1002/mp.16172.
35. Shi Z, Li J, Cao Q, Li H, Hu Q. Low-dose spectral CT denoising method via a generative adversarial network. Journal of Jilin University (Engineering and Technology Edition) 2020; 1:1-10.
36. Geng Z, Cao Z, Liu J. Recent advances in targeted antibacterial therapy basing on nanomaterials. Exploration 2023;3:20210117.
37. Zhang Y, Xu Y, Kong H, Zhang J, Chan HF, Wang J, Shao D, Tao Y, Li M. Microneedle system for tissue engineering and regenerative medicine. Exploration 2023;3:20210170.
38. Zhu F, Ge J, Gao Y, Li S, Chen Y, Tu J, Wang M, Jiao S. Molten salt electro-preparation of graphitic carbons. Exploration 2022;3:20210186.
39. Truong PL, Yin Y, Lee D, Ko SH. Advancement in COVID-19 detection using nanomaterial-based biosensors. Exploration 2023;3:20210232.
40. Du Y, Huo Y, Yang Q, Han Z, Hou L, Cui B, Fan K, Qiu Y, Chen Z, Huang W, Lu J, Cheng L, Cai W, Kang L. Ultrasmall iron-gallic acid coordination polymer nanodots with antioxidative neuroprotection for PET/MR imaging-guided ischemia stroke therapy. Exploration 2023;3:20220041.
41. Ke Y, Sheng N, Wei G, Wang K, Qin F, Guo J. Subject-aware image outpainting. Signal, Image and Video Processing 2023;17:2661-9.
42. Chen H, Zhang Y, Zhang W, Liao P, Li K, Zhou J, Wang G. Low-dose CT via convolutional neural network. Biomed Opt Express 2017;8:679-94.
43. Studholme C, Hill DL, Hawkes DJ. An overlap invariant entropy measure of 3D medical image alignment. Pattern Recognition 1999;32:71-86.

Cite this article as: Kong H, Yuan Z, Zhou H, Liang G, Yan Z, Cheng G, Hu Z. Synthetic high-energy computed tomography image via a Wasserstein generative adversarial network with the convolutional block attention module. Quant Imaging Med Surg 2023;13(7):4365-4379. doi: 10.21037/qims-22-947

