



The synthesis of high-energy CT images from low-energy CT images using an improved cycle generative adversarial network

Haojie Zhou^{1,2#}, Xinfeng Liu^{3#}, Haiyan Wang¹, Qihang Chen¹, Rongpin Wang³, Zhi-Feng Pang⁴, Yong Zhang⁵, Zhanli Hu¹

¹Lauterbur Research Center for Biomedical Imaging, Shenzhen Institutes of Advanced Technology, Chinese Academy of Sciences, Shenzhen, China; ²College of Software, Henan University, Kaifeng, China; ³Department of Radiology, Guizhou Provincial People's Hospital, Guiyang, China; ⁴College of Mathematics and Statistics, Henan University, Kaifeng, China; ⁵Department of Orthopaedic, Shenzhen University General Hospital, Shenzhen, China

Contributions: (I) Conception and design: H Zhou; (II) Administrative support: Z Hu; (III) Provision of study materials or patients: X Liu; (IV) Collection and assembly of data: Q Chen; (V) Data analysis and interpretation: H Zhou; (VI) Manuscript writing: All authors; (VII) Final approval of manuscript: All authors.

[#]These authors contributed equally to this work.

Correspondence to: Yong Zhang. Department of Orthopaedic, Shenzhen University General Hospital, Shenzhen, China. Email: thzhangyong@aliyun.com; Zhanli Hu. Lauterbur Research Center for Biomedical Imaging, Shenzhen Institutes of Advanced Technology, Chinese Academy of Sciences, Shenzhen, China. Email: zl.hu@siat.ac.cn.

Background: The dose of radiation a patient receives when undergoing dual-energy computed tomography (CT) is of significant concern to the medical community, and balancing the tradeoffs between the level of radiation used and the quality of CT images is challenging. This paper proposes a method of synthesizing high-energy CT (HECT) images from low-energy CT (LECT) images using a neural network that achieves an alternative to HECT scanning by employing an LECT scan, which greatly reduces the radiation dose a patient receives.

Methods: In the training phase, the proposed structure cyclically generates HECT and LECT images to improve the accuracy of extracting edge and texture features. Specifically, we combine multiple connection methods with channel attention (CA) and pixel attention (PA) mechanisms to improve the network's mapping ability of image features. In the prediction phase, we use a model consisting of only the network component that synthesizes HECT images from LECT images.

Results: Our proposed method was conducted on clinical hip CT image data sets from Guizhou Provincial People's Hospital. In a comparison with other available methods [a generative adversarial network (GAN), a residual encoder-to-decoder network with a visual geometry group (VGG) pretrained model (RED-VGG), a Wasserstein GAN (WGAN), and CycleGAN] in terms of metrics of peak signal-to-noise ratio (PSNR), structural similarity index measure (SSIM), normalized mean square error (NMSE), and a visual effect evaluation, the proposed method was found to perform better on each of these evaluation criteria. Compared with the results produced by CycleGAN, the proposed method improved the PSNR by 2.44%, the SSIM by 1.71%, and the NMSE by 15.2%. Furthermore, the differences in the statistical indicators are statistically significant, proving the strength of the proposed method.

Conclusions: The proposed method synthesizes high-energy CT images from low-energy CT images, which significantly reduces both the cost of treatment and the radiation dose received by patients. Based on both image quality score metrics and visual effects comparisons, the results of the proposed method are superior to those obtained by other methods.

Keywords: Computed tomography (CT); high-energy image synthesis; deep learning; cycle generative adversarial network

Submitted Feb 13, 2021. Accepted for publication Jul 02, 2021.

doi: 10.21037/qims-21-182

View this article at: <https://dx.doi.org/10.21037/qims-21-182>

Introduction

Dual-energy computed tomography (CT) has become an important noninvasive diagnostic technology which combines half high-energy CT (HECT) and half low-energy CT (LECT) scans to give patients less radiation, allowing more patient information to be obtained than when conventional CT scans are used (1,2). However, because spectral CT scans accumulate radiation and increase the risk of disease, exploring methods of reducing the radiation dose of CT scans is an important research area. Based on dual-energy CT (DECT), we have proposed a method of simultaneously synthesizing HECT scans by using only 1 HECT scan, which greatly reduces the radiation a patient receives and thus potentially expands the clinical application range of LECT scanning.

Among the state-of-the-art low-dose CT techniques, many traditional methods have been proposed to improve the quality of low-dose CT images, including (I) sinogram domain filtration methods (3-6); (II) iterative reconstruction methods, such as the total variation (TV) method and its variants (7-10); dictionary learning (DL) (11-14); and (III) image postprocessing methods, such as the nonlocal-mean (NLM) method (15) and the block-matching 3D (BM3D) algorithm (16). However, these methods are not ideal as they were all developed primarily for image denoising. In terms of noise, the results of HECT image are more complicated and include additional factors, such as reduced beam hardening, noise, and scattering, making research more difficult; therefore, traditional image-domain methods are unable to reconstruct special information in high-energy images. Moreover, X-ray projection data and its corresponding reconstruction algorithms are difficult to obtain, which limits the clinical application of projection data-based methods.

Deep learning techniques have developed rapidly in recent years. Applications of deep neural networks (DNNs) in medical image processing tasks have achieved impressive results, demonstrating their tremendous potential and providing new ideas for future research studies. Many

specialists have used sinogram domain data, image domain data, or a combination of both to enhance the quality of LECT images (13,17-22). For DECT images, Ma *et al.* (23) introduced a convolutional neural network (CNN) as a method of synthesizing pseudo-HECT images from LECT images. While CNN-based methods achieve high scores with respect to some evaluation indicators of image quality, visual observations suggest that they miss many details. CNN-based methods are flawed, as the texture of the generated image is oversmooth. In addition, the experimental window selected by Ma *et al.* lacked obvious contrast, and the region of interest (ROI) was unable to fully differentiate between HECT and LECT images. Alternative methods that use a generative adversarial network (GAN) have been proposed to solve this problem (24). A GAN learns deeper image features by producing a continuous confrontation between a generator model and a discriminator model to obtain a generated image that retains realistic details. In 2020, Yang *et al.* (25) proposed a GAN to synthesize LECT and HECT images from standard CT images to successfully compute the stopping power ratio. Yang *et al.* showed that a GAN can efficiently synthesize effective DECT images. The traditional GAN method uses a 1-directional mapping structure composed of a generator and a discriminator through which the network learns distribution mapping of LECT and HECT images. There is a weak mapping relationship between the 2 energies, and it can also be difficult to train the model due to strong artifacts and noise. In contrast, a CycleGAN (26) uses 2 generators and 2 discriminators to form a bidirectional mapping cycle structure. The network learns mutual mapping between the LECT and the HECT image domains to correct the extracted feature information, thereby generating more accurate results. The bidirectional cycle structure of the CycleGAN provides many new research opportunities, with related studies having applied this structure to explore its potential in medical processing and achieving promising results (27,28).

In this study, we propose a method for directly synthesizing HECT images from LECT images to

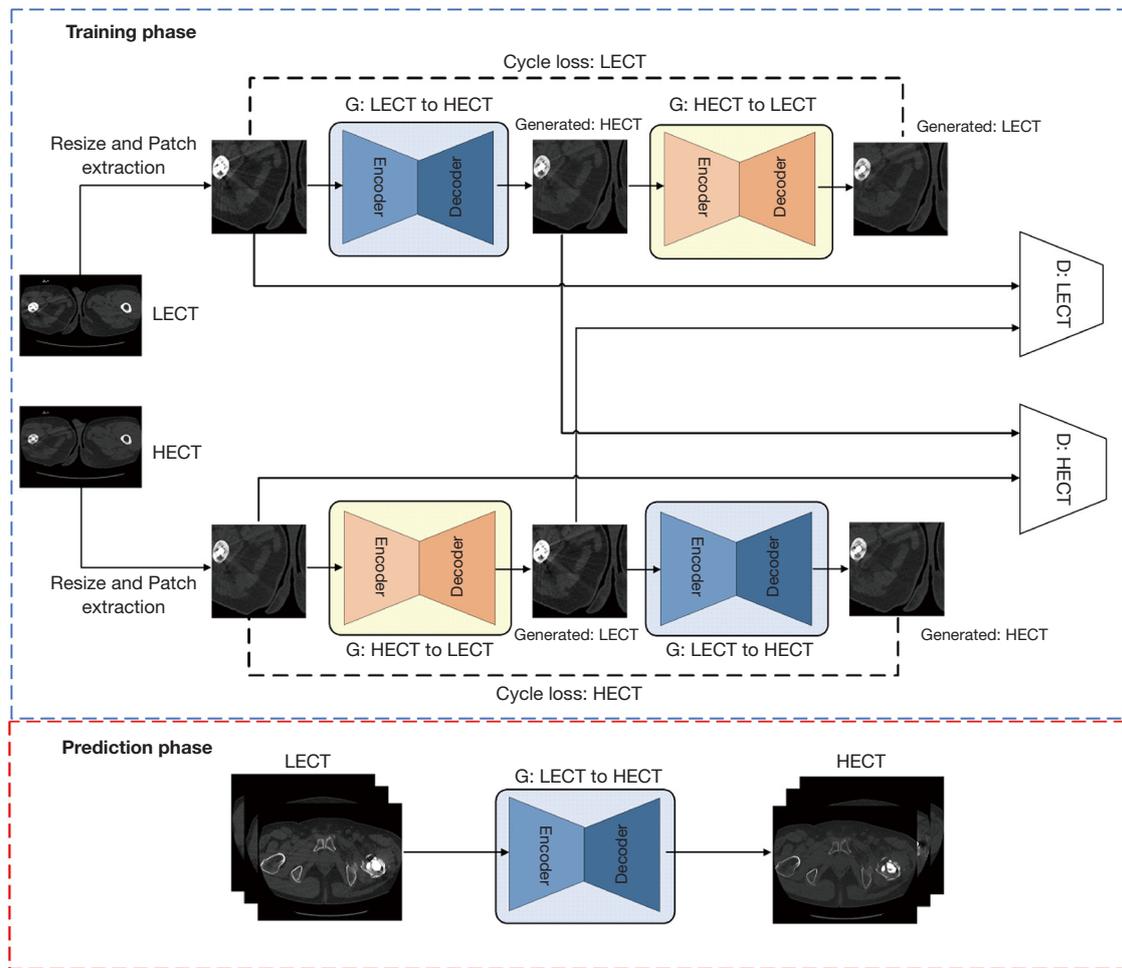


Figure 1 Overall workflow of the proposed method. G, generator; D, discriminator; LECT, low energy CT; HECT, high energy CT.

simulate the effect of DECT. The proposed method is based on the structure of CycleGAN but also adopts the connection strategies of ResNet (29), U-Net (30), and DenseNet (31), and employs attention mechanisms (32-34). The remainder of this paper is organized as follows: (I) introduces the network architecture in detail, (II) elaborates on the data used in the experiment and the experimental settings; (III) shows a DECT test data set for the performed experiments and compares and analyzes the results of the different methods, (IV) discusses the results of the overall study, and (V) summarizes the study findings.

Methods

In the image domain, an improved cycle generative

adversarial network is proposed as a method of synthesizing HECT images from LECT images. As shown in *Figure 1*, the proposed method is divided into 2 phases, a prediction phase and a training phase. The training phase includes 2 generators and discriminators forming a cyclic structure, in which one of the generators is used as the prediction phase. Specifically, the cyclic generator structure is used to extract the features of the relationships between the human tissues from both the forward and backward directions. Thus, the network learns the mapping functions between the HECT and LECT images. The discriminators are used to determine whether an input image is recognized as the target CT image.

During the training phase, the corresponding high-energy and low-energy CT input images from each batch are preprocessed. First, after the input image pairs are

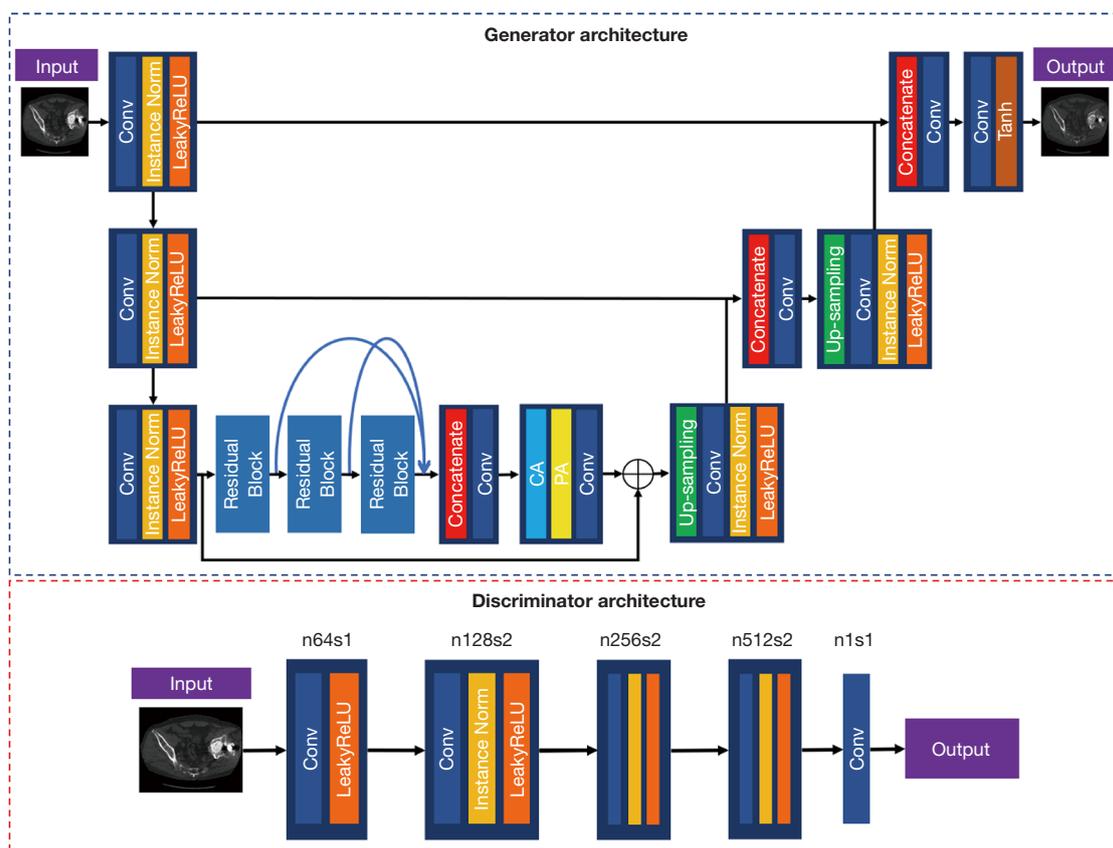


Figure 2 Network architecture.

resized to 256×256 pixel, a 128×128 pixel size region is randomly cropped from each pair of resized images. Random cropping can effectively prevent overfitting during network training. The cropped images are then sent to the network for training. Each pair of patches is put into the 2 generators. During the prediction phase, the low-energy CT images are input into the trained low-to-high generator model, which synthesizes high-energy CT images. At the prediction phase, the input images can be of any size.

Network architecture

The final model is based on the cyclic structure of the CycleGAN (26). The model adopts features from both U-Net (30) and DenseNet (31) and also applies the attention mechanism (32-34). This model aims to enhance the feature mapping capabilities between LECT and HECT images and to improve the stability of training. *Figure 2* shows the network architecture of the generators and discriminators of the proposed method in detail.

In the proposed generator network, we implement a U-shaped encoder-decoder structure, a popular convolutional neural network architecture in medical imaging. Specifically, there are 3 long connections with advanced features between this encoder and the decoder, in which different scale feature information is combined. In the encoder component, the down-sampling convolutional layer contains a convolution block (3×3 kernels with a stride of 2), an instance normalization (35) operation, and a leaky rectified linear unit (36) (LReLU) activation function. The input images fed to the backbone network after having also being fed to the 3 down-sampling layers.

The backbone network contains a skip connection and three residual blocks, with each residual block containing 3 basic blocks and 1 convolutional layer. A basic block consists of 2 convolutional layers, a feature attention module, and residual connections, with the entire block being used to extract image feature information. The low-level features are bypassed through the residual connection and combined with high-level features. The merged feature

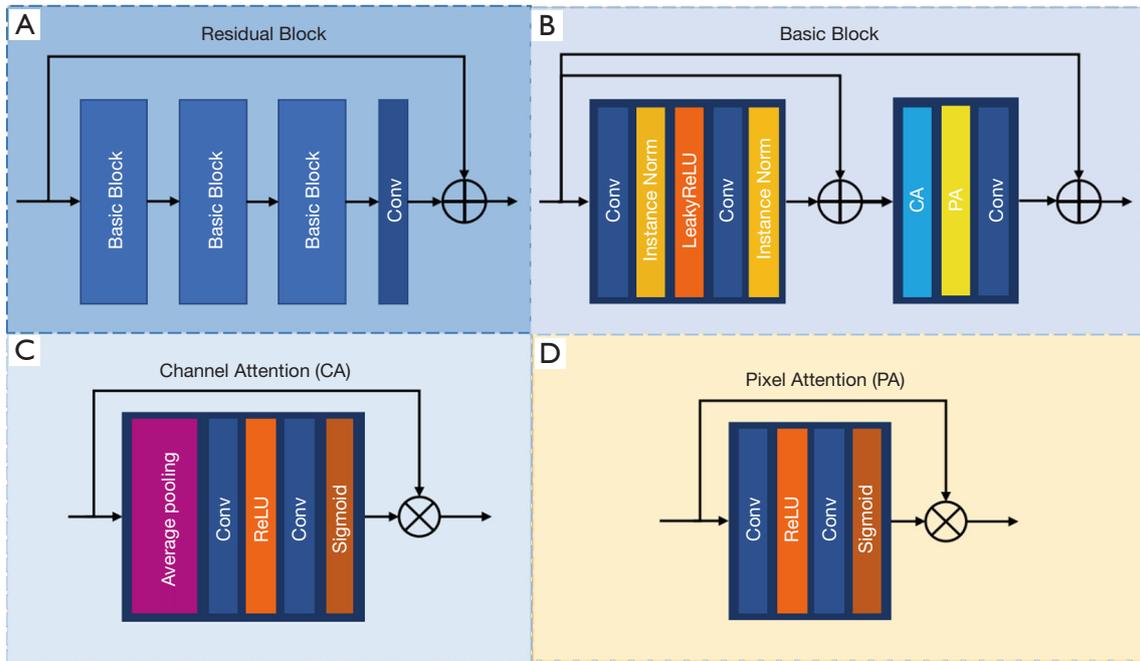


Figure 3 Details of the backbone structure.

maps are then transported to the feature attention blocks, which consist of channel attention (CA) and pixel attention (PA) modules. During the feature extraction process, the feature maps on some channels contain more important feature information than do others, with some pixels on the feature maps representing key feature information. The detailed structures of the backbone network are shown in *Figure 3*.

The decoder is connected behind the backbone network and consists of 2 decoding modules with upsampling operation capability, which doubles the size of the feature map. The connected feature maps are fused with upsampled results during this process. Finally, the output layer uses a convolutional layer and an activation function to inversely transform the input of the CT image.

The discriminator’s architecture adopts a patchGAN structure, which contains 5 convolutional layers. The first convolutional layers gradually reduce the input image size to 1/16th of the original image size and gradually increase the number of channels of the feature map to 512 via the middle convolution layers. The final convolutional layer combines all the feature maps to output a result. The patch size is set to 70×70 pixels, which effectively improves the efficiency of the discriminator.

Attention mechanism

The attention mechanism emphasizes the aforementioned key feature maps and pixel information by adjusting their weights, thereby causing the network to focus more on learning the key items of information. Some previous studies (32-34) in image restoration have introduced attention mechanisms and achieved excellent results. Learning from these previous studies, we chose to introduce attention mechanisms to the proposed network. The feature attention module includes both CA and PA, with the CA adopting global average pooling to adjust the weights of different channel feature maps:

$$g_c = \frac{1}{H \times W} \sum_{i=1}^H \sum_{j=1}^W f_{c-l}(i, j) \tag{1}$$

where $f_{c-l}(i, j)$ represents the value of the l -th channel of feature map f_c at (i, j) and the size of the feature maps is changed from $H \times W$ to 1×1 . Then, we obtain the channel weight W_{ca} as follows:

$$W_{ca} = \text{sigmoid} \left(\text{Conv} \left(\text{ReLU} \left(\text{Conv} (g_c) \right) \right) \right) \tag{2}$$

Finally, f_c' is multiplied elementwise with the channel weight W_{ca} :

$$f'_c = W_{ca} \otimes f_c \tag{3}$$

The PA module directly processes the input feature maps f_p through sigmoid and ReLU functions to change the number of feature maps to 1 and obtain the pixel weight W_{pa} :

$$W_{pa} = \text{sigmoid}\left(\text{Conv}\left(\text{ReLU}\left(\text{Conv}\left(f_p\right)\right)\right)\right) \tag{4}$$

Then, f_p and W_{pa} are multiplied elementwise to obtain the result f'_p :

$$f'_p = W_{pa} \otimes f_p \tag{5}$$

Loss function

The proposed method employs adversarial loss, cycle consistency loss, and reconstruction loss to correct the conversion between X the Y and image domains, where the data distributions are denoted as $x \sim P_{data}(x)$ and $y \sim P_{data}(y)$. LECT images are defined as the image domain X , and the HECT images are defined as the image domain Y .

The adversarial loss function of the generators and discriminators of the GAN and CycleGAN can be expressed as follows:

$$\mathcal{L}_{GAN-G}(G, D, X, Y) = E_{x \sim P_{data}(x)} \left[\log(D(G(x))) \right] \tag{6}$$

and:

$$\mathcal{L}_{GAN-D}(G, D, X, Y) = E_{x \sim P_{data}(x)} \left[\log(1 - D(G(x))) \right] + E_{y \sim P_{data}(y)} \left[\log(D(y)) \right] \tag{7}$$

where $E_{x \sim P_{data}(x)}$ represents the mathematical expectation when the input data are LECT images, with $E_{y \sim P_{data}(y)}$ denoting a similar meaning for HECT images. G represents a generator, D represents a discriminator, $G(x)$ represents the generated results, and $D(G(x))$ and $D(y)$ represent the discrimination results after inputting both a generated image and a real image into the discriminator. The entire adversarial loss function is obtained by adding both Eqs. [6] and [7]:

$$\mathcal{L}_{GAN}(G, D, X, Y) = \mathcal{L}_{GAN-G}(G, D, X, Y) + \mathcal{L}_{GAN-D}(G, D, X, Y) \tag{8}$$

As the entire adversarial loss function shows, the purpose of G is to produce results as close to the real image as possible, while the purpose of D is to distinguish $G(x)$ from y as often as possible.

Our work adopts the idea of least squares GAN

(LSGAN) (37). In doing so, we modified the cross entropy in the traditional adversarial loss function to calculate the distance between the result and the label, thus imposing greater penalties on generated data that are distant from the real data. This obtains more accurate generated images and improves the stability of training.

The generator $G_{low-high}$ and discriminator D_{high} have different adversarial loss functions. These adversarial loss function can be formulated as follows:

$$\mathcal{L}_{adv-G}(G_{low-high}, D_{high}, X, Y) = E_{x \sim P_{data}(x)} \left[\text{MSE}\left(D_{high}\left(G_{low-high}(x)\right), 1\right) \right] \tag{9}$$

where $D_{high}(G_{low-high}(x))$ denotes the judged result of the discriminator D_{high} for the generated image $G_{low-high}(x)$, and $\text{MSE}(D_{high}(G_{low-high}(x)), 1)$ denotes the mean squared error between $D_{high}(G_{low-high}(x))$ and the label of target image 1, thus measuring the closeness between the generated image and the target image. $E_{x \sim P_{data}(x)}$ and $E_{y \sim P_{data}(y)}$ represent the mathematical expectation. The adversarial loss function of D_{high} is formulated as follows:

$$\mathcal{L}_{adv-D}(D_{high}, X, Y) = \frac{1}{2} E_{x \sim P_{data}(x)} \left[\text{MSE}\left(D_{high}\left(G_{low-high}(x)\right), 0\right) \right] + \frac{1}{2} E_{y \sim P_{data}(y)} \left[\text{MSE}\left(D_{high}(y), 1\right) \right] \tag{10}$$

Where $D_{high}(y)$ denotes the judged result for the target image y , and $\text{MSE}(D_{high}(G_{low-high}(x)), 0)$ and $\text{MSE}(D_{high}(y), 1)$ denote the mean squared error between $D_{high}(G_{low-high}(x))$ and the label of generated image 0 and the mean squared error between $D_{high}(y)$ and the label of target image 1, respectively. These loss functions measure the accuracy with which the discriminator can distinguish between the generated and target images.

Correspondingly, the adversarial loss function of the generator $G_{high-low}$ and the discriminator D_{low} are respectively formulated as follows:

$$\mathcal{L}_{adv-G}(G_{high-low}, D_{low}, X, Y) = E_{y \sim P_{data}(y)} \left[\text{MSE}\left(D_{low}\left(G_{high-low}(y)\right), 1\right) \right] \tag{11}$$

$$\mathcal{L}_{adv-D}(D_{low}, X, Y) = \frac{1}{2} E_{y \sim P_{data}(y)} \left[\text{MSE}\left(D_{low}\left(G_{high-low}(y)\right), 0\right) \right] + \frac{1}{2} E_{x \sim P_{data}(x)} \left[\text{MSE}\left(D_{low}(x), 1\right) \right] \tag{12}$$

The meanings of these two equations are similar to those of the previous loss functions. Therefore, the entire adversarial loss function can be denoted as follows:

$$\begin{aligned} \mathcal{L}_{adv}(G, D, X, Y) = & \mathcal{L}_{adv-G}(G_{low-high}, D_{high}, X, Y) \\ & + \mathcal{L}_{adv-G}(G_{high-low}, D_{low}, X, Y) \\ & + \mathcal{L}_{adv-D}(D_{high}, X, Y) \\ & + \mathcal{L}_{adv-D}(D_{low}, X, Y) \end{aligned} \quad [13]$$

In theory, the generator can learn a variety of mappings between image domains X and Y ; consequently, applying only the adversarial loss may not guarantee that the generator has learned accurate mapping or is outputting desired results. Thus, the proposed method adopts the structure of CycleGAN and introduces 2 generators to generate cycle images, thus improving the learning accuracy. We define the cycle image x_{cyc} as the result of processing input x through both $G_{low-high}$ and $G_{high-low}$, described as $G_{high-low}(G_{low-high}(x))=x_{cyc}$. Correspondingly, the cycle image y_{cyc} can be described as $G_{high-low}(G_{low-high}(y))=y_{cyc}$.

Generating cyclic-consistent images through a bidirectional cycle reduces the probability space of the mapping functions, thereby guaranteeing that the generators learn precise mapping functions. The cycle consistency loss is used to correct cycle image generation and uses the L1 norm to calculate the difference between the source image and cycle image. The specific equation is formulated as follows:

$$\begin{aligned} \mathcal{L}_{cyc}(G_{low-high}, G_{high-low}) = & E_{x \sim p_{data}(x)} [\|x_{cyc} - x\|_1] \\ & + E_{y \sim p_{data}(y)} [\|y_{cyc} - y\|_1] \end{aligned} \quad [14]$$

Reconstruction loss was introduced to more directly and accurately measure the differences between the generated image and the target image during the network training. The work from Lim *et al.* (38) indicates that models for image restoration tasks trained using L1 loss achieve better results with respect to image quality. Therefore, we also chose to employ L1 loss for our proposed method. The reconstruction loss can be formulated as follows:

$$\begin{aligned} \mathcal{L}_{rec}(G_{low-high}, G_{high-low}) = & E_{(x \sim p_{data}(x), y \sim p_{data}(y))} [G_{low-high}(x) - y_1] \\ & + E_{(x \sim p_{data}(x), y \sim p_{data}(y))} [G_{high-low}(y) - x_1] \end{aligned} \quad [15]$$

The inclusion of reconstruction loss directly calculates the differences in pixel values and then returns them to the network, which prompts the network to generate a more accurate pixel value distribution during the next training cycle.

In summary, the complete loss function is as follows:

$$\begin{aligned} \mathcal{L}(G, D, X, Y) = & \mathcal{L}_{adv}(G, D, X, Y) \\ & + \lambda_{cyc} \mathcal{L}_{cyc}(G_{low-high}, G_{high-low}) \\ & + \lambda_{rec} \mathcal{L}_{rec}(G_{low-high}, G_{high-low}) \end{aligned} \quad [16]$$

where λ_{cyc} and λ_{rec} are hyperparameters that represent the weights of the cycle consistency loss and reconstruction loss, respectively.

Image evaluation

Referring to previously conducted studies, we use peak signal-to-noise ratio (PSNR), structural similarity index (SSIM), and normalized mean square error (NMSE) as metrics to quantify the performance of the proposed method. The generated image is defined as I_g , the target image is defined as I_t , and the PSNR, SSIM, and NMSE are denoted as follows:

$$PSNR(I_g, I_t) = 10 \cdot \log_{10} \left(\frac{MAX_I^2}{\|I_g - I_t\|_2^2} \right) \quad [17]$$

$$SSIM(I_g, I_t) = \frac{(2\mu_g \mu_t + c_1)(2\sigma_{g,t} + c_2)}{(\mu_g^2 + \mu_t^2 + c_1)(\sigma_g^2 + \sigma_t^2 + c_2)} \quad [18]$$

and:

$$NMSE(I_g, I_t) = \frac{\|I_g - I_t\|_2^2}{\|I_t\|_2^2} \quad [19]$$

Where MAX_I^2 represents the maximum value of the image pixels (255 pixels in an 8-bit image). The parameters μ_g and μ_t are the pixel averages of I_g and I_t ; σ_g^2 and σ_t^2 are the variances of I_g and I_t , respectively. $\sigma_{g,t}$ is the covariance of I_g and I_t . The parameters $C_1=k_1 \times L$ and $C_2=k_2 \times L$ are used to maintain stability. L is the dynamic range of the image pixels, set as $k_1=0.01$ and $k_2=0.03$. For the PSNR and SSIM, higher values indicate better results. For the NMSE, lower values indicate smaller pixel differences between the generated and target images, which indicates a better performance. In addition, the results of the metrics between the CT images generated using the proposed method and the ground truth CT images were compared with 4 other methods via paired Student's t test. The threshold of significance was set at 5% ($P<0.05$).

Materials and experimental setup

For this study, with the authorization of Guizhou Province

Table 1 Quantitative results of different methods on the Siemens dual-source CT test data set

Methods	PSNR	SSIM	NMSE
GAN	25.49*	0.5968*	0.0819*
RED-VGG	26.01*	0.6155*	0.0756*
WGAN	26.44*	0.6306*	0.0654*
CycleGAN	26.57*	0.6343*	0.0651*
Proposed	27.22*	0.6452*	0.0552*

*, denotes $P < 0.05$, corresponding to a significant difference. PSNR, peak signal-to-noise ratio; SSIM, structural similarity index measure; NMSE, normalized mean square error; GAN, generative adversarial network; RED-VGG, a residual encoder-to-decoder network with a visual geometry group pretrained model; WGAN, Wasserstein generative adversarial network; CycleGAN, cycle generative adversarial network.

People's Hospital, we used the Hospital's clinical CT data set to evaluate the performance of the proposed method. After collecting data from different patients, Guizhou Province People's Hospital reconstructed the corresponding 2D slice images to compile the clinical CT data set. The data set contained 14,617 pairs of 2D slice images from the hip CT scans of 33 patients. The raw sinogram data were acquired using a Siemens SOMATOM Force open-source CT scanner. For the training set, 13,000 pairs of images with obvious metal artifacts were selected. From the CT images of the remaining 8 patients, 1,000 pairs of images with obvious metal artifacts were also selected as the test set. For scanning with DECT, 2 X-ray tubes are set to different tube voltages and currents, so they scan and image at different energies. The tube voltage and current scanning parameters of the HECT scan are set to 150 kV and 200 mA, respectively, and the tube voltage and current scanning parameters of the LECT scan are set to 100 kV and 100 mA, respectively. The tube voltage determines the "hardness" of the X-rays (photon energy), and the product of the tube current and exposure time determines the amount of radiation (the number of photons emitted), with the product of these 3 factors referred to as the energy of the X-rays. Reducing the tube voltage correspondingly decreases the energy of the X-rays, which also reduces the radiation dose. The relationship between radiation dose and tube voltage has been previously discussed (39-41). When CT projection data are being acquired, each patient is scanned only once. During the scanning process, the LECT and HECT scans are performed simultaneously so that they obtain corresponding data, which are then used to reconstruct the image and avoid mismatches between the obtained LECT and HECT images. Our study was conducted in accordance with the Declaration of Helsinki

(as revised in 2013) and was approved by Guizhou Province People's Hospital. Informed consent was received from all included patients.

We set the hyperparameters λ_{cyc} and λ_{rec} in the loss function to 1 and 10, respectively. The proposed method used the Adam optimization algorithm (42) to adjust the network parameters, for which β_1 and β_2 are set to 0.5 and 0.999, respectively. The initial learning rate was set to 1×10^{-4} for the first 100 epochs and linearly decayed to 0 over the next 60 epochs. The input image batch size was 8. We implemented the models from PyTorch and ran them on a computer equipped with an NVIDIA GeForce GTX 1080Ti GPU (11.0 GB).

Results

To evaluate the performances of the proposed method, we compared its results with a GAN, a residual encoder-to-decoder network with a visual geometry group (VGG) pretrained model (43) (RED-VGG), a Wasserstein GAN (44) (WGAN), and a CycleGAN. *Table 1* summarizes the average measurement results for the different methods for the entire test set. The proposed method obtained a higher score on every index compared to the other methods. Furthermore, when compared with the better results from other algorithms, the proposed method improved the PSNR by 2.44%, the SSIM by 1.71%, and the NMSE by 15.2%. These differences were statistically significant.

To illustrate the performances of the different methods in more detail, *Figure 4* shows 2 test results from the original CT images of the left and right hips with different degrees of metal artifacts along with the generated results of different methods in the left and right hip images. The proposed method is observably superior to the other

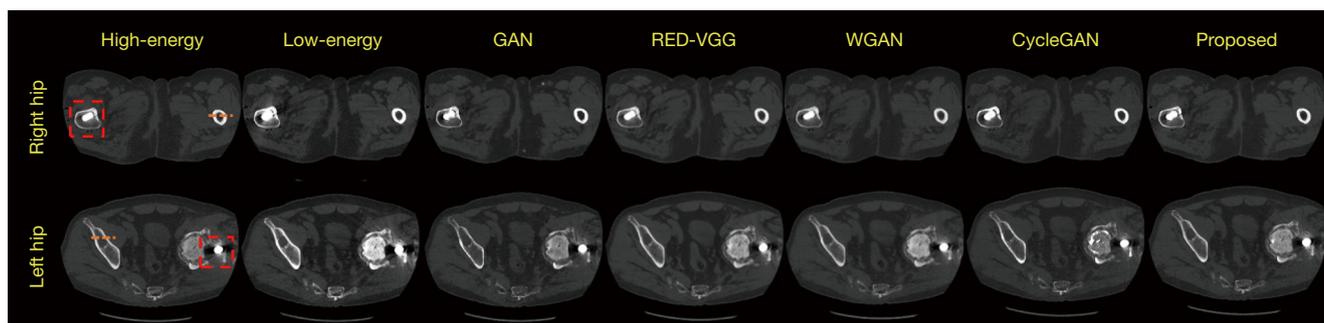


Figure 4 High- and low-energy CT images of the left and right hips. GAN, generative adversarial network; RED-VGG, a residual encoder-to-decoder network with a visual geometry group pretrained model; WGAN, Wasserstein generative adversarial network; CycleGAN, cycle generative adversarial network.

methods at preserving the structural information of other human tissues.

Figure 5 shows the profile results of the different methods at the position marked by the orange line in Figure 4. Based on the results of the left and right hips, the profile generated by the proposed method is closer to the target result. In terms of image quality, the values of the metrics in Table 2 indicate that the images generated by the proposed method in Figure 4 are more similar to an HECT image than are the other methods.

In addition to the restoration effect of the entire image, we are similarly concerned with restoring the details of the region around a metal implant in an image that has been obscured by metal artifacts. Therefore, further observations regarding the region of interest (ROI) were made. The ROI in the HECT image of the right hip in Figure 4 is denoted as ROI I, and the ROI in the HECT image of the left hip in Figure 4 denoted as ROI II. Figure 6 shows an enlarged view of the ROIs. The main metal artifacts in the LECT image are indicated by arrows; the other images use arrows to indicate the same region for comparison.

A comparison between the ROI I images shows that all 5 methods effectively reduced the metal artifacts in the areas indicated by the yellow arrows. However, the RED-VGG result was much poorer due to oversmoothing, causing considerable loss of texture information. The GAN model showed the same problem, and while WGAN was shown to effectively remove artifacts, there are changes to the texture of the CT images giving a less than satisfactory result. In the area indicated by the topmost yellow arrow, the proposed method more accurately distinguished the shape of the metal implant, with the effect of metal artifact reduction more closely resembling the target image. In addition, in the area

indicated by the red arrow, both the GAN and CycleGAN generated new artifacts, leading to obvious errors in the status information of the metal implant. The proposed method generated no artifacts in the same area; it retained the status information of the metal implant more completely, signifying its superior clinical application.

The ROI II images in Figure 6 show obvious metal artifacts exist in the HECT image, with the metal artifacts in ROI II shown to be much more observable than those in ROI I. As the main goal of this study was to synthesize HECT images, it was critical to determine which method generates a greater metal artifact reduction that more closely resembles a HECT result.

The overall results of the RED-VGG-, GAN-, and WGAN-generated images have oversmoothing effects, with the restoration effect in the areas indicated by the yellow arrows differing substantially from that of the target image. The generated results of both CycleGAN and the proposed method are not oversmoothed, and the restoration effects are similar to the areas indicated by the yellow arrows. However, in the area indicated by the red arrow, an obvious metal artifact appears in the generated result of CycleGAN, while the WGAN and proposed method are shown to accurately reduce the metal artifact at that location. WGAN-generated images are similar to the results of the proposed method in both removing metal artifacts and expressing the contrast between bones and muscles; however, they have more blurred edge contours.

This result is further confirmed by Figure 7, which shows that the difference between the synthetic high-energy CT images generated by the proposed method and the target high-energy CT images is minimal and evenly distributed, indicating that their data distributions are highly similar.

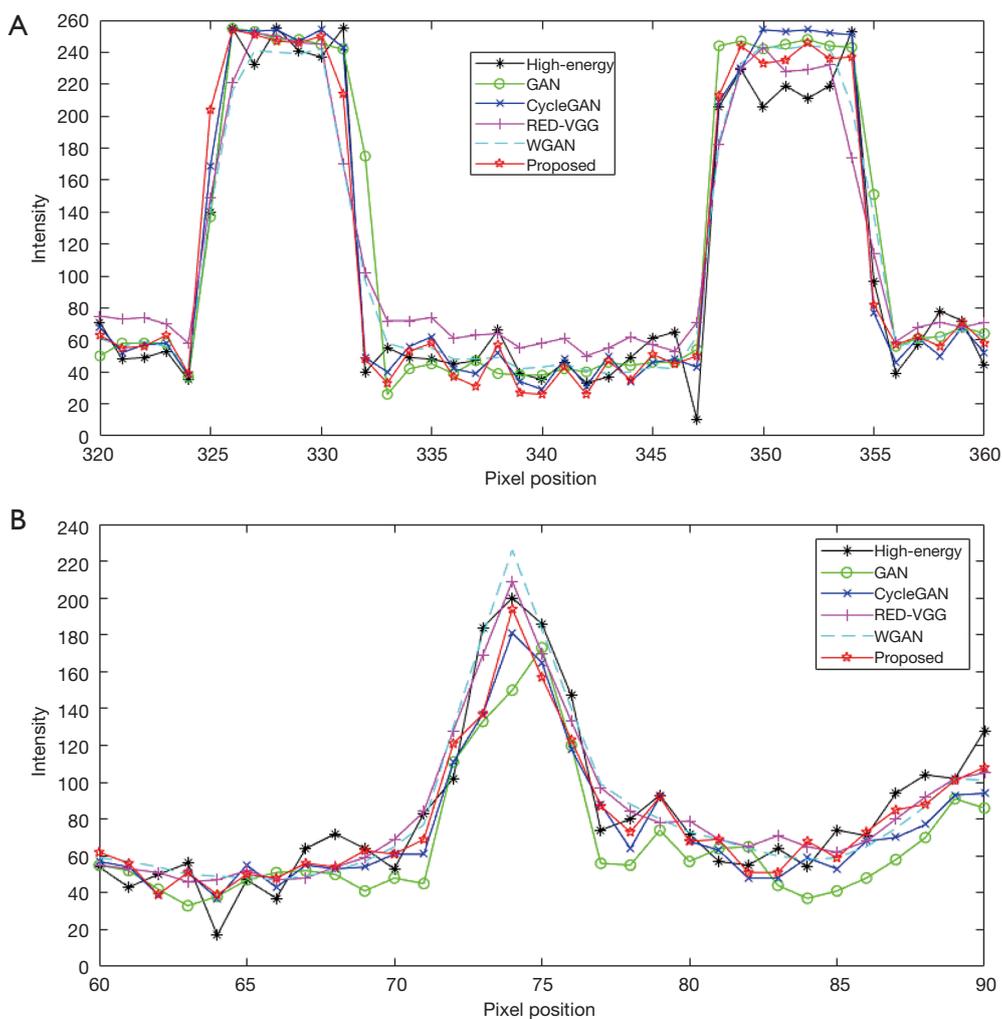


Figure 5 Profile result comparison of the different methods for a sample case of the right and left hips (orange outline in *Figure 4*). The methods represented by the different lines are shown in the legend. GAN, generative adversarial network; RED-VGG, a residual encoder-to-decoder network with a visual geometry group pretrained model; WGAN, Wasserstein generative adversarial network; CycleGAN, cycle generative adversarial network.

Table 2 Quantitative results of different methods in *Figure 4*

Methods	Right hip case			Left hip case		
	PSNR	SSIM	NMSE	PSNR	SSIM	NMSE
GAN	27.08	0.6455	0.0663	24.74	0.5660	0.0792
RED-VGG	27.39	0.6465	0.0621	24.97	0.5709	0.0783
WGAN	27.47	0.6537	0.0599	25.14	0.5834	0.0779
CycleGAN	27.58	0.6540	0.0591	24.83	0.5900	0.0776
Proposed	27.95	0.6639	0.0539	25.86	0.6004	0.0612

PSNR, peak signal-to-noise ratio; structural similarity index measure, normalized mean square error; GAN, generative adversarial network; RED-VGG, a residual encoder-to-decoder network with a visual geometry group pretrained model; WGAN, Wasserstein generative adversarial network; CycleGAN, cycle generative adversarial network.

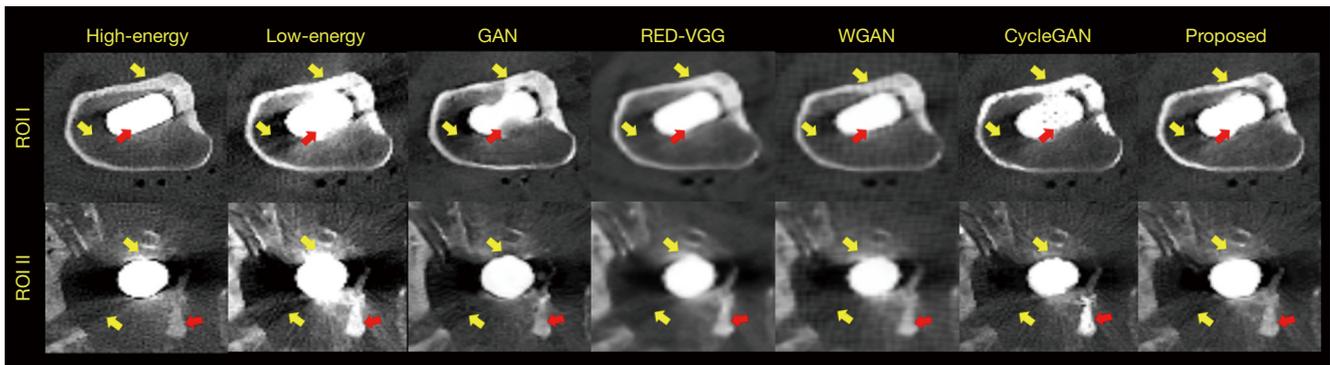


Figure 6 Enlarged views of the ROIs marked by the red-dotted box in *Figure 4*. ROI, the region of interest; GAN, generative adversarial network; RED-VGG, a residual encoder-to-decoder network with a visual geometry group pretrained model; WGAN, Wasserstein generative adversarial network; CycleGAN, cycle generative adversarial network.

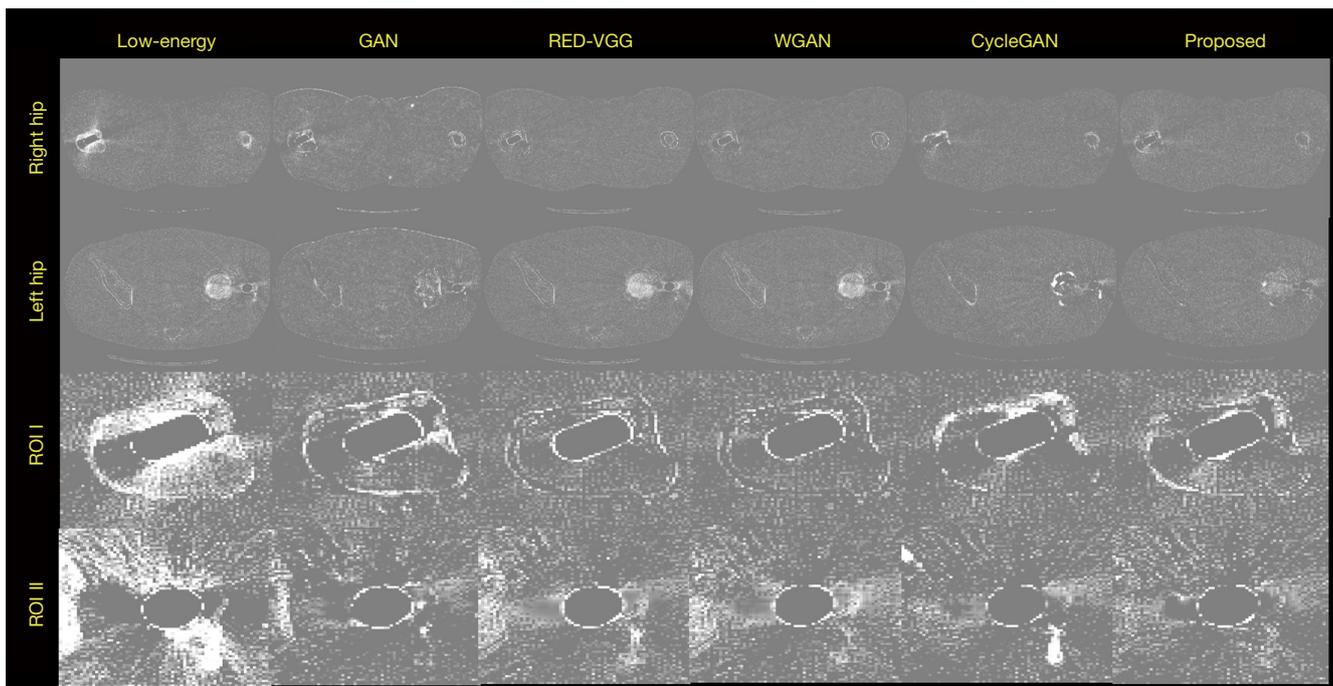


Figure 7 Enlarged difference images of the ROIs between the high-energy CT images and the other images. ROI, the region of interest; GAN, generative adversarial network; RED-VGG, a residual encoder-to-decoder network with a visual geometry group pretrained model; WGAN, Wasserstein generative adversarial network; CycleGAN, cycle generative adversarial network.

In contrast, noticeable differences exist in the distributions of data points between the images generated by the other methods and the HECT image.

Table 3 shows the image processing time required by the different methods. This comparison shows that CycleGAN was the fastest in processing each image, followed by proposed method; in contrast, the WGAN method was the

slowest and thus the most -time-consuming. Although there are differences in the computation times of the 5 methods, they all meet the time requirement for practical clinical application. The inference speed of the proposed model is in the same order of magnitude as the inference speeds of all methods, and they all have acceptable speeds for clinical applications.

Discussion

The proposed method derives HECT images from LECT images, reducing both the cost of CT usage and the radiation dose received by patients. Furthermore, the proposed method processes only image domain data, which is more convenient and avoids the difficulties involved in acquiring and processing sinogram domain data.

The combined results from *Table 1* and *Table 3* demonstrate that the proposed model provides significant

and stable improvements in quantitative results compared with the other methods. The inference speed of the proposed model is in the same order of magnitude as the inference speeds of all the other methods, with all possessing acceptable speeds for clinical application. *Figures 6-8* show that our improvement strategy is effective. We introduced adversarial, cycle consistency, and reconstruction losses to ensure the correctness of the opposing mappings between the 2 generators during model training, which improved the training stability. Our method effectively solves the oversmoothing phenomenon seen in RED-VGG- and GAN-generated images, while both CycleGAN and the proposed model display clearer edge details than does the WGAN model. We then combined the multilevel feature connection and image attention mechanisms in the cycle structure of CycleGAN to improve the quality of the generated CT images, which can, for example, improve the denoising result and contrast of tissue structures. Thus, it can be concluded that the proposed method generates more effective images than do the other described methods. In addition, regarding the processing effect of the proposed method on the different shapes and sizes of metal implants of the left and right hips in *Figure 4*, the proposed method can be applied to generalized cases with a variety of metal implant shapes and sizes.

Table 3 Average computation time required for image processing by the different methods

Methods	Computation time (ms/image)
GAN	83.6
RED-VGG	52.9
WGAN	87.9
CycleGAN	39.6
Proposed	61.5

GAN, generative adversarial network; RED-VGG, a residual encoder-to-decoder network with a visual geometry group pretrained model; WGAN, Wasserstein generative adversarial network; CycleGAN, cycle generative adversarial network.

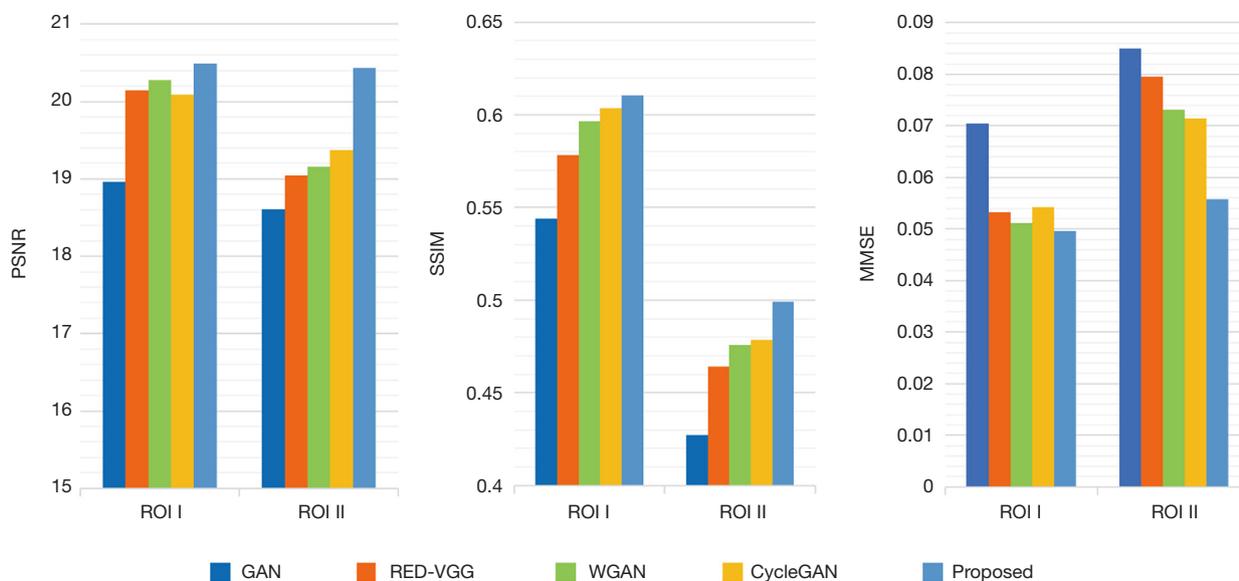


Figure 8 Quantitative results of different methods over the marked in ROIs in *Figure 6*. ROI, the region of interest; GAN, generative adversarial network; RED-VGG, a residual encoder-to-decoder network with a visual geometry group pretrained model; WGAN, Wasserstein generative adversarial network; CycleGAN, cycle generative adversarial network.

The limitation of our study is that it only focused on reconstructed hip CT images; however, the proposed method is not only confined to hip CT images as this model could also be applied to CT images of other body parts, such as knees or oral cavities, according to the data sets used to train the model. In future experiments, we plan to use additional data sets to verify the effectiveness of the algorithm and adjust the network structure to further improve the performance of the proposed model. Furthermore, our future studies will aim to optimize the network model to further explore its learning capabilities and extend its use to other imaging tasks, such as 3D reconstruction.

Conclusions

We propose a method of generating synthetic HECT images directly from LECT images with the goal of providing additional clinical information. By employing the proposed method, the effect of dual-energy CT can be synthesized from a low-energy scan, with the radiation doses patients receive being significantly reduced during CT examinations. The proposed model extracts accurate edge and texture features, and its cyclic structure condenses the mapping between images to obtain higher quality generated images. The trained model can be executed quickly, reducing the time required for image processing and enhancing its suitability for clinical application. This method is easily implemented, can be applied in a variety of environments, and produces results that can be obtained by inputting images directly into the pretrained model. Based on both image quality score metrics and visual effects comparisons, the proposed method achieved superior results and demonstrated its considerable potential.

Acknowledgments

The authors would like to thank the editor and anonymous reviewers for their constructive comments and suggestions.

Funding: This work was supported by the Shenzhen International Cooperation Research Project of China (GJHZ20180928115824168), the Guangdong International Science and Technology Cooperation Project of China (2018A050506064), the Natural Science Foundation of Guangdong Province in China (2020A1515010734), the Guangdong Special Support Program of China (2017TQ04R395), Basic research and free exploration project of Shenzhen Science and technology innovation

Commission of China (JCYJ20180305125332754), Natural Science Foundation of Shenzhen University General Hospital in China (SUGH2018QD009).

Footnote

Conflicts of Interest: All authors have completed the ICMJE uniform disclosure form (available at <https://dx.doi.org/10.21037/qims-21-182>). 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. Our study was conducted in accordance with the Declaration of Helsinki (as revised in 2013) and was approved by Guizhou Province People's Hospital. Informed consent was received from all included patients.

Open Access Statement: This is an Open Access article distributed in accordance with the Creative Commons Attribution-NonCommercial-NoDerivs 4.0 International License (CC BY-NC-ND 4.0), which permits the non-commercial replication and distribution of the article with the strict proviso that no changes or edits are made and the original work is properly cited (including links to both the formal publication through the relevant DOI and the license). See: <https://creativecommons.org/licenses/by-nc-nd/4.0/>.

References

1. Brenner DJ, Hall EJ. Computed tomography--an increasing source of radiation exposure. *N Engl J Med* 2007;357:2277-84.
2. Schenzle JC, Sommer WH, Neumaier K, Michalski G, Lechel U, Nikolaou K, Becker CR, Reiser MF, Johnson TR. Dual energy CT of the chest: how about the dose? *Invest Radiol* 2010;45:347-53.
3. Balda M, Hornegger J, Heismann B. Ray contribution masks for structure adaptive sinogram filtering. *IEEE Trans Med Imaging* 2012;31:1228-39.
4. Huang Z, Chen Z, Chen J, Lu P, Quan G, Du Y, Li C, Gu Z, Yang Y, Liu X, Zheng H, Liang D, Hu Z. DaNet: dose-aware network embedded with dose-level estimation for low-dose CT imaging. *Phys Med Biol* 2021;66:015005.
5. Wang J, Li T, Lu H, Liang Z. Penalized weighted least-squares approach to sinogram noise reduction and image

- reconstruction for low-dose X-ray computed tomography. *IEEE Trans Med Imaging* 2006;25:1272-83.
6. Zhang H, Li L, Wang L, Sun Y, Yan B, Cai A, Hu G. Computed tomography sinogram inpainting with compound prior modelling both sinogram and image sparsity. *IEEE Transactions on Nuclear Science* 2016;63:2567-76.
 7. Fu J, Feng F, Quan H, Wan Q, Chen Z, Liu X, Zheng H, Liang D, Cheng G, Hu Z. PWLS-PR: low-dose computed tomography image reconstruction using a patch-based regularization method based on the penalized weighted least squares total variation approach. *Quant Imaging Med Surg* 2021;11:2541-59.
 8. Hu Z, Zhang Y, Liu J, Ma J, Zheng H, Liang D. A feature refinement approach for statistical interior CT reconstruction. *Phys Med Biol* 2016;61:5311-34.
 9. Sidky EY, Pan X. Image reconstruction in circular cone-beam computed tomography by constrained, total-variation minimization. *Phys Med Biol* 2008;53:4777-807.
 10. Zhang Y, Wang Y, Zhang W, Lin F, Pu Y, Zhou J. Statistical iterative reconstruction using adaptive fractional order regularization. *Biomed Opt Express* 2016;7:1015-29.
 11. Xu Q, Yu H, Mou X, Zhang L, Hsieh J, Wang G. Low-dose X-ray CT reconstruction via dictionary learning. *IEEE Trans Med Imaging* 2012;31:1682-97.
 12. Chen Y, Shi L, Feng Q, Yang J, Shu H, Luo L, Coatrieux JL, Chen W. Artifact suppressed dictionary learning for low-dose CT image processing. *IEEE Trans Med Imaging* 2014;33:2271-92.
 13. Zhang W, Gao J, Yang Y, Liang D, Liu X, Zheng H, Hu Z. Image reconstruction for positron emission tomography based on patch-based regularization and dictionary learning. *Med Phys* 2019;46:5014-26.
 14. Xu M, Hu D, Luo F, Liu F, Wang S, Wu W. Limited angle X ray CT reconstruction using image gradient l0 norm with dictionary learning. *IEEE Transactions on Radiation and Plasma Medical Sciences* 2020;5:78-87.
 15. Chen Y, Gao D, Nie C, Luo L, Chen W, Yin X, Lin Y. Bayesian statistical reconstruction for low-dose X-ray computed tomography using an adaptive-weighting nonlocal prior. *Comput Med Imaging Graph* 2009;33:495-500.
 16. Kang D, Slomka P, Nakazato R, Woo J, Berman DS, Kuo CCJ, Dey D, editors. Image denoising of low-radiation dose coronary CT angiography by an adaptive block-matching 3D algorithm. *Medical Imaging 2013: Image Processing*; 2013: International Society for Optics and Photonics.
 17. 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.
 18. 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.
 19. Chang S, Chen X, Duan J, Mou X. A CNN based Hybrid Ring Artifact Reduction Algorithm for CT Images. *IEEE Transactions on Radiation and Plasma Medical Sciences* 2021;5:253-60.
 20. 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.
 21. Hu Z, Xue H, Zhang Q, Gao J, Zhang N, Zou S, Teng Y, Liu X, Yang Y, Liang D. DPIR-Net: Direct PET image reconstruction based on the Wasserstein generative adversarial network. *IEEE Transactions on Radiation and Plasma Medical Sciences* 2020 35-43.
 22. Xie S, Yang T. Artifact Removal in Sparse-Angle CT Based on Feature Fusion Residual Network. *IEEE Transactions on Radiation and Plasma Medical Sciences* 2020.
 23. Liao Y, Wang Y, Li S, He J, Zeng D, Bian Z, Ma J, editors. Pseudo dual energy CT imaging using deep learning-based framework: basic material estimation. *Medical Imaging 2018: Physics of Medical Imaging*; 2018: International Society for Optics and Photonics.
 24. Goodfellow I, Pouget-Abadie J, Mirza M, Xu B, Warde-Farley D, Ozair S, Courville A, Bengio Y, editors. Generative adversarial nets. *Advances in neural information processing systems*; 2014.
 25. Charyyev S, Wang T, Lei Y, Ghavidel B, Beitler JJ, McDonald M, Curran WJ, Liu T, Zhou J, Yang X. Learning-Based Synthetic Dual Energy CT Imaging from Single Energy CT for Stopping Power Ratio Calculation in Proton Radiation Therapy. *arXiv preprint arXiv:200512908* 2020.
 26. Zhu JY, Park T, Isola P, Efros AA, editors. Unpaired image-to-image translation using cycle-consistent adversarial networks. *Proceedings of the IEEE international conference on computer vision*; 2017.
 27. You C, Li G, Zhang Y, Zhang X, Shan H, Li M, Ju S, Zhao Z, Zhang Z, Cong W, Vannier MW, Saha PK, Hoffman EA, Wang G. CT Super-Resolution GAN Constrained by the Identical, Residual, and Cycle

- Learning Ensemble (GAN-CIRCLE). *IEEE Trans Med Imaging* 2020;39:188-203.
28. Yang H, Sun J, Carass A, Zhao C, Lee J, Xu Z, Prince J. Unpaired brain MR-to-CT synthesis using a structure-constrained CycleGAN. *Deep Learning in Medical Image Analysis and Multimodal Learning for Clinical Decision Support*. Springer; 2018. p. 174-82.
 29. He K, Zhang X, Ren S, Sun J, editors. Deep residual learning for image recognition. *Proceedings of the IEEE conference on computer vision and pattern recognition*; 2016.
 30. Ronneberger O, Fischer P, Brox T, editors. U-net: Convolutional networks for biomedical image segmentation. *International Conference on Medical image computing and computer-assisted intervention*; 2015: Springer.
 31. Huang G, Liu Z, Van Der Maaten L, Weinberger KQ, editors. Densely connected convolutional networks. *Proceedings of the IEEE conference on computer vision and pattern recognition*; 2017.
 32. Zhang Y, Li K, Li K, Wang L, Zhong B, Fu Y, editors. Image super-resolution using very deep residual channel attention networks. *Proceedings of the European Conference on Computer Vision (ECCV)*; 2018.
 33. Anwar S, Barnes N, editors. Real image denoising with feature attention. *Proceedings of the IEEE International Conference on Computer Vision*; 2019.
 34. Qin X, Wang Z, Bai Y, Xie X, Jia H, editors. FFA-Net: Feature Fusion Attention Network for Single Image Dehazing. AAAI; 2020.
 35. Ulyanov D, Vedaldi A, Lempitsky V. Instance normalization: The missing ingredient for fast stylization. *arXiv preprint arXiv:160708022* 2016.
 36. Maas AL, Hannun AY, Ng AY, editors. Rectifier nonlinearities improve neural network acoustic models. *Proc. icml*; 2013.
 37. Mao X, Li Q, Xie H, Lau RY, Wang Z, Paul Smolley S, editors. Least squares generative adversarial networks. *Proceedings of the IEEE international conference on computer vision*; 2017.
 38. Lim B, Son S, Kim H, Nah S, Mu Lee K, editors. Enhanced deep residual networks for single image super-resolution. *Proceedings of the IEEE conference on computer vision and pattern recognition workshops*; 2017.
 39. Bischoff B, Hein F, Meyer T, Hadamitzky M, Martinoff S, Schömig A, Hausleiter J. Impact of a reduced tube voltage on CT angiography and radiation dose: results of the PROTECTION I study. *JACC Cardiovasc Imaging* 2009;2:940-6.
 40. LI R, LIN J. Effect of 80 kV low tube voltage on image quality and radiation dose at cerebral computed tomography angiography. *Chinese Journal of Medical Imaging* 2013;21:177-80.
 41. Nakaura T, Kidoh M, Sakaino N, Utsunomiya D, Oda S, Kawahara T, Harada K, Yamashita Y. Low contrast- and low radiation dose protocol for cardiac CT of thin adults at 256-row CT: usefulness of low tube voltage scans and the hybrid iterative reconstruction algorithm. *Int J Cardiovasc Imaging* 2013;29:913-23.
 42. Kingma DP, Ba J. Adam: A method for stochastic optimization. *arXiv preprint arXiv:1412.6980* 2014.
 43. Kim B, Han M, Shim H, Baek J. A performance comparison of convolutional neural network-based image denoising methods: The effect of loss functions on low-dose CT images. *Med Phys* 2019;46:3906-23.
 44. 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.

Cite this article as: 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(1):28-42. doi: 10.21037/qims-21-182