Dynamic controllable residual generative adversarial network for low-dose computed tomography imaging

Zhenyu Xia[1], Jin Liu[1,2], Yanqin Kang[1,2], Yong Wang[1], Dianlin Hu[2,3], Yikun Zhang[2,3]

[1]School of Computer and Information, Anhui Polytechnic University, Wuhu, China; [2]Key Laboratory of Computer Network and Information Integration (Southeast University) Ministry of Education, Nanjing, China; [3]School of Computer Science and Engineering, Southeast University, Nanjing, China

Contributions: (I) Conception and design: Z Xia, J Liu; (II) Administrative support: J Liu, Y Kang, Y Wang; (III) Provision of study materials or patients: J Liu, Y Kang, Y Wang, D Hu, Y Zhang; (IV) Collection and assembly of data: Z Xia, J Liu, Y Kang; (V) Data analysis and interpretation: Z Xia, J Liu, Y Kang; (VI) Manuscript writing: All authors; (VII) Final approval of manuscript: All authors.

Correspondence to: Dr. Jin Liu. School of Computer and Information, Anhui Polytechnic University, No. 54, Middle Beijing Road, Wuhu 241000, China; Key Laboratory of Computer Network and Information Integration (Southeast University) Ministry of Education, No. 2, Sipailou, Nanjing 210096, China. Email: liujin@ahpu.edu.cn.

Background: Computed tomography (CT) imaging technology has become an indispensable auxiliary method in medical diagnosis and treatment. In mitigating the radiation damage caused by X-rays, low-dose computed tomography (LDCT) scanning is becoming more widely applied. However, LDCT scanning reduces the signal-to-noise ratio of the projection, and the resulting images suffer from serious streak artifacts and spot noise. In particular, the intensity of noise and artifacts varies significantly across different body parts under a single low-dose protocol.

Methods: To improve the quality of different degraded LDCT images in a unified framework, we developed a generative adversarial learning framework with a dynamic controllable residual. First, the generator network consists of the basic subnetwork and the conditional subnetwork. Inspired by the dynamic control strategy, we designed the basic subnetwork to adopt a residual architecture, with the conditional subnetwork providing weights to control the residual intensity. Second, we chose the Visual Geometry Group Network-128 (VGG-128) as the discriminator to improve the noise artifact suppression and feature retention ability of the generator. Additionally, a hybrid loss function was specifically designed, including the mean square error (MSE) loss, structural similarity index metric (SSIM) loss, adversarial loss, and gradient penalty (GP) loss.

Results: The results obtained on two datasets show the competitive performance of the proposed framework, with a 3.22 dB peak signal-to-noise ratio (PSNR) margin, 0.03 SSIM margin, and 0.2 contrast-to-noise ratio margin on the Challenge data and a 1.0 dB PSNR margin and 0.01 SSIM margin on the real data.

Conclusions: Experimental results demonstrated the competitive performance of the proposed method in terms of noise decrease, structural retention, and visual impression improvement.

Keywords: Low-dose computed tomography (LDCT); noise artifacts; dynamic controlled; generative adversarial network (GAN)

5272

Xia et al. Deep CNN for LDCT imaging

Introduction

X-ray computed tomography (CT) is a widely used medical imaging technique that can clearly display structural information and pathological conditions (1). However, exposure to X-ray CT results in potential cancer risk due to radiation damage. To ensure patient safety, a new method for clinical CT examinations is required that can reduce the radiation dose (2-4). This technique is desirable according to the "as low as reasonably achievable" (ALARA) principle (4,5). The procedure with a reduced X-ray radiation dose is called low-dose computed tomography (LDCT) imaging and can lead to severe noise artifact contamination in the reconstructed image, which can compromise most of clinical diagnostic tasks (6).

Over the past two decades, 3 major categories of LDCT imaging methods have been introduced, and all are currently being researched: (I) projection data filtering (7-9); (II) statistical iterative reconstruction (SIR) algorithms (10-12); and (III) CT image postprocessing (13-16). Alternative CT image postprocessing techniques that do not rely on raw projection and that can be easily integrated into the existing CT pipeline have been actively explored. Most image denoising and restoration methods can be directly applied to LDCT image postprocessing. For example, Schaap et al. proposed the anisotropic diffusion filter-based edge-preserving noise reduction technique for LDCT image processing (13), while Borsdorf et al. (14) proposed a wavelet correlation analysis method for noise suppression in LDCT images. In Watanabe et al. (15), a novel noise reduction and contrast enhancement method was proposed for lesion detection in abdomen LDCT images, while in Ma et al., a robust normal-dose, scan-induced, non-local-mean (NLM) filter method for preserving LDCT image spatial resolution and low-contrast structure was described (16). Kang et al. were the first to apply the block matching 3D (BM3D) to improve the quality of LDCT images (17). Chen et al., inspired by the theory of compressive sensing, adapted discriminative dictionary learning for LDCT noise artifact removal (18). Hasan et al. designed a blind source separation (BSS)-based method using multiframe LDCT image restoration (19). Despite their benefits, these methods generally fail to achieve a balance between image details, texture features, and noise artifact suppression. In processing noise and artifacts of different densities, oversmoothing or residual noise artifacts often occur. The difficulty in using postprocessing methods is due to the nonstationary noise and artifacts in LDCT images:

magnitudes may vary in different body parts and do not obey any specific distribution models.

More recently, deep learning (DL)-based methods have emerged as a promising alternative for LDCT image restoration. Without using strong prior models, these methods can automatically extract useful features for new data generation and achieve good performance over traditional algorithms (20). Many convolutional neural network (CNN)-based methods have been proposed for LDCT processing. For example, Kang et al. presented competitive performance from a lightweight CNN-based framework for LDCT processing (21). Inspired work in residual networks (ResNets), Chen et al. combined deconvolution and shortcut connections into an encoder-decoder CNN (22) that proved highly successful in LDCT image processing. Meanwhile, Zhang et al. proposed a denoising CNN (DNCNN) (23) that uses convolutional end-to-end residual learning to separate noise from noisy images. Collectively, these methods demonstrate the considerable potential of residual structures for image processing. However, the feature extraction ability of CNNs is restricted by datasets, hardware resources, and running time. To improve the feature-learning ability, Goodfellow et al. designed a generative model for directly drawing samples from the desired data distribution without requiring the explicit modeling of the underlying probability density function (24). With the help of discriminators, generative adversarial networks (GANs) can be a particularly valuable tool and have exhibited relatively excellent performance in medical image processing (25-30). For example, Radford et al. proposed a deep convolutional generative adversarial network (DCGAN) architecture that inputs noise data by series-upsampling operations and eventually generates an image from it, reducing the noise and preserving the edge details (25). Gulrajani et al. incorporated residual connections into both the generator and discriminator and designed a much deeper denoising network, with the experimental results showed that this technique can recover more structural details (26). Built upon on this effective strategy, the Wasserstein distance strategy and gradient penalty (GP) constraint were introduced into the GAN for cardiac CT image processing (27). To overcome the limitations of low-dose and routine-dose image pair-training preparation, the Noise2Noise model was proposed for LDCT image learning (28). To mitigate the problem of a lack of training data, a cycle-consistent GAN was proposed for LDCT denoising via the learning of unpaired image-to-image translation (29). Meanwhile, a similar

algorithm, Pix2Pix GAN (30) was developed for low-dose myocardial perfusion denoising.

However, the LDCT image degenerative process is related to the human tissue structure and does not obey any known noise distribution models in the image domain, making noise difficult to remove. For example, in one low-dose scanner protocol for the pelvis or shoulders, there is high attenuation of radiation passing in the lateral direction because of the scapulae and hipbone, resulting in serious streak artifacts in these areas but fewer artifacts in the lung. However, almost all network training requires matched training data pairs, and the degenerative process of testing data is the same as that of the training data (e.g., scanner protocol match, reconstruction algorithm match, and scanning region match). These factors can facilitate the acquisition of images useful for diagnosis under a unified processing framework, especially when using a low-dose scanner protocol, but remains challenging to implement (31). For example, Shan *et al*. designed a modularized encoder-decoder network based on the GAN learning framework that allows the radiologist to customize the depth for different sets of LDCT images (32); however, it is difficult to directly apply this method to real LDCT systems. By leveraging the attention mechanism, Liu *et al*. developed a multiscale reweighted strategy and convolutional coding neural network for LDCT processing (2). Although these studies involved network construction, the performance is limited for complex and image content prior-dependent degradation in LDCT image processing (33,34). In this study, we adopted a weight parameter to control the weighted sum of the residuals in ResNet and built a dynamic controllable ResNet as a generator. We chose the Visual Geometry Group Network-128 (VGG-128) network as a discriminator to improve the noise artifact suppression and feature retention ability. A hybrid loss function was specifically designed to improve network performance during the training process.

Methods

Our goal was to design a DL-based restoration model that takes the degraded LDCT image as inputs and outputs a high-quality processed image. Assuming that x_{LD} is the corresponding LDCT image and x_{LD} is the corresponding routine-dose CT (RDCT) image, their relation can be expressed by the following:

$$x_{LD} = T(x_{RD}) \qquad [1]$$

where $T(\bullet)$ represents a complex degradation process involving quantum noise, electronic noise, attenuation coefficient, error scattering, and other factors. These measurement data can be expressed by a complex Poisson noise model or Gaussian noise model (7,35,36). The attenuation coefficient integral along the X-ray path is the core factor that affects the noise intensity. The difference in the attenuation coefficient integral in different tissue parts is large. These differences often lead to mottle noise and streak artifacts during reconstruction. *Figure 1* depicts typical 2D views for a patient's LDCT images. The noise and artifact distribution of LDCT images is relatively complex, and different body parts have different levels of noise and artifact intensity. In the CT images of the shoulders and lungs, the streak artifact is obvious, and in the CT images of the abdomen, the speckle Poisson noise is more prominent than are the artifacts. Additionally, we can see that noise magnitudes in the LDCT images have nonstationary distributions throughout the tissue.

In this paper, we describe our work of developing a GAN with dynamic controllable residual (DR-GAN), which aims at modulating the degradation types and levels in LDCT images (with a principal focus on noise and artifacts). A flowchart of the proposed method is shown in *Figure 2*. To handle different noise and artifact distributions in LDCT images and improve the ResNet performance, the generator network consists of the basic subnetwork ResNet and the conditional subnetwork to control residual intensity via the degradation index module. In the GAN leaning framework, we chose the VGG-128 network as the discriminator to improve the noise artifact suppression and feature retention ability of the generator. Additionally, a hybrid loss function was specifically designed, which included the mean square error (MSE) loss, mean structural similarity index measure (MSSIM) loss, adversarial loss, and GP loss.

Dynamic controllable residual

A ResNet unit learns the local and global features via skip connections that combine different levels (37). One residual block can be expressed as follows:

$$Y = F(X, W_i) + X \qquad [2]$$

where X and Y are the input and output signal of the residual block, respectively; $F(X)$ denotes the residue feature mapping of the convolution layers; and W_i represents the convolution kernel of the i-th layer.
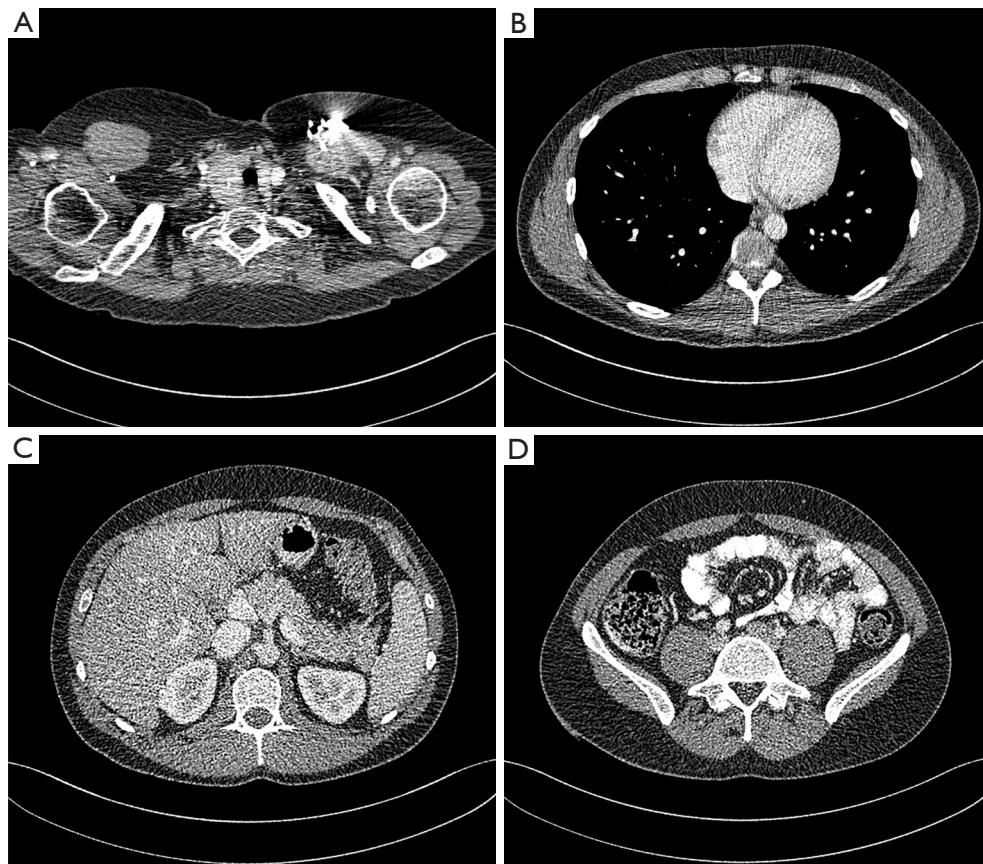
Figure 1 LDCT images of different body parts for the same scanner protocol. (A) Shoulder; (B) chest; (C) epigastrium; (D) hypogastrium. LDCT, low-dose computed tomography.
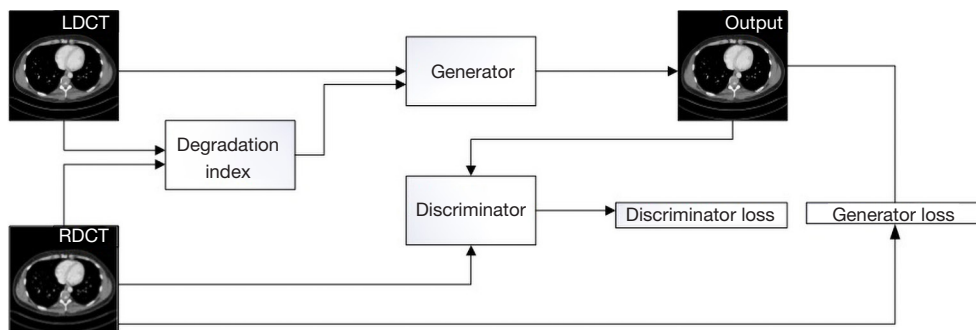

Figure 2 The workflow of the DR-GAN learning strategy for LDCT imaging. LDCT, low-dose computed tomography; RDCT, routine-dose computed tomography; DR-GAN, dynamic controllable residual generative adversarial network.

Through the introduction of skip connections, identity mapping can be realized. However, traditional residual block identity mapping has a single form, and it is difficult to balance the weights of different features during complex signal representation. It is therefore not suitable for degrading complex LDCT image processing tasks. Therefore, in our proposed method, a controllable weight is added to the conventional residual connection, thus creating a dynamically controllable residual module that can adapt to different strengths and form feature mapping.
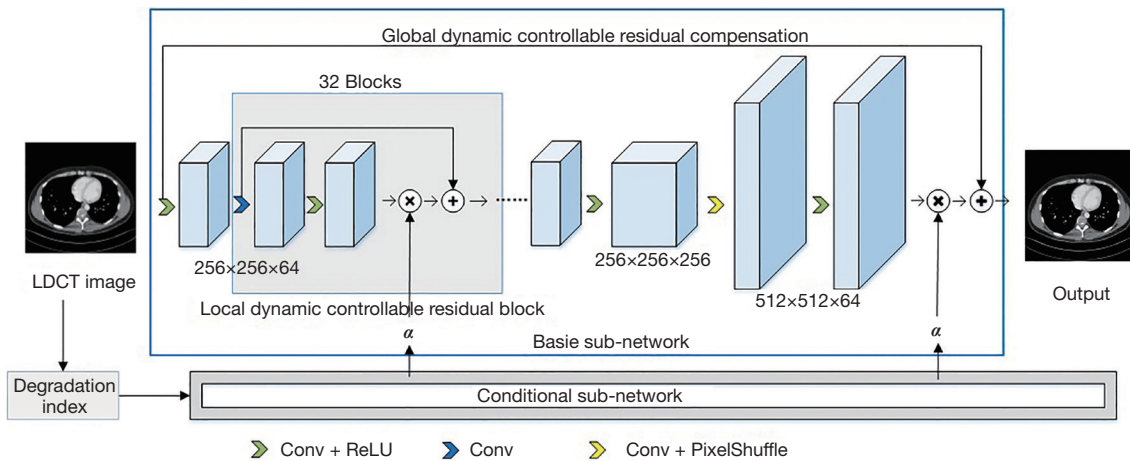
Figure 3 Generator network architecture. LDCT, low-dose computed tomography; Conv, convolution; ReLU, rectified linear unit.

By retaining the characteristics of the traditional residual, the dynamically controllable residual block can control the intensity residual component connection by introducing the weight α. The range of α is 0 to 1, which is the weight of residual mapping $F(X,W_i)$. The intensity of block output Y also comes between X and $F(X,W_i)+X$ (38-40). As shown in Eq. [3], the dynamic residual block can be expressed as follows:

$$Y = F(X,W_i) \times \alpha + X \qquad [3]$$

Generator

Inspired by the dynamic controllable strategy, we designed a dynamic controllable ResNet as a generator to improve the quality of different degraded LDCT images in a unified framework. The architecture of the generator G is shown in *Figure 3*. The generator G consists of 2 subnetworks and 1 module, the basic subnetwork, the conditional subnetwork, and the degradation index module. The basic subnetwork is the backbone network. The conditional subnetwork is an auxiliary network that is used to generate weights for different local dynamically controllable residual blocks in the basic subnetwork to assist the learning of the basic subnetwork from the degradation index module. The degradation index module provides the degradation type and level of the training data. Assume there are N degradation types $\{D_j\}_{j=1}^{N}$, then each degradation $D_j \in (0, d_j)$. Thus, we can obtain a degradation index vector for every LDCT image. Our goal was to design a restoration model that accepts the degraded image combined with the degradation index as inputs and generates the processed image. Taking any degradation type/level as input, the conditional subnetwork converts this into a condition vector and then obtains the controllable weights.

In the basic subnetwork, 32 local dynamic residual blocks are used to provide different intensities and different pattern features to improve the image quality of LDCT. One global dynamic residual block is used for similar tissue feature compensation. When the value of α is changed, the output shows different degrees of processing effect. The basic subnetwork mainly considers learning high-frequency residual information between low-dose and high-quality images, and residual features in most regions are close to 0. The local dynamic controlled residual block can avoid all the information of the training data and reduce the complexity and learning difficulty of the model. Adding local dynamic controllable residual connections to each network block can realize the feature extraction of noise and artifacts of a specific intensity and type. Enhanced structure preservation can be obtained in this way. Finally, global dynamic controllable residual connection, which helps accelerate model convergence and promote restoration performance, is used in the generator.

Based on previous literature (38-40), the conditional subnetwork was designed to consist of a stack of fully connected layers, which adds independent fully connected layers to each local dynamic controllable residual block and a global dynamic controllable residual block in the basic subnetwork. To generate the weights α that progress the basic subnetwork, each fully connected layer network needs to convert the degradation index of the corresponding
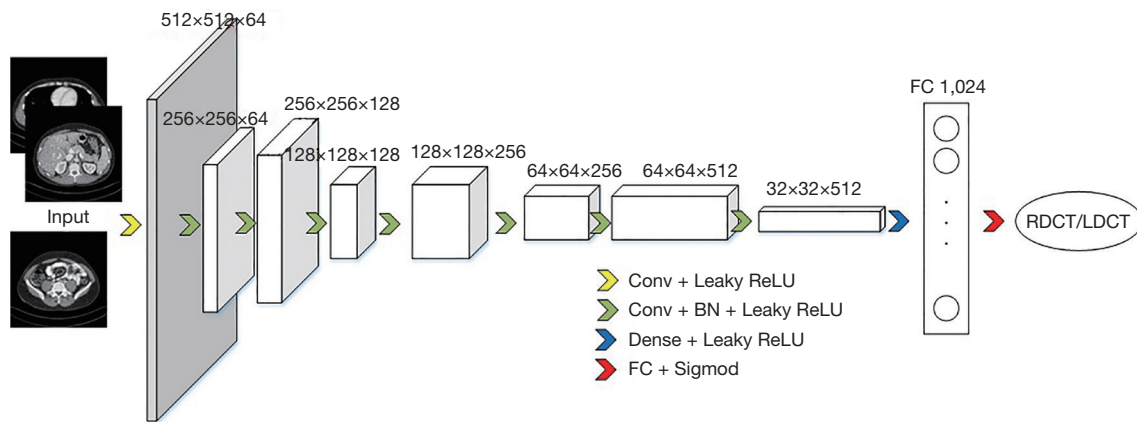
Figure 4 The discriminator network architecture. Conv, convolution; ReLU, rectified linear unit; BN, batch normalization; FC, full connection; RDCT, routine-dose computed tomography; LDCT, low-dose computed tomography.

image into a controllable weight α. The α dimension size is consistent with the feature channel of the residual block. The dimension size of the local controllable weights is set to 64, while the global controllable weight has a size of 1. The number of conditional subnetwork hidden layer neurons is 128 and 64 for local and global controllable weights, respectively.

Discriminator

The function of the discriminator is to promote the generator generate high-quality images in the generative adversarial learning framework. A modified VGG network was designed as discriminator D, as shown in *Figure 4*. The discriminator contains 8 convolution layers. From the second layer to the last layer, the convolution kernel size is 3×3. The stride size of convolution of the first, third, fifth, and seventh layers is 1 and that of the remaining layers is 2. After 1 convolution (Conv) and the leaky rectified linear unit (ReLU) layer, the next 7 layers use Conv + batch normalization (BN) + leaky ReLU operations to extract features at different levels of the image. In dense processing, the size of the convolution kernel is set to the same size as the previous feature map. Then, 1,024×32×32 convolutions are used to convolve the output characteristic diagram. The feature map is changed into a full connection (FC) layer in the form of 1,024 (FC 1,024). Finally, the FC + sigmoid operation is used to distinguish the true or false input image. By alternately training 2 networks, the generator produces a better estimation and generates a high-quality noise artifact-suppressed LDCT image.

Hybrid loss function

We further considered a hybrid loss function to optimize our model. In many previous studies (27,41), better performance was achieved through the integration of multiple objective functions. The widely used loss functions for image restoration are MSE and L1. In a previous study (42), structural similarity index metric (SSIM) was adopted to boost the structure feature closer to the ground truth. Similar to these methods, our proposed network adopted a hybrid consistency loss function. To maintain measurement consistency and learn more high-frequency details, we used the MSE as our main loss. To enhance the edge structural information and achieve a better visual effect, we introduced the MSSIM as an auxiliary loss. The adversarial loss, MSE loss, and MSSIM loss can improve the visual performance and mitigation oversmoothness.

To learn more high-frequency details, the MSE, which is also referred to as the *L2 norm*, is used as the loss function:

$$L_{MSE}\left(G\left(x_{LD}\right),x_{RD}\right)=\left\|G\left(x_{LD}\right)-x_{RD}\right\|_{2}^{2} \tag{4}$$

where $G(\bullet)$ is the generator G. To enhance the edge structural information, we introduce the MSSIM loss, which is formulated as follows:

$$L_{MSSIM}\left(G\left(x_{LD}\right),x_{RD}\right)=1-M_{SSIM}\left(G\left(x_{LD}\right),x_{RD}\right) \tag{5}$$

where $M_{SSIM}(\bullet)$ is calculated as the mean value of SSIM values in the patch set. The MSSIM value ranges from 0 to 1, with a larger value indicating better image quality. Therefore, the network training functions to minimize the following hybrid loss function L_{H}:

$$\min_G L_H = L_{MSE} + \alpha L_{MSSIM} \qquad [6]$$

where α denotes the weight of different loss terms.

The Wasserstein GAN with gradient penalty (WGAN-GP) optimizing strategy was introduced to mitigate the oversmoothing effect. The final LDCT processing objective function is formulated as follows:

$$\min_G \max_D L_{WGAN}(D,G) + \eta L_H(G) \qquad [7]$$

where is optimization objective of WGAN-GP, and η is the hyperparameter to balance the difference loss. During GAN training, the loss function of the generator may be constant at zero, and the gradient will disappear. To solve this problem, WGAN-GP uses the Wasserstein distance and GP to boost performance with the following objective function:

$$\min_G \max_D L_{WGAN}(D,G) = -E_{x_{RD}}\left[D(x_{RD})\right] + E_{x_{LD}}\left[D(G(x_{LD}))\right]$$
$$+ \lambda E_{\hat{x}}\left[\left(\left\|\nabla_{\hat{x}}\left[\nabla_{\hat{x}}D(\hat{x})\right]\right\|_2 - 1\right)^2\right] \qquad [8]$$

where $\hat{x}$ stands for the random interpolation sampling on the connection between x_{RD} and $G(x_{LD})$, and λ represents the hand-crafted penalty coefficient. As illustrated in *Figure 2*, DR-GAN includes 2 major components: the generator G and the discriminator D. Taking the LDCT image x_{LD} as input, the generator G produces an estimation $G(x_{LD})$ for the RDCT image x_{RD}. Then, the discriminator D guides the generator G to provide more realistic and high-quality images.

Comparison studies

Four methods were compared with DR-GAN, including filtered backprojection (FBP) (ramp filter), BM3D (17), residual encoder-decoder (RED) (22), and WGAN-GP (27). BM3D is known as a single-image block-matching iterative denoising method and here was applied for LDCT image processing. For BM3D, the noise type was additional white noise and pink noise, and the noise variance was 4×10^{-4} in our experiments. The RED method is a well-known LDCT image processing method trained by MSE loss; the number of convolution layers and deconvolution layers were set as 5, while the convolution size was set to 5×5. WGAN-GP is a postprocessing technique based on the generative model. The generator (without controllable weight) and discriminator have the same structure as does DR-GAN. WGAN-GP was trained by MSE and adversarial loss.

All parameters were set according to the papers in which they were originally described (17,22,27). All CNN-based methods were implemented in the TensorFlow DL framework. The platform was configured with an Intel(R) Core (TM) i5 3.00 GHz CPU and an NVIDIA Titan X graphical processing unit (GPU) with 12 GB of memory. To accelerate the convergence of the model, all training and testing were completed on a GPU.

Experimental setup

Challenge data

Use of the American Association of Physicists in Medicine (AAPM) Challenge data was authorized, and the data were downloaded from the 2016 National Institutes of Health (NIH)-AAPM-Mayo Clinic Low Dose CT Grand Challenge (https://www.Mayo.org/GrandChallenge/LowDoseCT/) (43). The Challenge data includes RDCT images from 10 patients (1 for validation, 1 for testing, and 8 for training). A fan-beam X-ray protocol was simulated to obtain the low-dose projections. The simulation geometric configuration was as follows: detector element size, 1.28 mm; detector element number, 736; distance from the source to the detector, 1,085.6 mm; distance from the source to the isocenter, 595 mm; and projection number per scanner cycle, 720. Poisson noise was added the routine-dose projection data to generate low-dose data (the X-ray incident photon intensity was $I_0 = 5\times10^4$ to simulate Poisson noise) (44). All images reconstructed with the analytic reconstruction (FBP, ramp filter) method had 512× 512 pixels, with each pixel measuring 0.8×0.8 mm^2 in size.

Real data

In this second set, real clinical data collected from United Imaging Healthcare (UIH) were used to further validate the performance of the proposed DR-GAN method (uCT-760 scanning unit). The dataset includes RDCT projections from 14 anonymous patients (12 for training, 1 for validation, and 1 for testing). A Chemical Inspection & Regulation Service (CIRS) phantom (3D Abdominal Phantom, Model 057A) CT dataset was adopted for quantitative analysis. The study was conducted in accordance with the Declaration of Helsinki (as revised in 2013). The protocol for scanning data collection and processing was approved by the institutional ethical review board of the UIH (No. 2015-07). Informed consent was obtained from all individual participants included in the study. Poisson noise was added into the routine-dose

projection to generate a dose level corresponding to 25% of the routine dose (approximately $I_0 = 1.2 \times 10^5$ incident photons from the X-ray source, 54 mA tube current, 120 kVp tube voltage, 0.9875 pitch; as provided by UIH) (45). The detector element size was 1.548×1.405 mm^2, and the detector element number was 936×80. The distance from the source to the detector was 950 mm, while the distance from the source to the isocenter was 570 mm. All images were reconstructed using the FBP method (B_SOFT_B kernel, helical interpolation) and had 512×512 pixels, with each pixel being 0.6836×0.6836 mm^2 in size.

Evaluation metrics

The experimental results were evaluated and analyzed from 2 aspects: visual assessment and quantitative assess. A quantitative analysis was performed using 2 indices: peak signal-to-noise ratio (PSNR), SSIM, and contrast-to-noise ratio (CNR). PSNR is used to measure the noise artifact suppression performance, SSIM to calculate the perceptual similarity performance, and CNR to measure the low-contrast detail preservation performance.

In this study, the PSNR, SSIM, and CNR indices were calculated via the following formulae.

$$\text{PSNR}(I, I_{RD}) = 20 \log_{10} \frac{L}{\text{MSE}(I, I_{RD})} \quad [9]$$

$$\text{SSIM}(I, I_{RD}) = \frac{(2\mu_{I_{RD}}\mu_I + C_1)(2\sigma_{II} + C_2)}{(\mu_{I_{RD}}^2 + \mu_I^2 + C_1)(\sigma_{I_{RD}}^2 + \sigma_I^2 + C_2)} \quad [10]$$

$$\text{CNR}(I_{ROI}, I_{BG}) = \frac{|\mu_{ROI} - \mu_{BG}|}{\sqrt{\sigma_{ROI}^2 - \sigma_{BG}^2}} \quad [11]$$

where I denotes the processed image; I_{RD} represents the reference data of the RDCT image; L is the value range of the image; MSE (•) is the MSE function; $\sigma_{I_{RD}}$ and σ_I are the standard deviations of the region of interest (ROI) in I_{RD} and I, respectively; $\mu_{I_{RD}}$ and μ_I are the corresponding mean values; σ_{II} is the corresponding covariance; C_1 and C_2 are 2 constant parameters; μ_{ROI} and μ_{BG} are the mean CT values of the ROI I_{ROI} and the background region I_{BG}, respectively; and σ_{ROI} and σ_{BG} are their respective standard deviations.

Parameter selection

The inputted data were all normalized to zero mean and unit variance. Following Yang *et al.* (27), the learning rate and exponential decay rates for the Adam algorithm were set as follows: $\lambda_r = 0.001$, $\beta_1 = 0.9$, and $\beta_2 = 0.999$. The $\lambda = 10$ in Eq. [8] was the default setting. The best values of the

parameters η and α in the hybrid loss functions were set as 1 and 10^4 in the Challenge data and real data experiment, respectively.

In our experiments, the discriminative feature representation (DFR) method was used to roughly evaluate the quality of LDCT images by 2 main degradation types: noise and artifacts (46). The main idea of DFR is to use 2 subdictionaries composed of different feature subdictionaries, representing tissue structure features and degradation features (streak artifacts or spot noise), to represent LDCT images. For noise degradation level estimation, the tissue attenuation feature subdictionary D_k^+ was trained from RDCT images, and the noise feature subdictionary D_k^- was trained from the difference data between the matched RDCT and LDCT by the K-singular value decomposition (KSVD) algorithm. To enhance the class-specific feature distinctiveness of D_k^+ and D_k^-, we applied the Fisher discrimination dictionary learning (FDDL) method to minimize the within-class scatter of different representation coefficients under the Fisher discrimination criterion. Thus, we could obtain 2 subdictionaries D^+ and D^- as the final tissue attenuation feature subdictionary and noise feature subdictionary, respectively. With the discriminative dictionary $D = [D^+; D^-]$, the calculated sparse coding [via the orthogonal matching pursuit (OMP) algorithm] was combined by the codes θ^+ and θ^- corresponding to the 2 subdictionaries D^+ and D^-, respectively. The noise degradation level was quantified using a ratio metric R between the summed codes associated with D^- and the summed weighted codes associated with D^+:

$$R = \frac{\sum_{m=1}^{M} \|\theta_m^-\|_1}{\sum_{m=1}^{M} \left(\|\theta_m^+\|_1 / (1 + \sigma_m)\right)} \quad [12]$$

where $\|\theta_m^+\|_1$ and $\|\theta_m^-\|_1$ denote the L1-norm of the code associated with the atoms in D^+ and D^- for each image patch, respectively; and σ_m is the standard deviation of each patch. The same process could be used to estimate the artifact degradation level. The artifact feature subdictionary D^- was trained from the difference images between matched RDCT and sparse view CT (360 views). In all the subsequent experiments, we adopted a dictionary with 1,024 atoms (512 atoms in D^+ and 512 atoms in D^-, atom size 16×16, maximum atom number 25). Finally, the degradation level evaluation results were demeaned, normalized, and formed as a degradation index. For example, (0.2, 0.3) represented a noise level of 0.2 and an artifact intensity of 0.3.
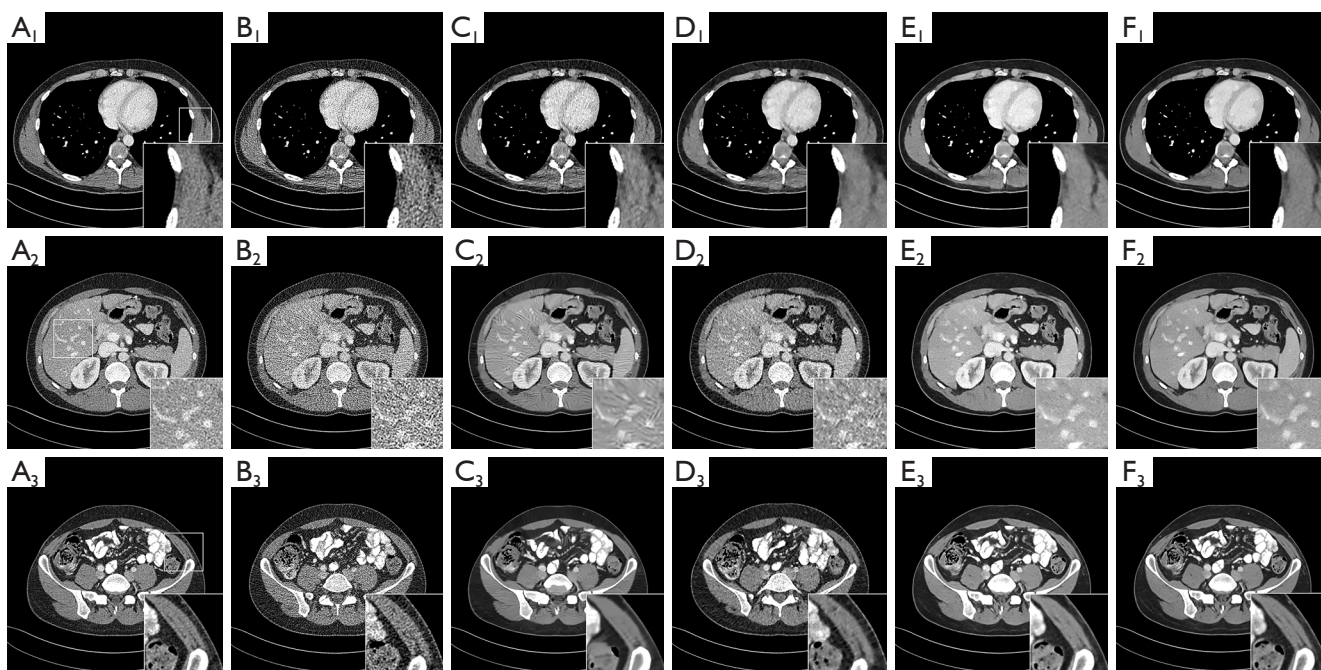
Figure 5 The chest, abdomen, and hypogastrium axial LDCT image results yielded by the different methods from the Challenge data. (A1-A3) Reference image (RDCT FBP image). (B1-B3) LDCT. (C1-C3) BM3D. (D1-D3) RED. (E1-E3) WGAN-GP. (F1-F3) DR-GAN. LDCT, low-dose computed tomography; RDCT, routine-dose computed tomography; FBP, filtered backprojection; BM3D, block matching 3D; RED, residual encoder-decoder; WGAN-GP, Wasserstein generative adversarial network with gradient penalty; DR-GAN, dynamic controllable residual generative adversarial network.

Results

Challenge data results

As shown in *Figure 5*, 3 axial slices were selected from the processed CT images for 1 patient case (L133). These LDCT images in the testing dataset had a rich, tiny structure and were disturbed by a large intensity range of strip artifacts and spot noise, making it difficult for doctors to diagnose the regular tissue or lesion regions. The classic BM3D method efficiently restored recovered the flat tissue region; however, the performance of the noise suppression ability was not good, and it also created some high-intensity artifacts. The RED method showed good tissue fidelity, but we saw some artifacts and noise distribution. However, the images processed by WGAN-GP suffered from a blurring effect and lost tissue texture in the abdominal CT images. A comparison of the images produced by BM3D, RED, and WGAN-GP in *Figure 5* demonstrates that the proposed DR-GAN can provide LDCT images with higher overall quality. In addition, the images generated by the proposed network appear closer to the reference images.

To demonstrate the performance of the proposed network from 3D coronal and sagittal views, 2 representative slices of results are shown in *Figure 6*, in which the blood vessels appear poorly identified in the LDCT FBP image. In the BM3D-processed images (*Figure 6*, C1-C2), the band artifacts appear to severely compromise the anatomical structure feature, and some artifacts and noise residual distributions in the RED results are apparent (*Figure 6*, D1-D2). WGAN and DR-GAN showed good noise suppression, and the structural feature was better maintained to a certain extent. Compared with other methods, the proposed method demonstrated excellent ability in removing noise and artifacts, and the processed image had better resolution and greater detail.

To further demonstrate the performance of DR-GAN, the difference maps relative to the RDCT FBP image are shown in *Figure 7*. It is clear that DR-GAN yielded the smallest difference from the RDCT FBP image in the overall region and less tissue structure difference in the sagittal view. Meanwhile, the BM3D and WGAN-GP processed images contained structural distortions
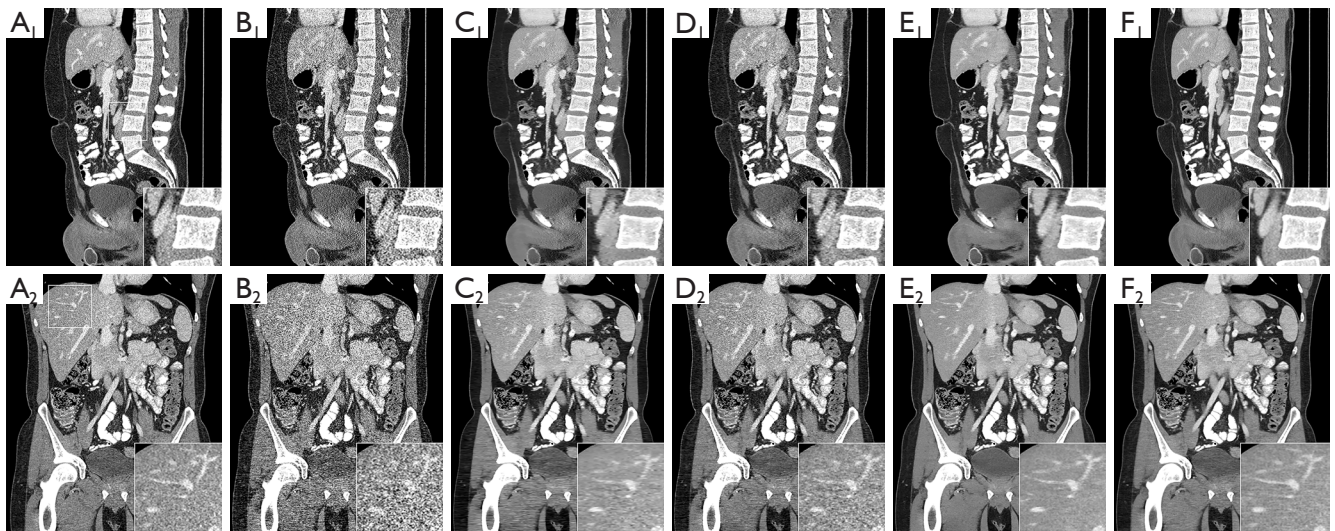
Figure 6 The sagittal and coronal LDCT image results of different methods on the Challenge data. (A1-A2) Reference image (RDCT FBP image). (B1-B2) LDCT. (C1-C2) BM3D. (D1-D2) RED. (E1-E2) WGAN-GP. (F1-F2) DR-GAN. LDCT, low-dose computed tomography; RDCT, routine-dose computed tomography; FBP, filtered backprojection; BM3D, block matching 3D; RED, residual encoder-decoder; WGAN-GP, Wasserstein generative adversarial network with gradient penalty; DR-GAN, dynamic controllable residual generative adversarial network.
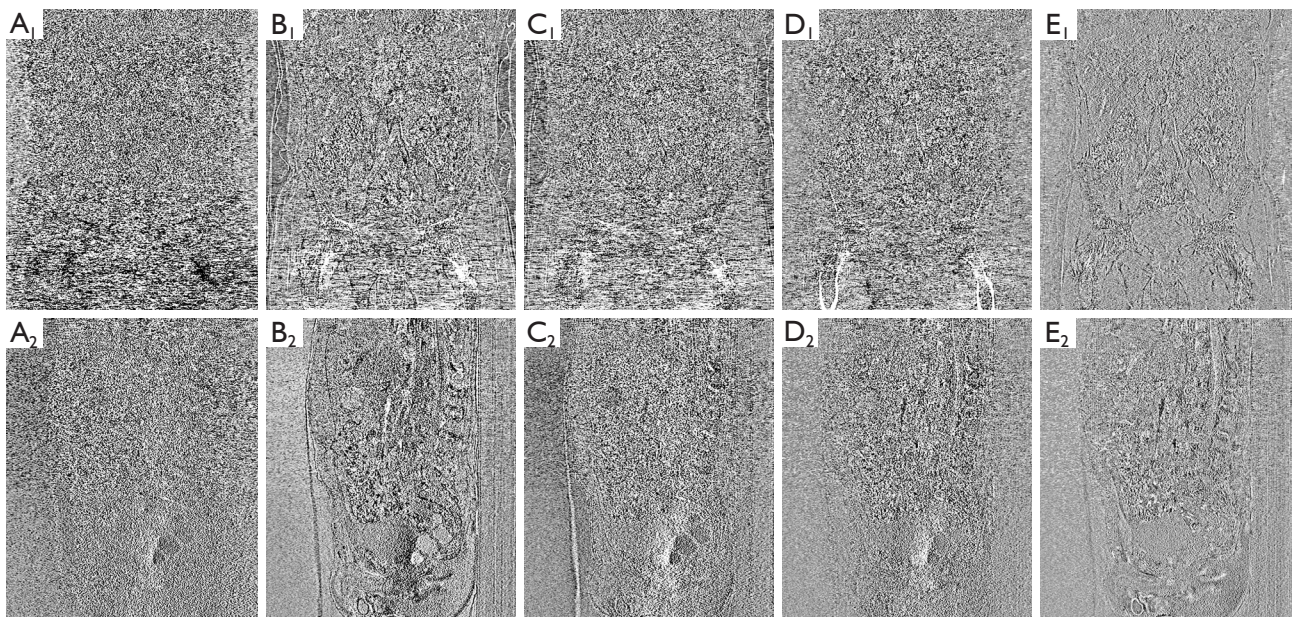

Figure 7 The difference images relative to the reference image (RDCT FBP image) for the Challenge data. The display window is (–50, 50) HU. (A1-A2) LDCT. (B1-B2) BM3D. (C1-C2) RED. (D1-D2) WGAN-GP; (E1-E2) DR-GAN. RDCT, routine-dose computed tomography; FBP, filtered backprojection; HU, Hounsfield unit; LDCT, low-dose computed tomography; BM3D, block matching 3D; RED, residual encoder-decoder; WGAN-GP, Wasserstein generative adversarial network with gradient penalty; DR-GAN, dynamic controllable residual generative adversarial network.
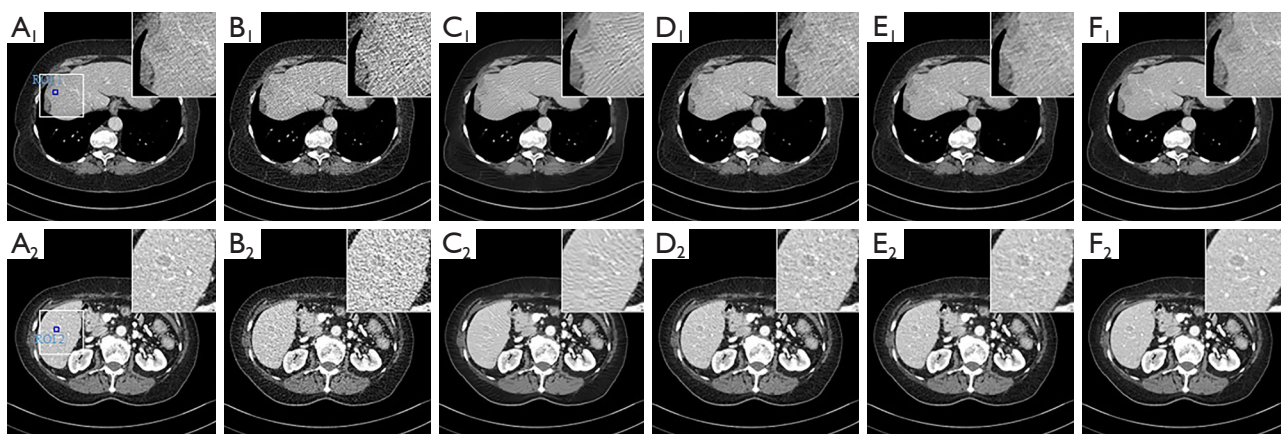
Figure 8 Examples of abdominal images with hepatic metastasis for the Challenge data. (A1-A2) Reference image (RDCT FBP image). (B1-B2) LDCT. (C1-C2) BM3D. (D1-D2) RED. (E1-E2) WGAN-GP. (F1-F2) DR-GAN. RDCT, routine-dose computed tomography; FBP, filtered backprojection; LDCT, low-dose computed tomography; BM3D, block matching 3D; RED, residual encoder-decoder; WGAN-GP, Wasserstein generative adversarial network with gradient penalty; DR-GAN, dynamic controllable residual generative adversarial network.

Table 1 Quantitative results associated with different methods for the Challenge data

Metrics	Test image	LDCT	BM3D	RED	WGAN-GP	DR-GAN
PSNR	Chest	33.931	37.144	37.855	38.899	42.925
	Abdomen	32.654	37.510	37.628	38.549	42.074
	Hypogastrium	34.017	39.414	39.318	40.425	42.862
	All images	34.080	37.675	38.262	39.291	42.511
SSIM	Chest	0.7747	0.8662	0.8957	0.9093	0.9485
	Abdomen	0.7057	0.8415	0.8601	0.8778	0.9355
	Hypogastrium	0.8193	0.9080	0.9315	0.9311	0.9576
	All images	0.7665	0.8719	0.8958	0.9092	0.9396
CNR ROI 1	(RDCT: 0.9879)	0.5576	0.9777	1.0352	1.4652	1.6484
CNR ROI 2	(RDCT: 0.8108)	0.6310	1.1397	1.0236	1.0557	1.3133

LDCT, low-dose computed tomography; BM3D, block matching 3D; RED, residual encoder-decoder; WGAN-GP, Wasserstein generative adversarial network with gradient penalty; DR-GAN, dynamic controllable residual generative adversarial network; PSNR, peak signal-to-noise ratio; SSIM, structural similarity index measurement; CNR, contrast-to-noise ratio; ROI, region of interest; RDCT, routine-dose computed tomography.

(*Figure 7*, *B1-C2*). The RED results had fewer image edge details but more residual noise in the CT images. This shows that the DR-GAN images were close to the RDCT FBP images.

Figure 8 shows the results for lesion CT images of 2 patients. Zoomed-in ROIs are also demonstrated. Specifically, *Figure 8* shows the CT images with low-contrast hepatic lesions (marked with blue boxes in *Figure 8*, *A1-A2*). From *Figure 8*, *A1-A2*, with the RDCT

images as references, we observe that the DR-GAN method yields the best image quality. We can also observe blurred features of low-contrast metastasis from the RED and WGAN-GP results (see the fourth and fifth columns in *Figure 8*). In addition, the metastases marked by the blue boxes were hard to detect (see *Figure 8*, *C1-E2*), but they could be discriminated with DR-GAN.

Table 1 presents the PSNR, SSIM, and CNR scores of the test results of the different methods. The CNR scores

5282

Xia et al. Deep CNN for LDCT imaging

Table 2 The subjective scores associated with the different methods (mean ± SD) for the Challenge data

Metric	Radiologist	RDCT	LDCT	BM3D	RED	WGAN-GP	DR-GAN
Noise suppression	R1	3.32±0.28	1.87±0.30	2.93±0.29	2.72±0.28	3.43±0.28	3.31±0.29
	R2	3.20±0.31	1.35±0.32	2.84±0.32	2.76±0.30	3.32±0.29	3.20±0.30
Artifact reduction	R1	3.52±0.23	1.63±0.31	2.27±0.32	3.17±0.29	3.61±0.28	3.55±0.30
	R2	3.41±0.25	1.41±0.31	2.15±0.29	3.04±0.29	3.50±0.28	3.42±0.29
Contrast retention	R1	3.65±0.24	1.73±0.32	2.28±0.33	3.13±0.30	3.42±0.32	3.51±0.29
	R2	3.49±0.26	1.52±0.34	2.09±0.34	2.92±0.31	3.27±0.29	3.40±0.29
Lesion discrimination	R1	3.41±0.25	1.54±0.31	2.33±0.31	2.96±0.28	3.23±0.28	3.39±0.27
	R2	3.30±0.26	1.32±0.32	2.12±0.32	2.85±0.29	3.01±0.28	3.18±0.28
Overall image quality	R1	3.53±0.24	1.61±0.31	2.34±0.31	2.94±0.28	3.41±0.29	3.48±0.29
	R2	3.40±0.27	1.40±0.32	2.23±0.32	2.83±0.29	3.25±0.28	3.32±0.29

SD, standard deviation; RDCT, routine-dose computed tomography; LDCT, low-dose computed tomography; BM3D, block matching 3D; RED, residual encoder-decoder; WGAN-GP, Wasserstein generative adversarial network with gradient penalty; DR-GAN, dynamic controllable residual generative adversarial network.

were calculated from the blue boxes in *Figure 8A1-A2*. As shown in *Table 1*, the FBP produced the lowest scores for LDCT imaging. The BM3D methods demonstrated some improvement but were unsatisfactory. The quantitative performances of WGAN-GP and DR-GAN were not significantly different. Consistent with the visualization results, DR-GAN obtained the best scores. The fluctuation range of the quantization scores of the proposed method was relatively small among the comparison methods, indicating that the robustness of the proposed method is superior.

For subjective image quality evaluation, 50 restored images with lesions were selected for a reader study, and a 5-point scale was adopted for subjective feature quantization. Two radiologists evaluated these images independently to provide their scores in *Table 2*. For all 5 indicators, the LDCT FBP images had the lowest scores. WGAN-GP produced substantially higher scores than did the other methods for noise artifact suppression. DR-GAN yielded slightly higher results than did RED and WGAN-GP for the contrast retention and lesion discrimination indicators. It was also found that the scores of the results from DR-GAN were closer to those of the RDCT images.

Real data results

In the real data experiments, all methods were retrained and readjusted for real data. *Figure 9* shows 3 axial slices

of 1 selected patient from the real data with the different methods applied. From the LDCT FBP image, we can see that the noise and artifacts seriously obfuscate the tissue in the image. Numerous strip artifacts can still be observed in the BM3D and RED results (see *Figure 9*, *C1-D3*, respectively). We found that the images generated by the WGAN-GP and DR-GAN had good visual effect (*Figure 9*, *E1-F3*). Moreover, the proposed DR-GAN was more promising in tiny tissue fidelity than was WGAN-GP (*Figure 9*, *F1-F3*).

Figure 10 shows the sagittal and coronal view images from the real data. We observed that the LDCT FBP images were severely distorted by band artifacts in the shoulder area. The image denoising method, BM3D, failed to suppress banding artifacts. As a deep DL method, RED showed good noise suppression, but it had residual artifacts and structural distortions. By implementing the Wasserstein distance, the WGAN-GP and DR-GAN mitigated the oversmoothing effect and improved the visual performance. Furthermore, by incorporating the dynamic controllable strategy and MSSIM loss into the generator, the DR-GAN exhibited a promising performance in structural preservation for the 3D coronal and sagittal view as compared to WGAN-GP.

Figure 11 shows the difference between the results of each method and the RDCT FBP image. Comparing the difference image from the different methods, we found that the intensity of the difference image processed by DR-
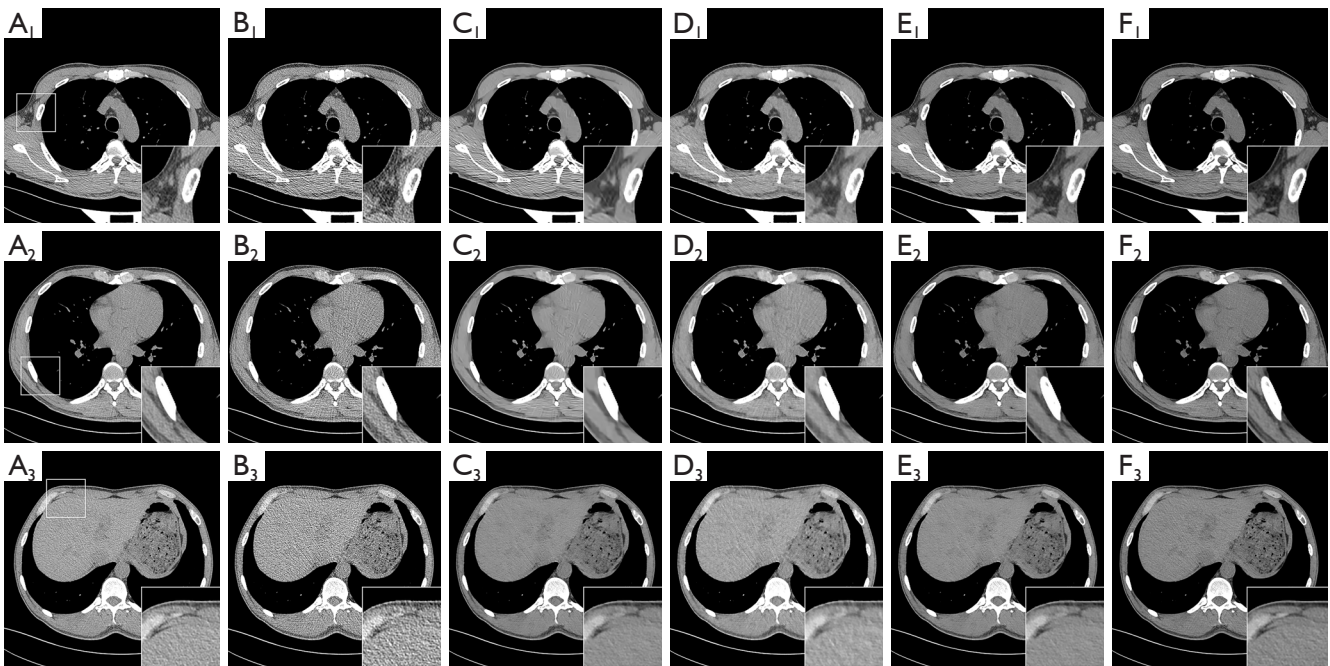
Figure 9 The shoulder, chest, abdominal axial LDCT image results of the different methods for the real data. (A1-A3) Reference image (RDCT FBP image). (B1-B3) LDCT. (C1-C3) BM3D. (D1-D3) RED. (E1-E3) WGAN-GP. (F1-F3) DR-GAN. LDCT, low-dose computed tomography; RDCT, routine-dose computed tomography; FBP, filtered backprojection; BM3D, block matching 3D; RED, residual encoder-decoder; WGAN-GP, Wasserstein generative adversarial network with gradient penalty; DR-GAN, dynamic controllable residual generative adversarial network.
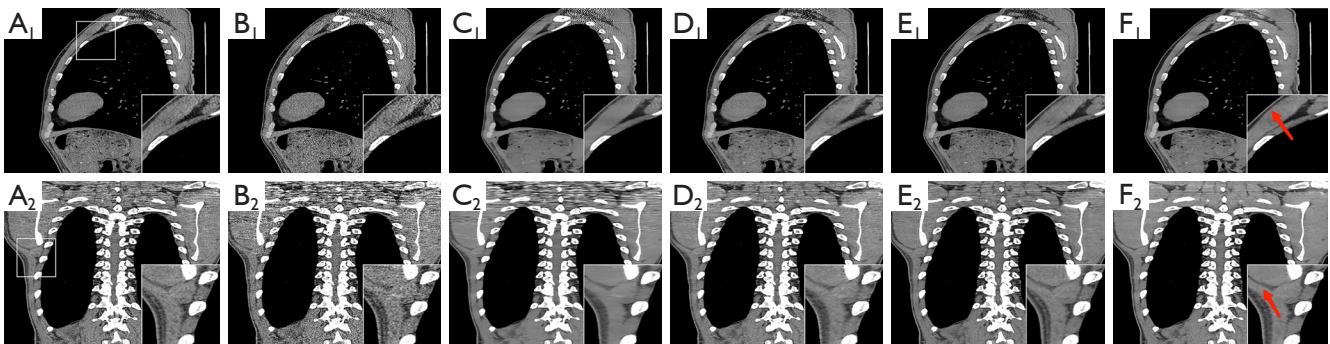

Figure 10 The sagittal and coronal LDCT image results of the different methods for the real data. (A1-A2) Reference RDCT FBP image. (B1-B2) LDCT. (C1-C2) BM3D. (D1-D2) RED. (E1-E2) WGAN-GP. (F1-F2) DR-GAN. The red arrows indicate obvious differences tissue. LDCT, low-dose computed tomography; RDCT, routine-dose computed tomography; FBP, filtered backprojection; BM3D, block matching 3D; RED, residual encoder-decoder; WGAN-GP, Wasserstein generative adversarial network with gradient penalty; DR-GAN, dynamic controllable residual generative adversarial network.

GAN was lower, indicating that the processed images were closer to the reference images than were those yielded by the comparison methods.

For the quantitative evaluation of the real dataset, we adopted the CIRS phantom data for comparison. *Figure 12, A1-A2* show the FBP-reconstructed RDCT images for reference. High-intensity noise artifacts can be observed for the LDCT FBP images (*Figure 12, B1-B2*),
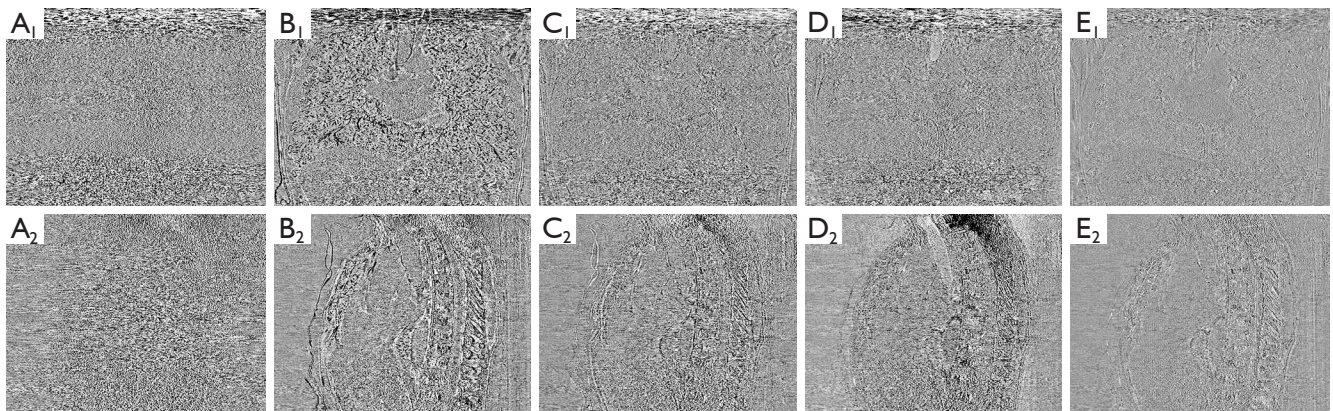
5284

Xia et al. Deep CNN for LDCT imaging

Figure 11 Difference images relative to the reference (RDCT FBP image for the real data. The display window is (–50, 50) HU. (A1-A2) LDCT. (B1-B2) BM3D. (C1-C2) RED. (D1-D2) WGAN-GP. (E1-E2) DR-GAN. RDCT, routine-dose computed tomography; FBP, filtered backprojection; HU, Hounsfield unit; LDCT, low-dose computed tomography; BM3D, block matching 3D; RED, residual encoder-decoder; WGAN-GP, Wasserstein generative adversarial network with gradient penalty; DR-GAN, dynamic controllable residual generative adversarial network.

Figure 12 The axial LDCT image results of different methods on the phantom data. (A1-A2) Reference RDCT FBP image. (B1-B2) LDCT. (C1-C2) BM3D. (D1-D2) RED. (E1-E2) WGAN-GP. (F1-F2) DR-GAN. ROI, region of interest; LDCT, low-dose computed tomography; RDCT, routine-dose computed tomography; FBP, filtered backprojection; BM3D, block matching 3D; RED, residual encoder-decoder; WGAN-GP, Wasserstein generative adversarial network with gradient penalty; DR-GAN, dynamic controllable residual generative adversarial network.
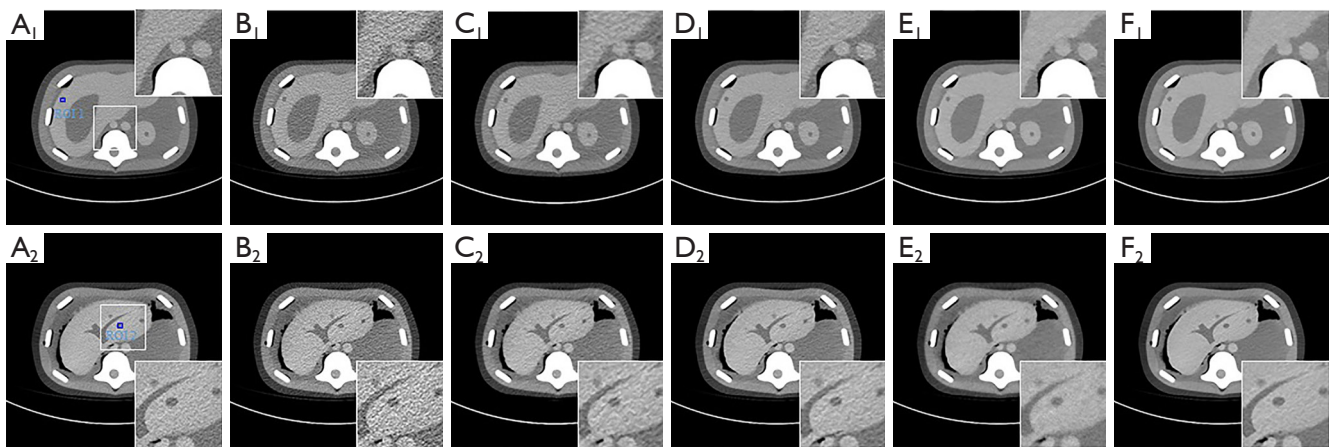
while for BM3D and RED, strip artifacts are present in the images (*Figure 12, C1-D2*), meaning that BM3D and RED could not obtain satisfactory results. Although WGAN-GP showed good ability in noise artifact suppression, its results suffered from a blurring phenomenon at the anatomical boundary (see the zoomed-in areas in *Figure 12, E1-E2*). Meanwhile, the DR-GAN methods (*Figure 12, F1-F2*) provided better LDCT image quality than did the other methods.

Figure 13 shows the Hounsfield unit (HU) intensity profiles across the virtual inner organ in 1 CIRS phantom slice for quantitative analysis in real low-dose data (54 mA tube current) and shows that the results of DR-GAN were closer to the reference RDCT FBP images than were those from other methods. The SSIM, PSNR, and CNR quantitative scores further demonstrated the processing
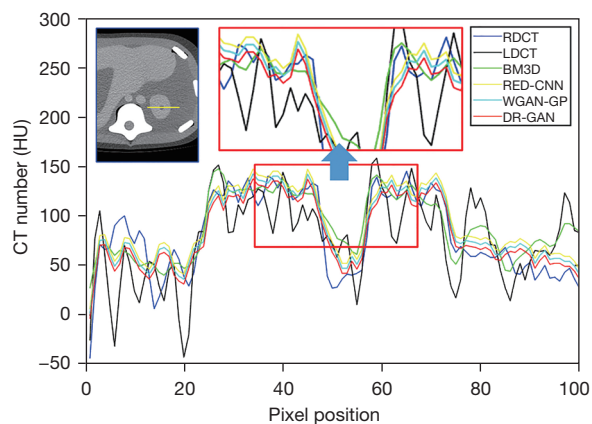
Figure 13 The CT intensity profiles of the specified yellow line in the CIRS phantom. CT, computed tomography; HU, Hounsfield unit; RDCT, routine-dose computed tomography; LDCT, low-dose computed tomography; BM3D, block matching 3D; RED-CNN, residual encoder-decoder convolutional neural network; WGAN-GP, Wasserstein generative adversarial network with gradient penalty; DR-GAN, dynamic controllable residual generative adversarial network; CIRS, Chemical Inspection & Regulation Service.

performance of the DR-GAN method (*Table 3*), with DR-GAN obtaining the highest score.

Ablation study

An ablation study was conducted to investigate the performance of the different elements of the DR-GAN network model, including the loss function and controllable residual. With reference to other impressive works (2,47-50), we adopted a progressive verification strategy for ablation analysis.

Effectiveness of loss function

As for the hybrid loss function of our DR-GAN training, a comparison of performance was completed with different weight parameters for the Challenge data. First, the model without adversarial learning (denoted as Dc-ResNet) and MSSIM loss terms was considered as a baseline. Then, MSSIM loss was introduced to the baseline model to build the new comparison model (denoted as Dc-ResNet+).

Table 3 Quantitative results associated with different methods for the real data

Metrics	Test image	LDCT	BM3D	RED	WGAN-GP	DR-GAN
PSNR	Chest	37.340	39.773	40.653	41.068	42.179
	Abdomen	37.564	40.120	40.367	41.228	41.763
	Hypogastrium	35.955	39.452	39.960	40.870	41.744
	All images	37.081	39.638	40.351	41.093	42.084
SSIM	Chest	0.8335	0.8944	0.9089	0.9163	0.9211
	Abdomen	0.8659	0.9198	0.9331	0.9387	0.9474
	Hypogastrium	0.8124	0.8914	0.9092	0.9171	0.9246
	All images	0.8395	0.8993	0.9141	0.9202	0.9313
CNR ROI 1	(RDCT: 0.8759)	0.7924	1.1791	0.9682	1.0641	1.1336
CNR ROI 2	(RDCT: 1.2467)	0.7807	1.2420	1.3364	1.6380	1.5951

LDCT, low-dose computed tomography; BM3D, block matching 3D; RED, residual encoder-decoder; WGAN-GP, Wasserstein generative adversarial network with gradient penalty; DR-GAN, dynamic controllable residual generative adversarial network; PSNR, peak signal-to-noise ratio; SSIM, structural similarity index measurement; CNR, contrast-to-noise ratio; ROI, region of interest; RDCT, routine-dose computed tomography.

Finally, the WGAN-GP framework was added to the Dc-ResNet+ model with the tuned GAN loss weight to construct our comparison method; that is, DR-GAN. Furthermore, we investigated the tradeoff between the quantitative score and the weight of the hybrid loss function

of our network. The quantitative results are shown in *Table 4*.

Compared with the 2 base models, DR-GAN obtained a lower MSE and PSNR. The main reason is that the base models were trained to minimize the pixel-wise MSE loss. Many studies have shown that MSE loss has an

Table 4 Quantitative results based on Challenge data for different model configurations

Metrics	Dc-ResNet	Dc-ResNet +	DR-GAN								
			$L_{WGAN} + \eta\, (L_{MSE} + \alpha L_{MSSIM})$								
	L_{MSE}	$L_{MSE} + \alpha L_{MSSIM}$	$\eta=0$	$\eta=10^{-1}$, $\alpha=0$	$\eta=10^{-1}$, $\alpha=10^2$	$\eta=10^{-1}$, $\alpha=10^4$	$\eta=10^0$, $\alpha=0$	$\eta=10^0$, $\alpha=10^2$	$\eta=10^0$, $\alpha=10^4$	$\eta=10^1$, $\alpha=10^2$	$\eta=10^1$, $\alpha=10^4$
PSNR	38.485	39.170	40.252	40.785	41.476	41.904	41.163	41.827	42.204	41.421	41.860
SSIM	0.9078	0.9166	0.9220	0.9274	0.9349	0.9392	0.9336	0.9387	0.9412	0.9337	0.9383

Dc-ResNet, dynamic controllable residual network; Dc-ResNet+, dynamic controllable residual network with MSSIM loss; DR-GAN, dynamic controllable residual generative adversarial network; MSN, mean square error; WGAN, Wasserstein generative adversarial network; PSNR, peak signal-to-noise ratio; SSIM, structural similarity index measurement; MSSIM, mean structural similarity index measurement.

Table 5 Quantitative results of the models with different dynamic controlled residuals

Dynamic controlled residual		PSNR	SSIM
Global	Local (number)		
√	√ (n=8)	41.423	0.9338
√	√ (n=16)	41.759	0.9379
√	√ (n=32)	42.134	0.9412
√	√ (n=48)	42.212	0.9488
×	√ (n=32)	41.019	0.9291
√	×	40.973	0.9285
×	×	40.610	0.9237

PSNR, peak signal-to-noise ratio; SSIM, structural similarity index measurement.

evident oversmoothing effect (48-50). Furthermore, DR-GAN obtained the best SSIM. The adversarial learning and MSSIM losses could effectively enhance the visual performance, generating textures similar to those of the reference image. Additionally, the comparison scores of the data shown in *Table 4* indicated that using only the MSE or MSSIM loss could yield optimized results. As seen in *Table 4*, the restored image quality was more sensitive to the MSE loss weight η than the MSSIM loss weight α.

Effectiveness of dynamic controllable residual strategy

Here, we discuss the effectiveness of the dynamic controllable residual strategy of training models with different network structures and without residual controllable weights. *Table 5* shows the quantitative scores of the test dataset. The performance was improved by introducing a dynamic controllable residual strategy,

verifying the effectiveness of DR-GAN. The model using the learnable controllable weight obtained better PSNR and SSIM scores.

Recent studies have suggested that deeper network architectures exhibit better performance for image processing tasks (37,50). We investigated the tradeoff between the quantitative score and number of local dynamic controllable residual blocks of our network. As seen in *Table 5*, adding more local dynamic controllable residual blocks yielded more obvious enhancement to the performance. To balance the tradeoff between processing effectiveness and efficiency, we chose the number of local dynamic controllable residual blocks n=32 as reasonable settings.

Convergence and computational efficiency

Additionally, the convergence properties of the DR-GAN model were analyzed by calculating the PSNR and training loss value. We set the weight of η to 1 based on the above experience, and the value of α was determined by parameter selection experiments. Specifically, the average PSNR values curve of the processed results obtained by using DR-GAN with different α and fixed η was generated. As shown in *Figure 14A*, the loss value of the generator declined rapidly before 50 epochs and converged to a constant stage. The loss value was smaller than that of the model without the GP. As shown in *Figure 14B*, when $\alpha=10^4$, the DR-GAN outperformed the other various α parameters with respect to PSNR.

The total numbers of parameters and the computational costs of different models were recorded for the Challenge data, with the results shown in *Table 6*. The execution times of all methods were calculated on the same GPU. For the DR-GAN network, approximately 17 hours were required
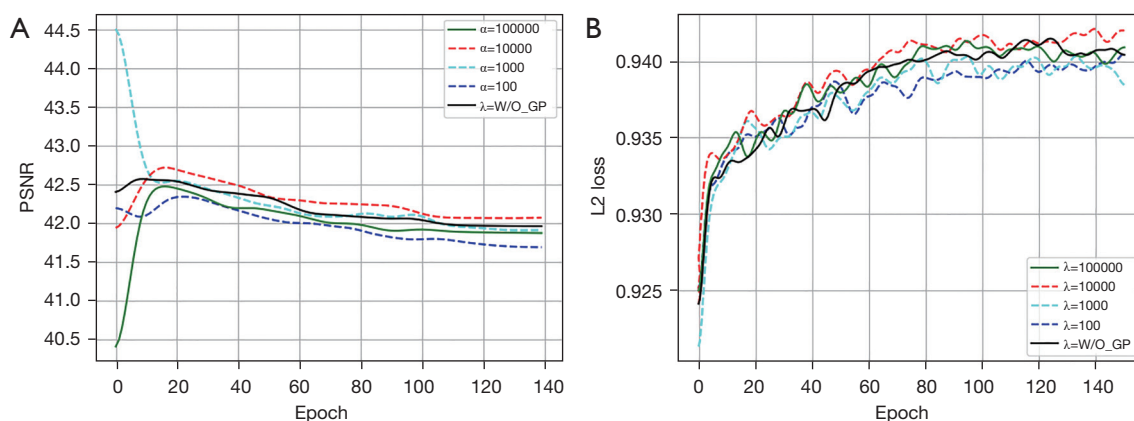
Figure 14 Convergence analysis of DR-GAN with different parameters. (A) PSNR. (B) L2 Loss. PSNR, peak signal-to-noise ratio; W/O_GP, without gradient penalty loss; DR-GAN, dynamic controllable residual generative adversarial network.

Table 6 Parameters and computational costs of the different methods

Method	BM3D	RED	WGAN-GP	DR-GAN
Parameters	–	1.8×10^6	5.2×10^7	5.8×10^7
Calculation	–	2.9×10^6	1.1×10^7	1.3×10^7
Training time (s/slice)	–	3,692.8	51,492	63,378
Test time (s/slice)	0.4405	0.1092	0.0971	0.1317

BM3D, block matching 3D; RED, residual encoder-decoder; WGAN-GP, Wasserstein generative adversarial network with gradient penalty; DR-GAN, dynamic controllable residual generative adversarial network.

for one training. However, with a well-trained model, the execution speed was quite efficient (about 0.13 seconds per slice). *Table 6* presents the total number of parameters and calculations required by the different deep networks. The DR-GAN network had more parameters and calculations than did the other networks. However, considering the performance improvement achieved over the WGAN-GP, the computational complexity involved in the DR-GAN was considerable, and the processing performance was better.

Performance robustness of different dose levels

In the above-described experiments, the noise levels of the training and test LDCT data were fixed and uniform. Nevertheless, the condition of the exact matched noise and artifacts (magnitudes/shapes) was not easy to achieve. However, the noise and artifact levels were different for different body parts even with the same CT scan

protocol. To analyze the robustness of DR-GAN, several combinations of noise levels in the training and testing datasets were simulated to generate the quantitative results (*Table 7*). For the dose levels in the Challenge data, 3 different dose levels were simulated via the Poisson noise model according to the literature (8). The incident photon intensity parameters were $I_0=8 \times 10^4$, $I_0=5 \times 10^4$, and $I_0=2 \times 10^4$, respectively. As shown in *Table 7*, the training datasets of RED, WGAN-GP, and DR-GAN were made for $I_0=5 \times 10^4$. RED+, WGAN-GP+ and DR-GAN+ denote the same networks as do RED, WGAN-GP, and DR-GAN with mixed training data at different dose levels.

From the processing results in *Table 7*, we can observe that the mixed dose levels in the training data led to better performance, and DR-GAN+ obtained the best performance in most situations. This supports the network training strategy: if an accurate dose level cannot be determined, the training data should be selected from a dose range close to the test data. Additionally, we can see that the DR-GAN result was still competitive in processing the cases of inconsistent LDCT images, demonstrating that the DR-GAN method, even if training is performed with a single dose level, offers a sound alternative in LDCT image processing.

Iterative reconstruction data study

In the above-described experiments, the training and test LDCT data used the matched FBP-reconstructed data. Nevertheless, some degraded types of matched CT data (tissue texture) were not always satisfactory. To analyze the robustness of date type, a cross-testing experiment

5288

Xia et al. Deep CNN for LDCT imaging

Table 7 Quantitative scores (mean) associated with different models for various noise levels

Dose level of testing data	Metrics	LDCT	BM3D	RED	WGAN-GP	DR-GAN	RED+	WGAN-GP+	DR-GAN+
$I_0=8\times10^4$	PSNR	36.778	41.572	43.539	44.270	44.683	44.700	44.626	44.841
	SSIM	0.8031	0.9118	0.9344	0.9369	0.9377	0.9412	0.9440	0.9466
$I_0=5\times10^4$	PSNR	34.080	37.675	38.262	39.291	42.511	39.645	39.427	42.574
	SSIM	0.7665	0.8719	0.8958	0.9092	0.9396	0.9017	0.9044	0.9391
$I_0=2\times10^4$	PSNR	29.718	33.795	35.371	36.043	36.837	37.344	37.872	38.116
	SSIM	0.6355	0.8156	0.8440	0.8536	0.8556	0.8646	0.8644	0.8731

LDCT, low-dose computed tomography; BM3D, block matching 3D; RED, residual encoder-decoder; WGAN-GP, Wasserstein generative adversarial network with gradient penalty; DR-GAN, dynamic controllable residual generative adversarial network; RED+, residual encoder-decoder network with mixed training data; WGAN-GP+, Wasserstein generative adversarial network with gradient penalty and mixed training data; DR-GAN+, dynamic controllable residual generative adversarial network and mixed training data; PSNR, peak signal-to-noise ratio; SSIM, structural similarity index measurement.
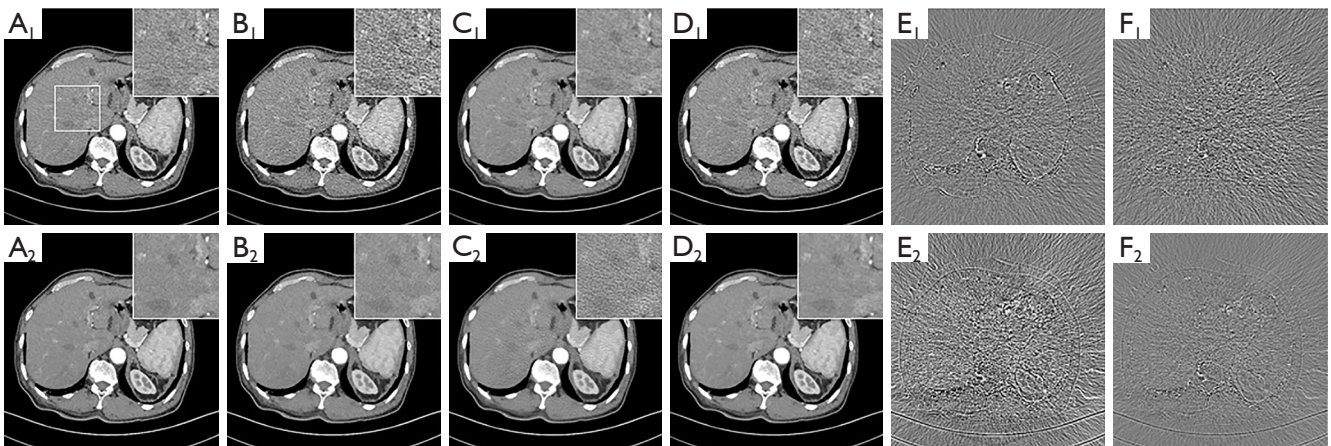

Figure 15 DR-GAN results and the different images. (A1) RDCT FBP image. (A2) RDCT RIO image. (B1) LDCT FBP image. (B2) LDCT RIO image. (C1) LDCT FBP image processed with DR-GAN using the trained FBP-reconstructed data. (C2) LDCT RIO image processed with DR-GAN using the trained FBP-reconstructed data. (D1) LDCT FBP image processed with DR-GAN using the trained RIO-reconstructed data. (D2) LDCT RIO image processed with DR-GAN using the trained RIO-reconstructed data. (E1-F2) Corresponding image difference with the RDCT FBP image. DR-GAN, dynamic controllable residual generative adversarial network; RDCT, routine-dose computed tomography; RIO, a commercial iterative reconstruction technique; LDCT, low-dose computed tomography; FBP, filtered backprojection.

was conducted between the FBP-reconstructed data and RIO-reconstructed data. RIO is a commercial iterative reconstruction algorithm deployed in the UIH CT scanner. *Figure 15* demonstrates the selected CT images from the different test data and the trained model.

In *Figure 15*, we can see that the matched training data achieved the best visual impression. This is verified by the different images in *Figure 15*, *D1-E2*, which show that the model trained using matched data provide the best results.

The main reason is that there was a large tissue and noise artifact texture difference between the FBP-reconstructed data and RIO-reconstructed data. There was almost no noise artifact feature in the RIO-reconstructed data. During network training, the weighting regarding the noise artifact feature removal may be very weak. This supports the following network training strategy: the training data and test data should be close in terms of noise artifact textures; that is, they should be from the same reconstruction

Table 8 Quantitative results (mean) associated with the different degradation indices

Degradation index	Estimation methods	Quantitative results (mean)		
		PSNR	SSIM	FSIM
Noise/artifacts	DFR/DFR	42.511	0.9396	0.9736
Noise/artifacts	DFR/SE	42.397	0.9383	0.9717
Noise/artifacts	PSNR/DFR	42.784	0.9414	0.9776
Image quality/artifacts	BIQI/DFR	42.443	0.9385	0.9736
Noise/image quality	DFR/BIQI	42.421	0.9380	0.9725
Noise/similarity	DFR/FSIM	42.627	0.9392	0.9740
Noise/similarity	PSNR/FSIM	42.848	0.9417	0.9774
Noise/artifacts/similarity	DFR/DFR/FSIM	42.823	0.9413	0.9769
Noise/artifacts/contrast	DFR/DFR/SE	42.466	0.9372	0.9731

PSNR, peak signal-to-noise ratio; SSIM, structural similarity index measurement; FISM, feature similarity index measure; DFR, discriminative feature representation; SE, subjective image quality evaluation; BIQI, blind image quality index.

methods or the same CT scanner.

Performance robustness of the different degradation indices

For most LDCT images, noise and artifacts are the main degradation types, and these 2 types of degradation are relatively independent. In the above-described experiments, we mainly used noise and artifact 2D degradation and the DFR evaluation method. Our degradation index module and network can be easily extended to higher dimension cases. To analyze the robustness of the degradation index module, several combinations of degradation types and degradation level evaluation methods in the training and testing were adopted to obtain the quantitative scores presented in *Table 8*. Here, we show 9 group examples with 5 degradation types: noise, artifact, contrast, similarity, and image quality. Degradation level evaluation methods include DFR, PSNR, feature similarity index measure (FSIM), blind image quality index (BIQI) (51), and subjective image quality evaluation (SE). PSNR and FSIM are reference evaluation indicators and are not available in real data testing. PSNR and FSIM indicators only serve as a reference for ideal situations. All evaluation results were demeaned, normalized and formed a degradation index vector.

From the processing results in *Table 8*, we can observe that the performance on three degradation types (the noise/artifacts/contrast group) decreases slightly, which is mainly due to insufficient training data and inaccurate evaluation. For degradation type selection, the performance of the

noise/image quality group is slightly lower than that of the noise/artifacts group and noise/similarity group but is not obvious in visualization. For degradation level evaluation method selection, the PSNR and FSIM evaluation indicators led to slightly better performance in terms of accuracy degradation scale during the training and testing stages. The SE evaluation indicator led to worse scores. Overall, the proposed network has robust performance for most degradation level evaluation methods. Experimentally, more degradation types (>2) need a greater amount of training data or complex networks, and the performance become degraded. With the computational efficiency and performance being taken into account, the noise and artifact degradation types and the DFR evaluation method can address most cases of LDCT image restoration tasks.

Discussion

CT imaging has become an important auxiliary medical diagnostic and treatment technology. To minimize exposure to X-ray radiation, research on high-performance LDCT image technologies has attracted substantial interest. However, low-dose scanning protocols (e.g., lowered milliampere/milliampere second settings) often lead to degraded images with increased mottle noise and streak artifacts. Most of the CNN methods use pixel-wise loss minimization and suffer from the following problems: blurring and oversmoothing phenomena, tiny structural deformation and noise, and artifact suppression

5290

Xia et al. Deep CNN for LDCT imaging

disequilibrium. To address the above Challenges, we propose a DR-GAN framework, including a dynamic controllable ResNet generator and VGG-128 network discriminator, to improve LDCT image quality.

Both the challenge and real data experimental results demonstrated the DR-GAN can provide a promising improvement in terms of noise artifact suppression and structural information preservation. Through ablation experiments, the components of the DR-GAN were confirmed to improve performance. A preliminary experiment of different dose level data further validated the robustness of the DR-GAN model. The results demonstrated that the proposed method has the potential to be a postprocessing solution when the dose level of training data and testing data do not match. An iterative reconstruction data study showed that DR-GAN remains valid for iterative reconstruction image input. These results verify the effectiveness of the proposed method.

Although the DR-GAN framework demonstrated promising results in LDCT image processing, some issues remain and should be addressed. First, residual noise artifacts can still emerge in the restored images (see *Figures 7,11*), and this problem deteriorates in the cases of lower-dose or high-intensity noise and artifacts. In the future, we will focus on how to further improve the separability of different noise and artifact features. Second, there are some hyperparameters involved in the DR-GAN, including λ, η, and α. These parameters are roughly selected based on the restored images. Hence, designing optimal parameters for automatic search algorithms for our network is also a challenging problem. Third, additional DR-GAN data validation is needed, which includes data from more patients and different CT scanners. Model evaluation will also be conducted in the context of other clinical tasks. Extended network applications, such as CT reconstruction (52), spectrum CT (53), and cone beam CT (CBCT) (54), can also be investigated in the future.

Conclusions

In this paper, we propose a DR-GAN framework to improve LDCT image quality. First, we adopted an adversarial learning strategy to alleviate oversmoothing and enhance the visual effect. Second, for the generator, a combined architecture consisting of the basic subnetwork and the conditional subnetwork was used to achieve dynamic controlled feature mapping. Furthermore, we chose the VGG-128 network as a discriminator to improve the noise artifact suppression and feature retention ability of the generator. Third, in the training process, a hybrid loss function was specifically designed to improve network performance. Experimental results demonstrated the competitive performance of the proposed method in terms of noise suppression, structural fidelity, and visual impression improvement.

Acknowledgments

Footnote

Conflicts of Interest: All authors have completed the ICMJE uniform disclosure form (available at https://qims. amegroups.com/article/view/10.21037/qims-22-1384/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). The protocol for scanning data collection and processing was approved by the institutional review board of the UIH (No. 2015-07). Informed consent was obtained from all individual participants included in the study.

References

1. Brenner DJ, Hall EJ. Computed tomography--an increasing source of radiation exposure. N Engl J Med 2007;357:2277-84.
2. Liu J, Kang Y, Xia Z, Qiang J, Zhang J, Zhang Y, Chen Y. MRCON-Net: Multiscale reweighted convolutional coding neural network for low-dose CT imaging. Comput Methods Programs Biomed 2022;221:106851.
3. Smith-Bindman R, Lipson J, Marcus R, Kim KP, Mahesh M, Gould R, Berrington de González A, Miglioretti DL. Radiation dose associated with common computed tomography examinations and the associated lifetime attributable risk of cancer. Arch Intern Med 2009;169:2078-86.
4. Shah NB, Platt SL. ALARA: is there a cause for alarm? Reducing radiation risks from computed tomography scanning in children. Curr Opin Pediatr 2008;20:243-7.
5. Gao Y, Liang Z, Zhang H, Yang J, Ferretti J, Bilfinger T, Yaddanapudi K, Schweitzer M, Bhattacharji P, Moore W. A Task-dependent Investigation on Dose and Texture in CT Image Reconstruction. IEEE Trans Radiat Plasma Med Sci 2020;4:441-9.
6. Boas FE, Fleischmann D. CT artifacts: causes and reduction techniques. Imaging in Medicine 2012;4:229-40.
7. Wang J, Lu H, Liang Z, Eremina D, Zhang G, Wang S, Chen J, Manzione J. An experimental study on the noise properties of x-ray CT sinogram data in Radon space. Phys Med Biol 2008;53:3327-41.
8. Luo S, Wu H, Sun Y, Li J, Li G, Gu N. A fast beam hardening correction method incorporated in a filtered back-projection based MAP algorithm. Phys Med Biol 2017;62:1810-30.
9. Liu J, Ma J, Zhang Y, Chen Y, Yang J, Shu H, Luo L, Coatrieux G, Yang W, Feng Q, Chen W. Discriminative Feature Representation to Improve Projection Data Inconsistency for Low Dose CT Imaging. IEEE Trans Med Imaging 2017;36:2499-509.
10. 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.
11. Liu J, Hu Y, Yang J, Chen Y, Shu H, Luo L, Feng Q, Gui Z, Coatrieux G. 3D feature constrained reconstruction for low-dose CT imaging. IEEE Trans Circuits Syst Video Technol 2018;28:1232-47.
12. Bao P, Xia W, Yang K, Chen W, Chen M, Xi Y, Niu S, Zhou J, Zhang H, Sun H, Wang Z, Zhang Y. Convolutional Sparse Coding for Compressed Sensing CT Reconstruction.

IEEE Trans Med Imaging 2019;38:2607-19.
13. Schaap M, Schilham AM, Zuiderveld KJ, Prokop M, Vonken EJ, Niessen WJ. Fast noise reduction in computed tomography for improved 3-D visualization. IEEE Trans Med Imaging 2008;27:1120-9.
14. Borsdorf A, Raupach R, Flohr T, Hornegger J. Wavelet based noise reduction in CT-images using correlation analysis. IEEE Trans Med Imaging 2008;27:1685-703.
15. Watanabe H, Kanematsu M, Miyoshi T, Goshima S, Kondo H, Moriyama N, Bae KT. Improvement of image quality of low radiation dose abdominal CT by increasing contrast enhancement. AJR Am J Roentgenol 2010;195:986-92.
16. Ma J, Huang J, Feng Q, Zhang H, Lu H, Liang Z, Chen W. Low-dose computed tomography image restoration using previous normal-dose scan. Med Phys 2011;38:5713-31.
17. Kang D, Slomka P, Nakazato R, Woo J, Berman D, Kuo C, Jay C, Dey D. Image denoising of low-radiation dose coronary CT angiography by an adaptive block-matching 3D algorithm. SPIE Med Phys 2013;8669:671-6.
18. Chen Y, Liu J, Hu Y, Yang J, Shi L, Shu H, Gui Z, Coatrieux G, Luo L. Discriminative feature representation: an effective postprocessing solution to low dose CT imaging. Phys Med Biol 2017;62:2103-31.
19. Hasan AM, Melli A, Wahid KA, Babyn P. Denoising Low-Dose CT Images Using Multiframe Blind Source Separation and Block Matching Filter. IEEE Trans Radiat Plasma Med Sci 2018;2:279-87.
20. Bie Y, Yang S, Li X, Zhao K, Zhang C, Zhong H. Impact of deep learning-based image reconstruction on image quality and lesion visibility in renal computed tomography at different doses. Quant Imaging Med Surg 2023;13:2197-207.
21. Kang E, Min J, Ye JC. A deep convolutional neural network using directional wavelets for low-dose X-ray CT reconstruction. Med Phys 2017;44:e360-75.
22. 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.
23. Zhang K, Zuo W, Chen Y, Meng D, Zhang L. Beyond a Gaussian Denoiser: Residual Learning of Deep CNN for Image Denoising. IEEE Trans Image Process 2017;26:3142-55.
24. Goodfellow IJ, Pouget-Abadie J, Mirza M, Xu B, Warde-Farley D, Ozair S, Courville A, Bengio Y. Generative Adversarial Nets. Proceedings of the 27th International Conference on Neural Information Processing Systems

5292

Xia et al. Deep CNN for LDCT imaging

2014;2:2672-80.

25. Radford A, Metz L, Chintala S. Unsupervised Representation Learning with Deep Convolutional Generative Adversarial Networks. arXiv 2015. arXiv:1511.06434.

26. Gulrajani I, Ahmed F, Arjovsky M, Dumoulin V, Courville A. Improved training of wasserstein gans. Adv Neural Inf Process Syst 2017;17:5768-78.

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. Hasan AM, Mohebbian MR, Wahid KA, Babyn P. Hybrid-Collaborative Noise2Noise Denoiser for Low-Dose CT Images. IEEE Trans Radiat Plasma Med Sci 2021;5:235-44.

29. Liu Y, Lei Y, Wang T, Fu Y, Tang X, Curran WJ, Liu T, Patel P, Yang X. CBCT-based synthetic CT generation using deep-attention cycleGAN for pancreatic adaptive radiotherapy. Med Phys 2020;47:2472-83.

30. Sun J, Du Y, Li C, Wu TH, Yang B, Mok GSP. Pix2Pix generative adversarial network for low dose myocardial perfusion SPECT denoising. Quant Imaging Med Surg 2022;12:3539-55.

31. Zhao F, Zeng Y, Peng G, Cao H, Liao J, Yu R, Peng S, Tan H. Noise of different tissues in chest CT scanning based on digital simulated technique. Chin J Tissue Eng Res 2012;16:1577-80.

32. Shan H, Zhang Y, Yang Q, Kruger U, Kalra MK, Sun L, Cong W, Wang G. 3-D Convolutional Encoder-Decoder Network for Low-Dose CT via Transfer Learning From a 2-D Trained Network. IEEE Trans Med Imaging 2018;37:1522-34.

33. Xia W, Shan H, Wang G, Zhang Y. Synergizing Physics/ Model-based and Data-driven Methods for Low-Dose CT. arXiv 2022. arXiv:2203.15725.

34. Kulathilake KASH, Abdullah NA, Sabri AQM, Lai KW. A review on Deep Learning approaches for low-dose Computed Tomography restoration. Complex Intell Systems 2023;9:2713-45.

35. 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.

36. Ma J, Liang Z, Fan Y, Liu Y, Huang J, Chen W, Lu H. Variance analysis of x-ray CT sinograms in the presence of electronic noise background. Med Phys 2012;39:4051-65.

37. He K, Zhang X, Ren S, Sun J. Deep Residual Learning for Image Recognition. Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR); 2016:770-8.

38. He J, Dong C, Qiao Y. Interactive multi-dimension modulation with dynamic controllable residual learning for image restoration. Proc of 16th ECCV 2020:53-68.

39. Cai H, He J, Qiao Y, Dong C. Toward interactive modulation for photo-realistic image restoration. Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR) Workshops; 2021:294-303.

40. Chen R, Zhang Y. Learning Dynamic Generative Attention for Single Image Super-Resolution. IEEE Trans Circuits Syst Video Technol 2022;32:8368-82.

41. Zhang Y, Hu D, Zhao Q, Quan G, Liu J, Liu Q, Zhang Y, Coatrieux G, Chen Y, Yu H. CLEAR: Comprehensive Learning Enabled Adversarial Reconstruction for Subtle Structure Enhanced Low-Dose CT Imaging. IEEE Trans Med Imaging 2021;40:3089-101.

42. Unal MO, Ertas M, Yildirim I. An unsupervised reconstruction method for low-dose CT using deep generative regularization prior. Biomed Signal Process Control 2022;75:103598.

43. Available online: https://www.aapm.org/GrandChallenge/ LowDoseCT/#noiseInsertion

44. Elbakri IA, Fessler JA. Statistical image reconstruction for polyenergetic X-ray computed tomography. IEEE Trans Med Imaging 2002;21:89-99.

45. Zabić S, Wang Q, Morton T, Brown KM. A low dose simulation tool for CT systems with energy integrating detectors. Med Phys 2013;40:031102.

46. Gu Y, Tang H, Lv T, Chen Y, Wang Z, Zhang L, Yang J, Shu H, Luo L, Coatrieux G. Discriminative feature representation for Noisy image quality assessment. Multimed Tools Appl 2020;79:7783-809.

47. Liu J, Zhang Y, Zhao Q, Lv T, Wu W, Cai N, Quan G, Yang W, Chen Y, Luo L, Shu H, Coatrieux JL. Deep iterative reconstruction estimation (DIRE): approximate iterative reconstruction estimation for low dose CT imaging. Phys Med Biol 2019;64:135007.

48. Yin X, Zhao Q, Liu J, Yang W, Yang J, Quan G, Chen Y, Shu H, Luo L, Coatrieux JL. Domain Progressive 3D Residual Convolution Network to Improve Low-Dose CT Imaging. IEEE Trans Med Imaging 2019;38:2903-13.

49. Zhang Y, Hu D, Hao S, Liu J, Quan G, Zhang Y, Ji X, Chen Y. DREAM-Net: Deep Residual Error Iterative Minimization Network for Sparse-View CT Reconstruction. IEEE J Biomed Health Inform

2023;27:480-91.

50. Hu D, Liu J, Lv T, Zhao Q, Zhang Y, Quan G, Feng J, Chen Y, Luo L. Hybrid-Domain Neural Network Processing for Sparse-View CT Reconstruction. IEEE Trans Radiat Plasma Med Sci 2021;5:88-98.

51. Moorthy AK, Bovik AC. A two-step framework for constructing blind image quality indices. IEEE Signal Process Lett 2010;17:513-6.

52. Hu D, Zhang Y, Liu J, Luo S, Chen Y. DIOR: Deep Iterative Optimization-Based Residual-Learning for Limited-Angle CT Reconstruction. IEEE Trans Med Imaging 2022;41:1778-90.

53. Zhu J, Su T, Zhang X, Yang J, Mi D, Zhang Y, Gao X, Zheng H, Liang D, Ge Y. Feasibility study of three-material decomposition in dual-energy cone-beam CT imaging with deep learning. Phys Med Biol 2022.

54. Hu D, Zhang Y, Liu J, Zhang Y, Coatrieux JL, Chen Y. PRIOR: Prior-Regularized Iterative Optimization Reconstruction For 4D CBCT. IEEE J Biomed Health Inform 2022;26:5551-62.