RESEARCH   **Open Access**

Disentangled generative adversarial network for low-dose CT

Wenchao Du, Hu Chen[*], Hongyu Yang and Yi Zhang[*]

* Correspondence: huchen@scu.
edu.cn; yzhang@scu.edu.cn
College of Computer Science,
Sichuan University, Chengdu
610065, China

Abstract

Generative adversarial network (GAN) has been applied for low-dose CT images to predict normal-dose CT images. However, the undesired artifacts and details bring uncertainty to the clinical diagnosis. In order to improve the visual quality while suppressing the noise, in this paper, we mainly studied the two key components of deep learning based low-dose CT (LDCT) restoration models—network architecture and adversarial loss, and proposed a disentangled noise suppression method based on GAN (*DNSGAN*) for LDCT. Specifically, a generator network, which contains the noise suppression and structure recovery modules, is proposed. Furthermore, a multi-scaled relativistic adversarial loss is introduced to preserve the finer structures of generated images. Experiments on simulated and real LDCT datasets show that the proposed method can effectively remove noise while recovering finer details and provide better visual perception than other state-of-the-art methods.

Keywords: Image reconstruction, Low-dose CT, Image denoising, Generative adversarial network
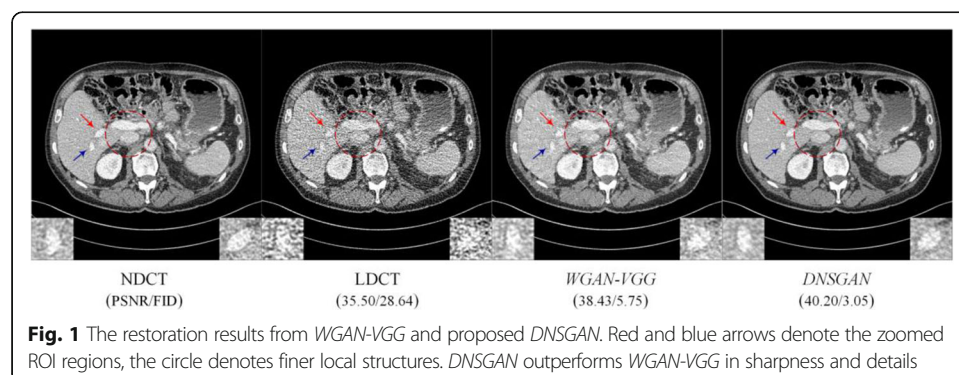
1 Introduction

Low-dose CT denoising has been a hot topic in medical imaging and numerous methods have been proposed to deal with this problem [1]. These algorithms could be approximately categorized into three groups according to the processing stage: Sinogram filtering, iterative reconstruction (IR), and image post processing methods. Sinogram filtering methods [2, 3] directly process the projection data, but any improper operations would result in undesired artifacts and loss of structural information or/and spatial resolution. IR methods [4–7] have the advantage in producing results with high peak signal-to-noise ratio (PSNR). However, the substantial computational cost and empirical parameter turning limit the extensive applications of this kind of methods in commercial scanners. Image post processing methods need not access the measurements and many methods [8–15] proposed for natural image restoration can be directly introduced for low-dose CT (LDCT) denoising, such as non-local means [8, 9], K-means singular value decomposition (*KSVD*) [10], and block-matching and 3D filtering (*BM3D*) [13]. However, due to the complexity of the statistical property

of noise in LDCT images, these methods cannot provide similar performance as that for natural images.

After the pioneering work was proposed by Chen et al. [16], deep neural network (DNN) approaches have brought a prosperous development in this field [17–19]. Various network architectures [20–23] have continuously improved the LDCT denoising performance. However, most of these methods utilize L2 norm as the target function, which produce results with high PSNR and structural similarity (SSIM) [24] but increase Fréchet inception distance (FID) [25] scores due to smoothed structural details. Since the PSNR metric does not completely coherent to the subjective evaluation of human observers, this fact may have uncertainly negative impact on clinical diagnosis. To circumvent this obstacle, generative adversarial network (GAN) and different loss functions were introduced to restore finer structural details as much as possible [26–31]. As the most representative one, *WGAN-VGG* [26], aided by stable Wasserstein GAN [32] training and perceptual loss [33], was proposed to encourage the network to favor solutions that look more like realistic normal-dose CT (NDCT) images. Although considerable improvements have been obtained, there still exists a noticeable gap between *WGAN-VGG* results and the NDCT images. One example is shown in Fig. 1. Although the result generated by *WGAN-VGG* has similar mottle-like noise, the distribution is quite different from the real NDCT image. The reason may lie in that most existing methods endeavor to transform LDCT images directly into corresponding normal-dose CT (NDCT) ones, which require a quite powerful model. Actually, an important problem was ignored that for LDCT, noise always adheres to the high frequency details. Therefore, the DNN-based methods with L2 norm tend to generate over-smooth results and the GAN-based methods introduce extra noise into the generated images, which would lead to better visualization but lower PSNR and higher FID scores.

In order to alleviate this contradiction, in this paper, we propose a novel disentangled noise suppressing method for LDCT. We disentangle the procedure of LDCT denoising into two stages, noise removal and structural detail enhancement, instead of one-step mapping. Specifically, we firstly transform the source distribution into an intermediate distribution by paying more attention on noise removal, which may lead to over-smooth results with varying degree. After that the intermediate distribution is transformed into the final target distribution. This process is implemented by recovering the finer details from the denoised images from last step. In addition, the proposed disentangled noise suppressing method is embedded into the framework of GAN [34], termed as *DNSGAN*, to further enhance the visual perception of reconstructed images.

Fig. 1 The restoration results from *WGAN-VGG* and proposed *DNSGAN*. Red and blue arrows denote the zoomed ROI regions, the circle denotes finer local structures. *DNSGAN* outperforms *WGAN-VGG* in sharpness and details

The main contribution of this paper can be summarized as that proposed a novel disentangled noise suppression method—*DNSGAN*. Instead of one-step mapping, *DNSG AN* is more effective to handle LDCT restoration with the divide and conquer strategy that decoupling image denoising into two stages—noise removal and structure enhancement. Proposed method achieved higher-quality image reconstruction and improved the generalization for kinds of noise-levels than other competitive methods, resulting in better balance between the details retaining and quantitative metrics.

The rest of this paper is organized as follows. In section 2, the proposed *DNSGAN* method is described in detail. The experimental results are demonstrated in section 3 and the final section concludes this paper.

2 Method

2.1 Noise reduction model

The general image restoration problem can be considered from the perspective of domain transform [35]. A source domain S and a target domain $\mathcal{T}$ contain samples from two different given distributions P_S and P_T respectively. $x \in S$ denotes the LDCT image from the source domain and $y \in \mathcal{T}$ denotes the corresponding NDCT image from the target domain where $x \sim P_S$, and $y \sim P_T$.

For the image restoration task, a generic denoising process for LDCT can be expressed as:

$$x = F(y) + \varepsilon \tag{1}$$

where $F : x \rightarrow y$ represents a nonlinear degrading process by the noise and ε stands for the additive part of noise and other unmodeled factors. Current methods based DNN focus on learning a nonlinear function $F^{\dagger}$ to directly map x into y, which can be expressed as:

$$F^{\dagger}(x) = \hat{y} \approx y \tag{2}$$

The general ideal is to find the optimal $F^{\dagger}$ to minimize the distance between P_S and P_T.

Unfortunately, since the noise in LDCT images does not obey any specific statistical distribution, the denoising operation will inevitably smooth the details to a certain degree, which makes it difficult to directly learn $F^{\dagger}$, even GAN is introduced to enforce stronger constraint. As a result, the result may quite depend on the specific form of loss function.

In order to solve this problem, inspired by the idea of domain transform, this process is disentangled into two steps: noise reduction and structural enhancement. The first step follows the general idea of learning based methods to learn an image-to-image denoising model and the second step is to recover the details smoothed by the first step. This is similar with the pre-upsampling image super-resolution models [36], which is upsampling the original image first and then recovering the details on the upsampled image. The denoised result y' after first step can be viewed as an intermediate result, which bridge the gap between the low-dose image x and normal-dose image y. Based on this consideration, Eq. 2 is reformulated as follow,

$$F^{\dagger}(x) = R\left(y^{'}\right) = \hat{y} \approx y, \;\; y^{'} = S(x) \tag{3}$$

where $S(\cdot) : x \in S \rightarrow y^{'} \in I$ represents the noise suppression process, which transforms the sample x into the intermediate domain I. $R(\cdot) : y^{'} \in I \rightarrow \hat{y} \in \mathcal{T}$ denotes the detail recovery process, which aims to enhance the structures and recover finer details from the denoised (probably over-smoothed) intermediate image.

2.2 Network architecture

The proposed network model follows the classical architecture of GAN, which contains a disentangled generator network and a relativistic multi-scale discriminator network. The generator is composed of two modules, a dynamic filter module for noise suppression and a structure enhancement module for detail recovery. The network architecture is shown in Fig. 2 and the details of each module are elaborated in the following subsections.
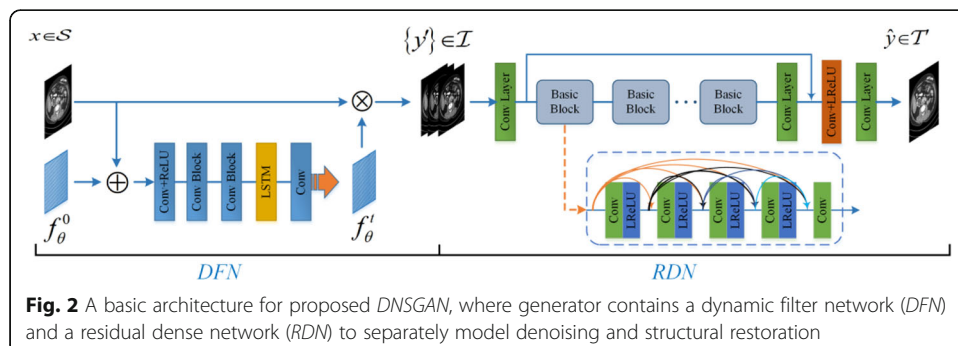
2.2.1 Noise removal module

Due to the nondeterminacy of the noise distribution in image domain, we propose to adopt dynamic filter network (DFN) [37], which is learned adaptively from the input data. The proposed noise suppressing model could be represented as:

$$y^{'} = S(x) = f_{\theta} \odot x \tag{4}$$

where $f_{\theta} = DFN(x)$, which denotes the output filter generated by DFN. $\theta \in \mathbb{R}^{s \times s}$ is the parameter set of the filter f. s is the filter size. f_{θ} is applied to the input as $y^{'} = f_{\theta} \odot x$, where $\odot$ is the point-wise multiplication operator.

In order to reduce the complexity of network structure while improving the performance of noise suppression, a LSTM unit is introduced into DFN to progressively generate dynamic filters. Furthermore, an adaptive strategy is used to guide the dynamic filter generation and we concatenated the last updated filter f_{θ}^{t-1} and current input as the updated input in each time step.

Considering that the DFN focuses on noise suppression, mean square error (MSE) loss function is utilized, which is formulated as:

Fig. 2 A basic architecture for proposed DNSGAN, where generator contains a dynamic filter network (DFN) and a residual dense network (RDN) to separately model denoising and structural restoration

$$\mathcal{L}_{dfn}\left(\left\{y^{'}\right\}, y\right) = \sum_{t=1}^{N} \lambda_t \mathcal{L}_{mse}\left(y^{'}_{f^t_\theta}, y\right) \qquad (5)$$

where $y^{'}_{f^t_\theta} = DFN(x \oplus f^{t-1}_\theta)$ is the updated image, and $\oplus$ denotes the channel-wise concatenation operation. f^0_θ is initialized with Gaussian distribution. To balance the training time and performance, we set $N = 3$ and $\lambda = [0.25, 0.5, 1]$ in our experiments.

2.2.2 Structure enhancement module

Inspired by the deep learning based works for image super-resolution [38, 39], our structural enhancement module used a residual dense network (*RDN*) [40] to recovery structural details, as shown in Fig. 2, which is similar with [39]. To further enhance the performance, we made the following improvements on [39]:

2.2.2.1 Richer input *RDN* aims to enhance the structure details for the denoised input. However, *DFN* tends to generate over-smoothed results with varying degrees. In order to avoid excessive details loss, we concatenated each $y^{'}_t$ at different time steps as input of *RDN*.

2.2.2.2 Lightweight backbone Compared with other networks for super-restoration task, our *RDN* module aims to recovery structural details from over-smoother inputs, which needs to pay more attention on the finer structures and details. Based on this consideration, we removed the up/down-sample operations in *RDN*, and used five residual dense blocks as backbone of *RDN*, which demonstrate powerful performance on detail recovery in our experiments.

2.2.2.3 Improved feature loss Considering that feature loss has been widely used for detail recovery, we borrowed the improved feature loss [39] into *RDN*. The features before the activation layer are utilized to enhance the performance of recovering details, which can avoid the inconsistent details due to the sparseness of activated feature. A pretrained VGG-19 [41] model was used for the feature loss.

As a result, the total loss for generator in *DNRGAN* is defined as:

$$\mathcal{L}_{Gen} = \lambda \mathcal{L}_{dfn} + \mathcal{L}_c + \eta \mathcal{L}_{fea} + \gamma \mathcal{L}_{G^{Ra}} \qquad (6)$$

where $\mathcal{L}_c = \mathbb{E}_x \|RDN(y^{'})\text{-}y\|_1$ is the content loss that evaluates the differences between the generated images and ground truth images, $\mathcal{L}_{fea}$ is the feature loss, $\mathcal{L}_{G^{Ra}}$ is the GAN loss, and λ, η, and γ are the balancing coefficients.

2.2.3 Relativistic PatchGAN

In order to reduce the complexity of the network and improve the visual quality of the generated image, we also made two modifications on the traditional discriminator architecture to enhance the training efficiency: (a) one is introducing the relativistic adversarial loss into discriminator, which mainly predict the probability that a real input is relatively more realistic than the fake input instead of a binary output, and (b) the other is using a multi-scale *PatctGAN* [42, 43] to simplify the network structure while enhance the performance of discriminator.

The traditional discriminator can be expressed as $D(x) = \sigma(C(x))$, where $\sigma(\cdot)$ is the sigmoid function and $C(x)$ is the non-transformed discriminator output. In our *DNSGAN*, a relativistic average discriminator [44] is used, referred as D_{Ra}, which is formulated as $D_{Ra}(y_r, y_f) = \sigma(C(y_r) - \mathbb{E}_{y_f}[C(y_f)])$, where y_r represents the NDCT image, y_f represents the generated denoised CT image, and $\mathbb{E}_{y_f}[\cdot]$ represents the averaging operation on all fake data in the mini-batch, as shown in Fig. 3.

The discriminator loss is then defined as:

$$\mathcal{L}_{D^{Ra}} = -\mathbb{E}_{y_r}\Big[\log\Big(D_{Ra}\Big(y_r, y_f\Big)\Big)\Big] - \mathbb{E}_{y_f}\Big[\log\Big(1 - D_{Ra}\Big(y_f, y_r\Big)\Big)\Big] \tag{7}$$

and the adversarial loss for generator is formulated with a symmetrical form:

$$\mathcal{L}_{G^{Ra}} = -\mathbb{E}_{y_r}\Big[\log\Big(1 - D_{Ra}\Big(y_r, y_f\Big)\Big)\Big] - \mathbb{E}_{y_f}\Big[\log\Big(D_{Ra}\Big(y_f, y_r\Big)\Big)\Big] \tag{8}$$

PatchGAN (Markovian discriminator) identifies each $N \times N$ image patch real or fake. It is more suitable for the tasks which focus on detail or texture preservation. Further, we introduced the relativistic discriminator to further enhance the performance of discriminator. Compared with standard *PatchGAN*, the relativistic *PatchGAN* loss in the proposed *DNSGAN* can be expressed as:

$$\min_G \; \max_{D_k} \; \sum_{k=1,2..}^{K} \mathcal{L}_{GAN}\Big(G^{Ra}, D_k^{Ra}\Big) \tag{9}$$

The relativistic discriminator contains five convolution layers and an average pooling layer, in our experiments and we selected two scaled patches from the last and penultimate layers to obtain the scores from real or fake samples.

Proposed method builds on an end-to-end learning architecture, which accepts arbitrary image size as input. Therefore, our method is trained using image patches and applied on the whole images. The details are provided in section 3 on experiments.

3 Experiments

This section presents the experimental setup and evaluates the performance of the proposed *DNRGAN*. Comprehensive experiments are set up with several competitive methods on two low-dose CT datasets respectively with simulated and real noise. In addition, peak noise-to-signal rate (PSNR), structural similarity (SSIM), and Fréchet inception distance (FID) are used to quantitatively evaluate the results. All these metrics were calculated based on the whole images.
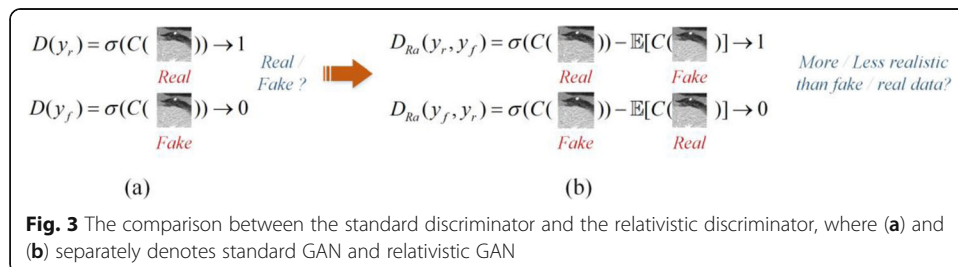
Fig. 3 The comparison between the standard discriminator and the relativistic discriminator, where (**a**) and (**b**) separately denotes standard GAN and relativistic GAN

3.1 Low-dose CT dataset with simulated noise

The Mayo clinic CT dataset [45] was used in our experiments, which is prepared for "the 2016 NIH-AAPM-Mayo Clinic Low Dose CT Ground Challenge" to evaluate competing LDCT image reconstruction algorithms. The dataset consists of 5936 normal-dose abdominal CT images with 512×512-pixel taken from 10 anonymous patients and corresponding simulated quarter-dose images after realistic noise insertion. The slice thickness and reconstruction interval in this dataset are 1.0 mm and 3.0 mm, respectively. The scanning tube potential and effective mAs used for this dataset were 120 kV and 200 mAs, respectively. All data were obtained on similar scanner models (Somatom Definition AS+, or Somatom Definition Flash operated in single-source mode, Siemens Healthcare, Forchheim, Germany). Please refer [45] for more details.

3.1.1 Experiment setting

We randomly selected 4000 slices from the LDCT images and corresponding NDCT images as training set, and the rest 1936 LDCT images were used as testing set. We generated approximately 120,000 samples with size of 128×128-pixel randomly cropped from the training set and validated the proposed model with the whole images in the testing set. The data in the experiments is normalized to [0, 1]. The batch size was set to 8. In order to speed up the training process, a PSNR-oriented model was trained. The learning rate is initialized as 2×10^{-4}. A GAN-based model is trained by fine-turning with learning rate 1×10^{-4}. For optimization, we used Adam algorithm [46] with $\beta_1 = 0.9$ and $\beta_2 = 0.99$. We implemented our model with the PyTorch framework [47] and trained on a NVIDIA Titan V GPU.

3.1.2 Components analysis

We first investigated the impacts of different modules and loss function combination for the proposed *DNSGAN* in noise suppression and structure recovery. For PSNR-oriented generator network, we first studied the effect of separate *DFN* module for noise removal, referred as *DNSN-DF*. For the enhancement module *RDN*, we mainly analyzed the effect on richer inputs, referred as *DNSN* and *DNSGAN*. For discriminator, we mainly focused on the factor of adversarial loss. A standard cross entropy loss was used in our proposed method for comparison. In addition, the improved feature loss was also analyzed. Table 1 gives the detailed descriptions on each variant combining different modules and loss functions.

A representative slice from testing set was selected to show the performance of method in Fig. 4. It is obvious that the methods with L2/L1 norm achieved smoother results, e.g., *DNSN-DF*, *DNSN-1*, and *DNSN*. Compared with *DNSN-DF*, *DNSN-1*, and *DNSN* with *RDN* module obtained more clear structures but smoother details, which resulted in higher FID scores, and revealed that the richer input is effective for structure enhancement. On the other hand, the methods with adversarial loss, such as *DNSGAN-1*, *DNSGAN*, *DNSGAN-CS*, and *DNSGAN-NF*, achieved better visual perception with lower FID scores. *DNSGAN-1* and *DNSGAN* had finer structures. In addition, improved feature loss promoted

Table 1 Summary of components and loss functions

Components	Loss function	Description
DNSN-DF	$\min \mathcal{L}_{dfn}$	Dynamic filter network with MSE loss
DNSN-1	$\min(\mathcal{L}_{dfn} + \mathcal{L}_c)$	Oriented-PSNR model with single input for RDN
DNSN	$\min(\mathcal{L}_{dfn} + \mathcal{L}_c)$	Oriented-PSNR model with multi-inputs for RDN
DNSGAN-CS	$\displaystyle \min_G \max_{D_k} \sum_{k=1,2..}^{K} \mathcal{L}_{GAN}^{Std}(G, D_k) + \lambda \mathcal{L}_{dfn} + \mathcal{L}_c + \eta \mathcal{L}_{fea}$	DNSGAN with cross entropy loss
DNSGAN-NF	$\displaystyle \min_G \max_{D_k} \sum_{k=1,2..}^{K} \mathcal{L}_{GAN}^{Ra}(G, D_k) + \lambda \mathcal{L}_{dfn} + \mathcal{L}_c$	DNSGAN without improved feature loss
DNSGAN-1	$\displaystyle \min_G \max_{D_k} \sum_{k=1,2..}^{K} \mathcal{L}_{GAN}^{Ra}(G, D_k) + \lambda \mathcal{L}_{dfn} + \mathcal{L}_c + \eta \mathcal{L}_{fea}$	DNSGAN with single-input for RDN module
DNSGAN	$\displaystyle \min_G \max_{D_k} \sum_{k=1,2..}^{K} \mathcal{L}_{GAN}^{Ra}(G, D_k) + \lambda \mathcal{L}_{dfn} + \mathcal{L}_c + \eta \mathcal{L}_{fea}$	DNSGAN with multi-input for RDN module

the structure recovery and artifact removal compared to DNSGAN-NF. The quantitative results from the whole testing set are shown in Table 2. It can be noticed that DNSN had the best PSNR and SSIM values, but DNSGAN achieved better balance between the visual perception and noise suppression.

3.1.3 Qualitative and quantitative results

In this section, *DNSN*, *DNSGAN-CS*, and *DNSGAN* were selected as our baselines to compare with other state-of-the-art methods including *BM3D*, *RedCNN* [21], and *WGAN-VGG*. A visualized result is given in Fig. 5. The zoomed regions (indicated by red and blue arrows) are used to visualize structural differences. All the methods presented powerful capacity of noise removal, but *BM3D*, *RedCNN*, and *DNSN* had smoother local details. *BM3D* even introduced extra artifacts compared with the other methods. *DNSN* achieved the best PSNR scores with only L1/L2 norm. The adversarial learning based methods brought better visual perception than the PSNR-oriented methods. However, *WGAN-VGG* generated some unpleasing artifacts. *DNSGAN-CS* achieved better qualitative results on noise reduction and structure restoration. The improved discriminator further enhanced the ability of model on details retaining.

In addition, we introduced the noise power spectrum (NPS) [48] to validate the performance of our method. We selected a structure-rich ROI area from the LDCT image, which was indicated by an orange rectangle in Fig. 5, to calculate the 2D and 1D NPS metrics and the results using different methods are shown in Fig. 6. All the methods presented the ability in noise removal to varying degree. However, undesired waxy artifacts leaded *BM3D* has a higher peak. Although *WGAN-VGG* brought better visual perception than *BM3D*, unpleasant details lead to a higher peak in 1D NPS curve and lower metrics (i.e., PSNR and SSIM). Instead, our method achieved a better trade-off between the noise removal and visual perception than other methods.

Fig. 4 Transverse CT image through the abdomen. The red and blue arrows denote zoomed ROI regions. The display window of this slice is [−180, 200] HU and quantitative metrics are from entire image. Zoomed in for better visualization
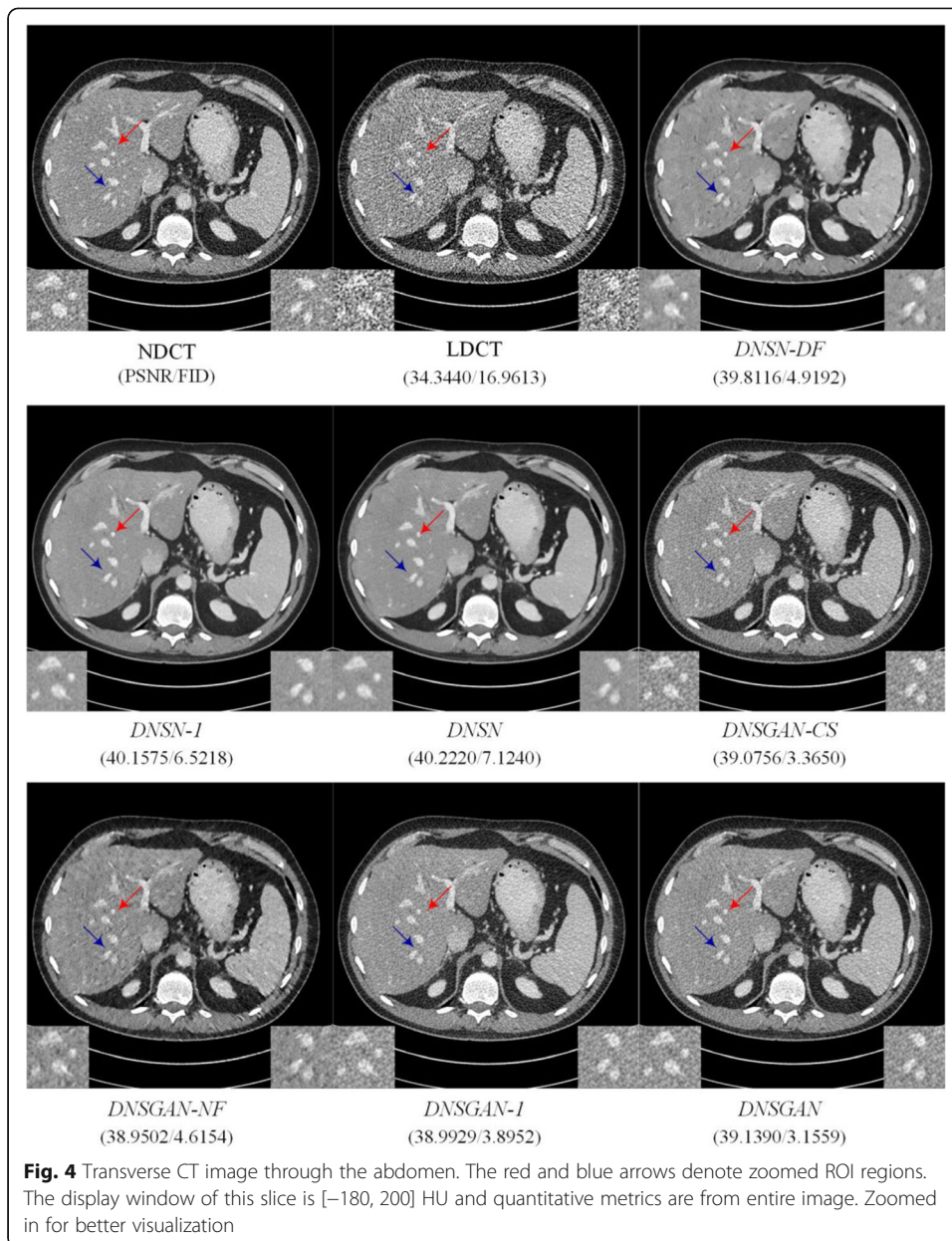
Figure 7 presents the results in coronal and sagittal planes with different methods. All the methods demonstrated similar trend to that in transverse plane and our method show the best balance between the fine structure recovery and noise reduction.
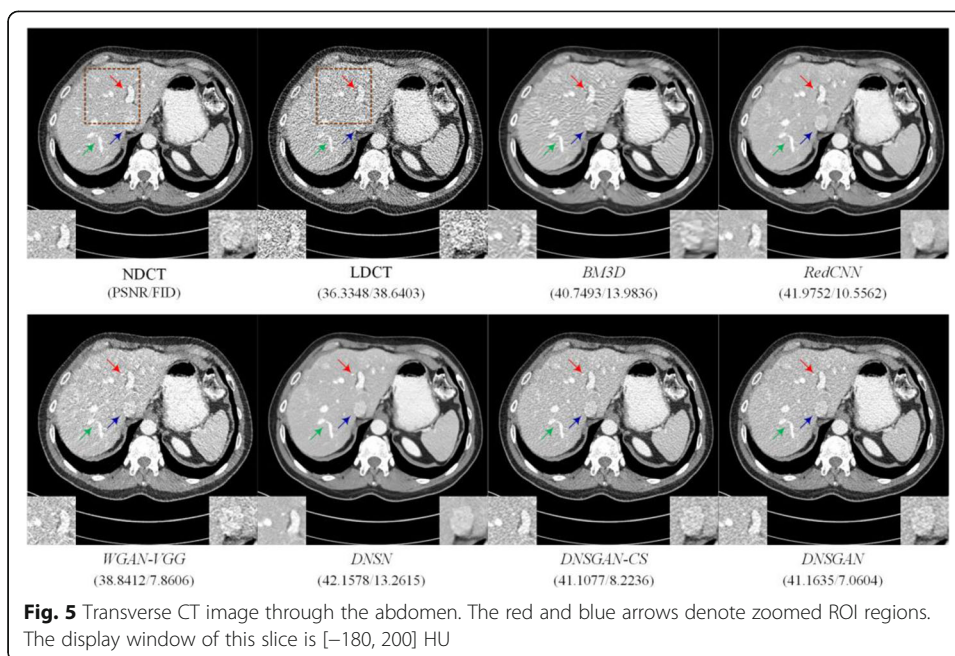
The quantitative results on the whole testing set are given in Table 3. *DNSN* achieved the higher PSNR and SSIM scores. The proposed *DNSGAN* achieved a better balance among these metrics.

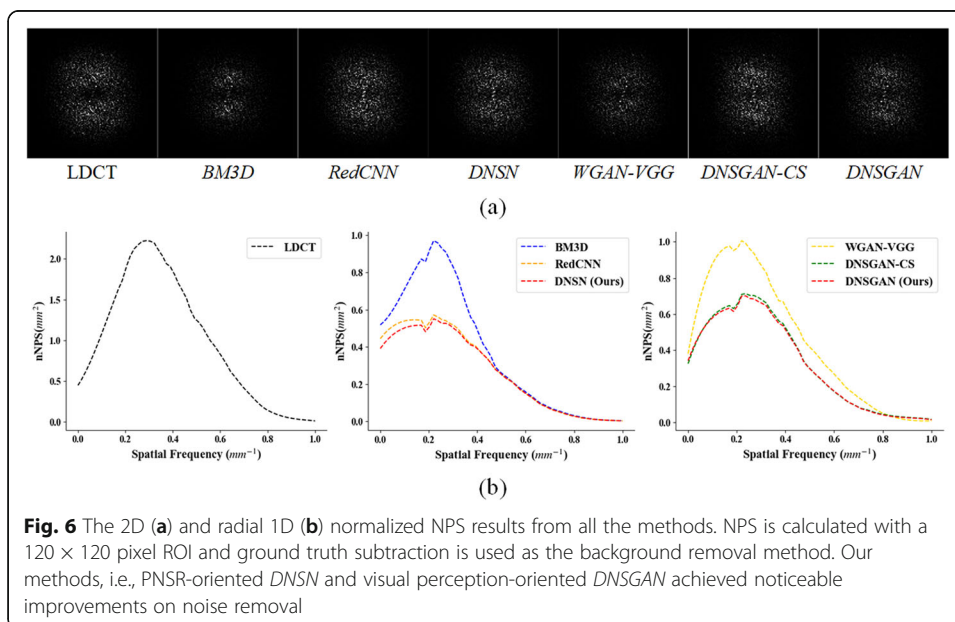### 3.2 Low-dose CT dataset with real noise

The proposed *DNSGAN* was also validated on a real low-dose CT dataset, Dongpu General Hospital (DGH) dataset, which contains 4872 one-sixth-dose head CT scans
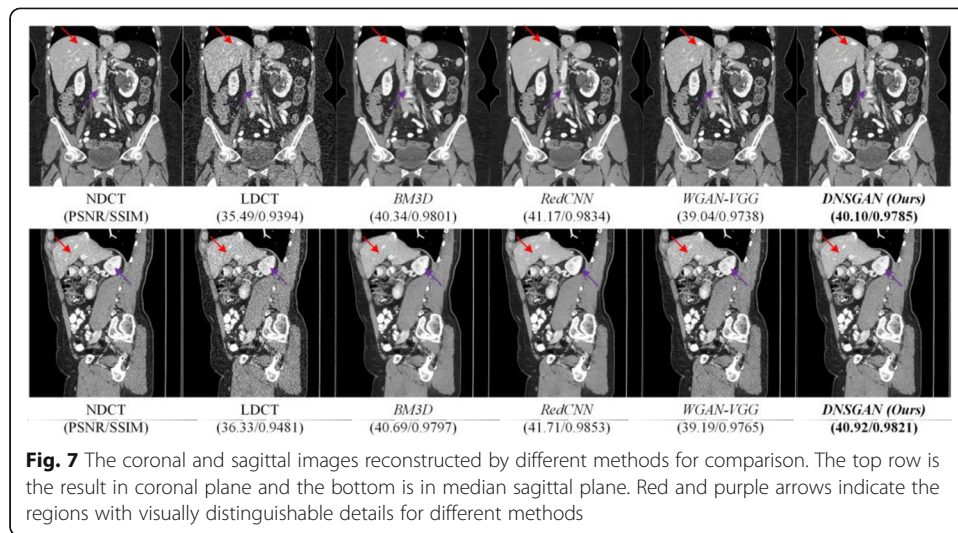
Table 2 The qualitative results of compared modules in the TestSet (mean±std)

	LDCT	DNSN-DF	DNSN-1	DNSN	DNSGAN-CS	DNRSAN-NF	DNSGAN-1	DNSGAN
PSNR	38.127±1.916	43.248±1.579	43.535±1.570	**43.630±1.587**	42.616±1.638	42.286±1.501	42.531±1.642	42.670±1.635
SSIM	0.961±0.016	0.989±0.004	0.989±0.004	**0.990±0.004**	0.988±0.005	0.987±0.005	0.987±0.005	0.987±0.005
FID	7.069	1.988	2.358	2.451	0.542	1.604	0.577	**0.503**

Fig. 5 Transverse CT image through the abdomen. The red and blue arrows denote zoomed ROI regions. The display window of this slice is [−180, 200] HU

with 512×512 pixels and corresponding normal-dose CT images from 11 patients with representative protocols. All data were obtained on same scanner (MinFound ScintCare CT16). Each head CT scan data from patients consist of three different scan thicknesses, e.g., 1.16 mm, 2.32 mm, and 4.64 mm. In addition, these CT scans are acquired by two different reconstructed kernels. Since the low-dose CT images and corresponding normal-dose CT images are not in perfect registration due to the error in the patient table re-positioning and uncertainty in the source angle initialization, in this experiment, we just validated the generalization performance of the proposed method on different noise-level datasets. The data in the experiments is normalized to [0 1].

Fig. 6 The 2D (**a**) and radial 1D (**b**) normalized NPS results from all the methods. NPS is calculated with a 120 × 120 pixel ROI and ground truth subtraction is used as the background removal method. Our methods, i.e., PNSR-oriented *DNSN* and visual perception-oriented *DNSGAN* achieved noticeable improvements on noise removal

Fig. 7 The coronal and sagittal images reconstructed by different methods for comparison. The top row is the result in coronal plane and the bottom is in median sagittal plane. Red and purple arrows indicate the regions with visually distinguishable details for different methods

3.2.1 Experiment setting

Considering DGH dataset contains varying scan thickness, which results in different noise levels in LDCT images. The dataset is divided into three parts according to the scan thickness, referred as DGH-L, DGH-M, and DGH-H, which separately denote different noise-level LDCT with the thickness of 4.64 mm, 2.32 mm, and 1.16 mm. In this experiment, we did not retrain models due to the non-ideal data situation. An alternative method was adopted that the pre-trained model on Mayo dataset was used to evaluate the results of DGH dataset, which is effective to evaluate the generalization ability of the proposed method for different data sources.

In addition, due to the lack of referenced images with accurate registration, PSNR and SSIM were abandoned in DGH dataset. Instead, the histogram of the gray-level, and FID were used to measure the capacity of model for noise removal and structure recovery.

3.2.2 Results for blind image restoration

One slice is selected from the DGH-L dataset and shown in Fig. 8. It is obvious that *BM3D* led to smoother result than *RedCNN* and *DNRN*. *WGAN-VGG*, *DNSGAN-CS*, and *DNSGAN* generate better results, but *WGAN-VGG* introduced extra artifacts at the edges. The histogram of DGH-L is illustrated in Fig. 9. All the methods tended to produce similar distribution with the ground truth in

Table 3 The qualitative results of compared methods in the TestSet (mean±std)

	LDCT	BM3D	RedCNN	WGAN-VGG	DNSN	DNSGAN-CS	DNSGAN
PSNR	38.127±1.916	41.932±1.427	43.496±1.577	40.367±1.471	**43.630±1.587**	42.616±1.638	42.670±1.635
SSIM	0.961±0.016	0.983±0.006	0.989±0.004	0.980±0.008	**0.990±0.004**	0.988±0.005	0.987±0.005
FID	7.069	5.203	2.027	1.052	2.451	0.542	**0.503**

Fig. 8 Transverse CT slice of the head from DGH-L. The display window is [−20, 90] HU. The red, blue, and green arrows indicate the richer detailed edges of local structure. Zoomed in for better visualization
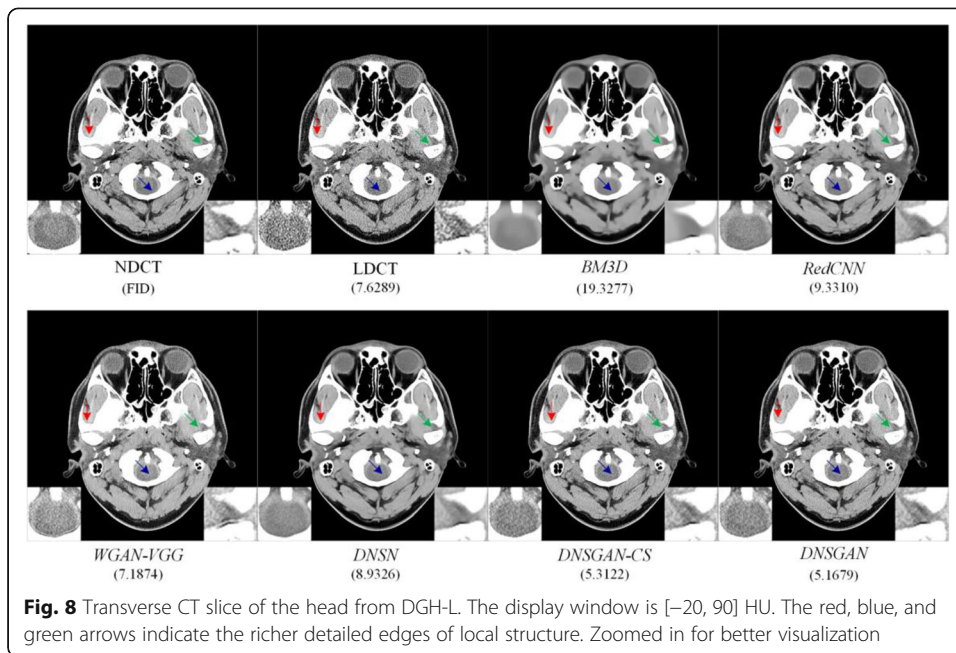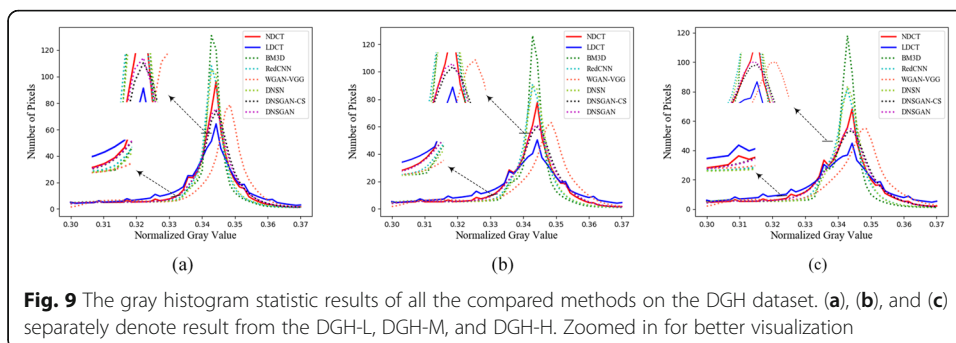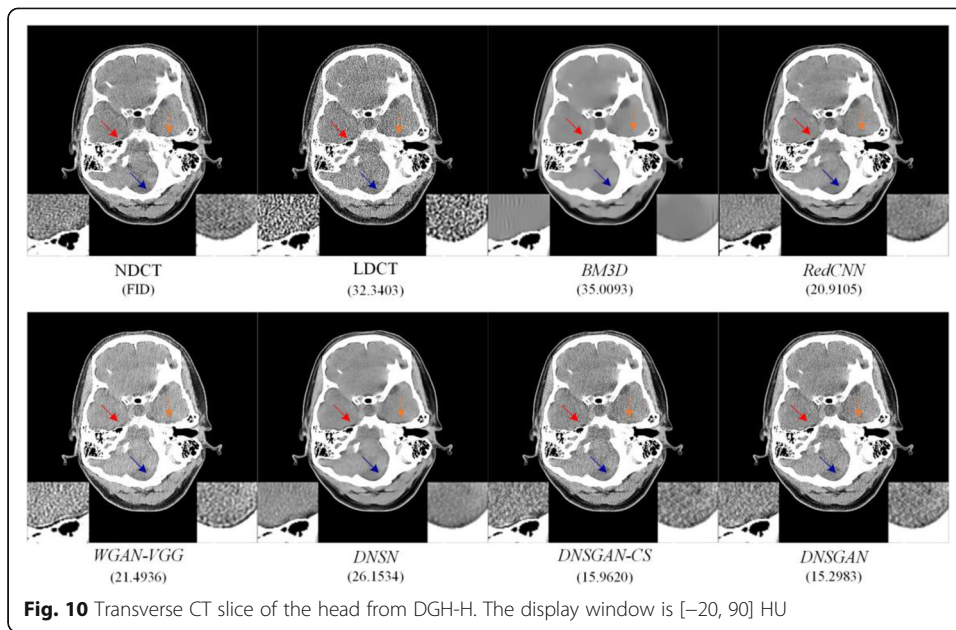
Fig. 9a. However, PSNR-oriented methods had smoother curves due to lack of finer details. Proposed *DNSGAN* fitted the curve of ground truth best. Similar trends can be observed from DGH-M and DGH-H in Fig. 9b and c.

In order to evaluate the robustness of the proposed method further, a slice with higher noise-level from DGH-H is shown in Fig. 10. *DNSGAN* still achieved the better metric than others. In addition, Table 4 gives the statistical results of FID produced by different methods on each datasets. All the methods presented strong ability on noise removal, but *BM3D* led to the worst FID value due to over-smoothing structures. Although *RedCNN* and *DNSN* had better results than *BM3D*, they had lower FID values than LDCT due to lack of finer details. GAN-based models achieved a better balance between the noise removal and detail preservation, but *WGAN-VGG* brought extra artifacts near the edges due to its poor generalization.

Furthermore, in Table 4, we can find that all the supervised learning based methods attained the best metric on DGH-M. However, CNN-based models

Fig. 9 The gray histogram statistic results of all the compared methods on the DGH dataset. (**a**), (**b**), and (**c**) separately denote result from the DGH-L, DGH-M, and DGH-H. Zoomed in for better visualization

Fig. 10 Transverse CT slice of the head from DGH-H. The display window is [−20, 90] HU

achieved better results on DGH-L and followed by DGH-H, but for GAN-based models, both *WGAN-VGG* and ours have opposite trend. Considering that all the methods were trained on Mayo dataset with specific low dose scans (e.g., quarter-dose), they tend to achieve better results on similar or lower noise levels, but for GAN-based methods, extra discriminative constraints provide more generalized ability on higher noise-level, which enables better results on DGH-H than DGH-L. Even so, the proposed *DNSGAN* still had the best metrics on all the datasets.

4 Conclusion

In this paper, we mainly propose a disentangled LDCT restoration model-*DNSG AN*, which explicitly decouples noise removal into two steps: noise suppression and structure recovery and achieves a better balance between quantitative metrics and visual perception than other state-of-the-art methods. In addition, some advanced techniques including dynamic filter network and residual dense network were introduced. Relativistic multi-scaled *PatchGAN* was also injected into the discriminator network to recover finer structures further. Experiments on both datasets with simulated and real noise respectively show that the proposed *DNSG AN* has competitive performance for LDCT restoration and strong generalization for different imaging protocols.

Table 4 The results of the FID for each methods on the DGH dataset

	LDCT	BM3D	RedCNN	WGAN-VGG	DNRN	DNRGAN-CS	DNRGAN
DGH-L	9.298779	12.592485	9.497512	8.939175	9.760384	8.738773	**8.738769**
DGH-M	9.226559	12.355052	9.376246	8.547891	9.695343	8.179189	**8.178490**
DGH-H	10.097047	13.119213	9.753763	8.831236	10.323053	8.351237	**8.336177**

Abbreviations

GAN: Generative adversarial network; LDCT: Low-dose CT; NDCT: Normal-dose CT; IR: Iterative reconstruction; PSNR: Peak signal-to-noise ratio; SSIM: Structural similarity; FID: Fréchet inception distance; DFN: Dynamic filter network; RDN: Residual dense network

Availability of data and materials

Please contact the author for DGH data request, and the Mayo dataset used in this work is available at http://www.aapm.org/GrandChallenge/LowDoseCT/.

Declarations

Competing interests

The authors declare that they have no competing interests.

References

1. H. Shan, A. Padole, F. Homayounieh, U. Kruger, R.D. Khera, C. Nitiwarangkul, et al., Competitive performance of a modularized deep neural network compared to commercial algorithms for low-dose CT image reconstruction. Nature Machine Intelligence **1**, 269–276 (2019). https://doi.org/10.1038/s42256-019-0057-9
2. J. Wang, H. Lu, T. Li, Z. Liang, *Sinogram noise reduction for low-dose CT by statistics-based nonlinear filters* (Image Processing, Medical Imaging, 2005). https://doi.org/10.1117/12.595662
3. J. Wang, T. Li, H. Lu, Z. Liang, Penalized weighted least-squares approach to sinogram noise reduction and image reconstruction for low-dose X-ray computed tomography. IEEE Trans on Medical Imaging **25**, 1272–1283 (2006). https://doi.org/10.1109/tmi.2006.882141
4. A.K. Hara, R.G. Paden, A.C. Silva, J.L. Kujak, H.J. Lawder, W. Pavlicek, Iterative reconstruction technique for reducing body radiation dose at CT: feasibility study. AJR Am. J. Roentgenol. **193**(3), 764–771 (2009). https://doi.org/10.2214/ajr.09.2397
5. M. Beister, D. Kolditz, W.A. Kalender, Iterative reconstruction methods in X-ray CT. Physica medica : PM : an international journal devoted to the applications of physics to medicine and biology : official journal of the Italian Association of Biomedical Physics **28**(2), 94–108 (2012). https://doi.org/10.1016/j.ejmp.2012.01.003
6. B.K. Man, S. Basu, Distance-driven projection and backprojection in three dimensions. Phys. Med. Biol. **49**(11), 2463–2475 (2004). https://doi.org/10.1088/0031-9155/49/11/024
7. I.A. Elbakri, J.A. Fessler, Segmentation-free statistical image reconstruction for polyenergetic x-ray computed tomography with experimental validation. Phys. Med. Biol. **48**, 2453–2468 (2003). https://doi.org/10.1088/0031-9155/48/15/314
8. Y. Liu, J. Ma, Y. Fan, Z. Liang, Adaptive-weighted total variation minimization for sparse data toward low-dose x-ray computed tomography image reconstruction. Phys. Med. Biol. **57**(23), 7923–7946 (2012). https://doi.org/10.1088/0031-9155/57/23/7923
9. Y. Chen, J. Ma, Q. Feng, L. Luo, P. Shi, W. Chen, Nonlocal prior Bayesian tomographic reconstruction. Journal of Mathematical Imaging and Vision **30**, 133–146 (2008). https://doi.org/10.1007/s10851-007-0042-5
10. J. Ma, J. Huang, Q. Feng, H. Zhang, H. Lu, Z. Liang, W. Chen, Low-dose computed tomography image restoration using previous normal-dose scan. Med. Phys. **38**(10), 5713–5731 (2011). https://doi.org/10.1118/1.3638125
11. Y. Chen, X. Yin, L. Shi, H. Shu, L. Luo, J.L. Coatrieux, C. Toumoulin, Improving abdomen tumor low-dose CT images using a fast dictionary learning based processing. Phys. Med. Biol. **58**(16), 5803–5819 (2013). https://doi.org/10.1088/0031-9155/58/16/5803
12. Y. Chen, L. Shi, Q. Feng, J. Yang, H. Shu, L. Luo, et al., Artifact suppressed dictionary learning for low-dose CT image processing. IEEE Trans on medical imaging **33**, 2271–2292 (2014). https://doi.org/10.1109/TMI.2014.2336860
13. P.F. Feruglio, C. Vinegoni, J. Gros, A. Sbarbati, R. Weissleder, Block matching 3D random noise filtering for absorption optical projection tomography. Phys. Med. Biol. **55**(18), 5401–5419 (2010). https://doi.org/10.1088/0031-9155/55/18/009
14. Y. Chen, Y. Zhang, J. Yang, Q. Cao, G. Yang, J. Chen, … & Q. Feng, Curve-like structure extraction using minimal path propagation with backtracking. IEEE Trans. Image Process., 25, 988-1003 (2015). Doi: https://doi.org/10.1109/tip.2015.2496279.
15. Y. Chen, Y. Zhang, H. Shu, J. Yang, L. Luo, J.L. Coatrieux, Q. Feng, Structure-adaptive fuzzy estimation for random-valued impulse noise suppression. IEEE Trans on Circuits and Systems for Video Technology **28**, 414–427 (2016). https://doi.org/10.1109/tcsvt.2016.2615444

16. H. Chen, Y. Zhang, W. Zhang, P. Liao, K. Li, J. Zhou, G. Wang, Low-dose CT via convolutional neural network. J. Biomedical optics express **8**(2), 679–694 (2017). https://doi.org/10.1364/BOE.8.000679

17. O. Ronneberger, P. Fischer, T. Brox, in *International Conference on Medical image computing and computer-assisted intervention*. U-net: convolutional networks for biomedical image segmentation (Springer, Cham, 2015, October), pp. 234–241. https://doi.org/10.1007/978-3-319-24574-4_28

18. J. Liu, Y. Hu, J. Yang, Y. Chen, H. Shu, L. Luo, ... & G. Coatrieux, 3D feature constrained reconstruction for low-dose CT imaging. IEEE Trans on Circuits and Systems for Video Technology, 28, 1232-1247 (2016). Doi: https://doi.org/10.1109/tcsvt.2016.2643009.

19. J. Liu, J. Ma, Y. Zhang, Y. Chen, J. Yang, H. Shu, et al., Discriminative feature representation to improve projection data inconsistency for low dose CT imaging. J. IEEE Trans on medical imaging **36**, 2499–2509 (2017). https://doi.org/10.1109/tmi.2017.2739841

20. E. Kang, J. Min, J.C. Ye, WaveNet: a deep convolutional neural network using directional wavelets for low-dose X-ray CT reconstruction. Med. Phys. **44**(10), e360–e375 (2017). https://doi.org/10.1002/mp.12344

21. H. Chen, Y. Zhang, M.K. Kalra, F. Lin, Y. Chen, P. Liao, et al., Low-dose CT with a residual encoder-decoder convolutional neural network. IEEE Trans on medical imaging **36**, 2524–2535 (2017). https://doi.org/10.1109/TMI.2017.2715284

22. W. Du, H. Chen, Z. Wu, H. Sun, P. Liao, Y. Zhang, Stacked competitive networks for noise reduction in low-dose CT. J. PloS one **12**, e0190069 (2017). https://doi.org/10.1371/journal.pone.0190069

23. H. Chen, Y. Zhang, Y. Chen, J. Zhang, W. Zhang, H. Sun, G. Wang, LEARN: learned experts' assessment-based reconstruction network for sparse-data CT. IEEE Trans on medical imaging **37**, 1333–1347 (2018). https://doi.org/10.1109/tmi.2018.2805692

24. Z. Wang, A.C. Bovik, H.R. Sheikh, E.P. Simoncelli, Image quality assessment: from error visibility to structural similarity. IEEE Trans. Image Process. **13**, 600–612 (2004). https://doi.org/10.1109/TIP.2003.819861

25. M. Heusel, H. Ramsauer, T. Unterthiner, B. Nessler, S. Hochreiter, in *Advances in Neural Information Processing Systems*. Gans trained by a two time-scale update rule converge to a local Nash equilibrium (2017), pp. 6626–6637

26. Q. Yang, P. Yan, Y. Zhang, H. Yu, Y. Shi, X. Mou, et al., Low-dose CT image denoising using a generative adversarial network with Wasserstein distance and perceptual loss. IEEE trans on medical imaging **37**, 1348–1357 (2018). https://doi.org/10.1109/TMI.2018.2827462

27. X. Yi, P. Babyn, Sharpness-aware low-dose CT denoising using conditional generative adversarial network. J. Digit. Imaging **31**(5), 655–669 (2018). https://doi.org/10.1007/s10278-018-0056-0

28. C. You, Q. Yang, L.G. Gjesteby, S.J. Li, Z. Zhang, et al., Structurally-sensitive multi-scale deep neural network for low-dose CT denoising. IEEE Access **6**, 41839–41855 (2018). https://doi.org/10.1109/ACCESS.2018.2858196

29. H. Shan, Y. Zhang, Q. Yang, U. Kruger, M.K. Kalra, L. Sun, et al., 3-D convolutional encoder-decoder network for low-dose CT via transfer learning from a 2-D trained network. IEEE Trans. Med. Imaging **37**(6), 1522–1534 (2018). https://doi.org/10.1109/TMI.2018.2832217

30. W. Du, H. Chen, P. Liao, H. Yang, G. Wang, Y. Zhang, Visual attention network for low-dose CT. IEEE Signal Processing Letters **26**(8), 1152–1156 (2019). https://doi.org/10.1109/LSP.2019.2922851

31. X. Yin, Q. Zhao, J. Liu, W. Yang, J. Yang, G. Quan, et al., Domain progressive 3D residual convolution network to improve low dose CT imaging. IEEE Trans on medical imaging (2019). https://doi.org/10.1109/tmi.2019.2917258

32. M. Arjovsky, S. Chintala, & L. Bottou, Wasserstein gan. arXiv preprint arXiv:1701.07875 (2017).

33. J. Johnson, A. Alahi, F. Li, in *European conference on computer vision*. Perceptual losses for real-time style transfer and super-resolution (Springer, Cham, 2016, October), pp. 694–711

34. I. Goodfellow, J. Pouget-Abadie, M. Mirza, B. Xu, D. Warde-Farley, S. Ozair, et al., in *Advances in neural information processing systems*. Generative adversarial nets (2014), pp. 2672–2680

35. B. Zhu, J.Z. Liu, S.F. Cauley, B.R. Rosen, M.S. Rosen, Image reconstruction by domain-transform manifold learning. Nature **555**(7697), 487 (2018). https://doi.org/10.1038/nature25988

36. C. Ledig, L. Theis, F. Huszár, J. Caballero, A. Cunningham, A. Acosta, et al., in *Proceedings of the IEEE conference on computer vision and pattern recognition*. Photo-realistic single image super-resolution using a generative adversarial network (2017), pp. 4681–4690

37. X. Jia, B. De Brabandere, T. Tuytelaars, L.V. Gool, in *Advances in Neural Information Processing Systems*. Dynamic filter networks (2016), pp. 667–675

38. C. Dong, C.C. Loy, K. He, X. Tang, Image super-resolution using deep convolutional networks. IEEE Trans on pattern analysis and machine intelligence **38**, 295–307 (2015). https://doi.org/10.1109/tpami.2015.2439281

39. X. Wang, K. Yu, S. Wu, J. Gu, Y. Liu, C. Dong, et al., in *Proceedings of the European Conference on Computer Vision (ECCV)*. Esrgan: Enhanced super-resolution generative adversarial networks (2018)

40. G. Huang, Z. Liu, L. Van Der Maaten, K.Q. Weinberger, in *Proceedings of the IEEE conference on computer vision and pattern recognition*. Densely connected convolutional networks (2017), pp. 4700–4708

41. K. Simonyan, A. Zisserman, Very deep convolutional networks for large-scale image recognition. arXiv preprint arXiv **1409**, 1556 (2014)

42. P. Isola, J.Y. Zhu, T. Zhou, A.A. Efros, in *Proceedings of the IEEE conference on computer vision and pattern recognition*. Image-to-image translation with conditional adversarial networks (2017), pp. 1125–1134

43. T.C. Wang, M.Y. Liu, J.Y. Zhu, A. Tao, J. Kautz, B. Catanzaro, in *Proceedings of the IEEE conference on computer vision and pattern recognition*. High-resolution image synthesis and semantic manipulation with conditional gans (2018), pp. 8798–8807

44. A. Jolicoeur-Martineau, The relativistic discriminator: a key element missing from standard GAN. arXiv preprint arXiv: 1807.00734 (2018).

45. AAPM, "Low dose CT grand challenge," 2017. [Online]. Available:http://www.aapm.org/GrandChallenge/LowDoseCT/#.

46. D. P. Kingma, & J. Ba, Adam: A method for stochastic optimization. arXiv preprint arXiv:1412.6980 (2014).

47. A. Paszke et al., in *Proc. Neural Inf. Process. Syst.*. Automatic differentiation in pytorch (2017)

48. M. Kijewski, P. Judy, The noise power spectrum of CT images. Phys. Med. Biol. **32**, 565–575 (1987)

Publisher's Note