DU-GAN: Generative Adversarial Networks With Dual-Domain U-Net-Based Discriminators for Low-Dose CT Denoising

Zhizhong Huang⬤, *Graduate Student Member, IEEE*, Junping Zhang⬤, *Member, IEEE*,
Yi Zhang⬤, *Senior Member, IEEE*, and Hongming Shan⬤, *Member, IEEE*

Abstract—Low-dose computed tomography (LDCT) has drawn major attention in the medical imaging field due to the potential health risks of CT-associated X-ray radiation to patients. Reducing the radiation dose, however, decreases the quality of the reconstructed images, which consequently compromises the diagnostic performance. Over the past few years, various deep learning techniques, especially generative adversarial networks (GANs), have been introduced to improve the image quality of LDCT images through denoising, achieving impressive results over traditional approaches. GAN-based denoising methods usually leverage an additional classification network, i.e., discriminator, to learn the most discriminate difference between the denoised and normal-dose images and, hence, regularize the denoising model accordingly; it often focuses either on the global structure or local details. To better regularize the LDCT denoising model, this article proposes a novel method, termed DU-GAN, which leverages U-Net-based discriminators in the GAN framework to learn both global and local differences between the denoised and normal-dose images in both image and gradient domains. The merit of such a U-Net-based discriminator is that it can not only provide the per-pixel feedback to the denoising network through the outputs of the U-Net but also focus on the global structure in a semantic level through the middle layer of the U-Net. In addition to the adversarial training in the image domain, we also apply another U-Net-based discriminator in the image gradient domain to alleviate the artifacts caused by photon starvation and enhance the edge of the denoised CT images. Furthermore, the CutMix technique enables the per-pixel outputs of the U-Net-based discriminator to provide radiologists with a confidence map to visualize the uncertainty of the denoised results, facilitating the LDCT-based screening and diagnosis. Extensive experiments on the simulated and real-world datasets demonstrate superior performance over recently published methods both qualitatively and quantitatively. Our source code is made available at https://github.com/Hzzone/DU-GAN.

Index Terms—Artifact removal, convolutional neural network (CNN), generative adversarial network (GAN), image translation, low-dose computed tomography (LDCT) denoising, U-Net, uncertainty estimation.

Zhizhong Huang and Junping Zhang are with the Shanghai Key Laboratory of Intelligent Information Processing and the School of Computer Science, Fudan University, Shanghai 200433, China (e-mail: zzhuang19@fudan.edu.cn; jpzhang@fudan.edu.cn).

Yi Zhang is with the College of Computer Science, Sichuan University, Chengdu 610065, China (e-mail: yzhang@scu.edu.cn).

Hongming Shan is with the Institute of Science and Technology for Brain-inspired Intelligence and the MOE Frontiers Center for Brain Science, Fudan University, Shanghai 200433, China, and also with the Shanghai Center for Brain Science and Brain-Inspired Technology, Shanghai 201210, China (e-mail: hmshan@fudan.edu.cn).

I. INTRODUCTION

COMPUTED tomography (CT) can provide the cross-sectional images of the internal body by the X-ray radiation, which is one of the most important imaging modalities in clinical diagnosis. Although CT plays an essential role in diagnosing diseases, the widespread use of CT is raising more and more public concerns toward its safety since CT-related X-ray radiation may cause unavoidable damage to the health of humans and induce cancers. Consequently, reducing the radiation dose of CT as low as reasonably achievable (also known as ALARA) is a well-accepted principle in CT-related research over the past decades [1]. The reduction of radiation dose, however, inevitably brings the noise and artifacts into the reconstructed images, severely compromising the subsequent diagnosis and other tasks such as low-dose computed tomography (LDCT)-based lung nodule classification [2].

A straightforward way to address this issue is to reduce the noise in the LDCT image [3], [4]. However, it remains a challenging problem due to its ill-posed nature. In recent years, various deep learning-based methods have been proposed for LDCT denoising [5]–[11], achieving impressive results. There are two key components in designing a denoising model: network architecture and loss function; the former one can determine the capacity of the denoising model, while the latter one can control how the denoised images visually look like [7]. Although different network architectures, such as 2-D convolutional neural networks (CNNs) [5], 3-D CNNs [7], [10], and residual encoder–decoder CNNs (RED-CNNs) [12], have been explored for LDCT denoising, literature has shown that the loss function is relatively more important than the network architecture as it has a direct impact on the image quality [7], [13].

One of the most popular loss functions is the mean-squared error (MSE), which computes the average of the squares of the per-pixel errors between the denoised and normal-dose images. Although gaining impressive performance in terms of peak signal-to-noise ratio (PSNR), MSE usually leads to oversmoothened images, which has been proven to poorly correlate with the human perception of image quality [14], [15]. In view of this observation, alternative loss functions, such as perceptual loss, ℓ_1 loss, and adversarial loss, have been investigated for LDCT denoising. Among them, adversarial loss has been shown to be a powerful one as it can dynamically measure the similarity between the denoised and normal-dose images during the training, which enables the denoised images to preserve more texture information from normal-dose one. The computation of the adversarial loss is based on the discriminator, which is a classification network to learn a representation differentiating the denoised images from the normal-dose images; it can measure the most discriminant difference either in a global or local level, depending on that one unit of the output of discriminator corresponds to the whole image or a local region. Such a discriminator is prone to forgetting previous difference because the distribution of synthetics samples shifts as the generator constantly changes through training, failing to maintain a powerful data representation to characterize the global and local image difference [16]. As a result, it often results in the generated images with discontinued and mottled local structures [17] or images with incoherent geometric and structural patterns [18]. In addition to the noise, LDCT images may contain severe streak artifacts caused by photon starvation, which may not be effectively removed through the loss function solely in the image domain.

To learn a powerful data representation to regularize the denoising model in the adversarial training, we propose a U-Net-based [19] discriminator in the generative adversarial network (GAN) framework for LDCT denoising, termed DU-GAN, which can simultaneously learn the global and local difference between the denoised and normal-dose images in image and gradient domains. More specifically, our proposed discriminator follows the U-Net architecture, including an encoder and a decoder network, where the encoder encodes the input to a scalar value focusing on the global structures, whereas the decoder reconstructs a per-pixel confidence map capturing the changes of local details between the denoised and normal-dose images. In doing so, it can provide not only the per-pixel feedback but also the global structural difference to the denoising network. In addition to the adversarial training in the image domain, we also apply another U-Net-based discriminator in the image gradient domain to alleviate the artifacts caused by photon starvation and enhance the edge of the denoised images. Moreover, to regularize the U-Net-based discriminator, we introduce the CutMix data augmentation to mix the denoised and normal-dose images. Consequently, the U-Net-based discriminator can provide radiologists with the per-pixel outputs as a confidence map to visualize the uncertainty of the denoised results, which can facilitate radiologists' screening and diagnosis when using the denoised LDCT images.

The benefits of the proposed DU-GAN are as follows.

1) Unlike existing GAN-based denoising methods that use a classification as the discriminator, the proposed DU-GAN utilizes a U-Net-based discriminator for LDCT denoising, which can simultaneously learn the global and local differences between the denoised and normal-dose images. Consequently, it can provide not only the per-pixel feedback but also the global structural difference to the denoising model.

2) In addition to adversarial training in the image domain, the proposed DU-GAN also performs adversarial training in the image gradient domains, which can alleviate the streak artifacts caused by photon starvation and enhance the edge of the denoised images.

3) The proposed DU-GAN can provide radiologists with a confidence map visualizing the uncertainty of the denoised results through the CutMix technique, which could facilitate radiologists' screening and diagnosis when using the denoised LDCT images.

4) Extensive experiments on simulated and real-world datasets demonstrate the effectiveness of the proposed method through both qualitative and quantitative comparisons.

The remainder of this article is organized as follows. We briefly survey the developments of the LDCT denoising methods and GANs in Section II. We present our LDCT denoising framework DU-GAN with dual-domain U-Net-based discriminators and then introduce the CutMix regularization technique as well as the network architectures and loss functions in our framework in Section III, followed by both qualitative and quantitative comparisons with the state-of-the-art methods on the simulated and real-world datasets in Section IV. Finally, we conclude this article in Section V.

II. RELATED WORK

This section briefly surveys the development of LDCT denoising and GANs.

A. LDCT Denoising

The noise reduction algorithms for LDCT can be summarized into three categories: 1) sinogram filtration; 2) iterative reconstruction; and 3) image postprocessing. As a significant difference from routine CT, the LDCT acquires noisy sinogram data from the scanner. A straightforward solution is to perform the denoising process on the sinogram data before image reconstruction, i.e., sinogram filtration-based methods [20]–[22]. Iterative reconstruction methods combine the statistics of raw data in the sinogram domain [23], [24] and the prior information in the image domain, such as total variation [25] and dictionary learning [26]; these pieces of generic information can be effectively integrated into the maximum likelihood and compressed sensing frameworks. These two categories, however, require the access to raw data that are typically unavailable from commercial CT scanner.

Different from the previous two categories, image postprocessing methods directly operate on the reconstructed images that are publicly available after removing patient
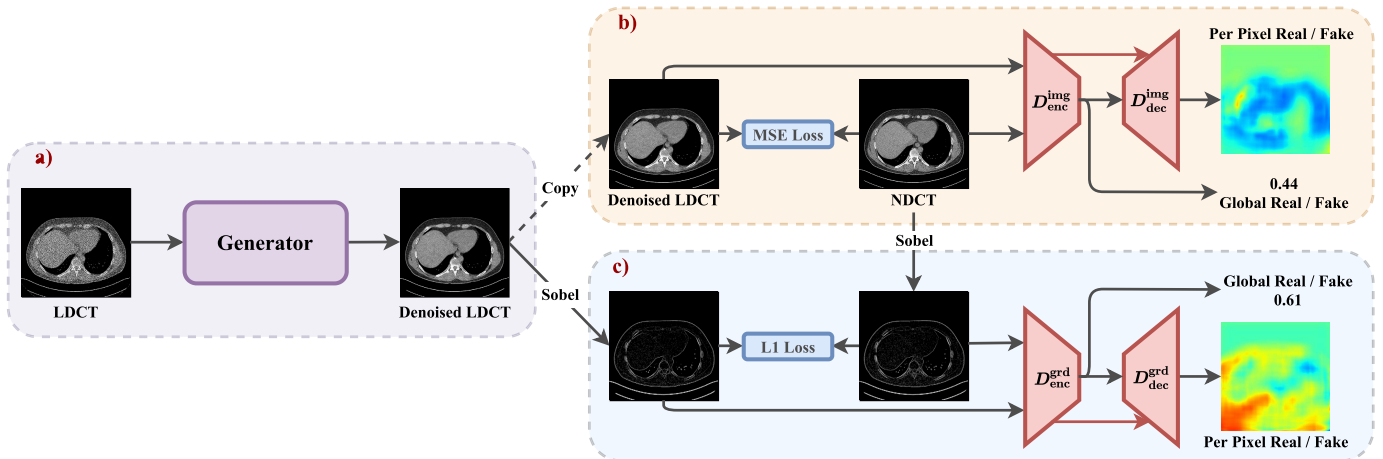
Fig. 1.   Overall framework of our proposed DU-GAN. The generator produces denoised LDCT images, and two independent branches with U-Net-based discriminators perform at the image and gradient domains. The U-Net-based discriminator provides both global structure and local per-pixel feedback to the generator. Furthermore, the image discriminator encourages the generator to produce photorealistic CT images, while the gradient discriminator is utilized for better edge and alleviating streak artifacts caused by photon starvation. (a) Denoising Process, (b) Image Domain Branch, and (c) Gradient Domain Branch.

privacy. Traditional methods, such as nonlocal means [27] and block-matching 3-D [28], however, lead to the loss of some critical structural details and result in oversmoothened denoised LDCT images. The rapid development of deep learning techniques has advanced many medical applications. In LDCT denoising, deep learning-based models have achieved impressive results [5], [7], [9], [10], [12], [29]. There are two critical components in designing a deep learning-based denoising model: network architecture and loss function; the former one determines the capacity of a denoising model, whereas the latter one controls how the denoised images visually look like. Although the literature has proposed several different network architectures for LDCT denoising such as 2-D CNNs [5], 3-D CNN [7], [10], RED-CNN [5], and cascaded CNN [12], the literature has shown that the loss function plays a relatively more important role than network architecture as it has a direct impact on the image quality [7], [13]. The simplest loss function is the MSE, which, however, has been shown to poorly correlate with the human perception of image quality [14], [15]. In view of this observation, alternative loss functions, such as perceptual loss, ℓ_1 loss, adversarial loss, or mixed loss functions, have been investigated for LDCT denoising. Among them, the adversarial loss has been shown to be a powerful one as it can dynamically measure the similarity between the denoised and normal-dose images during the training, which enables the denoised images to preserve more texture information from normal-dose one. The adversarial loss reflects either global or local similarity, depending on the design of discriminator.

Unlike the conventional adversarial loss, the adversarial loss used in this study is based on a U-Net-based discriminator, which can simultaneously characterize the global and local differences between the denoised and normal-dose images, better regularizing the denoising model, that is, DU-GAN enjoys both advantages of the per-pixel discriminator capturing the changes at pixel level and traditional classification discriminator focusing on global structures. In addition to the

adversarial loss in the image domain, the adversarial loss in the image gradient domain proposed in this article can alleviate the streak artifacts caused by photon starvation and enhance the edge of the denoised images.

B. Generative Adversarial Networks

As one of the most hot research topics in recent years, GANs [14] and their variants have been successfully applied to various tasks [30]–[32]. They typically consist of two networks: 1) a generator learning to capture the data distribution of training data and produce new samples that are indistinguishable from the real ones and 2) a discriminator attempting to distinguish real samples from fake ones produced by the generator. These two networks are trained alternatively, ending once the balance is achieved. In the context of LDCT denoising, the generator aims to produce photorealistic denoised results to fool the discriminator, whereas the discriminator tries to distinguish the real normal-dose CT (NDCT) images and denoised ones. To foster the stability of training GANs, various variants of GANs have been proposed, such as Wasserstein GAN (WGAN) [33], WGAN with gradient penalty (WGAN-GP) [34], and least-squares GANs [35].

In this article, we adopt the least-squares GANs [35], spectral normalization [36], and U-Net-based discriminator [16] to form the GANs framework for LDCT denoising. As a significant difference, our DU-GAN performs adversarial training in both image and gradient domains, which can reduce noise and alleviate streak artifacts simultaneously. We note that the proposed DU-GAN is also suitable for other variants of GANs, such as WGAN and WGAN-GP.

III. METHODOLOGY

Fig. 1 presents the proposed DU-GAN for LDCT denoising, which contains a denoising model as a generator and two U-Net-based discriminators in both image and gradient domains. We highlight that the U-Net-based discriminator is
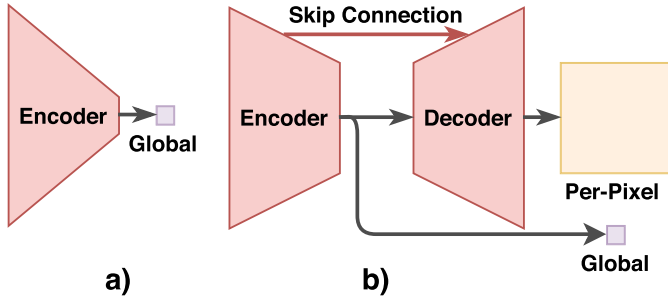
Fig. 2. Difference between (a) traditional classification discriminator and (b) U-Net-based discriminator. U-Net-based discriminator extends traditional one to capture the global and local information simultaneously.
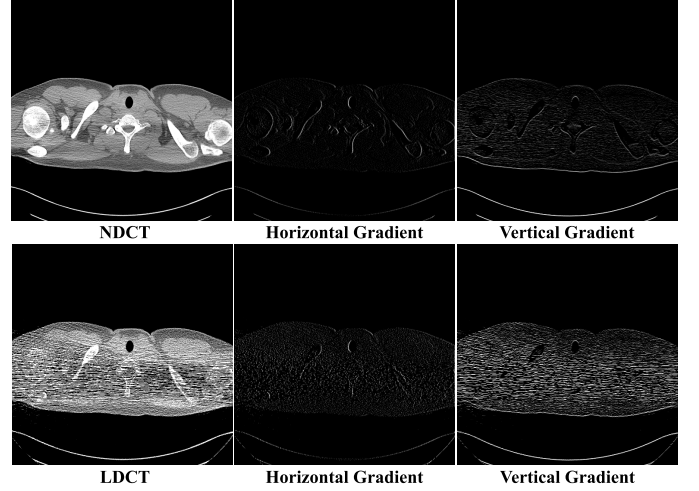

Fig. 3. Gradients of horizontal and vertical directions from a pair of LDCT and NDCT images. The streak artifacts can be easily captured in the gradient domain.

able to learn the global and local differences between denoised and normal-dose images. Next, we present all components, network architecture, and loss functions in detail, followed by their complexity.

A. Denoising Process

The denoising process is to learn a generative model G that maps an LDCT image $I_{\text{LD}} \in \mathbb{R}^{w \times h}$ of size $w \times h$ to its normal-dose CT (NDCT) counterpart $I_{\text{ND}} \in \mathbb{R}^{w \times h}$ by removing the noise in LDCT image. Formally, it can be written as

$$I_{\text{den}} = G(I_{\text{LD}}) \approx I_{\text{ND}} \tag{1}$$

where I_{den} denotes the denoised LDCT image. Typically, LDCT denoising can be seen as a specific image translation problem. Therefore, the GAN-based methods [7]–[9], [37] utilize the GANs to improve the visual quality of denoised LDCT images due to its strong capability of GANs in generating high-quality images. Different from the conventional GANs that take a noise vector to generate an image, our denoising model serves as the generator that only takes the LDCT image as the input. In this study, we used the RED-CNN [6] as the denoising model to demonstrate the effectiveness of the dual-domain U-Net-based discriminators in the adversarial training.

B. Dual-Domain U-Net-Based Discriminator

The GAN-based methods [7]–[9], [37] for LDCT denoising usually maintain the competition of GANs under the structural level, whose discriminator progressively downsamples the input into a scalar value and are trained with Wasserstein GANs [33], [34], as shown in Fig. 2(a). However, the discriminator is prone to forgetting previous samples because the distribution of synthetics samples shifts as the generator constantly changes during training, failing to maintain a powerful data representation to characterize the global and local image differences [16], [38].

To address the problems above, we introduce the U-Net-based discriminators in both image and gradient domains.

1) U-Net-Based Discriminator in the Image Domain: To learn a powerful data representation that can characterize both global and local differences, we design an LDCT denoising framework based on GANs to deal with LDCT denoising. Traditionally, U-Net contains an encoder, a decoder, and several skip connections copying the feature-maps from the encoder to the decoder to preserve high-resolution features,

which has demonstrated its state-of-the-art performance in many semantic segmentation tasks [39], [40] and image translation tasks [16], [31]. In the context of LDCT denoising, we highlight that U-Net and its variants are only used as the denoising model, which have not been explored as the discriminator. We adopt the U-Net to replace the standard classification discriminator in GANs to have a U-Net style discriminator that allows the discriminator to maintain both global and local data representation. Fig. 2(b) details the architecture of U-Net-based discriminator.

Here, we use D^{img} to denote the U-Net-based discriminator in the image domain. The encoder of D^{img}, $D_{\text{enc}}^{\text{img}}$, follows the traditional discriminator that progressively downsamples the input using several convolutional layers, capturing the global structure context. On the other hand, the decoder $D_{\text{dec}}^{\text{img}}$ performs progressive upsampling with skip connections from encoder $D_{\text{enc}}^{\text{img}}$ in a reverse order, further enhancing the ability of discriminator to draw the local details of real and fake samples. Furthermore, the discriminator loss is computed from the outputs of both $D_{\text{enc}}^{\text{img}}$ and $D_{\text{dec}}^{\text{img}}$, while the traditional discriminator used in previous works [7], [8], [37] only classifies the inputs into being real and fake from the encoder. In doing so, the U-Net-based discriminator can provide more informative feedback to the generator, including both local per-pixel and global structural information. In this article, we employ the least-squares GANs [35] rather than conventional GANs [14] for the discriminators to stabilize the training process and improve the visual quality of denoised LDCT. Formally, the discriminator loss for D^{img} from both $D_{\text{enc}}^{\text{img}}$ and $D_{\text{dec}}^{\text{img}}$ can be written as

$$\mathcal{L}_{D^{\text{img}}} = \underbrace{\mathbb{E}_{I_{\text{ND}}}\big[D_{\text{enc}}^{\text{img}}(I_{\text{ND}}) - 1\big]^2 + \mathbb{E}_{I_{\text{LD}}}\big[D_{\text{enc}}^{\text{img}}(I_{\text{den}})\big]^2}_{\text{global adversarial}}$$
$$+ \underbrace{\mathbb{E}_{I_{\text{ND}}}\Big[D_{\text{dec}}^{\text{img}}(I_{\text{ND}}) - 1\Big]^2 + \mathbb{E}_{I_{\text{LD}}}\Big[D_{\text{dec}}^{\text{img}}(I_{\text{den}})\Big]^2}_{\text{local adversarial}} \tag{2}$$

where 1 is the decision boundary of least-squares GANs.

2) U-Net-Based Discriminator in the Gradient Domain: However, the competition in the image domain alone is only able to force the generator toward generating photorealistic denoised LDCT images; it is insufficient to encourage better edge for keeping the pathological changes of original NDCT images and alleviate the streak artifacts caused by photon starvation in LDCT. Previous methods such as [9] measure the different MSE in the gradient domain, which may be insufficient to enhance the edge as MSE tends to blur image. To this end, we propose to perform an additional GANs competition in the gradient domain, where our motivation is presented in Fig. 3. Specifically, the streaks and edge in CT images are highlighted in their horizontal and vertical gradient magnitudes. Therefore, another branch of the gradients estimated by a Sobel operator [41] is performed aside from the image branch, which encourages better edge information and alleviates streak artifacts. Similar to (2), we can define the discriminator loss in the gradient domain $\mathcal{L}_{D^{\text{grd}}}$, where D^{grd} represents the discriminator in the gradient domain.

3) Dual-Domain U-Net-Based Discriminators: Combining the U-Net-based discriminators in the image and gradient domains, two independent GANs competitions are maintained during training. The overall framework of our proposed LDCT denoising model is shown in Fig. 1. In detail, the generator is to denoise an LDCT image, which is then fed into two independent discriminators operating in the image and gradient domains. The discriminator D^{img} in the image domain branch penalizes the generator generating photorealistic denoised LDCT, whereas the discriminator D^{grd} in the gradient domain branch encourages better edge while alleviating streak artifacts caused by photon starvation. In addition, the discriminator in each branch employs a U-Net-based architecture to encourage the generator focusing both global structure and local details, which can also boost the interpretability of the denoising process with the per-pixel confidence map output by $D^{\text{img}}_{\text{dec}}$ and $D^{\text{grd}}_{\text{dec}}$. Finally, the dual-domain U-Net-based discriminator loss can be defined as follows:

$$\mathcal{L}_{D^{\text{dud}}} = \mathcal{L}_{D^{\text{img}}} + \mathcal{L}_{D^{\text{grd}}}. \tag{3}$$

C. CutMix Regularization

The discriminator D^{img} suffers from the decreasing capability in recognizing the local differences between real and fake samples as the training goes, which may unexpectedly harm the denoising performance. Besides, the discriminator is supposed to focus on structure change at the global level and local details at the per-pixel level. To address these issues, we adopt the CutMix augmentation technique to regularize the discriminator inspired by [16] and [42], which can empower the discriminator to learn the intrinsic difference between real and fake samples. Specifically, CutMix technique generates a new training image from two images by cutting patches from the one and pasting them to another. We define this augmentation technique in the context of LDCT denoising as follows:

$$\begin{aligned} \boldsymbol{I}_{\text{mix}} &= \text{mix}(\boldsymbol{I}_{\text{ND}}, \boldsymbol{I}_{\text{den}}, \boldsymbol{M}) \\ &= \boldsymbol{M} \odot \boldsymbol{I}_{\text{ND}} + (1 - \boldsymbol{M}) \odot \boldsymbol{I}_{\text{den}} \end{aligned} \tag{4}$$

where $\boldsymbol{M} \in \{0, 1\}^{w \times h}$ is a binary mask controlling how to mix the NDCT and denoised images and $\odot$ represents the elementwise multiplication.

The mixed samples should be regarded as fake samples globally by the encoder $D^{\text{img}}_{\text{enc}}$ since the CutMix operation has destroyed the global context of NDCT image; otherwise, the CutMix may be introduced to denoise LDCT images during the training of GANs, causing undesirable denoising. Similarly, $D^{\text{img}}_{\text{dec}}$ should be able to recognize the mixed area to provide the generator with accurate per-pixel feedback. Therefore, the regularization loss of CutMix can be formulated as

$$\mathcal{L}_{\text{reg}} = \mathbb{E}_{\boldsymbol{I}_{\text{mix}}} \left[\left[D^{\text{img}}_{\text{enc}}(\boldsymbol{I}_{\text{mix}}) \right]^2 + \left[D^{\text{img}}_{\text{dec}}(\boldsymbol{I}_{\text{mix}}) - \boldsymbol{M} \right]^2 \right] \tag{5}$$

where $\boldsymbol{M}$ used in CutMix also serves as the ground truth for $D^{\text{img}}_{\text{dec}}$.

Furthermore, to penalize the outputs of discriminator to be consistent with the per-pixel predictions after the CutMix operation, we further introduce another consistency loss following [16] to regularize the discriminator with CutMix operation, which can be written as

$$\mathcal{L}_{\text{con}} = \left\| D^{\text{img}}_{\text{dec}}(\boldsymbol{I}_{\text{mix}}) - \text{mix}\left(D^{\text{img}}_{\text{dec}}(\boldsymbol{I}_{\text{ND}}), D^{\text{img}}_{\text{dec}}(\boldsymbol{I}_{\text{den}}), \boldsymbol{M} \right) \right\|^2_F \tag{6}$$

where $\|\cdot\|_F$ represents the Frobenius norm.

During training, the binary mask $\boldsymbol{M}$ is generated following the same pipeline as [42] and [43]. Specifically, we first sample the combination ratio r from beta distribution Beta(1, 1) and then uniformly sample the top-left coordinates of the bounding box of cropping regions from $\boldsymbol{I}_{\text{ND}}$ to $\boldsymbol{I}_{\text{den}}$, with preserving the r ratio. Similar to [42] and [43], we employ a probability p_{mix} to control whether to apply the CutMix regularization technique for each mini-batch samples, which is empirically set to 0.5. Fig. 4 shows the visual results of D^{img} with CutMix regularization technique. It can be observed that the outputs of $D^{\text{img}}_{\text{dec}}$ are the spatial combination of the real and generated patches with respect to the real/fake classification score. Therefore, the results have demonstrated the strong discriminative capability of the U-Net-based discriminator in accurately learning per-pixel differences between real and generated samples, even though they are cut and mixed together to fool the discriminator. Besides learning the per-pixel local details, $D^{\text{img}}_{\text{enc}}$ can accurately predict the proportion of real patches, i.e., the mixed ratio, as it is to focus on the global structures.

D. Network Architecture

As we described above, our proposed method follows the GANs framework to optimize the generator effectively for LDCT denoising, with the U-Net-based discriminator focusing on both global structures and local details, and an extra gradient branch encouraging better boundaries and details. In this section, we describe the network architectures of the generator and U-Net-based discriminator.

1) RED-CNN-Based Generator: In this article, we employ RED-CNN [6] as the generator of our framework for LDCT denoising since this article mainly focuses on the adversarial
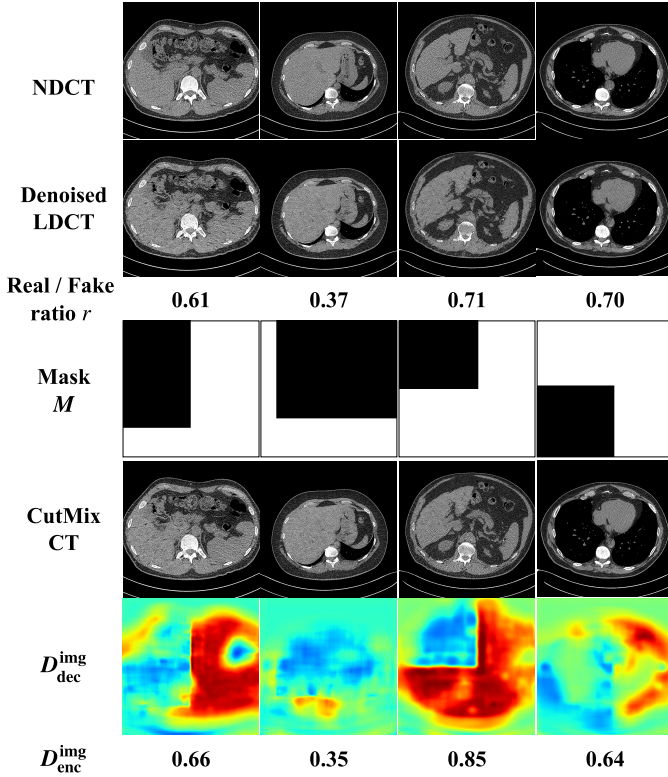
Fig. 4. Illustration of the CutMix regularization for the U-Net-based discriminator, i.e., D^{img}. We randomly sample the ratio r and top-left coordinates of the bounding box to form the mask M controlling where to crop. $D^{\text{img}}_{\text{dec}}$ is able to effectively capture the pixel differences between NDCT and denoised LDCT images, while $D^{\text{img}}_{\text{enc}}$ can predict the mixed ratio. Note that the blue color of $D^{\text{img}}_{\text{dec}}$ indicates the lower confidence score. Therefore, a well-trained discriminator can provide radiologists with a confidence map showing the uncertainty of the denoised results.

loss from dual-domain U-Net-based discriminators. The main difference from [6] is that our framework is optimized in GANs manner, while the vanilla RED-CNN suffers the problem of oversmoothened LDCT images with MSE. Specifically, RED-CNN employs the U-Net architecture but removes the downsampling/upsampling operations to prevent information loss. We stack ten (de)convolutional layers at both encoder and decoder, each of which has 32 filters for the sake of the computation cost, followed by a rectified linear unit (ReLU) activation function. There are in total ten residual skip connections. It is important to note that although RED-CNN is adopted as the generator in our framework, the proposed method can also be adapted to other GAN-based methods such as CPCE [7] and WGAN-VGG [8] with only changing the discriminators.

2) U-Net-Based Discriminator: As detailed in Section III-B, there are two independent discriminators in both image and gradient domains, each of which follows a U-Net architecture. Specifically, D_{enc} has six downsampling ResBlocks [44] with increasing number of filters, i.e., 64, 128, 256, 512, 512, and 512. At the bottom of D_{enc}, a fully connected layer is used to output the global confidence score. Similarly, D_{dec} used the same number of ResBlocks in a reverse order to process the bilinearly upsampled features and the skip residuals of the same resolution, followed by a 1×1 convolutional layer to output the per-pixel confidence

map. Most importantly, a spectral normalization layer [36] and a leaky ReLU activation with a slope of 0.2 for negative input follow each convolutional layer of D except the last one.

We note that the network architectures of the generator and discriminator were proposed in the literature; we did not propose a new network architecture to achieve the performance gain. One of our key contributions is to use the U-net as the discriminator in dual domain to capture both local details and global structures for LDCT denoising.

E. Loss Functions

1) Adversarial Loss: In this section, we employ the sum of these two branches as the adversarial loss, which is defined in the context of least-squares GANs as follows:

$$
\mathcal{L}_{\text{adv}} = \mathbb{E}\Bigg[\underbrace{\left[D^{\text{img}}_{\text{enc}}(\boldsymbol{I}_{\text{den}}) - 1\right]^2 + \left[D^{\text{img}}_{\text{dec}}(\boldsymbol{I}_{\text{den}}) - 1\right]^2}_{\text{image domain}}
$$
$$
+ \underbrace{\left[D^{\text{grd}}_{\text{enc}}(\nabla(\boldsymbol{I}_{\text{den}})) - 1\right]^2 + \left[D^{\text{grd}}_{\text{dec}}(\nabla(\boldsymbol{I}_{\text{den}})) - 1\right]^2}_{\text{gradient domain}} \Bigg]
$$

(7)

where ∇ denotes the Sobel operator to obtain the image gradient.

2) Pixelwise Loss: To encourage the generator output of the denoised LDCT images that match the NDCT images with both pixel level and gradient level, we adopt an pixelwise loss between the NDCT images and denoised LDCT images, which includes a pixel loss and gradient loss for each branch, as shown in Fig. 1. The additional gradient loss can encourage to better preserve edge information at the pixel level. The two losses can be written as

$$
\mathcal{L}_{\text{img}} = \mathbb{E}_{(\boldsymbol{I}_{\text{LD}}, \boldsymbol{I}_{\text{ND}})} \|\boldsymbol{I}_{\text{den}} - \boldsymbol{I}_{\text{ND}}\|^2_F \tag{8}
$$
$$
\mathcal{L}_{\text{grd}} = \mathbb{E}_{(\boldsymbol{I}_{\text{LD}}, \boldsymbol{I}_{\text{ND}})} |\nabla(\boldsymbol{I}_{\text{den}}) - \nabla(\boldsymbol{I}_{\text{ND}})|. \tag{9}
$$

Note that we employ the MSE in pixel level rather than the feature level using the pretrained model [7], [8] for the sake of computation cost, and the absolute mean error in gradient level as the gradients is much sparser than pixels.

3) Final Loss: To encourage the generator to generate photorealistic denoised LDCT images with better edge information and alleviate streak artifacts, the final loss function to optimize the generator G is expressed as

$$
\mathcal{L}_G = \lambda_{\text{adv}}\mathcal{L}_{\text{adv}} + \lambda_{\text{img}}\mathcal{L}_{\text{img}} + \lambda_{\text{grd}}\mathcal{L}_{\text{grd}} \tag{10}
$$

where λ_{adv}, λ_{img}, and λ_{grd} are the weights for $\mathcal{L}_{\text{adv}}$, $\mathcal{L}_{\text{img}}$, and $\mathcal{L}_{\text{grd}}$, respectively. Here, we empirically determine the hyperparameters in a sequential way. First, with only pixelwise loss, our proposed DU-GAN reduces to RED-CNN since discriminators are not included during training. Although fast convergence, only optimizing the MSE loss leads to oversmoothing and blurred results, causing the loss of structural details. We set λ_{img} to be 1. Second, we tune λ_{adv} to control the

importance of adversarial loss to capture the texture details. We start from a small value for λ_{adv} and then gradually increase the importance of the adversarial loss and visualize the denoising results. Finally, we tune λ_{grd} to capture edge information with a large value as the gradients are much sparser than pixels.

The discriminators D^{img} and D^{grd} are optimized by minimizing the following mixed loss:

$$\mathcal{L}_D = \mathcal{L}_{D^{\text{dud}}} + \mathcal{L}_{\text{reg}} + \mathcal{L}_{\text{con}}. \tag{11}$$

Note that we employ the same loss function in (11) to optimize both D^{img} and D^{grd}, but they are independent to each other and D^{grd} has an additional Sobel operator to compute the gradients.

F. Complexity of DU-GAN

Next, we discuss the complexity of the DU-GAN in terms of hyperparameters and computational costs. First, compared to MSE-based methods that directly optimize mean-squared loss, DU-GAN is a GAN-based method that introduces an additional adversarial loss to the training process. Compared to vanilla GAN-based methods with a traditional classification discriminator, DU-GAN proposed to use the U-Net-based discriminator to focus on both local details and global structures. Furthermore, DU-GAN also introduces another gradient branch along the original pixel branch to encourage clear boundaries. Therefore, there is only one extra hyperparameter to control the importance of the gradient branch. Second, the main computational costs of DU-GAN come from the proposed U-Net-based discriminator and gradient branch. However, such computational costs are affordable considering the better denoising quality and performance for our DU-GAN and only happen during the training stage, that is, the inference efficiency is still the same as the traditional ones.

IV. EXPERIMENTS

This section presents the datasets, implementation details, qualitative and quantitative evaluations, uncertainty visualization, and ablation study.

A. Datasets

1) Simulated Dataset: The LDCT dataset used in this study was originally for the 2016 NIHAAPM-Mayo Clinic LDCT Grand Challenge and lately released in [45]. It provides scans from three regions of the body with different simulated low doses, i.e., head with 25% of normal-dose, abdomen with 25%, and chest with 10%. In our experiments, we used the 25% abdomen and 10% chest datasets, named Mayo-25% and Mayo-10%, respectively. We evaluated our method on abdomen scans for comparisons with most previous works and conducted experiments on chest scans since 10% of normal-dose at chest is rather challenging compared to 25% of normal-dose at abdomen. For each dataset, we randomly select 20 patients for training and another 20 patients for testing, no identity overlapping between training and testing. In detail, 300k and 64k image patches were randomly selected from each set. For more information about this dataset, please refer to [45].

2) Real-World Dataset: The real-world dataset from [37] includes 850 CT scans of a deceased piglet obtained by a GE scanner (Discovery CT750 HD). The dataset provides CT scans of the normal dose, 50%, 25%, 10%, and 5% dose with a size of 512×512, 708 of which is served for training while the left for testing. We evaluated our method on 5% LDCTs as it is the most challenging dose, where the dataset is named Piglet-5%. We randomly selected 60k and 12k image patches from training and testing sets, respectively. For more information about this dataset, please refer to [37].

B. Implementation Details

Following [7], [8], and [46], we employed the image patches with a size of 64×64 and a window of $[-300, 300]$ to train all models with emphasis on tissue CT window, which are then directly applied to the whole image for visualization and testing. Note that we excluded those image patches that were mostly air. During training, all images are linearly normalized to [0, 1].

During training, we trained the model with a maximum of 100k iterations and with a mini-batch of size 64 on one NVIDIA V100 GPU. All networks in the proposed framework are initialized with the He initialization [48] and optimized by the Adam optimization method [49] with a fixed learning rate of 10^{-4}. The hyperparameters in the loss functions were empirically set as follows: λ_{adv} was 0.1, λ_{img} was 1, and λ_{grd} was 20. We implemented four deep learning-based methods, including RED-CNN [6], WGAN-VGG [8], CPCE-2D [7], Q-AE [46], and CNCL [47] with the reference of official source code.

C. Qualitative Evaluations

To demonstrate the effectiveness of the proposed method in generating photorealistic denoised results with faithful details, Fig. 5 shows the representative results from three different datasets, and Fig. 6 shows the results of one neck CT slice with strong streak artifacts. The regions of interest (ROIs) marked by the red rectangles are zoomed below.

All methods present visually well-denoised results to some degrees. However, RED-CNN and Q-AE oversmoothed and blurred the LDCT images as they are optimized by the MSE loss that tends to average the results, causing the loss of structural details. Although WGAN-VGG and CPCE-2D have greatly improved the visual fidelity, as expected, due to the use of adversarial loss, minor streak artifacts can still be observed since their traditional classification discriminator only provide the generator with global structure feedback. Besides, they employed the perceptual loss in the high-level feature space to suppress the blurriness resulting from MSE loss. The perceptual loss, however, can only preserve the structures of NDCT images since some local details may be lost after processed by a pretrained model. For example, the low attenuation lesions in Fig. 5 and the bones in Fig. 6 are less clear by WGAN-VGG and CPCE-2D, while they can be easily observed in NDCT as well as the results of our methods. Most importantly, the small structures with their boundaries are consistently
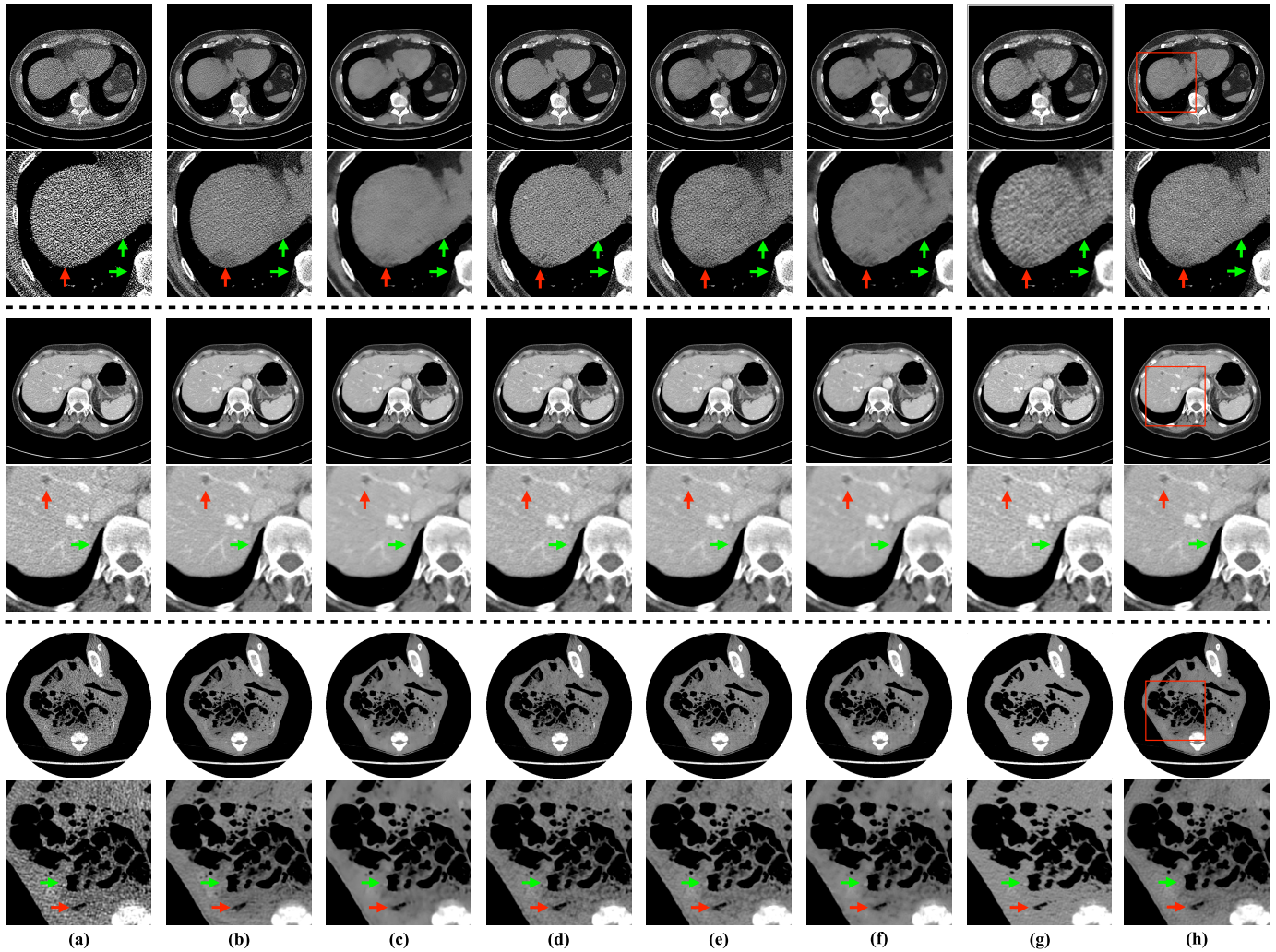
Fig. 5. Transverse CT images from the Mayo-10%, Mayo-25%, and Piglet-5%: (a) LDCT, (b) NDCT, (c) RED-CNN, (d) WGAN-VGG, (e) CPCE-2D, (f) Q-AE, (g) CNCL, and (h) DU-GAN (ours). Zoomed ROI of the red rectangle is shown below the full-size one. The display window is [−160, 240] HU for better visualization. Red arrow indicates low attenuation lesion. Green arrow indicates the white edge artifacts shown in some baseline algorithms while not shown in our method.
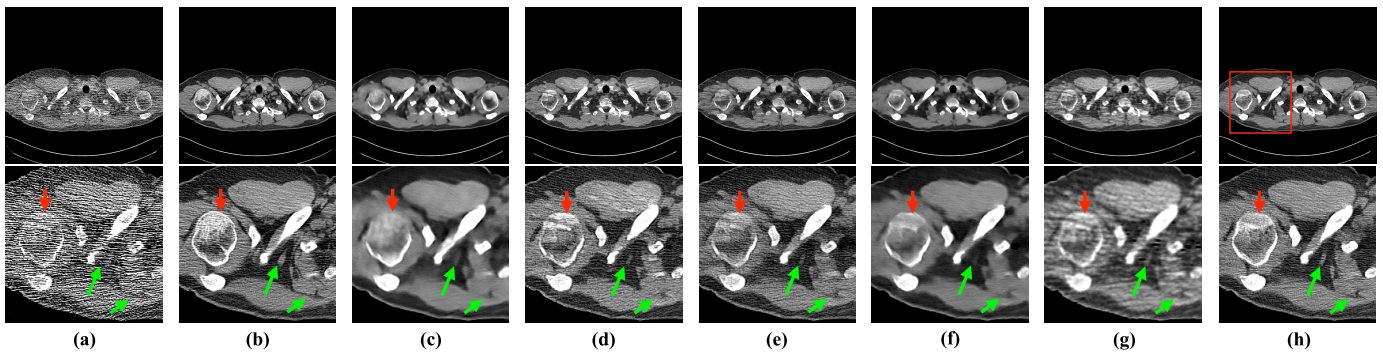

Fig. 6. Transverse neck CT images from the Mayo-10%: (a) LDCT, (b) NDCT, (c) RED-CNN, (d) WGAN-VGG, (e) CPCE-2D, (f) Q-AE, (g) CNCL, and (h) DU-GAN (ours). Zoomed ROI of the red rectangle is shown in the second row. The display window is [−160, 240] HU for better visualization. Red arrow indicates bone area, while green arrow indicates a small structure.

preserved with a clear visual fidelity. This benefits from the well-designed dual-domain U-Net-based discriminators, which can provide feedback of both global structures and local details to the generator, compared to the traditional classification discriminator used in WGAN-VGG and CPCE-2D with only structure information. Besides, the gradient domain branch can also encourage the denoising model to better preserve edge information.

TABLE I

QUANTITATIVE COMPARISONS OF DIFFERENT METHODS ON THE TESTING SETS OF TWO SIMULATED DATASETS AND ONE REAL-WORLD DATASET. THE BEST RESULTS AMONG MSE- AND GAN-BASED METHODS ARE MARKED IN BOLD

|	Method	Mayo-10%			Mayo-25%			Piglet-5%		
		PSNR↑	RMSE↓	SSIM↑	PSNR↑	RMSE↓	SSIM↑	PSNR↑	RMSE↓	SSIM↑
	LDCT	14.6382	0.1913	0.6561	31.5517	0.0283	0.8639	28.7279	0.0395	0.8587
MSE-based	RED-CNN [6]	**23.1388**	**0.0721**	**0.7249**	34.5740	**0.0196**	**0.9236**	26.9691	0.0450	**0.9318**
	Q-AE [46]	21.3149	0.0884	0.7045	**34.6477**	0.0197	0.9215	**29.7081**	**0.0331**	0.9317
GAN-based	WGAN-VGG [8]	20.3922	0.0992	0.7029	33.2910	0.0226	0.9092	**30.3787**	**0.0318**	0.9232
	CPCE-2D [7]	20.1435	0.0899	0.7295	33.0612	0.0232	0.9125	28.5329	0.0379	0.9211
	CNCL [47]	21.8964	0.0852	0.7110	32.4967	0.0243	0.9048	28.5673	0.0383	0.9132
	DU-GAN (ours)	**22.3075**	**0.0802**	**0.7489**	**34.6186**	**0.0196**	**0.9196**	29.8598	0.0325	**0.9345**
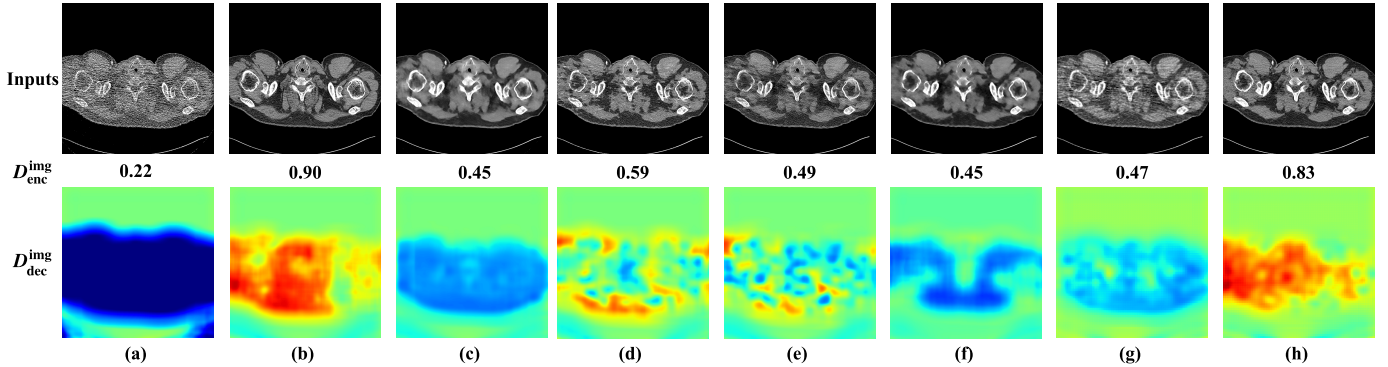
Fig. 7. Uncertainty visualization of applying the trained discriminator to the outputs of different methods: (a) LDCT, (b) NDCT, (c) RED-CNN, (d) WGAN-VGG, (e) CPCE-2D, (f) Q-AE, (g) CNCL, and (h) DU-GAN (ours). The display window is $[-160, 240]$ HU for better visualization. Note that the blue color of D_{dec}^{img} indicates the lower confidence score, while red color indicates higher confidence score.

Beyond encouraging better edge, Fig. 6 also shows its impressive performance in dealing with the LDCT images with strong streak artifacts caused by photon starvation. Compared to the baseline methods that produce strongly blurry and ghosted denoised results, our method can effectively address this problem in the following aspects.

1) Streak artifacts can be easily detected by the gradient domain branch.
2) Once detected, the dual-domain U-Net discriminators can fill the occlusion area by adversarial training to alleviate the impact of streak artifacts.

In summary, all of these results further validate the superiority of our methods.

D. Quantitative Evaluations

For quantitative evaluations, we adopted three widely used metrics, including PSNR, structural similarity (SSIM), and root-mean-square error (RMSE). More specifically, PSNR and RMSE measure the denoising performance at pixel level, while SSIM computes the SSIM within a window. Table I presents the results of different methods. First, RED-CNN and Q-AE are MSE-based denoising methods as they are directly trained with solely MSE loss. Although they achieve better PSNR and RMSE results, the visual results in Figs. 5 and 6 confirm that MSE-based methods produce oversmoothed results compared to the NDCT images, leading to lose of structural information [7], [8], [50]. Note that the oversmoothed denoising results lead to a lower SSIM score. Second, WGAN-VGG, CPCE-2D,

and our DU-GAN are GAN-based methods. CPCE-2D performed better than WGAN-VGG due to the conveying path since WGAN-VGG has to reconstruct the denoised results from the input LDCT images. Obviously, our method performs the best in terms of SSIM score with high visual fidelity, while the PSNR and RMSE are also better than WGAN-VGG and CPCE-2D, indicating the superior denoising performance of our method while better structural fidelity.

Though DU-GAN used the same network architecture of the RED-CNN as the denoising model, their qualitative and quantitative differences directly come from the adversarial training and dual-domain U-Net-based discriminators. The results of DU-GAN preserve more structural details that are important for diagnosis, at the cost of compromising the quantitative metrics such as PSNR and RMSE. We note that PSNR and RMSE are pixelwise metrics, poorly correlating with human perception of image quality [14].

E. Uncertainty Visualization

Fig. 4 shows the proposed discriminator with the U-Net architecture and CutMix regularization can robustly learn the per-pixel differences of local details between NDCT and denoised LDCT images by the decoder and also focus on the global structures by the encoder. With this well-trained discriminator, we can provide the radiologist with a confidence map showing the uncertainty of the denoised results since it is to learn the distribution of real samples, i.e., NDCT images. Therefore, we directly applied the trained discriminator D^{img}

TABLE II

ABLATION STUDY OF COMPONENT ANALYSIS. OUR METHOD IS THE BASELINE METHOD WITH U-NET DISCRIMINATOR IN THE IMAGE DOMAIN, CUTMIX REGULARIZATION, AND THE U-NET DISCRIMINATOR IN THE GRADIENT DOMAIN (DUAL-DOMAIN). THE BEST RESULTS ARE MARKED IN BOLD

Method	PSNR↑	RMSE↓	SSIM↑
Baseline	21.4988	0.0871	0.7365
+ U-Net Based Discriminator	22.1214	0.0816	0.7454
+ CutMix Regularization	21.7894	0.0844	0.7477
Ours (+ Dual-Domain)	**22.3075**	**0.0802**	**0.7489**

TABLE III

ABLATION STUDY OF DIFFERENT DISCRIMINATORS ON TESTING SET OF MAYO-10% DATASET. THE BEST RESULTS ARE MARKED IN BOLD

Method	PSNR↑	RMSE↓	SSIM↑
Patch	21.4988	0.0871	0.7365
Global	22.6810	0.0760	0.7262
Pixel	**23.1102**	**0.0724**	0.7343
U-Net	22.1214	0.0816	**0.7454**

to the LDCT, NDCT, and the denoised LDCT images of different methods.

Fig. 7 shows the uncertainty visualization. Obviously, the discriminator can accurately distinguish the LDCT from NDCT images on both global score and per-pixel confidence. As both RED-CNN and Q-AE oversmoothen the LDCT images, the abdomen area of transverse CT image becomes better than LDCT images on the confidence map, according to the results of $D_{\text{dec}}^{\text{img}}$. This also explains why RED-CNN and Q-AE have the lowest global score of $D_{\text{enc}}^{\text{img}}$, which indicates that the discriminator can robustly detect the blurriness in the CT images. Furthermore, although CPCE-2D can produce more clear denoised results than RED-CNN, the streak artifacts significantly compromised the quality of the denoised results. Similarly, the WGAN-VGG has learned more local details than CPCE-2D, but it still cannot handle the impact of the streak artifacts. On the contrary, the proposed method can produce the most photorealistic denoised results with the highest global score. Compared to the traditional classification discriminator used in CPCE-2D and WGAN-VGG, our DU-GAN can provide the generator with the per-pixel feedback by learning the local detail differences. It can be seen from the per pixel of $D_{\text{dec}}^{\text{img}}$. In other words, we achieve a more smooth per-pixel confidence, indicating that the discriminator cannot distinguish the real and fake samples at the per-pixel level.

F. Ablation Study

In this section, we conducted the ablation study of our method to fully explore the proposed method in terms of the importance of different components, the architectures of discriminator, and the different patch sizes. The ablation study was done on the testing set of Mayo-10% dataset, which includes a total of 6590 slices from 20 patients.

1) Components Analysis: We investigate the impact of the U-Net-based discriminator in the image domain, CutMix regularization, and dual-domain training (i.e., with gradient branch) by gradually applying them to the baseline method. Similar to WGAN-VGG and CPCE-2D, the baseline method only includes the traditional classification discriminator with the same hyperparameters for a fair comparison.

Table II presents the quantitative results for ablation study. First, replacing the traditional classification discriminator with a U-Net-based discriminator can simultaneously provide the generator with both global structure and local per-pixel feedback, which leads to a significant increase in terms of SSIM.

Second, when we further use the CutMix technique to regularize the U-Net-based discriminator, the mixed samples can boost the discriminant capacity of discriminator and make the discriminator more focus on the local details, leading to the increased SSIM score and a slightly decreased PSNR and RMSE. Finally, further adding the U-net-based discriminator in the gradient domain into the method above forming the dual-domain training yields our method. Specifically, the additional gradient-domain training can help our method remove the streak artifacts and encourage more clear edge in the denoised LDCT images. As a result, it can effectively improve all metrics, including the PSNR and RMSE in pixel space and SSIM in visual similarity.

2) Architectures of Discriminator: Since the architectures of discriminator play a critical role in the training of GANs, it is worthwhile studying the advantage of the U-Net-based discriminator over other classical discriminator architectures, such as patch discriminator [31], pixel discriminator [31], and traditional global discriminator. Compared to the traditional classification discriminator that classifies the real and fake samples at the image level, the patch discriminator focuses on the image patches. Due to the patch training of LDCT denoising, this discriminator architecture can be seen as the patch discriminator. The discriminator with seven convolutional layers and one fully connected layer is regarded as the global discriminator. On the other hand, the pixel discriminator [31] contains 7 1×1 convolutional layers to penalize the generator at per-pixel level. For fair comparisons, we trained patch and pixel discriminator with image patches and trained the global discriminator with the whole images with the size of 512×512, respectively. Table III shows that the combination of global and pixel information in a U-Net-based discriminator produces the best SSIM score. This indicates the advantage of U-net-based discriminator for LDCT denoising over other classical discriminator architectures, such as patch discriminator, pixel discriminator, and traditional global discriminator. Instead of pixel discriminator only capturing per-pixel difference and traditional classification discriminator only focusing on global structure, the U-Net-based discriminator has the advantages of both worlds, yielding better quantitative results and denoising quality.

3) Patch Size: Due to the U-Net architecture of the discriminator, it is also important to analyze the influence of the patch size during training. However, it is very difficult to directly train the denoising model from scratch. Therefore, we trained our model with the image size of 64×64, 128×128, 256×256, and 512×512, and we fine-tuned the generator based

TABLE IV
ABLATION STUDY OF PATCH SIZES ON THE TESTING SET OF MAYO-10%
DATASET. THE BEST RESULTS ARE MARKED IN BOLD

Size	PSNR↑	RMSE↓	SSIM↑
64×64	**22.3075**	**0.0802**	**0.7489**
128×128	22.0218	0.0826	0.7479
256×256	21.8478	0.0843	0.7467
512×512	21.7254	0.0855	0.7441

on the model trained on previous smaller size. Table IV shows that a small patch size can achieve better performance because the larger patch sizes may introduce training difficulties with less training samples.

V. CONCLUSION AND DISCUSSION

In this article, we proposed a novel DU-GAN for LDCT denoising. The introduced U-Net-based discriminator can not only provide the per-pixel feedback to the denoising network but also focus on the global structure. We further add an extra U-Net-based discriminator into the gradient domain, which can enhance the edge information and alleviate the streak artifacts caused by photon starvation. We also examined that the CutMix technique can boost the training of discriminator, which can provide the radiologists with a confidence map on the uncertainty of the denoised results. Extensive experiments demonstrated the effectiveness of the proposed method through visual comparison and quantitative comparison.

Although a different architecture or model size could certainly affect results, our DU-GAN has demonstrated its generalization ability on two simulated LDCT datasets of different doses and one real-world dataset, with the same network architecture and hyperparameters. Our ablation study validates each component and their relative importance of all components should be consistent on a new dataset, which indicates that our DU-GAN can be easily adapted to different scenarios.

We acknowledge some limitations in this work. First, we used the qualitative and quantitative comparisons to evaluate the image quality. A human reader study may be needed to further validate its potential in clinical diagnosis, though there are significant differences between the proposed and other baseline methods. Second, the U-net-based discriminator can provide radiologists with a confidence map of the denoised images. How this helps radiologists in clinical routine could be examined with specific tasks such as liver lesion diagnosis, which can be further studied as a future direction. Third, DU-GAN could introduce slightly more computational cost during training since it employs the U-Net as the discriminator and adopts a dual-domain training strategy. However, we emphasize that the extra computational cost is relatively affordable as DU-GAN trains the whole framework based on the 64×64 image patches instead of the original image size 512×512. Therefore, we believe that the computational cost can be significantly reduced. We note that the extra computational cost only happens for the training stage, that is, the inference efficiency is still the same as the traditional one as the dual-domain discriminators are not involved during

the testing stage. Finally, we only validated DU-GAN with two dose levels in this article and it is worth further validating DU-GAN with lower radiation dose.

In conclusion, the proposed DU-GAN achieves better denoising performance than other GAN-based models and has great potential for clinical use with uncertainty visualization.

REFERENCES

[1] N. B. Shah and S. L. Platt, "ALARA: Is there a cause for alarm? Reducing radiation risks from computed tomography scanning in children," *Current Opinion Pediatrics*, vol. 20, no. 3, pp. 243–247, 2008.

[2] Y. Lei, Y. Tian, H. Shan, J. Zhang, G. Wang, and M. K. Kalra, "Shape and margin-aware lung nodule classification in low-dose CT images via soft activation mapping," *Med. Image Anal.*, vol. 60, Feb. 2020, Art. no. 101628.

[3] F. Attivissimo, G. Cavone, A. M. L. Lanzolla, and M. Spadavecchia, "A technique to improve the image quality in computer tomography," *IEEE Trans. Instrum. Meas.*, vol. 59, no. 5, pp. 1251–1257, May 2010.

[4] G. Wang, J. C. Ye, and B. De Man, "Deep learning for tomographic image reconstruction," *Nature Mach. Intell.*, vol. 2, no. 12, pp. 737–748, Dec. 2020.

[5] H. Chen *et al.*, "Low-dose CT via convolutional neural network," *Biomed. Opt. Exp.*, vol. 8, no. 2, pp. 679–694, 2017.

[6] H. Chen, Y. Zhang, M. K. Kalra, F. Lin, Y. Chen, P. Liao, J. Zhou, and G. Wang, "Low-dose CT with a residual encoder-decoder convolutional neural network (red-CNN)," *IEEE Trans. Med. Imag.*, vol. 36, no. 12, pp. 2524–2535, Jun. 2017.

[7] H. Shan *et al.*, "3-D convolutional encoder-decoder network for low-dose CT via transfer learning from a 2-D trained network," *IEEE Trans. Med. Imag.*, vol. 37, no. 6, pp. 1522–1534, Jun. 2018.

[8] Q. Yang *et al.*, "Low-dose CT image denoising using a generative adversarial network with Wasserstein distance and perceptual loss," *IEEE Trans. Med. Imag.*, vol. 37, no. 6, pp. 1348–1357, Jun. 2018.

[9] H. Shan *et al.*, "Competitive performance of a modularized deep neural network compared to commercial algorithms for low-dose CT image reconstruction," *Nature Mach. Intell.*, vol. 1, no. 6, pp. 269–276, 2019.

[10] J. M. Wolterink, T. Leiner, M. A. Viergever, and I. Išgum, "Generative adversarial networks for noise reduction in low-dose CT," *IEEE Trans. Med. Imag.*, vol. 36, no. 12, pp. 2536–2545, Dec. 2017.

[11] G. Wang, J. C. Ye, K. Mueller, and J. A. Fessler, "Image reconstruction is a new frontier of machine learning," *IEEE Trans. Med. Imag.*, vol. 37, no. 6, pp. 1289–1296, Jun. 2018.

[12] D. Wu, K. Kim, G. El Fakhri, and Q. Li, "A cascaded convolutional neural network for X-ray low-dose CT image denoising," 2017, *arXiv:1705.04267*.

[13] H. Zhao, O. Gallo, I. Frosio, and J. Kautz, "Loss functions for image restoration with neural networks," *IEEE Trans. Comput. Imag.*, vol. 3, no. 1, pp. 47–57, Mar. 2017.

[14] I. Goodfellow *et al.*, "Generative adversarial nets," in *Proc. Adv. Neural Inf. Process. Syst.*, 2014, pp. 2672–2680.

[15] Z. Wang, A. C. Bovik, H. R. Sheikh, and E. P. Simoncelli, "Image quality assessment: From error visibility to structural similarity," *IEEE Trans. Image Process.*, vol. 13, no. 4, pp. 600–612, Apr. 2004.

[16] E. Schonfeld, B. Schiele, and A. Khoreva, "A U-Net based discriminator for generative adversarial networks," in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit.*, Jun. 2020, pp. 8207–8216.

[17] C. H. Lin, C.-C. Chang, Y.-S. Chen, D.-C. Juan, W. Wei, and H.-T. Chen, "COCO-GAN: Generation by parts via conditional coordinating," in *Proc. IEEE/CVF Int. Conf. Comput. Vis.*, Oct. 2019, pp. 4512–4521.

[18] H. Zhang, I. Goodfellow, D. Metaxas, and A. Odena, "Self-attention generative adversarial networks," in *Proc. Int. Conf. Mach. Learn.*, 2019, pp. 7354–7363.

[19] O. Ronneberger, P. Fischer, and T. Brox, "U-Net: Convolutional networks for biomedical image segmentation," in *Proc. Int. Conf. Med. Image Comput. Comput. Assist. Intervent.* Munich, Germany: Springer, 2015, pp. 234–241.

[20] J. Wang, H. Lu, T. Li, and Z. Liang, "Sinogram noise reduction for low-dose CT by statistics-based nonlinear filters," *Proc. SPIE*, vol. 5747, p. 2059, Apr. 2005.

[21] J. Wang, T. Li, H. Lu, and Z. Liang, "Penalized weighted least-squares approach to sinogram noise reduction and image reconstruction for low-dose X-ray computed tomography," *IEEE Trans. Med. Imag.*, vol. 25, no. 10, pp. 1272–1283, Oct. 2006.

[22] A. Manduca *et al.*, "Projection space denoising with bilateral filtering and CT noise modeling for dose reduction in CT," *Med. Phys.*, vol. 36, no. 11, pp. 4911–4919, 2009.

[23] S. Ramani and J. A. Fessler, "A splitting-based iterative algorithm for accelerated statistical X-ray CT reconstruction," *IEEE Trans. Med. Imag.*, vol. 31, no. 3, pp. 677–688, Mar. 2012.

[24] W. Wu, J. Shi, H. Yu, W. Wu, and V. Vardhanabhuti, "Tensor gradient L_0-norm minimization-based low-dose CT and its application to COVID-19," *IEEE Trans. Instrum. Meas.*, vol. 70, pp. 1–12, 2021.

[25] X. Zheng, S. Ravishankar, Y. Long, and J. A. Fessler, "PWLS-ULTRA: An efficient clustering and learning-based approach for low-dose 3D CT image reconstruction," *IEEE Trans. Med. Imag.*, vol. 37, no. 6, pp. 1498–1510, Jun. 2018.

[26] Q. Xu, H. Yu, X. Mou, L. Zhang, J. Hsieh, and G. Wang, "Low-dose X-ray CT reconstruction via dictionary learning," *IEEE Trans. Med. Imag.*, vol. 31, no. 9, pp. 1682–1697, Sep. 2012.

[27] J. Ma *et al.*, "Low-dose computed tomography image restoration using previous normal-dose scan," *Med. Phys.*, vol. 38, no. 10, pp. 5713–5731, 2011.

[28] P. F. Feruglio, C. Vinegoni, J. Gros, A. Sbarbati, and R. Weissleder, "Block matching 3D random noise filtering for absorption optical projection tomography," *Phys. Med. Biol.*, vol. 55, no. 18, p. 5401, 2010.

[29] M. Li, W. Hsu, X. Xie, J. Cong, and W. Gao, "SACNN: Self-attention convolutional neural network for low-dose CT denoising with self-supervised perceptual loss network," *IEEE Trans. Med. Imag.*, vol. 39, no. 7, pp. 2289–2301, Jul. 2020.

[30] T. Karras, T. Aila, S. Laine, and J. Lehtinen, "Progressive growing of GANs for improved quality, stability, and variation," in *Proc. Int. Conf. Learn. Represent.*, 2018, pp. 1–26.

[31] P. Isola, J.-Y. Zhu, T. Zhou, and A. A. Efros, "Image-to-image translation with conditional adversarial networks," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jul. 2017, pp. 1125–1134.

[32] A. Guo, L. Fang, M. Qi, and S. Li, "Unsupervised denoising of optical coherence tomography images with nonlocal-generative adversarial network," *IEEE Trans. Instrum. Meas.*, vol. 70, pp. 1–12, 2021.

[33] M. Arjovsky, S. Chintala, and L. Bottou, "Wasserstein generative adversarial networks," in *Proc. Int. Conf. Mach. Learn.*, 2017, pp. 214–223.

[34] I. Gulrajani, F. Ahmed, M. Arjovsky, V. Dumoulin, and A. C. Courville, "Improved training of Wasserstein GANs," in *Proc. Adv. Neural Inf. Process. Syst.*, 2017, pp. 5769–5779.

[35] X. Mao, Q. Li, H. Xie, R. Y. K. Lau, Z. Wang, and S. P. Smolley, "Least squares generative adversarial networks," in *Proc. IEEE Int. Conf. Comput. Vis. (ICCV)*, Oct. 2017, pp. 2794–2802.

[36] T. Miyato, T. Kataoka, M. Koyama, and Y. Yoshida, "Spectral normalization for generative adversarial networks," in *Proc. Int. Conf. Learn. Represent.*, 2018, pp. 1–26.

[37] X. Yi and P. Babyn, "Sharpness-aware low-dose CT denoising using conditional generative adversarial network," *J. Digit. Imag.*, vol. 31, no. 5, pp. 655–669, Oct. 2018.

[38] T. Chen, X. Zhai, M. Ritter, M. Lucic, and N. Houlsby, "Self-supervised GANs via auxiliary rotation loss," in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jun. 2019, pp. 12154–12163.

[39] Z. Zhou, M. M. R. Siddiquee, N. Tajbakhsh, and J. Liang, "UNet++: A nested U-Net architecture for medical image segmentation," in *Deep Learning in Medical Image Analysis and Multimodal Learning for Clinical Decision Support*. Granada, Spain: Springer, 2018, pp. 3–11.

[40] Y. Han and J. C. Ye, "Framing U-Net via deep convolutional framelets: Application to sparse-view CT," *IEEE Trans. Med. Imag.*, vol. 37, no. 6, pp. 1418–1429, Jun. 2018.

[41] N. Kanopoulos, N. Vasanthavada, and R. L. Baker, "Design of an image edge detection filter using the Sobel operator," *IEEE J. Solid-State Circuits*, vol. SSC-23, no. 2, pp. 358–367, Apr. 1988.

[42] S. Yun, D. Han, S. J. Oh, S. Chun, J. Choe, and Y. Yoo, "CutMix: Regularization strategy to train strong classifiers with localizable features," in *Proc. IEEE Int. Conf. Comp. Vis.*, Oct. 2019, pp. 6023–6032.

[43] H. Zhang, M. Cisse, Y. N. Dauphin, and D. Lopez-Paz, "mixup: Beyond empirical risk minimization," in *Proc. Int. Conf. Learn. Represent.*, 2018, pp. 1–13.

[44] K. He, X. Zhang, S. Ren, and J. Sun, "Deep residual learning for image recognition," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, Jun. 2016, pp. 770–778.

[45] T. R. Moen *et al.*, "Low-dose CT image and projection dataset," *Med. Phys.*, vol. 48, no. 2, pp. 902–911, Feb. 2021.

[46] F. Fan *et al.*, "Quadratic autoencoder (Q-AE) for low-dose CT denoising," *IEEE Trans. Med. Imag.*, vol. 39, no. 6, pp. 2035–2050, Jun. 2020.

[47] M. Geng *et al.*, "Content-noise complementary learning for medical image denoising," *IEEE Trans. Med. Imag.*, early access, Sep. 16, 2021, doi: 10.1109/TMI.2021.3113365.

[48] K. He, X. Zhang, S. Ren, and J. Sun, "Delving deep into rectifiers: Surpassing human-level performance on ImageNet classification," in *Proc. IEEE Int. Conf. Comput. Vis.*, Dec. 2015, pp. 1026–1034.

[49] D. P. Kingma and J. Ba, "Adam: A method for stochastic optimization," 2014, *arXiv:1412.6980*.

[50] J. Johnson, A. Alahi, and L. Fei-Fei, "Perceptual losses for real-time style transfer and super-resolution," in *Proc. Eur. Conf. Comp. Vis.* Amsterdam, The Netherlands: Springer, 2016, pp. 694–711.

Zhizhong Huang (Graduate Student Member, IEEE) received the B.S. degree from Sichuan University, Chengdu, China, in 2019. He is currently pursuing the Ph.D. degree with the School of Computer Science, Fudan University, Shanghai, China.

His research interests include machine/deep learning, computer vision, face aging, and image translation.

Junping Zhang (Member, IEEE) received the B.S. degree in automation from Xiangtan University, Xiangtan, China, in 1992, the M.S. degree in control theory and control engineering from Hunan University, Changsha, China, in 2000, and the Ph.D. degree in intelligent systems and pattern recognition from the Institution of Automation, Chinese Academy of Sciences, Beijing, China, in 2003.

He has been a Professor with the School of Computer Science, Fudan University, Shanghai, China, since 2006. His research interests include machine learning, image processing, biometric authentication, and intelligent transportation systems.

Dr. Zhang was an Associate Editor of *IEEE Transactions on Intelligent Transportation Systems* from 2010 to 2018. He has been an Associate Editor of IEEE INTELLIGENT SYSTEMS since 2009.

Yi Zhang (Senior Member, IEEE) received the bachelor's, master's, and Ph.D. degrees from the College of Computer Science, Sichuan University, Chengdu, China, in 2005, 2008, and 2012, respectively.

From 2014 to 2015, he was with the Department of Biomedical Engineering, Rensselaer Polytechnic Institute, Troy, NY, USA, as a Post-Doctoral Researcher. He is currently a Full Professor with the College of Computer Science, Sichuan University, where he is also the Director of the Deep Imaging Group (DIG). His research interests include medical imaging, compressive sensing, and deep learning.

Hongming Shan (Member, IEEE) received the B.S. degree in information and computing science from the Shandong University of Technology, Zibo, China, in 2011, and the Ph.D. degree in machine learning from Fudan University, Shanghai, China, in 2017.

From 2017 to 2020, he was a Post-Doctoral Research Associate with the Rensselaer Polytechnic Institute, Troy, NY, USA. He is currently an Associate Professor with the Institute of Science and Technology for Brain-Inspired Intelligence, Fudan University, and a "Qiusuo" Research Leader with the Shanghai Center for Brain Science and Brain-Inspired Technology, Shanghai. His research interests include machine/deep learning, computer vision, biomedical imaging, and brain-inspired intelligence.