# Speckle Noise Reduction for OCT Images Based on Image Style Transfer and Conditional GAN

Yi Zhou, Kai Yu, Meng Wang, Yuhui Ma, Yuanyuan Peng, Zhongyue Chen, Weifang Zhu, Fei Shi, and Xinjian Chen, *Senior Member, IEEE*

***Abstract*—Raw optical coherence tomography (OCT) images typically are of low quality because speckle noise blurs retinal structures, severely compromising visual quality and degrading performances of subsequent image analysis tasks. In our previous study (Ma *et al.*, 2018) we have developed a Conditional Generative Adversarial Network (cGAN) for speckle noise removal in OCT images collected by several commercial OCT scanners, which we collectively refer to as scanner T. In this paper, we improve the cGAN model and apply it to our in-house OCT scanner (scanner B) for speckle noise suppression. The proposed model consists of two steps: 1) We train a Cycle-Consistent GAN (CycleGAN) to learn style transfer between two OCT image datasets collected by different scanners. The purpose of the CycleGAN is to leverage the ground truth dataset created in our previous study. 2) We train a mini-cGAN model based on the PatchGAN mechanism with the ground truth dataset to suppress speckle noise in OCT images. After training, we first apply the CycleGAN model to convert raw images collected by scanner B to match the style of the images from scanner T, and subsequently use the mini-cGAN model to suppress speckle noise in the style transferred images. We evaluate the proposed method on a dataset collected by scanner B. Experimental results show that the improved model outperforms our previous method and other state-of-the-art models in speckle noise removal, retinal structure preservation and contrast enhancement.**

***Index Terms*—Optical coherence tomography, Speckle noise reduction, Generative adversarial networks.**

## I. Introduction

OPTICAL coherence tomography (OCT) is a method of acquiring and processing optical signals. It can scan optical scattering media such as biological tissues, and the obtained three-dimensional image resolution can reach the micron level. The first generation of OCT was time-domain OCT (TD-OCT) [2], which used temporal information related to position of a moving reference mirror to encode the position of each cut. In 1993, for the first time, OCT was used to scan retinal structures in vivo [3]. In 2002, spectral-domain OCT (SD-OCT) was developed [4], which simultaneously acquired all information from a single one-dimensional scan (Ascan) of tissue by evaluating frequency spectrum of the interference between reflected light and a stationary reference mirror. The use of fixed mirrors has allowed the scanning processing to be much faster, thus greatly increasing the data availability. A two-dimensional image (Bscan) can be achieved by laterally combining a series of many Ascans.

The quality of an OCT imaging system is usually categorized by its sensitivity, which is defined as the weakest signal that can be detected in the noise range. Signal can be amplified by increasing incident power or exposure time of the detector. However, the power is limited by imaging safety constraints, and an increase in exposure time leads to a significant increase in data acquisition time, resulting in strong autocorrelation of the signal and blurred image reconstructions. Therefore, most OCT systems acquire data with the shortest allowed exposure time and the highest acceptable incident power, and improve image quality through data post-processing.

Similar to many other imaging modalities, OCT images suffer from speckle noise, which is inherently granular in OCT. Speckle noise can degrade quality of OCT images, impair visual interpretation and diagnosis, and make automated image analysis such as retinal layer segmentation for various layer thickness measurements challenging. Therefore, speckle noise reduction is critical for both visual interpretation and automatic analysis of retinal OCT. However, speckle noise is inherently a randomly modulated signal caused by temporal and spatial coherence of light waves, the basis of interferometry, on which the OCT technique is built. As such, speckles have dual roles of a noise source and an information carrier for microstructure of tissue. A complete elimination of speckles is undesirable.

Speckle suppression techniques are usually divided into two categories: hardware compensation and algorithmic suppression. Typically, hardware speckle noise compensation methods are based on the acquisition of multiple frames with uncorrelated noise that can be averaged out to improve the signal-to-noise ratio (SNR). Duan *et al.* [5] proposed interleaved OCT (iOCT),

Manuscript received December 30, 2020; revised March 17, 2021; accepted April 17, 2021. Date of publication April 21, 2021; date of current version January 5, 2022. This study was supported in part by the National Key R&D Program of China under Grant 2018YFA0701700, and in part by the National Nature Science Foundation of China under Grant 61622114. *(Corresponding author: Xinjian Chen.)*

Yi Zhou, Kai Yu, Meng Wang, Yuhui Ma, Yuanyuan Peng, Zhongyue Chen, Weifang Zhu, and Fei Shi are with the School of Electronics and Information Engineering, Soochow University, Jiangsu 215006, China (e-mail: zhouyi.zura@gmail.com; 578069383@qq.com; wangmeng9218@126.com; mayuhui@nimte.ac.cn; yypeng@stu.suda.edu.cn; chenzy@suda.edu.cn; wfzhu@suda.edu.cn; shifei@suda.edu.cn).

Xinjian Chen is with the School of Electronics and Information Engineering and the State Key Laboratory of Radiation Medicine and Protection, Soochow University, Jiangsu 215006, China (e-mail: xjchen@suda.edu.cn).

Digital Object Identifier 10.1109/JBHI.2021.3074852

for spatial and angular compounding to restrain speckle noise. Based on strain compounding, Kennedy *et al.* [6] presented a speckle reduction technique for OCT. Cheng *et al.* [7] proposed a dual-beam angular compounding method based on angular compounding to reduce speckle noise and improve SNR. However, the main drawbacks inherent in hardware speckle noise compensation methods include longer data acquisition time, more complex data acquisition processes, and increased complexity in the optomechanical design of OCT imaging systems.

Algorithmic suppression aims to reduce speckle noise through post-processing of OCT images while maintaining image resolution, contrast, and edge fidelity. Many traditional algorithmic methods have been proposed for automatic OCT denoise and enhancement, and can be categorized as: 1). Digital filter based techniques, such as Lee filter [8], Kuan filter [9], enhanced Lee filter [10], adaptive Wiener filter [11], weighted median filter [12], anisotropic diffusion filtering [13], [14], non-local methods [15], [16], block-matching and 3D filtering methods (BM3D) [17], 2). Sparse transform-based methods based on different transformations including wavelets [18], [19], curvelet transform [20], dictionary learning [21], [22], and 3). Statistical model-based methods [23]–[25] and methods based on low-rank decomposition [26], [27].

In recent years, deep learning has injected new vitality into image denoising and enhancement. Remez *et al.* [28] proposed a fully convolutional neural network (CNN) architecture that allows us to take advantage of the progressive nature of the denoising process. Yang *et al.* [29] proposed to expand the computation pipeline of the BM3D algorithm into a CNN structure. Zhang *et al.* [30] used residual learning strategies and batch normalization to improve the feedforward denoising CNN (DnCNN) to target Gaussian denoising (blind Gaussian denoising) at unknown noise levels. Dong *et al.* [31] proposed a denoising prior driven network for image restoration. However, all the above work utilized an additive noise model. The noise of our OCT image is far from additive. Therefore, these models are not suitable for speckle noise in OCT images.

In our previous study, Ma *et al.* [1] proposed a cGAN model [32] to achieve speckle noise removal, which is the first application of image-to-image cGAN network to OCT speckle noise reduction. The cGAN model does not assume any specific distribution model and learns it from data. Through the competition between a generator and a discriminator, the cGAN model can achieve speckle noise reduction and contrast enhancement simultaneously. The model was applied to OCT images collected by commercial scanners (scanner T) and achieved excellent speckle noise suppression results. The cGAN model was trained with the paired ground truth datasets created by scanner $T$ using the method described in Section II-E. Different from GAN models, cGAN model has conditional inputs to guide data generation and belongs to supervised learning, making it a good model candidate for applications where paired ground truth datasets are available for training.

Recently, we have developed an in-house OCT scanner (scanner $B$) and its raw images are also subjected to the speckle noise blurring. In this paper, we improve our previous method and develop a mini-cGAN model based on the PatchGAN mechanism [32] to remove speckle noise for scanner $B$. It is noticed that the image style (histogram) of scanner B is different from that of scanner $T$. To leverage the existing ground truth dataset for training, we first train a CycleGAN [33] to learn the bi-directional style transfer between scanners T and B. After training, we apply the CycleGAN model to images from scanner B to match the style of images from scanner $T$. Finally, we train the mini-cGAN model with the existing ground truth dataset for denoising of images obtained by scanner $B$. The main contributions of the paper are as follows:

- A novel denoising method is proposed to suppress speckle noise in retinal OCT images and achieves superb performances in retinal structure preservation and contrast enhancement.
- A CycleGAN model is implemented for image histogram and style matching to leverage ground truth datasets collected by different scanners for training.
- A miniaturised patch cGAN is developed that significantly improves speckle noise removal and subsequent retinal layer segmentation in OCT images.

## II. Method

The diagram of the proposed work is illustrated in Fig. 1, which consists of two independent networks: a CycleGAN for image style transfer and a mini-cGAN for denoising. In the training phase, we use unpaired datasets, $B$ and $T$, collected by the scanners B and T, respectively, for training. For CycleGAN, we train two generators, $G_B$ and $G_T$, to perform bi-directional image style transfer between $B$ and $T$. CycleGAN also has two discriminators $D_B$ and $D_T$, to discriminate real and generated image pairs, $T$ vs. $G_B(B)$ and $B$ vs. $G_T(T)$, respectively. For mini-cGAN, $T_{GT}$ are ground truth images obtained from $T$ by the method described in Section II-E. The mini-cGAN model takes $T_{GT}$ and $T$ as inputs to produce $G(T, Z)$ as denoised images, where $Z$ denotes noise. The discriminator $D$ tries to identify real and fake ground truth image pairs, $T_{GT}$ and $G(T, Z)$. In the testing phase, we first use the trained CycleGAN model to covert $B$ as $G_B(B)$ that match the style of images $T$. We then input $G_B(B)$ to the mini-cGAN's generator $G$ to produce denoised images $G(G_B(B), Z)$. The testing flowchart is shown in Fig. 2.

### A. Overview of CycleGAN and cGAN

GAN models are trained to learn distribution from real images and generate realistic images by randomly sampling from the learned distribution. However, GANs do not have control on the contents of the generated data. In contrast to original GAN models, conditional GAN (cGAN) models add a condition $K$ to the generator to control the generated data. The condition $K$ can be any format including category labels or images, and data used to train cGAN models are always paired with its condition. CycleGAN models aim to form a bi-directional universal mapping between data domain $T$ and data domain $B$, and the goal of learning is style transfer between domains rather than a specific one-to-one data mapping, i.e., the cGAN's dependence on data pairing is eliminated.

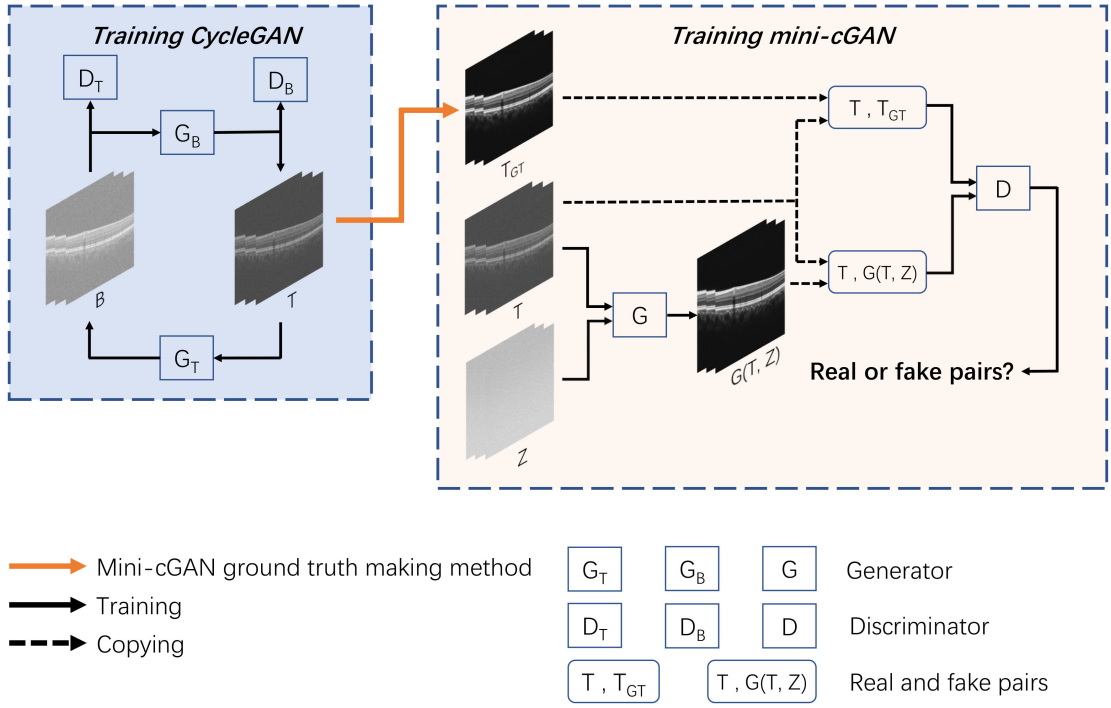


Fig. 1. Overview of the proposed method in the training phase. We use unpaired datasets, $B$ and $T$, collected by the scanners B and T, respectively, for training. For CycleGAN, we train two generators, $G_B$ and $G_T$, to perform bi-directional image style transfer between $B$ and $T$. CycleGAN also has two discriminators $D_B$ and $D_T$, to discriminate real and generated image pairs, $T$ vs. $G_B(B)$ and $B$ vs. $G_T(T)$, respectively. For mini-cGAN, $T_{GT}$ are ground truth images obtained by the method described in Section II-E. The mini-cGAN model takes $T_{GT}$ and $T$ as inputs to produce $G(T, Z)$ as denoised images, where $Z$ denotes noise. The discriminator $D$ tries to identify real and fake ground truth image pairs, $T_{GT}$ and $G(T, Z)$.

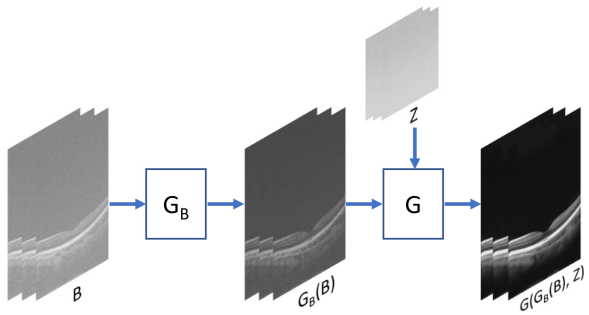

Fig. 2. Flowchart of the proposed method in the testing phase. We first input images, $B$, from scanner B, into generator $G_B$ to obtain the style transferred images $G_B(B)$, and then input $G_B(B)$ into the mini-cGAN's generator, $G$, to obtain denoised images $G(G_B(B), Z)$, where $Z$ denotes noise.

## B. Objective Functions

*1) CycleGAN Loss:* The GAN loss is applied to both generators, $G_T$: $T \to B$ and $G_B$: $B \to T$. For the generator $G_T$ and its discriminator $D_B$, the loss function $\mathcal{L}_{GAN}$ is expressed as:

$$\mathcal{L}_{GAN}(G_T, D_B, T, B) = \mathbb{E}_{b\sim P_{data}(b)}[\log D_B(b)] + \mathbb{E}_{t\sim P_{data}(t)}[\log(1 - D_B(G_T(t)))], \quad (1)$$

where $b \in B$ and $t \in T$ are individual image samples from $B$ and $T$, respectively. $G_T$ tries to generate images $G_T(t)$ that look similar to images from $B$, while $D_B$ aims to distinguish between translated samples $G_T(t)$ and real sample $b$.

Individual GAN loss usually cannot guarantee that the learned function will map a particular input $t$ to the desired output $b$. To further reduce the space of possible mapping functions, Zhu *et al.* [33] argued that the learned mapping function should be cycle-consistent: for each image $t \in T$, the image translation cycle should be able to bring $t$ back to the original domain, i.e., $t \to G_T(t) \to G_B(G_T(t)) \approx t$, and call it forward cyclic consistency. Correspondingly, $G_T$ and $G_B$ should also satisfy backward cyclic consistency, $b \to G_B(b) \to G_T(G_B(b)) \approx b$. A cycle consistency loss of $\mathcal{L}_{cyc}$ is proposed as,

$$\mathcal{L}_{cyc}(G_T, G_B) = \mathbb{E}_{t\sim P_{data}(t)}[\|G_B(G_T(t)) - t\|_1] + \mathbb{E}_{b\sim P_{data}(b)}[\|G_T(G_B(b)) - b\|_1]. \quad (2)$$

The full objective function of CycleGAN can be shown as follows:

$$\mathcal{L}_{CycleGAN}(G_T, G_B, D_T, D_B) = \mathcal{L}_{GAN}(G_T, D_B, T, B) + \mathcal{L}_{GAN}(G_B, D_T, B, T) + \lambda L_{cyc}(G_T, G_B), \quad (3)$$

where $\lambda$ controls the relative importance of the two objectives, $\mathcal{L}_{GAN}$ and $\mathcal{L}_{cyc}$. The overall goal is to solve,

$$G_T^*, G_B^* = \arg \min_{G_T, G_B} \max_{D_T, D_B} \mathcal{L}(G_T, G_B, D_T, D_B), \quad (4)$$

where $G_T^*$ and $G_B^*$ represent the optimized generator of $T \to B$ and $B \to T$, respectively.

*2) mini-CGAN Loss:* The objective function of the mini-cGAN can be shown as follows [32]:

$$\mathcal{L}_{cGAN}(G,D) = \mathbb{E}_{t,t_{gt}\sim P_{data}(t,t_{gt})}[\log D(t,t_{gt})] + \mathbb{E}_{t\sim P_{data}(t),z\sim P_z(z)}[\log(1-D(t,G(t,z)))], \quad (5)$$

where $t \in T$ is the input image, $z$ is the random noise, $t_{gt} \in T_{GT}$ is the corresponding ground truth for $t$, $D(t,t_{gt})$ is the output of the discriminator, and $G(t,z)$ is the output image of the generator. The first term is the expected value calculated for all training images and the corresponding ground truth, and the second term is the expected value calculated for all training images and random noise.

In training, $G$ tries to minimize the objective function while $D$ tries to maximize it, leading to the following optimization problem,

$$G^* = \arg\min_G \max_D L(G,D). \quad (6)$$

where $G^*$ represents the optimized generator.

It has proven that it is useful to combine the objective function of cGAN with the traditional loss (such as $L_1$ or $L_2$ loss) [32]. The traditional loss functions incorporate spatial information and can improve quality of the generated images. The $L_2$ loss punishes errors caused by uniform distribution in space leading to blurred images, while $L_1$ loss punishes errors caused by sparse distribution in space resulting in sharper images. The $L_1$ loss can be formulated as,

$$\mathcal{L}_{L_1}(G) = \mathbb{E}_{t,t_{gt},z}[\|t_{gt} - G(t,z)\|_1], \quad (7)$$

We integrate the $L_1$ loss into the final objective optimization function of cGAN as,

$$G^* = \arg\min_G \max_D L(G,D) + \alpha\mathcal{L}_{L_1}(G). \quad (8)$$

where $\alpha$ is a controlling parameter.

## C. Network Structure

*1) Implementation of CycleGAN:* In this paper, the generators of CycleGAN are two U-Net networks [34]. U-Net has the encoder-decoder architecture with skip connections between the symmetric layers in the encoder and the decoder. The discriminators of CycleGAN are two PatchGAN models, which are used to identify whether or not an extracted image patch is from a particular domain. The building blocks for both U-Net and PatchGAN are typical layers used in deep learning models including convolutional, InstanceNorm [35] and LeakyReLU [36].

Our goal is to perform style transfer for the original image, $B_{ori}$, and leave the structure and pathology information unaltered. The structural and pathology information are reflected at both the local and global scales in the image. U-Net's symmetrical direct connections at different resolution levels provide an effective solution to preserve those information during OCT image style transfer.

Unlike traditional discriminators that discriminate the probability of the truthfulness of an entire image, PatchGAN attempts to identify the truthfulness of each $p \times p$ patch in the image. It has been shown that $p$ can be much smaller than the full size of the image and still produce high-quality results. Smaller $p$ has fewer parameters, runs faster, and can be applied to arbitrarily large images. Such discriminator regards the image as a Markov random field and assumes independence between pixels separated by more than the patch diameter. The Markov discriminator tends to preserve high-resolution information leading to sharper results.

During training, the Adam solver [37] is used to optimize both the generator and the discriminator. We use the standard training method to alternately optimize the generator and the discriminator. In testing, only the generators are used.

*2) Implementation of mini-CGAN:* In this paper, we propose a mini-cGAN network for denoising, in which the generator and the discriminator utilize the PatchGAN mechanism to work with image patches rather than whole OCT images. The advantage of PatchGAN is that it requires much less training data because one whole image can generate many image patches for training. We use a U-Net network as the backbone for the generator, which takes $p_s \times p_s$ patches from the original image and generates same sized denoised image patches. Since most of the import retinal information in OCT images are located in the certain regions in the image, image patches are sampled from those regions with a high probability.

In the testing phase, we input patches with a size of $p_s \times p_s$ to the generator and scan through the whole image to build the entire denoised image. We set $n_o$, the number of overlapped patches at each pixel, as a controllable parameter. The final output image is the average of all the denoised overlapping patches. Detailed steps are described as follows.

1) Pad the original size $H \times W$ to $H_{pad} \times W_{pad}$, where $H_{pad} = (H + p_s + p_h)$, $W_{pad} = (W + p_s + p_w)$, and $p_h = p_s - H\%p_s$, $p_w = p_s - W\%p_s$, % is the remainder operation. Both of $H_{pad}$ and $W_{pad}$ are divisible by $p_s$.
2) Determine the top-left corner coordinates of the $i^{th}$ box, $(x_i, y_i)$, as $x_i = y_i = p_s \div n_o \times \Delta_i$, where $\Delta_i$ is the number of cycles from 0 to $(n_o - 1)$ and the box has a size of $(H_{pad} - p_s) \times (W_{pad} - p_s)$.
3) Split the box into non-overlapping patches with a size of $p_s \times p_s$, input all patches into the trained generator and stitch output patches into a single output image. Repeat this step to get $n_o$ output images.
4) Average the $n_o$ output images and remove the padded parts to get the final denoised image.

Due to the specificity of our scanner B, reflective artifacts caused by light signal reconstruction in the choroid area below retina in OCT image need to be removed appropriately. Our random patch training method resolves this issue because the input content of the network is the patch for each OCT image. Although most patches are located in the retina of the OCT, they also cover a variety of background areas, which allows the network to learn to remove artifacts appeared in the background.

## D. Data

We collected 3D volumes and slices raw OCT data from scanners B and T, and the study was approved by the Institutional Review Board of Suzhou University and informed consent was obtained from all subjects.

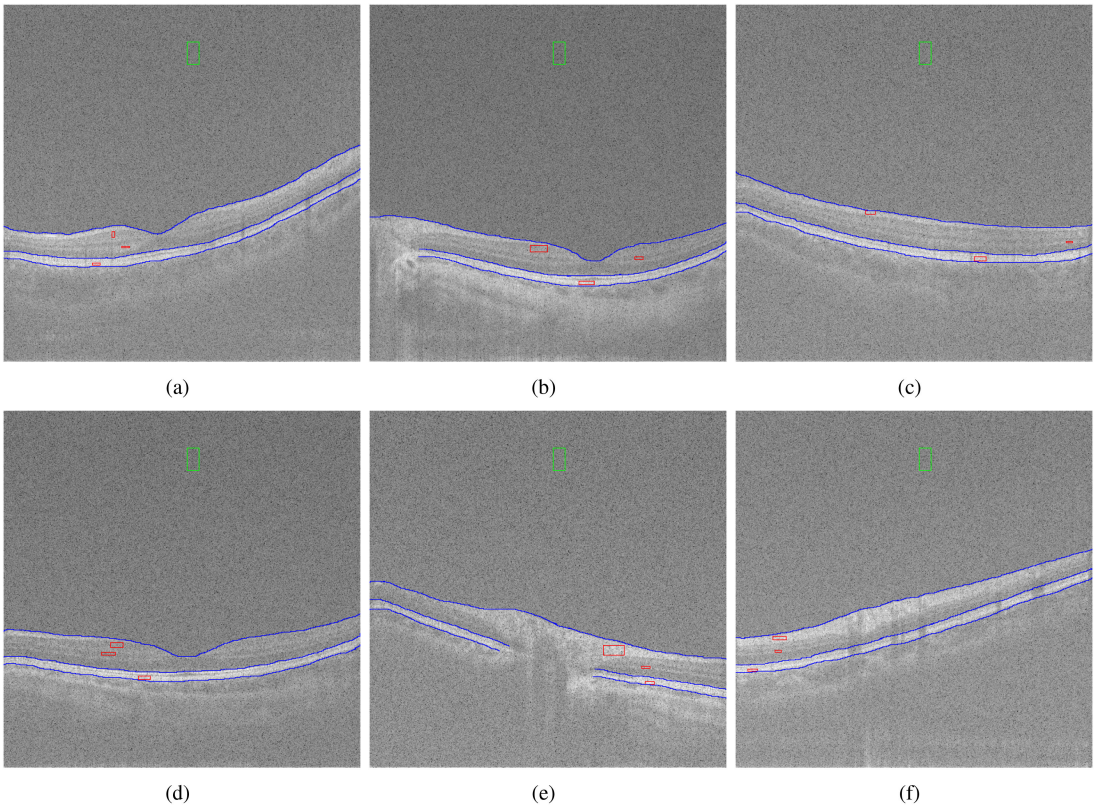


Fig. 3. Manually defined ROIs on 6 original noisy Bscans for performance metrics computation. Red and green rectangles represent signal and background regions defined for SNR and CNR calculation, respectively. Blue curves are boundaries defined for EPI computation. (a) and (b) are six-line scanning mode in macular pattern, (c) and (d) are macular region scanning mode in glaucoma pattern, and (e) and (f) are optic disc region scanning mode in glaucoma pattern, corresponding to the three different scanning modes as listed in Table I.

TABLE I
SPECIFICATIONS OF DATA $B$ FROM SCANNER B

| Pattern | Scanning Mode | Center wavelength (nm) | Physical size (mm) | Pixel size | Slice direction |
|---|---|---|---|---|---|
| Macular | Six-line | 840 | 9 | $1000 \times 1024 \times 18$ | Every 60 degrees |
| Glaucoma | Macular region | 840 | $6 \times 6$ | $512 \times 1024 \times 100$ | Horizontal |
| | Optic disc region | | $6 \times 6$ | $512 \times 1024 \times 100$ | Horizontal |

*1) Data From Scanner B:* Data from a non-commercial in-house OCT scanner B were acquired by scanning the central region of the 3D macular with a center wavelength of 840 nm. We collected three different scanning modes based on two patterns from 10 individuals. Fig. 3 shows 6 Bscan images for performance evaluation. All the data came from normal eyes. Table I lists the specifications of scanning modes and patterns.

*2) Data From Scanner T:* Scanner T consists of four commercial scanners, Topcon DRI-1 Atlantis (Japan Topcon Corporation), Topcon 3D OCT 2000 (Japan Topcon Corporation), Topcon 3D OCT 1000 (Japan Topcon Corporation), Zeiss Cirrus 4000 (German Carl Zeiss Corporation). We collected three datasets by using scanner T [1]: Training dataset1, Training dataset2 and Testing datasets.

Training dataset1 was acquired by scanning a 3D macular center region with a center wavelength of 1050 nm from Topcon DRI-1 Atlantis. The data volume size is $512 \times 992 \times 256$ (width $\times$ height $\times$ number of slices), which corresponds to the retina of the size of 6 mm $\times$ 2.6 mm $\times$ 6 mm. 20 acquisitions were repeated at short intervals from normal eyes.

Training dataset2 was acquired by scanning a 3D macular center region with a center wavelength of 840 nm from Topcon 3D OCT 2000. The data volume size is $512 \times 885 \times 128$ (width $\times$ height $\times$ number of slices), corresponding to the retina of the size of 6 mm $\times$ 2.3 mm $\times$ 6 mm. 20 acquisitions were repeated at short intervals from two eyes of a normal person.

The testing datasets contains 2D images of 9 different scanning modes from 4 different scanners. The data came from both normal and pathological eyes. Table II briefly lists the specifications of the 2 groups of training datasets and 9 testing data volumes, more details can be found in [1].

## E. Ground Truth for Training

*1) CycleGAN Training Data:* Though histogram matching can quickly match histograms between image domains, it does not produce a general mapping that can be applied directly to unseen images. In this study, we train a CycleGAN model that learns a style transfer mapping function from images collected by scanner B to these by scanner T. Since CycleGAN does not

TABLE II
SPECIFICATIONS OF DATA $T$ FROM SCANNER T

| Dataset | Scanner | Center wavelength (nm) | Bscan size (pixels/mm) | Normal/Pathological |
|---|---|---|---|---|
| Training dataset1 (Ground Truth) | Topcon DRI-1 | 1050 | $512 \times 992/6 \times 2.6$ | Normal |
| Training dataset2 | Topcon 2000 | 840 | $512 \times 885/6 \times 2.3$ | |
| Testing datasets | Topcon DRI-1 | 1050 | $512 \times 992/6 \times 2.6$ | Normal |
| | | | $512 \times 992/12 \times 2.6$ | Normal |
| | | | $512 \times 992/6 \times 2.6$ | Pathological (CSC) |
| | Topcon 1000 | 840 | $512 \times 480/6 \times 1.68$ | Normal |
| | | | $512 \times 480/6 \times 1.68$ | Pathological (CSC) |
| | Topcon 2000 | 840 | $512 \times 885/6 \times 2.3$ | Normal |
| | | | $512 \times 885/6 \times 2.3$ | |
| | Zeiss Cirrus 4000 | 840 | $512 \times 1024/6 \times 2$ | Pathological (PM) |
| | | | $512 \times 1024/6 \times 2$ | Pathological (CSC) |

require paired images for training, we use unpaired datasets collected by the two scanners for training.

*2) mini-CGAN Training Data:* The proposed mini-cGAN model requires paired raw/clean images for training, where raw images contain noises while clean images are the denoised counterparts. We utilized an existing dataset created by Ma *et al.* [1] for training. In the dataset, the denoised images were produced by registering and averaging Bscans images repeatedly acquired at the same location from the same person. We briefly describe the process as follows.

The 3D OCT acquisition scanning mode was performed on the same eye of the same person $M$ times. During the acquisition process, the person kept the eyeball as still as possible to make the differences among different acquisitions minimal. Assuming each scan volume contains $N$ images, $M \times N$ slice images for the person were produced in total. To create a training image pair, a random image was first selected from one of the volumes as the noised image. The $2n$ adjacent slices and itself from the same volume, and the $2n+1$ adjacent slices from the remaining $M-1$ volumes were selected, registered by affine transformation, and averaged to create the denoised counterpart. There were in total $L = (2n+1) \times M - 1$ slices registered and averaged to created the denoised image.

Because human eye cannot be completely still, averaging all $L$ slices often results to a blurred image. MSSIM score [38] between the noised image and its neighbor slice was used to rank the $L$ slices and only the first $l$ slices were kept to perform the registration and averaging. In addition, a piecewise linear image stretch was performed on the denoised image for contract enhancement. Intensity values less than the average of background region set to 0, and then all intensities were scaled to the interval of [0, 255].

The training "dataset1" and "dataset2" listed in Table II were used to created the ground truth. $M = 20$, $n = 3$ and $l = 60$ were chosen for "dataset1" while $M = 20$, $n = 3$, and $l = 40$ were chosen for "dataset2". These parameters were chosen by trial and error to obtain the best quality for denoised images. There are 256 image pairs in each of the datasets and all pairs were combined to train the mini-cGAN model.

## F. Evaluation Metrics

We utilize four performance metrics to evaluate different models including signal-to-noise ratio (SNR), contrast-to-noise ration (CNR), speckle suppression index (SSI) [15] and edge preservation index (EPI). Six regions of interest (ROIs) areas were manually defined to compute the performance metrics as shown in Fig. 3. Green rectangles represent background regions. Red rectangles denote signal regions (located in the retinal neural fiber layer (RNFL), inner retina, and the retinal pigment epithelium (RPE) complex, respectively). Blue boundaries (upper boundary of RNFL, inner-outer retina boundary and the lower boundary of RPE) denote the locations where EPI is calculated.

*1) Signal-to-Noise Ratio (SNR):* SNR reflects noise level in image, and it is defined as,

$$SNR = 10 \lg \left( \frac{\sigma_s^2}{\sigma_b^2} \right), \tag{9}$$

where $\sigma_s$ and $\sigma_b$ denote the standard deviation of the signal and background region, respectively.

*2) Contrast-to-Noise Ratio (CNR):* CNR is the contrast between signal region and background region,

$$CNR = 10 \lg \left( \frac{|\mu_i - \mu_b|}{\sqrt{\sigma_i^2 + \sigma_b^2}} \right), \tag{10}$$

where $\mu_i$ and $\sigma_i$ denote the mean and standard deviation of $i^{th}$ signal region in the image, while $\mu_b$ and $\sigma_b$ denote the mean and standard deviation of background region. In our experiments, the average CNR was computed over the 3 signal ROIs.

*3) Speckle Suppression Index (SSI):* SSI measures ratio between noise and denoised images. SSI metric tends to be less than 1 when noise is well reduced,

$$SSI = \frac{\sigma_o}{\mu_o} \times \frac{\mu_d}{\sigma_d}, \tag{11}$$

where $\mu_o$ and $\sigma_o$ denote the mean and standard deviation of a noisy image, $\mu_d$ and $\sigma_d$ denote the mean and standard deviation of the denoised version.

*4) Edge Preservation Index (EPI):* EPI reflects the extent of details of edge preserved in the image after denoising. EPI in the longitudinal direction is defined as,

$$EPI = \frac{\sum_i \sum_j |I_d(i+1,j) - I_d(i,j)|}{\sum_i \sum_j |I_o(i+1,j) - I_o(i,j)|}, \tag{12}$$

where $I_o$ and $I_d$ represent a noisy image and the denoised version, while $i$ and $j$ represent coordinates in the longitudinal and lateral direction in the image. EPI may not be an accurate indicator for edge-preservation if calculated over the entire image, since the gradient will become smaller in homogeneous regions after denoising. In our experiments, the neighborhood was set

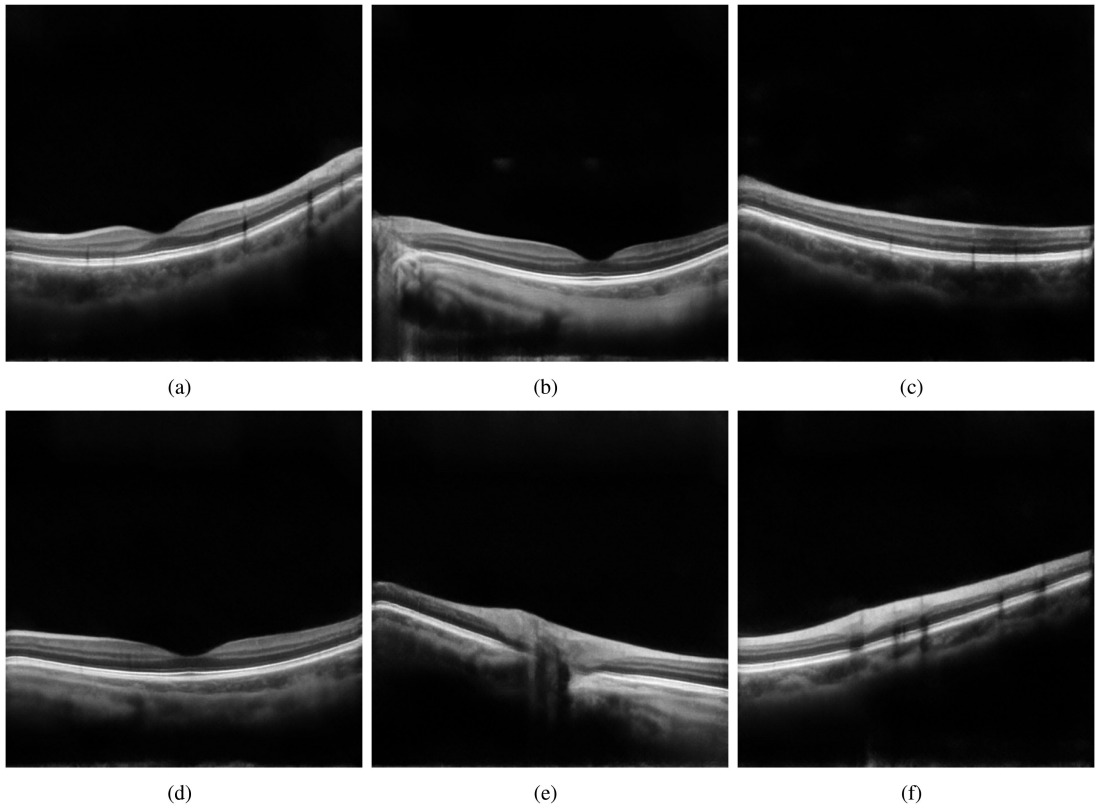


Fig. 4. Denoised Bscans by the proposed method, corresponding to the Bscans in Fig. 3.

as a band with height of 7 pixels centered at the boundaries as shown in Fig. 3.

## III. Experiments and Results

### A. Implementation Details

*1) Cyclegan:* We found that when data augmentation is used during training, it would make the loss difficult to converge and cause the generator to learn incorrect style transfer, i.e., generating retinal structure artifacts in the background region, which is undesirable. Therefore, we used the data from the scanner T without data augmentation to perform style transfer.

The generator of CycleGAN consists of a 7-layer U-Net, which has 7 up and 7 down layers. For the discriminator networks, we used a $70 \times 70$ PatchGAN, which aimed to classify whether or not $70 \times 70$ overlapping image patches were real or fake. The batch size was set to 1, and the number of training epochs was set to 100. The proposed CycleGAN was implemented using Pytorch and was trained using NVIDIA GTX 1660Ti GPU with 6 G memory.

*2) Mini-Cgan:* In cGAN training, we used data augmentation to improve training efficiency, including flipping along the transverse axis, scaling, rotation, and non-rigid transformations. Training data size was increased by a factor of two after the augmentation. For all the experiments, we set $\lambda = 10$ in [3] and $\alpha = 100$ in Eq [8], respectively, and utilized the Adam solver with an initial learning rate of 0.0002 and a momentum of 0.5 for optimization.

During training, we set patch size $p_s$ to 128. Small $p_s$ values will cause the model to ignore large structures of the retina, and large $p_s$ values result inadequate noise removal. We set the number of overlapping $n_o$ to 32. We found that if $n_o$ was too small, there were mesh-like noises presented in resulted images. This is because each output patch produced a distinct boundary, and the average of few patches resulted distinguish boundaries. While $n_o$ was too large, such as 64, resulted much longer computational time without significant performance improvement. The number of layers of the generator was a 5-layer U-Net, and the discriminator was an $8 \times 8$ PatchGAN. The batch size was set to 60, and the number of training epochs was 2000. The proposed mini-cGAN was implemented using Pytorch and was trained using NVIDIA GTX 1080Ti GPU with 12 G memory.

### B. Experimental Results

In this section, we will first show the results by the proposed model and qualitatively assess it by inspection. We then quantitatively compare it with other methods using the performance metrics described in Section II-F.

*1) Results by the Proposed Model:* Fig. 4 shows denoised images for the Bscans shown in Fig. 3. Through visual inspection, we observe that the proposed method can well retain the overall structures and details of the retina while suppressing speckle noise. The background is almost uniformly black, and contrasts among different layers of the retina are clear. The proposed model worked for different patterns, scanning modes,

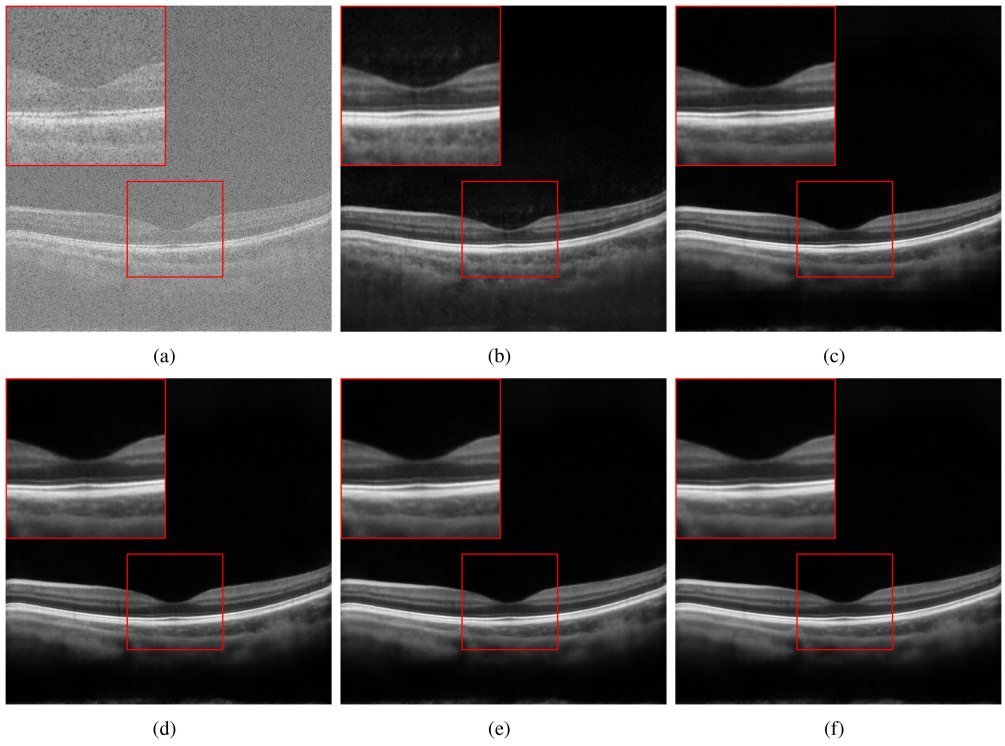


Fig. 5. Results for (d) Bscan in Fig. 3, a 150 × 150 region cropped from every images in the same position. (a) Original image, (b) Our previous method [1], (c) CycleGAN + mini-cGAN($n_o = 8$) (d) CycleGAN + mini-cGAN($n_o = 16$) (e) CycleGAN + mini-cGAN($n_o = 32$) (f) CycleGAN + mini-cGAN($n_o = 64$).

TABLE III
EVALUATION METRICS FOR DIFFERENT SETTINGS (MEAN ± STANDARD)

| Method | SNR | CNR | SSI | EPI |
|---|---|---|---|---|
| Original | 0.066±0.623 | 3.704±0.973 | 1.000±0.000 | 1.000±0.000 |
| Our Previous Method [1] | 20.740±4.564 | 11.780±0.964 | 0.113±0.009 | 0.917±0.074 |
| CycleGAN + mini-cGAN ($n_o$=8) | 19.695±3.047 | 12.019±1.325 | 0.090±0.011 | 0.992±0.072 |
| CycleGAN + mini-cGAN ($n_o$=16) | 20.082±2.882 | 12.273±1.417 | 0.090±0.012 | 0.993±0.074 |
| CycleGAN + mini-cGAN ($n_o$=32) | 20.857±2.813 | 12.459±1.528 | **0.089±0.011** | **0.998±0.073** |
| CycleGAN + mini-cGAN ($n_o$=64) | **21.506±2.774** | **12.593±1.463** | 0.089±0.012 | 0.995±0.073 |

different resolutions, and obtained qualitatively good results by inspection.

*2) Comparison to Our Previous Work:* We ran the proposed model with different values for $n_o$ and compared it to our previous method, Edge-cGAN [1]. Results are shown in Fig. 5 and Table III. Fig. 5 shows the denoised results from different runs. It is observed that denoised images by the proposed method have better contrast among different layers of the retina and darker background regions as compared to the result by our previous method (Fig. 5(b)). Table III lists the quantitative performance metrics of the competing methods. Both of our previous method and the proposed method significantly improved upon the original images. The proposed method outperformed our previous method in all the four performance metrics and achieved best SNR and CNR with $n_o = 64$, and best SSI and EPI with $n_o = 32$.

TABLE IV
STATISTICAL ANALYSIS ($p$-VALUE) OF THE PROPOSED METHOD VERSE OUR PREVIOUS METHOD

| Method | SNR | CNR | SSI | EPI |
|---|---|---|---|---|
| CycleGAN + mini-cGAN ($n_o$=8) | \ | 0.142 | 0.006 | 0.001 |
| CycleGAN + mini-cGAN ($n_o$=16) | \ | 0.060 | 0.007 | 0.001 |
| CycleGAN + mini-cGAN ($n_o$=32) | 0.078 | 0.027 | 0.006 | 0.000 |
| CycleGAN + mini-cGAN ($n_o$=64) | 0.050 | 0.035 | 0.006 | 0.000 |

Via $T$-test, we investigated the statistical significance of the performance improvements by the proposed method over our previous method. As can be seen in Table IV, all of the SSI and EPI improvements are statistically significant with $p$-values less

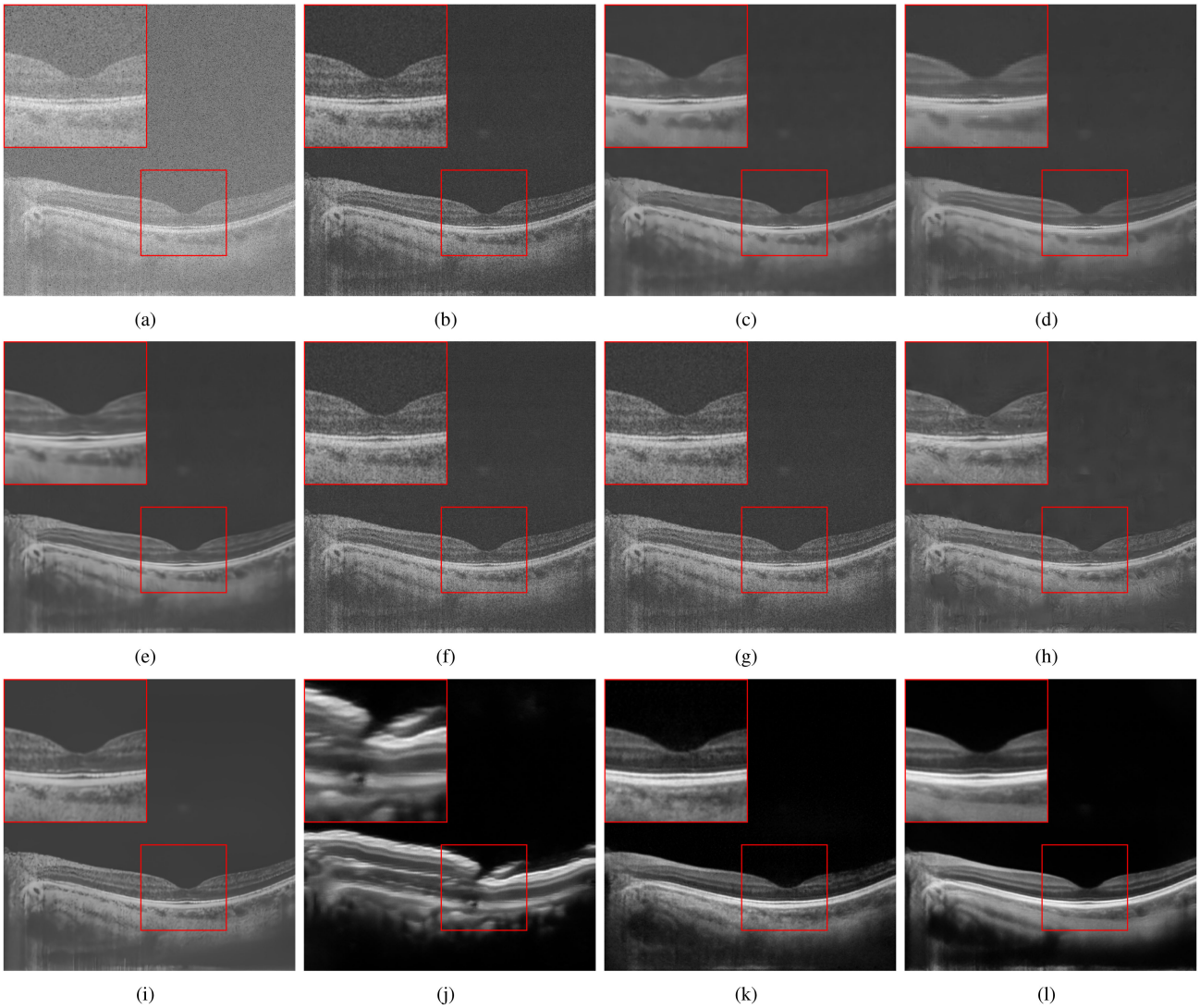


Fig. 6. Results for (b) Bscan in Fig. 3, a 150 × 150 region cropped from every images in the same position. (a) Original image (b) Original image after CycleGAN (c) NLM (d) BM3D (e) STROLLR (f) K-SVD (g) MAP (h) DnCNN (i) DPDNN (j) NAGAN + mini-cGAN ($n_o = 32$) (k) Our previous method [1] (l) Proposed (CycleGAN + mini-cGAN ($n_o$ =32)).

than 0.05. CNR improvements are significant for $n_0$ equaled 32 and 64. However, all SNR improvements are not significant.

*3) Comparison to the State of the Arts:* We compared the proposed method with the state-of-the-art methods, including non-local means (NLM) [39], block-matching and 3D filtering (BM3D) [40], sparsifying transform learning and low-rank method (STROLLR) [41], 3D complex wavelet based K-SVD for OCT denoising [19], maximum-a-posteriori (MAP) estimation based on local statistical model for OCT denoising [25], deep CNN with residual learning (DnCNN) [30], denoising prior driven DNN for image restoration (DPDNN) [31], noise adaptation generative adversarial network (NAGAN) [42], and our previous method: the edge-sensitive cGAN [1]. In these experiments, parameters for each of the methods were set so that it can achieve best result for the application.

Fig. 6 shows denoised images by the competing methods. Visual inspection reveals that the proposed method achieved the best result (Fig. 6(l)), our previous method ranks the second (Fig. 6(k)), NAGAN had serious artifacts (Fig. 6(j)), and all other methods either over-smoothed the input image (NLM, BM3D, and STROLLR, Fig. 6(c)–(e)) or still presented speckle noise in the results (Fig. 6(f)–(i)). Table V shows qualitative performance metrics indicating that the proposed method achieved the best CNR, SSI and EPI. DPDNN achieved the best SNR among all competing methods.

*4) Application in Retinal Layer Segmentation:* To test if the denoised OCT images can improve subsequent image analysis, we conducted retinal layer segmentation on original and denoised OCT images using the context pyramid fusion network (CPFNet) [43]. We collected 6272 retinal OCT B-scan images from 49 individuals, with 128 B-scan images acquired per person. The layer1 (from nerve fiber layer to outer plexiform layer), layer2 (from outer nuclear layer to myoid zone layer) and layer3 (from ellipsoid zone layer to retinal pigment epithelium layer) were manually labeled as ground truth under the supervision of a senior ophthalmologist. We divided the data into training and test sets with a ratio of 39 vs 10 individuals. Example segmentation results and average Dice coefficient (DSC) are shown in Fig. 7

TABLE V
EVALUATION METRICS FOR DIFFERENT DENOISING METHODS ON BSCANS WITH CYCLEGAN (MEAN ± STANDARD)

| Method | SNR | CNR | SSI | EPI |
|---|---|---|---|---|
| Original | 0.066±0.623 | 3.704±0.973 | 1.000±0.000 | 1.000±0.000 |
| NLM | 15.827±6.487 | 10.864±2.278 | 0.613±0.104 | 0.524±0.156 |
| BM3D | 6.975±6.182 | 11.229±1.166 | 0.607±0.096 | 0.427±0.082 |
| STROLLR | 16.851±5.008 | 11.339±2.009 | 0.612±0.101 | 0.455±0.082 |
| K-SVD | 8.153±2.570 | 6.219±0.927 | 0.548±0.084 | 0.878±0.132 |
| MAP | 6.873±2.764 | 6.996±1.272 | 0.561±0.083 | 0.738±0.087 |
| DnCNN | 11.510±3.496 | 8.291±1.078 | 0.598±0.100 | 0.601±0.122 |
| DPDNN | **25.600±3.991** | 10.660±2.807 | 0.604±0.101 | 0.509±0.111 |
| NAGAN + mini-cGAN ($n_o$=32) | 23.092±2.443 | 10.387±1.811 | 0.096±0.007 | 0.407±0.090 |
| Our Previous Method [1] | 20.740±4.564 | 11.780±0.964 | 0.113±0.009 | 0.917±0.074 |
| Proposed (CycleGAN + mini-cGAN ($n_o$=32)) | 20.857±2.813 | **12.459±1.528** | **0.089±0.011** | **0.998±0.073** |

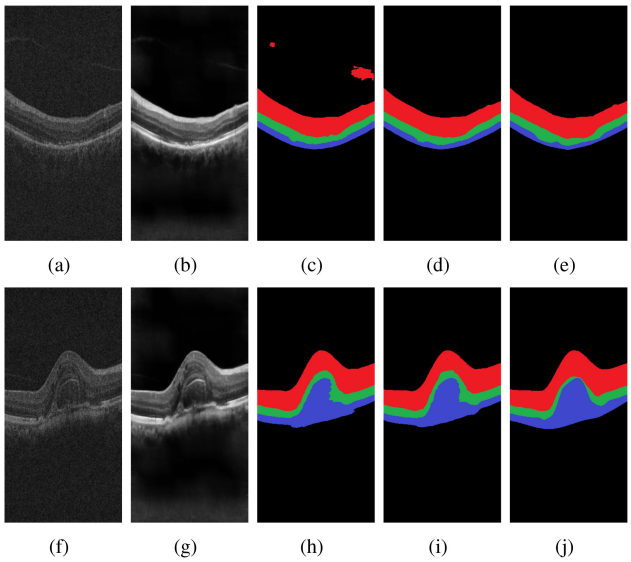


Fig. 7. Retinal layer segmentation results. (a) and (f) Original images, (b) and (g) Denoised images, (c) and (h) Segmentation results of original images, (d) and (i) Segmentation results of denoised images, (e) and (j) Ground truth. Red, green and blue regions represent layer1, layer2 and layer3, respectively.

TABLE VI
THE DICE COEFFICIENT OF SEGMENTATION (%)

| Target | Original | Denoised |
|---|---|---|
| Layer1 | 96.60 | **96.88** |
| Layer2 | 89.94 | **90.44** |
| Layer3 | 89.53 | **89.62** |
| Mean | 92.02 | **92.31** |

and Table VI, respectively. It is observed that the segmentation results on the denoised images have better structural information of retinal layers, and segmentation accuracies in terms of DSC are improved as compared with these on the original images.

## IV. DISCUSSION

In our previous study [1], we developed low-quality noise image and high-quality denoised image pair datasets for training. In the current investigation, we leveraged the training datasets to improve image quality of our in-house scanner (scanner B) by converting style of the raw image by scanner B to match the style of scanner T, with which the training datasets were created. The image style conversion was achieved by the CycleGAN model that learned different styles from unpaired images by both scanners. In addition, we modified the cGAN model in our previous study to a mini-GAN structure that achieved sharper details in the denoised images. The enhanced model outperformed our previous model in term of all the four performance metrics as shown in Table III.

The proposed method achieved the best visual quality as compared to the state-of-the-art methods (Fig. 6). NLM, BM3D and STROLLR suffered from excessive smoothing, leading to loss of information along regions with barely visible borders among layers. K-SVD and MAP did not effectively remove noise in both background and intra-retinal regions. MAP performed worse in preserving retinal structures (Fig. 6(g)), which may be due to unstable speckle noise parameter prediction. DnCNN performed well in retina regions but left some artifacts in background. DPDNN achieved good performances in background while the interlayer details were not well maintained. NAGAN had serious artifacts, which severely blurred retinal structures, though it performed well for contrast enhancement. As compared to the proposed method (Fig. 6(l)), our previous model did not preserve details among the layers under retina (Fig. 6(k)).

The quantitative evaluation metrics of the competing methods are listed in Table V. All methods improved image quality upon the original noise image as measured by all the four performance metrics. The proposed method achieved best results in CNR, SSI, and EPI, while DPDNN obtained the best SNR. The metrics of CNR, SSI, and EPI represent contrast, speckle suppression and edge preservation performances in the resulted image, respectively, indicating that the proposed method can preserve detail information while suppressing noise.

DPDNN over-smoothed input images and obtained better SNRs. DPDNN is a mean squared error (MSE) based deep learning model while the proposed method is based on the GAN loss. As pointed out by Ledig *et al.* [44], minimizing MSE maximizes SNR because the solution is given by the average of many possible candidate solutions, while GAN based optimization only sample one candidate as solution. Therefore, it is not surprising that DPDNN over-smoothed the image with a high SNR but could not preserve detail information. In practice, trade-offs must be made to best serve for specific applications.

CycleGAN has been applied in medical imaging field for cross modal synthesis such as generating CT-like images based on MR images [45], denoising [46], [47] and registration [48]–[50]. The cGAN model has been applied for supervised or semi-supervised learning for registration [51], denoising [52], reconstruction [53] and superresolution [54]. In the proposed method, we integrated CycleGAN and cGAN based on the PatchGAN mechanism to achieve cross-domain image denoising and contrast enhancement. The proposed model achieved a balance between noise suppression and contrast enhancement.

The objectives of speckle noise suppression in OCT images are to improve visual quality of the image and subsequent retinal layer segmentation performances. From Fig. 5 and Fig. 6, we can see the visual quality of denoised images is significantly improved as compared with the original images. Fig. 7 shows that the layer structure information in the denoised images are much sharper than those in the original images and the improved structure information did improve the retinal layer segmentation accuracies as shown in Table VI.

Our study has limitations. First, the proposed method is not an end-to-end learning model, and requires two separate steps to train the two modules. Second, the final result during testing is the average of multiple overlapped patches produced by the trained mini-cGAN model, which increases the testing complexity. As our future work, we will investigate an integrated model that incorporates image style transfer and noise suppression into a unified framework. In addition, we will explore methods to reduce parameters in the network to improve its efficiency.

## V. Conclusion

In this paper, we proposed a new speckle noise reduction method for retinal OCT images. The method consists of a CycleGAN model for image style transfer between two datasets and a mini-cGAN network for image denoising and enhancement. The mini-cGAN network was based on the PatchGAN mechanism in both training and testing. The proposed method achieved better results in speckle noise suppression, edge preservation, and contrast enhancement on retinal OCT images collected from different scanning modes, as compared to our previous method and other competing state-of-the-art models.

## References


[1] Y. Ma, X. Chen, W. Zhu, X. Cheng, D. Xiang, and F. Shi, "Speckle noise reduction in optical coherence tomography images based on edge-sensitive cgan," *Biomed. Opt. Exp.*, vol. 9, no. 11, pp. 5129–5146, 2018.

[2] D. Huang *et al.*, "Optical coherence tomography," *Science*, vol. 254, no. 5035, pp. 1178–1181, 1991.

[3] E. A. Swanson *et al.*, " *In vivo* retinal imaging by optical coherence tomography," *Opt. Lett.*, vol. 18, no. 21, pp. 1864–1866, 1993.

[4] M. Wojtkowski *et al.*, " *In vivo* human retinal imaging by fourier domain optical coherence tomography," *J. Biomed. Opt.*, vol. 7, no. 3, pp. 457–463, 2002.

[5] L. Duan, H. Y. Lee, G. Lee, M. Agrawal, G. T. Smith, and A. K. Ellerbee, "Single-shot speckle noise reduction by interleaved optical coherence tomography," *J. Biomed. Opt.*, vol. 19, no. 12, 2014, Art. no. 120501.

[6] B. F. Kennedy, T. R. Hillman, A. Curatolo, and D. D. Sampson, "Speckle reduction in optical coherence tomography by strain compounding," *Opt. Lett.*, vol. 35, no. 14, pp. 2445–2447, 2010.

[7] W. Cheng, J. Qian, Z. Cao, X. Chen, and J. Mo, "Dual-beam angular compounding for speckle reduction in optical coherence tomography," *Proc. Opt. Coherence Tomogr. Coherence Domain Opt. Methods Biomed. XXI*, vol. 10053, 2017, Art. no. 100532Z.

[8] J.-S. Lee, "Speckle analysis and smoothing of synthetic aperture radar images," *Comput. Graph. Image Process.*, vol. 17, no. 1, pp. 24–32, 1981.

[9] D. T. Kuan, A. A. Sawchuk, T. C. Strand, and P. Chavel, "Adaptive noise smoothing filter for images with signal-dependent noise," *IEEE Trans. Pattern Anal. Mach. Intell.*, no. 2, pp. 165–177, 1985.

[10] A. Lopes, R. Touzi, and E. Nezry, "Adaptive speckle filters and scene heterogeneity," *IEEE Trans. Geosci. Remote Sens.*, vol. 28, no. 6, pp. 992–1000, 1990.

[11] J. S. Lim, *Two-Dimensional Signal and Image Processing*. Englewood Cliffs, NJ, USA: Prentice-Hall, Inc., 1990.

[12] A. Hojjatoleslami and M. R. Avanaki, "OCT skin image enhancement through attenuation compensation," *Appl. Opt.*, vol. 51, no. 21, pp. 4927–4935, 2012.

[13] Z. Cao and X. Zhang, "PDE-based non-linear anisotropic diffusion techniques for medical image denoising," in *Proc. Spring Congr. Eng. Technol.*, 2012, pp. 1–4.

[14] P. Puvanathasan and K. Bizheva, "Interval type-II fuzzy anisotropic diffusion algorithm for speckle noise reduction in optical coherence tomography images," *Opt. Exp.*, vol. 17, no. 2, pp. 733–746, 2009.

[15] J. Aum, J.-h. Kim, and J. Jeong, "Effective speckle noise suppression in optical coherence tomography images using nonlocal means denoising filter with double gaussian anisotropic kernels," *Appl. Opt.*, vol. 54, no. 13, pp. D43–D50, 2015.

[16] X. Zhang, L. Li, F. Zhu, W. W. Hou, and X. Chen, "Spiking cortical model-based nonlocal means method for speckle reduction in optical coherence tomography images," *J. Biomed. Opt.*, vol. 19, no. 6, 2014, Art. no. 0 66005.

[17] B. Chong and Y.-K. Zhu, "Speckle reduction in optical coherence tomography images of human finger skin by wavelet modified BM3D filter," *Opt. Commun.*, vol. 291, pp. 461–469, 2013.

[18] F. Zaki, Y. Wang, H. Su, X. Yuan, and X. Liu, "Noise adaptive wavelet thresholding for speckle noise removal in optical coherence tomography," *Biomed. Opt. Exp.*, vol. 8, no. 5, pp. 2720–2731, 2017.

[19] R. Kafieh, H. Rabbani, and I. Selesnick, "Three dimensional data-driven multi scale atomic representation of optical coherence tomography," *IEEE Trans. Med. Imag.*, vol. 34, no. 5, pp. 1042–1062, May 2014.

[20] Z. Jian, L. Yu, B. Rao, B. J. Tromberg, and Z. Chen, "Three-dimensional speckle suppression in optical coherence tomography based on the curvelet transform," *Opt. Exp.*, vol. 18, no. 2, pp. 1024–1032, 2010.

[21] L. Fang, S. Li, Q. Nie, J. A. Izatt, C. A. Toth, and S. Farsiu, "Sparsity based denoising of spectral domain optical coherence tomography images," *Biomed. Opt. Exp.*, vol. 3, no. 5, pp. 927–942, 2012.

[22] L. Fang, S. Li, D. Cunefare, and S. Farsiu, "Segmentation based sparse reconstruction of optical coherence tomography images," *IEEE Trans. Med. Imag.*, vol. 36, no. 2, pp. 407–421, Feb. 2017.

[23] A. Wong, A. Mishra, K. Bizheva, and D. A. Clausi, "General bayesian estimation for speckle noise reduction in optical coherence tomography retinal imagery," *Opt. Exp.*, vol. 18, no. 8, pp. 8338–8352, 2010.

[24] A. Cameron, D. Lui, A. Boroomand, J. Glaister, A. Wong, and K. Bizheva, "Stochastic speckle noise compensation in optical coherence tomography using non-stationary spline-based speckle noise modelling," *Biomed. Opt. Exp.*, vol. 4, no. 9, pp. 1769–1785, 2013.

[25] M. Li, R. Idoughi, B. Choudhury, and W. Heidrich, "Statistical model for OCT image denoising," *Biomed. Opt. Exp.*, vol. 8, no. 9, pp. 3903–3917, 2017.

[26] I. Kopriva, F. Shi, and X. Chen, "Enhanced low-rank sparsity decomposition for speckle reduction in optical coherence tomography," *J. Biomed. Opt.*, vol. 21, no. 7, 2016, Art. no. 076008.

[27] J. Cheng *et al.*, "Speckle reduction in 3D optical coherence tomography of retina by a-scan reconstruction," *IEEE Trans. Med. Imag.*, vol. 35, no. 10, pp. 2270–2279, Oct. 2016.

[28] T. Remez, O. Litany, R. Giryes, and A. M. Bronstein, "Class-aware fully convolutional gaussian and poisson denoising," *IEEE Trans. Image Process.*, vol. 27, no. 11, pp. 5707–5722, Nov. 2018.

[29] D. Yang and J. Sun, "BM3D-Net: A convolutional neural network for transform-domain collaborative filtering," *IEEE Signal Process. Lett.*, vol. 25, no. 1, pp. 55–59, Jan. 2017.

[30] K. Zhang, W. Zuo, Y. Chen, D. Meng, and L. Zhang, "Beyond a gaussian denoiser: Residual learning of deep CNN for image denoising," *IEEE Trans. Image Process.*, vol. 26, no. 7, pp. 3142–3155, Jul. 2017.

[31] W. Dong, P. Wang, W. Yin, G. Shi, F. Wu, and X. Lu, "Denoising prior driven deep neural network for image restoration," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 41, no. 10, pp. 2305–2318, Oct. 2018.

[32] 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.*, 2017, pp. 1125–1134.

[33] J.-Y. Zhu, T. Park, P. Isola, and A. A. Efros, "Unpaired image-to-image translation using cycle-consistent adversarial networks," in *Proc. IEEE Int. Conf. Comput. Vis.*, 2017, pp. 2223–2232.

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

[35] D. Ulyanov, A. Vedaldi, and V. Lempitsky, "Instance normalization: The missing ingredient for fast stylization," 2016, *arXiv:1607.08022*.

[36] A. L. Maas, A. Y. Hannun, and A. Y. Ng, "Rectifier nonlinearities improve neural network acoustic models," in *Proc. icml*, vol. 30, no. 1, 2013, pp. 3–8.

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

[38] 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.

[39] A. Buades, B. Coll, and J.-M. Morel, "A non-local algorithm for image denoising," *Proc. IEEE Comput. Soc. Conf. Comput. Vis. Pattern Recognit.*, vol. 2, 2005, pp. 60–65.

[40] K. Dabov, A. Foi, V. Katkovnik, and K. Egiazarian, "Image denoising by sparse 3-d transform-domain collaborative filtering," *IEEE Trans. Image Process.*, vol. 16, no. 8, pp. 2080–2095, Aug. 2007.

[41] B. Wen, Y. Li, and Y. Bresler, "When sparsity meets low-rankness: Transform learning with non-local low-rank constraint for image restoration," in *Proc. IEEE Int. Conf. Acoust., Speech Signal Process.*, 2017, pp. 2297–2301.

[42] T. Zhang *et al.*, "Noise adaptation generative adversarial network for medical image analysis," *IEEE Trans. Med. Imag.*, vol. 39, no. 4, pp. 1149–1159, Apr. 2019.

[43] S. Feng *et al.*, "CPFNET: Context pyramid fusion network for medical image segmentation," *IEEE Trans. Med. Imag.*, vol. 39, no. 10, pp. 3008–3018, Oct. 2020.

[44] C. Ledig *et al.*, "Photo-realistic single image super-resolution using a generative adversarial network," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, 2017, pp. 4681–4690.

[45] J. M. Wolterink *et al.*, "Deep MR to CT synthesis using unpaired data," in *Proc. Int. Workshop Simul. Synth. Med. Imag.*, 2017, pp. 14–23.

[46] E. Kang, H. J. Koo, D. H. Yang, J. B. Seo, and J. C. Ye, "Cycle-consistent adversarial denoising network for multiphase coronary ct angiography," *Med. Phys.*, vol. 46, no. 2, pp. 550–562, 2019.

[47] C. You *et al.*, "CT super-resolution GAN constrained by the identical, residual, and cycle learning ensemble (GAN-CIRCLE)," *IEEE Trans. Med. Imag.*, vol. 39, no. 1, pp. 188–203, Jan. 2020.

[48] C. Tanner, F. Ozdemir, R. Profanter, V. Vishnevsky, E. Konukoglu, and O. Goksel, "Generative adversarial networks for MR-CT deformable image registration," 2018, *arXiv:1807.07349*.

[49] D. Mahapatra, B. Antony, S. Sedai, and R. Garnavi, "Deformable medical image registration using generative adversarial networks," in *Proc. IEEE 15th Int. Symp. Biomed. Imag.*, 2018, pp. 1449–1453.

[50] D. Mahapatra, Z. Ge, S. Sedai, and R. Chakravorty, "Joint registration and segmentation of xray images using generative adversarial networks," in *Proc. Int. Workshop Mach. Learn. Med. Imag.*, 2018, pp. 73–80.

[51] P. Yan, S. Xu, A. R. Rastinehad, and B. J. Wood, "Adversarial image registration with application for MR and TRUS image fusion," in *Proc. Int. Workshop Mach. Learn. Med. Imag.*, 2018, pp. 197–204.

[52] J. Wang, Y. Zhao, J. H. Noble, and B. M. Dawant, "Conditional generative adversarial networks for metal artifact reduction in CT images of the ear," in *Proc. Int. Conf. Med. Image Comput. Comput.-Assist. Interv.*, 2018, pp. 3–11.

[53] H. Liao, Z. Huo, W. J. Sehnert, S. K. Zhou, and J. Luo, "Adversarial sparse-view CBCT artifact reduction," in *Proc. Int. Conf. Med. Image Comput. Comput.-Assist. Interv.*, 2018, pp. 154–162.

[54] I. Sánchez and V. Vilaplana, "Brain MRI super-resolution using 3D generative adversarial networks," 2018, *arXiv:1812.11440.*