Multiscale denoising generative adversarial network for speckle reduction in optical coherence tomography images

Xiaojun Yu[a], **Chenkun Ge**[a], **Mingshuai Li**[a], **Muhammad Zulkifal Aziz**[a], **Jianhua Mo**[b] and **Zeming Fan**[a,*]

[a]Northwestern Polytechnical University, School of Automation, Xi'an, China
[b]Soochow University, School of Electronics and Information Engineering, Suzhou, China

Abstract

Purpose: Optical coherence tomography (OCT) is a noninvasive, high-resolution imaging modality capable of providing both cross-sectional and three-dimensional images of tissue microstructures. Owing to its low-coherence interferometry nature, however, OCT inevitably suffers from speckles, which diminish image quality and mitigate the precise disease diagnoses, and therefore, despeckling mechanisms are highly desired to alleviate the influences of speckles on OCT images.

Approach: We propose a multiscale denoising generative adversarial network (MDGAN) for speckle reductions in OCT images. A cascade multiscale module is adopted as MDGAN basic block first to raise the network learning capability and take advantage of the multiscale context, and then a spatial attention mechanism is proposed to refine the denoised images. For enormous feature learning in OCT images, a deep back-projection layer is finally introduced to alternatively upscale and downscale the features map of MDGAN.

Results: Experiments with two different OCT image datasets are conducted to verify the effectiveness of the proposed MDGAN scheme. Results compared those of the state-of-the-art existing methods show that MDGAN is able to improve both peak-single-to-noise ratio and signal-to-noise ratio by 3 dB at most, with its structural similarity index measurement and contrast-to-noise ratio being 1.4% and 1.3% lower than those of the best existing methods.

Conclusions: Results demonstrate that MDGAN is effective and robust for OCT image speckle reductions and outperforms the best state-of-the-art denoising methods in different cases. It could help alleviate the influence of speckles in OCT images and improve OCT imaging-based diagnosis.

 [DOI: 10.1117/1.JMI.10.2.024006]

Keywords: optical coherence tomography; medical and biological imaging; image despeckling; generative adversarial network.

Paper 22204GRR received Aug. 4, 2022; accepted for publication Mar. 13, 2023; published online Mar. 30, 2023.

1 Introduction

Optical coherence tomography (OCT) is a low-coherence interferometry-based imaging modality capable of providing depth-resolved microstructure images of biological tissues.[1] Owing to its noninvasive and high-resolution properties, OCT has widely been adopted for various disease diagnoses, especially in ophthalmology.[2] Due to its low-coherence process resulting from the coherent addition of photons scattered back with random phase and amplitude, however, OCT inevitably suffers from speckle noise. Speckle largely reduces the image quality and degrades the

*Address all correspondence to Zeming Fan, fanzeming@nwpu.edu.cn

2329-4302/2023/$28.00 © 2023 SPIE

accuracy of OCT imaging-based disease diagnoses.[3] Therefore, speckle reduction in OCT images is highly desired to improve the clinic utility of OCT.

Over the past years, OCT speckle reduction has attracted extensive research interest, and numerous despeckling methods have been proposed.[4] Typically, OCT image despeckling methods can be categorized into hardware-based and software-based ones according to the different types of techniques adopted.[5] Specifically, the hardware-based methods usually include angular compounding,[6,7] frequency compounding,[8] polarization-,[9] and spatial-diversity-based[10] system designs. Although such methods are effective and robust, they are not readily to be implemented for existing commercial OCT systems. For example, for those angular compounding and polarization-diversity-based methods, extensive modifications must be made to the OCT systems, and some special materials or/and designs are typically required. Such requirements increase both system complexity and costs, making such schemes backward incompatible and could not be extended to the existing commercial OCT systems. While for both frequency compounding and spatial-diversity-based methods, although extensive hardware changes are not necessarily needed, their system spatial resolutions are typically sacrificed during the imaging process, which thus degrades the system performance.

Contrarily, the software-based denoising methods can be easily implemented onto the commercial systems since no special requirements are needed on the system design. Typically, the software-based techniques could be further categorized into model-based and deep learning-based ones. The model-based methods usually try to restore details of the potential clean images using optimization schemes, such as nonlocal filter,[11,12] sparse coding,[13] effective prior,[14] wavelet transform,[15] and low rank.[16] The nonlocal weighted sparse representation (NWSR),[17] as well as the block-matching and 3D filtering (BM3D)[18] schemes are typical model-based despeckling methods. Although BM3D employs a block level estimation for denoising based on image self-similarity, NWSR utilizes the sparse representation of multiple similar noisy and denoised patches to improve patch estimation. However, it is worth noting that BM3D usually suffers from the edge ringing effect when processing those images with high complexity and low contrast, and NWSR vectorized patches may disrupt the structures of the reconstructed images in certain cases. Both methods suffer from deficiency in preserving detailed structures in OCT images.

In recent years, as artificial intelligence is receiving increasing research interest, various deep learning-based denoising methods have also been proposed in the literature. Tajmirriahi et al. implemented a lightweight convolution network as deep autoencoders (AEs) to simulate the latest state-of-the-art method in OCT image denoising. Results show that the AE has good performance in speckle OCT denoising.[19] Anoop et al.[20] proposed a cascaded convolutional neural network (CNN) architecture for OCT image despeckling and also adopted it eliminates the impacts of noises on OCT datasets obtained with different devices.

More recently, various methods with either unsupervised or semisupervised training schemes have also been proposed.[21,22] By utilizing up and downsampling networks to generate denoising and super-resolution OCT images concurrently, Qiu et al.[23] proposed a semisupervised learning method called N2NSR-OCT for both denoising and image super-resolution. In addition, Ni et al.[24] proposed a speckle-modulating mechanism, namely, Sm-Net OCT, to extract speckle properties for OCT image speckle removing with generative adversarial network (GAN). Although it demonstrated that Sm-Net OCT helps improve both image quality and imaging depth, it suffers from heavy training loads, especially for those large-scale images. Such is because for Sm-Net OCT, the speckle patterns play a key role in model training, yet they are typically difficult to extract and characterize. Recently, we also proposed a generative adversarial network with multiscale convolution and dilated convolution res-network (MDR-GAN) for OCT despeckling.[25] By utilizing the convolution and dilated convolution res-network blocks to improve the network learning ability, MDR-GAN achieved satisfactory despeckling effects. It is worth noting that the generator of MDR-GAN consists of three parts, namely, feature extraction, feature mapping, and feature reconstruction, which would impose complexity onto the system design. Huang et al.[26] used an unsupervised method, namely, DRGAN-OCT, for speckle reduction without employing matched image pairs. By decomposing the noisy images into content and noise spaces with an encoder first, and then adopting a generator to predict the denoised

image contents with the extracted features, DRGAN-OCT saves the number of clean images required for network model training.

More recently, various methods with either unsupervised or semisupervised training schemes have also been proposed.[21,22] However, it is worth noting that, although such semisupervised or unsupervised deep-learning schemes could help alleviate the requirement for the large number of clean images utilized by the supervised training schemes, their overall despeckling effects are typically less satisfactory. Furthermore, due to the complicated deep-learning schemes employed for training, those semisupervised and unsupervised methods are usually complicated and computationally extensive. Therefore, in clinical practice, network architectures with both relatively simpler training strategies and stronger learning abilities are highly desired to achieve satisfactory denoising effects with a limited number of training datasets.

This paper proposes a multiscale denoising generative adversarial network (MDGAN) for OCT speckle reductions. Specifically, a cascade multiscale module (CMSM) is proposed to recover the multiscale image features and increase the network learning capacity first, and then a deep back-projection (DBP) layer is employed to upscale and downscale the feature maps alternatively. After that, a loss function is finally devised to regenerate the high-frequency image information. The main contributions of this paper are as follows.

- MDGAN, which requires only a limited number of clean and noisy image pairs, is proposed for OCT image despeckling.
- A CMSM is proposed to recover the multiscale OCT image features while increasing the network learning capability.
- The spatial attention mechanism (SAM) combining with different loss functions is presented in the training scheme for regeneration of the most common details in OCT images.
- Experiments are conducted to compare MDGAN with state-of-the-art OCT despeckling methods in different cases for performance verifications.

The remainder of this paper is organized as follows. Section 2 introduces the MDGAN network architecture and the proposed training scheme in detail. Section 3 describes the experimental setup and the performance metrics. Section 4 presents the experimental results. Section 5 concludes this paper.

2 Method

2.1 *Generative Adversarial Network Architecture*

GANs have widely been used for denoising, super-resolution, classification, and other related image processing areas. For a denoising GAN, the image sets are typically expressed as with x_i and y_i being an image pair. Assuming that x is a noisy image and y is the corresponding noise-free image, and the objective of denoising GAN is to find the mapping between the input image x and its corresponding noise-free image y.

In the proposed GAN model, we define the generator and the discriminator as G and D, respectively. The discriminator D inputs a high-dimension vector, such as a picture, and outputs a scalar. The more realistic the input picture is, the larger the scalar. Initially, the clean image is fed to the discriminator, which results in a higher scalar, while the generated denoised image gets a lower scalar. After several training epochs, the scalar value of the generated denoised image input into the discriminator will increase. Once the quality of the denoised image $G(x)$ generated by G is high enough, the discriminator D will regard it as realistic, and the training process tends to balance. The principle of the GAN model can be expressed with the following min–max optimization to optimize the generator G and the discriminator D:

$$\min_G \max_D V(G,D) = E_{y\sim P_y}[\log(D(y))] + E_{x\sim P_x}[\log(1-D(x))], \tag{1}$$

where $E[\bullet]$ denotes the expectation function, whereas P_y and P_x denote the real and the noisy data distributions, respectively. When the GAN model reaches the maximum and minimum optimization, both generator and discriminator reach an equilibrium state.

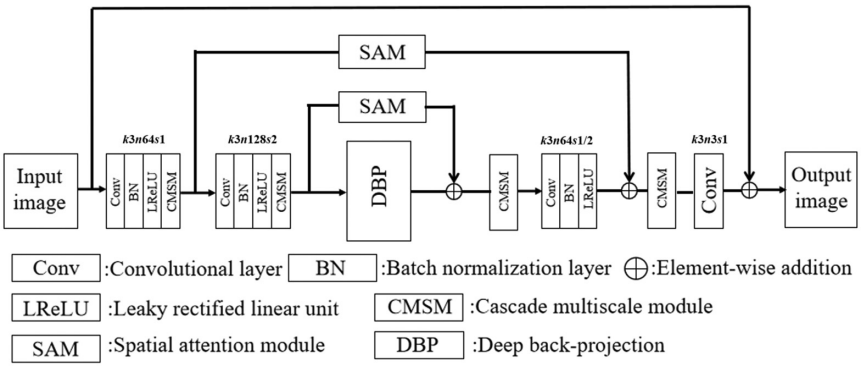

Fig. 1 The generator architecture of the proposed MDGAN, wherein *k* represents the kernel size of the convolution layers, *n* is the filter number, and *s* represents the stride.

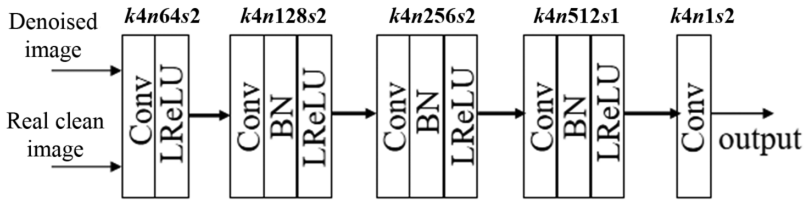

Fig. 2 The discriminator architecture of the proposed MDGAN, namely, PatchGAN used to determine whether an input image is a real clean image or a denoised image.

Figure 1 shows the schematic of the proposed generator of network architecture in MDGAN. As seen, the network is based upon a U-Net[27] with skip connections. Specifically, in such an architecture, the SAM and the element-wise addition function are employed, while the DBP layer is placed between the encoder and decoder. Element Conv is the convolutional layer, while BN and LReLU represent the batch normalization[28] and leaky rectified linear unit,[29] respectively. In MDGAN, the CMSMs are employed to capture the multiscale features, whereas SAMs are used to refine the denoised image, and the DBP layer alternatively upscales and downscales the feature maps to capture image information. Furthermore, the k, n, and s denote the kernel size, filter number, and stride, respectively. To further balance the network performances and computation cost, appropriate parameters k or n are selected, and the discriminator D in MDGAN adopts a 70×70 PatchGAN as shown in Fig. 2 to distinguish between the real clean and the denoised images.[30] The residual learning method is also introduced to link the input and output feature maps.[31]

2.2 *Cascade Multiscale Module*

Based upon a multiscale cross-work, the proposed CMSM module is employed to improve information flow and capture feature.[32] This study applies a series of 3×3 convolutional layers to capture the multiscale image features. Figure 3(a) shows that a CMSM contains three single multiscale modules (MSMs). The input feature maps flow into those MSMs continuously, whereas the output feature maps are concatenated together and directed into a 1×1 convolutional layer. The MSM is shown in Fig. 3(b), wherein the input feature map is uniformly split into four subsets, with x_i being the i'th subset input. Compared with the original feature map, the channel number in each subset x_i is reduced to 1/4th of its original. The four subset outputs are concatenated into a 3×3 convolutional layer denoted as $K_i(\cdot)$, and therefore, the MSM can be defined as

$$y_i = \begin{cases} x_i, & i = 1 \\ K_i(x_i), & i = 2 \\ K_i(x_i + y_{i-1}), & 2 < i \leq 4 \end{cases}, \tag{2}$$

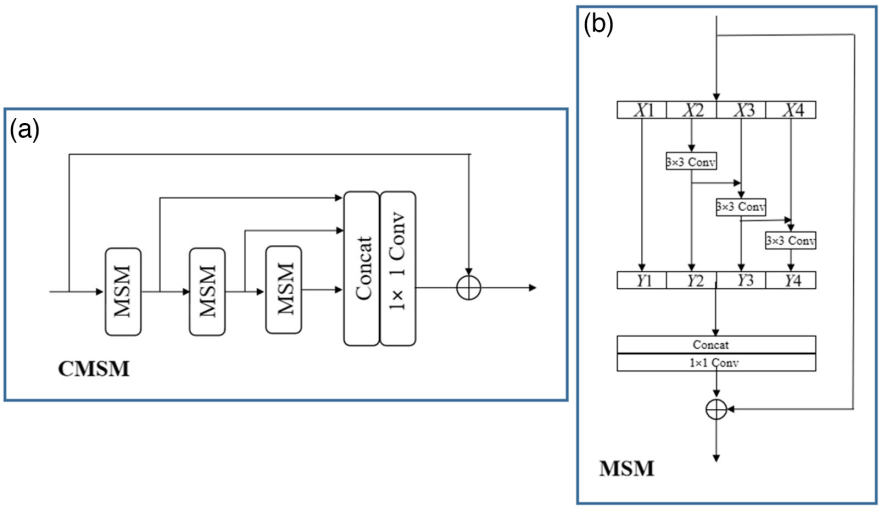

Fig. 3 The network architecture of the (a) CMSM and (b) an MSM.

where y_i is the output of $K_i(\cdot)$. Owing to the balanced connection, MSM gradually decreases the gap between the input and the output feature maps, which helps gain a large receptive field for the output feature map. Finally, the four splits with different scales are concatenated into a 1×1 convolutional layer.

2.3 *Spatial Attention Mechanism*

Figure 4 presents the spatial attention module (SAM).[33] The main objective of utilizing SAM module is to refine the denoising results with enhanced computational speed. Assume a feature map $F \in R^{C\times H\times W}$ is input into SAM, where C, H, and W denote the channels, height, and width of F, respectively, then the average and max-pooling functions are adopted to perform down-sampling onto F to generate double one-channel maps. Finally, such maps are concatenated and fed into a 7×7 convolutional layer to generate a new 2D spatial attention map $M \in R^{1\times H\times W}$. The operation process of SAM could be mathematically expressed as follows:

$$M_s(F) = \sigma(f^{7\times 7}([\text{AvgPool}(F); \text{MaxPool}(F)])), \tag{3}$$

where σ is the sigmoid function, and $f^{7\times 7}$ represents a 7×7 convolutional layer. Finally, the output feature map of SAM can be given as follows:

$$F' = M_s(F) \otimes F. \tag{4}$$

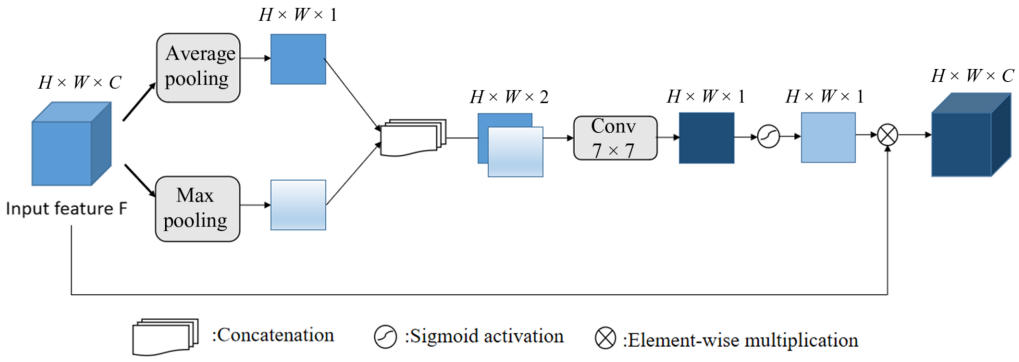

Fig. 4 The spatial attention module with average pooling and max pooling to refine the denoising results.

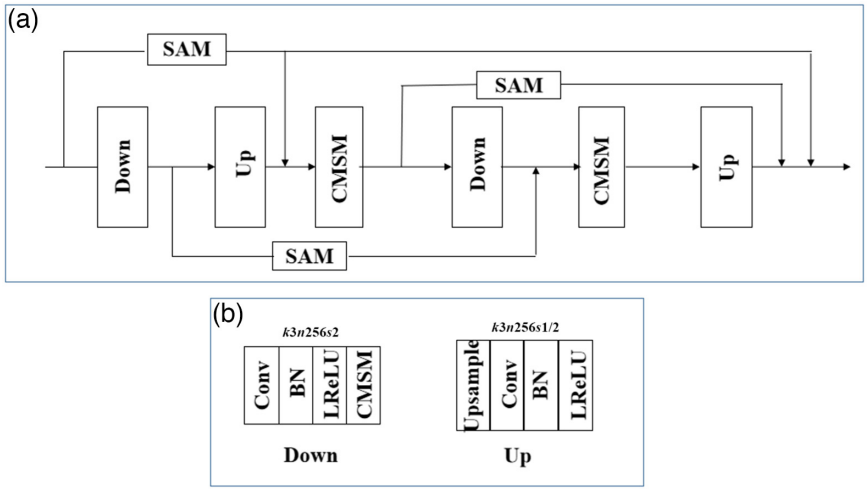

Fig. 5 (a) Simplified network architecture of the deep back-projection layer. (b) Architecture of the down and the up blocks. The parameters *k*, *n*, and *s* represent the convolution layer kernel size, the filter number, and the stride number, respectively.

2.4 *Deep Back-Projection Layer*

A DBP layer is adopted in MDGAN to exploit the upsampling and downsampling layers, which helps identify the mutual relation of noisy and clean image pairs.[34] As shown in Fig. 5(a), the upsampling and downsampling layers are utilized to alternatively upscale and downscale the feature maps. Specifically, in the downsampling layer, the kernel size, filter number, and stride number are 3, 256, and 2, respectively. The upsampling layer with an upsampling block presents a transposed convolutional layer with the same configuration as that of the downsampling layer. Moreover, the connection between two layers employs an SAM model to better capture the visual structures, whereas the CMSM module again is utilized to find sufficient information for high-frequency detail refinements in Fig. 5(b).

2.5 *Object Function*

The MDGAN discriminator D determines whether a real clean or denoised image is more realistic.[35] In this study, the least square GAN, which adopts the least square loss to minimize the divergence of Pearson, χ^2 is used to estimate the distribution between the denoised images and the real clean images. Generally, the adversarial loss is used to restore the high-frequency image information and reduce the blurring effects caused by L_2 loss. Therefore, the objective functions of D and G are described, respectively, as follows:

$$\begin{aligned}\min_D V_{\mathrm{LSGAN}}(D) &= 0.5E_{y\sim P_y}[(D(y)-1)]^2 + 0.5E_{x\sim P_x}[(D(G(x)))]^2 \\ \max_G V_{\mathrm{LSGAN}}(G) &= 0.5E_{x\sim P_x}[(D(G(x))-1)]^2,\end{aligned} \tag{5}$$

where y and x denote a real clean image and its corresponding noisy image, respectively, whereas $D(y)$ and $G(x)$ represent the discriminator and generator accordingly.

For generator G, the mean-square-error (MSE) loss, also named L_2 loss, is used to maintain the image details and structure contents. The mean error loss could realize a pixel-wise error minimization between the denoised image $G(x)$ and its real clean image y. The L_2 loss function is described as follows:

$$L_{\mathrm{MSE}} = \frac{1}{N}\sum_{i=1}^{N}\|y - G(x)\|^2. \tag{6}$$

For image denoising in this study, since L_2 loss does not usually match the image quality perceived by human eyes, the restored images suffer from the loss in image details. Since L_2 loss does not usually match the image quality perceived by human eyes, the restored images usually suffer from the loss of image details. In addition, although the L2 loss could help reduce the pixel content gap between the denoised and the clean images, it is expected that the feature representation of the denoised and the clean images can be reduced. In this study, such an objective is obtained by employing a pretraining network like VGG-16,[31] a low-weight auxiliary VGG loss function is encapsulated into the generator loss for image despeckling. The VGG loss changes the computational space from the image into the feature domain, and thus it helps address the above issues. The VGG loss is defined as follows in this study:

$$L_{\text{vgg}} = \frac{1}{N}\sum_{i=1}^{N} \|\text{VGG}_{16}(y) - \text{VGG}_{16}(G(x))\|, \tag{7}$$

where VGG_{16} presents the VGG-16 network. In addition, the edge loss is also adopted to retrieve sharp edge information, and it is described as follows:

$$L_{\text{edge}} = \frac{\sum_i \sum_j |T_x(i+1,j) - T_x(i,j)|}{\sum_i \sum_j |T_y(i+1,j) - T_y(i,j)|}, \tag{8}$$

where T_x and T_y are two-dimensional matrices representing the denoised image and its corresponding real clean counterpart, whereas i and j represent the i'th row and j'th column of a two-dimensional image matrix.

In summary, the overall generator loss function combined with L_{edge}, L_{vgg}, L_{MSE}, and L_{GAN} is formulated:

$$L_G = \alpha L_{\text{MSE}} + \beta L_{\text{GAN}} + \gamma L_{\text{vgg}} + \lambda L_{\text{edge}}, \tag{9}$$

where α, β, γ, and λ are the coefficients of corresponding losses. In this study, they are hyperparameters with fixed values chosen empirically in the experiments. The values of each part will be shown in Sec. 3.2

It is worth noting that, when GAN is employed for image-to-image transformation, hallucinations typically occur when those images are of different domains. To address such an issue, a loss function, e.g., L1-loss, is usually employed in literature during the image conversion process to eliminate their influences. In this study, however, since only retinal images are employed for verifying the effectiveness of MDGAN, and those retinal images are only transformed from the noisy ones to the clean ones within the same domain, the generated hallucination will not seriously impact on the denoising results. The features of the noise images and the ground-truth images correspond to each other, and thus the hallucination problem will not change the structure of the denoised images. The VGG loss is utilized with a small weight for MDGAN, whereas the MSE loss is mainly used to maintain strong similarity between image pairs in this study.

3 Experiments

3.1 *Dataset*

Experiments with two different publicly available OCT datasets were conducted to verify the effectiveness of MDGAN. Both datasets are comprised of central foveal images, which are collected by the same SD-OCT imaging systems from Bioptigen, Inc. (Durham, North Carolina, United States) with an axial resolution of ~4.5 μm per pixel in tissue.[36] The first one is named as dataset CF, and the second one is called dataset SS in this study. Specifically, the first dataset CF contains 17 retinal OCT image pairs acquired from normal and abnormal subjects, and each image pair includes a noisy SD-OCT image and a corresponding ground-truth image obtained by registering and averaging several B-scans acquired at the same position. As the size of each

image was different, all images are cropped with an anchor point of geometric image center at a resolution of 500×950 (height $\times$ width) to facilitate the network training and testing process.

In the experiments, there are 17 image pairs within dataset CF: seven pairs were removed for their limited qualities, leaving only 10 pairs remaining for experiments, of which eight image pairs were randomly chosen for network training, while the remaining two pairs were used for testing. Specifically, by traversing each of the eight image pairs with a 256×256 window and a stride of 50, a total number of 2552 patch pairs were generated for training, and such obtained noisy patches were directed to the generator for training to realize extracted feature mapping between the noisy images and the real clean ones. Finally, the remaining two image pairs in dataset CF were processed by the network for comparisons. For fair comparisons, the existing mechanisms were implemented following exactly the way that they were reported, and all the parameters were tuned to achieve their respective best performances. The same 2552 patch pairs and the two noisy images were also adopted for their training and testing, respectively, and those noisy images processed by those existing methods were compared with those by MDGAN quantitatively and qualitatively in the testing phase.

The dataset SS contains five OCT image pairs with a size of 448×800 (height $\times$ width). Similarly, after removing one mismatched image pair, only four were left and employed for experiments. In the experiments, since both datasets CF and SS are collected by the same OCT device, it is assumed that the speckle distribution patterns of the two datasets are the same, even though image sizes are different for the two datasets. Therefore, the MDGAN model trained by dataset CF is employed for processing those images in dataset SS in the testing phase. All performance metrics and those images used for comparisons are shown in the following sections.

3.2 *Parameter Setting*

The generator and discriminator of MDGAN are optimized using the adaptive momentum estimation (Adam) optimizer, with $\beta_1 = 0.5$ and $\beta_2 = 0.999$. All the other network parameters are also marked in Figs. 1 and 2. Specifically, in the training part, the learning rate was set to be first, and then it was gradually decreased to with a decaying factor of 0.1 in every 50 epochs. Such a training process terminates until the losses of both the generator and the discriminator reach a balanced state that is set for the model. In this study, the total number of steps for the whole training is set to be 200, whereas the coefficients of the generator loss were empirically chosen to be $\alpha = 1$, $\beta = 0.001$, $\gamma = 0.006$, and $\lambda = 0.0005$ to achieve better performance. The networks were implemented in Python with PyTorch framework, and all experiments were conducted on a workstation (Intel Xeon W-2145 CPU at 3.70 GHz) and were accelerated by an NVIDIA Quadro GPU with 5 GB memory.

3.3 *Evaluation Metrics*

The MDGAN performance is assessed with six indicators, including peak-single-to-noise ratio (PSNR), edge preservation index (EPI), the equivalent number of looks (ENL), contrast-to-noise ratio (CNR), structural similarity index measurement (SSIM), and signal-to-noise ratio (SNR). PSNR, EPI, and SSIM are computed for the entire image, whereas ENL, SNR, and CNR are measured within several regions of interest (ROIs). The denoising efficacy of MDGAN is validated by comparing MDGAN with the other state-of-the-art denoising methods. A brief overview of those performance metrics is as follows.

3.3.1 *Peak signal-to-noise ratio*

The PSNR is the main metric that measures the similarity between the denoised image and the reference image, and it can be expressed as follows:

$$\mathrm{PSNR}(r, g) = 10 \log_{10}(255^2 / \mathrm{MSE}(r, g)), \tag{10}$$

$$\mathrm{MSE}(r, g) = \frac{1}{MN} \sum_{i=1}^{M} \sum_{j=1}^{N} (r_{ij} - g_{ij})^2, \tag{11}$$

where $r_{i,j}$ and $g_{i,j}$ represent the pixel values at the corresponding coordinates of clean region and denoise region, respectively, whereas M and N are the height and width of the image.

3.3.2 *Edge preservation index*

The EPI indicates the extent of image edge detail preservations after processing, and it is defined as follows:

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

where I_d denotes the denoised image, I_n denotes the noisy image, and i, j represent the i'th row and j'th column of an image. Generally, a higher EPI value implies better edge preservations.

3.3.3 *Equivalent number of looks*

ENL is a typical parameter used for evaluating OCT image speckle reductions, and it measures the smoothness of the homogeneous region of the denoised image. ENL is calculated over the background ROI of each test image as follows:

$$\mathrm{ENL} = \frac{\mu_b^2}{\sigma_b^2}, \tag{13}$$

where μ_b and σ_b represent the mean and standard deviation of selected background ROI in each image, respectively.

3.3.4 *Contrast-to-noise ratio*

CNR measures the contrast between the signal and the background regions, and it is defined as follows:

$$\mathrm{CNR} = \frac{1}{m}\sum_{i=1}^{m}\left[10\log_{10}\left(\frac{\mu_i - \mu_b}{\sqrt{\sigma_i^2 + \sigma_b^2}}\right)\right], \tag{14}$$

where m is the number of all selected signal ROIs; μ_i and σ_i are the mean and standard deviation of the i'th selected ROI, whereas μ_b and σ_b are the mean and standard deviation of the select background ROI.

3.3.5 *Structural similarity index measurement*

SSIM is a full reference metric being widely used for image quality evaluation. For an image x and an image y, SSIM between them could be calculated as follows:

$$\mathrm{SSIM}(i,b) = \frac{(2\mu_i\mu_b + C_1)(2\sigma_{ib} + C_2)}{(\mu_i^2 + \mu_b^2 + C_1)(\sigma_i^2 + \sigma_b^2 + C_2)}, \tag{15}$$

where μ_b/μ_i and σ_b/σ_i are the mean and standard deviation of a clean/denoised region, respectively, whereas σ_{ib} denotes the cross correlation between the clean and denoised regions. C_1 and C_2 are the positive stabilizing constants.

3.3.6 *Signal-to-noise ratio*

SNR is the ratio of signal mean to background standard deviation, which is defined as

$$\mathrm{SNR} = \frac{\mu_{\mathrm{sig}}}{\sigma_{\mathrm{bg}}}, \tag{16}$$

where μ_{sig} and σ_{bg} are the mean of signal region and the standard deviation background region, respectively. Note that the SNR here is not the same as the definition for OCT signals,[37] instead, it is defined for image analysis only with an arbitrary unit.[38]

4 Results and Discussion

Experiments with two publicly available OCT retinal image datasets, namely, dataset CF and dataset SS, are conducted to verify the effectiveness of MDGAN in this study.[36] The images from different datasets are also processed by the state-of-the-art existing denoising methods, including BM3D,[18] Sm-Net OCT,[24] MDR-GAN,[25] DRGAN-OCT,[26] and DnCNN,[31] for comparisons in different cases. In this study, two OCT retinal images from dataset CF and four from dataset SS were processed for comparison.

4.1 *Experiment with the Dataset CF*

Two OCT images as shown in Figs. 6(a) and 7(a) from dataset CF are employed for experiments to verify the effectiveness of MDGAN, and those images that are denoised with the different methods are shown in Figs. 6 and 7.

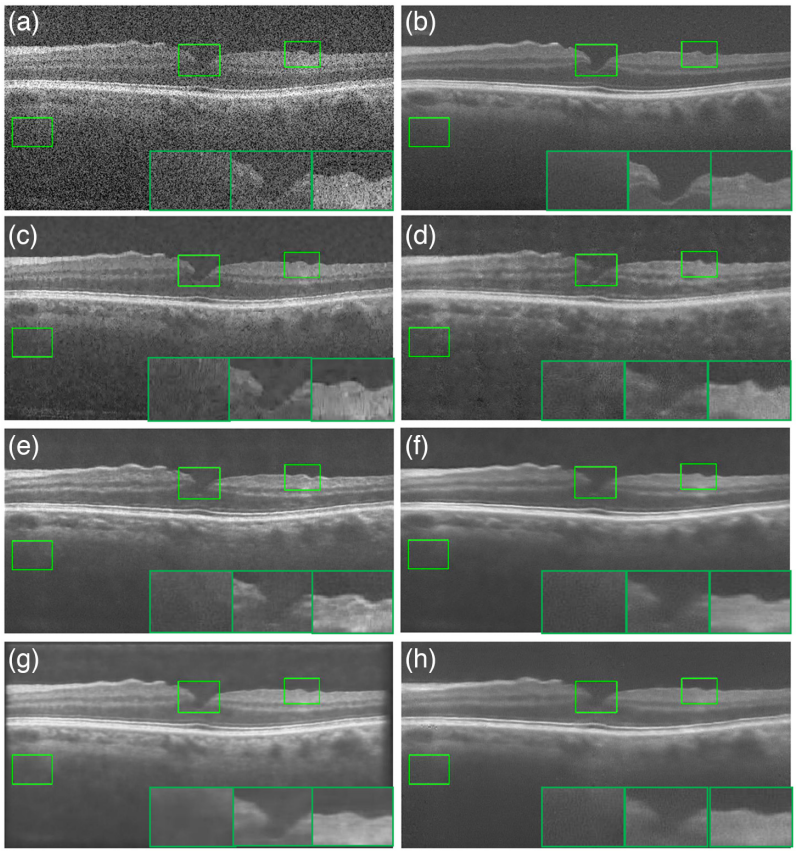

Fig. 6 The OCT retinal image from dataset CF that is processed by different denoising methods: (a) original noisy image, (b) ground-truth image, (c) BM3D, (d) DnCNN, (e) Sm-Net OCT, (f) MRD-GAN, (g) DR-GAN OCT, and (h) MDGAN.

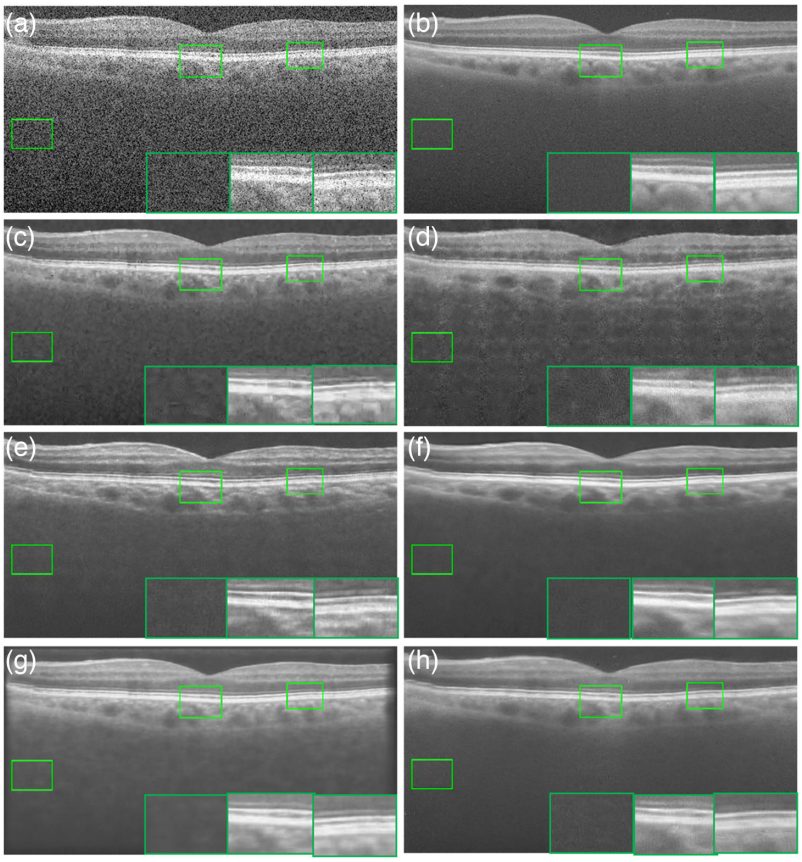

Fig. 7 Another OCT retinal image selected from dataset CF that is processed with different methods: (a) original noisy image, (b) ground-truth image, (c) BM3D, (d) DnCNN, (e) Sm-Net OCT, (f) MRD-GAN, (g) DR-GAN OCT, and (h) MDGAN.

Results in Fig. 6 demonstrate that all existing methods achieve satisfactory results in speckle reductions, and the performances of different denoising effects are somewhat different. Specifically, as shown in Figs. 6(c)–6(g), the speckles are largely suppressed as compared with the original noisy image shown in Fig. 6(a), and the performances of Sm-Net OCT are better than those of BM3D and DnCNN, with better visual effects achieved. Due to the complexity of the image and the limited number of training image pairs, however, there still exist some speckles in the obtained images, and therefore, the image is not smooth enough. Figure 6(f) shows that the images processed by MDR-GAN are natural, and the background noises are also well suppressed. By utilizing a small number of image pairs for training, DR-GAN OCT achieves certain denoising results as shown in Fig. 6(g). However, due to the unsupervised learning scheme adopted, such denoising effects are still relatively limited, especially for the background noises. Figure 6(h) shows the image denoised by MDGAN. As seen, MDGAN achieves satisfactory despeckling effects, wherein the speckles are largely suppressed, and therefore, the obtained image in Fig. 6(h) is much smoother with clearer structural details as compared with those processed by the other schemes, as shown in Figs. 6(c)–6(g).

The typical ROIs in those figures are also selected and marked with green rectangles for comparisons. As seen in the enlarged ROIs of Figs. 6(c)–6(g), although speckles are largely suppressed, speckle residues still exist in the background, and the key structures are a bit blurred, which thus deteriorates the image visual effects. In contrast, as shown in Fig. 6(h) by MDGAN, speckles in the background regions are almost eliminated, and the structure details in those images are well reserved, and thus the visual effects are much better.

The same results can also be observed in Figs. 7(c)–7(h). Such results convincingly demonstrate that the proposed MDGAN scheme is capable of speckle reductions while reserving image structure details.

4.2 *Experiment with the Dataset SS*

To further verify the effectiveness and robustness of MDGAN, experiments are also conducted with the other four image pairs from dataset SS, and those obtained images are compared with those ones processed by the other existing denoising schemes in different cases. Since those retinal images from datasets SS and CF are collected by the same OCT device, it is reasonably assumed that the noise distribution patterns are the same for all images within those two datasets, and therefore, the MDGAN network trained by dataset CF is also employed to process those images from dataset SS.

Figure 8 presents a randomly selected image from dataset SS that is processed with different denoising methods. As seen, similar results are obtained as those of the ones from dataset CF. Again, it could be observed in Figs. 8(c)–8(s) that the images processed by BM3D, DnCNN, Sm-Net OCT, MDR-GAN, DR-GAN OCT, and MDGAN, achieve better visual effects as compared with the original noisy one. As can be seen, the images processed by those mechanisms look much smoother, and those speckles are largely suppressed, or even eliminated, in background regions. Among those existing methods, MDR-GAN achieves the best visual effects. Due to the limited number of image pairs for training, however, there still exist speckle residues, especially for BM3D and DnCNN, and thus, blurring effects are introduced, which hides the detailed image microstructures. In contrast, for the image processed by MDGAN, it could be observed that speckles are largely suppressed, and thus the image visual effects are comparable with those of the ones processed by MDR-GAN, DnCNN, and Sm-Net OCT, i.e., Figs. 8(d)–8(f), respectively. It could also be observed in enlarged ROIs shown in Fig. 8(h), the speckles in the background are almost eliminated, while the image microstructures are well preserved.

Similar results could also be observed with the other image randomly chosen from dataset SS, as shown in Fig. 9. As can be seen, the speckles in Fig. 9(f) are largely suppressed, and its

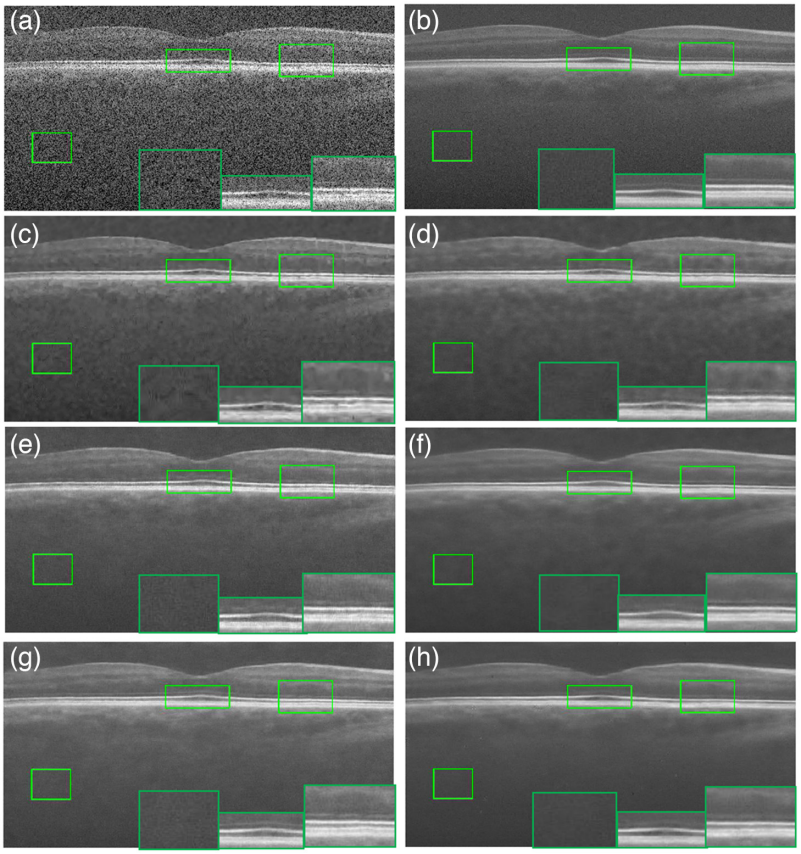

Fig. 8 Results of a OCT retinal image randomly selected from the dataset SS that is processed different denoising methods: (a) original noisy image, (b) ground-truth image, (c) BM3D, (d) DnCNN, (e) Sm-Net OCT, (f) MRD-GAN, (g) DR-GAN OCT, and (h) MDGAN.

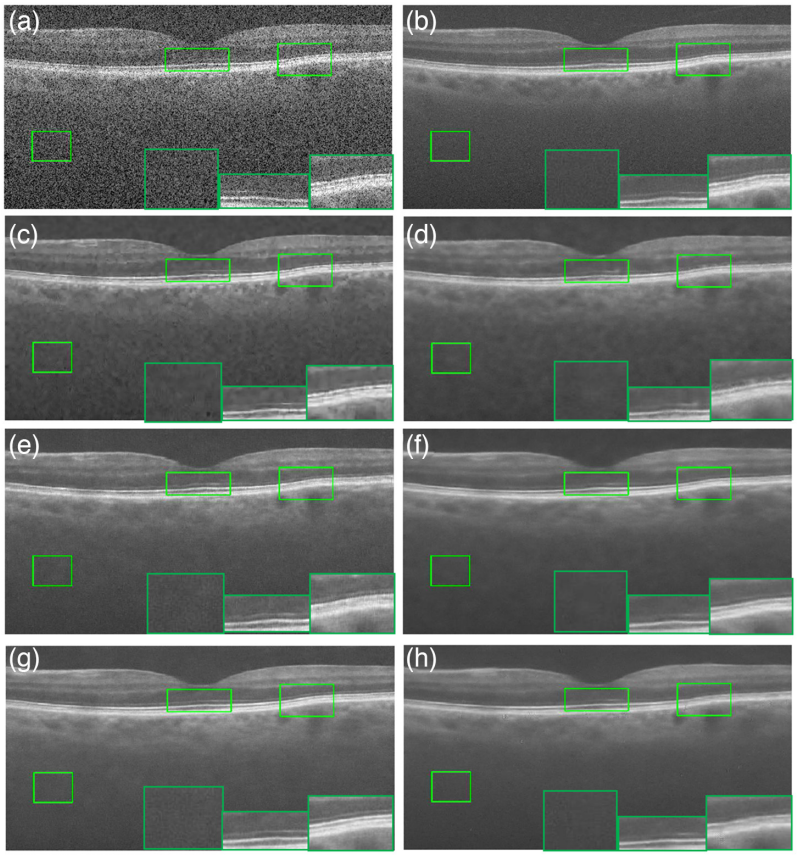

Fig. 9 Another OCT retinal image randomly selected from dataset SS that is denoised by the different denoising methods: (a) original noisy image, (b) ground-truth image. (c) BM3D, (d) DnCNN, (e) Sm-Net OCT, (f) MRD-GAN, (g) DR-GAN OCT, and (h) MDGAN.

visual effects are better those in Figs. 9(c)–9(g) and are even comparable to the ground-truth image as shown by Fig. 9(b). Meanwhile, the image structural details, as illustrated by the ROIs shown in the green rectangles, are also well preserved and could be clearly seen. Such results again convincingly demonstrate that MDGAN is not only capable of suppressing speckles in OCT images but also could preserve the image details in the image denoising process.

It is also worth noting, however, that at high zoom-in levels for those images obtained by MDGAN, there still exist small artifacts in some areas, as shown in Figs. 8(h) and 9(h). The reason for such artifacts is mainly because of the strong learning capability of MDGAN. As shown in Figs. 8(b) and 9(b), there exist minimal speckles in the ground-truth image, and once it is used for training, such speckles could be recognized as structural details, which thus produce artifacts in some areas, although such artifacts would not impact on the overall denoising results. Our next-step work is to eliminate such artifacts.

4.3 *Other Metrics with the Two Datasets*

To measure the performances of different mechanisms, the other performance metrics, e.g., EPI, ENL, and CNR, are also calculated for those images processed with different denoising methods. The ROIs in each figure are marked manually at a same position, wherein the green areas denote the background and the signal ROIs, respectively, and all the metrics are averaged over those images in each test dataset.

Table 1 compares the metrics of different methods with those images from dataset CF. As seen in Table 1, MDGAN ranks first in PSNR and SSIM among all those despeckling methods. Specifically, for PSNR, MDGAN outperforms BM3D, SM-Net OCT, and DnCNN by 0.70, 2.51, and 0.94 dB, respectively, and it is also 0.08 dB higher than MDR-GAN, and 6.56 dB higher than

Table 1 Quantitative results of the dataset CF with different methods.

	PSNR (dB)	SSIM (a.u.)	EPI (a.u.)	SNR (a.u.)	CNR (a.u.)	ENL (a.u.)
BM3D[18]	26.9797	0.5790	0.8311	31.4884	1.8031	122.1865
DnCNN[31]	26.7410	0.5258	**0.9073**	29.4314	1.7396	85.6478
Sm-Net OCT[24]	25.1684	0.6413	0.8005	34.2933	1.6645	280.4071
MDR-GAN[25]	27.5995	0.6776	0.7485	**36.1875**	1.4422	402.6848
DR-GAN OCT[26]	21.1168	0.6384	0.3980	35.4881	**1.9773**	**530.6097**
MDGAN	**27.6790**	**0.6788**	0.6969	35.3061	0.7699	526.9630
GT	—	1	0.9001	32.6777	2.1387	180.5701

Note: The bold values denote the best values that were obtained for a certain metric among all those de-speckling schemes been compared.

DR-GAN OCT. And for SNR, MDGAN is 12.12% and 19.96% higher than BM3D and DnCNN, respectively, and it is also 331% and 515% higher for ENL, respectively. Such results prove that MDGAN is effective in speckle reductions in OCT images.

It is also worth noting that for EPI, although MDGAN is slightly lower than those of the other methods, it is still quite similar to those of the existing methods. Such results indicate that all those methods could effectively preserve the image edges in their denoising processes. However, due to the different smoothing effects introduced by the different reference denoising region selection strategies, ENL varies for each method in their despeckling processes. Specifically, since MDGAN processes the selected background regions one by one and the loss functions impose a constraint when processing the region edges, MDGAN ranks fifth among all methods when processing the high-resolution images from dataset CF.

The same image processing procedure has also been applied onto images from dataset SS, and the performance metrics are calculated. As shown in Table 2, MDGAN performs the best in SSIM, SNR, and ENL, and it also achieves the second-highest PSNR among all denoising methods, demonstrating that MDGAN is effective and robust in speckle reductions for OCT images. It is also worth noting that the other metrics in Table 2, except for PSNR, are a bit different from those in Table 1. For example, MDGAN is almost 3.3 times that of BM3D for ENL. The reason for such observations is that different methods may change the image signal intensities in their despeckling process, and since the resolution of those images in dataset SS is relatively low, ENL ranks first and CNR decreases slightly. In addition, the average processing time required by different methods is also recorded as shown in Table 3. Results show that MDGAN takes about

Table 2 Quantitative results of the dataset SS with different methods.

	PSNR (dB)	SSIM (a.u.)	EPI (a.u.)	SNR (a.u.)	CNR (a.u.)	ENL (a.u.)
BM3D[18]	29.3074	0.6785	0.6669	34.1420	1.9459	216.5167
DnCNN[31]	29.3784	0.7033	0.6419	36.5897	**2.3019**	425.7947
Sm-Net OCT[24]	26.4423	0.6178	**0.8533**	34.5856	2.0481	263.4014
MDR-GAN[25]	**30.1599**	0.7112	0.6060	38.0569	2.2734	611.8077
DR-GAN OCT[26]	24.9696	0.6581	0.8373	35.6786	2.1290	391.8944
MDGAN	29.5970	**0.7205**	0.4953	**38.9112**	1.5851	**729.6187**
GT	—	1	0.8980	33.1317	1.9396	183.7867

Note: The bold values denote the best values that were obtained for a certain metric among all those de-speckling schemes been compared.

Table 3 The average processing time required to process an image from dataset SS.

Method	BM3D	DnCNN	Sm-Net OCT	MDR-GAN	DR-GAN	MDGAN
Times (s)	15.54	2.02	1.06	2.59	2.26	4.64

4.64 s on average to process an image with a size of 448×800 from dataset SS, which could be expected to meet the time required for clinical diagnosis after certain optimizations.

To better compare, the performances among different methods, the metrics of all those methods are normalized to their respective best ones. Results presented in Figs. 10(a) and 10(b) demonstrate that MDGAN achieves the highest PSNR and SSIM for dataset CF, while it ranks first in SSIM, SNR, and ENL for dataset SS among all those despeckling methods compared. While for the other metrics, MDGAN performs similarly as compared with the other methods. Such results demonstrate that MDGAN is effective and robust for speckle reductions in OCT images.

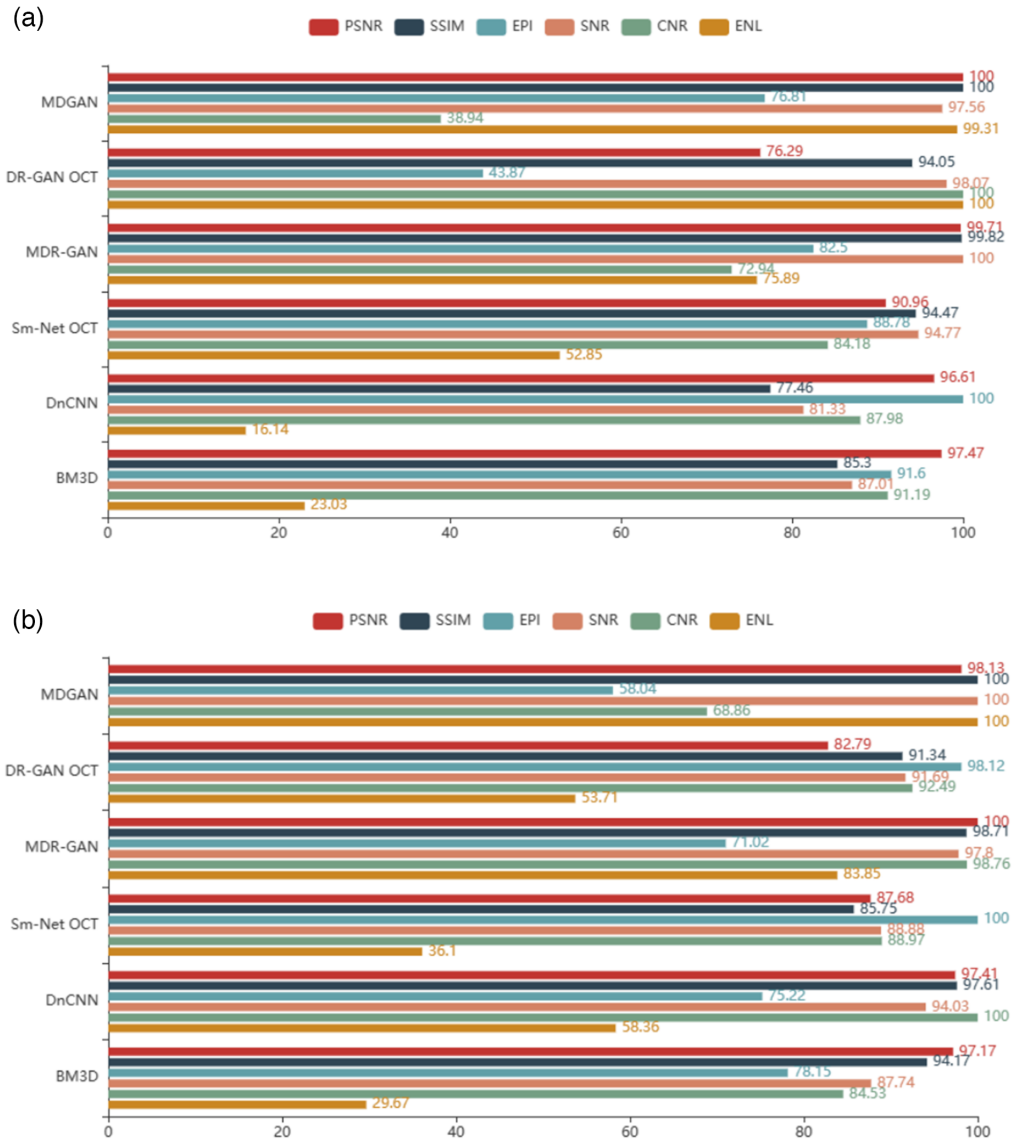

Fig. 10 Performances of different methods obtained with different datasets: results with (a) dataset CF and (b) dataset SS.

4.4 *Ablation Experiment*

To verify the effectiveness of each network module and loss functions, ablation experiments are also carried out with the network architecture designed. With different modules integrated, Table 4 shows the different performance metrics of MDGAN for an image randomly chosen from dataset SS. As seen, all those three modules, i.e., SAM, DBP, and CMSM, could improve the performances of PSNR, SSIM, SNR, CNR, and ENL, whereas DBP helps improve PSNR better, and SAM helps improve SSIM, SNR, and ENL. Among the three modules, DBP impacts on PSNR the most, and without such a module, the PSNR could be reduced by 4.54%. On the contrary, SAM module has the most influences on SSIM, SNR, and ENL. As seen, without such a module, such metrics could be reduced by 11.49%, 9.80%, and 58.02%, respectively.

We also conducted ablation experiments on the design of loss function. We will set up three different loss functions for ablation experiments. The specific ablation results are shown in Table 5. From the numerical results, no matter which combination of losses is missing, it will affect the balance of network training and produce poor quality denoising images. Among them, the trained model without L_{vgg} and L_{edg} will perform worst in PSNR, SSIM, SNR, and ENL. Introduction of loss function L_{vgg} could improves PSNR, SSIM, SNR, and ENL, which increases 1.33%, 0.79%, 0.52%, and 1.65% compared with the trained model without L_{vgg} and L_{edg}, respectively. Similarly, loss function L_{edg} also increased on these indicators, which improves 2.27%, 3.47%, 5.01%, and 35.93%, respectively. Finally, combining all loss functions will lead to greater growth, which raises 5.04%, 13.14%, 11.87%, 8.33%, and 158.81% on PSNR, SSIM, SNR, CNR, and ENL compared with the trained model without L_{vgg} and L_{edg}, respectively. Therefore, these data can show that the added loss function of L_{vgg} and L_{edg} is very effective for network training.

Table 4 Quantitative results of ablation experiment on SS without different modules.

	PSNR (dB)	SSIM (a.u.)	EPI (a.u.)	SNR (a.u.)	CNR (a.u.)	ENL (a.u.)
MDGAN without SAM	28.1248	0.6148	**0.8719**	33.5411	2.2505	212.3067
MDGAN without DBP	27.8378	0.6327	0.8405	34.7992	2.1988	300.8238
MDGAN without CMSM	27.9035	0.6286	0.8542	34.4206	**2.2893**	267.0651
MDGAN	**29.5970**	**0.7205**	0.4953	**38.9112**	1.5851	**729.6187**
GT	—	1	0.8980	33.1317	1.9396	183.7867

Note: The bold values denote the best values that were obtained for a certain metric among all those de-speckling schemes been compared.

Table 5 Quantitative results of ablation experiment on SS with different loss functions.

	PSNR (dB)	SSIM (a.u.)	EPI (a.u.)	SNR (a.u.)	CNR (a.u.)	ENL (a.u.)
	28.4245	0.6218	**0.8681**	33.4324	1.9630	198.6409
	28.6881	0.6352	0.8421	34.9026	**2.1903**	265.6429
	28.0498	0.6139	0.8636	33.2366	2.0955	195.4224
$L_{\text{MSE}} + L_{\text{GAN}} + L_{\text{edg}} + L_{\text{vgg}}$	**29.5970**	**0.7205**	0.4953	**38.9112**	1.5851	**729.6187**
GT	—	1	0.8980	33.1317	1.9396	183.7867

Note: The bold values denote the best values that were obtained for a certain metric among all those de-speckling schemes been compared.

5 Conclusion

In summary, MDGAN is proposed for speckle reduction in OCT images. With a CMSM being employed to utilize the multiscale context and a SAM being utilized to refine the denoised images, the proposed MDGAN scheme is effective and robust in OCT speckle reductions. Extensive experiments with two different OCT image datasets are conducted to validate the effectiveness of MDGAN. Results show that MDGAN is comparable to the best existing state-of-the-art methods in terms of both visual effect and quantitative metrics. However, as the fully supervised learning scheme is adopted in MDGAN, and a limited number of clean images must be cropped to generate sufficient training images, the overall architecture of MDGAN is a bit complex. In future work, our objective is to reduce the complexity of the network architecture by employing self-supervised deep-learning schemes for OCT image speckle reductions.

Disclosures

The authors declare no potential conflicts of interests.

Acknowledgments

This work was supported in part by the Key Research and Development Program of Shaanxi (Grant No. 2021SF-342), the Guangdong Basic and Applied Basic Research Foundation (Grant No. 2021B1515120013), the National Natural Science Foundation of China (Grant No. 61705184), the Key Research Project of Shaanxi Higher Education Teaching Reform (Grant No. 21BG005).

Data, Materials, and Code Availability

The dataset used in the experiment can be obtained from Ref. 36, and the relevant code of the experiment is published at https://github.com/GE-123-cpu/MDGAN.

References

1. D. Huang et al., "Optical coherence tomography," *Science* **254**(5035), 1178–1181 (1991).
2. W. Drexler and J. G. Fujimoto, "State-of-the-art retinal optical coherence tomography," *Prog. Retinal Eye Res.* **27**(1), 45–88 (2008).
3. A. V. D'Amico et al., "Optical coherence tomography as a method for identifying Benign and malignant microscopic structures in the prostate gland," *Urology* **55**(5), 783–787 (2000).
4. A. Desjardins et al., "Speckle reduction in oct using massively-parallel detection and frequency-domain ranging," *Opt. Express* **14**(11), 4736–4745 (2006).
5. J. Xu et al., "Wavelet domain compounding for speckle reduction in optical coherence tomography," *J. Biomed. Opt.* **18**(9), 096002 (2013).
6. M. Wang et al., "Upconversion nanoparticle powered microneedle patches for transdermal delivery of siRNA," *Adv. Healthcare Mater.* **9**(2), 1900635 (2020).
7. D. Cui et al., "Multifiber angular compounding optical coherence tomography for speckle reduction," *Opt. Lett.* **42**(1), 125–128 (2017).
8. M. Pircher et al., "Speckle reduction in optical coherence tomography by frequency compounding," *J. Biomed. Opt.* **8**(3), 565–569 (2003).
9. M. Pircher et al., "Three dimensional polarization sensitive OCT of human skin in vivo," *Opt. Express* **12**(14), 3236–3244 (2004).
10. G. R. Wilkins, O. M. Houghton, and A. L. Oldenburg, "Automated segmentation of intraretinal cystoid fluid in optical coherence tomography," *IEEE Trans. Biomed. Eng.* **59**(4), 1109–1114 (2012).

11. A. Buades, B. Coll, and J.-M. Morel, "A non-local algorithm for image denoising," in *IEEE Comput. Soc. Conf. Comput. Vision and Pattern Recognit.*, Vol. 2, pp. 60–65 (2005).
12. P. Sudeep et al., "Enhancement and bias removal of optical coherence tomography images: an iterative approach with adaptive bilateral filtering," *Comput. Biol. Med.* **71**, 97–107 (2016).
13. M. Elad and M. Aharon, "Image denoising via sparse and redundant representations over learned dictionaries," *IEEE Trans. Image Process.* **15**(12), 3736–3745 (2006).
14. D. Zoran and Y. Weiss, "From learning models of natural image patches to whole image restoration," in *Int. Conf. Comput. Vision*, IEEE, pp. 479–486 (2011).
15. X. Sang et al., "Speckle noise reduction mechanism based on dual-density dual-tree complex wavelet in optical coherence tomography," in *IEEE 5th Optoelectron. Global Conf. (OGC)*, IEEE, pp. 190–192 (2020).
16. S. Gu et al., "Weighted nuclear norm minimization with application to image denoising," in *Proc. IEEE Conf. Comput. Vision and Pattern Recognit.*, pp. 2862–2869 (2014).
17. A. Abbasi et al., "Optical coherence tomography retinal image reconstruction via nonlocal weighted sparse representation," *J. Biomed. Opt.* **23**(3), 036001 (2018).
18. K. Dabov et al., "Image restoration by sparse 3D transform-domain collaborative filtering," *Proc. SPIE* **6812**, 681207 (2008).
19. M. Tajmirriahi et al., "A lightweight mimic convolutional auto-encoder for denoising retinal optical coherence tomography images," *IEEE Trans. Instrum. Meas.* **70**, 1–8 (2021).
20. B. Anoop et al., "A cascaded convolutional neural network architecture for despeckling OCT images," *Biomed. Signal Process. Control* **66**, 102463 (2021).
21. J. J. Rico-Jimenez et al., "Real-time OCT image denoising using a self-fusion neural network," *Biomed. Opt. Express* **13**(3), 1398–1409 (2022).
22. M. Wang et al., "Semi-supervised capsule cGAN for speckle noise reduction in retinal OCT images," *IEEE Trans. Med. Imaging* **40**(4), 1168–1183 (2021).
23. B. Qiu et al., "N2NSR-OCT: simultaneous denoising and super-resolution in optical coherence tomography images using semisupervised deep learning," *J. Biophotonics* **14**(1), e202000282 (2021).
24. G. Ni et al., "SM-Net OCT: a deep-learning-based speckle-modulating optical coherence tomography," *Opt. Express* **29**(16), 25511–25523 (2021).
25. X. Yu et al., "A generative adversarial network with multi-scale convolution and dilated convolution res-network for OCT retinal image despeckling," *Biomed. Signal Process. Control* **80**, 104231 (2023).
26. Y. Huang et al., "Noise-powered disentangled representation for unsupervised speckle reduction of optical coherence tomography images," *IEEE Trans. Med. Imaging* **40**(10), 2600–2614 (2020).
27. O. Ronneberger, P. Fischer, and T. Brox, "U-Net: convolutional networks for biomedical image segmentation," *Lect. Notes Comput. Sci.* **9351**, 234–241 (2015).
28. S. Ioffe and C. Szegedy, "Batch normalization: accelerating deep network training by reducing internal covariate shift," in *Int. Conf. Mach. Learn.*, pp. 448–456, PMLR (2015).
29. A. L. Maas, A. Y. Hannun, and A. Y. Ng, "Rectifier nonlinearities improve neural network acoustic models," in *Proc. ICML*, Citeseer Vol. **30**, p. 3 (2013).
30. P. Isola et al., "Image-to-image translation with conditional adversarial networks," in *Proc. IEEE Conf. Comput. Vision and Pattern Recognit.*, pp. 1125–1134 (2017).
31. K. Zhang et al., "Beyond a Gaussian denoiser: residual learning of deep CNN for image denoising," *IEEE Trans. Image Process.* **26**(7), 3142–3155 (2017).
32. Y. Hu et al., "Single image super-resolution via cascaded multi-scale cross network," arXiv:1802.08808 (2018).
33. S. Woo et al., "CBAM: convolutional block attention module," *Lect. Notes Comput. Sci.* **11211**, 3–19 (2018).
34. M. Haris, G. Shakhnarovich, and N. Ukita, "Deep back-projection networks for super-resolution," in *Proc. IEEE Conf. Comput. Vision and Pattern Recognit.*, pp. 1664–1673 (2018).

35. X. Wang et al., "ESRGAN: enhanced super-resolution generative adversarial networks," *Lect. Notes Comput. Sci.* **11133**, 63–79 (2018).
36. L. Fang et al., "Fast acquisition and reconstruction of optical coherence tomography images via sparse representation," *IEEE Trans. Med. Imaging* **32**(11), 2034–2049 (2013).
37. J. F. De Boer et al., "Improved signal-to-noise ratio in spectral-domain compared with time-domain optical coherence tomography," *Opt. Lett.* **28**(21), 2067–2069 (2003).
38. J. T. Bushberg and J. M. Boone, *The Essential Physics of Medical Imaging*, Lippincott Williams & Wilkins (2011).

Xiaojun Yu received his PhD from Nanyang Technological University, Singapore, in 2015, where he was a postdoctoral research fellow from 2015 to 2017. He is currently an associate professor at the Northwestern Polytechnical University, China. His main research interests include high-resolution optical coherence tomography (OCT) and its imaging applications.

Chenkun Ge received his bachelor's and master's degrees from Fuzhou University and the Northwest Polytechnic University, respectively. He is currently pursuing his doctoral degree from the same university. His research interests include OCT imaging and OCT image processing.

Mingshuai Li received his BS degree from Xi'an University of Technology, China. He is currently working toward his MEng degree from the School of Automation, Northwestern Polytechnical University, Xi'an Shaanxi, China. His research interests include OCT imaging and OCT image processing.

Muhammad Zulkifal Aziz his BS degree from the Government College University, Lahore, Pakistan, and his master's degree from the Northwestern Polytechnical University, Xi'an, China. He is currently working toward his PhD and his main research interests include machine learning, brain computer interface, and EEG signal processing. He has won the best conference paper award at IEEE ICICN 2021 and is also the recipient of outstanding postgraduate student and outstanding thesis awards at the same university.

Jianhua Mo received his BEng and his MEng degrees from Zhejiang University, China, and his PhD from the National University of Singapore in 2011, all in engineering. He did his postdoctoral training at VU University Amsterdam before joining Soochow University, Suzhou, China, in 2014. He is currently an associate professor at Soochow University. His research interests include optical imaging and spectroscopy techniques for biomedical applications and nondestructive inspection in industry.

Zeming Fan received his PhD from Xi'an University of Technology, China in 2002. After three years, working as a postdoctoral fellow at Yongji Motor Factory, he joined Northwestern Polytechnical University as an associate professor. He has been a full professor at the same university since 2018. His main research interests include artificial intelligence and robotic control.