



DN-GAN: Denoising generative adversarial networks for speckle noise reduction in optical coherence tomography images

Zailiang Chen^{a,b}, Ziyang Zeng^{a,b}, Hailan Shen^{a,*}, Xianxian Zheng^{a,b}, Peishan Dai^{a,b}, Pingbo Ouyang^{a,c}

^a School of Computer Science and Engineering, Central South University, Hunan, Changsha, 410083, China

^b Hunan Engineering Research Center of Machine Vision and Intelligent Medicine, Central South University, Changsha, 410083, China

^c The Second Xiangya Hospital of Central South University, Changsha, 410011, China

ARTICLE INFO

Article history:

Received 10 November 2018

Received in revised form 15 July 2019

Accepted 22 July 2019

Available online 29 August 2019

Keywords:

Denoising generative adversarial networks (DN-GAN)

Speckle reduction

Detail preservation

Optical coherence tomography (OCT)

Layer segmentation

ABSTRACT

Optical coherence tomography (OCT) is an efficient noninvasive bioimaging technique that can measure retinal tissue. Considering the changes in the acquisition environment during imaging, the OCT images are affected by granular speckle noise, thereby reducing the image quality. In this paper, an efficient method based on generative adversarial network is proposed to reduce the speckle noise and preserve the texture details. The proposed model consists of two components, that is, a denoising generator and a discriminator. The denoising generator learns how to map the noise image to the ground truth. The discriminator learns as a loss function to compare the differences between the ground truth and the image reconstructed by the generator. A number of repeated densely sampled B-scan OCT images are used with multi-frame registration to train the denoising generator. The original OCT images are denoised by a trained generator to quickly and efficiently obtain improved quality. Results showed that the proposed method outperforms the other popular methods, and achieves a better denoising effectiveness.

© 2019 Elsevier Ltd. All rights reserved.

1. Introduction

Optical coherence tomography (OCT) is a noninvasive imaging technology with many applications in the field of ophthalmology for the quantitative analysis of eye tissues [1]. OCT helps doctors in diagnosing ophthalmic diseases, such as glaucoma, age-related macular degeneration (AMD), diabetic macular edema (DME), and anemia [2–5]. OCT is a high-resolution imaging technology that is susceptible to the acquisition environment. Imaging light is multiply scattered and coherently superimposed by the intraocular tissue that can generate the speckle noise, which is a common problem in OCT images [6]. Removing speckle noise is challenging due to three factors. First, speckle noise is a multiplicative noise instead of the additive white Gaussian noise [7]. Estimating the noise distribution by statistical models is difficult. Second, when imaging different tissues of the eye, the intensity of speckle noise is also different. Finally, speckle noise not only appears in the background

but also exists in the retina, thereby severely degrading the OCT imaging quality for clinical diagnosis and analysis.

The methods proposed to reduce the speckle noise can mainly be divided into two categories: hardware-based methods and post-processing methods. Hardware-based methods include spatial [8] and angle averaging [9] methods. The cost and acquisition time are inevitably increased. Post-processing methods are flexible and easy to implement. Early post-processing methods use smoothing filters to reduce speckle noise in OCT images, such as averaging and median filters [10], adaptive Wiener [11], and complex diffusion denoising filters [12]. Wavelet transform-based methods are also effective means to reduce noise [13,14]. Although these filter-based methods reduce the speckle noise to a certain extent, they also cause a loss in image detail. Fang et al. [15] proposed a dictionary learning method by using sparse decomposition and high peak-signal-to-noise ratio (PSNR) image to remove the speckle noise. Another improved sparse decomposition method [16] does not need prior high-quality image, and therefore the time of training the dictionary is reduced. Baghaie et al. [17] utilized sparse and low-rank decomposition to denoise and align the multiple frames simultaneously. In summary, the sparse decomposition has prospective applications in denoising, but it requires additional time to process a single image. Another common method uses sta-

* Corresponding author.

E-mail addresses: xxyczl@csu.edu.cn (Z. Chen), zengziyang@csu.edu.cn (Z. Zeng), hn.shl@126.com (H. Shen), xxzheng@csu.edu.cn (X. Zheng), daipeishan@163.com (P. Dai), ouyangpingbo@csu.edu.cn (P. Ouyang).

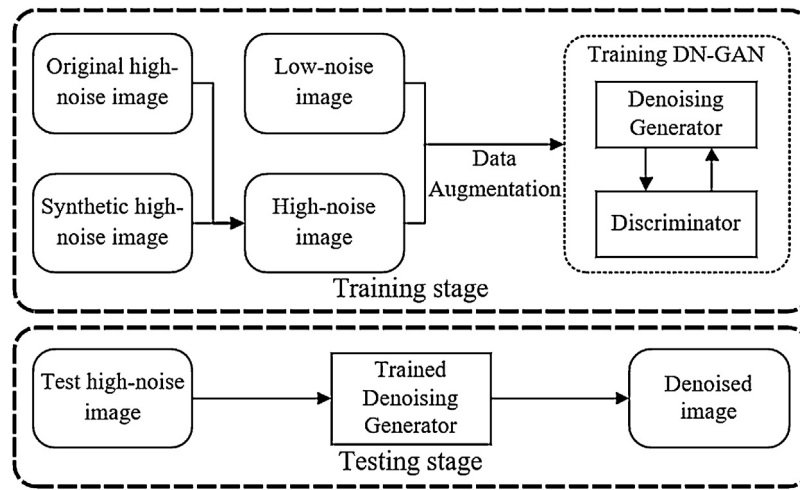


Fig. 1. Flowchart of the proposed speckle noise reduction method.

tistical models to estimate speckle noise distribution. Jesus et al. [18] used a generalized Gamma distribution to set the probability model. Li et al. [19] combined the empirical speckle statistics with the Huber variant of the general variation regularization to set up the statistical model. These methods achieve good denoising performance, but the major drawback of statistics-based methods is that there are many tunable parameters needed to be defined, and the fact that their performances rely on the models with assumptions. Block-matching 3D (BM3D) [20] and K-singular value decomposition (K-SVD) [21] are widely used in noise reduction. These algorithms do not require reference images and tend to have the best convenience in practice. However, they both require a long time to deal with considerable amount of noise image, and prone to flake noise in the region with high noise intensity.

Deep learning method based on generative adversarial network (GAN) [22] has shown advanced performance in many fields. GAN consists of a generator and a discriminator. The adversarial training between the two models provides the generator imagination ability to output realistic images. In the image restoration problem, super-resolution GAN (SRGAN) [23] achieves the best performance in the natural image by using the adversarial and pixel-wise losses to recover the detailed information for 4x upscaling factors. SRGAN-based networks are also used in medical image. Chen et al. [24] utilized 3D densely connected super-resolution network (DCSRN) as a generator to restore the details of brain magnetic resonance images. Yang et al. [25] used U-net as a generator for dealiasing magnetic resonance images and improving the definition. The target of super-resolution, dealiasing, and denoising problems is to recover the detail of the image as much as possible. But these methods are ineffective in denoising tasks in OCT images due to the huge difference in different types of images.

In this study, a denoising GAN (DN-GAN) model is proposed for reducing the speckle noise of OCT image. Denoising generator is guided by the discriminator to learn the mapping between the noised image and ground truth, which can not only remove the speckle noise in the background and retina but also preserve the edges and details to enhance the image quality. The contributions of our study are following:

- The adversarial training is used to automatically extract features of low-noise (LN) OCT images, and encode them into the manifold space to reduce the perceptual difference between high-noise (HN) images and LN images, thereby outputting clear images with accurate and realistic details by a denoising generator.

- The frequency loss is first added to the loss function to suppress the noise in the high-frequency region and preserve the structural information in the low frequency-region, thereby making this network suitable for this problem.
- A new denoising framework, that is, DN-GAN, with an efficient generator and few parameters is designed. The generator is improved by adding the context-encoding module to enhance the features that are beneficial for speckle reduction. The performance of DN-GAN surpasses those of the popular networks used for image reconstruction.

2. Method

To denoise the HN image automatically, the proposed DN-GAN learns the mapping relationship between HN and LN images in the end-to-end process. DN-GAN consists of a denoising generator and a discriminator. The HN images are reconstructed by generator that makes the prediction image close to the LN image which is indistinguishable for discriminator. The goal of the discriminator is to identify the difference as much as possible between the two types of images. With iterative adversarial training, the quality of prediction image is gradually being sufficiently realistic and noise-free. The overall flowchart of our proposed method is shown in Fig. 1.

2.1. Background

A denoising problem can be formulated as a prediction process to transform HN images to LN images. Supposed that the HN image we collected is X , and the ideal LN image is Y . Speckle noise Z is considered as multiplicative noise, and this relationship can be expressed as the following Eq. (1):

$$X = Y \cdot Z \quad (1)$$

With regard to a statistical model, the goal is to estimate the noise distribution Z and obtain Y by an inverse operation. However, if we can obtain Y by a pre-processing method, DN-GAN can automatically extract the visual features of X , and find an inverse mapping function f that presents as $\hat{Y} = f(X)$. Hence, prior information, such as noise distribution, is unnecessary. Then, the image \hat{Y} recovered by any algorithm cannot be equal to Y , and the gap R can be presented as follows:

$$R = \hat{Y} - Y = u - v \quad (2)$$

where u is the residual noise in \hat{Y} , and v represents the unrecovered detail. The goal for DN-GAN is to minimize the gap R . Therefore, a

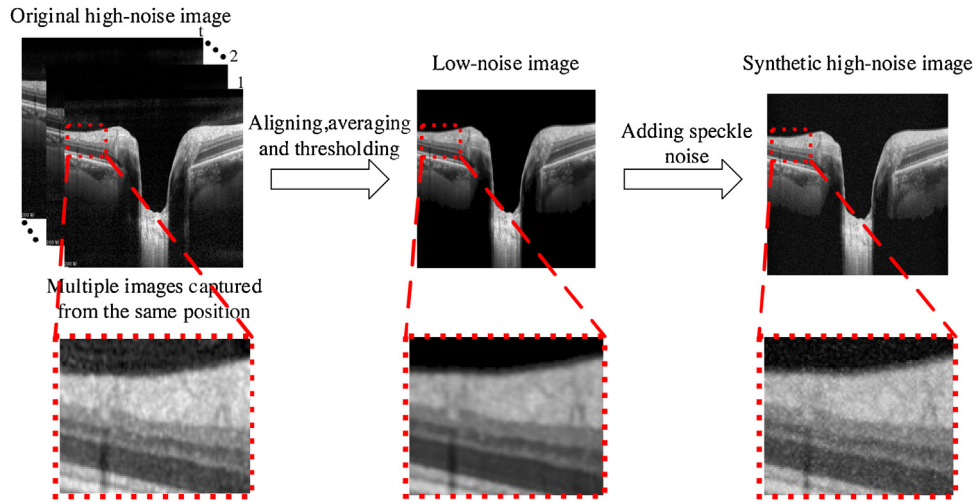


Fig. 2. Illustration of the pre-processing step for training data.

clean image Y can be obtained from \hat{Y} by subtracting u and adding v .

2.2. Pre-processing for training data

High-quality LN images are considered as ground truth in our study. The enhanced depth imaging OCT (EDI-OCT) possesses a deeper tissue than the spectral domain OCT (SD-OCT). By contrast, the OCT images of important fundus tissues include the optic nerve head (ONH) and the macula. The deep retinal tissue of ONH image, such as lamina cribrosa, is a significant factor for early glaucoma detection. On the contrary, the macula OCT image lacks such tissue. Thus, the EDI-OCT images of the ONH are used as training data to strengthen the ability of the model to process images with different tissues. Then, a scheme of the pre-processing method is shown in Fig. 2.

The detailed steps of pre-processing are described as follows:

- 1) Multiple images-captured from the same position. To get ideal LN images, a customized scanning pattern is employed with a number of repeated densely B-scans captured from an extremely close positions in the spatial domain. Given the randomness of the speckle scattering direction and eye movement, certain differences between single-frame images, such as the contrast imbalance or the difference in noise intensity, will occur.
- 2) Aligning, averaging, and thresholding. The differences as described above are used to align and average all the single-frame images according to the registration method [26]. A spline pyramid hierarchy is utilized to provide a consistent interpolation and filtration model for applying transformations during registrations, and then generate the final aligned OCT image, so that the speckle noise is suppressed considerably, and the contrast imbalance caused by imaging light attenuation is reduced. Then the pixel intensity of the retinal nerve fibre layer (RNFL) and choroid is used as a threshold to remove the residual noise and dissociative molecules in the background.
- 3) Adding speckle noise. For the HN image, to simulate the randomness of real noise, and enhance the robustness of DN-GAN to process images with different noise intensities, inspired by the methods [27], a noise model is used for the LN image to synthesize new HN images. The synthetic noise is computed by creating a random matrix with the same size and density probability distribution of the original noise. Then, low-pass Gaussian filters and a multi-scale average pooling operation are used to smooth this matrix, which can enhance the random-

ness of synthetic noise compared with original noise. Finally, a sliding window with overlapping stride crops the LN, HN, and synthetic HN images into small patches. Conventional data augmentation process (e.g., image flipping, rotating, and shifting) is also applied to these patches.

The speckle noise added to the LN image must be similar to original noise. Given that the synthetic method [27] is proposed for the SD-OCT image, the parameters are adjusted such that the noise model fits EDI-OCT image. The details and probability density functions of the synthetic and original noises are shown in Fig. 3.

2.3. Training a DN-GAN

In general, the supervised learning reconstruction methods use absolute error (L1 loss) or mean square error (L2 loss) to reduce the pixel-wise difference between the generated image and ground truth that it may overly smooth the image and lack of coherent detail. Thus, the high texture detail cannot be effectively reconstructed. Adversarial training provides a perceptual-based loss function that can measure the difference in data distribution of the manifold, thereby making the generated image close to the real image with an improved visual effect [23].

As described above, the adversarial network consists of two components. With regard to denoising generator, DenseNet possesses a good performance in image classification [28] by reusing the densely skip connection and feature map of the convolutional layer, thus obtaining the best accuracy with few parameters. Our DN-GAN model is inspired by some previous studies [24,28] to use DenseNet as a denoising generator. The proposed generator consists of four DenseBlocks. Each DenseBlock contains four convolution layers and three elementwise layers that connects the output of each convolutional layer to the output of all previous layers; the four DenseBlocks follow the same principle. The difference is that a reduction block is added after each elementwise layer to fuse the feature map, thereby resulting in the same input dimension for different DenseBlocks. Inspired by the trick of reusing the features and the method [29], the context encoding block is added after each reduction block, therefore increasing the weight of useful feature maps. By comparison, the feature maps with less importance are resulted in low weight, thus making the input of DenseBlock reasonable. Then, each convolutional layer of the generator is followed by a BatchNorm layer and a Relu activation function.

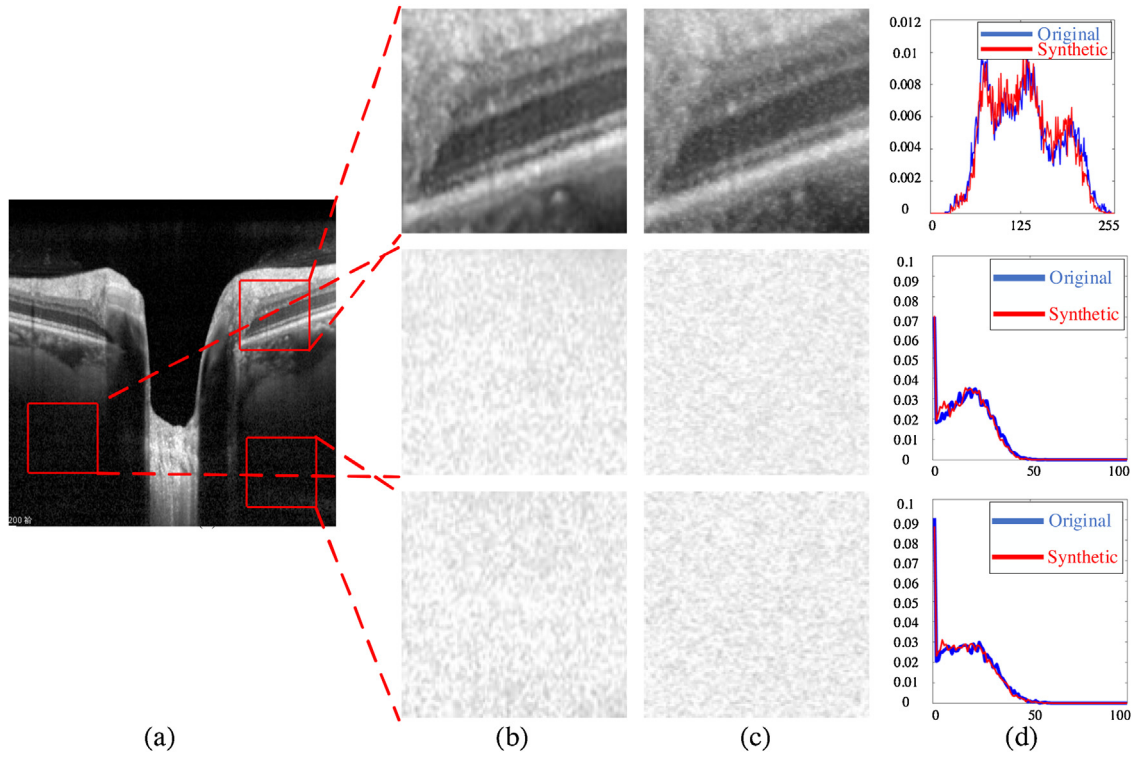


Fig. 3. The details of the patches of original and synthetic noise. (a) Original HN image. (b) Selection of noise patches from the Original HN image. (c) Synthetic noise patches based on original noise patches. (d) Comparison of probability density functions between synthetic patches and original patches. The horizontal axis represents the probability, and the vertical one indicates the grayscale. To get a clear visual effect of original and synthetic noise, the color of background region is reversed.

The standard CNN classification model is used as the discriminator following SRGAN [30], which contains eight convolutional layers, with the BatchNorm layer and the Leaky Relu activation function. The last two layers are dense layers, and the activation function is Relu and Sigmoid, respectively. Fig. 4 shows the overall framework of our DN-GAN architecture.

2.4. Loss function design

To improve the perceptual quality of reconstructed images, the perceptual and content losses are designed for training the denoising generator. The content loss consists of two parts, namely, pixel-wise and frequency domain L1 losses. The two loss types can be expressed as follows:

$$\min_{\theta_G} \text{loss}_{\text{spatial}}(\theta_G) = \frac{\sum_{y=1}^H \sum_{x=1}^W |I_{x,y}^T - I_{x,y}^G|}{HW} \quad (3)$$

$$\min_{\theta_G} \text{loss}_{\text{frequency}}(\theta_G) = \frac{\sum_{y=1}^H \sum_{x=1}^W |f_{x,y}^T - f_{x,y}^G|}{HW} \quad (4)$$

where H and W denote the height and width of the image; x and y are the pixel coordinates; T and G indicate the target LN and generated patches, respectively; I and f are the spatial and frequency domains of the patches of the image. On the other hand, the perceptual loss named loss_{adv} calculated by the discriminator is used to measure the difference between the data distributions [22], which can be expressed as:

$$\min_{\theta_G} \text{loss}_{\text{adv}}(\theta_G) = -\log(D_{\theta_D}(G_{\theta_G}(x_{HR}))) \quad (5)$$

where θ_G and θ_D denote the weights and biases of generator and discriminator respectively, and x_{HR} is the patch of noise image. The

final loss function of our proposed method is given by Eq. (5), as shown below:

$$\text{loss} = \alpha \text{loss}_{\text{spatial}} + \beta \text{loss}_{\text{frequency}} + \gamma \text{loss}_{\text{adv}} \quad (6)$$

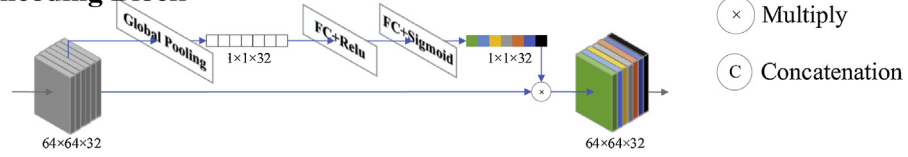
where α , β , and γ indicate the weights for balancing the different losses. In this study, the weighted sum of the three losses above is used as a loss function for DN-GAN. In general, the noise is considered in the high-frequency region of the image, which is relatively incoherent with useless information. Almost speckle noise is located in the noncentral region of the frequency domain. Therefore, the loss function is improved by adding the frequency loss to measure the difference in a frequency domain between the LN and generated images as an additional constraint to the $\text{loss}_{\text{spatial}}$.

In the field of image restoration, the high-level perceptual feature extracted from the VGG network [32] pretrained on ImageNet dataset [33] is also a part of the perceptual loss [23]. It can be used to reduce the euclidean distance between the feature representations of a generated image and a reference image in a high-dimensional feature space. However, in this study, OCT and natural images are significantly different. The features extracted by VGG will not be representative and may increase the training cost. Thus, VGG loss is disregarded. By contrast, we find that L1 loss is preferable than L2 loss in our study.

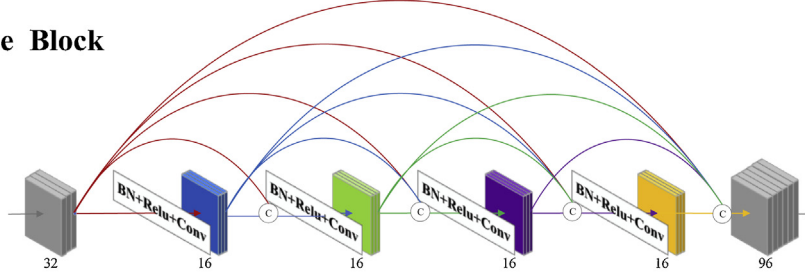
2.5. Networks and training settings

The models are implemented in Tensorlayer on a server with NVIDIA GTX Titan XP GPU. All the convolutional layer sizes of DN-GAN are 3×3 , and the number of filters is 16. The feature maps output by the reduction block are reduced to 32. Based on the magnitude of three losses, the values of the hyperparameters α , β , and γ are predefined to maintain a balance among the different losses. For the network, ADAM optimizer with a learning rate 10^{-4} and a

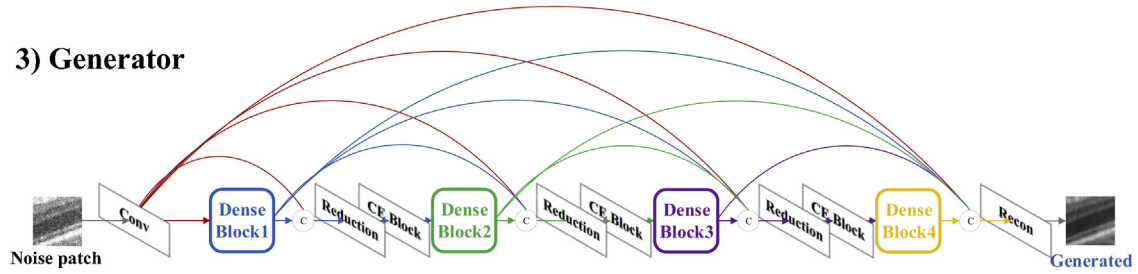
1) Context Encoding Block



2) Dense Block



3) Generator



4) Discriminator

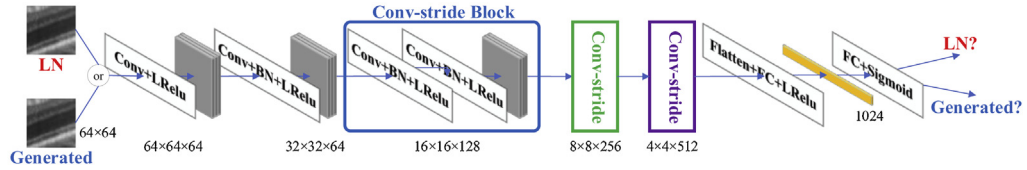


Fig. 4. The architecture of Context Encoding Block, DenseBlock and DN-GAN. The CE Block in Generator is Context Encoding Block. The Reduction before Context Encoding Block compress the input feature map number to 32 by a 1×1 convolution. The final Reconstruction layer is 1×1 convolution with Tanh activation. The discriminator is standard CNN architecture with 8 convolutional layers.

10^{-1} decay factor for every 10 epochs is utilized to minimize the total loss with a batch size of 16.

3. Experiments

3.1. Datasets

The OCT datasets are collected from the device Heidelberg's Spectralis SD-OCT from Germany in the Second Xiangya Hospital of the Central South University. The image resolution is 768 pixels \times 496 pixels. The datasets are collected on the ONH of 36 samples, in which each sample possessed more than 50 single-frame images. A total of 36 datasets are obtained. Each dataset includes three image types, namely, LN, synthetic HN, and original HN images, consisting of 20 original and 20 synthetic HN images. The 36 datasets are randomly divided into two parts, that is, 25 datasets (70%) for training and 11 datasets (30%) for test. Then, the training datasets are cropped into small patches with size 64 pixels \times 64 pixels.

The training data is preprocessed separately. Given that our generator is a fully convolutional network, test image is directly

reconstructed by the generator without cropping into patches. Its size is set to 480 pixels \times 480 pixels by cropping the redundant background and useless region, which is convenient for testing various methods. On the other hand, to test the generalization of DN-GAN, a small amount of macula OCT images with size 630 pixels \times 496 pixels are collected that are lacking in our dataset. The boundaries of retinal layers are labeled by ophthalmologist.

3.2. DN-GAN variations and comparison methods

To test the effectiveness of frequency loss, training data with synthetic noise, and the generator with context encoding block, the following DN-GAN variations are compared: (1) pixel perceptual (case 1): the model with pixel-wise L1 and perceptual losses; (2) pixel-frequency-perceptual-synthetic noise-context encoding block (case 2): the model with pixel-wise L1 loss, frequency domain loss, perceptual loss, synthetic noise, and context encoding block; (3) pixel-perceptual-synthetic noise (case 3): the model without frequency domain loss and context encoding block; (4) pixel-perceptual-context encoding block (case 4): the model without frequency domain and synthetic noise; and (5) pixel-perceptual-

Table 1

The quantitative evaluation of SSIM, PSNR and FBE for different variations in DN-GAN.

	Case 1 PSNR↑/SSIM↑/FBE↓	Case 2 PSNR↑/SSIM↑/FBE↓	Case 3 PSNR↑/SSIM↑/FBE↓	Case 4 PSNR↑/SSIM↑/FBE↓	Case 5 PSNR↑/SSIM↑/FBE↓
Validation set#1	26.4140/0.8756/7.6312	27.3052/0.8873/4.8579	27.2425/0.8891/5.2648	27.4327/0.8911/4.3585	27.8131/0.9002/3.6392
Validation set#2	26.7742/0.8633/7.4956	27.3812/0.8892/4.7413	27.0348/0.8750/5.3972	27.2643/0.8850/4.5101	28.1651/0.9015/3.6212
Mean	26.5941/0.8695/7.5634	27.3432/0.8889/4.7996	27.1387/0.8821/5.3310	27.3485/0.8881/4.4343	27.9891/0.9009/3.6302

Table 2

The quantitative evaluation of SSIM, PSNR and FBE for different generators tested in different datasets. Running Time is the time cost required by the model for a batch. Validation set#1 and Validation set#2 consist of original HN image and synthetic noise image, respectively.

Methods	Validation set#1 PSNR↑/SSIM↑/FBE↓	Validation set#2 PSNR↑/SSIM↑/FBE↓	Parameters (Generator)	Running Time(s)
SRResNet [23]	27.1099 / 0.8766 / 3.7871	27.0009 / 0.8745 / 3.8461	0.4090 M	0.2633
DCSRN [24]	27.1015 / 0.8741 / 4.2102	27.0146 / 0.8767 / 4.3412	0.1522 M	0.2661
GAN-UNet [25]	27.1302 / 0.8800 / 6.6421	26.6802 / 0.8808 / 6.7124	0.2613 M	0.2976
GAN-SRResNet [23]	27.6341 / 0.8901 / 3.7198	27.6446 / 0.9014 / 3.6958	0.4090 M	0.3156
DN-GAN	27.8131 / 0.9002 / 3.6393	28.1651 / 0.9015 / 3.6125	0.1522 M	0.3039

frequency (case 5): the model without synthetic noise and context encoding block.

At the same time, our model is also compared with different non-GAN and GAN models, such as SRResNet [23], GAN-SRResNet [23], DCSRN [24], and GAN-UNet [25]. The last two upsampling layers of SRResNet and GAN-SRResNet are removed, and five residual blocks are set to approximate the DN-GAN parameters. DCSRN also adds a context encoding block after the reduction block. GAN-UNet possesses five convolution layers to fit the size of patches from our training data. For all GAN models, the training data consists of original and synthetic noise data, and the loss function is consistent with DN-GAN. Non-GAN models are training without the discriminator and perceptual loss. Finally, DN-GAN is compared with the traditional OCT denoising methods, including MSBTD [15], SBSDI [16], BM3D [20], K-SVD [21], and Tikhonov [30].

3.3. Results

3.3.1. Quantitative measures of performance

In general, peak signal to noise ratio (PSNR) and structural similarity index (SSIM) [23] are used to compare the performance of denoising method by measuring the similarity between denoised image and reference image. Moreover, as stated above, the image with speckle noise has significant difference compared with noise-free images in frequency domain. To depict the mean square difference between the reference image and denoised image in frequency domain, we propose a metric named frequency-based error (FBE) which can be represented by Eq. (4). In addition, contrast-to-noise ratio (CNR) are employed for comparison and analysis in regions of interest (ROI), which are widely used in OCT denoising methods [11,16]. CNR represents the noise level of ROI compared with background region, which is defined as follows:

$$CNR = \frac{\mu_R - \mu_B}{\sqrt{\sigma_R^2 + \sigma_B^2}} \quad (7)$$

where μ_R and σ_R are mean and standard deviation of ROI (e.g., red box #1-3 in Fig. 7); μ_B and σ_B are mean and standard deviation of the background region (e.g., blue box #4 in Fig. 7), respectively.

Given that most regions of the OCT images are in the background, the traditional method cannot completely remove the background noise. Thus, the lamina cribrosa and structure of retina are selected manually as ROI to measure the performance of different methods.

3.3.2. DN-GAN variations

Table 1 tabulates the quantitative comparison results of DN-GAN variations. Validation set#1 and Validation set#2 consist of original HN image and synthetic noise image, respectively.

Table 1 shows the effects of the different experimental schemes on PSNR, SSIM and FBE. The result demonstrates that the training set with synthetic noise enhances the diversity of the data and the robustness of the model. By contrast, the context encoding block enhances the useful information by reusing the feature map and suppressing ineffective ones to recover detailed information. Frequency domain loss preserves the low-frequency information and suppresses speckle noise in high-frequency regions. The denoising performance is improved in case 5 comparing to the other experimental schemes.

3.3.3. Comparison methods based on the neural network

The qualitative analyses of the GAN and non-GAN models on PSNR, SSIM and FBE are shown in Table 2. The validation sets are the same, as shown in Table 1. As described in Section 3.2, the structure of these models is adjusted. GAN models trained the generator and discriminator at the same time. Thus, the running time of GAN model is longer than that of the non-GAN model. Our DN-GAN model with the minimum parameters achieves the best performance among the tested models. However, DN-GAN has a deeper structure than the other models, and the training time is slightly longer than GAN-UNet.

The denoising results by different models are shown in Fig. 5. The second row is the details of retina, and the third row demonstrates the lamina cribrosa. For non-GAN models, SRResNet and DCSRN only train a generator without a discriminator. Thus, the details of the generated image are not effectively restored due to lack of perceptual loss, and the running time is relatively shorter comparing to GAN-based model. The edges of the retina are unclear; in particular, the pore paths of the lamina cribrosa are blurred. Although GAN-UNet can restore the details effectively, much residual noise remained in the retina and background. GAN-SRRes Net has a similar performance to DN-GAN, but the intensity and contrast of the image are low that is similar to SRResNet. As indicated by the red arrow in Fig. 5, there is an anomalous region. Due to the residual connection by addition of the last convolution layer and input, these two models cannot recover the details of pixel with high intensity. In contrast, DN-GAN not only enhance the contrast between the retinal layers but also preserves the details of lamina cribrosa. It means that DN-GAN can get better performance with less additional time cost and less parameters.

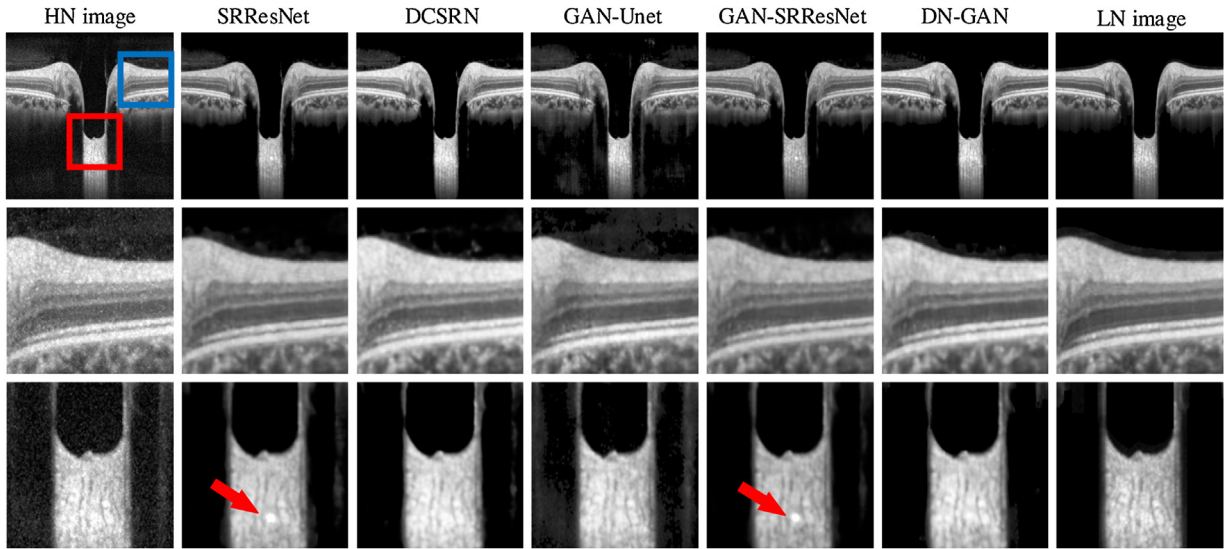


Fig. 5. Original image (HN image) and reconstruction results by the SRResNet, DCSRNet, GAN-Unet, GAN-SRResNet, DN-GAN, and the reference image (LN image). The middle row is the detail of retina in blue box and the bottom row is the detail of lamina cribrosa in red box. Best viewed with color and zoom-in.

Table 3

The quantitative evaluation of SSIM, PSNR, FEB and CNR for different methods calculated in ROI of test OCT images.

Methods	PSNR↑(dB)	SSIM↑	FEB↓	CNR↑(dB)	Test time↓(s)
Origin	22.8245	0.7057	4.1536	2.0716	–
K-SVD [21]	23.6568	0.8178	1.2653	2.2551	38.80
BM3D [20]	23.4504	0.8158	1.4869	2.2336	2.51
Tikhonov [30]	23.6865	0.8675	1.3727	2.2504	1.65
MSBTD [15]	24.0533	0.8699	1.7097	2.2010	337.83
SBSDI [16]	24.6939	0.7732	1.1778	2.2580	26.36
DN-GAN	25.4566	0.8808	1.1626	2.4719	0.05

3.3.4. Comparison methods based on the traditional methods

The performance of different methods is tested on the original HN image, as shown in Fig. 6. According to the details displayed by the red and blue boxes, it can easily be observed that K-SVD and BM3D suppress the noise effectively throughout the entire image, but dark spots are left in the area with high-intensity noise. Tikhonov retains complete details and texture information, but the denoising effect is insignificant. SBSDI can preserve the edge information effectively, but structures with additional details are removed due to the overcompensation. MSBTD also performs well in low-intensity areas but leads to ambiguous details. However, DN-GAN achieves a balanced performance in denoising and detail preservation comparing to other methods due to its imagination ability.

The quantitative evaluation of the performance of these methods is shown in Table 3. Given that the traditional methods process directly on images instead of regenerating an image, the background noise cannot be completely removed. Thus, the PSNR, SSIM, FEB and CNR are calculated in the ROI without the background. The ROI are shown in Fig. 7, where the lamina cribrosa and retinal area are on the both sides of the fovea. Quantitative evaluation is averaged for three ROI (e.g., red box #1-3 in Fig. 7). The background region (e.g., blue box #4 in Fig. 7) is utilized to calculate CNR. As a result, in Table 3, our proposed method is better than the other methods, and the test time is the shortest that it is more suitable for practical applications.

3.3.5. Visual analysis of retinal data

One important application of OCT is the segmentation of retinal layers to detect certain ophthalmic diseases (e.g., AMD and DME). These diseases may affect the appearance of certain retinal lay-

ers, such as the thickness of RNFL, retinal pigment epithelial, and other layers. An automatic graph search algorithm [31] is used to segment the retinal layers, which depend on the contrast among the different layers. Thus, the algorithm is highly susceptible to noise with inaccurate segmentation results. The layer segmentation algorithm is performed on the macula OCT images processed by different denoising methods. The results of denoised images are shown in Fig. 8(b–f). The segmentation algorithm by default uses a Gaussian filter as the denoising method, as shown in Fig. 8(c). The Gaussian filter is replaced by the comparative methods of our evaluation shown in Fig. 8(d–f).

As shown in Fig. 8(a), the two red boxes are the details of the center of the macula and retinal structure. The layer segmentation result (Fig. 8[b]) without any pre-processing is susceptible to noise. Hence, the area boundary with jitter do not correspond with the ground truth. Gaussian filtering (Fig. 8[c]) and K-SVD (Fig. 8[e]) can slightly reduce these jitters, but the loss of details causes an inaccuracy in the boundary of RNFL layer. BM3D (Fig. 8[d]) removes the noise effectively, and obtains the relative smoothness of area boundary. Influenced by the excessive denoising and loss of detailed information, segmentation result of the fovea area is inaccurate. Fig. 8(f) shows our proposed DN-GAN, which obtains the best similarity comparing to the ground truth. DN-GAN eliminates the inaccuracy of the upper boundary of the outer plexiform layer caused by artifacts, noise and false edges. To quantitatively measure the spatial overlap with ground truth, we use the dice coefficient as a metric, which is a general measure for medical segmentation [34,35], defined as follows:

$$Dice = \frac{2 \sum_x^{\Omega} p_x g_x}{\sum_x^{\Omega} p_x^2 + \sum_x^{\Omega} g_x^2} \quad (8)$$

where p_x and g_x indicate the prediction result and ground truth in region Ω of OCT, respectively. Then, x is the point of true positive in region Ω .

As shown in Table 4, the mean Dice coefficient of our DN-GAN outperforms the other pre-processing methods that the prediction result has maximum overlap area comparing to the ground truth. DN-GAN achieves the optimal balance between denoising and detail preservation.

On the other hand, to improve the visibility of OCT image, the clinician may slightly adjust the imaging angle of the image dur-

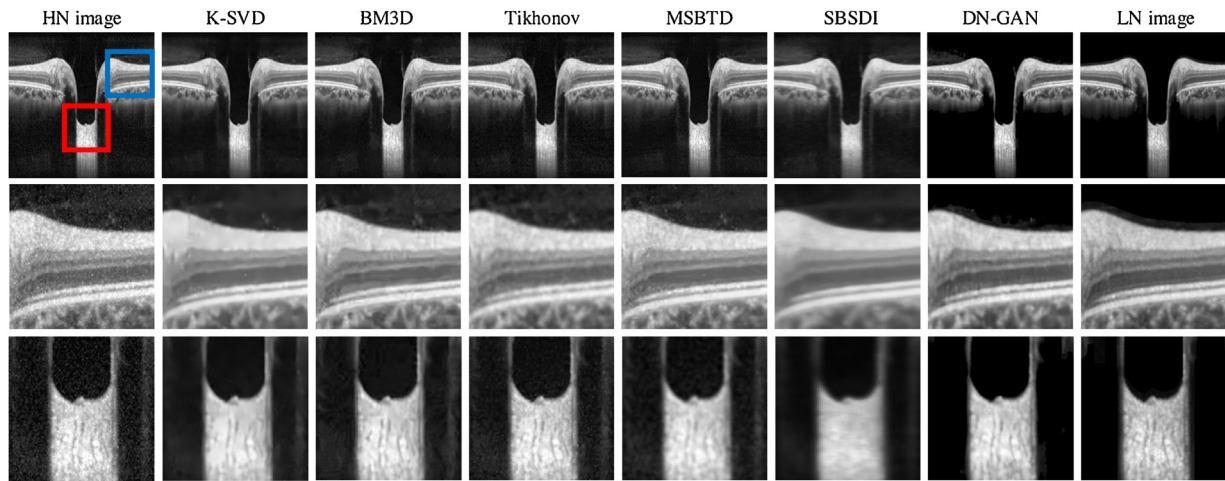


Fig. 6. Original image (HN image) and reconstruction results by the K-SVD, BM3D, Tikhonov, MSBTD, SBSDI, DN-GAN and the reference image (LN image). The middle row is the detail of retina in blue box and the bottom row is the detail of lamina cribrosa in red box. Best viewed with color and zoom-in.

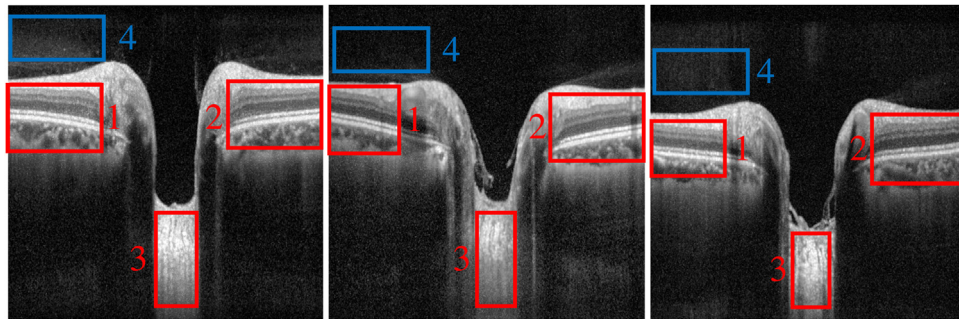


Fig. 7. ROI and background regions selected from the original image.

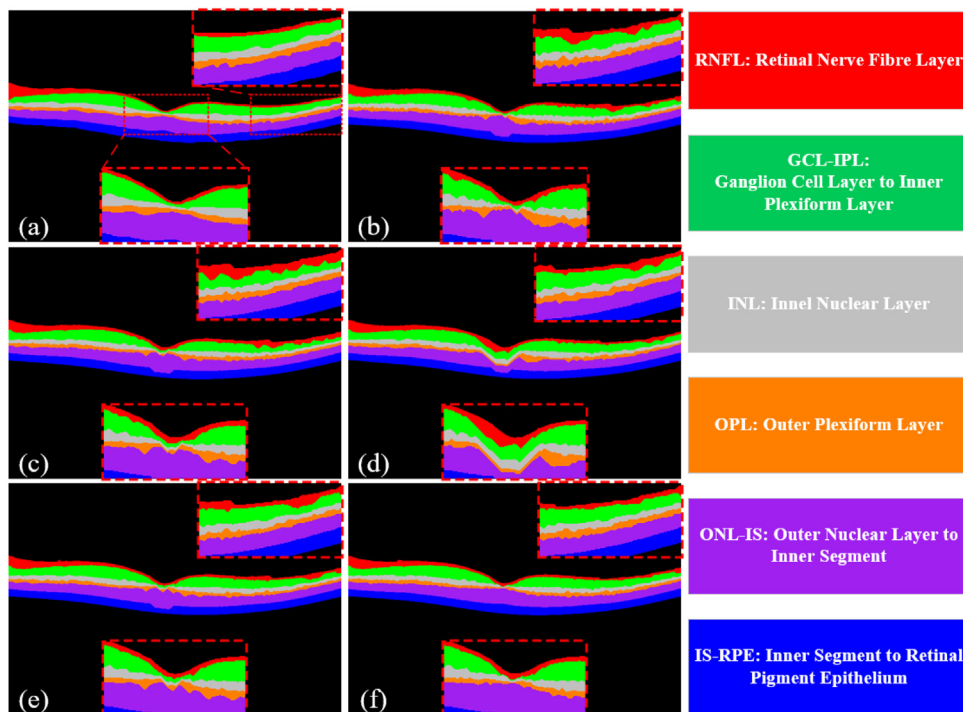


Fig. 8. Semantic segmentation result by different pre-processing methods. (a) Ground truth, (b) segmentation result without any pre-processing, and the segmentation results by (c) Gaussian filter, (d) BM3D, (e) K-SVD, (f) DN-GAN. The retinal layers corresponding to each color are presented to the right. Best viewed with color and zoom-in.

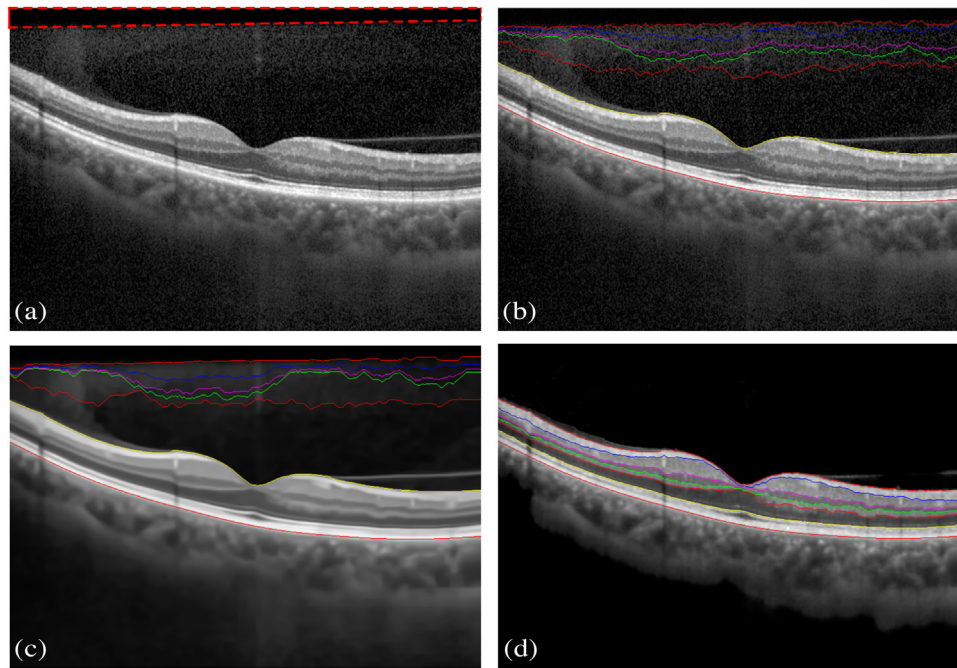


Fig. 9. Layer segmentation result. (a) Original image, (b) segmentation result without any pre-processing, and the results by the (c) BM3D, (d) DN-GAN. Best viewed with zoom-in.

Table 4

The quantitative evaluation of Dice coefficient for different methods.

Methods	RNFL	GCL-IPL	INL	OPL	ONL-IS	IS-RPE	Mean
Origin	0.8135	0.9012	0.8641	0.7952	0.9247	0.9231	0.8703
K-SVD [21]	0.7874	0.8886	0.8836	0.8510	0.9498	0.9458	0.8844
BM3D [20]	0.8092	0.8807	0.8262	0.7601	0.9016	0.9461	0.8540
Gauss [31]	0.7913	0.8905	0.8856	0.8529	0.9517	0.9465	0.8864
DN-GAN	0.8682	0.9429	0.9054	0.8631	0.9682	0.9662	0.9190

ing acquisition step, thereby possibly generating a false edge in the upper or lower part of the image (e.g., the red dotted box in Fig. 9(a)). If the grayscale gradient of the edge is similar to the boundary of retina, it may cause the graph search algorithm to fail, as shown in Fig. 9(b). Traditional denoising methods, such as the BM3D algorithm, only reduce but not eliminate the grayscale gradient in the edge of the regions. Our method can remove the background noise and avoid this issue, as shown in Fig. 9(d).

4. Conclusions

In this paper, to improve the quality of the OCT image, the DN-GAN model is presented to automatically reduce the speckle noise. To enhance the DN-GAN performance, the synthetic data is added to increase the diversity of the training data, and added the context encoding block for generator to strengthen the useful features beneficial for denoising. The frequency and spatial domain losses are first used to suppress speckle noise further. For speckle reduction in the OCT image, compared with other methods, our proposed method possesses several advantages. 1) Compared with other GAN and non-GAN models, DN-GAN achieves the best denoising performance with fewer parameters, and the generated images are more realistic than those of the other models. 2) Compared with traditional methods, our proposed model achieves a balanced performance in denoising and detail preservation. The running time is also the shortest, which is highly convenient for practical applications. 3) Visual analysis on the segmentation of the retinal images shows that the proposed method effectively improves the performance of the segmentation method, which

demonstrates that DN-GAN has clinical value in practical application.

Although the results prove that the proposed method obtains better performance than those of the other existing approaches, DN-GAN also possesses some limitations. Structural information may be removed as noise in the areas with low pixel intensity due to the attenuation of incident light, thereby resulting in the loss of information of the choroid. This research direction is the topic we will work in our next work. In the future, our method will be applied to 3D-OCT images by using the repeated information of adjacent frames to achieve improved denoising effect without collecting the dense B-scan images. By contrast, DN-GAN also has the potential to perform image denoising and super-resolution simultaneously only by downsampling the HN image and adding the upsampling layer as the last layer to generator. The high-quality magnification of the entire image or certain areas can be achieved, which is beneficial in observing the anomalous areas with small lesions.

Acknowledgments

This research was supported by the National Natural Science Foundation of China under Grant No. 61672542 and No. 61772556.

Declaration of Competing Interest

The authors declared that they have no conflicts of interest to this work.

References

- [1] David Huang, Eric A. Swanson, Charles P. Lin, Joel S. Schuman, William G. Stinson, Warren Chang, Michael R. Hee, Thomas Flotte, Kenton Gregory, Carmen A. Puliafito, et al., Optical coherence tomography, *Science* 254 (5035) (1991) 1178–1181.
- [2] Folkert K. Horn, Christian Y. Mardin, Robert Laemmer, Delia Baleanu, Anselm M. Juenemann, Friedrich E. Kruse, Ralf P. Tornow, Correlation between local glaucomatous visual field defects and loss of nerve fiber layer thickness

- measured with polarimetry and spectral domain oct, *Invest. Ophthalmol. Vis. Sci.* 50 (5) (2009) 1971–1977.
- [3] Ilse Krebs, Stefan Hagen, Werner Brannath, Paulina Haas, Irene Womastek, Gabriella de Salvo, Siamak Ansari-Shahrezaei, Susanne Binder, Repeatability and reproducibility of retinal thickness measurements by optical coherence tomography in age-related macular degeneration, *Ophthalmology* 117 (8) (2010) 1577–1584.
 - [4] U. Rajendra Acharya, Muthu Rama Krishnan Mookiah, Joel E.W. Koh, Jen Hong Tan, Sulatha V. Bhandary, A. Krishna Rao, Yuki Hagiwara, Chua Kuang Chua, Augustinus Laude, Automated diabetic macular edema (dme) grading system using dwt, dct features and maculopathy index, *Comput. Biol. Med.* 84 (2017) 59–68.
 - [5] Zailiang Chen, Yufang Mo, Pingbo Ouyang, Hailan Shen, Dabao Li, Rongchang Zhao, Retinal vessel optical coherence tomography images for anemia screening, *Med. Biol. Eng. Comput.* 57 (4) (2019) 953–966.
 - [6] Joseph M. Schmitt, S.H. Xiang, Kin Man Yung, Speckle in optical coherence tomography, *J. Biomed. Opt.* 4 (1) (1999) 95–106.
 - [7] Andras A. Lindenmaier, Leigh Conroy, Golnaz Farhat, Ralph S. DaCosta, Costel Flueraru, I. Alex Vitkin, Texture analysis of optical coherence tomography speckle for characterizing biological tissues in vivo, *Opt. Lett.* 38 (8) (2013) 1280–1282.
 - [8] Mohammad R.N. Avanaki, Ramona Cernat, Paul J. Tadrous, Taran Tatla, Adrian Gh Podoleanu, S. Ali Hojjatoleslami, Spatial compounding algorithm for speckle reduction of dynamic focus oct images, *IEEE Photonics Technol. Lett.* 25 (15) (2013) 1439–1442.
 - [9] J.M. Schmitt, Array detection for speckle reduction in optical coherence microscopy, *Phys. Med. Biol.* 42 (7) (1997) 1427.
 - [10] Jadwiga Rogowska, Mark E. Brezinski, Image processing techniques for noise removal, enhancement and segmentation of cartilage oct images, *Phys. Med. Biol.* 47 (4) (2002) 641.
 - [11] Aydogan Ozcan, Alberto Bilenca, Adrien E. Desjardins, Brett E. Bouma, Guillermo J. Tearney, Speckle reduction in optical coherence tomography images using digital filtering, *JOSA A* 24 (7) (2007) 1901–1910.
 - [12] Rui Bernardes, Cristina Maduro, Pedro Serranho, Adérito Araújo, Sílvia Barbeiro, José Cunha-Vaz, Improved adaptive complex diffusion despeckling filter, *Opt. Express* 18 (23) (2010) 24048–24059.
 - [13] Markus A. Mayer, Anja Borsdorf, Martin Wagner, Joachim Hornegger, Christian Y. Mardin, Ralf P. Tornow, Wavelet denoising of multiframe optical coherence tomography data, *Biomed. Opt. Express* 3 (3) (2012) 572–589.
 - [14] Tsuyoshi Murakami, Koichi Ogawa, Speckle noise reduction of optical coherence tomography images with a wavelet transform, in: 2018 IEEE 14th International Colloquium on Signal Processing & Its Applications (CSPA), IEEE, 2018, pp. 31–34.
 - [15] Leyuan Fang, Shutao Li, Qing Nie, Joseph A. Izatt, Cynthia A. Toth, Sina Farsiu, Sparsity based denoising of spectral domain optical coherence tomography images, *Biomed. Opt. Express* 3 (5) (2012) 927–942.
 - [16] Leyuan Fang, Shutao Li, Ryan P. McNabb, Qing Nie, Anthony N. Kuo, Cynthia A. Toth, Joseph A. Izatt, Sina Farsiu, Fast acquisition and reconstruction of optical coherence tomography images via sparse representation, *IEEE Trans. Med. Imaging* 32 (11) (2013) 2034–2049.
 - [17] Ahmadreza Baghaie, Roshan M. D'souza, Yu Zeyun, Sparse and low rank decomposition based batch image alignment for speckle reduction of retinal oct images, 2015 IEEE 12th International Symposium on Biomedical Imaging (ISBI) (2015) 226–230.
 - [18] Danilo A. Jesus, D. Robert Iskander, Assessment of corneal properties based on statistical modeling of oct speckle, *Biomed. Opt. Express* 8 (1) (2017) 162–176.
 - [19] Muxingzi Li, Ramzi Idoughi, Biswarup Choudhury, Wolfgang Heidrich, Statistical model for oct image denoising, *Biomed. Opt. Express* 8 (9) (2017) 3903–3917.
 - [20] Chong Bo, Yong-Kai Zhu, Speckle reduction in optical coherence tomography images of human finger skin by wavelet modified bm3d filter, *Opt. Commun.* 291 (2013) 461–469.
 - [21] Raheleh Kafieh, Hossein Rabbani, Ivan Selesnick, Three dimensional data-driven multi scale atomic representation of optical coherence tomography, *IEEE Trans. Med. Imaging* 34 (5) (2014) 1042–1062.
 - [22] Ian Goodfellow, Jean Pouget-Abadie, Mehdi Mirza, Bing Xu, David Warde-Farley, Sherjil Ozair, Aaron Courville, Yoshua Bengio, Generative adversarial nets, *Adv. Neural Inf. Process. Syst.* (2014) 2672–2680.
 - [23] Christian Ledig, Lucas Theis, Ferenc Huszar, Jose Caballero, Andrew Cunningham, Alejandro Acosta, Andrew Aitken, Alykhan Tejani, Johannes Totz, Zehan Wang, et al., Photo-realistic single image super-resolution using a generative adversarial network, *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition* (2017) 4681–4690.
 - [24] Yuhua Chen, Feng Shi, Anthony G. Christodoulou, Yibin Xie, Zhengwei Zhou, Li Debiao, Efficient and accurate mri super-resolution using a generative adversarial network and 3d multi-level densely connected network, in: *International Conference on Medical Image Computing and Computer-Assisted Intervention*, Springer, 2018, pp. 91–99.
 - [25] Guang Yang, Simiao Yu, Hao Dong, Greg Slabaugh, Pier Luigi Dragotti, Xujiong Ye, Fangde Liu, Simon Arridge, Jennifer Keegan, Yike Guo, et al., Dagan: deep de-aliasing generative adversarial networks for fast compressed sensing mri reconstruction, *IEEE Trans. Med. Imaging* 37 (6) (2017) 1310–1321.
 - [26] Philippe Thevenaz, Urs E. Ruttimann, Michael Unser, A pyramid approach to subpixel registration based on intensity, *IEEE Trans. Image Process.* 7 (1) (1998) 27–41.
 - [27] Pedro Serranho, Cristina Maduro, Torcato Santos, José Cunha-Vaz, Rui Bernardes, Synthetic oct data for image processing performance testing, in: 2011 18th IEEE International Conference on Image Processing, IEEE, 2011, pp. 401–404.
 - [28] Gao Huang, Zhuang Liu, Laurens Van Der Maaten, Kilian Q. Weinberger, Densely connected convolutional networks, *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition* (2017) 4700–4708.
 - [29] Jie Hu, Li Shen, Gang Sun, Squeeze-and-excitation networks, *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition* (2018) 7132–7141.
 - [30] Gabriel T. Chong, Sina Farsiu, Sharon F. Freedman, Neeru Sarin, Anjum F. Koreishi, Joseph A. Izatt, Cynthia A. Toth, Abnormal foveal morphology in ocular albinism imaged with spectral-domain optical coherence tomography, *Arch. Ophthalmol.* 127 (1) (2009) 37–44.
 - [31] Stephanie J. Chiu, Xiao T. Li, Peter Nicholas, Cynthia A. Toth, Joseph A. Izatt, Sina Farsiu, Automatic segmentation of seven retinal layers in sdopt images congruent with expert manual segmentation, *Opt. Express* 18 (18) (2010) 19413–19428.
 - [32] Karen Simonyan, Andrew Zisserman, Very deep convolutional networks for large-scale image recognition, *Comput. Sci.* (2014).
 - [33] Olga Russakovsky, Jia Deng, Hao Su, Jonathan Krause, Sanjeev Satheesh, Sean Ma, Zhiheng Huang, Andrej Karpathy, Aditya Khosla, Michael Bernstein, et al., Imagenet large scale visual recognition challenge, *Int. J. Comput. Vis.* 115 (3) (2015) 211–252.
 - [34] Fausto Milletari, Nassir Navab, Seyed-Ahmad Ahmadi, V-net: fully convolutional neural networks for volumetric medical image segmentation, in: 2016 Fourth International Conference on 3D Vision (3DV), IEEE, 2016, pp. 565–571.
 - [35] Zailiang Chen, Peng Peng, Hailan Shen, Hao Wei, Pingbo Ouyang, Xuanchu Duan, Region segmentation strategy for bruch's membrane opening detection in spectral domain optical coherence tomography images, *Biomed. Optics Express* 10 (2) (2019) 526–538.