

Unsupervised Denoising of Optical Coherence Tomography Images with Nonlocal-Generative Adversarial Network

Anjing Guo, *Student Member, IEEE*, Leyuan Fang, *Senior Member, IEEE*, Min Qi, and Shutao Li, *Fellow, IEEE*

Abstract—Deep learning for image denoising has recently attracted considerable attentions due to its excellent performance. Since most of current deep learning based denoising models require a large number of clean images for training, it is difficult to extend them to the denoising problems when the reference clean images are hard to acquire (e.g., optical coherence tomography (OCT) images). In this paper, we propose a novel unsupervised deep learning model called as nonlocal-generative adversarial network (Nonlocal-GAN) for OCT image denoising, where the deep model can be trained without reference clean images. Specifically, considering that the background areas of OCT images mainly contain pure real noise samples, we creatively train a discriminator to distinguish background real noise samples from the fake noise samples generated by the denoiser, i.e., the generator, and then the discriminator will guide the generator for denoising. To further enhance denoising performance, we introduce a nonlocal means layer into the generator of the Nonlocal-GAN model. Furthermore, since nearby several OCT B-scans have strong correlations, we also propose a Nonlocal-GAN-M model to utilize the high correlations within nearby B-scans. Extensive experimental results on clinical retinal OCT images demonstrate the effectiveness and efficiency of the proposed method.

Index Terms—Optical coherence tomography, generative adversarial networks, deep learning, image denoising.

I. INTRODUCTION

OPTICAL coherence tomography (OCT) is a non-invasive 3D tomography imaging technique, which can provide a high resolution view of the superficial structures in tissue at micro-meter level [1]–[3]. As an effective diagnostic tool, OCT has been widely used in diagnostic ophthalmology [4]–[6]. However, the OCT imaging process is inevitably interfered by sample-based speckle and detector noise, which degrades the image quality, and thus seriously affects the subsequent processing and analysis of the OCT images.

In the last decades, many denoising models have been introduced to reconstruct high quality medical images [7]–[18]. Traditional reconstruction approaches were based on

This work was supported by the National Natural Science Fund of China under Grants 61890962 and 61922029, the Science and Technology Plan Project Fund of Hunan Province under Grant CX2018B171, 2017RS3024, 2018TP1013, the Science and Technology Talents Program of Hunan Association for Science and Technology under Grants 2017TJ-Q09, and the National Key R&D Program of China under Grant 2018YFB1305200.

A. Guo, L. Fang, and S. Li are with the College of Electrical and Information Engineering, Hunan University, Changsha, 410082, and also with the Key Laboratory of Visual Perception and Artificial Intelligence of Hunan Province, Changsha, 410082, China (email: anjing_guo@hnu.edu.cn; fangleyuan@gmail.com; shutao_li@hnu.edu.cn).

M. Qi is with the Department of Plastic Surgery, Xiangya Hospital, Central South University, Changsha, 410082, China (email: qimin05@csu.edu.cn).

the smoothness prior and processed images in the spatial domain (e.g., anisotropic filtering, Tikhonov filtering, and total variation [8], [9]). Other approaches transformed the input degraded images into other domains (e.g., using the wavelet [15], [16], dual tree complex wavelet [17], and curvelets [18]). Recently, nonlocal denoising methods [19]–[21] have been introduced, which can exploit the self-similarity in the image to obtain excellent denoising performance. In [19], the nonlocal means (NLM) method first searched the similar pixels for the current processed pixel, and then estimated the current pixel by a weighted average of the similar pixels. In the block matching and 3D collaborative filtering (BM3D) [20] method, similar image patches were stacked together and jointly filtered in the 3D transform domain. The block matching and 4D collaborative filtering (BM4D) method proposed in [21] was the extension version of BM3D, which can be conducted on 3D image volumes. These two methods can provide very promising denoising results for 2D/3D natural images/volumes. In addition, some sparsity based methods using dictionary learning (K-SVD) [22] have also been applied for noise reduction in OCT images [23]–[25]. In [23], a multiscale sparsity based tomographic denoising (MSBTD) method was proposed for OCT image denoising. In [25], the segmented retinal layer information was utilized to achieve the efficient dictionary learning and nonlocal denoising. However, the denoised images obtained by these methods often show smooth appearance, which loses some meaningful pathology details in the original noisy images.

Recently, deep learning has demonstrated to be a powerful method for image recognition [26]–[32] and restoration problems [33]–[42]. The advantage of deep learning based denoising methods was that they can preserve more details in denoising, which was benefited from their deeper architectures and quantity of learnable parameters. In [33], a shallow convolutional neural network (CNN) was firstly introduced for natural image denoising. In [35], Zhang *et al.* proposed a DnCNN model for natural image denoising, which introduced residual learning [29] and batch normalization [43] to boost the denoising performance. To enhance the flexibility of CNN denoise model, in [36], Zhang *et al.* further presented a fast and flexible denoising convolutional neural network (FFDNet) with a tunable noise level map as the input. The denoising results obtained by the above deep learning-based methods are very convincing. However, these methods require a large number of reference clean images to train their deep neural networks. In real clinical OCT image application, the reference

clean images are not available. Though we can adopt densely sampling imaging strategy (i.e., repeatedly imaging same position with multiple times) to obtain noise-reduced averaged image for the training of CNN [38], it will greatly increase acquisition time and some motions will also create heavy blurs in the final averaged image. Therefore, it is difficult to directly extend these deep learning models for OCT image denoising.

In addition, generative adversarial network (GAN) [44] has also been successfully applied for image denoising problems [39]–[41]. In [39], the authors proposed to generate numerous simulated noise patches with a GAN, and then added them to clean natural images to construct training pairs for blind image denoising (GCBDD). In [40], an extra adversarial loss was introduced to low-dose CT image denoising (GAN-MSE), which can obtain clearer CT images compared with the way that minimizes the mean squared error between low-dose and routine-dose images directly. Similarly, Yang *et al.* in [41] introduced a Wasserstein GAN [45] with perceptual loss (GAN-WDP) to denoise low-dose CT images. In these models, they still used clean natural images or the routine-dose (full-dose) CT images. Inspired by the concept that GAN can be trained with unpaired adversarial samples, we attempt to build a network which can learn to denoise OCT images without reference clean images by exploiting the characteristics of original noisy OCT images.

In this paper, a new model termed as nonlocal-generative adversarial network (Nonlocal-GAN) is proposed for OCT image denoising, which can learn image characteristics and denoise in an unsupervised manner. Unlike previous GAN based image restoration models, the discriminator of Nonlocal-GAN can learn the characteristics of the noise in original noisy OCT images and then guide the generator for denoising. Specifically, the proposed Nonlocal-GAN first uses the detail parts of the OCT images to construct the detail noise dataset, and the background parts to construct the background noise dataset. Then, a nonlocal means layer is incorporated into the generator part of the GAN, which can sufficiently exploit the nonlocal similarity of the OCT image. Subsequently, the detail and background noise datasets are utilized for training the Nonlocal-GAN. The generator of the network is used for denoising and estimating simulated noise samples to challenge the discriminator, while the discriminator is then employed to identify real noise samples from the estimated noise samples. After training process, the generator of Nonlocal-GAN can denoise testing OCT images in an end-to-end fashion.

The remainder of this paper is organized as follows. Section II briefly reviews the related works. Section III first introduces the proposed Nonlocal-GAN model, and then uses it for OCT image denoising. Experimental results on clinical OCT data are presented in Section IV. Section V concludes this paper and suggests some future works.

II. RELATED WORKS

A. Nonlocal Means for Image Denoising

The nonlocal means is an effective denoising tool, which can exploit the self-similarities inherent to images [19]. Specifically, for each patch p_i cropped from the noisy image, the

nonlocal means method first searches patches with the highest similarity to p_i from a large window. The similarity can be measured by the squared ℓ_2 -norm of the intensity differences between two patches:

$$d(p_i, p_j) = \|p_i - p_j\|_2^2, j \in \Lambda, \quad (1)$$

where Λ is a set containing the indexes of all patches in the whole searching window. Then, these found patches are processed by a weighted average filtering or patch similarity penalty to achieve denoising [46]. The nonlocal block matching and weighted filtering have demonstrated to be useful to improve the denoising performance [19]–[21].

B. Generative Adversarial Network

In [44], Goodfellow *et al.* first introduced GAN, which includes two components: a discriminator D and a generator G . The discriminator, $D(x)$, computes the probability that the input x is sampled from data distribution p_{data} rather than generator distribution p_G . Meanwhile, G learns a distribution p_G and samples $G(z)$ from p_G with an input noise variable z , which is drawn from a prior $p_z(z)$. Then the virtual sample $G(z)$ is utilized to challenge D . Both G and D can be implemented by neural networks and trained by solving the following minimax optimization:

$$\min_G \max_D V(G, D) = \mathbb{E}_{x \sim p_{data}(x)} [\log(D(x))] + \mathbb{E}_{z \sim p_z(z)} [\log(1 - D(G(z)))]. \quad (2)$$

The learning follows an iterative procedure, where the discriminator and generator are alternatively updated. Given any generator G , the training criterion for D is to maximize

$$V(G, D) = \int_x p_{data}(x) \log(D(x)) + p_G(x) \log(1 - D(x)) dx. \quad (3)$$

The maximum of function $y \rightarrow a \log(y) + b \log(1 - y)$ in $[0, 1]$ can be achieved at $\frac{a}{a+b}$, and the optimal D is

$$D_G^*(x) = \frac{p_{data}(x)}{p_{data}(x) + p_G(x)}. \quad (4)$$

Substituting $D_G^*(x)$ into equation (2), we can get

$$\begin{aligned} C(G) &= \max_D V(G, D) \\ &= \mathbb{E}_{x \sim p_{data}} [\log D_G^*(x)] + \mathbb{E}_{x \sim p_G} [\log(1 - D_G^*(x))] \\ &= -\log(4) + KL \left(p_{data} \parallel \frac{p_{data} + p_G}{2} \right) + \\ &KL \left(p_G \parallel \frac{p_{data} + p_G}{2} \right), \end{aligned} \quad (5)$$

where KL is the Kullback-Leibler divergence. Furthermore we get can get

$$C(G) = -\log(4) + 2 \cdot JSD(p_{data} \parallel p_G), \quad (6)$$

where JSD is the Jensen-Shannon divergence, the JSD is always non-negative and 0 only when $p_{data} = p_G$, i.e., the G perfectly replicate the distribution of real data. In the training process of GAN, G makes p_G approach p_{data} .

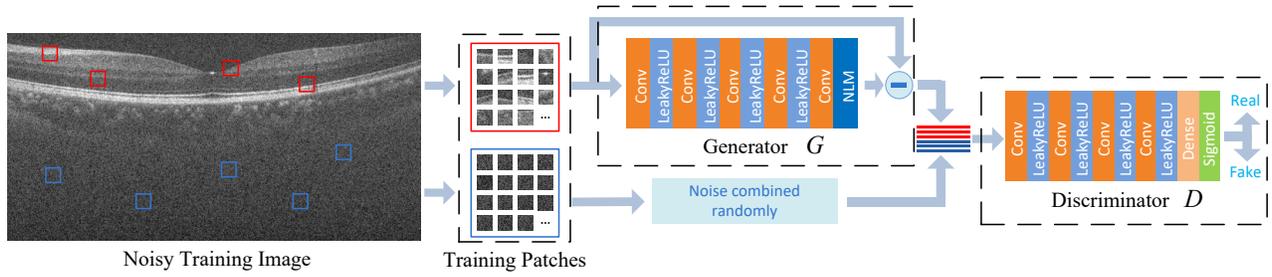


Fig. 1. The framework of Nonlocal-GAN. The Nonlocal-GAN consists of two components: a generator G and a discriminator D . The generator is a CNN based denoising model, which generates noise samples from the noisy detail patches. The discriminator is a CNN based classification model, which attempts to distinguish the background real noise samples from the synthetic noise samples created by the generator.

III. PROPOSED NONLOCAL-GAN FOR OCT IMAGE DENOISING

In this paper, we propose the Nonlocal-GAN method, which can be trained without the reference clean images and denoise in an end-to-end way. The Nonlocal-GAN method is composed of four main parts: a) creation of training datasets; b) structure of Nonlocal-GAN; c) training of Nonlocal-GAN; and d) denoising with Nonlocal-GAN. The framework of Nonlocal-GAN is illustrated in Fig. 1.

A. Creation of Training Datasets

To construct training datasets, we propose an automatic method to divide the noisy OCT image into two parts. First, consider the noise model:

$$y = x + v, \quad (7)$$

where y is the noisy image, x is the reference clean image which obeys distribution p_r , and v is residual between y and x that obeys distribution p_n , i.e., the noise. The previous GAN based image restoration methods [40], [41] use the discriminator D to identify whether the input image is sampled from p_r or denoised by generator G . In this way, G can learn a distribution p_G which approaches p_r , and thus G can efficiently restore high quality images from the degraded images. However, these methods require the reference clean image x , which means p_r is known, that is hard to obtain in OCT image denoising problem.

Meanwhile, if we know the noise distribution p_n , we can make D identify real noise samples from the synthetic noise samples generated by G , and then guide G to denoise. In this way, a two-player minimax game can also be well constructed, and the training criterion for D can be formulated as

$$V(G, D) = \mathbb{E}_{x \sim p_n(x)} [\log(D(x))] + \mathbb{E}_{x \sim p_G(x)} [\log(1 - D(x))]. \quad (8)$$

As shown in Fig. 1, we can get these synthetic noise samples by subtracting the denoised images from original OCT images.

In original OCT images, there are many areas with pure noise, i.e., the background areas, which are perfect samples of distribution p_n . Hence, to construct training dataset, considering that the luminance of detail parts is much higher than the background parts in OCT images, we first compute the mean value of the whole OCT image. Then, we can obtain a lot of

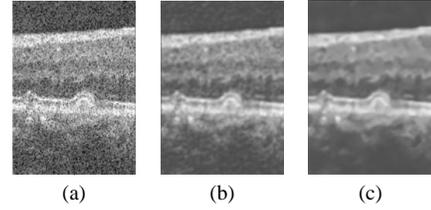


Fig. 2. The comparison of denoising results: (a) original noisy region; (b) region denoised by G without the nonlocal means layer; (c) region denoised by G with the nonlocal means layer.

patches with size $w \times w$ (e.g., 33×33) by scanning the whole image with stride s (e.g., 6). Next, we divide these patches into two datasets by comparing the mean value of patches with the mean value of the whole image, i.e., the detail noise dataset $Data_{dn}$ and the background noise dataset $Data_{bn}$. The constraints of detail noise patches and background noise patches can be separately formulated as,

$$Mean(p) \geq (1 + \alpha) \cdot Mean(I), \quad (9)$$

$$Mean(p) \leq (1 - \alpha) \cdot Mean(I), \quad (10)$$

where p represents a patch, I represents the whole image, and α is a scale factor.

B. Structure of Nonlocal-GAN

1) *Generator G* : The generator G is a CNN based model, which transforms the input detail noise patch $I_{dn} \subset Data_{dn}$ into the residual patch $G(I_{dn})$. The CNN in G predicts clean patch I_c , and then we can get $G(I_{dn})$ by the subtraction of I_c from I_{dn} .

The CNN in G contains five convolution layers with small kernel (of size 3×3). The number of convolution kernels is 64 for the first four convolution layers, and 1 for the last. Except the final layer, all convolution layers use leaky rectified linear activation function (LeakyReLU), which can increase training stability [38, 43]. The weights are initialized with a normal distribution ($\mu = 0.0, \sigma = 0.001$), and the biases are initialized with zero.

2) *Nonlocal means layer*: After the last convolution layer, the I_c predicted by CNN in G may still have noise. Therefore, a nonlocal means layer (see the “NLM” layer in Fig. 1) is further incorporated into the CNN to remove the residual

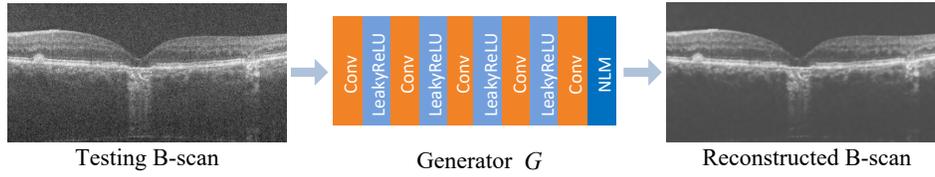


Fig. 3. The structure of testing process of Nonlocal-GAN-S method. In the Nonlocal-GAN-S method, each testing B-scan is sent to the well-trained G for denoising directly.

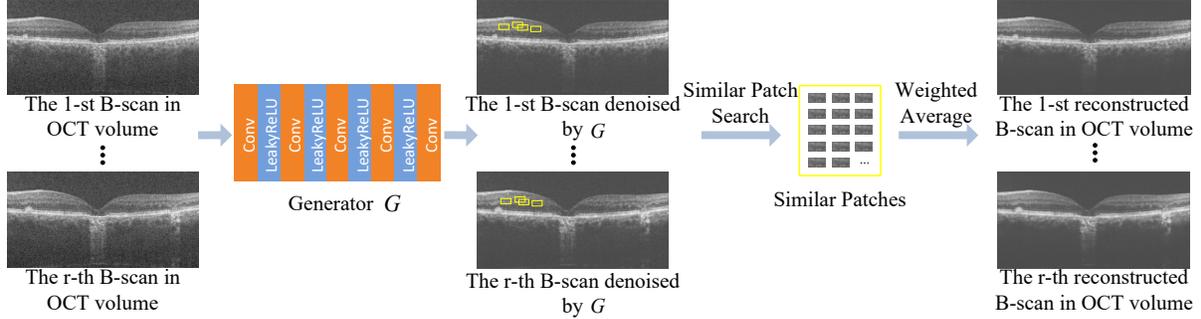


Fig. 4. The structure of testing process of Nonlocal-GAN-M method. In the Nonlocal-GAN-M method, the nonlocal means layer in G is removed, and all the B-scans in the OCT volume are sent to G as a batch. Then, the nonlocal means operation is conducted among the denoised B-scans to remove the residual noise.

noise. Specifically, for each template patch p_i extracted from I_c , the nonlocal means layer first computes the similarities of all patches with p_i in a searching window. The similarity between two patches can be measured by equation (1). Then, all patches in the searching window are processed by a weighted average filtering to achieve denoising. The process is similar to the NLM method in [19]. The effect of the nonlocal means layer is shown in Fig. 2.

3) *Discriminator D*: The discriminator D is a conventional classification network based on CNN. The input of D is the residual $G(I_{dn})$ or the subset I_{bn} sampled from $Data_{bn}$.

The discriminator contains four convolution layers with kernel size 5×5 . The first two convolution layers have 64 kernels, while the remaining two layers have 32 kernels. The strides of all convolution layers are 1, 2, 1 and 2 [47], respectively. All convolution layers use LeakyReLU activation functions. The resulting feature maps are flattened and connected to an output node. Then, a sigmoid activation is utilized to determine whether the input is a real noise sample (label 1) or not (label 0). The initialization of weights and biases in D is the same as the generator G .

C. Training of Nonlocal-GAN

The framework shown in Fig. 1 contains two parts: generator G with output $G(I_{dn})$, and discriminator D with output $D(G(I_{dn}))$ or $D(I_{bn})$. The loss for G is defined as

$$\ell_G = \ell(D(G(I_{dn})), 1), \quad (11)$$

where $\ell(D(G(I_{dn})), 1)$ is the binary cross-entropy between the discriminator's prediction and the label 1 that the generator wants the discriminator to predict. In other words, the goal of G is to denoise and provide the residual patches (the virtual noise samples) which can confuse D .

For discriminator D , it has an adversarial goal to the generator and aims to correctly distinguish the $G(I_{dn})$ generated by G or the I_{bn} sampled from $Data_{bn}$. Hence, the discriminator minimizes

$$\ell_D = \ell(D(I_{bn}), 1) + \ell(D(G(I_{dn})), 0), \quad (12)$$

where $\ell(D(I_{bn}), 1)$ is the binary cross-entropy between the discriminator's decision on I_{bn} and its target label 1, and $\ell(D(G(I_{dn})), 0)$ is the binary cross-entropy between the discriminator's decision on $G(I_{dn})$ and its target label 0.

Specifically, to construct more background noise samples from the limited $Data_{bn}$, we randomly arrange pixels in background noise patches at each iteration. All relevant parameters in the generator and discriminator are simultaneously optimized using the stochastic gradient descent (SGD) optimizer, with a learning rate of 0.0005 and a moment of 0.9. The method was implemented in Tensorflow [48] and Keras.

D. Denoising with Nonlocal-GAN

In testing process, we propose two methods for OCT image denoising: Nonlocal-GAN-S and Nonlocal-GAN-M.

1) *Nonlocal-GAN-S*: The structure of Nonlocal-GAN-S is shown in Fig. 3. Since there is no fully connected layer in generator G and the convolution operation can be conducted on the images with random sizes, each B-scan in the 3D OCT volume can be sent to G directly. After the nonlocal means layer, we can get the denoised B-scan.

2) *Nonlocal-GAN-M*: Considering that each OCT volume includes a number of B-scans, and nearby slices in the OCT volume have strong correlations, we further propose the Nonlocal-GAN-M method to utilize the high correlations within several nearby B-scans. Fig. 4 illustrates the structure of

Nonlocal-GAN-M. Specifically, we first remove the nonlocal means layer in G . Then, all the B-scans of an OCT volume are sent into G as a batch, and we can get the denoised B-scans which may still have noise. After that, one nonlocal means operation is conducted among these denoised B-scans. Specifically, for the patch p_i^r in the r -th slice of the B-scans volume, we first compute the similarities of J nearby overlapped patches $\{p_i^{r,1}, \dots, p_i^{r,J}\}$ with the template patch p_i^r in each slice using the similarity measure in equation (1). Then, for all nearby patches $\{p_i^{1,1}, \dots, p_i^{1,J}, \dots, p_i^{r,1}, \dots, p_i^{r,J}\}$, we apply a weighted average on these patches to obtain an averaged patch $p_i^{Ave,r}$,

$$p_i^{Ave,r} = \sum_{n=1}^r \sum_{j=1}^J w_i^{n,j} p_i^{n,j}, \quad (13)$$

where $w_i^{n,j}$ is the weight for the patch $p_i^{n,j}$, computed as

$$w_i^{n,j} = \frac{\exp\left(-\|p_i^{n,j} - p_i^r\|_2^2/h^2\right)}{\sum_{n=1}^r \sum_{j=1}^J \exp\left(-\|p_i^{n,j} - p_i^r\|_2^2/h^2\right)}, \quad (14)$$

where h is a predetermined scalar. In this way, we can use the nonlocal similarity information among different B-scans to remove the residual noise further.

IV. EXPERIMENTS

A. Datasets

Our experiments used two retinal OCT volume datasets. Both the two datasets were imaged by a Bioptigen SDOCT system (Durham, NC, USA) with an axial resolution of 4.5 μm per pixel in tissue. These datasets were attained with adherence to the tenets of the Declaration of Helsinki and were captured in the A2A SD-OCT Study [49].

The first dataset was originally introduced in the works [24], [25] which includes OCT images from 18 subjects with normal eye and age macular degeneration (AMD)¹. Each subject has 5 nearby B-scans (each is of size 450×900) near the fovea.

The second dataset was previously used in the work [50], which consists of macular SDOCT images from 20 subjects with severe AMD². Each subject has 11 adjacent B-scans (each is of size 512×1000) near the fovea.

B. Compared Methods and Parameters Setting

The proposed Nonlocal-GAN method³ is compared with some well-known denoising methods: NLM [19], BM3D [20], K-SVD [22], BM4D [21], segmentation based sparse reconstruction (SSR) [25], GCBD [39], GAN-MSE [40], DnCNN [35], GAN-WDP [41], and denoising GAN (DeGAN). In the DeGAN method, to show the effectiveness of the nonlocal

means layer in G , we removed the nonlocal means layer of the Nonlocal-GAN-S method in training and testing process. Similarly, to show the effectiveness of the GAN part in our models, we compared our models with the NLM [19] method. The K-SVD [22] is a popular sparsity based denoising method, which trains an universal dictionary on the input noisy image. The BM3D [20] is a benchmark denoising method, which exploits patch self-similarities of the input image by nonlocal searching and 3D collaborative filtering. The BM4D [21] is the extension version of BM3D, which can be conducted on 3D image volumes. The SSR [25] method is a segmentation-based reconstruction framework with sparse representation for OCT image denoising, which exploits patch similarities within each segmented layer to enhance the reconstruction performance. The other four compared methods are deep learning based models: the GCBD [39] method which generates plenty of simulated noise samples with a WGAN for blind image denoising, the DnCNN [35] method which is a single CNN denoising model, the GAN-MSE [40] and GAN-WDP [41] methods which are trained with extra adversarial loss and perceptual loss.

For the NLM [19], BM3D [20], K-SVD [22], GCBD [39], GAN-MSE [40], DnCNN [35], GAN-WDP [41], DeGAN and Nonlocal-GAN-S methods, we separately denoised each B-scan of the OCT volume. For the BM4D [21], SSR [25] and Nonlocal-GAN-M methods, these methods required several B-scans of an OCT volume in denoising. Considering that slices from farther distances might have large differences, we chose 5 nearby slices of the volume, i.e., r was set to 5, and α was set to 0.15. For the nonlocal means layer of Nonlocal-GAN-S and NLM method [19], the scalar h in equation (14), size of search window, and similarity square neighborhood were set to 10, 10×10 , and 5×5 , respectively. For the Nonlocal-GAN-M method, the scalar h in equation (14) was set to 10. To reduce computing cost, the size of nearby patches and J were selected to 5×10 and 50, respectively. The parameters of other compared methods: K-SVD [22], BM3D [20], BM4D [21] and SSR [25] were set to the default values as in their references. For the deep learning based methods GCBD [39], GAN-MSE [40], DnCNN [35], GAN-WDP [41], since the clean OCT images are hard to acquire, as indicated in [25], we averaged the nearby B-scans in an OCT volume to get the low-noise version of the foveal noisy B-scan, and then used the averaged images as reference clean images. In these methods, the division of training and testing data was consistent with our proposed Nonlocal-GAN-S and Nonlocal-GAN-M methods.

C. Quantitative Metrics

Unlike simulated experiments in natural image denoising, the reference clean image of each OCT B-scan is not available, and therefore, the conventional metrics PSNR and structural similarity index (SSIM) [51] which need the reference images can not be used for evaluating OCT image denoising. In this paper, we choose the following nonreference metrics to evaluate denoising performance.

1) *Signal to Noise Ratio (SNR)*: the SNR is a global performance measure which has been widely used to evaluate

¹Dataset was downloaded at: http://people.duke.edu/~sf59/Fang_TMI_2013.htm

²Dataset was downloaded at: http://people.duke.edu/~sf59/Chiu_IOVS_2011_dataset.htm

³The source code of the proposed Nonlocal-GAN method will be released at <https://sites.google.com/site/leyuanfang/home>

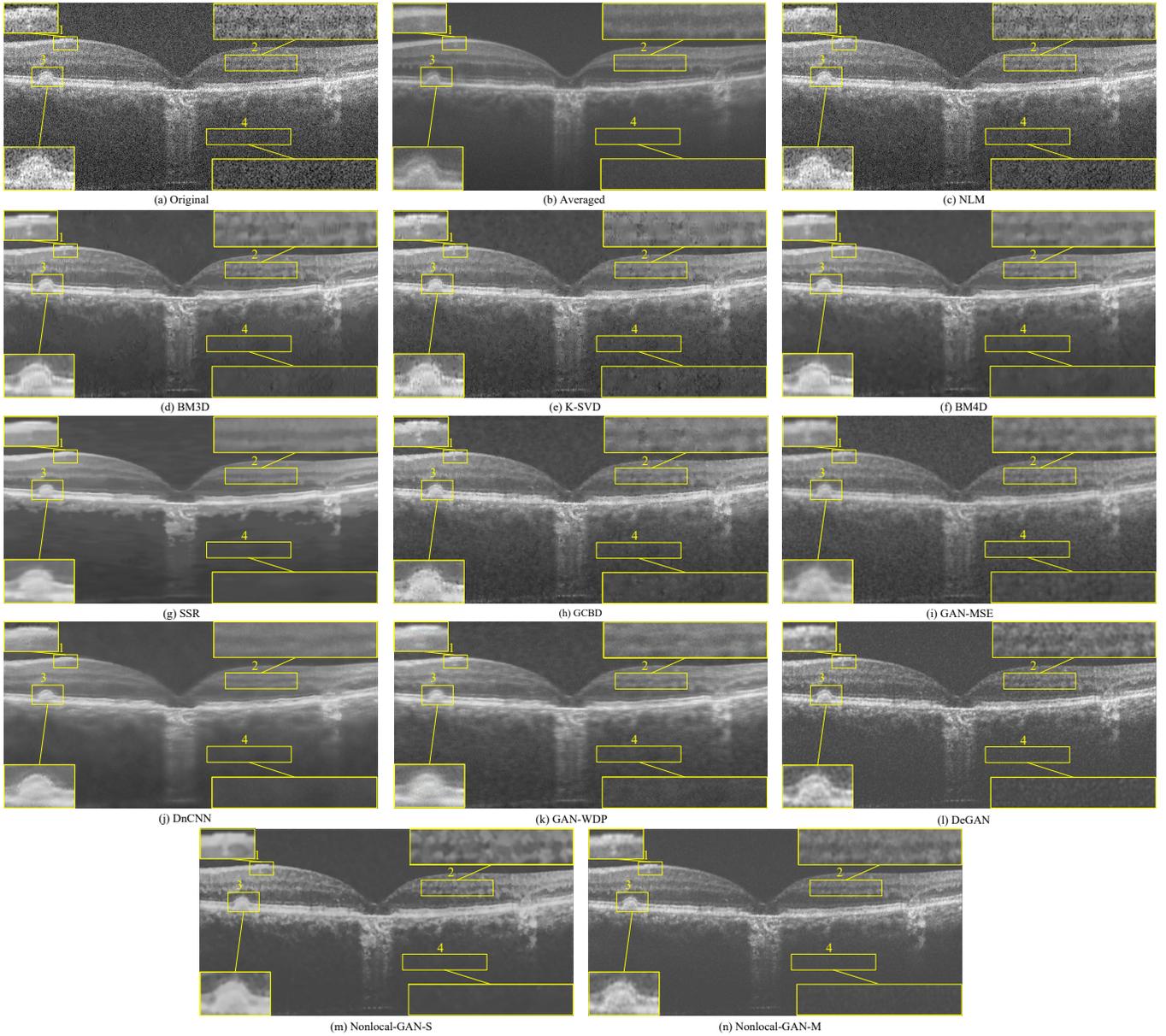


Fig. 5. The first testing OCT image selected from the first dataset and its denoising results using the NLM [19], BM3D [20], K-SVD [22], BM4D [21], SSR [25], GCBD [39], GAN-MSE [40], DnCNN [35], GAN-WDP [41], DeGAN, the proposed Nonlocal-GAN-S and Nonlocal-GAN-M methods.

denoising performance when the reference clean images are not available [8], [52]. The SNR can be calculated as

$$SNR = 20 \log (I_{max} / \sigma_B), \quad (15)$$

where I_{max} is the maximum value of the denoised image, and σ_B is the standard deviation of noise in the background region.

2) *Equivalent number of looks (ENL)*: ENL [8], [52] is a commonly used performance measure for speckle suppression, which measures smoothness in regions that appear to be homogeneous. Here, we calculate it only in background region, as

$$ENL = \frac{\mu_B^2}{\sigma_B^2}, \quad (16)$$

where μ_B and σ_B denote mean value and standard deviation of the background region, respectively. A larger ENL indicates a stronger speckle smoothing in the corresponding region.

3) *Contrast to Noise Ratio (CNR)*: CNR [8], [25], [52] is also a nonreference metric for speckle suppression, which measures the contrast between regions of interest and background noise. In the m -th region of interest (ROI), CNR can be calculated as

$$CNR_m = 10 \log \frac{\mu_m - \mu_B}{\sqrt{\sigma_m^2 + \sigma_B^2}}, \quad (17)$$

where μ_m and σ_m^2 denote mean value and variance of m -th ROI, μ_B and σ_B^2 denote mean value and variance of the background region, respectively. We average the CNR measure over the M ROIs:

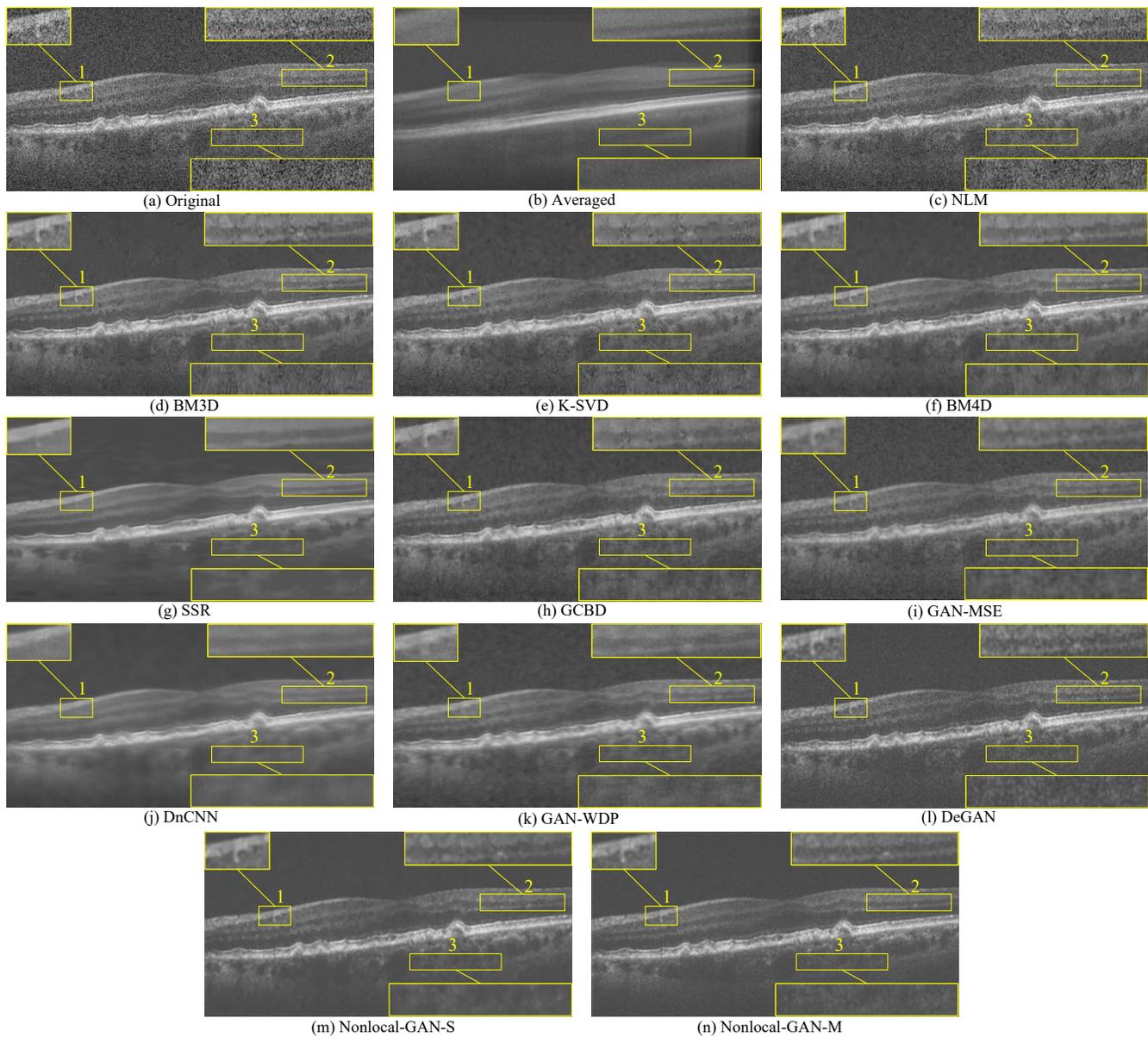


Fig. 6. The second testing OCT image selected from the first dataset and its denoising results using the NLM [19], BM3D [20], K-SVD [22], BM4D [21], SSR [25], GCBD [39], GAN-MSE [40], DnCNN [35], GAN-WDP [41], DeGAN, the proposed Nonlocal-GAN-S and Nonlocal-GAN-M methods.

$$CNR = \frac{1}{M} \sum_{m=1}^M CNR_m. \quad (18)$$

D. Results Comparison

For the first dataset, we used 2 volumes for training (with patch size 33×33 and stride 6), and the other 16 volumes for testing. Here, we show two visual comparisons of testing OCT images. For the first testing image, Fig. 5 provides visual comparison of denoising results obtained from the NLM [19], BM3D [20], K-SVD [22], BM4D [21], SSR [25], GCBD [39], GAN-MSE [40], DnCNN [35], GAN-WDP [41], DeGAN, Nonlocal-GAN-S and Nonlocal-GAN-M methods. As can be observed, the results from the NLM method exhibit serious noisy appearance. The BM3D, K-SVD, and BM4D methods

significantly remove noise, but show some visible artifacts (see the areas in boxes #3 and #4 in Fig. 5). The SSR and DnCNN methods can get clean images without visible artifacts, but the results seem to be over-smooth (see the areas in boxes #1 and #3 in Fig. 5). The results of GCBD, GAN-MSE, GAN-WDP and DeGAN methods still have some noise in the denoised image (see the areas in boxes #2 and #3 in Fig. 5). By contrast, our two methods Nonlocal-GAN-S and Nonlocal-GAN-M can obtain clearer results and preserve more details (see the areas in boxes #1-#4 in Fig. 5). For the second testing OCT image, Fig. 6 shows visual comparison of denoising results obtained from different methods. Similarly, the Nonlocal-GAN-S and Nonlocal-GAN-M methods can get better visual performance compared with other methods (see the areas in boxes #1-#3 of the Fig. 6).

Average quantitative results (over 16 foveal images in the

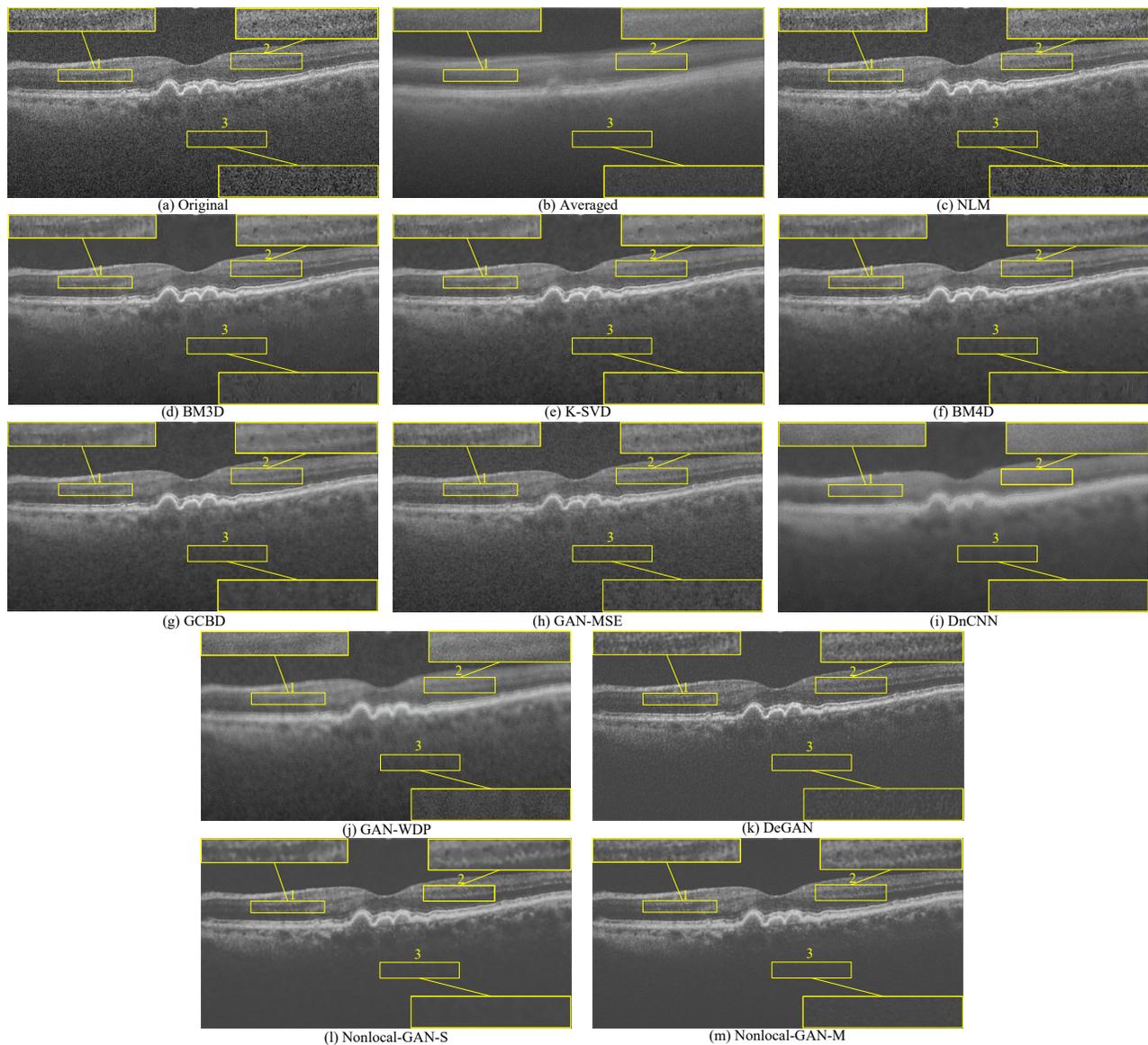


Fig. 7. The first testing OCT image selected from the second dataset and its denoising results using the NLM [19], BM3D [20], K-SVD [22], BM4D [21], GCBD [39], GAN-MSE [40], DnCNN [35], GAN-WDP [41], DeGAN, the proposed Nonlocal-GAN-S and Nonlocal-GAN-M methods.

first dataset) of all the test methods are tabulated in the Table I. The proposed Nonlocal-GAN-S and Nonlocal-GAN-M methods consistently outperform the compared methods in terms of three evaluation metrics (i.e., SNR, ENL and CNR). The results show that the Nonlocal-GAN-S and Nonlocal-GAN-M methods have higher ENL, which means that our methods can better remove the noise compared with other methods. In addition, our methods can obtain high performance in terms of SNR and CNR. In our two methods, though the Nonlocal-GAN-S has better quantitative results, by utilizing the similarity information of several continuous B-scans, the results of Nonlocal-GAN-M have clearer layer structures than that of Nonlocal-GAN-S. Furthermore, the averaged running time (over 16 foveal images) of all the test methods are also reported in Table I. All the compared methods as well as the proposed two methods Nonlocal-GAN-S and Nonlocal-GAN-

M were executed on a desktop PC with an i7-6850K CPU at 3.60 GHz, 64 GB of RAM and a NVIDIA Titan X Pascal GPU. The proposed Nonlocal-GAN-S and Nonlocal-GAN-M methods only require 0.11 and 0.66 seconds to denoise one B-scan, which benefit from their end-to-end architectures and GPU's acceleration.

For the second dataset, since each volume consists of 11 B-scans, we just used 1 volume for training (with patch size 33×33 and stride 6), and the other 19 volumes for testing. As subjects in this dataset have severe AMD disease, the layers are very hard to be segmented and thus the layer segmented based SSR method is not used for comparison. For the testing images, Figs. 7 and 8 provide visual comparisons of denoising results obtained from the NLM [19], BM3D [20], K-SVD [22], BM4D [21], GCBD [39], GAN-MSE [40], DnCNN [35], GAN-WDP [41], DeGAN, our Nonlocal-

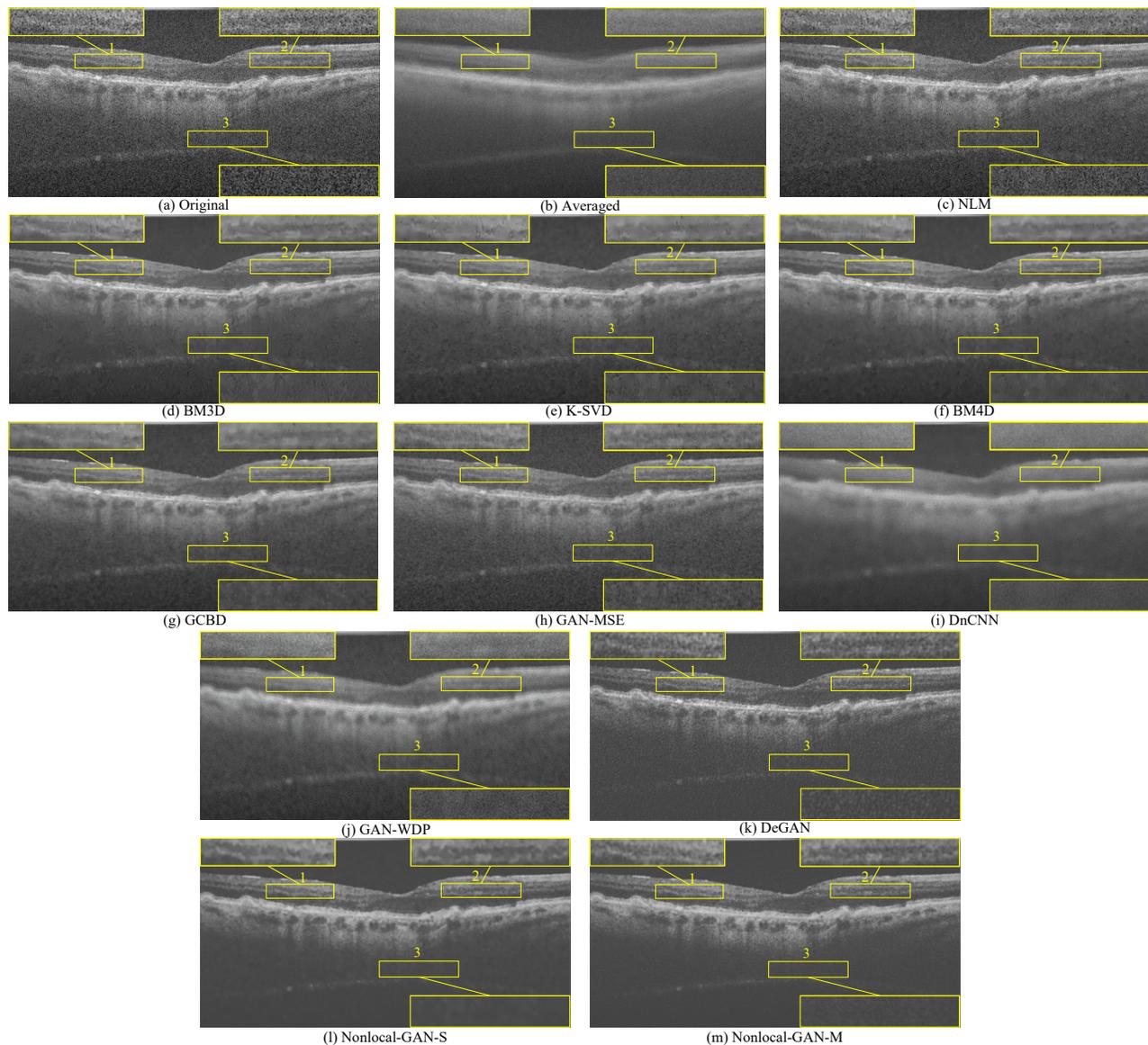


Fig. 8. The second testing OCT image selected from the second dataset and its denoising results using the NLM [19], BM3D [20], K-SVD [22], BM4D [21], GCBD [39], GAN-MSE [40], DnCNN [35], GAN-WDP [41], DeGAN, the proposed Nonlocal-GAN-S and Nonlocal-GAN-M methods.

GAN-S and Nonlocal-GAN-M methods. As can be seen, the BM4D, Nonlocal-GAN-S and Nonlocal-GAN-M can get better visual performance, and the results of Nonlocal-GAN-M show more layer information. Average quantitative results (over 19 foveal images in the second dataset) of all the test methods are tabulated in the Table II. Similarly, the results of Nonlocal-GAN-S and Nonlocal-GAN-M methods outperform the compared methods.

E. Effect of Template Search Patch Size

In this section, we analyze the effect of the template search patch size on the performance (in the SNR) of the proposed Nonlocal-GAN-S and Nonlocal-GAN-M methods. For the Nonlocal-GAN-S method, the patch size was varied from 3×3 to 9×9 . The performance of the proposed Nonlocal-GAN-S method under varied patch sizes is shown in Fig. 9(a). For

the Nonlocal-GAN-M method, the patch size was varied from 3×6 to 7×14 . The performance of the proposed Nonlocal-GAN-M method under varied patch sizes is shown in Fig. 9(b). As can be seen, the different patch sizes have little effect on our experimental results, which demonstrates that the proposed methods are generally robust to the size of the selected similar patches.

V. CONCLUSIONS

In this paper, we presented a nonlocal-generative adversarial network (Nonlocal-GAN) for 3D OCT image denoising. By learning the characteristics of the noise in the original noisy OCT images, the Nonlocal-GAN can be trained in an unsupervised way. Experiments demonstrated the superiority of the proposed methods over several state-of-the-art denoising approaches. Meanwhile, the proposed model can also get

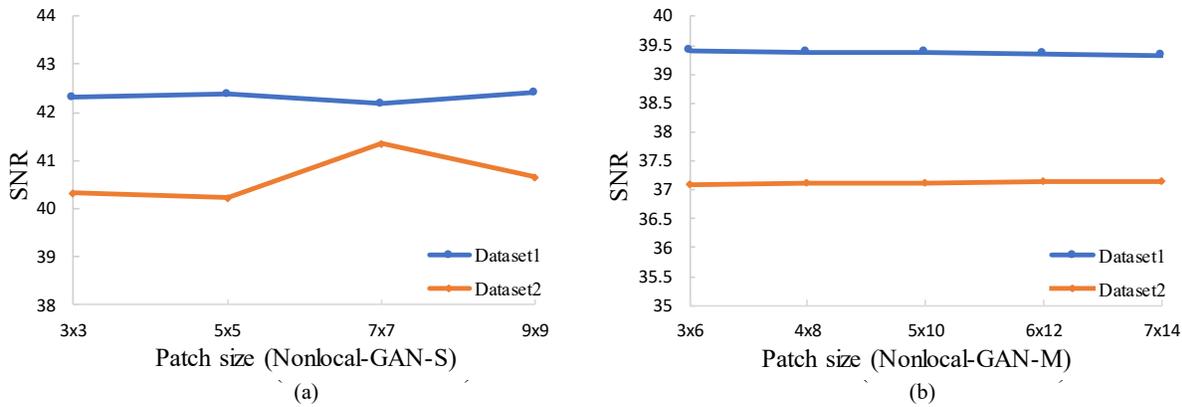


Fig. 9. Effect of patch size on the performance of (a) Nonlocal-GAN-S and (b) Nonlocal-GAN-M methods.

TABLE I
QUANTITATIVE RESULTS IN AVERAGED SNR, ENL, CNR, TOGETHER WITH RUNNING TIME (IN SECONDS) FOR 16 FOVEAL IMAGES FROM THE FIRST DATASET RECONSTRUCTED BY DIFFERENT METHODS

Method	SNR	ENL	CNR	Running Time
NLM [19]	20.13	5.71	3.66	101.25
BM3D [20]	29.95	62.90	6.83	3.69
K-SVD [22]	29.73	60.04	7.01	2.64
BM4D [21]	30.85	79.96	7.13	3.87
SSR [25]	31.13	94.92	7.77	1.40
G CBD [39]	31.68	95.28	6.98	0.24
GAN-MSE [40]	28.20	59.53	6.77	0.10
DnCNN [35]	32.06	123.09	7.38	0.24
GAN-WDP [41]	29.80	65.55	7.00	0.04
DeGAN	35.74	384.36	5.78	0.04
Nonlocal-GAN-S	42.35	1990.70	8.08	0.11
Nonlocal-GAN-M	39.37	913.29	7.80	0.66

TABLE II
QUANTITATIVE RESULTS IN AVERAGED SNR, ENL, CNR, TOGETHER WITH RUNNING TIME (IN SECONDS) FOR 19 FOVEAL IMAGES FROM THE SECOND DATASET RECONSTRUCTED BY DIFFERENT METHODS

Method	SNR	ENL	CNR	Running Time
NLM [19]	20.06	5.24	3.79	127.95
BM3D [20]	29.88	57.76	6.47	4.74
K-SVD [22]	29.33	53.15	6.57	3.12
BM4D [21]	31.00	76.13	6.64	5.08
G CBD [39]	32.16	99.31	6.66	0.28
GAN-MSE [40]	24.93	23.41	5.68	0.12
DnCNN [35]	28.36	57.21	7.05	0.28
GAN-WDP [41]	24.86	18.82	5.90	0.06
DeGAN	33.77	214.48	5.43	0.06
Nonlocal-GAN-S	40.14	981.29	7.41	0.14
Nonlocal-GAN-M	37.12	465.28	7.04	0.86

better visual performance compared with some deep learning based blind image denoising methods. Our approach can effectively remove noise while preserving more useful details and clearer layer structures, which are critical for physicians' precise diagnosis. In addition, high efficiency in the denoising process shows that our methods can be easily applied on the real OCT imaging systems.

In our future works, we will try to improve the performance of the discriminator in the Nonlocal-GAN, since a well-trained discriminator can guide the generator for denoising more

effectively. In addition, we will also extend our denoising models to other real noise datasets, where the reference clean images are not available.

ACKNOWLEDGMENT

The authors thank the editors and all of the anonymous reviewers for their constructive feedback and criticisms.

REFERENCES

- [1] D. Huang, E. A. Swanson, C. P. Lin, J. S. Schuman, W. G. Stinson, W. Chang, M. R. Hee, T. Flotte, K. Gregory, C. A. Puliafito, and J. G. Fujimoto, "Optical coherence tomography," *Science*, vol. 254, no. 5035, pp. 1178–1181, Nov. 1991.
- [2] Y. Mao, C. Flueraru, S. Chang, D. P. Popescu, and M. G. Sowa, "High-quality tissue imaging using a catheter-based swept-source optical coherence tomography systems with an integrated semiconductor optical amplifier," *IEEE Trans. Instrum. Meas.*, vol. 60, no. 10, pp. 3376–3383, Oct. 2011.
- [3] J.-H. Han and S. M. Yoon, "Integrated near-infrared dual-modality imaging with common path fiber optic system," *IEEE Trans. Instrum. Meas.*, vol. 60, no. 12, pp. 3958–3961, Dec. 2011.
- [4] S. Lu, C. Y.-l. Cheung, J. Liu, J. H. Lim, C. K.-s. Leung, and T. Y. Wong, "Automated layer segmentation of optical coherence tomography images," *IEEE Trans. Biomed. Eng.*, vol. 57, no. 10, pp. 2605–2608, Oct. 2010.
- [5] G. R. Wilkins, O. M. Houghton, and A. L. Oldenburg, "Automated segmentation of intraretinal cystoid fluid in optical coherence tomography," *IEEE Trans. Biomed. Eng.*, vol. 59, no. 4, pp. 1109–1114, Apr. 2012.
- [6] Y. Sun and C. Xue, "Automated diagnosing of nevus flammeus using oct raw signal," in *Proc. Int. Conf. Comput. Inf. Appl.*, 2010, pp. 453–456.
- [7] A. B. Wu, E. Lebed, M. V. Sarunic, and M. F. Beg, "Quantitative evaluation of transform domains for compressive sampling-based recovery of sparsely sampled volumetric oct images," *IEEE Trans. Biomed. Eng.*, vol. 60, no. 2, pp. 470–478, Feb. 2012.
- [8] A. Ozcan, A. Bilencu, A. E. Desjardins, B. E. Bouma, and G. J. Tearney, "Speckle reduction in optical coherence tomography images using digital filtering," *J. Opt. Soc. Am.*, vol. 24, no. 7, pp. 1901–1910, Jul. 2007.
- [9] W. Zhao and H. Lu, "Medical image fusion and denoising with alternating sequential filter and adaptive fractional order total variation," *IEEE Trans. Instrum. Meas.*, vol. 66, no. 9, pp. 2283–2294, Sep. 2017.
- [10] F. Attivissimo, G. Cavone, A. M. L. Lanzolla, and M. Spadavecchia, "A technique to improve the image quality in computer tomography," *IEEE Trans. Instrum. Meas.*, vol. 59, no. 5, pp. 1251–1257, May. 2010.
- [11] X. Liu and J. U. Kang, "Compressive sd-oct: the application of compressed sensing in spectral domain optical coherence tomography," *Opt. Exp.*, vol. 18, no. 21, pp. 22010–22019, Oct. 2010.
- [12] D. Xu, N. Vaswani, Y. Huang, and J. U. Kang, "Modified compressive sensing optical coherence tomography with noise reduction," *Opt. Lett.*, vol. 37, no. 20, pp. 4209–4211, Aug. 2012.

- [13] J. Cheng, D. Tao, Y. Quan, D. W. K. Wong, G. C. M. Cheung, M. Akiba, and J. Liu, "Speckle reduction in 3d optical coherence tomography of retina by a-scan reconstruction," *IEEE Trans. Instrum. Meas.*, vol. 35, no. 10, pp. 2270–2279, Oct. 2016.
- [14] J. Cheng, L. Duan, D. W. K. Wong, M. Akiba, and J. Liu, "Speckle reduction in optical coherence tomography by matrix completion using bilateral random projection," in *Proc. Int. Conf. IEEE Eng. Med. Bio. Soc.*, 2014, pp. 186–189.
- [15] H. Rabbani, R. Nezafat, and S. Gazor, "Wavelet-domain medical image denoising using bivariate laplacian mixture model," *IEEE Trans. Biomed. Eng.*, vol. 56, no. 12, pp. 2826–2837, Dec. 2009.
- [16] H. Rabbani, M. Vafadust, P. Abolmaesumi, and S. Gazor, "Speckle noise reduction of medical ultrasound images in complex wavelet domain using mixture priors," *IEEE Trans. Biomed. Eng.*, vol. 55, no. 9, pp. 2152–2160, Sep. 2008.
- [17] S. Chitchian, M. A. Fiddy, and N. M. Fried, "Denoising during optical coherence tomography of the prostate nerves via wavelet shrinkage using dual-tree complex wavelet transform," *J. Biomed. Opt.*, vol. 14, no. 1, pp. 14–31, Jan. 2009.
- [18] Z. Jian, L. Yu, B. Rao, B. J. Tromberg, and Z. Chen, "Three-dimensional speckle suppression in optical coherence tomography based on the curvelet transform," *Opt. Exp.*, vol. 18, no. 2, pp. 1024–1032, Jan. 2010.
- [19] A. Buades, B. Coll, and J.-M. Morel, "A non-local algorithm for image denoising," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, vol. 2, 2005, pp. 60–65.
- [20] K. Dabov, A. Foi, V. Katkovnik, and K. Egiazarian, "Image denoising by sparse 3-d transform-domain collaborative filtering," *IEEE Trans. Image Process.*, vol. 16, no. 8, pp. 2080–2095, Aug. 2007.
- [21] M. Maggioni, V. Katkovnik, K. Egiazarian, and A. Foi, "Nonlocal transform-domain filter for volumetric data denoising and reconstruction," *IEEE Trans. Image Process.*, vol. 22, no. 1, pp. 119–133, Jan. 2012.
- [22] M. Elad and M. Aharon, "Image denoising via sparse and redundant representations over learned dictionaries," *IEEE Trans. Image Process.*, vol. 15, no. 12, pp. 3736–3745, Dec. 2006.
- [23] L. Fang, S. Li, Q. Nie, J. A. Izatt, C. A. Toth, and S. Farsiu, "Sparsity based denoising of spectral domain optical coherence tomography images," *Biomed. Opt. Exp.*, vol. 3, no. 5, pp. 927–942, May. 2012.
- [24] L. Fang, S. Li, R. P. McNabb, Q. Nie, A. N. Kuo, C. A. Toth, J. A. Izatt, and S. Farsiu, "Fast acquisition and reconstruction of optical coherence tomography images via sparse representation," *IEEE Trans. Med. Imag.*, vol. 32, no. 11, pp. 2034–2049, Nov. 2013.
- [25] L. Fang, S. Li, D. Cunefare, and S. Farsiu, "Segmentation based sparse reconstruction of optical coherence tomography images," *IEEE Trans. Med. Imag.*, vol. 36, no. 2, pp. 407–421, Feb. 2017.
- [26] A. Krizhevsky, I. Sutskever, and G. E. Hinton, "Imagenet classification with deep convolutional neural networks," in *Proc. Adv. Neural Inf. Process. Syst.*, 2012, pp. 1097–1105.
- [27] K. Simonyan and A. Zisserman, "Very deep convolutional networks for large-scale image recognition," in *Proc. Int. Conf. Learn. Representations*, 2015, pp. 106–120.
- [28] J. Long, E. Shelhamer, and T. Darrell, "Fully convolutional networks for semantic segmentation," in *Proc. IEEE Int. Conf. Comput. Vis. Pattern Recognit.*, 2015, pp. 3431–3440.
- [29] K. He, X. Zhang, S. Ren, and J. Sun, "Deep residual learning for image recognition," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, 2016, pp. 770–778.
- [30] J. Lei, Q. Liu, and X. Wang, "Deep learning-based inversion method for imaging problems in electrical capacitance tomography," *IEEE Trans. Instrum. Meas.*, vol. 67, no. 9, pp. 2107–2118, Sep. 2018.
- [31] B. Taji, A. D. Chan, and S. Shirmohammadi, "False alarm reduction in atrial fibrillation detection using deep belief networks," *IEEE Trans. Instrum. Meas.*, vol. 67, no. 5, pp. 1124–1131, May. 2017.
- [32] G. Jiang, H. He, P. Xie, and Y. Tang, "Stacked multilevel-denoising autoencoders: A new representation learning approach for wind turbine gearbox fault diagnosis," *IEEE Trans. Instrum. Meas.*, vol. 66, no. 9, pp. 2391–2402, Sep. 2017.
- [33] V. Jain and S. Seung, "Natural image denoising with convolutional networks," in *Proc. Adv. Neural Inf. Process. Syst.*, 2009, pp. 769–776.
- [34] H. C. Burger, C. J. Schuler, and S. Harmeling, "Image denoising: Can plain neural networks compete with BM3D?" in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, 2012, pp. 2392–2399.
- [35] K. Zhang, W. Zuo, Y. Chen, D. Meng, and L. Zhang, "Beyond a gaussian denoiser: Residual learning of deep cnn for image denoising," *IEEE Trans. Image Process.*, vol. 26, no. 7, pp. 3142–3155, Jul. 2017.
- [36] K. Zhang, W. Zuo, and L. Zhang, "FFDNet: Toward a fast and flexible solution for cnn-based image denoising," *IEEE Trans. Image Process.*, vol. 27, no. 9, pp. 4608–4622, Sep. 2018.
- [37] S. Guo, Z. Yan, K. Zhang, W. Zuo, and L. Zhang, "Toward convolutional blind denoising of real photographs," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, 2019, pp. 1712–1722.
- [38] A. Abbasi, A. Monadjemi, L. Fang, H. Rabbani, and Y. Zhang, "Three-dimensional optical coherence tomography image denoising through multi-input fully-convolutional networks," *Comput. Biol. Med.*, vol. 108, no. 4, pp. 1–8, May. 2019.
- [39] J. Chen, J. Chen, H. Chao, and M. Yang, "Image blind denoising with generative adversarial network based noise modeling," in *Proc. IEEE Int. Conf. Comput. Vis. Pattern Recognit.*, 2018, pp. 3155–3164.
- [40] J. M. Wolterink, T. Leiner, M. A. Viergever, and I. Išgum, "Generative adversarial networks for noise reduction in low-dose CT," *IEEE Trans. Med. Imag.*, vol. 36, no. 12, pp. 2536–2545, Dec. 2017.
- [41] Q. Yang, P. Yan, Y. Zhang, H. Yu, Y. Shi, X. Mou, M. K. Kalra, Y. Zhang, L. Sun, and G. Wang, "Low-dose CT image denoising using a generative adversarial network with Wasserstein distance and perceptual loss," *IEEE Trans. Med. Imag.*, vol. 37, no. 6, pp. 1348–1357, Jun. 2018.
- [42] S. Anwar and N. Barnes, "Real image denoising with feature attention," [Online]. Available: <https://arxiv.org/abs/1904.07396>
- [43] S. Ioffe and C. Szegedy, "Batch normalization: Accelerating deep network training by reducing internal covariate shift," in *Proc. Int. Conf. Mach. Learn.*, 2015, pp. 448–456.
- [44] I. Goodfellow, J. Pouget-Abadie, M. Mirza, B. Xu, D. Warde-Farley, S. Ozair, A. Courville, and Y. Bengio, "Generative adversarial nets," in *Proc. Adv. Neural Inf. Process. Syst.*, 2014, pp. 2672–2680.
- [45] M. Arjovsky, S. Chintala, and L. Bottou, "Wasserstein generative adversarial networks," in *Proc. Int. Conf. Mach. Learn.*, 2017, pp. 214–223.
- [46] A. Buades, B. Coll, and J.-M. Morel, "A review of image denoising algorithms, with a new one," *Multiscale Model. Simul.*, vol. 4, no. 2, pp. 490–530, 2005.
- [47] J. T. Springenberg, A. Dosovitskiy, T. Brox, and M. Riedmiller, "Striving for simplicity: The all convolutional net," 2015, pp. 200–207.
- [48] M. Abadi, P. Barham, J. Chen, Z. Chen, A. Davis, J. Dean, M. Devin, S. Ghemawat, G. Irving, M. Isard *et al.*, "Tensorflow: A system for large-scale machine learning," [Online]. Available: <https://arxiv.org/abs/1603.04467>
- [49] S. Farsiu, S. J. Chiu, R. V. O'Connell, F. A. Folgar, E. Yuan, J. A. Izatt, C. A. Toth, A.-R. E. D. S. . A. S. D. O. C. T. S. Group *et al.*, "Quantitative classification of eyes with and without intermediate age-related macular degeneration using optical coherence tomography," *Ophthalmology*, vol. 121, no. 1, pp. 162–172, Jan. 2014.
- [50] S. J. Chiu, J. A. Izatt, R. V. O'Connell, K. P. Winter, C. A. Toth, and S. Farsiu, "Validated automatic segmentation of amd pathology including drusen and geographic atrophy in sd-oct images," *Invest. Ophthalmol. Vis. Sci.*, vol. 53, no. 1, pp. 53–61, Jan. 2012.
- [51] Z. Wang, A. C. Bovik, H. R. Sheikh, and E. P. Simoncelli, "Image quality assessment: from error visibility to structural similarity," *IEEE Trans. Image Process.*, vol. 13, no. 4, pp. 600–612, Apr. 2004.
- [52] A. Pizurica, L. Jovanov, B. Huysmans, V. Zlokolica, P. De Keyser, F. Dhaenens, and W. Philips, "Multiresolution denoising for optical coherence tomography: a review and evaluation," *Curr. Med. Imag. Rev.*, vol. 4, no. 4, pp. 270–284, Apr. 2008.