Research Article | Vol. 9, No. 11 | 1 Nov 2018 | BIOMEDICAL OPTICS EXPRESS 5129
Biomedical Optics EXPRESS

# Speckle noise reduction in optical coherence tomography images based on edge-sensitive cGAN

YUHUI MA,[1,4] XINJIAN CHEN,[1,2,4] WEIFANG ZHU,[1,2,3] XUENA CHENG,[1] DEHUI XIANG,[1,2] AND FEI SHI[1,2,3,*]

[1]School of Electronic and Information Engineering, Soochow University, Suzhou 215006, China
[2]State Key Laboratory of Radiation Medicine and Protection, Soochow University, Suzhou 215123, China
[3]Collaborative Innovation Center of IoT Technology and Intelligent Systems, Minjiang University, Fuzhou 350108, China
[4]contributed equally
*shifei@suda.edu.cn

**Abstract:** Speckle noise in optical coherence tomography (OCT) impairs both the visual quality and the performance of automatic analysis. Edge preservation is an important issue for speckle reduction. In this paper, we propose an end-to-end framework for simultaneous speckle reduction and contrast enhancement for retinal OCT images based on the conditional generative adversarial network (cGAN). The edge loss function is added to the final objective so that the model is sensitive to the edge-related details. We also propose a novel method for obtaining clean images for training from outputs of commercial OCT scanners. The results show that the overall denoising performance of the proposed method is better than other traditional methods and deep learning methods. The proposed model also has good generalization ability and is capable of despeckling different types of retinal OCT images.



## 1 Introduction

Optical coherence tomography (OCT) generates cross-sectional images of ocular biological tissue in micron resolution [1] and has become an essential tool for imaging of retina. Speckle noise, caused by multiple forward and backward scattering of light waves, is the main quality degrading factor in OCT images. The presence of speckle noise often obscures subtle but important morphological details and thus is detrimental to clinical diagnosis. It also affects the performance of automatic analysis methods intended for objective and accurate quantifications. Although the imaging resolution, speed and depth of OCT has been greatly improved over the last two decades, speckle noise, as an intrinsic problem to the imaging technique, has not been well solved. The most common despeckling approach adopted in commercial scanners is Bscan averaging. A high quality image can be obtained by averaging registered multiple Bscans acquired from the same position. However, this approach is currently impractical for 3D scans due to the long acquisition time for overlapping Bscans. In this paper, we focus on another category of speckle reduction techniques, which utilize software-based image processing algorithms to reconstruct an enhanced image.

A large number of image processing algorithms for OCT denoising have been proposed so far, which can be roughly divided into several categories with some overlapping: the partial differential equation (PDE) based methods such as anisotropic diffusion filtering [2,3], block matching based methods such as non-local means (NLM) [4,5] or block matching and 3D filtering (BM3D) [6], sparse transform based methods based on wavelets [7,8], curvelets [9], or dictionary learning [10,11], statistical model based methods [12–14], and low rank decomposition based methods [15,16].

#340602 https://doi.org/10.1364/BOE.9.005129
Journal © 2018 Received 27 Jul 2018; revised 7 Sep 2018; accepted 16 Sep 2018; published 2 Oct 2018

Biomedical Optics EXPRESS

Though the main aim of OCT denoising is to reduce the grainy appearance in homogeneous areas, another important issue is preservation of image details, especially the edges, because edges are the most vital information needed for both visual inspection and automatic analysis such as segmentation. As the noise level is high in OCT images, many spatial filters tend to oversmooth the image, resulting in reduced contrast at the edges. Block matching based methods can result in edge distortions caused by disagreement of the edges in different blocks. Transform-based methods also tend to produce artifacts with the shape of transform basis near the edges.

Recently, deep learning provides new ideas for image denoising. Mao et al. [17] proposed very deep convolutional encoder-decoder networks (RED-Net) with symmetric skip connections. Tai et al. [18] proposed a persistent memory network (MemNet). Zhang et al. [19] proposed residual learning of deep convolutional neural network (DnCNN) for natural image denoising. The network was designed to predict the residual image from the noisy input. Cai et al. [20] borrowed the idea, improved the network structure with residue module and applied it to OCT image denoising. However, in all these works, the additive noise assumption is used. Therefore these models are not most suitable for speckle noise in OCT images.

In this paper, we aim to remove speckle noise in Bscans from 3D OCT volumes exported from commercial retinal OCT scanners. Figure 1 shows the flowchart of the proposed method. We treat image denoising as an image-to-image translation problem, and propose a method based on conditional generative adversarial network (cGAN) [21] to achieve the goal. Trained by noisy images and corresponding high quality images obtained by registration and averaging, and with the competition of the generator and the discriminator, the network is able to learn the underlying clean structures of retinas. To our best knowledge, it is the first time that the image-to-image cGAN network is applied to OCT speckle noise reduction. The contributions of our work are listed as follows.

- We introduce a new edge loss into the objective function of cGAN and make the network sensitive to the edge information, thus achieving good edge preservation while smoothing the homogeneous areas.
- We propose a method for obtaining high quality training images that works for common users of commercial OCT scanners.
- By preprocessing of the training images, we make the deep network an end-to-end framework that achieves simultaneous speckle noise reduction and contrast enhancement.
- By data augmentation, we make the deep network capable of handling both OCT image from normal and pathological subjects, and also data from different types of scanners.

## 2 Methods

### *2.1 Overview of conditional adversarial networks*

The conditional adversarial networks have been proved a good image-to-image translation model for tasks such as label-to-photo conversion, colorization, and semantic segmentation [21]. Different from the original GAN which learn a mapping from random noise to output image, the cGAN generates the output image conditioned on an observed image. cGAN consists of two modules with opposite goals: the generator G that extracts features of the observed image x and produces the corresponding fake image $y_{fake}$, and the discriminator D that classifies between real pairs (x, $y_{real}$) and fake pairs (x, $y_{fake}$). The model structure is illustrated in Fig. 2. Essentially, cGAN aims to learn a mapping from x and random vector z

Biomedical Optics EXPRESS

to the ground truth $y_{real}$: G: {x,z}→y under the competition between the generator G and the discriminator D.

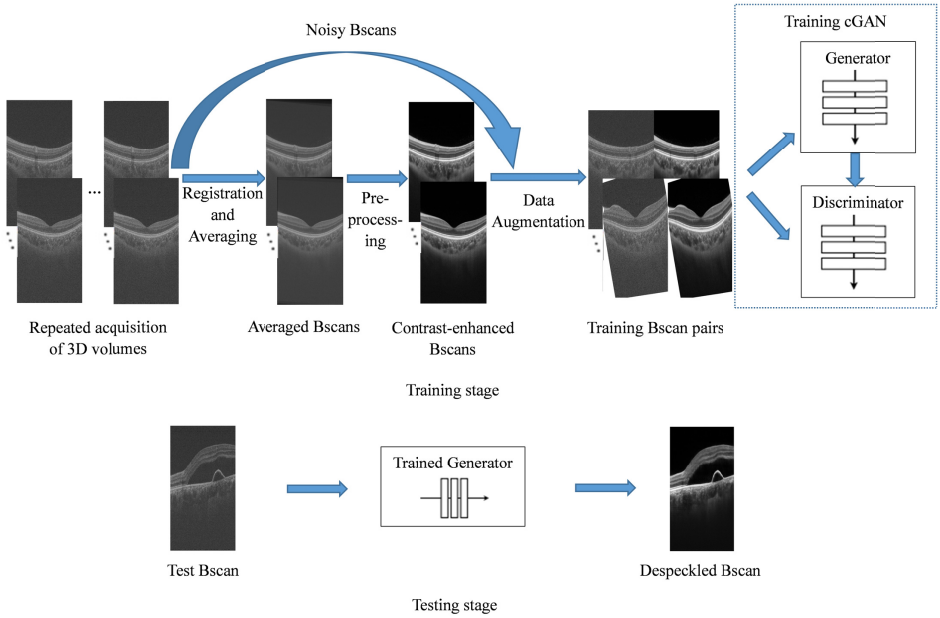


Fig. 1. Flowchart of the proposed speckle noise reduction method.

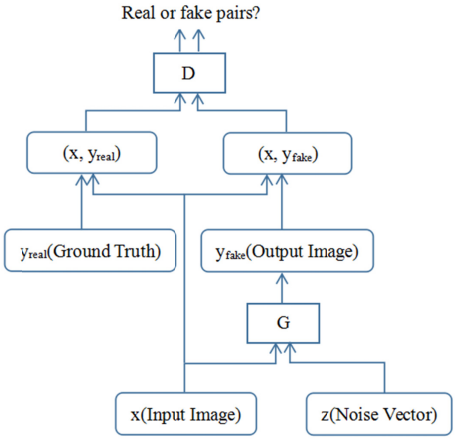


Fig. 2. Model structure of cGAN. G tries to generate fake images that fool D, while D tries to identify the fake pairs.

### *2.2 Objective function*

The objective function of cGAN can be described as follows [21]:

$$L_{cGAN}(G,D) = E_{x,y\sim p_{data}(x,y)}\left[\log D(x,y)\right] + E_{x\sim p_{data}(x),z\sim p_z(z)}\left[\log\left(1-D\left(x,G(x,z)\right)\right)\right] \quad (1)$$

where $x$ is the input observed image, $z$ is random noise, $y$ is the corresponding ground truth for $x$, $D(x,y)$ is the output of the discriminator, and $G(x,z)$ is the output image of the generator. The first term is the expectation calculated over all training image pairs, and the

Biomedical Optics EXPRESS

second term is the expectation calculated over all training observed images and random vectors.

In training, G tries to minimize the objective against D that tries to maximize it, resulting in the following optimization:

$$G^* = \arg\min_G \max_D L_{cGAN}(G, D) \tag{2}$$

where $G^*$ represents the resulted optimized generator.

Previous approaches to cGAN have found it beneficial to combine the cGAN's objective with a traditional loss, such as L1 or L2 distance, so that the generated image is more similar to the ground truth. L1 distance encourages errors that are sparsely distributed in space while L2 distance encourages errors that are uniformly distributed in space. Therefore L1 distance results in less blurring than L2 distance [21].

$$L_{L1}(G) = E_{x,y\sim p_{data}(x,y),z\sim p_z(z)}\left[\left\|y - G(x,z)\right\|_1\right] \tag{3}$$

By adding the L1 distance, the optimization becomes

$$G^* = \arg\min_G \max_D L_{cGAN}(G, D) + \alpha L_{L1}(G) \tag{4}$$

where $\alpha$ is a weighting parameter.

In this paper, we further modified the objective function by adding a loss that is explicitly related to the edge information, to deal with the difficulty of edge-preserving while despeckling. The edge loss is defined as follows:

$$L_{Edge}(G) = E_{x,y\sim p_{data}(x,y),z\sim p_z(z)}\left[-\log\frac{\sum_{i,j}\left|G(x,z)_{i+1,j} - G(x,z)_{i,j}\right|}{\sum_{i,j}\left|y_{i+1,j} - y_{i,j}\right|}\right] \tag{5}$$

where $i$ and $j$ represent coordinates in the longitudinal and lateral direction in the B-scan image. The edge loss measures the edge similarity between generated image and the ground truth, which is inspired by the edge preservation index (EPI). As the retina has a layered structure, the longitudinal gradient is more important than the lateral one. Therefore, considering the simplicity of the model, only longitudinal gradient is used in calculating the edge loss.

Thus the final optimization is performed as:

$$G^* = \arg\min_G \max_D L_{cGAN}(G, D) + \alpha L_{L1}(G) + \beta L_{Edge}(G) \tag{6}$$

where $\alpha$ and $\beta$ are the weighting parameters.

### 2.3 Implementation of cGAN

In this paper, the "U-Net" [22], a kind of encoder-decoder structure with skip connections between symmetric layers in the encoder and decoder stacks, is used as the main framework of the generator, and PatchGAN, that identifies real or fake pairs based on patches in an image, is adopted as the discriminator architecture. Modules of the form Convolution-BatchNorm-ReLU [23] are the basic components of both generator and discriminator. Detailed structures are given in the Appendix.

In view of our task of despeckling OCT images, the despeckled OCT image shares the structure information with the corresponding noisy OCT image, which requires the structure of the output image of the generator remains aligned with that of the input image. This is a mapping problem from a high resolution input grid to a high resolution output grid.

Biomedical Optics EXPRESS

Symmetric skip connections of U-Net provides an effective solution for the problem, which helps to produce the despeckled OCT image with more details due to the combination of low-level and high-level information.

In general, the discriminator in GANs outputs the probability that its input is real based on the whole image. Different from traditional discriminators, PatchGAN tries to identify whether each p × p patch in an image is real or fake. Such a discriminator regards the image as a Markov random field assuming independence between pixels separated by more than a patch diameter. The discriminator is run convolutionally across the whole image, averaging all responses of patches to achieve the final probability. One of the advantages of PatchGAN is that it can be applied to arbitrary-size images.

At training time, the Adam solver is applied to optimize the two adversarial networks. We adopt the standard approach of training GANs: optimization is carried out on the discriminator and the generator alternately [24]. At testing time, only the trained generator is used.

### *2.4 Ground truth for training*

For deep network training, the input pairs of original noisy image and clean ground truth image are needed. However, for OCT image despeckling, there's no ground truth image readily available. As mentioned, good despeckling results can be obtained by averaging Bscans repeatedly acquired at the same location. Though commercial scanners such as Topcon DRI-1 offer such scanning protocol, it only outputs the final high quality image. The averaging calculation is completed by the proprietary software,and the raw noisy image is not available to common users. Here we propose an alternative way of obtaining training images pairs that is practical for any commercial OCT scanner users. The high quality images are obtained through registration and averaging of Bscans from multiple OCT volumes.

*M* 3D OCT volumes are obtained repeatedly from the same normal eye, with minimal eye movement between different acquisitions. One volume is randomly picked as the target image and denoted as $V_1$, while other volumes are denoted as $V_2 \cdots V_M$. Let $B_k^m$ denote the *k*th Bscan in $V_m$. For a certain Bscan $B_i^1$ in the target volume, from all volumes, put the 2*N*+1 Bscans with indices closest to $i$ into a set: $\{B_k^m \mid m=1,\cdots,M, k=\min[\max(1,i-N),K-2N], \cdots,\max[2N+1,\min(i+N,K)]\}-\{B_i^1\}$, where $K$ represents the total number of Bscans in one OCT volume. Then all Bscans in this set are registered to $B_i^1$ using affine transformation. From the (2*N*+1)*M*-1 registered images, *L* images with the highest mean structural similarity index (MSSIM) [25] scores to $B_i^1$ are selected and averaged together with $B_i^1$. Then $B_i^1$ and the averaging result forms a noisy-clean image pair. Repeat this for all Bscans in the target volume, and we have a whole set of training samples at different locations of the retina. This procedure can be repeated for multiple eyes to obtain a larger training set.

In this method, we assume Bscans in a 3D volume within a small range share similar retinal structures. Registration is needed to remove possible misalignment in structure caused by eye movement or slight difference in scanning locations between different scans. The MSSIM measure ensures the best aligned images are averaged, to prevent blurring in the averaged results. The registration is performed using the imregister function of MATLAB (Mathworks, version 2012a and later). It is a multi-resolution registration method based on pixel intensities. The transform parameters are optimized by minimizing the mean square error of pixel intensities between the target image and the transformed image, using the gradient descent method. An image pyramid is built with decreased resolution by a factor of 2 in each dimension. The parameters are first optimized at the coarsest level of the pyramid and then successively refined on the next level, until getting back to the original full resolution image. In our experiments, the number of pyramid levels was set to 3. For gradient descent

Biomedical Optics EXPRESS

method, the maximum iteration number on each level was set as 500, and the maximum and minimum step length were set as 6.25e-02 and 5e-4, respectively.

In our method, to achieve both denoising and contrast enhancement, we further perform a piece-wise linear transform to the pixel intensities of the training images. Intensities less than the mean of the background region are mapped to 0, and the rest intensities are scaled to [0, 255]. Figure 3 shows Bscans from the target volume, the corresponding averaged Bscans and the enhanced Bscans used as ground truth.

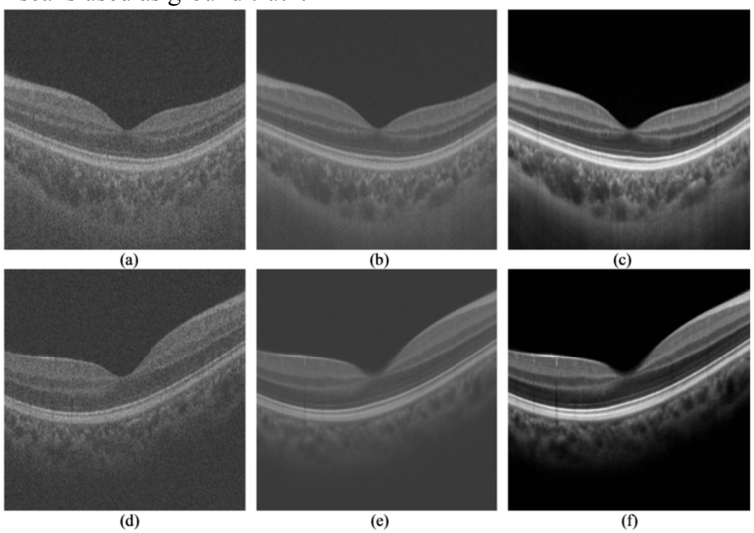


Fig. 3. Ground truth for training. Row 1: training data 1; Row 2: training data 2. (a)(d) Bscans from the original target volume. (b)(e) Corresponding Bscans after registration and averaging. (c)(f) Corresponding Bscans after contrast enhancement.

## 3. Experimental setup

### *3.1 Data*

In our experiments, two groups of training data, corresponding to the two mainstream center wavelength used in commercial retinal OCT scanners were used. The training data 1 were acquired using the Topcon DRI-1 Atlantis (Topcon Corporation, Japan) with center wavelength of 1050 nm, with 3D macular-centered scanning protocol. The volume size was 512 × 992 × 256 (width × height × Bscans) corresponding to 6 × 2.6 × 6 $mm^3$. $M$ = 20 repeated acquisitions were obtained with short intervals from a normal eye. To calculate the ground truth, $N$ was set as 3, and $L$ was set as 60. When $L$ was increased, the resulting averaged images appeared smoother, but more details were lost and the edges became more blurred. The value 60 was chosen at a balancing point in this tradeoff. With the method in 2.4, 256 training Bscan pairs were obtained.

The training data 2 were acquired using the Topcon 3D OCT 2000 (Topcon Corporation, Japan) with center wavelength of 840 nm, with 3D macular-centered scanning protocol. The volume size was 512 × 885 × 128 (width × height × Bscans) corresponding to 6 × 2.3 × 6 $mm^3$. $M$ = 20 repeated acquisitions were obtained with short intervals from each of two normal eyes. To calculate the ground truth, $N$ was set as 3, and $L$ was empirically set as 40. $L$ was set smaller than that of training data 1 because of the lower resolution in the slow axis. The larger spacing between Bscans made less Bscans well aligned with the target Bscan after registration, and larger $L$ would lead to blurred edges. 128 × 2 training Bscan pairs were obtained, which makes the two training data sets the same size.

We tested the proposed method on images acquired by the same scanner as the training data under the same or different scanning protocol, as well as images acquired by different

Biomedical Optics EXPRESS

scanners. The data came from both normal and pathological eyes. Table 1 listed the specifications of the 2 groups of training data volumes and 9 testing data volumes. The pathological data were from patients with central serous chorioretinopathy (CSC) or pathological myopia (PM). Intra-retinal fluids, neurosensory retinal detachment (NRD) or pigment epithelial detachment (PED) may appear in some Bscans. For each volume, 4 Bscans, two in the center and two in the peripheral area, were selected for quantitative evaluation.

**Table 1. Specifications of training and testing OCT data**

| | Scanner | Center wave-length (nm) | Longitudinal resolution in tissue (μm) | Lateral resolution in tissue (μm) | Bscan size (pixels/mm) | Pixel size (μm) | Location | Normal/ Pathological |
|---|---|---|---|---|---|---|---|---|
| training 1 | Topcon DRI-1 | 1050 | 20 | 8 | 512 × 992/6 × 2.6 | 11.72 × 2.6 | macula | Normal |
| training 2 | Topcon 2000 | 840 | 20 | 5~6 | 512 × 885/6 × 2.3 | 11.72 × 2.6 | macula | Normal |
| testing 1 | Topcon DRI-1 | 1050 | 20 | 8 | 512 × 992/6 × 2.6 | 11.72 × 2.6 | macula | Normal |
| testing 2 | | | | | 512 × 992/12 × 2.6 | 23.44 × 2.6 | macula + ONH | Normal |
| testing 3 | | | | | 512 × 992/6 × 2.6 | 11.72 × 2.6 | macula | Pathological (CSC) |
| testing 4 | Topcon 1000 | 840 | 20 | 6 | 512 × 480/6 × 1.68 | 11.72 × 3.5 | macula | Normal |
| testing 5 | | | | | 512 × 480/6 × 1.68 | 11.72 × 3.5 | macula | Pathological (CSC) |
| testing 6 | Topcon 2000 | 840 | 20 | 5~6 | 512 × 885/6 × 2.3 | 11.72 × 2.6 | macula | Normal |
| testing 7 | | | | | 512 × 885/6 × 2.3 | 11.72 × 2.6 | ONH | Normal |
| testing 8 | Zeiss Cirrus 4000 | 840 | 15 | 5 | 512 × 1024/6 × 2 | 11.72 × 1.95 | macula | Pathological (PM) |
| testing 9 | | | | | 512 × 1024/6 × 2 | 11.72 × 1.95 | macula | Pathological (CSC) |

All OCT data were uncompressed raw data exported from the scanners. For all acquisitions, we chose the 3D scanning mode with maximum number of Bscans provided by the scanner. In these modes, the output Bscans were the original acquisition, not averaged ones over several repetitions.

The study was approved by the Institutional Review Board of Soochow University, and informed consent was obtained from all subjects.

### *3.2 Implementation details*

Data augmentation was used to allow the model to learn different characteristics of the testing data. Flipping in the lateral direction was used to simulate the symmetry of right and left eye. Different scaling factors were applied to simulate the four types of pixel size (geometric size of the Bscan divided by the number of pixels in corresponding dimensions) of testing data. Rotation was used to simulate different inclination of the retina in the OCT image. Non-rigid transformation was used to simulate the deformation caused by pathologies. These processing are applied randomly, and the training data is augmented with a factor of two.

In the experiment, the Adam solver with initial learning rate 2e-4 and momentum 0.5 was applied to optimize the two adversarial networks. The weighting parameters are selected as $\alpha = 100$ and $\beta = 1$, so that the L1 loss and edge loss are of the same order of magnitude. As tested, too large weight for the edge loss might make the training difficult to converge. The batch size was set as 1 and the number of training epochs was set as 100. The proposed method were coded in Python based on Tensorflow and trained using the NVIDIA GTX Titan X GPU with 12G memory.

Biomedical Optics EXPRESS

### 3.3 Evaluation metrics

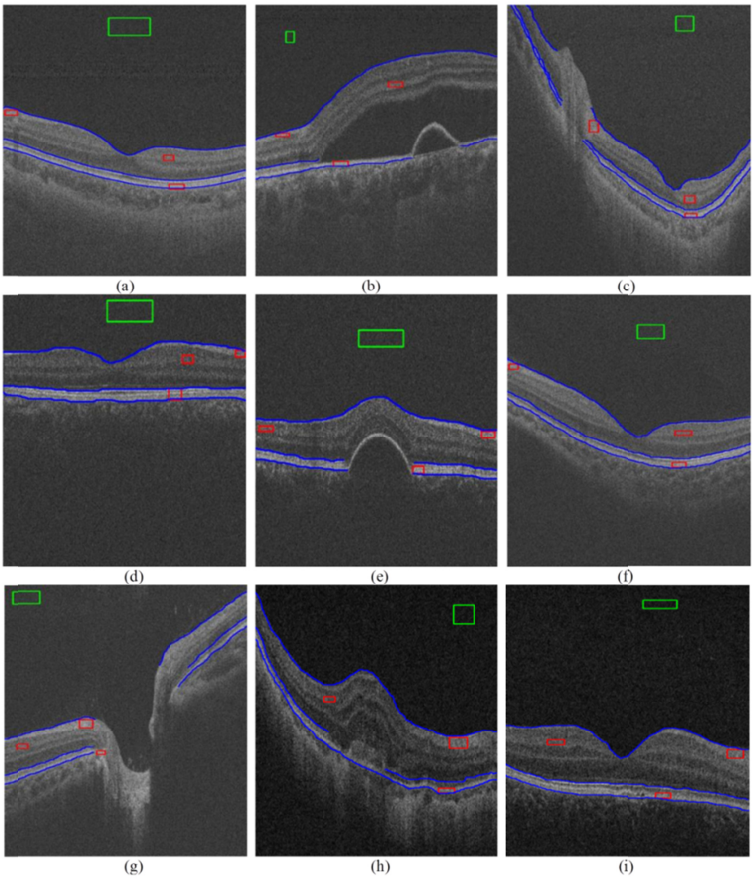


Fig. 4. Original noisy OCT images with selected ROIs and boundaries marked. Three signal regions (red) and one background region (green) are manually selected for calculating SNR, CNR and ENL. Three boundaries (blue) are manually delineated for calculating EPI. Panel (a) to (i) correspond to testing data 1 to 9 listed in Table 1.

To quantitatively compare the performance of different denoising algorithms on OCT images, four performance metrics are computed in the experiments: signal-to-noise ratio (SNR), contrast-to-noise ratio (CNR), equivalent number of looks (ENL) and edge preservation index (EPI). For calculating these metrics, regions of interest (ROIs) are manually defined. One Bscan of the original image from each of the 9 testing data are shown in Fig. 4, where the green rectangle represents the background region, three red rectangles represent the signal regions, located in the retinal neural fiber layer (RNFL), inner retina, and the retinal pigment epithelium (RPE) complex, respectively, and the three boundaries in blue (upper boundary of RNFL, inner-outer retina boundary and the lower boundary of RPE) denote the locations where EPI is calculated. The performance metrics are introduced as follows.

#### 3.3.1 Signal-to-noise ratio (SNR)

SNR is a suitable criterion to reflect noise level in the image, defined as follows:

$$SNR = 10\log_{10}\left(\frac{\max(I)^2}{\sigma_b^2}\right) \tag{7}$$

Biomedical Optics EXPRESS

where max($I$) represents the maximum pixel intensity of the image I, and $\sigma_b$ is the standard deviation of the background region.

### 3.3.2 Contrast-to-noise ratio (CNR)

CNR is a measure of the contrast between the region of signal and the noisy background region in the image. CNR of the $i$-th signal region is calculated as:

$$CNR_i = 10\log_{10}\left(\frac{|\mu_i - \mu_b|}{\sqrt{\sigma_i^2 + \sigma_b^2}}\right) \tag{8}$$

where $\mu_i$ and $\sigma_i$ denote the mean and standard deviation of $i$-th signal region in the image, while $\mu_b$ and $\sigma_b$ denote the mean and standard deviation of the background region.

In our experiments, the average CNR is computed over the 3 signal ROIs.

### 3.3.3 Equivalent number of looks (ENL)

ENL is commonly used to measure smoothness of the homogeneous region in the image. ENL over $i$-th ROI in an image can be calculated as:

$$ENL_i = \frac{\mu_i^2}{\sigma_i^2} \tag{9}$$

where $\mu_i$ and $\sigma_i$ denote the mean and standard deviation of $i$-th signal ROI in the image.

In our experiments, the average ENL is computed over the 3 signal ROIs.

### 3.3.4 Edge preservation index (EPI)

EPI is a performance measure that reflects the extent of maintaining details of edge in the image after denoising. EPI in the longitudinal direction is defined as:

$$EPI = \frac{\sum_i \sum_j \left| I_d(i+1,j) - I_d(i,j) \right|}{\sum_i \sum_j \left| I_o(i+1,j) - I_o(i,j) \right|} \tag{10}$$

where $I_o$ and $I_d$ represent the noisy image and the denoised image, while $i$ and $j$ represent coordinates in the longitudinal and lateral direction in the image. This index may not be an accurate indicator of edge-preservation if calculated over the entire image, since after denoising, the gradient will become smaller in homogeneous regions. Therefore we only calculate the sums in (9) in the neighborhood of image boundaries. In our experiments, the neighborhood was set as a band with height of 7 pixels centered at the boundaries shown in Fig. 4.

## 4 Results

Figure 5 shows denoised Bscans from the 9 test data obtained using training data 1, corresponding to those in Fig. 4. By visual inspection, we can see that the proposed edge-sensitive cGAN works well for the data with different resolution, obtained at different retinal locations, and both for normal and pathological retina. The retinal structures are preserved well while speckle noise is suppressed. The contrast between layers is also enhanced. After denoising, the background is homogeneous and almost black. The highest pixel intensities occur at the RPE layer. For Bscans in which the choroid is also visualized, such as in Fig. 5(a)(b)(c)(f)(g), both the capillary and the large vessels can be observed much more clearly.

In order to further evaluate the effectiveness of the proposed edge sensitive cGAN, we design three groups of comparative experiments. The first group is aimed at comparing different objective functions. The second group studies the performance achieved by different

Biomedical Optics EXPRESS

training data and the performance on testing data acquired by different scanners. The third group is comparative analysis of various existing denoising methods.

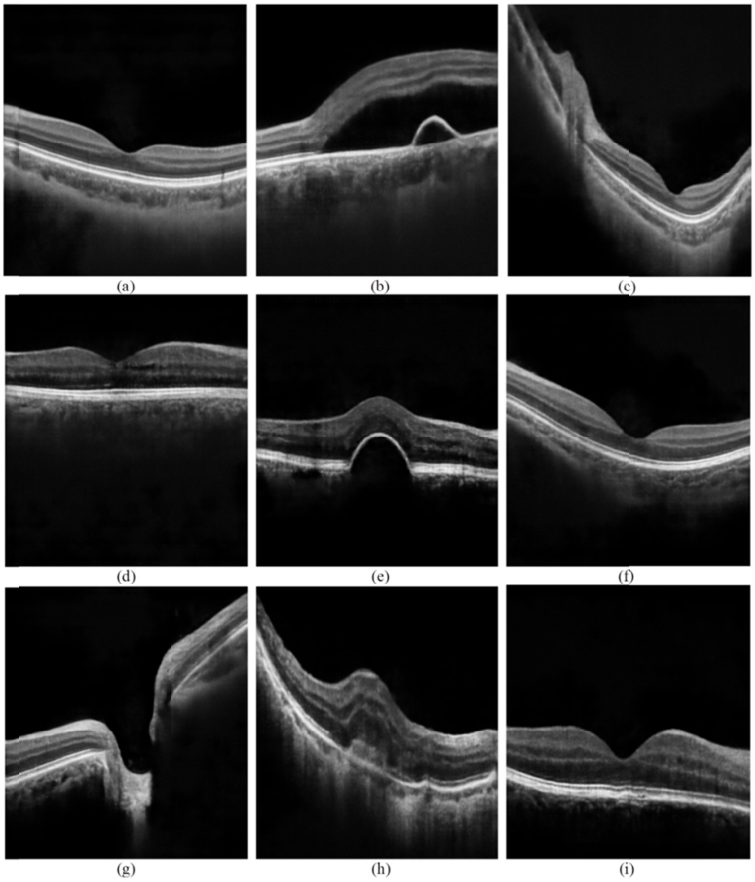


Fig. 5. Denoised Bscans by the proposed method, corresponding to the Bscans in Fig. 4.

### *4.1 Performance of different objective functions*

We first study the effect of different loss function on denoising results. The network was trained using only the cGAN term as in Eq. (2), cGAN term added with L1 distance as in Eq. (4), and the edge-sensitive objective function, i.e., cGAN term combined with both L1 distance and edge loss as in Eq. (6), respectively. Training data 1 were used in this study. All evaluation metrics listed in this subsection are the mean values calculated for the total 36 testing Bscans, with 4 from each of the 9 testing volumes.

Figure 6 shows the denoising results of two Bscans. We can see that results using cGAN term only have artifacts inside the layers. Some fine details are blurred, and there's occasional boundary loss. By adding the L1 norm, the intra-layer smoothness is improved. By adding the edge loss, the contrast at the edge is further enhanced, and more fine details are preserved. Table 2 shows the average performance metrics for results obtained by the three objective functions. The results of cGAN + L1, and of edge-sensitive objective function are higher in SNR, CNR and ENL than those of basic cGAN loss, which is compliant with the better visual quality. By adding the edge loss, the ENL and EPI increased with slight decrease of SNR and CNR, caused by a slight increase of the standard deviation of the background region and the signal region. Generally, the edge-sensitive objective function obtains the best overall denoising performance.

Biomedical Optics EXPRESS

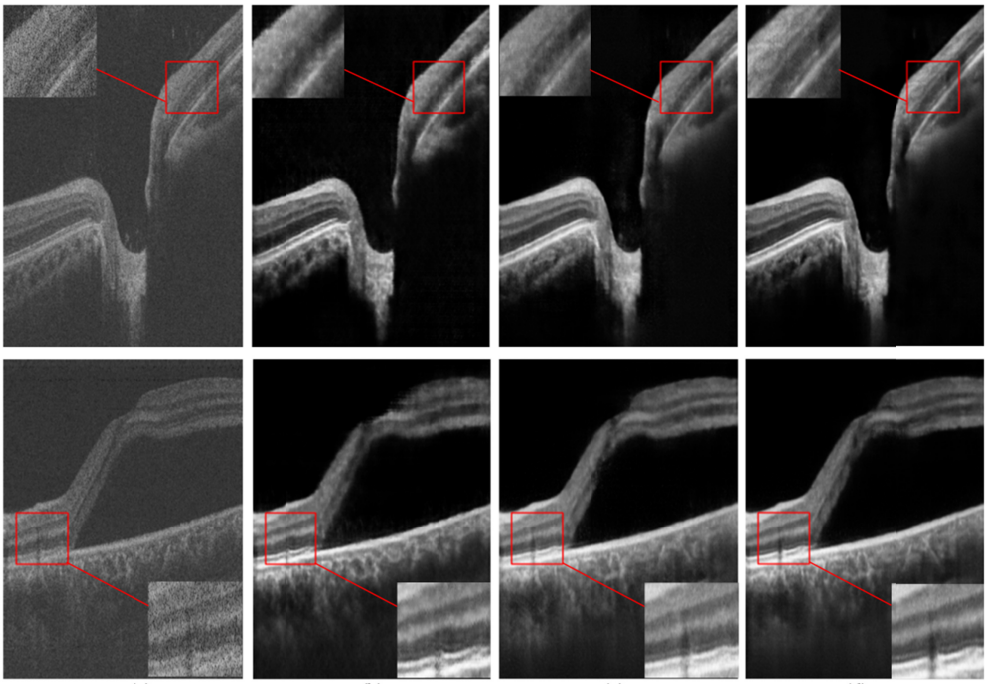


Fig. 6. Results for two Bscans obtained by different objective functions. (a) Original images (b) Results using cGAN term only (c) Results using cGAN + L1 term (d) Results using the edge-sensitive objective function.

**Table 2. Evaluation metrics in average for different objective functions.**

| | SNR | CNR | ENL | EPI |
|---|---|---|---|---|
| Original | 26.50 ± 1.04 | 4.55 ± 0.80 | 34.81 ± 12.71 | 1.00 ± 0.00 |
| cGAN | 44.76 ± 12.29 | 9.10 ± 0.83 | 92.10 ± 35.39 | 1.01 ± 0.17 |
| cGAN + L1 | **50.18 ± 9.07** | **9.99 ± 0.77** | 131.84 ± 41.90 | 1.00 ± 0.20 |
| Edge-sensitive objective function | 49.57 ± 5.01 | 9.85 ± 1.02 | **133.96 ± 60.11** | **1.06 ± 0.18** |

### *4.2 Performance for data from different scanners*

We trained two networks using training data 1 and 2, respectively. The resulting performance metrics are calculated for testing data obtained by different scanners, respectively, and are shown in Fig. 7. The total averaged metrics obtained by the two training sets are also shown in Fig. 7 and listed in Table 3. For SNR, the results obtained by training data 2 are higher, which can be attributed to the lower background variances of this training set. For CNR and ENL, the values are higher when the training data and the test data are obtained using the same center wavelength. The ENL values have biggest variations across different training and testing data, partly because it is calculated using the squared values of the pixel statistics. The ENL values are also strongly related to those of the original noisy Bscans. For EPI, the results obtained by training data 2 are lower. This is because the edges are more blurred in this training set, caused by registering and averaging Bscans with larger spacing.

Biomedical Optics EXPRESS

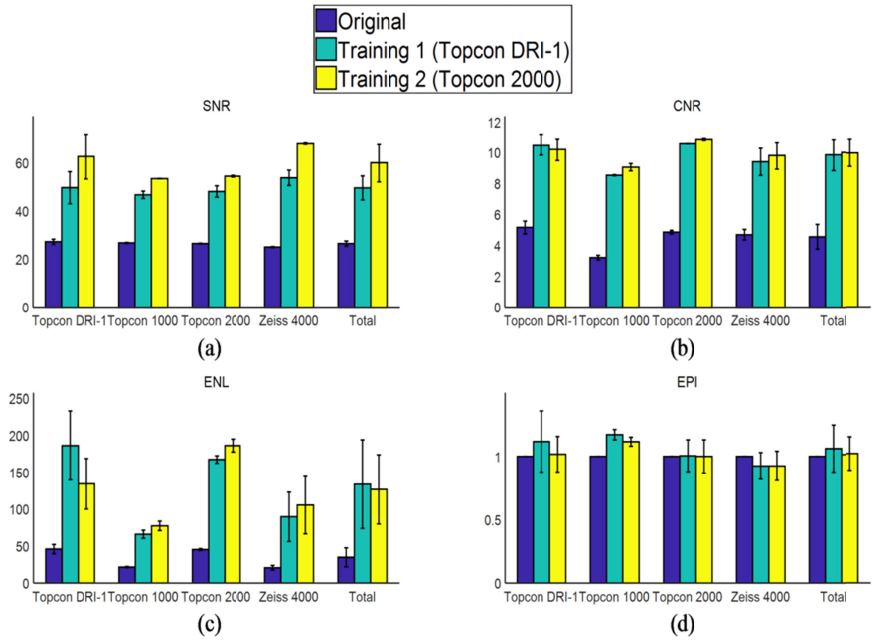


Fig. 7. Evaluation metrics obtained by different training data and for test data from different scanners. (a)SNR (b)CNR (c)ENL (d)EPI

Examples of denoised Bscans obtained by the two training sets are shown in Fig. 8(i)(j) and Fig. 9(i)(j). As for visual qualities, both training sets results in similar appearance for the tissues inside the retina. However, in the choroid, the training data 1 recovers more details than training data 2. This can be explained by the longer imaging depth of the Topcon DRI-1 scanner with center wavelength of 1050nm. Thus more information of the choroid can be learned from training data 1 than training data 2 (See Fig. 3).

### *4.3 Comparison to state-of-the-art*

We compared the proposed method with state-of-the-art approaches for general image denoising and for OCT despeckling, including non-local means (NLM) [26], block-matching and 3D filtering (BM3D) [27], sparsifying transform learning and low-rank method (STROLLR) [28], deep CNN with residual learning (DnCNN) [19], 3D complex wavelet based K-SVD for OCT denoising [8], maximum-a-posteriori (MAP) estimation based on local statistical model for OCT denoising [14], Resnet-based method for OCT denoising [20]. In these experiments, parameters for NLM [26] and BM3D [28] were tuned to reach a balance between speckle removal and edge preservation, while parameters of other methods were set to default values as in the corresponding references. For deep learning based methods [19,20], the same training data and data augmentation method was used. For other methods for comparison, the denoised results were contrast enhanced by linearly mapping the range of pixel intensities to [0,255].

Figure 8 shows results for one Bscan from testing data 1, which is acquired in exactly the same way as the training data 1. The quality of the raw data is relatively good. Figure 9 shows results for one Bscan from testing data 8, which is from a scanner by a different manufacturer (Carl Zeiss Meditech, Germany). The image quality of the raw data is the lowest, caused by the pathology. The average performance metrics for all test data are listed in Table 3. We can see that the proposed method with training data 1 can obtain the best visual quality. The proposed method with training data 1 is also higher in performance metrics than most of the methods compared, except for K-SVD. Visually inspected, results of NLM, BM3D and

Biomedical Optics EXPRESS

STROLLR all have artifacts inside the retinal layers and near the boundaries. For BM3D, the background regions are not homogeneous. For K-SVD, the results in Fig. 8(e) is oversmoothed with many image details blurred, which is the reason of the high SNR and ENL, and also result in low EPI. However, the results in Fig. 9(e) is under-smoothed, which might be caused by the difficulty of dictionary learning from the low quality image. The result of MAP in Fig. 8(f) is under-smoothed while in Fig. 9(f) is a bit oversmoothed. This might be caused by the unstable estimation of speckle parameters for different images. Moreover, the background noise is not removed well and the contrast is low. The results of DnCNN present vertical artifacts, and it almost fails for testing data 8. This shows the poor generalization ability of the network, probably due to limited training samples. The results of ResNet have distortions at the edges, and the contrast between layers is lower than that of the proposed method. The proposed method with training data 1 obtains good denoising results for testing data 1. Especially, among all the methods compared, it best recovers the thin layer above the RPE complex, known as external limiting membrane (ELM), which can be viewed more clearly in the zoomed cropped image. For testing data 8, the result is a bit blurred, but still better than other methods. Therefore in summary, combining both subjective and objective evaluation criteria, the proposed edge-sensitive cGAN can obtain best results among the methods compared. With training data 1, it improved the SNR, CNR and ENL by 87%, 116% and 285%, respectively, with respect to the original image. While many denoising methods reduce the EPI, it improved the EPI by 6%, which means the edges are enhanced. With training data 2, it improved the SNR, CNR and ENL by 127%, 120% and 265%, respectively, with respect to the original image. The mean EPI is comparative to the original image, indicating that the edges are mostly preserved.

**Table 3. Evaluation metrics in average for different denoising methods.**

| | SNR | CNR | ENL | EPI |
|---|---|---|---|---|
| Original | 26.50±1.04 | 4.55±0.80 | 34.81±12.71 | 1.00±0.00 |
| NLM[26] | 44.56±2.88 | 6.11±1.44 | 63.56±43.67 | 1.04±0.09 |
| BM3D[27] | 34.80±1.76 | 8.36±0.80 | 111.15±46.15 | 0.81±0.10 |
| STROLLR[28] | 41.86±2.09 | 8.18±1.41 | 123.91±98.29 | 0.80±0.09 |
| K-SVD[8] | 50.07±2.12 | 9.24±1.87 | **260.58±326.15** | 0.79±0.13 |
| MAP[14] | 31.73±0.73 | 7.33±1.28 | 128.44±54.76 | 0.75±0.09 |
| DnCNN[19] | 37.38±5.82 | 7.26±1.24 | 59.13±13.27 | 0.78±0.11 |
| ResNet[20] | 35.81±3.85 | 8.85±1.26 | 115.96±48.49 | 0.86±0.09 |
| Proposed (training 1) | 49.57±5.01 | 9.85±1.02 | 133.96±60.11 | **1.06±0.**18 |
| Proposed (training 2) | **60.09±8.00** | **10.01±0.87** | 126.91±46.99 | 1.01±0.13 |

## 5. Discussion and conclusions

In this paper, we propose an end-to-end deep learning framework that achieves simultaneous speckle reduction and contrast enhancement for retinal OCT images. The method is based on the image-to-image cGAN structure with a new edge-sensitive objective function. Unlike previous deep networks proposed for denoising [17–20] which try to estimate the noise residue from the noisy input, the cGAN learns the mapping from the noisy image to the clean image through the competition of generator and discriminator, and thus is not limited by the additive noise assumption. By introduction of the edge loss function, the method achieves a balanced performance in speckle reduction and structure preservation.

A novel method is proposed for obtaining the ground truth images based on multiple volumetric scans of the same eye, which is easy to implement for users of commercial

Biomedical Optics EXPRESS

scanners. Through registration and averaging of the best aligned images, high quality Bscans paired with the original ones are computed. Their contrast is further enhanced, so that the trained deep network is empowered with the ability of contrast enhancement in addition to denoising. In this paper, the training data was obtained from only one normal eye. Data augmentation allows the network to adapt to other types of OCT images with limited training samples.

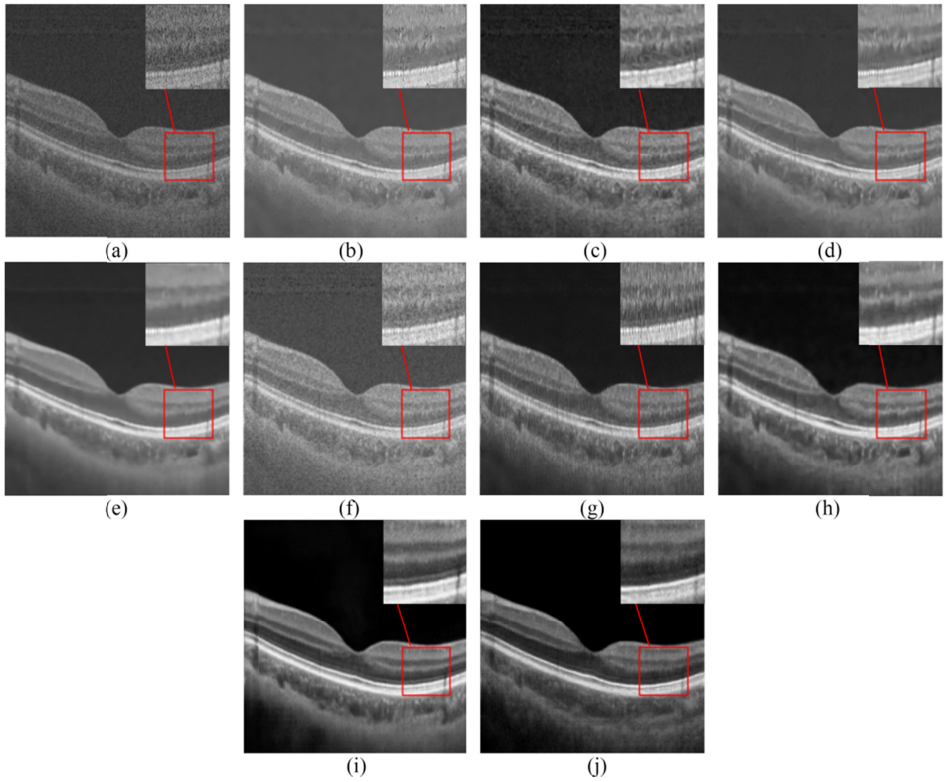


Fig. 8. Results for one Bscan of test data 1. (a) Original image (b) NLM(c) BM3D (d) STROLLR (e) K-SVD (f) MAP (g) DnCNN (h) ResNet (i) Proposed(training 1) (j) Proposed(training 2).

The proposed method was tested with 2 sets of training data, obtained by two scanners with different center wavelength, and 9 sets of testing data, obtained by different scanners, with different resolutions, at different retina locations, and from both normal and pathological eyes. There are some variations in the resulting quality metrics for different combinations of training and testing data. For some metrics, slightly higher values are obtained when the training and testing data share the same wavelength, but the effect is not overwhelming. For both training data, good results can be obtained both qualitatively and quantitatively for all testing data. In general, the despeckling performance is more related to the quality of the clean images generated for training, and also to the quality of the original noisy images. Theoretically, it is preferred to train the network with data from the same scanner as the OCT image to be denoised, so that the characteristics of speckles can be learned accurately. However, in practice, we find training data from a different scanner can also get good despeckling performance, as long as high quality ground truth images can be obtained. Data from scanners with higher resolution in the slow axis, and also with deeper penetration are recommended, which can achieve enhanced edge information and good despeckling performance both in the retina and in the choroid region. Additionally, although only training data from one or two eyes (256 training Bscan pairs) are proved effective in this study, better

Biomedical Optics EXPRESS

performance can be expected if more training data are provided, considering the nature of deep learning networks.

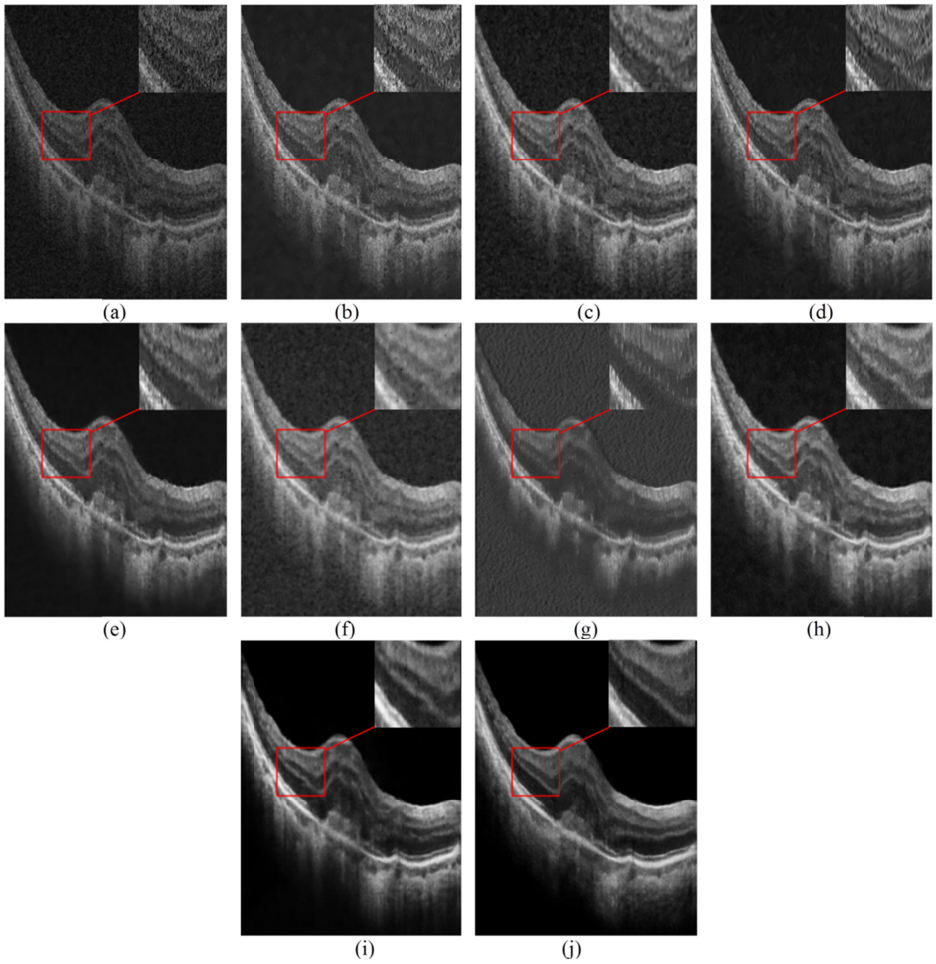


Fig. 9. Results for one Bscan of test data 8. (a) Original image (b) NLM(c) BM3D (d) STROLLR (e) K-SVD (f) MAP (g) DnCNN (h) ResNet (i) Proposed(training 1) (j) Proposed(training 2).

The results are compared with state-of-the-art methods proposed for natural images or OCT images both qualitatively and quantitatively. These methods include block-matching based methods, wavelet based method, probabilistic method, sparsity and low-rank based method, and deep learning methods. Upon comparison, our results have the best visual quality, especially in preserving detailed retinal structures. The proposed method also achieved consistently good performance on all test images. As far as the performance metrics are considered, the proposed method achieves the highest mean CNR and EPI values, while the SNR and ENL values are higher than all other methods except K-SVD [8]. However, based on examination of the resulting images, we find that the high SNR and ENL values of K-SVD are the results of oversmoothing with significant loss of image details. Therefore we argue that any evaluation metric alone cannot fully represent the quality of the image.

Considering the difficulty in evaluating OCT despeckling results, in the future, we will study the effect of proposed OCT image despeckling under the context of specific applications. First, as OCT images are mostly visually inspected in routine diagnosis, we can use manual gradings from clinical experts to judge whether despeckling is helpful for

Biomedical Optics EXPRESS

clinicians. Secondly, as OCT despeckling acts as the preprocessing step for automatic OCT image analysis, we will study how the performance of tasks such as segmentation is improved by the proposed despeckling method.

We don't list the time cost of each comparative methods here because it is unfair to compare other methods run on CPU or even with MATLAB codes that are not optimized in efficiency with the deep learning methods run on GPU. Still, the testing stage of deep learning methods has been proved very fast. The proposed method only requires an average of 0.22 seconds for denoising one Bscan, which can readily meet the real-time demand of clinical practice.

In conclusion, we have proposed an efficient and effective method that aims for speckle noise reduction in 3D OCT volumes exported from commercial retinal OCT scanners. The method achieves speckle noise suppression, edge preservation and contrast enhancement simultaneously. This method can be also extended to enhancement of other medical image modalities such as ultrasound image and low-dose CT image.

## Appendix: Detailed structures of cGAN

The overall structure of U-shape generator is illustrated in Fig.10. It is a kind of encoder-decoder structure with symmetric skip connections. All convolution and deconvolution layers apply 4×4 spatial filters with stride 2. Each layer adopts BatchNorm except the first convolutional layer of the encoder. All ReLUs in the encoder are leaky with slope 0.2, while those in the decoder are not. The random noise z is implicitly implemented as dropout with rate 0.5, i.e., randomly dropping some outputs by the probability of 0.5, in the first three layers of the decoder. The dropout can also prevent overfitting effectively during training. Tanh is used as the activation function of the last layer in the decoder.

The discriminator architecture called PatchGAN is shown in Fig. 11. PatchGAN inputs real pairs or fake pairs, and produce the corresponding outputs. It has five convolution layers. All ReLUs in the first four layers are leaky with slope 0.2. The middle three layers adopt BatchNorm. 4×4 spatial filters with stride 2 are applied in the first three layers except for those in last two layers with stride 1. For this design, the size of PatchGAN's receptive field, i.e., the size of the patch p is set as 70, which makes PatchGAN have fewer parameters and run faster than traditional discriminators and still produce high quality results [21]. Sigmoid is used as the activation function of the last layer to achieve the purpose of identification. In the final 62×62 image, each pixel represents the probability that the corresponding 70×70 patch in the input is identified as real.

Biomedical Optics EXPRESS

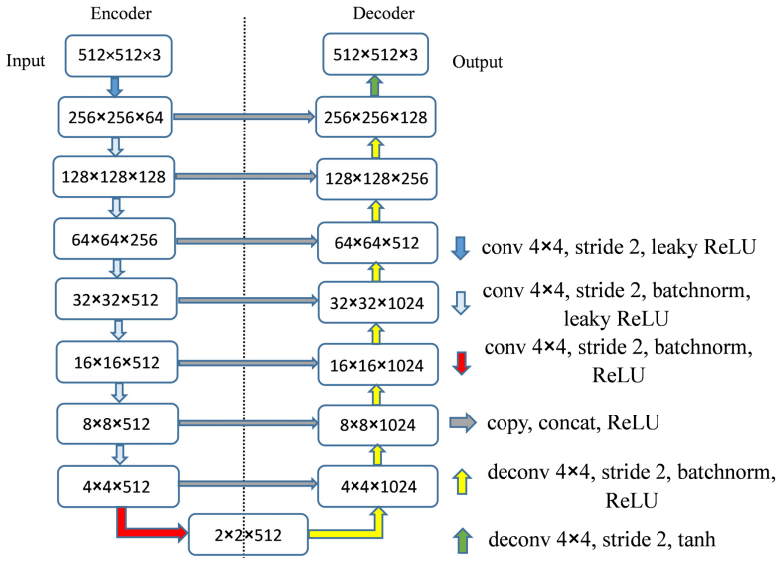


Fig. 10. U-shape architecture of the generator.

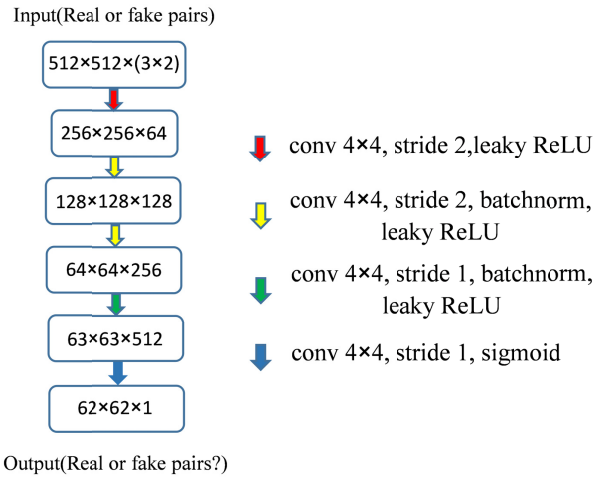


Fig. 11. Architecture of the discriminator: PatchGAN.

## Funding

The National Basic Research Program of China (973 Program) (2014CB748600); the National Natural Science Foundation of China (NSFC) (61622114, 61771326, 61401294, 81401472, 61401293, 81371629); Collaborative Innovation Center of IoT Technology and Intelligent Systems,Minjiang University (No: IIC1702).

## Disclosures

The authors declare that there are no conflicts of interest related to this article.

## References

1. J. M. Schmitt, S. H. Xiang, and K. M. Yung, "Speckle in optical coherence tomography," J. Biomed. Opt. **4**(1), 95–105 (1999).
2. H. M. Salinas and D. C. Fernández, "Comparison of PDE-based nonlinear diffusion approaches for image enhancement and denoising in optical coherence tomography," IEEE Trans. Med. Imaging **26**(6), 761–771

Biomedical Optics EXPRESS

(2007).
3. P. Puvanathasan and K. Bizheva, "Interval type-II fuzzy anisotropic diffusion algorithm for speckle noise reduction in optical coherence tomography images," Opt. Express **17**(2), 733–746 (2009).
4. J. Aum, J. H. Kim, and J. Jeong, "Effective speckle noise suppression in optical coherence tomography images using nonlocal means denoising filter with double Gaussian anisotropic kernels," Appl. Opt. **54**(13), 13–14 (2015).
5. X. Zhang, L. Li, F. Zhu, W. Hou, and X. Chen, "Spiking cortical model-based nonlocal means method for speckle reduction in optical coherence tomography images," J. Biomed. Opt. **19**(6), 066005 (2014).
6. B. Chong and Y. K. Zhu, "Speckle reduction in optical coherence tomography images of human finger skin by wavelet modified BM3D filter," Opt. Commun. **291**(6), 461–469 (2013).
7. F. Zaki, Y. Wang, H. Su, X. Yuan, and X. Liu, "Noise adaptive wavelet thresholding for speckle noise removal in optical coherence tomography," Biomed. Opt. Express **8**(5), 2720–2731 (2017).
8. R. Kafieh, H. Rabbani, and I. Selesnick, "Three dimensional data-driven multi scale atomic representation of optical coherence tomography," IEEE Trans. Med. Imaging **34**(5), 1042–1062 (2015).
9. 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. Express **18**(2), 1024–1032 (2010).
10. 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. Express **3**(5), 927–942 (2012).
11. L. Fang, S. Li, D. Cunefare, and S. Farsiu, "Segmentation based sparse reconstruction of optical coherence tomography images," IEEE Trans. Med. Imaging **36**(2), 407–421 (2017).
12. A. Wong, A. Mishra, K. Bizheva, and D. A. Clausi, "General Bayesian estimation for speckle noise reduction in optical coherence tomography retinal imagery," Opt. Express **18**(8), 8338–8352 (2010).
13. A. Cameron, D. Lui, A. Boroomand, J. Glaister, A. Wong, and K. Bizheva, "Stochastic speckle noise compensation in optical coherence tomography using non-stationary spline-based speckle noise modelling," Biomed. Opt. Express **4**(9), 1769–1785 (2013).
14. M. Li, R. Idoughi, B. Choudhury, and W. Heidrich, "Statistical model for OCT image denoising," Biomed. Opt. Express **8**(9), 3903–3917 (2017).
15. I. Kopriva, F. Shi, and X. Chen, "Enhanced low-rank + sparsity decomposition for speckle reduction in optical coherence tomography," J. Biomed. Opt. **21**(7), 076008 (2016).
16. C. 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. Med. Imaging **35**(10), 2270–2279 (2016).
17. X. J. Mao, C. Shen, and Y. B. Yang, "Image restoration using very deep convolutional encoder-decoder networks with symmetric skip connections," In *Proceedings of International Conference on Neural Information Processing Systems (NIPS),* (2016).
18. Y. Tai, J. Yang, X. Liu, and C. Xu, "MemNet: a persistent memory network for image restoration," In *Proceedings of IEEE International Conference on Computer Vision (ICCV)*, 4549–4557(2017).
19. 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. **26**(7), 3142–3155 (2017).
20. N. Cai, F. Shi, Y. Gu, D. Hu, Y. Chen, and X. Chen, "A ResNet-based universal method for speckle reduction in optical coherence tomography images," In *Proceedings of IEEE International Symposium on Biomedical Imaging (ISBI),* (2018).
21. P. Isola, J. Y. Zhu, T. Zhou, and A. A. Efros, "Image-to-image translation with conditional adversarial networks," In *Proceedings of Computer Vision and Pattern Recognition (CVPR)*, 5967–5976(2017).
22. O. Ronneberger, P. Fischer, and T. Brox, "U-Net: convolutional networks for biomedical image segmentation," In *Proceedings of Medical Image Computing and Computer-Assisted Intervention (MICCAI)*, 234–241(2015).
23. S. Ioffe and C. Szegedy, "Batch Normalization: accelerating deep network training by reducing internal covariate shift," arXiv:1502.03167v3, (2015).
24. I. J. Goodfellow, J. Pouget-Abadie, M. Mirza, B. Xu, D. Warde-Farley, S. Ozair, A. Courville, and Y. Bengio, "Generative adversarial nets," In proceedings of International Conference on Neural Information Processing Systems(NIPS), 2672–2680(2014).
25. 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. **13**(4), 600–612 (2004).
26. A. Buades, B. Coll, and J. M. Morel, "A non-local algorithm for image denoising," In *Proceedings of Computer Vision and Pattern Recognition (CVPR)*, 60–65 (2005).
27. K. Dabov, A. Foi, V. Katkovnik, and K. Egiazarian, "Image denoising by sparse 3-D transform-domain collaborative filtering," IEEE Trans. Image Process. **16**(8), 2080–2095 (2007).
28. B. Wen, Y. Li, and Y. Bresler, "When sparsity meets low-rankness: transform learning with non-local low-rank constraint for image restoration," In proceedings of IEEE International Conference on Acoustics, Speech and Signal Processing(ICASSP), 2297–2301(2017).