Physics in Medicine & Biology

DeSpecNet: a CNN-based method for speckle reduction in retinal optical coherence tomography images

To cite this article before publication: Fei Shi *et al* 2019 *Phys. Med. Biol.* in press https://doi.org/10.1088/1361-6560/ab3556

DeSpecNet: a CNN-based Method for Speckle Reduction in Retinal Optical Coherence Tomography Images

Fei Shi[a,+], **Ning Cai**[b,e,+], **Yunbo Gu**[b,e], **Dianlin Hu**[b,e], **Yuhui Ma**[a], **Yang Chen**[b,d,e,*], **Xinjian Chen**[a,c,*]

[a]School of Electronics and Information Engineering, Soochow University, Suzhou, China
[b]Laboratory of Image Science and Technology, School of Computer Science and Engineering, Southeast University, Nanjing, China
[c]State Key Laboratory of Radiation Medicine and Protection, Soochow University, Suzhou, China
[d]Centre de Rechercheen Information Biomedicale Sino-Francais (LIA CRIBs), Rennes, France
[e]The Key Laboratory of Computer Network and Information Integration (Southeast University), Ministry of Education, Nanjing, China

Abstract. Speckle is a major quality degrading factor in optical coherence tomography (OCT) images. In this work we propose a new deep learning network for speckle reduction in retinal OCT images, termed DeSpecNet. Unlike traditional algorithms, the model can learn from training data instead of manually selecting parameters such as noise level. The proposed deep convolutional neural network (CNN) applies strategies including residual learning, shortcut connection, batch normalization and leaky rectified linear units to achieve good despeckling performance. Application of the proposed method to the OCT images shows great improvement in both visual quality and quantitative indices. The proposed method provides good generalization ability for different types of retinal OCT images. It outperforms state-of-the-art methods in suppressing speckles and revealing subtle features while preserving edges.

Keywords: optical coherence tomography, speckle reduction, deep learning, residual learning.

Corresponding authors: *Yang Chen, chenyang.list@seu.edu.cn *Xinjian Chen, xjchen@suda.edu.cn
+ indicates these authors contributed equally.

1 Introduction

Optical coherence tomography (OCT) generates cross-sectional imaging of biological tissue in micron resolution and has become a routine technique in diagnosis of retinal diseases. Speckles, although contain certain information, often reduce the contrast and obscure subtle structures, and thus impair clinical diagnosis.[1] Speckles also affect the performance of automatic OCT image analysis methods, such as retinal layer segmentation[2,3]or pathological region segmentation.[4,5] To improve the utility of OCT images both for human observation and for computer analysis, efficient and effective speckle reduction methods are needed. The aim is to reduce the granular appearance while preserving the fine structure of retinal tissues.

Speckle reduction methods can be divided into two categories: hardware based and digital signal processing based methods. Hardware based methods such as angular compounding,[6,7] spatial compounding,[8,9] frequency compounding[10] aim to produce uncorrelated speckle patterns that can be cancelled by averaging. These approaches need specially designed acquisition systems, and cannot be directly applied to commercial OCT scanners. One temporal compounding approach that has been successfully integrated into some commercial scanners can achieve a high quality B-scan by averaging multiple B-scans acquired at the same location. However, the multiple acquisitions largely prolong the total imaging time, and the number of such despeckled B-scans is limited. Digital signal processing based methods are mostly post-processing methods that solely rely on modifying the obtained OCT images. Numerous algorithms have been proposed for this purpose. Median filter and Wiener filter are among the early attempts for OCT despckling.[11] They are easy to implement but often cause blurred edges. Partial differential equation (PDE) based methods such as anisotropic diffusion can provide better results, but there is the problem of over-fitting and over-smoothing. Wavelet/curvelet-based thresholding methods are widely applied.[12–16] However, artifacts may appear at the edges because the image features cannot be fully represented by the fixed wavelet basis, and the thresholds are difficult to choose. Dictionary learning based methods are therefore proposed.[17–19] The atoms in the dictionary can represent the clean image and speckles with less regular structures are removed. But the duration for learning a dictionary can be long. Non-local means (NLM), as an effective image denoising method, was applied to OCT despeckling.[20,21] BM3D[22] which combines the idea of block matching and wavelet thresholding was also proposed for OCT despeckling.[23] These methods depending on similar image patches may cause edge distortion because the patches are not always well matched. Statistical modelling are often used for image denoising and also for OCT despeckling.[24–26] However, as the stochastic

property of speckles are difficult to characterize, the proposed model may not be robust for different data. Recently, methods based on low rank decomposition have been proposed.[27,28] Their performance is affected by the accuracy of the low rank + sparsity matrices model, and the decomposition is time-costly and can only reach suboptimal results. In summary, structure-preserving speckle reduction for OCT images remains an open problem.

Deep learning networks, especially deep convolutional neural networks (CNNs) have achieved high performance in many computer vision tasks, and also showed potential in image denoising. Zhang *et al.* proposed a convolutional neural network called DnCNN for natural image denoising.[29] The idea of residual learning[30] is adopted, where the network learns the noise residue from the noisy image. By subtracting the noise residue, a denoised image is obtained. Batch normalization technique is further introduced to stabilize and enhance the training performance. Deep learning methods were also proposed for despeckling in synthetic aperture radar[31,32] or ultrasound imaging.[33,34] Wang *et al.*[31] proposed a division residue CNN to learn the multiplicative speckle component from the input image. In this work, we also apply residue learning and batch normalization, but improve the network using the shortcut connectivity blocks and leaky rectified linear units. Furthermore, a particular method for obtaining the training data is proposed for OCT images. Unlike many traditional methods which require input or estimation of parameters such as noise level, the proposed network, named DeSpecNet, is able to learn the characteristics of speckles from the training data. In experiment, it achieves good speckle reduction performance for different types of retinal OCT images with limited number of training samples.

The rest of the paper is organized as follows. Section 2 describes the despeckling framework, including data acquisition and preparation, the network structure, and implementation details. Section 3 presents experimental results and comparisons with state-of-the-art denoising methods. Dis-

cussions and conclusions are given in Section 4.

2 Method

2.1 Method overview

In this paper, we propose a deep learning based framework that works for speckle reduction in each frame (called B-scan) from a 3D volume output by commercial OCT scanners. As speckles can be modeled as multiplicative to the latent clean image,[15,27] the output of commercial OCT scanners, which is in logarithmic scale, can be modeled as $\mathbf{y} = \mathbf{x} + \mathbf{s}$, where $\mathbf{y}$ represents the B-scan image with speckle, $\mathbf{x}$ denotes the latent clean image, and $\mathbf{s}$ denotes the residue corresponding to pure speckles. In the proposed method we train the convolutional network to predict $\mathbf{s}$ from $\mathbf{y}$. The training and testing workflow is shown in Figure 1. Both training and testing images go through a preprocessing procedure called flattening to make the retinal structures more aligned. In the test stage, the output despeckled images are de-flattened to restore the original shape. To further enhance the contrast, the intensities of the output image are linearly stretched to the full dynamic range.

2.2 Training data acquisition and preparation

In this work ground truth denoised images for training are generated using a combination of spatial and temporal compounding methods. Temporal compounding is achieved by averaging B-scans from multiple OCT volumes of the same eye, and spatial compounding by averaging adjacent B-scans in each volume. In order to avoid compounding low quality data (for instance due to eye motion), B-scans are registered, and only images with sufficient structural similarity are compounded.

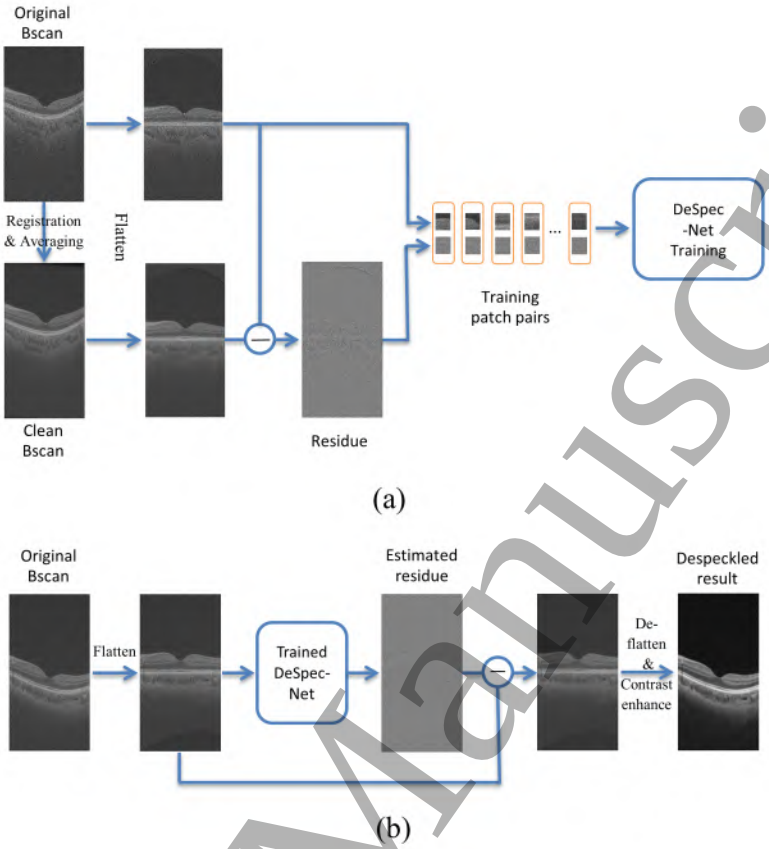

Fig 1 The workflow of the proposed method: (a) training stage (b) testing stage

Specifically, M repeated OCT volumes are acquired during stable fixation. One of these volumes is randomly selected as the target volume and its B-scans are referred to as target B-scans. For each target B-scan (where possible due to edge effects), select the corresponding B-scan and the surrounding $N-1$ B-scans from all volumes. These $NM-1$ B-scans are then registered and the structural similarity (SSIM) index[35] calculated for each compared with the target B-scan. The L B-scans most similar to the target B-scan are averaged to form the temporally and spatially compounded ground truth. The flow diagram is shown in Figure 2.

In this paper, we set the parameters as $M=20$, $N=7$, and $L=10$ so that the "ground truth" image is balanced in regional smoothness and edge sharpness. The registration is achieved using

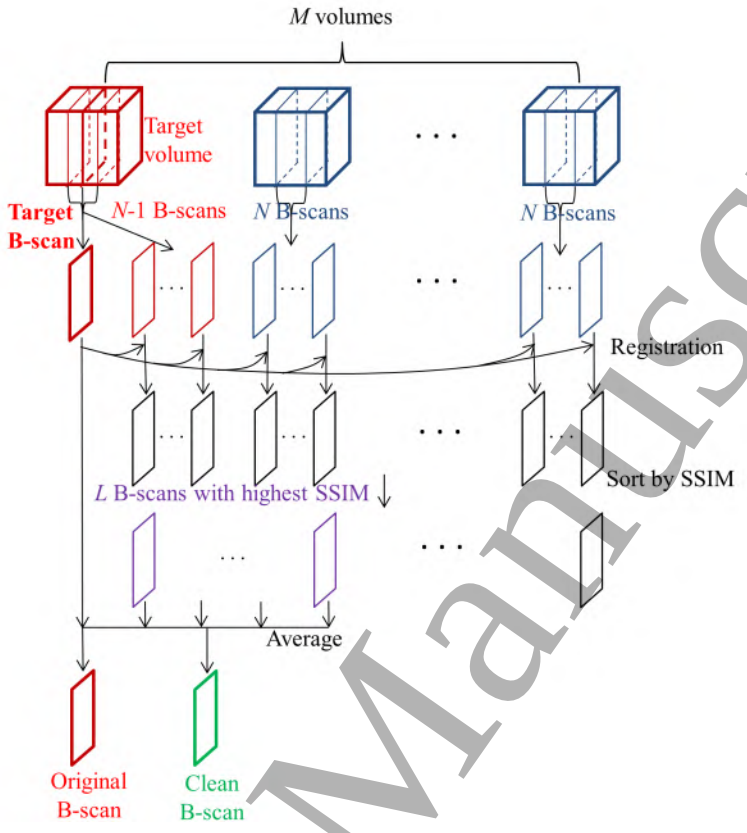

Fig 2 The flow diagram of training data computing.

the imregister routine in MATLAB (Mathworks, version 2012a and later). This routine implements an intensity-based multi-resolution registration algorithm. Transform parameters are first obtained in low resolution, and successively refined in higher resolutions. The similarity measure is the mean square error of pixel intensities between the target image and the transformed image. The transform parameters are optimized using the gradient descent method. In our experiments, the affine transformation model was used. The number of resolution levels was set to three. For gradient descent method, the maximum iteration number on each level was set as 500, and the maximum and minimum step length were set as 0.0625 and 0.0005, respectively.

In our experiments, two types of scanners and three normal eyes, each randomly chosen from

one subject, were involved for training data preparation. The first scanner was Topcon Atlantis DRI-1 SS-OCT scanner (Topcon, Tokyo, Japan) with center wavelength of 1050nm. Macula-centered OCT volumes were acquired repeatedly for one eye, each with $992 \times 512 \times 256$ (height $\times$ width $\times$ B-scans) voxels corresponding to $2.6 \times 6 \times 6$ mm^3. The second scanner was Topcon OCT-2000 SD-OCT scanner (Topcon, Tokyo, Japan) with center wavelength of 840nm, and macula-centered OCT volumes were acquired repeatedly for each of two eyes, each with $885 \times 512 \times 128$ (height $\times$ width $\times$ B-scans) voxels corresponding to $2.3 \times 6 \times 6$ mm^3. Different types of OCT scanners were used to allow the network to adapt to different speckle characteristics. As the second scanner gave less B-scans per volume, data from one more eye was needed so that the final training data from the two scanners were balanced. For both types of data, the parameters used for calculating the "ground truth" were the same.

Note that all OCT images used in the experiments were uncompressed raw data directly exported from the scanners. More specification of the scanners are listed in Table 1.

The bottom of the retinal pigment epithelium (RPE) layer are then detected from the clean image volume by a 3D multi-scale graph search method.[36] Using this surface as reference, both the original and the clean B-scans in the training set are "flattened" by circularly shifting each column so that the RPE bottom becomes a flat surface in the resulting image. This flattening procedure is also applied to the test B-scans, so that the different pose of the retina can be compensated, and the retinal structures in the training and testing images are better aligned. As the size of speckles is small, at the sampling rate of commercial OCT scanners, the speckles in OCT images do not show high spatial correlations. Moreover, the proposed network is trained to learn the pixel-to-pixel correspondence between input and output. Therefore this change of spatial arrangements of the pixels will not affect the performance of the model. The training pairs before and after flattening

are shown in Figure 3.

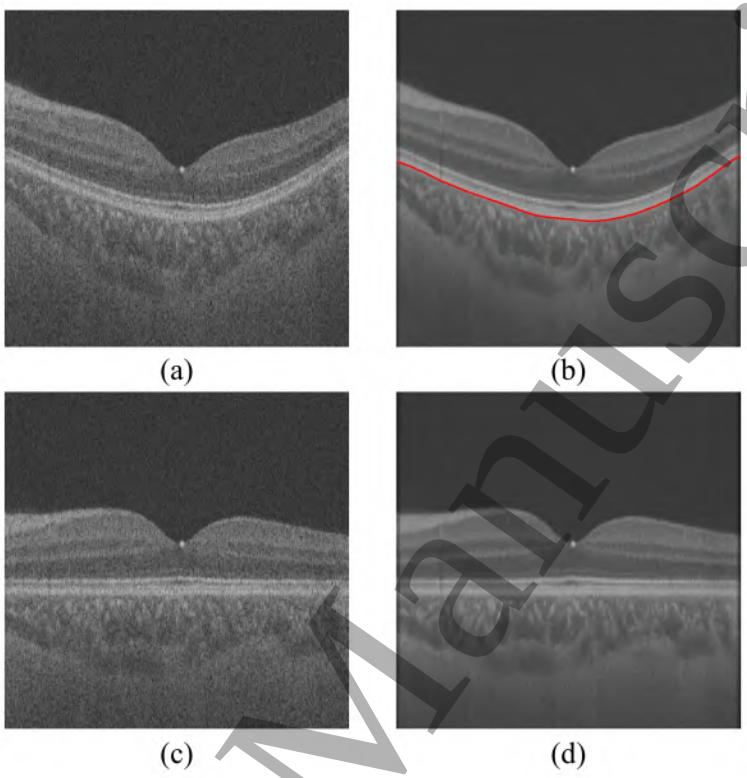

Fig 3 Training image preparation results: (a) Original B-scan. (b) Clean B-scan obtained by registration and averaging. The red curve denotes the detected RPE bottom. (c) Original B-scan after flattening. (d) Clean B-scan after flattening.

2.3 Network architecture

Four strategies are used in designing the network architecture, including residual learning, shortcut connection, batch normalization (BN) and Leaky rectified linear units (Leaky-ReLU). As is shown in Figure 4, the proposed DeSpecNet consists of 17 layers, in sequence: two Conv-BN-Leaky-ReLU, three shortcut blocks each with three layers, five Conv-BN-Leaky-ReLU and in the end one Conv. Note that, different than CNN for classification, the last layer is a convolution layer so that the output is an image instead of a classification result. The kernel size of convolution layers is

kept constant as 3×3 with stride 1. No padding is applied for convolution, so that the feature map size decreases throughout the layers. For all but the last layer, the channel number is 128. More explanations are given as follows.

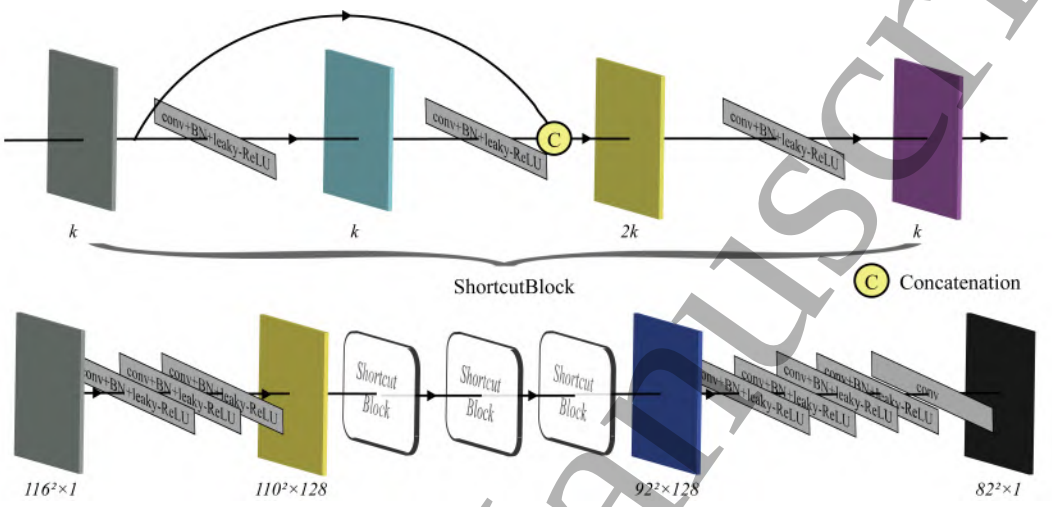

Fig 4 Network architecture for the proposed DeSpecNet.

2.3.1 Residual learning

Residue learning was first proposed by He *et al.*[30] to solve the performance degradation problem for very deep networks. As the residues are usually close to zero, the residual mapping is much easier to learn than the original mapping which is close to identity mapping. This assumption also holds for the case of despeckling. The proposed method formulates the whole network as a residue learning block, which learns the speckle residue from the original B-scan. Let $(\mathbf{x}_i, \mathbf{y}_i)$ denote the ith training image patch pair ($i = 1, \cdots, N$), and $\mathcal{R}(\mathbf{y}_i; \Theta)$ denote the predicted speckle residue given speckled image patch $\mathbf{y}$ and trainable parameters Θ, The loss function of the proposed DeSpecNet is formulated as:

$$l(\Theta) = \frac{1}{N} \sum_{i=1}^{N} \|\mathcal{R}(\mathbf{y}_i; \Theta) - (\mathbf{y}_i - \mathbf{x}_i)\|_1, \tag{1}$$

Here the L1 instead of L2 loss is used. As the L1 loss tolerates more outliers, and edges can be seen as outliers as opposed to smooth regions, the L1 loss is helpful to maintain edge sharpness in the output image.

2.3.2 Shortcut connectivity block

The proposed network applies three shortcut connectivity blocks (shortcut block) to further enhance the feature extraction performance. As is shown in Figure 4, each shortcut block consists of three Conv-BN-LeakyReLU connection in sequence. In each block, the input of first convolutional layer and the output of the second Leaky ReLU are concatenated in channel dimension, as input of the third convolutional layer. Similar to dense blocks,[37] shortcut blocks allow feature reuse among adjacent layers. The shortcut connection makes the early-stage feature maps more closely connected to the output, and thus more efficiently updated with gradients of the loss function, consequently leading to more compact and more accurate models. It also helps to avoid performance degradation or overfitting, and require fewer parameters and less computation to achieve comparable performance than the traditional CNN. Therefore, it helps the proposed network to achieve good despeckling performance with limited size of training data.

2.3.3 Batch normalization and Leaky-ReLU

Batch normalization[38] was proposed to alleviate the internal covariate shift by scaling and shifting before the nonlinearity activation. It has been widely used in classification or image generation tasks. By reducing the statistical differences between training samples, BN has several advantages, such as accelerating training, preventing gradient vanishing or explosion. It also serves as regularization term that enhances the generalization ability of the network.

Leaky-ReLU is a variant of ReLU. While ReLU can only output feature maps with positive elements, leaky-ReLU keeps the non-linearity and preserves feature with negative number.leaky-ReLU may benefit the training process by preventing dead neurons with zero output and zero gradient.

2.4 Implementation details

In training, the input B-scans are cropped to 116×116 overlapping patches with stride of 50. Some boundary regions are excluded to avoid artifacts caused by registration-averaging or flattening. The overlapping sampling can be seen as a data augmentation method that helps to effectively utilize the training data. A total of 90,112 sample patches were generated for training. The Adam optimizer was used for training. The initial learning rate was 0.0001. Truncated normal initializer was used for weight initialization. The training batch size was 26. The network was trained for 35,000 steps, when convergence was reached. Note that in testing, no cropping was applied, and the whole B-scan was input into the trained network. The proposed method were coded in Python based on Tensorflow and run on a PC with Intel Xeon CPU E2-2683 v3 @ 2.00GHz with 64G RAM, accelerated using the NVIDIA GTX Titan X GPU with 12G memory.

3 Experimental Results

3.1 Testing data

We tested our data on eleven OCT volumes obtained from four types of OCT scanners, from t-wo different manufacturers, covering the two types of center wavelength (840nm and 1050nm). Table 1 listed the specifications of the two groups of training data volumes and eleven test data volumes. The testing scans were from normal or pathological eyes. The pathological eyes were

from subjects with central serous chorioretinopathy (CSC), pathological myopia (PM), age-related macular degeneration(AMD) or diabetic macular edema(DME). From each test OCT volume, four B-scans, including two in the peripheral area and two close to the center, were selected for quantitative evaluation. All OCT data were uncompressed and unprocessed raw data exported from the scanners. The study was approved by the Institutional Review Board of Soochow University, and informed consent was obtained from all subjects.

Table 1 Specifications of training and testing data

	Scanner	Center wave-length (nm)	B-scan size (pixels)	Location	Normal/ Pathological
training 1	Topcon DRI-1	1050	512×992	macula	Normal
training 2	Topcon 2000	840	512×885	macula	Normal
testing 1	Topcon DRI-1	1050	512×992	macula	Normal
testing 2			512×992	macula+ONH	Normal
testing 3			512×992	macula	Pathological(CSC)
testing 4	Topcon 1000	840	512×480	macula	Normal
testing 5			512×480	macula	Pathological(CSC)
testing 6	Topcon 2000	840	512×885	macula	Normal
testing 7			512×885	ONH	Normal
testing 8			512×885	macula	Pathological(AMD)
testing 9			512×885	macula	Pathological(DME)
testing 10	Zeiss Cirrus 4000	840	512×1024	macula	Pathological (PM)
testing 11			512×1024	macula	Pathological(CSC)

3.2 Methods for comparison

We compare the proposed method with state-of-the-art approaches for general image denoising and for OCT despeckling, including non-local means (NLM),[39,40] block-matching and 3D filtering (BM3D),[22,41] sparsifying transform learning and low-rank method (STROLLR),[42,43] deep CNN with residual learning (DnCNN),[29,44] 3D complex wavelet based K-SVD for OCT denoising,[19,45]

maximum-a-posteriori (MAP) estimation based on local statistical model for OCT denoising.[26,46] In these experiments, parameters for NLM and BM3D 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 NLM, the template window size was 7, the search window size was 21, and the filter strength was 10. For BM3D, the noise standard deviation was set as 40.

As the network is trained to simulate spatial/temporal compounding, we implement an intra-volume compounding method for comparison. For each B-scan, the nearest L B-scans are registered to it, and they are averaged to get the despeckling result. The number of averaged B-scans and the registration method are exactly the same as in training data computation.

We also compare the results with three variations of the network. The first model uses L2 loss instead of L1 loss. The second model replaces each shortcut block with a residue block, which means simply replacing the concatenation inside the block with summation. The third model uses ReLU instead of Leaky ReLU.

For fair comparison, all resulting images are enhanced by linear intensity mapping to the full dynamic range.

3.3 Qualitative Evaluation

The despeckling results for B-scans from the eleven testing volumes are shown in Figure 5. It can be seen that the proposed method performs well for all cases, removing the speckles in different regions while preserving the edges and structure details. Both the retina layers and the choroid vessels are visually enhanced.

Despeckling results for two B-scans are further shown in Figure 6 and 7, compared with

results by other methods. It can be seen that the block matching methods, NLM and BM3D, both result in artifacts inside the retinal layers and blurred boundaries between layers. STROLLER gives better appearance inside the retinal layers, but has distortion near strong edges. K-SVD gives oversmoothed results for testing data 1. It works better for testing data 7, but the edges are also blurred. The MAP method does insufficient smoothing for both cases. The results by intra-volume compounding method still have speckles left inside layers, and edges can be blurred, especially when the retinal structure have big changes across adjacent B-scans, such as in testing data 7. In general, the CNN-based methods work much better than the other methods. All the five methods are able to recover the entire external limiting membrane, a thin structure above the RPE complex with high intensity, which is barely discernible in the original image. However, for DnCNN, there are sometimes overshooting artifacts near strong edges, such as the black shadows in the concave part shown in Figure 7(h). Using L2 loss, the results are slightly blurred than those of the proposed method. For the method with res-blocks, there are sometimes dot-like artifacts in both background and retinal regions, such as shown in Figure 6(h). For the method with ReLU, the resulting retinal regions are less smoothed than the proposed method with Leaky ReLU.

3.4 Quantitative Evaluation

In this section, four performance indices are used to quantitatively compare the performance of different denoising algorithms for the test images, including signal-to-noise ratio (SNR), contrast-to-noise ratio (CNR), equivalent number of looks (ENL) and edge preservation index (EPI). For calculating the indices, a background region of interest (ROI) , three signal ROIs and three retinal boundaries are manually selected, as shown in Figure 5. The background ROI (in green) is a rectangular region randomly selected above the retina. Three signal ROIs (in red) are located

in the retinal neural fiber layer (RNFL), inner retina, and the retinal pigment epithelium (RPE) complex, respectively. Three boundaries (in blue) are the upper boundary of RNFL, inner-outer retina boundary and the lower boundary of RPE.

The indices are calculated as follows:

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

$$CNR_i = 10\log_{10}\left(\frac{|\mu_i - \mu_b|}{\sqrt{\sigma_i^2 + \sigma_b^2}}\right), \tag{3}$$

$$ENL_i = \frac{\mu_i^2}{\sigma_i^2}, \tag{4}$$

where $\max(I)$ is the maximum pixel intensity of the B-scan I, μ_b and σ_b denote the mean and standard deviation of the background region, and μ_i and σ_i denote the mean and standard deviation of i-th signal region. By definition, SNR measures the homogeneity of the background, CNR reflects both the contrast between foreground and background, and the homogeneity of both areas, and ENL represents the signal strength and homogeneity of foreground regions. In our experiment, mean CNR and ENL are calculated over the three signal regions.

$$EPI = \frac{\sum_i \sum_j \left|I_d\left(i+1,j\right) - I_d\left(i,j\right)\right|}{\sum_i \sum_j \left|I_o\left(i+1,j\right) - I_o\left(i,j\right)\right|} \tag{5}$$

where I_o and I_d represent the original image and the despeckled image, while i and j represent coordinates in the vertical and horizontal direction in the image. EPI is a measure of edge sharpness, but it can be also high in areas of strong noise. With the layered retinal structure, the vertical gra-

dients are much larger than the horizontal ones. Therefore only vertical gradients are considered in calculating EPI. To focus on the edges, the EPIs are calculated in the neighborhoods of the three manually delineated boundaries.

The mean and standard deviation of performance indices for different methods, calculated over the 44 test B-scans, are listed in Table 2. NLM has high SNR, indicating that it performs well in background denoising, but the low CNR and ENL show that it doesn't work well in retinal regions. The high EPI may be result of artifacts near boundaries. BM3D is low in SNR, with poor performance in background. STROLLR is quite balanced in all indices, but the CNR, ENL and EPI are all lower than the proposed DeSpecNet. K-SVD has very high SNR, CNR and ENL, but low EPI, which can be associated with the oversmoothed results. MAP is low in all indices, especially SNR and EPI. The intra-volume compounding gives low SNR and CNR. In general, the CNN-based methods obtain high performance indices. Compared with DnCNN, the proposed DespecNet with shortcut block and leaky ReLU is higher in CNR, ENL and EPI. Using L1 instead of L2 loss, the proposed method is higher in SNR, CNR, and EPI, but lower in ENL. This indicates that the L2 loss results in more smoothed retinal regions but also blurs the edges. By using shortcut block instead of res-block, the SNR is improved while CNR and ENL decreases a bit. By using leaky ReLU instead of ReLU, the SNR, CNR, and ENL are increased. It can be concluded that the proposed DeSpecNet has the best overall performance regarding the quantitative indices.

3.5 Computational time

The computational time of all deep learning based methods are listed in Table. 3. For all networks, training is finished until the loss converges. The testing time for each B-scan is the average for 2000 B-scans with size 992×512. As DnCNN has the most simple structure, it requires the short-

Table 2 Comparison of performance indices.

	SNR	CNR	ENL	EPI
Original	26.16±1.27	4.32±1.07	34.10±12.86	1.00±0.00
NLM	44.02±3.04	5.82±1.67	58.63±43.55	1.06±0.10
BM3D	34.57±1.76	8.18±0.96	102.66±48.75	0.80±0.11
STROLLR	41.65±2.05	8.03±1.56	114.95±98.18	0.79±0.10
K-SVD	49.69±2.18	8.93±2.04	223.70±320.53	0.78±0.13
MAP	31.22±1.33	7.00±1.48	119.49±56.08	0.76±0.09
Intra-volume Compounding	33.24±2.03	8.45±1.16	107.12±45.83	0.90±0.12
DnCNN	40.72±2.58	9.47±1.36	154.17±83.84	0.81±0.11
Shortcut block + Leaky ReLU + L2 loss	39.27±4.58	9.31±1.66	186.92±116.67	0.83±0.08
Res-block + Leaky ReLU + L1 loss	38.42±3.37	9.73±1.68	173.98±111.22	0.91±0.09
Shortcut block + ReLU + L1 loss	39.37±4.06	9.16±1.60	124.02±66.09	0.91±0.09
DeSpecNet (Shortcut block + Leaky ReLU + L1 loss)	40.17±6.00	9.67±1.72	166.23±97.52	0.91±0.09

est training and testing time. The four model variations including the proposed DeSpecNet take longer time to converge. However, as the training is done off-line, the several hour training time is acceptable. For testing, the proposed network takes slightly longer than the other model variations. Other methods compared requires no off-line training. With the current implementation (in MATLAB or mixed MATLAB/C), their testing time ranges from seconds to minutes. We don't list their time cost in detail here because it is unfair to compare the time cost of methods with different software platforms, especially when all deep learning methods use GPU for acceleration. Still, by the nature of deep learning, the proposed method is easily accelerated by parallel computation, and the testing time can reach the real-time requirement of clinical applications.

4 Discussion and Conclusions

In this paper, we propose a deep learning network for speckle reduction in OCT B-scans. Residue learning is adopted to make the model more stable and easier to optimize. Compared with DnC-

Table 3 Comparison of computational time.

Method	Training (min)	Testing(s)
DnCNN	141	0.11
Shortcut block + Leaky ReLU + L2 loss	300	0.36
Res-block + Leaky ReLU + L1 loss	385	0.35
Shortcut block + ReLU + L1 loss	306	0.34
DeSpecNet (Shortcut block + Leaky ReLU + L1 loss)	356	0.38

NN[29] which is mainly based on Conv-BN-ReLU modules, the proposed DeSpecNet has the same depth, but is improved with shortcut connectivity blocks, leaky ReLU and L1 loss function, and the number of channels are doubled. Experimental results show that the proposed method outperforms DnCNN. The advantages of using L1 loss instead of L2 loss, using concatenation for the shortcut connection instead of summation as in residue blocks, and using leaky ReLU instead of ReLU, are also demonstrated experimentally. The better visual quality and performance indices indicate better feature extraction and representation ability of the proposed network.

The proposed method also well outperforms other state-of-the-art methods for image denoising or OCT despeckling. It appears that many methods designed for natural images, such as NLM,[39] BM3D[22] and STROLLR[42] do not work well with OCT images. The poor performance may be caused by the difficulty of finding matching blocks or learning the underlying models with the complexity and dominance of OCT speckles. Methods designed for OCT images, such as K-SVD[19] and MAP[26] may work for certain types of OCT images as reported, but the generalization ability is poor. By contrast, the proposed method achieves good despeckling results for all types of OCT images tested, including images taken at different retina locations, with different resolutions, and both normal and pathological cases.

It is also showed that the proposed method outperforms the intra-volume compounding method.

Learning from the results of multi-volume spatial/temporal compounding, the proposed method offers a way to simulate the despeckling effect offered by temporal compounding. Although it cannot reach the performance of real temporal compounding by repeated scanning, it has the advantage of software-based methods as discussed in Introduction.

Additionally, the means for obtaining "clean" images as ground truth and the patch-based overlapping sampling strategy in training are specifically designed for OCT images. The registration-averaging method for obtaining training data only depends on the output of commercial scanners and a number of normal eyes as subjects, and good performance is obtained by limited number of samples. Based on patch-based training, the model can automatically select representative features that helps to decrease training loss, providing generalization ability for different types of OCT images.

In this work, we use three eyes for computing the training data and show that, even with such a limited number of training subjects, the model works for a variety of test images. Three is the minimum requirement to balance the data of two optical systems with different center wavelength. If more volunteers are available, we can scan the same number of subjects with both scanners, and discard half of the B-scans from the first scanner. In this way both the number of subjects and the number of B-scans are balanced between the two scanners. By including more subjects for training, the generalization ability of the network can be surely improved.

Besides, we only use normal eyes for training, because the registration-average method do not work well for images from pathological eyes. First, it's difficult for patients to maintain stable fixation during multiple acquisitions, leading to big differences between volumes. Then, pathologies often cause drastic change in structure between adjacent B-scans. Furthermore, images from pathological eyes are often with lower quality. All these factors can make registration difficult

and high quality training set hard to obtain. In the future, we will investigate advanced registration methods or other alternative ways to obtain speckled/clean image pairs for pathological eyes. Inclusion of these data in training will probably improve the performance of the model.

In summary, by applying the proposed DeSpecNet, the quality of OCT B-scans are improved effectively, with speckles suppressed, edges preserved and contrast enhanced simultaneously. It is promising that this preprocessing can help with manual inspection of the OCT as well as improving the performance of following automatic OCT analysis methods. Furthermore, accelerated by GPU, the computational time can readily meet the real-time demand of clinical practice. In the future, we will further exploit the improvement of the method by incorporating fuzzy metrics, sparse coding and hybrid domain processing into the proposed strategy.[47–52] We will also further investigate the contribution of the proposed OCT despeckling method to specific applications, such as manual diagnosis of certain retinal pathologies, and automatic segmentation of retinal structures or lesions.

Disclosures

The authors have no relevant financial interests in the manuscript and no other potential conflicts of interest to disclose.

Acknowledgments

This work was supported in part by the State Key Project of Research and Development Plan under Grant 2017YFA0104302, Grant 2017YFC0109202 and 2017YFC0107900, in part by the National Natural Science Foundation of China (NSFC) under Grant 61622114, 81530060, 61871117 and 61771326, and in part by the National Basic Research Program of China (973 Program) under Grant 2014CB748600.

The authors thank Dr. Haoyu Chen, Joint Shantou International Eye Center, China and Dr. Songtao Yuan, Jiangsu Provincial Hospital, China for providing some of the test data.

References

1 J. M. Schmitt, S. Xiang, and K. M. Yung, "Speckle in optical coherence tomography," *Journal of Biomedical Optics* **4**(1), 95–106 (1999).

2 K. Yu, F. Shi, E. Gao, *et al.*, "Shared-hole graph search with adaptive constraints for 3D optic nerve head optical coherence tomography image segmentation," *Biomedical Optics Express* **9**(3), 962 (2018).

3 D. Xiang, H. Tian, X. Yang, *et al.*, "Automatic segmentation of retinal layer in OCT images with choroidal neovascularization," *IEEE Transactions on Image Processing* **27**(12), 5880 – 5891 (2018).

4 W. Zhu, L. Zhang, F. Shi, *et al.*, "Automated framework for intraretinal cystoid macular edema segmentation in three-dimensional optical coherence tomography images with macular hole," *Journal of Biomedical Optics* **22**(7), 76014 (2017).

5 J. Guo, W. Zhu, F. Shi, *et al.*, "A framework for classification and segmentation of branch retinal artery occlusion in SD-OCT," *IEEE Transactions on Image Processing* **26**(7), 3518 – 3527 (2017).

6 N. Iftimia, B. E. Bouma, and G. J. Tearney, "Speckle reduction in optical coherence tomography by "path length encoded" angular compounding," *Journal of Biomedical Optics* **8**(2), 260–263 (2003).

7 A. E. Desjardins, B. J. Vakoc, W. Y. Oh, *et al.*, "Angle-resolved optical coherence tomogra-

phy with sequential angular selectivity for speckle reduction," *Optics Express* **15**(10), 6200 (2007).

8 B. F. Kennedy, T. R. Hillman, A. Curatolo, *et al.*, "Speckle reduction in optical coherence tomography by strain compounding," *Optics Letters* **35**(14), 2445 (2010).

9 D. Alonso-Caneiro, S. A. Read, and M. J. Collins, "Speckle reduction in optical coherence tomography imaging by affine-motion image registration," *Journal of Biomedical Optics* **16**(11), 116027 (2011).

10 M. Pircher, E. Gotzinger, R. Leitgeb, *et al.*, "Speckle reduction in optical coherence tomography by frequency compounding," *Journal of Biomedical Optics* **8**(3), 565–569 (2003).

11 A. Ozcan, A. Bilenca, A. E. Desjardins, *et al.*, "Speckle reduction in optical coherence tomography images using digital filtering," *Journal of the Optical Society of America A Optics Image Science & Vision* **24**(7), 1901–10 (2007).

12 H. Rabbani, R. Nezafat, and S. Gazor, "Wavelet-domain medical image denoising using bivariate Laplacian mixture model.," *IEEE Transactions on Biomedical Engineering* **56**(12), 2826–2837 (2009).

13 M. A. Mayer, A. Borsdorf, M. Wagner, *et al.*, "Wavelet denoising of multiframe optical coherence tomography data," *Biomedical Optics Express* **3**(3), 572–589 (2012).

14 F. Zaki, Y. Wang, H. Su, *et al.*, "Noise adaptive wavelet thresholding for speckle noise removal in optical coherence tomography," *Biomedical Optics Express* **8**(5), 2720–2731 (2017).

15 Z. Jian, Z. Yu, L. Yu, *et al.*, "Speckle attenuation in optical coherence tomography by curvelet shrinkage," *Optics Letters* **34**(10), 1516 (2009).

16 B. Rao, B. J. Tromberg, L. Yu, *et al.*, "Three-dimensional speckle suppression in optical coherence tomography based on the curvelet transform," *Optics Express* **18**(2), 1024–1032 (2010).

17 L. Fang, S. Li, Q. Nie, *et al.*, "Sparsity based denoising of spectral domain optical coherence tomography images," *Biomedical Optics Express* **3**(5), 927–942 (2012).

18 L. Fang, S. Li, D. Cunefare, *et al.*, "Segmentation based sparse reconstruction of optical coherence tomography images," *IEEE Transactions on Medical Imaging* **36**(2), 407–412 (2017).

19 R. Kafieh, H. Rabbani, and I. Selesnick, "Three dimensional data-driven multi scale atomic representation of optical coherence tomography," *IEEE Transactions on Medical Imaging* **34**(5), 1042–1062 (2015).

20 X. Zhang, L. Li, F. Zhu, *et al.*, "Spiking cortical model-based nonlocal means method for speckle reduction in optical coherence tomography images," *Journal of Biomedical Optics* **19**(6), 066005 (2014).

21 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," *Applied Optics* **54**(13), 13–14 (2015).

22 K. Dabov, A. Foi, V. Katkovnik, *et al.*, "Image denoising by sparse 3-d transform-domain collaborative filtering," *IEEE Transactions on Image Processing* **16**(8), 2080–2095 (2007).

23 B. Chong and Y. K. Zhu, "Speckle reduction in optical coherence tomography images of human finger skin by wavelet modified BM3D filter," *Optics Communications* **291**(6), 461–469 (2013).

24 M. Akshaya, W. Alexander, D. A. Clausi, *et al.*, "General Bayesian estimation for speckle noise reduction in optical coherence tomography retinal imagery," *Optics Express* **18**(8), 8338–52 (2010).

25 A. Cameron, D. Lui, A. Boroomand, *et al.*, "Stochastic speckle noise compensation in optical coherence tomography using non-stationary spline-based speckle noise modelling.," *Biomedical Optics Express* **4**(9), 1769–1785 (2013).

26 M. Li, R. Idoughi, B. Choudhury, *et al.*, "Statistical model for OCT image denoising," *Biomedical Optics Express* **8**(9), 3903–3917 (2017).

27 I. Kopriva, F. Shi, and X. Chen, "Enhanced low-rank + sparsity decomposition for speckle reduction in optical coherence tomography.," *Journal of Biomedical Optics* **21**(7), 76008 (2016).

28 J. Cheng, D. Tao, Q. Ying, *et al.*, "Speckle reduction in 3D optical coherence tomography of retina by A-scan reconstruction," *IEEE Transactions on Medical Imaging* **35**(10), 2270–2279 (2016).

29 K. Zhang, W. Zuo, Y. Chen, *et al.*, "Beyond a Gaussian denoiser: Residual learning of deep cnn for image denoising," *IEEE Transactions on Image Processing* **26**(7), 3142–3155 (2017).

30 K. He, X. Zhang, S. Ren, *et al.*, "Deep residual learning for image recognition," in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, 770–778 (2016).

31 P. Wang, H. Zhang, and V. M. Patel, "SAR image despeckling using a convolutional neural network," *IEEE Signal Processing Letters* **24**(12), 1763–1767 (2017).

32 Q. Zhang, Q. Yuan, J. Li, *et al.*, "Learning a dilated residual network for SAR image despeckling," *Remote Sensing* **10**(2), 196 (2018).

33 F. Dietrichson, E. Smistad, A. Ostvik, *et al.*, "Ultrasound speckle reduction using generative adversial networks," in *2018 IEEE International Ultrasonics Symposium (IUS)*, IEEE (2018).

34 D. Mishra, S. Chaudhury, M. Sarkar, *et al.*, "Ultrasound image enhancement using structure oriented adversarial network," *IEEE Signal Processing Letters* **25**(9), 1349–1353 (2018).

35 Z. Wang, A. C. Bovik, H. R. Sheikh, *et al.*, "Image quality assessment: from error visibility to structural similarity," *IEEE Transactions on Image Processing* **13**(4), 600–612 (2004).

36 F. Shi, X. Chen, H. Zhao, *et al.*, "Automated 3-D retinal layer segmentation of macular optical coherence tomography images with serous pigment epithelial detachments.," *IEEE Transactions on Medical Imaging* **34**(2), 441–452 (2015).

37 H. Gao, L. Zhuang, and K. Q. Weinberger, "Densely connected convolutional networks," in *2017 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, (2017).

38 S. Ioffe and C. Szegedy, "Batch normalization: Accelerating deep network training by reducing internal covariate shift," *arXiv preprint arXiv:1502.03167* (2015).

39 A. Buades, B. Coll, and J.-M. Morel, "A non-local algorithm for image denoising," in *Computer Vision and Pattern Recognition, CVPR 2005. IEEE Computer Society Conference on*, **2**, 60–65, IEEE (2005).

40 J. V. Manjon-Herrera and A. Buades, "Non-Local Means Filter." https://ww2.mathworks.cn/matlabcentral/fileexchange/13176-non-local-means-filter (2008).

41 K. Dabov, A. Danieyan, and A. Foi, "BM3D MATLAB Software." http://www.cs.tut.fi/~foi/GCF-BM3D/ (2014).

42 B. Wen, Y. Li, and Y. Bresler, "When sparsity meets low-rankness: Transform learning with non-local low-rank constraint for image restoration," in *IEEE International Conference on Acoustics*, (2017).

43 B. Wen, Y. Li, and Y. Bresler, "Image Denoising Codes using STROLLR learning." https://github.com/wenbihan/strollr2d_icassp2017 (2018).

44 K. Zhang, W. Zuo, Y. Chen, *et al.*, "DnCNN." https://github.com/cszn/DnCNN (2018).

45 R. Kafieh, "K-SVD Denoise for OCT." https://sites.google.com/site/rahelekafieh/research/state-of-the-art-method-for-oct-denoising (2012).

46 M. Li, R. Idoughi, B. Choudhury, *et al.*, "OCT Denoising Code." https://github.com/vccimaging/OCTDenoising (2017).

47 Y. Wei, H. Zhang, J. Yang, *et al.*, "Low-dose CT image postprocessing based on residual convolutional network," *IEEE Access* **5**, 24698–24705 (2017).

48 C. Yang, Y. Zhang, H. Shu, *et al.*, "Structure-adaptive fuzzy estimation for random-valued impulse noise suppression," *IEEE Transactions on Circuits & Systems for Video Technology* **28**(2), 414 427 (2018).

49 J. Liu, J. Ma, Y. Zhang, *et al.*, "Discriminative feature representation to improve projection data inconsistency for low dose CT imaging," *IEEE Transactions on Medical Imaging* **36**(12), 2499 – 2509 (2017).

50 Y. Chen, L. Shi, Q. Feng, *et al.*, "Artifact suppressed dictionary learning for low-dose CT image processing," *IEEE Transactions on Medical Imaging* **33**(12), 2271–2292 (2014).

51 Y. Chen, J. Ma, Q. Feng, *et al.*, "Nonlocal prior Bayesian tomographic reconstruction," *Journal of Mathematical Imaging & Vision* **30**(2), 133–146 (2008).

52 X. Yin, Q. Zhao, J. Liu, *et al.*, "Domain progressive 3d residual convolution network to improve low dose CT imaging," *IEEE Transactions on Medical Imaging* (2019). DOI: 10.1109/TMI.2019.2917258.

Fei Shi is an associate professor at Soochow University, Suzhou, China. She received her Ph.D. degree in electrical engineering from Polytechnic University, United States in 2006. She has co-authored over 40 papers in internationally recognized journals and conferences. Her current research interests include OCT image despeckling, segmentation, and classification.

Ning Cai received the B.S degree from Anhui University in 2016. He is currently pursuing the M.S. degree in computer science with Southeast University. His interests are in image processing and deep learning. He is now working on medical image noise reduction algorithms.

Yang Chen received the M.S. and Ph.D. degrees in biomedical engineering from First Military Medical University, China, in 2004 and 2007, respectively. Since 2008, he has been a faculty member with the Department of Computer Science and Engineering, Southeast University, China. His recent work concentrates on medical image reconstruction, image analysis, pattern recognition, and computerized-aid diagnosis.

Xinjian Chen is a distinguished professor at Soochow University, Suzhou, China. He received Ph.D. degree from the Chinese Academy of Sciences in 2006. He conducted postdoctoral research in University of Pennsylvania, National Institute of Health, and University of Iowa, USA from 2008 to 2012. He has published over 100 top international journal and conference papers. He

has also been granted with 6 patents. His current research focus is medical image processing and analysis.

Biographies of the other authors are not available.

List of Figures

List of Tables

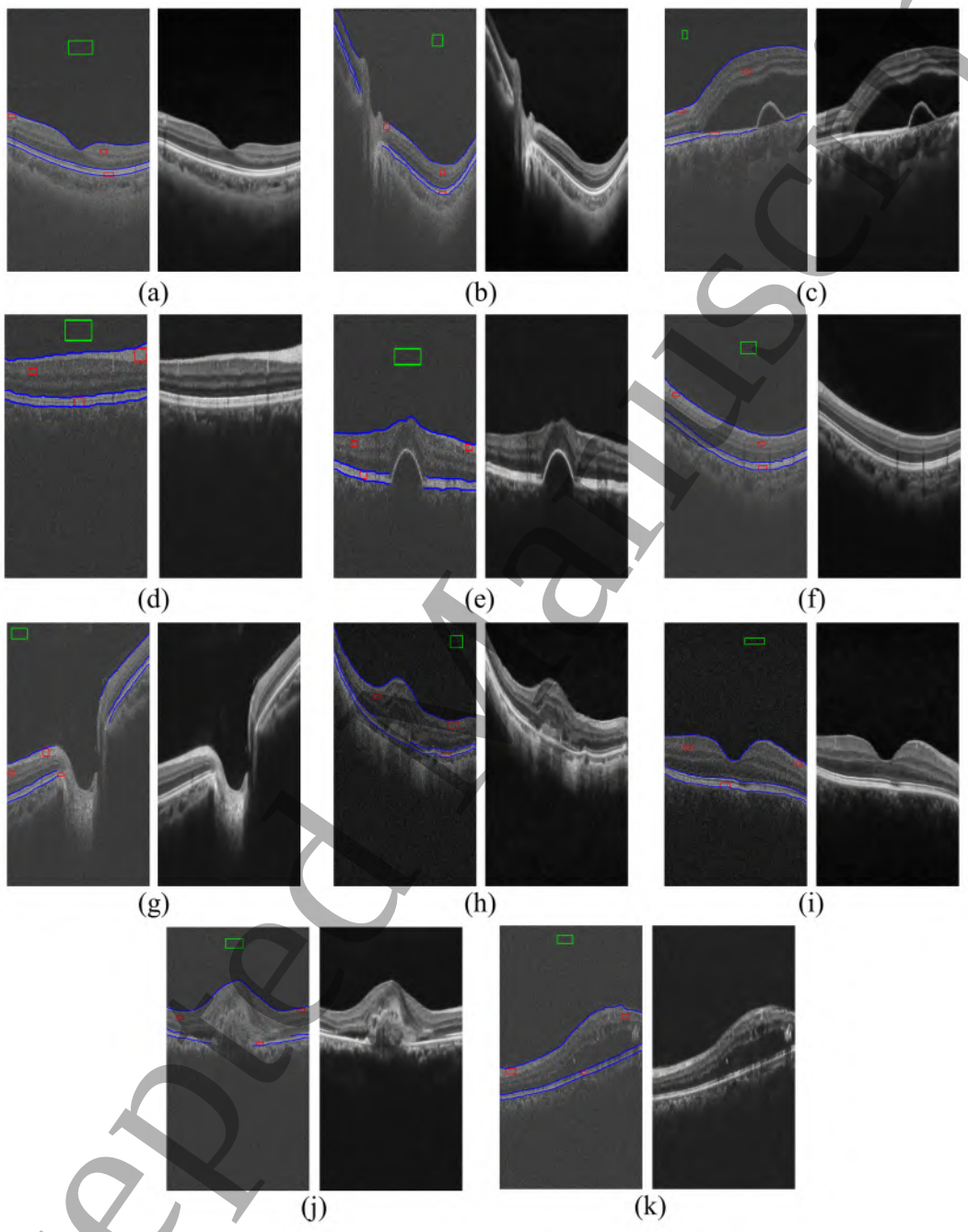

Fig 5 Despeckling results for the B-scans from the eleven testing volumes. (Some are scaled for display purpose) (a)-(k) corresponds to testing data 1-11, respectively. For each panel, left: original B-scan, right: despeckled B-scan. In each original B-scan, the green rectangle represents the background ROI, the red rectangles represents the signal ROIs, and the blue curves represent boundaries for calculating EPI.

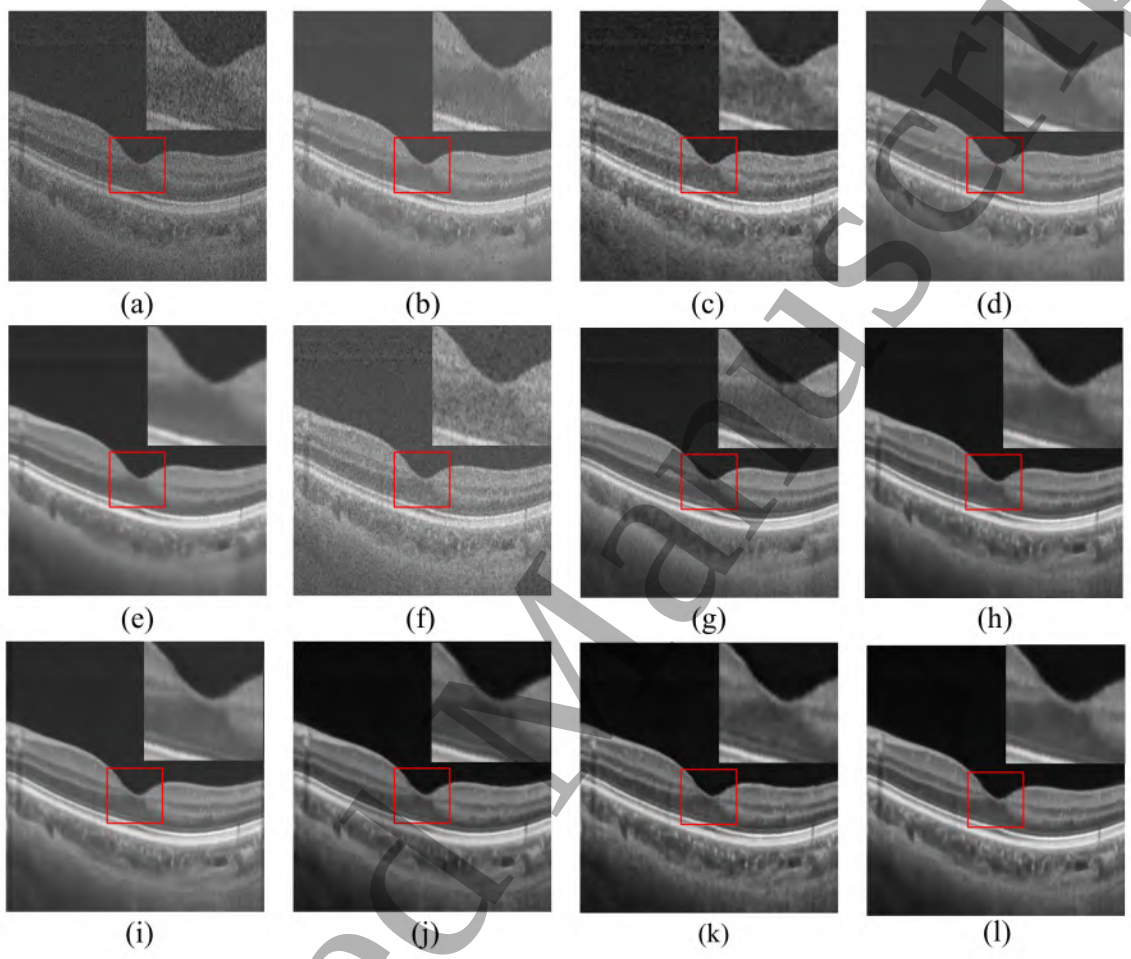

Fig 6 Results for one B-scan of testing data 1. Part of the background are cropped. The regions in red rectangle are zoomed. (a)Original image (b) NLM(c) BM3D (d) STROLLR (e) K-SVD (f) MAP (g) Intra-volume compounding (h) DnCNN (i) Shortcut block + Leaky ReLU + L2 loss (j) Res-block + Leaky ReLU + L1 loss (k) Shortcut block + ReLU + L1 loss (l) Proposed DeSpecNet(Shortcut block + Leaky ReLU + L1 loss)

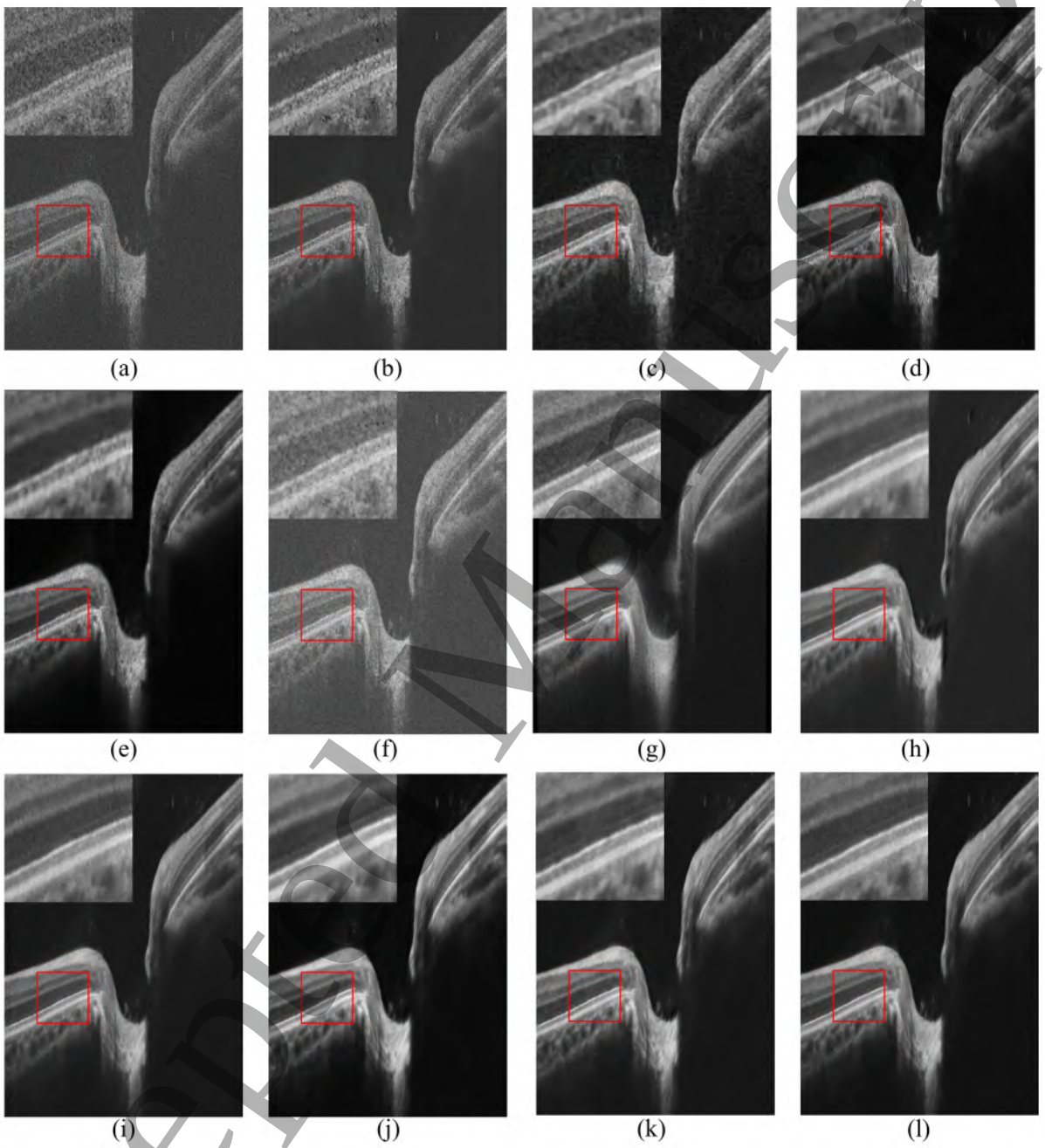

Fig 7 Results for one B-scan of testing data 7. Part of the background are cropped. The regions in red rectangle are zoomed. (a)Original image (b) NLM(c) BM3D (d) STROLLR (e) K-SVD (f) MAP (g) Intra-volume compounding (h) DnCNN (i) Shortcut block + Leaky ReLU + L2 loss (j) Res-block + Leaky ReLU + L1 loss (k) Shortcut block + ReLU + L1 loss (l) Proposed DeSpecNet(Shortcut block + Leaky ReLU + L1 loss)