Research Article Vol. 14, No. 6 / 1 Jun 2023 / Biomedical Optics Express 2773

Biomedical Optics EXPRESS

Self-supervised Blind2Unblind deep learning scheme for OCT speckle reductions

XIAOJUN YU,[1,2] CHENKUN GE,[1] MINGSHUAI LI,[1] MIAO YUAN,[1] LINBO LIU,[3] JIANHUA MO,[4] PERRY PING SHUM,[5] AND JINNA CHEN[5,*]

[1] *School of Automation, Northwestern Polytechnical University, Xi'an, Shaanxi, 710072, China*
[2] *Research & Development Institute of Northwestern Polytechnical University in Shenzhen, Shenzhen, Guangzhou, 51800, China*
[3] *School of Electrical and Electronic Engineering, Nanyang Technological University, 639798, Singapore*
[4] *School of Electronics and Information Engineering, Soochow University, Suzhou, Jiangsu, 215006, China*
[5] *Department of Electrical and Electronic Engineering, Southern University of Science and Technology, Shenzhen, Guangdong, 518055, China*
[*] *chenjn@sustech.edu.cn*

Abstract: As a low-coherence interferometry-based imaging modality, optical coherence tomography (OCT) inevitably suffers from the influence of speckles originating from multiply scattered photons. Speckles hide tissue microstructures and degrade the accuracy of disease diagnoses, which thus hinder OCT clinical applications. Various methods have been proposed to address such an issue, yet they suffer either from the heavy computational load, or the lack of high-quality clean images prior, or both. In this paper, a novel self-supervised deep learning scheme, namely, Blind2Unblind network with refinement strategy (B2Unet), is proposed for OCT speckle reduction with a single noisy image only. Specifically, the overall B2Unet network architecture is presented first, and then, a global-aware mask mapper together with a loss function are devised to improve image perception and optimize sampled mask mapper blind spots, respectively. To make the blind spots visible to B2Unet, a new re-visible loss is also designed, and its convergence is discussed with the speckle properties being considered. Extensive experiments with different OCT image datasets are finally conducted to compare B2Unet with those state-of-the-art existing methods. Both qualitative and quantitative results convincingly demonstrate that B2Unet outperforms the state-of-the-art model-based and fully supervised deep-learning methods, and it is robust and capable of effectively suppressing speckles while preserving the important tissue micro-structures in OCT images in different cases.

© 2023 Optica Publishing Group under the terms of the Optica Open Access Publishing Agreement

1. Introduction

Optical coherence tomography (OCT) is a noninvasive high-resolution imaging modality and is capable of providing both cross-sectional and three-dimensional (3D) tissue microstructure images with a resolution of up to a few micrometers [1]. Over the past few decades, OCT has been adopted for various applications for its noninvasive and high-resolution properties [2]. Due to its inherent low-coherence interferometry nature, however, OCT inevitably suffers from the influences of speckles that are introduced by multiple forward and backscattering of illumination light. Speckles hide the critical tissue microstructures and reduce the contrast of OCT images, and make it difficult to identify tissue structure boundaries, which thus reduce the accuracy of disease diagnoses, and finally, hinder OCT clinical applications [3].

To alleviate the influences of speckles to OCT imaging, various despeckling methods have been proposed in literature over the past decades. Such methods can be roughly categorized into model-based and deep-learning-based ones [4]. Specifically, the model-based methods usually devise filters to reduce speckles according to certain image properties, such as inter-image

#481870 https://doi.org/10.1364/BOE.481870
Journal © 2023 Received 23 Nov 2022; revised 2 Mar 2023; accepted 9 Mar 2023; published 18 May 2023

Biomedical Optics EXPRESS

similarities, image sparsity, and some other domain components. According to the filtering scheme been adopted, the model-based methods could be further divided into spatial domain and transform domain ones. Non-local mean (NLM) filtering [5], total variation (TV) regularization [6], block-matching and three-dimensional filtering (BM3D) [7], non-local weighted sparse representation (NWSR) [8], and the two-step iteration method (TSI) [9] are typical spatial domain model-based methods. NLM filtering calculates neighborhood pixel weights based on image self-similarity [5], while BM3D employs a block level estimation for denoising [7]. TV regularization combines the inter-frame low rank prior and the intra-frame total anisotropic variation prior for despeckling [10], whereas NWSR adopts a sparse representation of multiple similar noisy and denoised patches to improve estimation of a new patch [8]. TSI divides OCT speckle noise into additive and multiplicative components and then adopts a two-step filtering scheme to suppress such components iteratively. Although satisfactory results could be achieved with those model-based methods, limitations still exist. BM3D may lose both edge boundaries and texture details for images with high complexity and low contrast, while NWSR destroys structural details of the reconstructed images using vectorization patches. Both NLM and TV regularization suffer from excessive smoothing in certain areas, and TSI computational efficiency must be improved. In contrast, the transform domain mechanisms of despeckling are different. Both wavelet and curvelet transform methods are converting time-domain images into Fourier or wavelet domain, followed by separating clean signals from the noises with appropriate filters designed [11,12]. However, since these methods are based mainly on the assumption that there are no spectrum overlaps between the clean and noisy signals in transform domain, despeckling efficiency is the main problem due to the dual role of speckles in OCT imaging.

With the rapid development advancement of artificial intelligence (AI) in recent years, various deep-learning methods have also been proposed for OCT despeckling [13]. The main advantage of these methods is that they could effectively remove speckles while retaining structural details using deep networks with appropriate learning schemes. Convolutional neural networks (CNNs) have already shown their excellence in OCT despeckling [14]. For example, Zhang *et al*. presented a deep CNN network, namely DnCNN, for image denoising [15]. With a batch normalization (BN) layer to address the gradient dispersion effect and a residual learning scheme to improve network learning ability, DnCNN outperforms the model-based methods under different noise levels. Super-Resolution CNN (SRCNN) is another deep learning method that has been proposed for image processing [16]. By direct learning the end-to-end mapping between corresponding low- and high-resolution images, SRCNN is typically used for both image super-resolution and denoising. Since both DnCNN and SRCNN use bicubic interpolation scheme in their training sessions, the computational load is relatively high. It is also worth noting that most of the learning-based denoising schemes are fully supervised, wherein noisy and clean image pairs are needed for training. In practice, however, it is difficult to acquire noise-free ground truth images for two reasons. First, due to the dual role of speckles in OCT imaging, it is impossible to separate the clean images from their noisy peers. Second, the possible solution, i.e., repeated frame acquisition at the same tissue position, is time-consuming and prone to motion artifacts, especially for *in vivo* imaging. Therefore, simple and effective despeckling schemes are highly desired in clinical practice.

To address the above issues with those fully supervised denoising schemes, another possible solution is self-supervised denoising (also named unsupervised denoising), which is attracting increasing research interests recently [17]. Ulyanov *et al*. proposed a single image depth learning model, namely, deep image prior (DIP), for image restoration with a deep network architecture [18]. By utilizing an implicit regularization strategy to fit corrupted images, DIP is effective for denoising, especially for images with low noise level. DIP achieves satisfactory performances, but it is shown to be less competitive as compared with the typical model-based BM3D. Another self-supervised denoising method, namely Self2Self, was also demonstrated by training a dropout

Biomedical Optics EXPRESS

denoiser with a Bernoulli sampler-generated pair and averaging over the predicted multiple instances [19]. Although Self2Self has no prerequisite but a single noisy image, it takes a long time for model training, which thus makes it unsuitable for real-time denoising. To save training time and overcome the information loss, a Blind2Unblind method was also proposed [20]. By employing a global-aware mask mapper for perception and training acceleration, and a re-visible loss-based training strategy for non-blind denoising, Blind2Unblind achieves satisfactory results without any noise information or model prior.

In recent years, some un-supervised or self-supervised methods have also been proposed for speckle noise reduction in OCT images. By decomposing the noisy images into content and noise spaces with an encoder first, and then adopting a generator to predict the denoised image contents with the extracted features, Huang *et al.* reported the first unsupervised method, namely DRGAN-OCT, for speckle reduction with a small number of network training only, yet without employing any matched image pairs [21]. Guo *et al.* proposed to employ GAN discriminator to distinguish the real noisy samples from the fake ones and then use the NLM method to improve the denoising performance [22]. Although both methods require no ground truth images, multiple adjacent similar B-scan images are still required. Zhou *et al.* [23] proposed to combine cross-scale CNN with an intra-/inter-patch-based transformer for unsupervised OCT despeckle. With the former extracting the local features while the latter to extract and merge the local and global features, such a method obtains good performances by using a reconstruction network to produce the final denoised result. Since both CNN and transformer are required for feature extractions, however, it is relatively computationally extensive with higher processing time needed. More recently, two new self-supervised methods have also been proposed. Specifically, Li *et al.* proposed a method, namely, MAP-SNR, for speckle reduction [24]. By randomly selecting adjacent pixel blocks from the original noise image to generate two similar subsampling ones as input and target, while a self-supervised strategy to map the relationship between adjacent pixels, MAP-SNR achieves reliable contribution in speckle noise reduction for single OCT image. On the contrary, by replacing the frame average method with 3 frames fusion images through pre-training, Rico-Jimenez *et al.* proposed a self-fusion neural network for real-time OCT image denoising [25]. Although results show that these methods achieve satisfactory results, a number of adjacent frame datasets are still needed for learning/fusion. Self-/unsupervised denoising methods are expected to be a potential candidate for real-time OCT despeckling owing to their relative lower requirement on supervision datasets, however, the balancing between performances and system complexity has always been an insurmountable problem, and therefore, simpler and more effective mechanism are highly desired in practice.

Due to the dual role of speckles in OCT images, which are regarded to be both the information carrier and the noises that affect the structure details, the loss function is divergent when using Blind2Unblind for OCT image training, and therefore, the Blind2Unblind cannot be employed directly for OCT despeckling. Inspired by the self-supervised Blind2Unblind denoising strategy, an improved self-supervised Blind2Unblind scheme, namely, Blind2Unblind network with refinement strategy (B2Unet), together with a new global-aware mask mapper and a re-visible loss function is devised for OCT speckle reductions. Specifically, the new re-visible loss function is devised to address the loss divergence issue of Blind2Unblind for OCT despeckling, while a novel refinement module is also devised and integrated onto the Blind2Unblind inference unit to improve the network performances. By dividing each noisy image into several mini-blocks first, and then set some pixels in those mini-blocks to be blind spots to generate volumetric noisy images with the devised global masks, the B2Unet finally feeds the volumetric noisy images, which are sampled by the global mask mapper, into the denoising network for despeckled image generation. The main contributions of this paper are as follows.

1. A self-supervised B2Unet network with a new global mask mapper and a re-visible loss function is proposed for the first time, to the best of our knowledge, for OCT speckle reductions.

Biomedical Optics EXPRESS

2. A new loss function is devised to address the divergence issue with B2Unet training, while its upper and lower convergence bounds are also analyzed theoretically.

3. A refinement strategy is devised and integrated onto B2Unet inference model to improve denoising performances of the overall network.

4. Experiments with different OCT images, with both down- and up-sampling processing, are conducted to compare B2Unet with those of the state-of-the-art existing methods for verification.

The rest of this paper is organized as follows. Section 2 presents the proposed B2Unet architecture. Section 3 theoretically analyzes the feasibility of the devised re-visible loss and gives its lower and upper bounds. Experiments are presented in Section 4 to compare B2Unet with the state-of-the-art existing methods. Section 5 concludes the whole paper.

2. Method

It is reported that pixels in OCT images do not exist independently, instead, the values of each pixel and its surrounding ones are conditionally related [26]. Hence, if a mask mapper is applied to an OCT image to break the correlation between their pixels, a neural network can employ the information of these masked pixels to infer their true values via the surrounding pixels. In this study, we introduce the masking strategy into the Blind2Unblind training model, and devise a B2Unet network for OCT despeckling [20].

2.1. B2Unet model

Figure 1 shows the diagram of B2Unet network. As shown, it consists of a training and an inference unit. For the training scheme, a noisy volume with blind spot mask is Ω_y generated first using a noisy OCT image y and a global masker $\Omega(\cdot)$. Such a noisy volume is then fed into a denoising network $f_\theta(\cdot)$ to generate a denoised volume $f_\theta(\Omega_y)$, which again is sampled by a global mask mapper to generate a pseudo denoised image $g(f_\theta(\Omega_y))$. Meanwhile, the original noisy image y is also input to the same denoising network $f_\theta(\cdot)$ without gradient updating to obtain a denoised image $f_\theta(y)$. Finally, the original noisy image y, the denoised image $f_\theta(y)$, as well as the generated pseudo denoised image are processed together to generate a denoised image by minimizing the re-visible loss. In this way, the B2Unet model is trained, which not only eliminates the information loss caused by blind spots, but also ensures the satisfactory training convergence. While for the inference scheme, While for the inference scheme, the noisy OCT images are input into the well-trained despeckling network model first to generate the denosied images, which are then input into a refinement module to generate the desired despeckled images. In this study, the denoising network adopts a typical U-net for training.

2.2. Global masker and global-aware mask mapper

Various methods have been proposed to train the hidden pixels with manual masking [27]. However, it is worth noting that those existing optimization functions focus only on the masking regions, and it may cause certain problems, such as reduced accuracy and slow convergence rate. To address such issues for B2Unet, a global masker is adopted to mask the noisy image first, and then all masks are re-projected into the same image after denoising.

Figure 2 presents the devised masking strategy. As we can seen, a mask image, the size of which is the same as that of the noisy image y, is divided into several blocks with 2×2 cells first, and then two pixels are randomly chosen as the blind spots in each cell, and the mask is obtained by randomly masking each 2×2 cell in image y. Meanwhile, the noisy image y is filtered by a kernel with both stride and padding to be 1 to generate y_c. By conducting $y_c \times mask$, y_m is created and y_{inv} is generated sequentially by performing y $\times(1 - mask)$. Finally, the masked image Ω_y is obtained after summing y_m and y_{inv}.

To improve then information exchange between the mask regions, a global mask mapper is adopted to map the denoised images on blind spots, which helps increasing the accuracy of noise

Biomedical Optics EXPRESS

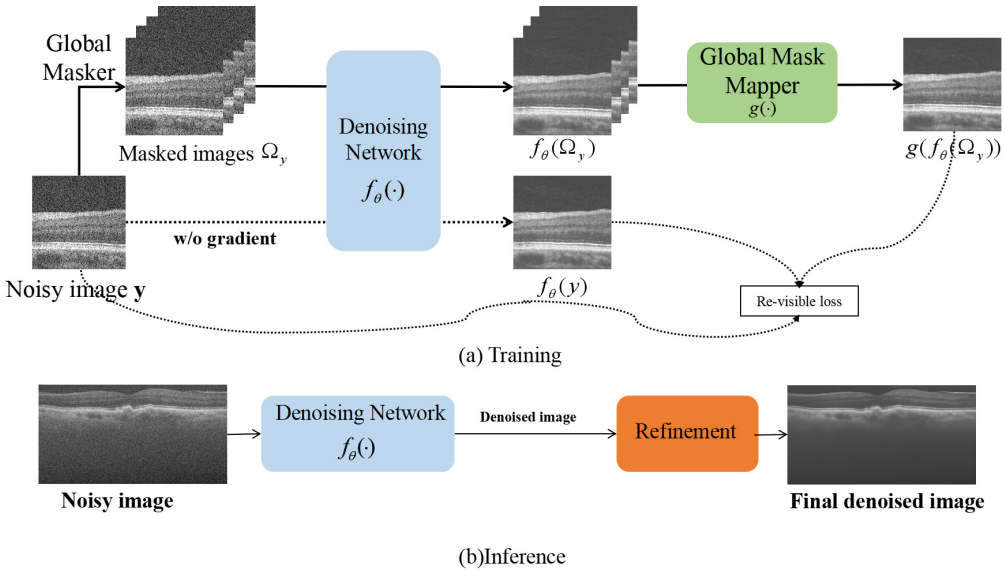

Fig. 1. Architecture diagram of the proposed B2Unet model.

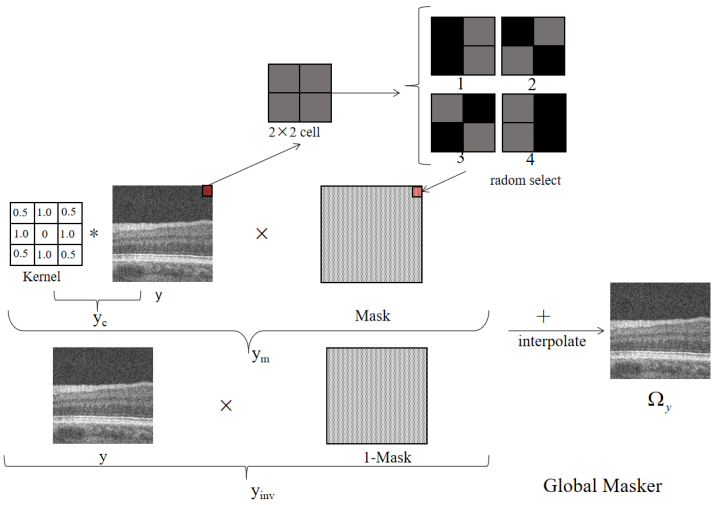

Fig. 2. Workflow of the global masker.

removal and speeding up the manual mask training. The workflow of this process is shown in Fig. 3, wherein the denoised images $f_\theta(\Omega_y)_i$ are multiplied by their corresponding masks first, and then, the pixel values are summarized to synthesize the pseudo denoised image $g(f_\theta(\Omega_y))$.

2.3. Denoising refinement module

Due to the randomness of the mask images been selected for the global masker during training, the denoised images obtained by using the denoising network alone would not be satisfactory results, since some of those denoised images may still contain speckle residues. To address such an issue, a denoising refinement module as shown in Fig. 4 is devised and integrated into inference unit of B2Unet. As seen, the denoised image is multiplied by four mask images first,

Biomedical Optics EXPRESS

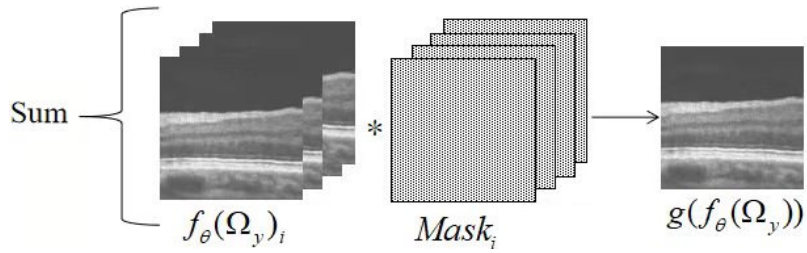

Fig. 3. Workflow of the global mask mapper.

and then the generated four images are input into the trained denoising network again. Finally, the four images generated by the denoising network are averaged to generate the final refined image. The mask generation in the refinement module of inference unit is similar to that in the training unit. While the mask image with the same size as the denoised image is divided into several blocks with 2×2 cells. For each block, the blind spot in the 2×2 cells of the mask image are selected one by one sequentially in a clockwise direction.

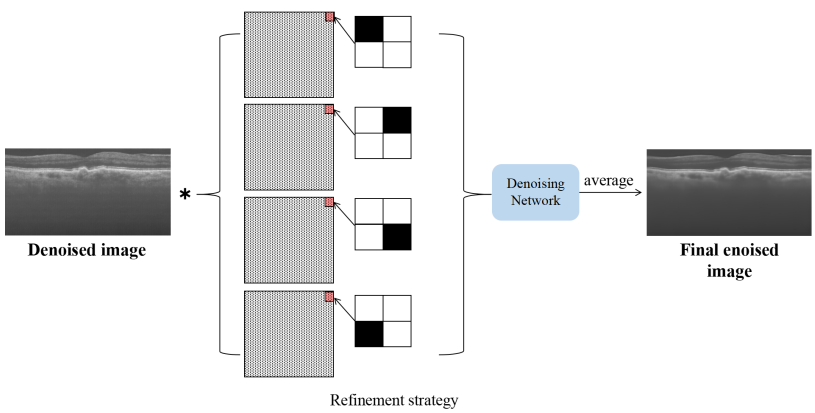

Fig. 4. Workflow of the refinement module in the inference unit.

2.4. Re-visible loss

To optimize B2Unet for OCT speckle reductions, a new re-visible loss function is designed. Specifically, the denoised images with blind spot denoising and non-blind spot denoising are combined to improve training stability, and the re-visible loss can be expressed as below,

$$L = \left\|g\left(f_\theta\left(\Omega_y\right)\right) - \hat{f}_\theta\left(y\right)\right\|_2^2 + (\lambda - 1)\left\|\hat{f}_\theta\left(y\right) - y\right\|_2^2 + 2\left\|g\left(f_\theta\left(\Omega_y\right)\right) - y\right\|_1 \cdot \left\|\hat{f}_\theta\left(y\right) - y\right\|_1 \tag{1}$$

where $f_\theta(\cdot)$ denotes the denoising network, $g(\cdot)$ denotes the result of global mask mapper, λ is a variable hyper-parameter that helps to avoid the divergence in model training. $\hat{f}_\theta(y)$ indicates that the noisy image y itself is obtained during the training without gradient updating.

3. Theory

By adopting blind spot schemes, self-supervised methods require less information, although the denoising effects are less significant as compared with the fully supervised ones.

Biomedical Optics EXPRESS

In this study, our main objective is to transform the invisible blind spots into visible ones [20], and the devised loss function for multi-task denoising is as follows,

$$arg\min_{\theta} E_y \left\| g\left(f_\theta\left(\Omega_y\right)\right) - y \right\|_2^2 + \lambda \cdot \left\| f_\theta\left(y\right) - y \right\|_2^2 \tag{2}$$

where Ω_y is the noisy masked volume that contains blind spots at all positions in image y. $g\left(\cdot\right)$ is the devised global-aware mask mapper, which denoises all the pixels at blind spots and maps them to a pseudo denoised image.

In this study, Eq. (2) is adopted as the objective training function for identity mapping. Furthermore, both the blind spots and the non-blind denoising schemes are combined to formulate a re-visible loss function for denoising, and the inequality between such two schemes are illustrated as follows,

$$\begin{aligned}
&\left\| g\left(f_\theta\left(\Omega_y\right)\right) - f_\theta(y) \right\|_2^2 \leq \left\| \left| g\left(f_\theta\left(\Omega_y\right)\right) - y \right| - \left| f_\theta(y) - y \right| \right\|_2^2 \\
&= \left\| g\left(f_\theta\left(\Omega_y\right)\right) - y \right\|_2^2 + \left\| f_\theta(y) - y \right\|_2^2 - 2\left\| (g\left(f_\theta\left(\Omega_y\right)\right) - y)^T (f_\theta(y) - y) \right\|_1 \\
&= \left\| g\left(f_\theta\left(\Omega_y\right)\right) - y \right\|_2^2 + \lambda \left\| f_\theta(y) - y \right\|_2^2 + (1-\lambda)\left\| f_\theta(y) - y \right\|_2^2 - 2\left\| (g\left(f_\theta\left(\Omega_y\right)\right) - y)^T (f_\theta(y) - y) \right\|_1
\end{aligned} \tag{3}$$

To avoid that the objective training function learns a single identity mapping relationship only, $f_\theta\left(y\right)$ is expected not to participate in the back propagation, and thus, the denoising network $f_\theta\left(y\right)$ in Eq (2) is replaced with a model $\hat{f}_\theta\left(y\right)$ that has no gradient updating. Hence, we convert the loss function Eq. (2) into the one as below,

$$arg\min_{\theta} E_y \left\| g\left(f_\theta\left(\Omega_y\right)\right) - \hat{f}_\theta\left(y\right) \right\|_2^2 + (\lambda - 1)\left\| \hat{f}_\theta\left(y\right) - y \right\|_2^2 + 2\left\| (g\left(f_\theta\left(\Omega_y\right)\right) - y)^T (\hat{f}_\theta\left(y\right) - y) \right\|_1 \tag{4}$$

Furthermore, to ensure that $\hat{f}_\theta\left(y\right)$ participates in gradient updating implicitly, the objective function is expected to guarantee that the derivative of $f_\theta\left(\Omega_y\right)$ contains $\hat{f}_\theta\left(y\right)$ to fulfil the non-blind requirements. Meanwhile, the optimized objective function should also converge during the training process. Hence, the derivative of the first term in Eq. (4) could be obtained as $2[g(f_\theta(\Omega_y))]^T(g\left(f_\theta\left(\Omega_y\right)\right) - \hat{f}_\theta\left(y\right))$.

To simplify the subsequent discussion, we set,

$$\tau\left(y\right) = \left\| g\left(f_\theta\left(\Omega_y\right)\right) - \hat{f}_\theta\left(y\right) \right\|_2^2 + (\lambda - 1)\left\| \hat{f}_\theta\left(y\right) - y \right\|_2^2 + 2\left\| (g\left(f_\theta\left(\Omega_y\right)\right) - y)^T (\hat{f}_\theta\left(y\right) - y) \right\|_1 \tag{5}$$

Moreover, as the following equation usually holds,

$$\begin{aligned}
&\left\| \left| \hat{f}_\theta\left(y\right) - g\left(f_\theta\left(\Omega_y\right)\right) \right| - \sqrt{\lambda - 1}\left| \hat{f}_\theta\left(y\right) - y \right| \right\|_2^2 = \\
&\left\| g\left(f_\theta\left(\Omega_y\right)\right) - \hat{f}_\theta\left(y\right) \right\|_2^2 + (\lambda - 1)\left\| \hat{f}_\theta\left(y\right) - y \right\|_2^2 - 2\sqrt{\lambda - 1}\left\| (\hat{f}_\theta(y) - g(f_\theta(\Omega_y)))^T (\hat{f}_\theta(y) - y) \right\|_1
\end{aligned} \tag{6}$$

To achieve ideal denoising effect, for $\lambda > 1$, we have

$$\left\| \left| \hat{f}_\theta\left(y\right) - g\left(f_\theta\left(\Omega_y\right)\right) \right| - \sqrt{\lambda - 1}\left| \hat{f}_\theta\left(y\right) - y \right| \right\|_2^2 = \tau(y) - 2\left\| A1 \right\|_1 - 2\sqrt{\lambda - 1}\left\| A2 \right\|_1 \tag{7}$$

where $A1 = (g\left(f_\theta\left(\Omega_y\right)\right) - y)^T(\hat{f}_\theta\left(y\right) - y)$, and $A2 = (\hat{f}_\theta(y) - g(f_\theta(\Omega_y)))^T(\hat{f}_\theta(y) - y)$. Since $\|\cdot\|_1 \geqslant 0$, we can further obtain,

$$\begin{aligned}
\tau\left(y\right) &\geq \left\| \left| \hat{f}_\theta\left(y\right) - g\left(f_\theta\left(\Omega_y\right)\right) \right| - \sqrt{\lambda - 1}\left| \hat{f}_\theta\left(y\right) - y \right| \right\|_2^2 \\
&\geq \left\| \hat{f}_\theta\left(y\right) - g\left(f_\theta\left(\Omega_y\right)\right) - \sqrt{\lambda - 1}(\hat{f}_\theta\left(y\right) - y) \right\|_2^2
\end{aligned} \tag{8}$$

Biomedical Optics EXPRESS

Hence, when $\tau(y)$ is the minimum, the denoiser converges to f_θ^*, while the optimal solution $\tilde{x}$ of $arg\ \min_\theta \left\| \hat{f}_\theta(y) - g\left(f_\theta\left(\Omega_y\right)\right) - \sqrt{\lambda - 1}\left(\hat{f}_\theta(y) - y\right) \right\|_2^2$ can be obtained,

$$\tilde{x} = \hat{f}_\theta^*(y) - \frac{\hat{f}_\theta^*(y) - g\left(f_\theta^*\left(\Omega_y\right)\right)}{\sqrt{\lambda - 1}} \tag{9}$$

The above results indicate that, once given a noisy image y, both $\lim_{\lambda \to 2} \tilde{x} = g\left(f_\theta^*\left(\Omega_y\right)\right)$ and $\lim_{\lambda \to \infty} \tilde{x} = \hat{f}_\theta^*(y)$ hold for Eq (9), and thus, $g\left(f_\theta^*\left(\Omega_y\right)\right) \le \tilde{x} \le \hat{f}_\theta^*(y)$. Furthermore, as the denoised image x is generated from the noisy images y only, the above limits still hold for Eq. (9) , then we can set λ to be $\lambda \to \infty$, and finally, the optimal $\tilde{x}$ converges to its upper limit.

In this study, we set B2Unet re-visible loss function as follows,

$$\begin{aligned} arg \min_\theta E_y \left\| g\left(f_\theta\left(\Omega_y\right)\right) - \hat{f}_\theta(y) \right\|_2^2 + (\lambda - 1)\left\| \hat{f}_\theta(y) - y \right\|_2^2 + \\ 2\left\| (g\left(f_\theta\left(\Omega_y\right)\right) - y)^T (\hat{f}_\theta(y) - y) \right\|_1 + \triangle \end{aligned} \tag{10}$$

where $\triangle$ is used for regularization and it could be set as below,

$$\triangle = 2\left\| g\left(f_\theta\left(\Omega_y\right)\right) - y \right\|_1 \cdot \left\| \hat{f}_\theta(y) - y \right\|_1 - 2\left\| (g\left(f_\theta\left(\Omega_y\right)\right) - y)^T (\hat{f}_\theta(y) - y) \right\|_1 \tag{11}$$

In this way, the blind spots can be visible to B2Unet with the re-visible loss function Eq. (1).

4. Experiments

To verify the effectiveness of B2Unet scheme, experiments are carried out to compare it with those state-of-the-art existing methods, such as two step iteration (TSI) [9], NWSR [8], SRCNN [9], DRGAN [21] and MAP-SNR [24]. For fair comparisons, all existing methods are implemented the same way as they were reported, and all their parameters are tuned to achieve their respective best performances in the study.

4.1. Datasets

The public OCT retinal image dataset provided in [28,29] is adopted for experiments. Such a dataset is collected by a Bioptigen SD-OCT (Durham, NC, USA) with an axial resolution of 4.5μm per pixel [30], and it contains four subsets, of which the first three were collected from the human eyes, while the last one was acquired from mouse. We denote them to be D1, D2, D3 and D4, respectively. The D1 and D3 contain noisy-clean image pairs, while D2 and D4 contain noisy images only. Since all four subset images are collected with the same OCT, we reasonably assume that the speckle distribution patterns are the same.

For fair comparisons, both DnCNN and SRCNN are trained with the noisy-clean image pairs from D1, i.e., three noisy-clean image pairs are randomly chosen from D1 first, and then each image pair is cropped into 70 patches, with a size of 256×256 pixels, in a stride of 50 to generate a new training dataset, hence, totally 210 new noisy-clean image pairs are obtained. While for B2Unet, three noisy images are randomly chosen from D1 first, and then they are cropped in the same way as the 210 noisy patches above obtained for training.

4.2. Parameter setting

In this study, B2Unet adopts an U-net architecture that is optimized by an adaptive momentum estimation (Adam) with a learning rate of 1e-4 and a training epoch number of 20 [31,32]. While for the hyper-parameter λ in the re-visible loss, it is initially set to be 3, and later is updated with an increment of 0.1 per epoch in first ten rounds. Finally, it is fixed to be 4 till the end of

Biomedical Optics EXPRESS

training. All the denoising schemes are implemented in Python with PyTorch framework, and all experiments are conducted on a workstation (Intel Xeon W-2145 CPU @3.70GHz) accelerated by an NVIDIA GeForce RTX 3060Ti GPU with 8G memory.

4.3. Performance metrics

Different metrics are employed for performance evaluations. Since no clean images are available in D2 and D4, quantitative evaluations, such as peak signal-to-noise ratio (PSNR) and structural similarity (SSIM) [33], are utilized for evaluating in this study.

Signal-to-Noise Ratio (SNR): SNR is a typical global performance metric that is defined to be the ratio of the signal mean to the background standard deviation, i.e.,

$$SNR = 20log(I_{max}/\sigma_B) \tag{12}$$

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

Equivalent Number of Looks (ENL): ENL is a metric typically utilized to measure the smoothness of the homogeneous image regions, and it measures the image background region in this study, and it is defined as below,

$$ENL = \mu_B^2/\sigma_B^2 \tag{13}$$

where μ_B and σ_B denote the mean and standard deviation of an image background region B, respectively. The larger ENL is, the smoother the corresponding region.

Variance (VAR): VAR is sensitive to noise, and it is defined as below in this study,

$$VAR = \sum_j \sum_i |I(i,j) - \mu|^2 \tag{14}$$

where $I(i,j)$ denotes the pixel intensity at the i-th row and j-th column of an image I, and μ is the mean of the image I. For each OCT image, since the background region occupies a large portion, the pixel intensity in despeckled image background should be similar with each other. Hence, the smaller VAR is, the better the denoising effect would be.

Contrast to noise ratio (CNR): CNR is the ratio of image contrast to noise, measuring the ratio of peak signal intensity to that of the background, and it is defined as follows,

$$CNR = \frac{1}{n}\sum_{i=1}^{n} 10\log\left(\frac{|\mu_i - \mu_B|}{\sqrt{\sigma_i^2 + \sigma_B^2}}\right) \tag{15}$$

where μ_i and σ_i^2 denote the mean and variance of a selected region i, while μ_B and σ_B^2 are the mean and variance of a background region B.

4.4. Results with retina image dataset

B2Unet is trained with the 210 noisy image patches that are cropped from three random noisy images in D1 first, and then, it is employed to process the other subset images. The remaining seven noisy images in D1 with a size of 950×500 are processed by B2Unet for evaluation.

Fig. 5 presents an image from D1 that is processed with different despeckling methods. As shown, satisfactory results are obtained with different methods. Specifically, among those model-based methods, TSI achieves the least significant effect, while NWSR performs better. As illustrated by Fig. 5(d), extensive speckle residues still exist in the denoised image, and the artifacts also contaminates the image structural details. While in Figs. 5(c), the speckles are

Biomedical Optics EXPRESS

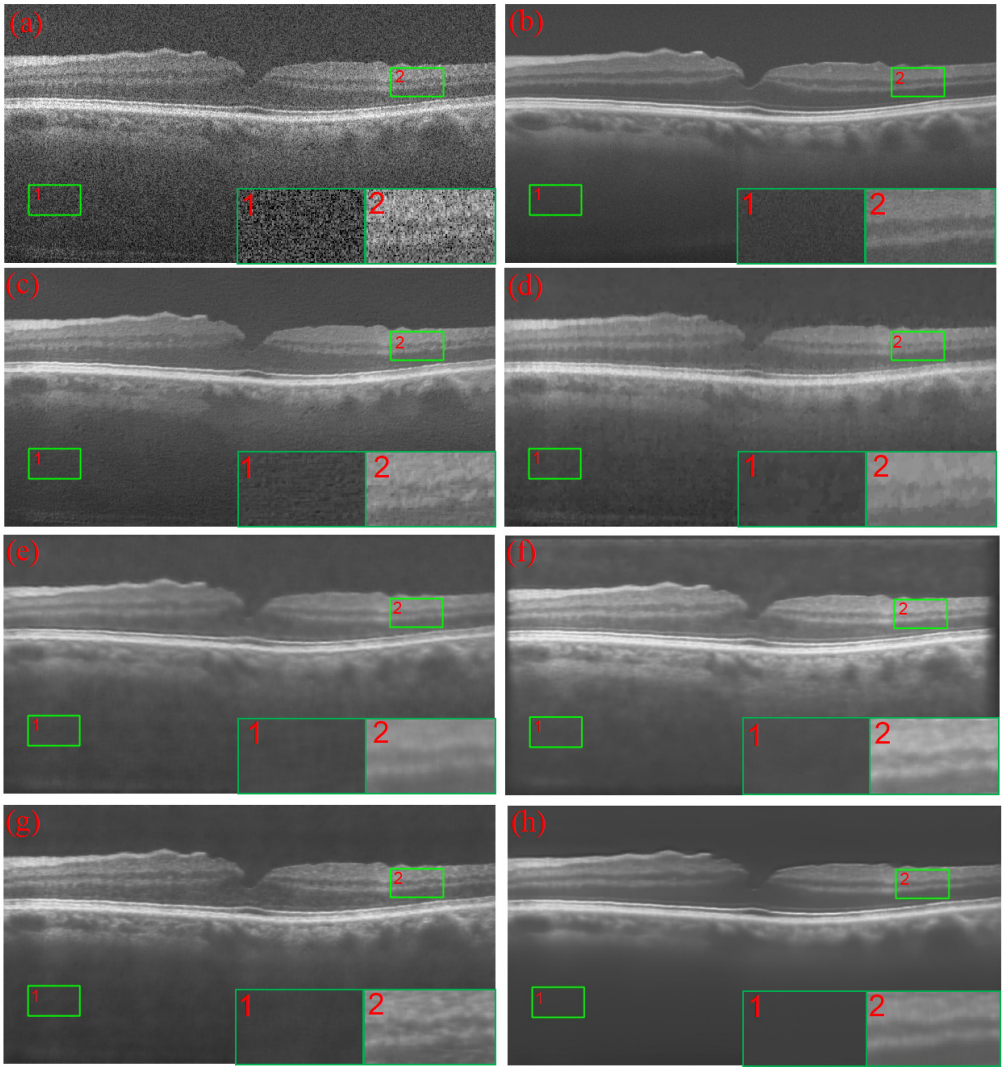

Fig. 5. An OCT image from subset D1 used for experiment. (a) The original image from D1 for processing. (b) The corresponding real clean image. The image despeckled by (c) NWSR, (d) TSI, (e) SRCNN, (f)DRGAN, (g)MAP-SNR, (h)B2Unet.

largely suppressed, making the overall image smooth. However, as the speckle distribution variance must be known as a priori for NWSR, the contrasts of structural details appear to be reduced for NWSR. Similarly, the structural boundaries in Fig. 5(d) by TSI are also affected by artifacts, wherein the image is excessively smoothed.

Figure 5(e) presents the image processed by SRCNN. Results show that, as a fully supervised deep-learning method, SRCNN suppresses speckles effectively, with which influences speckles are largely reduced, while the tissue structural details are well preserved. The image looks smooth with the tissue details being clearly resolved, and it is even better than those processed by the model-based methods. In clinical practice, however, since noisy-clean image pairs are difficult to obtain prior, the applications of fully supervised deep-learning methods could be hindered.

Biomedical Optics EXPRESS

Figures 5(f) and 5(g) show two images that are processed by the un-supervised and self-supervised methods, i.e., DRGAN and MAP-SNR, respectively. Both images demonstrate that the overall denoising effects of DRGAN and MAP-SNR are excellent since the speckles are largely suppressed and tissue microstructures are well preserved, and they are even comparable with those fully supervised methods. Due to the lack of sufficient clean images for training, however, there still exist some over-denoised areas in the background region for DRGAN in Fig. 5(f), and some noise residuals in Fig. 5(g) denoised by MAP-SNR.

To alleviate the difficulties of acquiring clean images, the proposed B2Unet is designed to learn from a single noisy image for despeckling. Figure 5(h) shows the image denoised with B2Unet. As seen, B2UNet largely suppresses speckles with the influence of speckles being nearly negligible. Specifically, as shown in the left inset in Fig. 5(h), speckles in the image background regions are nearly eliminated, and denoising results are comparable to those of the clean image shown in Fig. 5(b). It is also worth mentioning that, after being despeckled, the image details are well preserved, while the tissue structural details are still comparable to those of the clean one, e.g., both the image resolution and the structural details illustrated in the right inset in Fig. 5(h) are comparable to those in Fig. 5(b). Such comparisons indicate that B2Unet is effective in despeckling and its obtained results are comparable with those existing methods.

To further verify the effectiveness of B2Unet, different metrics, e.g., SNR, ENL, CNR and VAR, are also calculated for images in D1 for quantitative assessment. Specifically, since those images have their corresponding clean counterparts, both PSNR and SSIM are also calculated. Table 1 presents a quantitative comparison between those methods. Results in Table 1 show that B2Unet performs the best among all those methods with different metrics. For example, its SSIM, SNR, and ENL outperform those of the self-supervised MAP-SNR scheme by 1.4%, 27.8% and 862.8%, respectively. In addition, its SSIM, SNR and ENL outperform the fully-supervised method SRCNN by 2.8%, 24.4% and 595.7%, respectively. Such results convincingly demonstrate that B2Unet is effective for despeckling, especially for removing the background noises, and its performances are comparable to those of the state-of-the-art existing methods.The SSIM value obtained by this method ranks first among all methods, indicating that the image structural details of our method are also well preserved in the despeckling process.

Table 1. Quantitative assessments with D1 and D2 for different methods.

	Subject	Number	Size	Method	PSNR	SSIM	SNR	ENL	CNR	VAR(10e8)
Data1	10	10*4	950×500	Noisy image	17.986	0.0838	18.1153	3.7369	0	8.0300
				NWSR	27.2954	0.5730	31.4536	111.8567	3.3357	3.5231
				TSI	27.835	0.6693	33.5246	235.1185	**3.9109**	3.3464
				SRCNN	27.3115	0.7127	38.8163	890.2302	3.2697	2.5368
				DRGAN	19.5065	0.6591	37.6299	638.7958	3.7403	5.5284
				MAP-SNR	**28.1088**	0.7222	39.3228	643.2975	3.1690	3.0680
				B2Unet	27.0875	0.7325	**48.3045**	**6193.5509**	2.8456	**2.3276**
				clean	\	\	32.4701	142.1576	3.2902	3.5368
Data2	39	39*4	450×450	Noisy image	\	\	17.8404	3.9811	-0.9947	3.6453
				NWSR	\	\	32.6035	146.0455	1.3459	1.616
				TSI	\	\	40.0417	1343.8894	**2.2217**	1.3239
				SRCNN	\	\	40.8091	1377.9681	1.3485	**1.2858**
				DRGAN	\	\	34.4314	271.7027	1.4646	1.8590
				MAP-SNR	\	\	41.8373	1153.4238	1.3708	1.5510
				B2Unet	\	\	**48.4294**	**5417.3030**	0.9268	1.3548

Biomedical Optics EXPRESS

Since all images in Datasets D1 and D2 are acquired with the same OCT device, we reasonable assume that speckle distribution pattern in the images of D2 is the same as that in D1, and thus, B2Unet trained by D1 is adopted directly for image despeckling for D2. Figure 6 presents an image from D2 that is processed with different despeckling methods. As seen, the denoising performances vary among different methods, and the overall performances of those model-based methods are less significant as compared with those of deep learning ones. From Figs. 6(a)–6(h), it could be observed that, among those model-based methods, NWSR performs much better than TSI with the denoised image shown in Fig. 6(c) are much smoother than Fig. 6(d), since blurring effects caused by excessive smoothing appear in the image processed by TSI, as shown in Fig. 6(d).

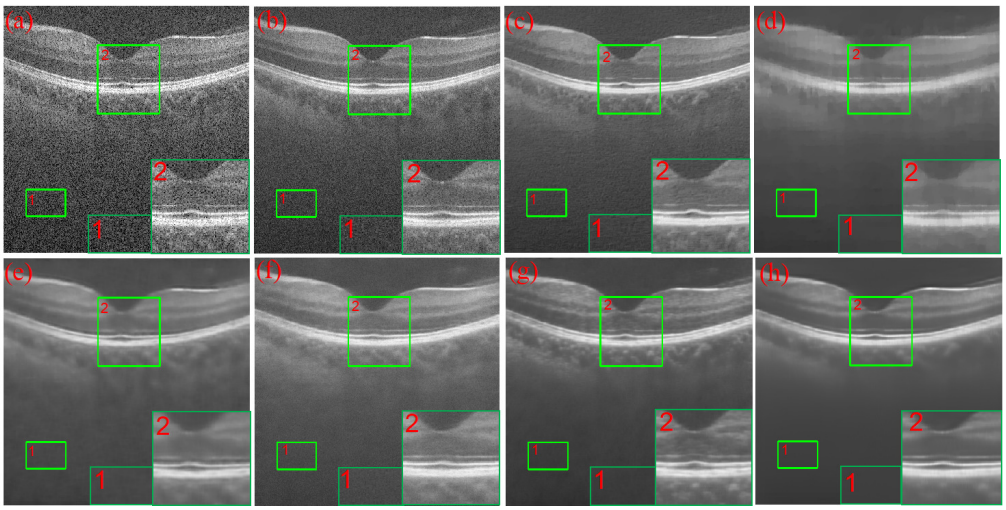

Fig. 6. An OCT image from D2 for verification. (a) Original noisy image, (b) Image averaged over four consecutive frames, and image processed by (c)NWSR, (d) TSI, (e) SRCNN, (f) DRGAN, (g) MAP-SNR, (h) B2Unet.

Results in Fig. 6(e) show that the denoising effect could be largely improved when employing the fully supervised deep learning schemes. As shown in Figs. 6(e), by utilizing SRCNN, the image speckles are largely reduced while the structural details, especially the layered boundaries, are well preserved. Figures 6(f) and 6(g) show the images that are denoised by the un-supervised DRGAN and the self-supervised MAP-SNR, respectively. Results in Fig. 6(f) show that DRGAN help preserve the detailed layered structures, yet there still exist small speckles in the background area, while Fig. 6(g) illustrates that MAP-SNR largely suppresses the speckle noise in the background areas with the structural details being well preserved, yet there exist some speckle residuals in the complex structural area. While for B2Unet, one can also observe that its denoising effects are much better than those of NWSR, TSI, while are comparable to those fully supervised or self-supervised learning methods, i.e., DnCNN and MAP-SNR. Specifically, as shown in the larger inset of each figure, the layered structural details are well preserved in the image by B2Unet, and they are comparable to or even better than those in the images processed by the fully supervised method, e.g., Fig. 6(e) by SRCNN. While for the background regions, as shown by the smaller inset in each image, the one processed by B2Unet is smooth, while the speckles in background regions are nearly eliminated. Such results convincingly demonstrate that B2Unet is able to alleviate the influence of speckles using a single noisy image only.

Table 1 also presents a quantitative comparison between B2Unet and the existing denoising methods with dataset D2. As we can see, B2Unet achieves the highest SNR and ENL among

Biomedical Optics EXPRESS

all those methods, which is 15.7% and 369.7% higher than that of MAP-SNR. It is also worth mentioning that both CNR and VAR of B2Unet are satisfactory and comparable to those of the existing methods. Such results conclude that the denoising ability of B2Unet with D1 is also fully applicable to D2.

Experiments are also carried out with datasets D3 and D4 to verify the effectiveness of B2Unet. Again, B2Unet trained by D1 is employed for processing the images from D3 and D4, wherein D3 contains a ground truth image for reference while D4 does not. Specifically, for D3, all 18 human retinal images are adopted for testing, with one being randomly selected for comparisons. While for D4, four noisy images collected at the same position are present, and the average of four consecutive frames is utilized as a reference clean image for comparisons.

Figure 7 presents an image from D3 that is processed by different methods. As seen, the performances vary among different methods. Specifically, among those methods, NWSR performs better than TSI. As shown in Figs. 7(c) and 7(d), such obtained images are even comparable to ground truth clean image shown in Fig. 7(b). While when deep learning schemes are employed, satisfactory results can also be achieved. As seen, Figs. 7(e) and 7(h) that are processed by SRCNN and B2Unet are comparable to the ground truth clean image in Fig. 7(b), and the one processed by B2Unet is also comparable to those by SRCNN. In Figs. 7(e) and 7(h), it could also be noticed that the structural details, as illustrated by the larger inset, are well preserved, while the speckles in the image background, as shown in the smaller inset, are almost eliminated. Figures 7(f) and 7(g) show the images denoised by DRGAN and MAP-SNR, respectively. The observations are similar to those for D1. The image denoised by DRGAN preserves the important image structures, yet the background areas still contain some speckles. In contrast, the image denoised by MAP-SNR achieves satisfactory denoising effects for both image structures and background areas, and they are comparable to the ones by B2Unet shown in Fig. 7(h). Results in Fig. 7(h) demonstrate that B2Unet achieves satisfactory speckle suppression in OCT images.

Figure 8 presents an image of D4 processed by different methods. As shown, TSI achieves the best performances, followed by MAP-SNR and B2Unet as shown in Figs. 8(g) and 8(h). It is worth noting that SRCNN achieves satisfactory denoising results in Figs. 8(e), yet the structural details are not well preserved due to the blurring effects caused by excessive smoothing. In contrast, DRGAN, MAP-SNR and B2Unet achieves a trade-off between the denoising effects and the image detail preservations. As shown in the insets of Figs. 8(f), 8(g), and 8(h), although speckle suppressions are not as good as those in Fig. 8(e), the layered structural boundaries could be clearly resolved. Such results again convincingly prove the effectiveness of B2Unet.

Table 2 presents the quantitative comparison between B2Unet and the other existing methods for dataset D3 and D4. Specifically, B2Unet achieves the best SSIM, SNR and ENL, and it is 1.4%, 13.1% and 263.2% higher than those of MAP-SNR for dataset D3. For both D3 and D4, it could be noticed that B2Unet largely outperforms both the fully supervised and the unsupervised deep learning methods for all metrics, demonstrating that B2Unet is robust and effective for speckle reductions in OCT image. Furthermore, since no clean image is required by B2Unet prior, it is expected that B2Unet could be widely adopted as a feasible tool for OCT denoising, and thus, would be of great potential in clinical applications.

4.5. Results with other datasets

To further verify the robustness of B2Unet, some other dataset, i.e., swine eye images [34], and those acquired by our lab-customize OCT systems *in vivo*, e.g., thumbnail images [35], skin images [35], airway epithelia image [36] and diseased retina image [37], are also employed in experiments. B2Unet trained by D1 again is utilized for despeckling. Since those OCT devices are different, B2Unet is also trained with the corresponding *in vivo*images for comparisons. The model training parameters are kept the same as those used with D1.

Biomedical Optics EXPRESS

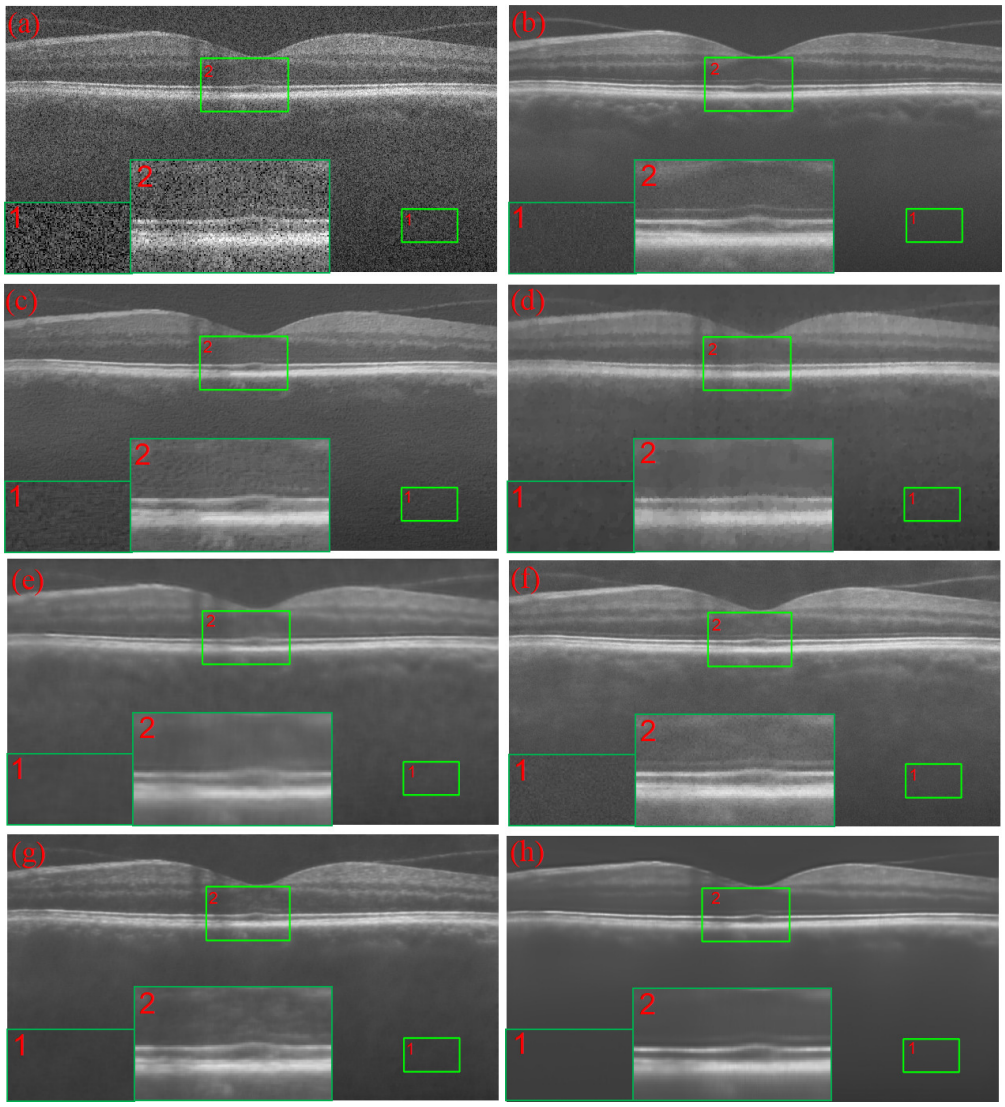

Fig. 7. A random image from D3 used for experiment. (a) The original image, (b) ground truth clean image. The image processed by (c) NWSR, (d) TSI, (e) SRCNN, (f) DRGAN, (g) MAP-SNR, (h)B2Unet.

Figures 9 presents those images denoised with B2Unet that are trained with different datasets. Specifically, Figs. 9(a1-a5) present the original images that are acquired by different OCT systems, while Figs. 9(b1-b5) and Figs. 9(c1-c5) show the images processed by B2Unet that are trained by D1 [28,29] and *in vivo*images, respectively. As shown in Figs. 9(a1-a5), extensive speckles exist in the original noisy images, which largely degrades image quality, and thus, hides the image details. While after been denoised by B2Unet, however, the images' qualities are largely improved. As we can see in Figs. 9(b1-b5) and 9(c1-c5), all images become smooth with image details been well preserved, and the image speckles are largely removed. Specifically, for the image background as shown in the green rectangles, the speckles are largely suppressed or even removed.

Biomedical Optics EXPRESS

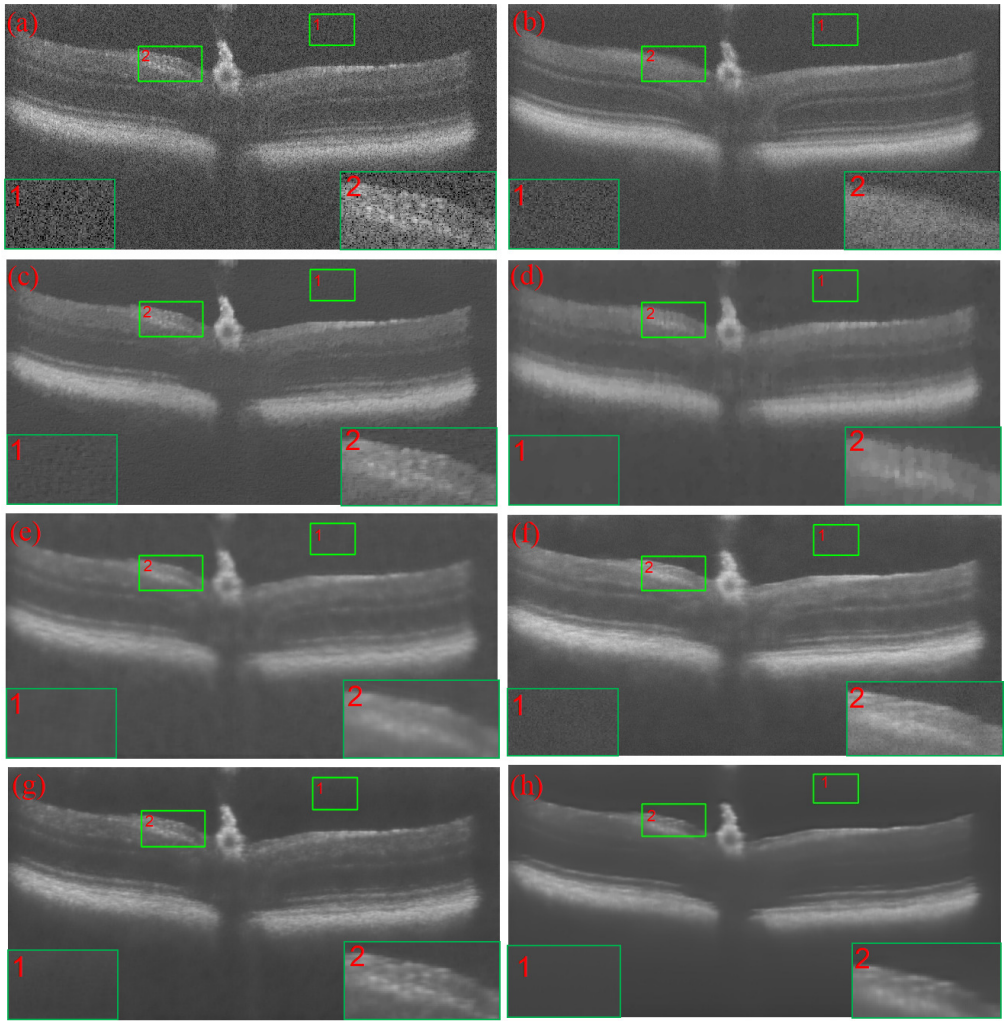

Fig. 8. A random image from D4 used for experiment. (a) The original image, (b) reference clean image averaged over four consecutive frames. The image processed by (c) NWSR, (d) TSI, (e) SRCNN, (f) DRGAN, (g) MAP-SNR, (h)B2Unet.

When further comparing Figs. 9(b1-b5) with Figs. 9(c1-c5), one can also observe that, visual effects of those images denoised by B2Unet that trained by D1 are much better than those processed by B2Unet that trained with *in vivo* images. Performance metrics are also calculated in Table 3. Results show that metrics of the images processed by B2Unet trained with D1 are much better, yet the ones trained with *in vivo*images are not satisfactory. The main reason is probably because of the speckle distributions of those images. As shown, speckles in those *in vivo*images are large and their distributions are dense over the whole images, which largely hides the tissue microstructures. Therefore, B2Unet training is not satisfactory if such *in vivo*images are used, and the despeckling effects are limited. In contrast, when B2Unet is trained by D1, both image visual effects and quantitative metrics are better. This is because speckles in those retinal images are relatively uniform, and the distributions are relative less significant in the image background. In such a case, the image speckles could be largely suppressed once B2Unet

Biomedical Optics EXPRESS

Table 2. Quantitative results of data sets D3 and D4 with different methods.

	Subject	Number	Size	Method	PSNR	SSIM	SNR	ENL	CNR	VAR(10e8)
Data3	18	18*4	900×450	Noisy image	17.7850	0.0904	17.9173	3.7261	-0.8984	7.8228
				NWSR	26.9449	0.5560	31.7876	108.9344	1.5913	3.8106
				TSI	**27.5151**	0.6552	34.7992	293.4748	**2.0721**	3.5564
				SRCNN	26.8732	0.6739	38.6413	868.5034	1.6280	**2.8740**
				DRGAN	23.7349	0.5781	33.6506	217.3121	1.5944	4.2609
				MAP-SNR	27.2381	0.6845	39.9935	752.9737	1.5034	3.6045
				B2Unet	26.8162	**0.6944**	**45.2387**	**2735.1920**	0.9640	3.1687
				clean	\	\	32.4320	119.7168	1.9751	3.7819
Data4	1	1*4	1000×450	Noisy image	\	\	19.4950	7.3610	-1.5774	5.9387
				NWSR	\	\	34.8093	396.7630	0.9676	2.4219
				TSI	\	\	**42.7879**	**3277.5277**	**1.7007**	2.2457
				SRCNN	\	\	40.9488	2162.296	1.5263	**1.963**
				DRGAN	\	\	34.4897	315.6014	1.2869	3.7258
				MAP-SNR	\	\	35.2933	395.6756	1.1288	2.5343
				B2Unet	\	\	42.0168	2080.3318	0.2780	2.4118

is adopted for despeckling. Such results again demonstrate that B2Unet is robust and effective for OCT speckle reductions.

Table 3. Performances of B2Unet trained with different image datasets.

Data type	B2Unet trained with	SNR	CNR	ENL	VAR(10e8)
Swine	Noisy image dataset	27.2339	2.7919	19.8199	5.6871
	Retinal image dataset	**40.2378**	**3.2017**	**773.9742**	**2.8797**
	The corresponding image dataset	22.8309	0.7868	7.1429	10.3808
Skin	Noisy image dataset	23.914	2.6375	4.6249	8.2738
	Retinal image dataset	**39.0532**	2.5532	**444.6741**	**4.5256**
	The corresponding image dataset	33.6214	**3.2207**	53.239	6.1829
Thumbnail	Noisy image dataset	24.4208	**4.029**	2.8338	29.1556
	Retinal image dataset	**43.2567**	3.5132	**882.0393**	**10.7958**
	The corresponding image dataset	25.1632	3.2495	2.607	27.0152
Airway epithelia	Noisy image dataset	57.075	**2.6073**	0	14.2979
	Retinal image dataset	57.129	2.344	**17429.1881**	**7.7463**
	The corresponding image dataset	**73.0101**	1.6488	0	15.3711
Diseased retina	Noisy image dataset	27.877	1.3077	3.4388	3.4031
	Retinal image dataset	**45.4441**	0.5904	**1349.2797**	**1.2003**
	The corresponding image dataset	31.8686	**1.4156**	6.2667	3.0201

4.6. Influences of speckle distributions

When employing supervised learning schemes for OCT despeckling, the images utilized for both training and testing are typically from the same datasets acquired by the same OCT, which thus may have similar noise distributions. While for self-supervised B2Unet, although its despeckling is influenced by the images employed for training, it is robust and effective in processing those acquired by other OCTs *in vivo*, as long as those images used for training have similar speckle

Biomedical Optics EXPRESS

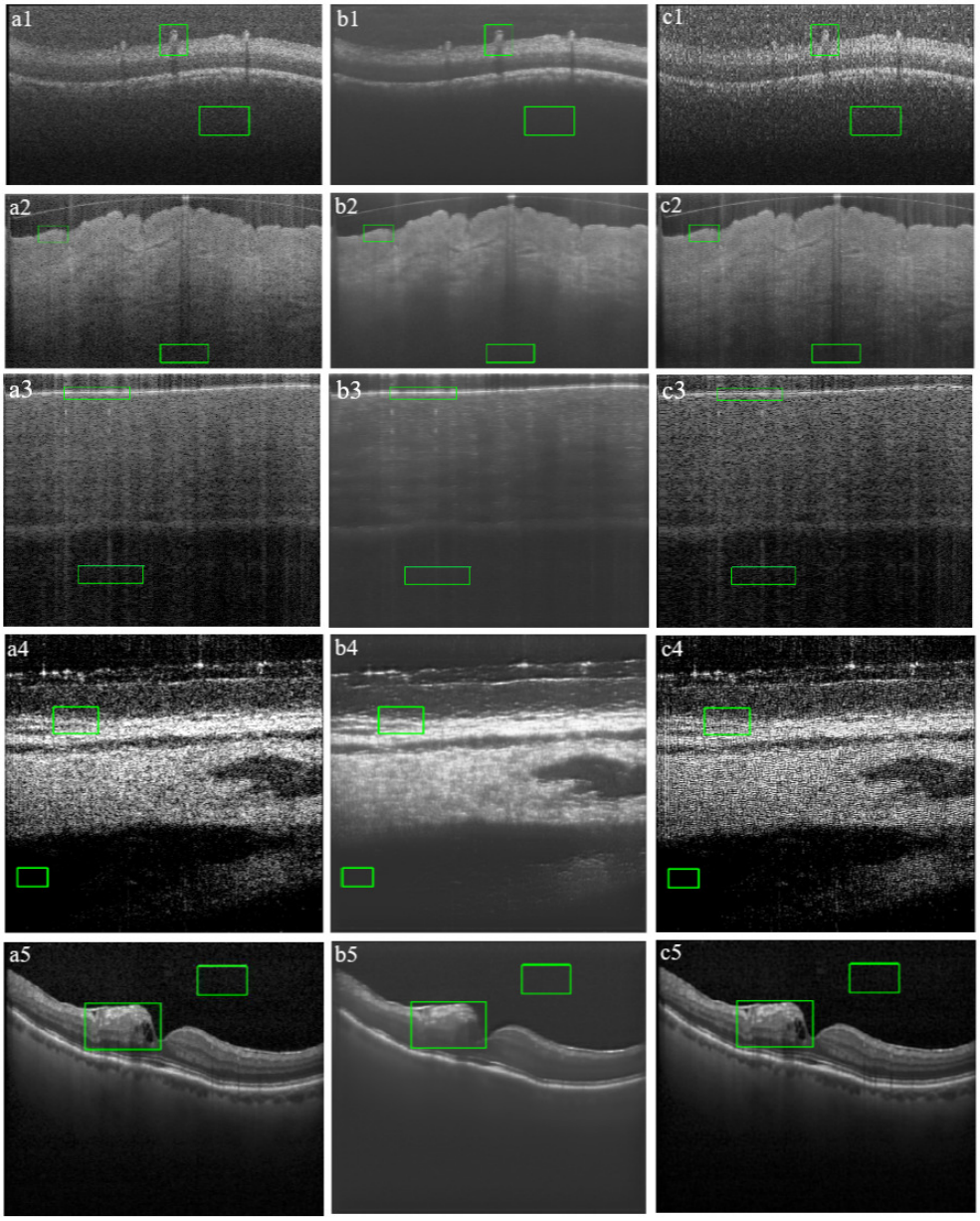

Fig. 9. B2Unet for *in vivo*OCT image despeckling. (a1-a5) The original image acquired by different OCT systems *in vivo*. (b1-b5) The images despeckled by B2Unet trained by D1. (c1-c5) The noisy image despeckled by B2Unet trained with the corresponding *in vivo*images. a1: swine eye image, a2: skin image; a3: thumbnail image; a4: airway epithelia image; a5: diseased retinal image. The green rectangles denote the background and structure areas utilized for quantitative analysis.

distributions [28,29]. To further evaluate the influences of speckle distributions, those *in vivo* images are further up- and down-sampled first, and then are utilized for B2Unet training, and finally, B2Unet is employed for OCT speckle reductions.

Biomedical Optics EXPRESS

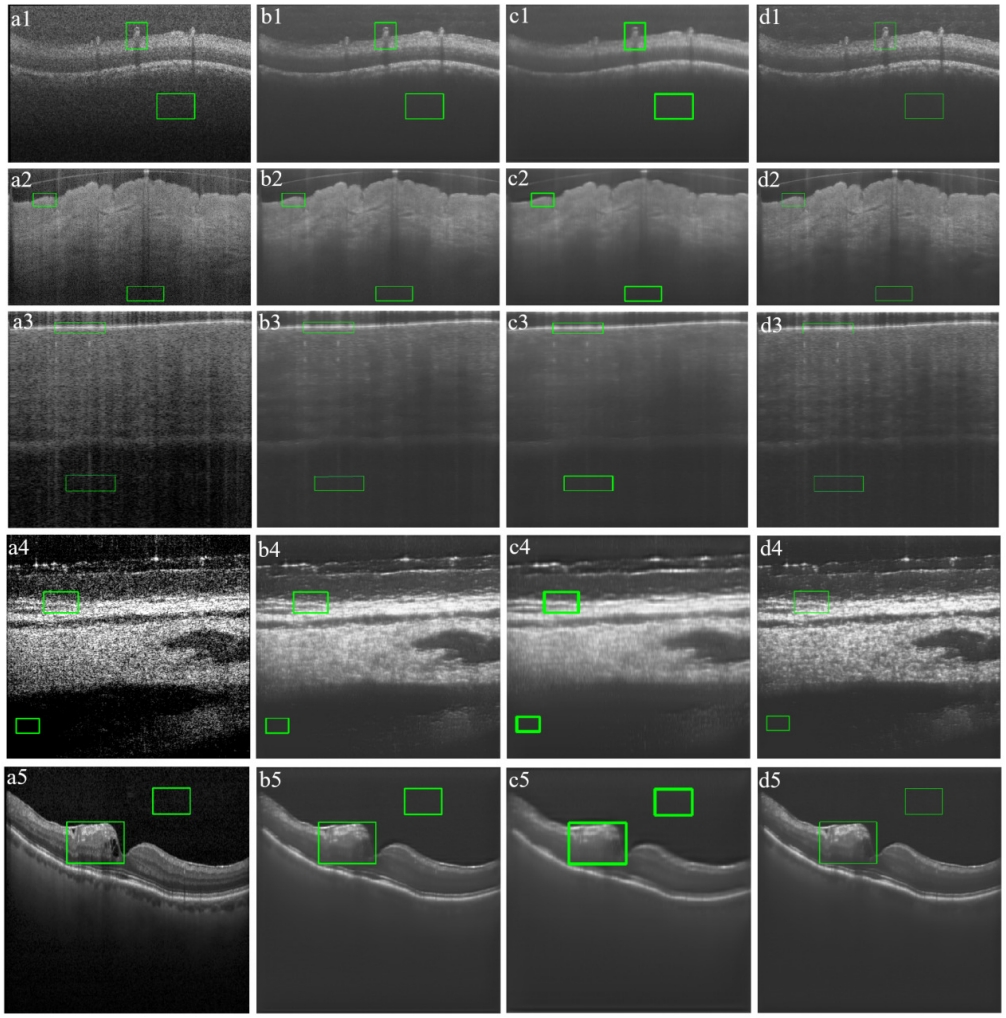

Fig. 10. B2Unet is employed for despeckling *in vivo* images. Column (a1-a5) original noisy images, and columns (b-d) images despeckled by B2Unet trained by (b1-b5) D1, (c1-c5) down-sampled, and (d1-d5) up-sampled images. Rows 1-5 denote the swine eye, skin, thumbnail, airway epithelia, and diseased retinal images. The green rectangle represents the background and the structure areas for metrics calculation.

Figure 10 presents the *in vivo* images that are processed with B2Unet trained with retinal images from D1, i.e., the original, down-sampled, and up-sampled ones. Results in Fig. 10 demonstrate that the best visual effects are achieved for those down-sampled images. Specifically, as is shown, speckles in those down-sampled images, i.e., Figs. 10(c1-c6), are largely suppressed, and thus, the obtained images are smooth, while the structural details are well preserved. On the contrary, visual effects of those up-sampled images are less significant as speckle residuals still exist. The reason for such results is that down-sampling operation reduces the overall image size, while the image speckles are concentrated, making them easier to suppress.

Performance metrics are also calculated for quantitative evaluation. Table 4 presents a comparison between those images that are processed with B2Unet that are trained with different image datasets. Results show that, those down-sampled images achieve the best CNR and the

Biomedical Optics EXPRESS

lowest ENL and VAR among all those obtained images, while SNR of the down- and up-sampled images are close to that of the original image. Such visual effects and quantitative results again demonstrate that B2Unet is robust and effective for speckle reductions in OCT images with dense noise distributions. For the dense speckle noise distribution, it can better reflect the characteristics of noise removal visually, but there is no rule in the evaluation indicators. Whether it is denoising after up-sampling or down-sampling, each has its advantages and disadvantages. Although the denoising effect is obvious after down-sampling, it will lead to smaller image size and no effective tissue microstructure can be observed. The denoising after up-sampling will lead to larger image size and better observation of tissue microstructure, but it can not achieve effective denoising results. For image denoising, the best denoising effect is to use the model trained on the same noise distribution data to denoise. The denoising after up-sampling or down-sampling can only be applied to some special scenes.

Table 4. Performances of B2Unet for despeckling different noisy images.

Data type	Process	SNR	CNR	ENL	VAR(10e8)
Swine image(single)	noisy	14.5747	-2.0967	0.7788	17.467
	down-sampling	**33.1322**	**4.6824**	**166.1589**	**0.7448**
	normal	28.7817	4.0795	50.9005	4.8892
	up-sampling	30.2956	3.3906	61.2274	17.2702
Swine image(average)	noisy	27.2339	2.7919	19.8199	5.6187
	down-sampling	40.4153	**3.3979**	**1023.3177**	**0.6787**
	normal	40.2378	3.2017	773.9432	2.8797
	up-sampling	**42.3083**	2.8893	1021.14	11.2885
Skin	noisy	23.914	2.6375	4.6249	8.2738
	down-sampling	40.096	**2.6983**	713.5036	**1.0322**
	normal	39.0532	2.5532	444.6741	4.5256
	up-sampling	**43.3615**	2.4485	**1106.1574**	17.4727
Thumbnail	noisy	24.4208	**4.029**	2.8338	29.1556
	down-sampling	**46.2869**	3.56	**1730.3891**	**2.6581**
	normal	43.2567	3.5132	882.0393	10.7958
	up-sampling	42.3377	3.3817	692.2494	42.801
Airway epithelia	noisy	57.075	**2.6073**	0	14.2979
	down-sampling	48.6905	2.0471	2715.6615	**1.7387**
	normal	57.129	2.344	17429.1881	7.7463
	up-sampling	**65.6328**	2.0413	**124189.8947**	29.39128
Diseased retina	noisy	27.877	**1.3077**	3.4388	3.4031
	down-sampling	43.3646	0.9415	1071.7812	**0.2679**
	normal	45.4441	0.5904	1349.2797	1.2003
	up-sampling	**47.4353**	0.4807	**2500.2002**	4.8021

4.7. Influences of re-visible loss

Experiments are also carried out to verify the effectiveness of re-visible loss function by changing the variable λ . Specifically, experiments are conducted by increasing rate of λ at each step. First, we evaluate the influences λ by setting it as a constant. For experiments, we set λ to be 3, 4, 5, 6, 7, respectively, and then, employ the corresponding re-visible loss functions for B2Unet training, and calculate both SNR and ENL for images from D1. Results are plotted in Fig. 11(a). As we

Biomedical Optics EXPRESS

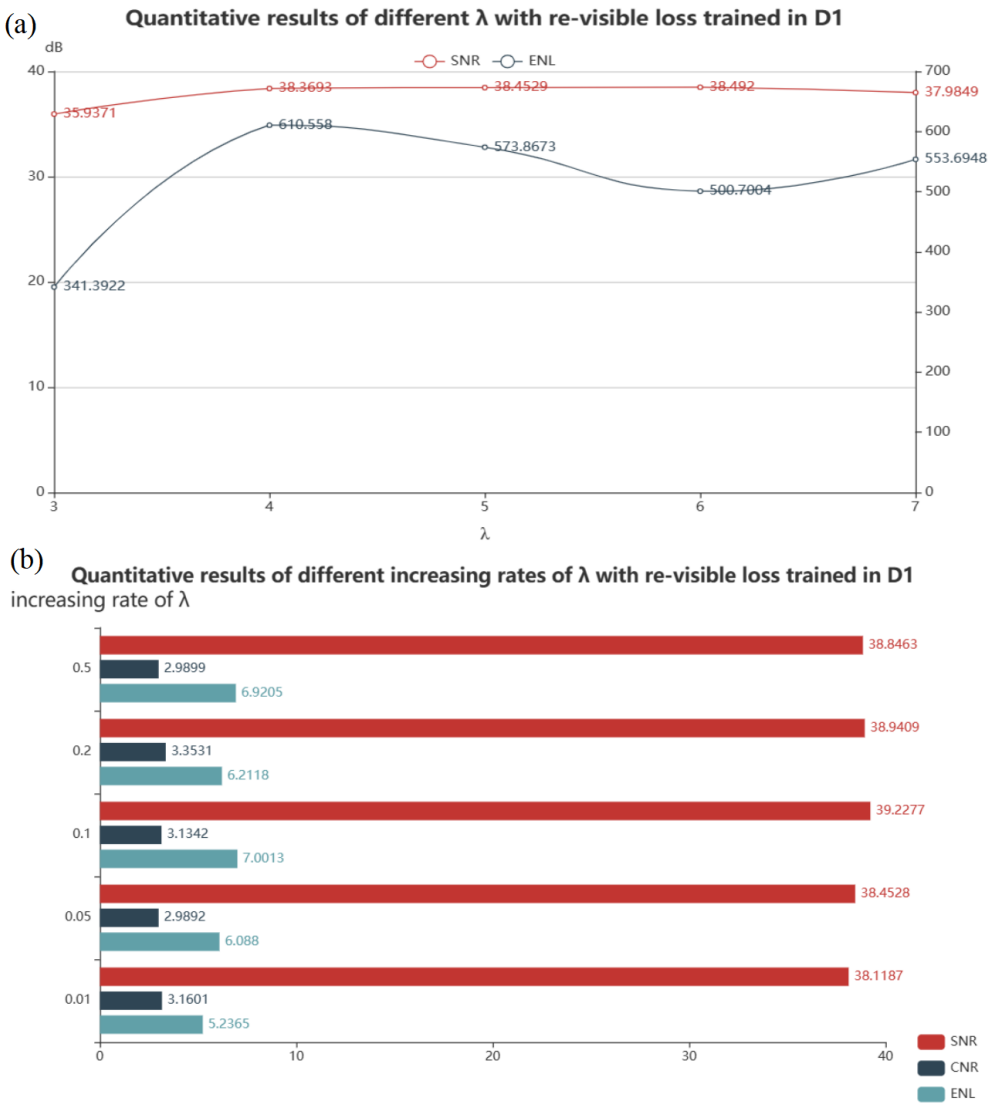

Fig. 11. Quantitative results of different λ with re-visible loss trained in D1. And quantitative results of different increasing rates of λ with re-visible loss trained in D1.

can see, with an increasing variable λ, both SNR and ENL increase significantly from 3 to 4, and then stay more or less unchanged when λ increases from 4 to 7. Such results show that increasing λ in the re-visible loss function would improve the despeckling quality of OCT images and prove the reliability of the theoretical derivation again.

Furthermore, we also evaluate the influences of the increasing rate of λ at each step in the training process. In the testing process, we set the initial value of λ to be 3, and then increase it by 0.01, 0.05, 0.1, 0.2, 0.5 respectively at each step, and finally, fix it to be constant after 10 steps. The metrics are calculated for the images processed by B2Unet that is trained with different λ. Results in Fig. 11(b) demonstrate that a small increase in λ during the training process can improve the denoising effects of B2Unet, yet the improvement is quite small with the metrics being quite similar to each other for different increasing rate of λ. In this study, we choose

Biomedical Optics EXPRESS

an increasing rate of 0.1 for λ to achieve a balance between the computational loads and the despeckling effects.

4.8. Influences of refinement strategy

By comparing the denoised images output from the denoising network and those from refinement unit in Fig. 1(b), experiments are also carried out to verify the effectiveness of refinement module in B2Unet inference unit with different image datasets.

Figure 12 tests the effectiveness of refinement module on image denoising for D1. Figures 12(a) and 12(b) are noisy image and its corresponding clean image, while Figs. 12(c) and 12(d) are the images denoised by B2Unet without and with the refinement strategy. Results in Figs. 12(c) and 12(d) demonstrate that, without the refinement strategy, there exists some speckle residues in the image background in Fig. 12(c), while when the refinement module is adopted, the image background is clean while the tissue structural details are well preserved. Therefore, it can be concluded that the refinement module helps improve noise reductions, especially for removing those noise residues.

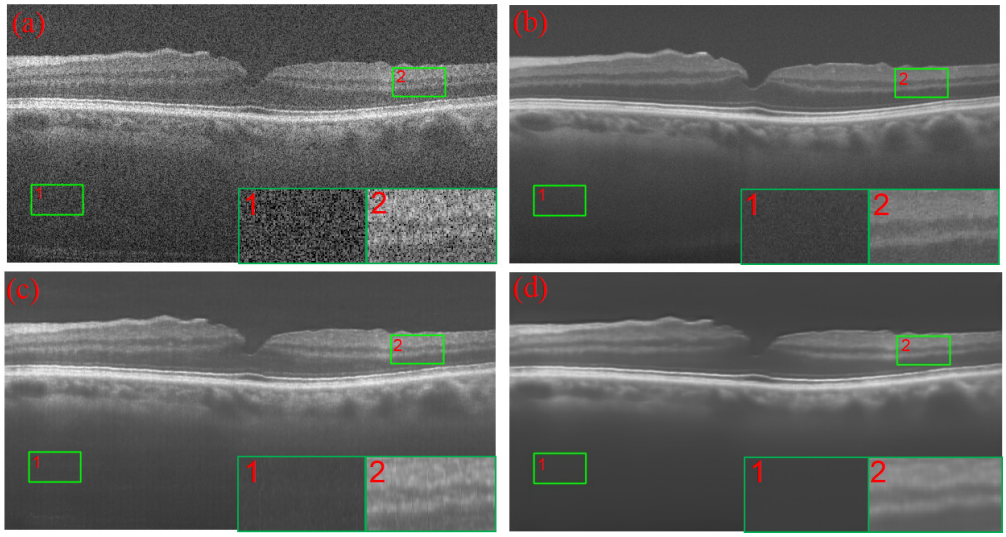

Fig. 12. A noise image from D1 used for experiment. (a) The original image, (b) clean image. the image processed by B2Unet (c) without refinement, (d) with refinement.

Table 5. Quantitative assessments with D1 and D2 for different methods.

	Refinement	PSNR	SSIM	SNR	ENL	CNR	VAR(10e8)
Data1	✗	27.4017	0.6872	39.2277	700.1321	3.1342	3.0924
	✓	27.0875	**0.7325**	**48.3045**	**6193.5509**	2.8456	**2.3276**
Data2	✗	\	\	41.2243	1107.3134	1.2172	1.6419
	✓	\	\	**48.4294**	**5417.3030**	0.9268	**1.3548**
Data3	✗	26.3662	0.6470	38.6376	161.1556	1.3473	3.7995
	✓	**26.8162**	**0.6944**	**45.2387**	**2735.1920**	0.9640	**3.1687**
Data4	✗	\	\	39.7118	1173.2750	1.1159	2.6008
	✓	\	\	**42.0168**	**2080.3318**	0.2780	**2.4118**

Biomedical Optics EXPRESS

Quantitative comparisons are also performed to compare the inference unit with and without refinement strategy for OCT despeckling. Results in Table 5 show that the refinement module helps improve all metrics except for CNR. For example, metrics SSIM, SNR, and ENL improved by 6.6%, 23.1% and 784.6%, respectively for D1. Similar results could also be obtained with other datasets, demonstrating the effectiveness of the refinement strategy for despeckling.

5. Conclusion

In summary, this paper studies speckle reductions in OCT images, and a novel self-supervised deep leaning method, namely the B2Unet, is proposed for the first-time for speckle reductions in OCT images, to the best of our knowledge. By utilizing a global-aware mask mapper to improve image perception, and a new re-visible loss function to facilitate network training, B2Unet is designed to employ a single noisy image for network training. Experiments with different image datasets are conducted to compare B2Unet with the state-of-the-art existing methods. Influence of OCT speckle distributions is also evaluated in different cases. Both qualitative and quantitative results show that B2Unet is effective and robust in suppressing image speckles while retaining the layered microstructural details and tissue morphologies in OCT images, and its performances are comparable to, or even better than those of the fully supervised deep learning methods in different cases when the image speckles are comparable to those used for B2Unet training. Owing to its effectiveness and robustness, B2Unet is of great significance to clinical applications and is expected to be a viable tool for OCT imaging-based diagnosis, especially for the cases when real clean images are not available or hard to acquire prior.

Funding. National Natural Science Foundation of China (62220106006); Basic and Applied Basic Research Foundation of Guangdong Province (2021B1515120013); Key Research and Development Projects of Shaanxi Province (2021SF-342); Key Research Project of Shaanxi Higher Education Teaching Reform (21BG005).

Acknowledgments. The authors would like to acknowledge the continuous support from Guangdong Key Laboratory of Integrated Optoelectronics and Intellisense.

Disclosures. The authors declare no conflicts of interest.

Data Availability. Data underlying the results presented in this paper are not publicly available at this time but may be obtained from the authors upon reasonable request.

References

1. D. Huang, E. A. Swanson, C. P. Lin, J. S. Schuman, W. G. Stinson, W. Chang, M. R. Hee, T. Flotte, K. Gregory, C. A. Puliafito, and J. Fujimoto, "Optical coherence tomography," Science **254**(5035), 1178–1181 (1991).
2. W. Drexler and J. G. Fujimoto, "State-of-the-art retinal optical coherence tomography," Prog. Retinal Eye Res. **27**(1), 45–88 (2008).
3. A. V. D'Amico, M. Weinstein, X. Li, J. P. Richie, and J. Fujimoto, "Optical coherence tomography as a method for identifying benign and malignant microscopic structures in the prostate gland," Urology **55**(5), 783–787 (2000).
4. A. Desjardins, B. Vakoc, G. Tearney, and B. Bouma, "Speckle reduction in OCT using massively-parallel detection and frequency-domain ranging," Opt. Express **14**(11), 4736–4745 (2006).
5. A. Buades, B. Coll, and J.-M. Morel, "Non-local means denoising," Image Processing On Line **1**, 208–212 (2011).
6. A. Chambolle, "An algorithm for total variation minimization and applications," J. Math. Imaging Vis. **20**(1/2), 163–177 (2004).
7. K. Dabov, A. Foi, V. Katkovnik, and K. Egiazarian, "Bm3d image denoising with shape-adaptive principal component analysis," in *SPARS'09-Signal Processing with Adaptive Sparse Structured Representations* (2009).
8. A. Abbasi, A. Monadjemi, L. Fang, and H. Rabbani, "Optical coherence tomography retinal image reconstruction via nonlocal weighted sparse representation," J. Biomed. Opt. **23**(03), 1 (2018).
9. X. Wang, X. Yu, X. Liu, S. Chen, S. Chen, N. Wang, and L. Liu, "A two-step iteration mechanism for speckle reduction in optical coherence tomography," Biomedical Signal Processing and Control **43**, 86–95 (2018).
10. L. Bian, J. Suo, F. Chen, and Q. Dai, "Multiframe denoising of high-speed optical coherence tomography data using interframe and intraframe priors," J. Biomed. Opt. **20**(3), 036006 (2015).
11. H. Rabbani, R. Nezafat, and S. Gazor, "Wavelet-domain medical image denoising using bivariate Laplacian mixture model," IEEE Trans. Biomed. Eng. **56**(12), 2826–2837 (2009).
12. J.-L. Starck, E. J. Candés, and D. L. Donoho, "The curvelet transform for image denoising," IEEE Trans. on Image Process. **11**(6), 670–684 (2002).

Biomedical Optics EXPRESS

13. C. Tian, L. Fei, W. Zheng, Y. Xu, W. Zuo, and C.-W. Lin, "Deep learning on image denoising: an overview," Neural Networks **131**, 251–275 (2020).
14. S. Lefkimmiatis, "Universal denoising networks: a novel cnn architecture for image denoising," in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition* (2018), pp. 3204–3213.
15. 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. on Image Process. **26**(7), 3142–3155 (2017).
16. C. Dong, C. C. Loy, K. He, and X. Tang, "Image super-resolution using deep convolutional networks," IEEE Trans. Pattern Anal. Mach. Intell. **38**(2), 295–307 (2016).
17. S. Shurrab and R. Duwairi, "Self-supervised learning methods and applications in medical imaging analysis: a survey," arXiv, arXiv:2109.08685 (2021).
18. D. Ulyanov, A. Vedaldi, and S. Victor, "Lempitsky: deep image prior," in *Computer Vision and Pattern Recognition (CVPR)* vol. 1, (2018).
19. Y. Quan, M. Chen, T. Pang, and H. Ji, "Self2self with dropout: Learning self-supervised denoising from single image," in *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition* (2020) pp. 1890–1898.
20. Z. Wang, J. Liu, G. Li, and H. Han, "Blind2unblind: self-supervised image denoising with 594 visible blind spots," in *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition* (2019), pp. 2129–2137.
21. Y. Huang, W. Xia, Z. Lu, Y. Liu, H. Chen, J. Zhou, L. Fang, and Y. Zhang, "Noise-powered disentangled representation for unsupervised speckle reduction of optical coherence tomography images," IEEE Trans. Med. Imaging **40**(10), 2600–2614 (2021).
22. A. Guo, L. Fang, M. Qi, and S. Li, "Unsupervised denoising of optical coherence tomography images with nonlocal-generative adversarial network," IEEE Trans. Instrum. Meas. **70**, 1 (2020).
23. Q. Zhou, M. Wen, M. Ding, and X. Zhang, "Unsupervised despeckling of optical coherence tomography images by combining cross-scale cnn with an intra-patch and inter-patch based transformer," Opt. Express **30**(11), 18800–18820 (2022).
24. Y. Li, Y. Fan, and H. Liao, "Self-supervised speckle noise reduction of optical coherence tomography without clean data," Biomed. Opt. Express **13**(12), 6357–6372 (2022).
25. J. J. Rico-Jimenez, D. Hu, E. M. Tang, I. Oguz, and Y. K. Tao, "Real-time OCT image denoising using a self-fusion neural network," Biomed. Opt. Express **13**(3), 1398–1409 (2022).
26. A. Krull, T.-O. Buchholz, and F. Jug, "Noise2void-learning denoising from single noisy images," in *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition* (2019), pp. 2129–2137.
27. I. Y. Chun, D. Park, X. Zheng, S. Y. Chun, and Y. Long, "Self-supervised regression learning using domain knowledge: Applications to improving self-supervised denoising in imaging," arXiv, arXiv:2205.04821 (2022).
28. L. Fang, S. Li, R. P. McNabb, Q. Nie, A. N. Kuo, C. A. Toth, J. A. Izatt, and S. Farsiu, "Fast acquisition and reconstruction of optical coherence tomography images via sparse representation," IEEE Trans. Med. Imaging **32**(11), 2034–2049 (2013).
29. 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).
30. S. Farsiu, S. J. Chiu, R. V. O'Connell, F. A. Folgar, E. Yuan, J. A. Izatt, and C. A. Toth, and Age-Related Eye Disease Study 2 Ancillary Spectral Domain Optical Coherence Tomography Study Group, "Quantitative classification of eyes with and without intermediate age-related macular degeneration using optical coherence tomography," Ophthalmology **121**(1), 162–172 (2014).
31. O. Ronneberger, P. Fischer, and T. Brox, "U-net: convolutional networks for biomedical image segmentation," in *International Conference on Medical image computing and computer-assisted intervention* (Springer, 2015), pp. 234–241.
32. D. P. Kingma and J. B. Adam, "A method for stochastic," *Optimization. In, ICLR*, vol. 5 (2015).
33. Z. Wang, A. C. Bovik, H. R. Sheikh, and E. P. Simoncelli, "Image quality assessment: from error visibility to structural similarity," IEEE Transactions on Image Processing, **13**(4), 600–612 (2004).
34. M. A. Mayer, A. Borsdorf, M. Wagner, J. Hornegger, C. Y. Mardin, and R. P. Tornow, "Wavelet denoising of multiframe optical coherence tomography data," Biomed. Opt. Express **3**(3), 572–589 (2012).
35. W. Liu, Y. H. Ong, X. J. Yu, J. Ju, C. M. Perlaki, L. B. Liu, and Q. Liu, "Snapshot depth sensitive Raman spectroscopy in layered tissues," Opt. Express **24**(25), 28312–28,325 (2016).
36. L. Liu, K. K. Chu, G. H. Houser, B. J. Diephuis, Y. Li, E. J. Wilsterman, S. Shastry, G. Dierksen, S. E. Birket, and M. Mazur, "Method for quantitative study of airway functional microanatomy using micro-optical coherence tomography," PLoS One **8**(1), e54473 (2013).
37. P. Mooney, "Retinal OCT images (optical coherence tomography)," Kaggle Dataset, 2018, http://dx.doi.org/10.17632/rscbjbr9sj.2.