Research Article | Vol. 15, No. 2 / 1 Feb 2024 / Biomedical Optics Express 1233
Biomedical Optics EXPRESS

Self-supervised Self2Self denoising strategy for OCT speckle reduction with a single noisy image

CHENKUN GE,[1] XIAOJUN YU,[1,2,*] MIAO YUAN,[1] ZEMING FAN,[1] JINNA CHEN,[3] PERRY PING SHUM,[1] AND LINBO LIU[4]

[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]Department of Electrical and Electronic Engineering, Southern University of Science and Technology, Shenzhen, Guangdong, 518055, China
[4]School of Electrical and Electronic Engineering, Nanyang Technological University, 639798, Singapore
*XJYU@nwpu.edu.cn

Abstract: Optical coherence tomography (OCT) inevitably suffers from the influence of speckles originating from multiple scattered photons owing to its low-coherence interferometry property. Although various deep learning schemes have been proposed for OCT despeckling, they typically suffer from the requirement for ground-truth images, which are difficult to collect in clinical practice. To alleviate the influences of speckles without requiring ground-truth images, this paper presents a self-supervised deep learning scheme, namely, Self2Self strategy (S2Snet), for OCT despeckling using a single noisy image. Specifically, in this study, the main deep learning architecture is the Self2Self network, with its partial convolution being updated with a gated convolution layer. Specifically, both the input images and their Bernoulli sampling instances are adopted as network input first, and then, a devised loss function is integrated into the network to remove the background noise. Finally, the denoised output is estimated using the average of multiple predicted outputs. Experiments with various OCT datasets are conducted to verify the effectiveness of the proposed S2Snet scheme. Results compared with those of the existing methods demonstrate that S2Snet not only outperforms those existing self-supervised deep learning methods but also achieves better performances than those non-deep learning ones in different cases. Specifically, S2Snet achieves an improvement of 3.41% and 2.37% for PSNR and SSIM, respectively, as compared to the original Self2Self network, while such improvements become 19.9% and 22.7% as compared with the well-known non-deep learning NWSR method.

1. Introduction

Optical coherence tomography (OCT) is an optical imaging modality based on low coherence interferometry, which provides high-resolution images of biological tissue microstructures [1]. Due to its non-invasive and high-resolution characteristics, OCT had been widely in various areas, especially in ophthalmology [2]. However, due to its low-coherence nature, OCT images suffer from speckles that are caused by multiple forward and backward scattering of illumination light. Speckle noise largely reduces the quality of OCT images, impacting on the accuracy of disease diagnosis [3]. To improve the quality of OCT image for disease diagnoses, various denoising methods have been proposed in literature over the past decades [4].

The traditional image denoising methods could be roughly divided into two categories, i.e, spatial domain denoising schemes, e.g., anisotropic filtering [5], non-local means filtering [6], and total variation [7], etc., and transform domain denoising schemes, e.g., wavelet transform [8], and curvelet transform [9]. Specifically, the former process OCT images directly in the spatial domain, while the latter process OCT images in the transform domain. Buades *et al.* proposed a non-local means (NLM) filtering method, which calculates the weight of neighborhood pixels according to

#515520 https://doi.org/10.1364/BOE.515520
Journal © 2024 Received 11 Dec 2023; revised 13 Jan 2024; accepted 16 Jan 2024; published 30 Jan 2024

Biomedical Optics EXPRESS

the image self-similarity [6]. The block-matching and 3D filtering (BM3D) employs the idea of block-wise estimation to denoise the images [10], wherein similar blocks are searched and stacked in the local area using a hard threshold denoising method, and those areas are aggregated according to the mean weight, and finally are denoised with the Wiener collaborative filtering [11]. The non-local weighted sparse representation (NWSR) adopts a sparse representation of multiple similar noisy and denoised patches to estimate each of the new patches [12]. However, it is worth noting that BM3D usually suffers from edge blurring effects when processing those images with high complexity and low contrast, while NWSR vectorization patch may destroy the structures of the reconstructed images in certain cases, causing some meaningful pathological details to be lost in the denoised images. Some other algorithms have also been proposed. For example, Wang *et al.* proposed a two-step iteration (TSI) method [13], while Yu *et al.* presents a noise statistical distribution analysis-based two-step filtering mechanism for OCT despeckling [14]. However, both methods suffer from their computational complexity since they decompose the speckles into additive and multiplicative ones, and suppress them sequentially in an iterative manner.

In recent years, deep learning techniques have emerged as an excellent tool for OCT despeckling owing to their ability to retain structural details in the denoising process. Jain *et al.* proposed to use a convolutional neural network (CNN) to denoise natural images for the first time and achieved satisfactory results as compared with those conventional methods [15]. Zhang *et al.* proposed a deep CNN network, named DnCNN, to reduce noises in OCT image [16]. It adopts a residual learning scheme to improve the network learning ability together with a batch normalization layer to address the gradient dispersion effect. However, it is worth noting that most of the current deep learning based denoising methods are supervised, wherein either a number of noisy and clean image pairs or sophisticated style-transferring training schemes are required for training [17–20].

To address such an issue, some other methods with semi-supervised training schemes have also been proposed. For those schemes, however, clean images are still required, which are usually difficult to obtain, especially for those intraoperative *in vivo* imaging systems. The wide application of such methods are largely hindered. Therefore, it is of great significance to devise deep learning schemes without requiring any clean images, i.e., the denoising network should be trained with noisy images only, and more and more attentions are paid to the unsupervised denoising learning methods. Ulyanov *et al.* proposed a deep learning model named deep image prior (DIP) for single image recovery, yet its performances are not competitive as compared with BM3D [21]. Zhou *et al.*[22] proposed a unsupervised learning method by using the sub-sampled noisy and denoised images to build a Neighbor2Neighbor loss and a PNLM loss for speckle reduction in OCT images [23]. Li *et al.* proposed a self-supervised scheme, namely MAP-SNR, by mapping the adjacent pixel blocks from the original noisy and the trained images [24]. Rico-Jimenez *et al.* devised a self-fusion neural network for real-time denoising of OCT images using three pre-trained frames [25]. Huang *et al.* adopted an unsupervised method, i.e., DRGAN, for speckle reduction without using image pairs, instead, employing a small amount of clean images for network training [26]. Yu *et al.* also proposed B2Unet to denoise OCT images based on the Bulin2Unblind mechanism, and achieved satisfactory denoising results [27].

Currently, although various semi-supervised or unsupervised schemes have been proposed, their performances are still not as good as those supervised learning ones, and few solutions using a single noisy image have been presented in literature. Therefore, this paper presents a new self-supervised Self2Self strategy, namely, S2Snet, utilizing a single noisy image for OCT denoising. By utilizing a gated convolutional layer to improve the Self2Slef network [28–30] denoising performances, and a devised loss function to remove the background noise, S2Snet can acquire the detailed structures while remove the background noise of the whole image effectively. The main contributions of this study are as follows,

Biomedical Optics EXPRESS

1. A Self2Self network together with a gated convolution layer is employed for OCT despeckling using a single noisy image and its Bernoulli sampled instances.
2. A loss function for background noise suppression is devised and integrated onto the overall loss function to improve OCT despeckling performances.
3. Experiments with various OCT datasets are conducted to compare S2Snet with both the existing self-supervised learning schemes and those non-deep learning methods to verify its effectiveness in different cases.

The rest of this paper is organized as follows. Section 2 briefly introduces the main principle of the proposed self-supervised denoising mechanism. Section 3 presents the proposed S2Snet strategy, including its training and denoising schemes. Section 4 presents the denoising results obtained with some clinical datasets. Section 5 concludes the whole paper.

2. Principle

2.1. Denoising principles

The main purpose of image denoising is to preserve the image structural details while remove the background noises. Typically, an OCT noisy image y could be modeled as,

$$y = x + n \tag{1}$$

where x denotes the ground truth image, and n denotes the random noise. The neural network denoiser developed in this study is denoted by $F(\cdot)$, which can be trained with a single noisy image for denoising. Hence, such a process could be denoted as follows,

$$F(\cdot) : y \rightarrow x \tag{2}$$

With the denoising neural network being interpreted as a Bayesian estimator, then its prediction accuracy could be measured by mean square error (MSE) as below,

$$MSE = bias^2 + variance \tag{3}$$

It is noted that reducing the number of training samples to a single image would largely increase variance, which is detrimental to denoising. Therefore, it is important to minimize the above variance for self-supervised learning.

To reduce the variance of the Bayesian estimator, Self2Self network [28] introduces a dropout-based ensemble, which is a regularization technique widely being used in deep neural networks [31], assuming that the activation value of a neuron stops working with a certain probability p when propagating forward. Owing to the model uncertainty introduced by dropout [32], the predictions of such models may have a certain degree of statistical independence, and the average of these predictions will reduce the variance of the results. Therefore, such an assumption makes the model more generalized, since it will not rely too much on some local features, instead, it utilizes a single neural network to approximate several neurons.

2.2. Basic idea

Although dropout is introduced into the original Self2Self network to maintain the image structural details with a single image, there would still be some noise residue in the background. To address such a problem, the original loss function of Self2Self is amended.

When using Self2Self network to denoise an image, it is found that the initial result has no background noise, while the image structure details are relatively fuzzy. In the training process, however, although the resultant image details become clearer gradually, there still exist some noises in the background. Such results indicate that when OCT images are trained with Slef2Self network, the image structures are generated first, followed by image background, and therefore,

Biomedical Optics EXPRESS

if the learning of image background could be reduced, the image background noise would be largely reduced in the achieved image. Inspired by such an idea, a new loss function is devised for the Self2Self denoising network, wherein both a self-prediction loss and a background noise attenuation loss are integrated in the network loss function.

A bernoulli sampled instance $\hat{y}$ of an image y with probability p is defined as below,

$$\hat{y}[i,j] = \begin{cases} y[i,j], p \\ 0, 1-p \end{cases} \tag{4}$$

where $[i,j]$ is position of image pixels.

The independent Bernoulli sampled instances of y are divided into two sets $\{\hat{y}_m\}_m$ and $\{\tilde{y}_n\}_n$. The training and testing processes of S2Snet are summarized below.

Training: S2Snet is trained by minimizing the following loss function with Bernoulli dropout:

$$\min \sum_m \left[L\left(F\left(\hat{y}_m\right), y - \hat{y}_m\right) + L\left(F\left(y\right), F\left(\hat{y}_m\right)\right) \right] \tag{5}$$

Testing: Each $\tilde{y}_n$ is input into a trained model with Bernoulli dropout to generate a predicted $\tilde{x}_n$. The denoising result is the average of all predictions.

3. Method

In this section, the S2Snet network architecture as well as its training and denoising schemes are introduced.

3.1. S2Snet network architecture

Fig. 1 shows the S2Snet architecture, which consists of an encoder-decoder network structure. As seen in Fig. 1, the size of an OCT noisy patch is set to be 256×256×1, and 256×256×48 features could be obtained via the encoder with gated convolution [29], which is then processed by the following six encoder blocks. The first 5 encoder blocks contain a gated convolution layer, a leaky rectified linear unit (LReLU), and a max pooling layer with 2×2 kernel and a stride of 2. The last encoder block includes a gated convolution layer and a LReLU. The number of output channels of every encoder block is set to be 48, and thus, the size of the last encoder finally output is 8×8×48. All the gated convolution layers are using 3×3 convolution kernel, and each LReLU is set to be 0.1. The decoder part has 5 decoder blocks, each of which contains an up-sampling layer, a convolution layer with dropout, and a LReLU. Each up-sampling layer adopts 2 scaling factors for decoding. The number of output channels for each of the first four decoder blocks is set to be 96. The final decoder block utilizes three convolution layers with a dropout and a LReLU to obtain output image with a size of 256×256×1.

The whole network structure is quite similar to that of the Self2Self network, with the partial convolution being updated with a gated convolution in the encoder block, which helps improve the effectiveness and efficiency of network model training.

The gated convolution is shown in Fig. 2 (a), and it learns soft mask automatically from the input image data, which is formulated as :

$$\begin{cases} Gating_{y,x} = \sum\sum W_g \cdot I \\ Feature_{y,x} = \sum\sum W_f \cdot I \\ O_{y,x} = \theta\left(Feature_{y,x}\right) \cdot \sigma\left(Gating_{y,x}\right) \end{cases} \tag{6}$$

where σ is sigmoid function, and therefore, the output gated values are in between zero and one. θ is defined as an activation function, e.g., ReLU, ELU and LeakyReLU. Denote each pixel

Biomedical Optics EXPRESS

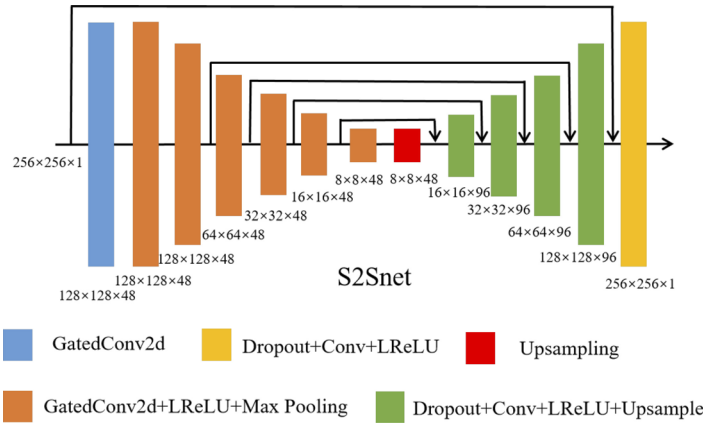

Fig. 1. Architecture of proposed S2Snet network.

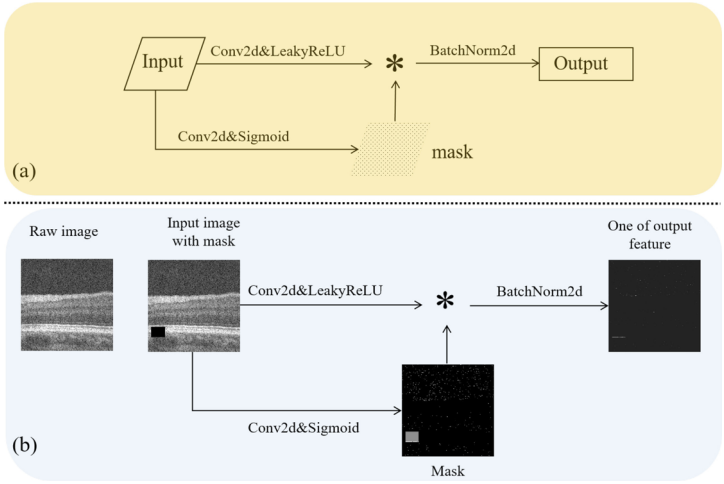

Fig. 2. (a) Illustration of gated convolution, (b) Visualizing results of gated convolution on noise image with masked region.

located at (y, x) of an output map to be $O_{y,x}$, and I to be the input feature. W_g and W_f are two different convolution layers.

Figure 2(b) depicts the visual results of the gated convolution layer, wherein initialized gated convolution is applied to the noisy image with masks and the resulted single-channel feature map could be obtained. It illustrates that, despite the utilization of gated convolution with random parameters, the generated mask still closely aligns with the original masked region. If a loss function is introduced for training, the gated convolution layer can learn from the features in the region. It proves that gated convolution enables the learning of a dynamic feature selection mechanism for every channel and spatial location. Hence, it not only learns to select features based on the background and mask, taking into account semantic segmentation in certain channels, but also learns to emphasize the masking areas across different channels even in deeper layers, thereby improving the generation of inpainting results.

Biomedical Optics EXPRESS

3.2. Training scheme

Fig. 3 illustrates the training procedure of the proposed method. In this process, both the input image and its Bernoulli sampled image are fed simultaneously into the S2Snet with shared weight. This yields two different denoised images. To facilitate training, self-prediction loss and background noise attenuation loss are computed. Next, the relevant details are explained later.

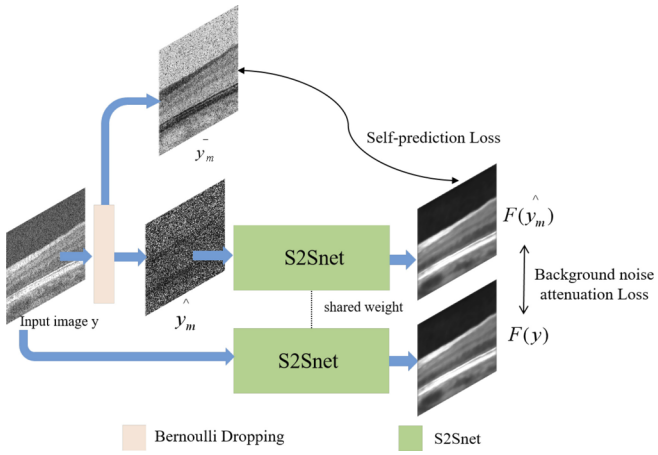

Fig. 3. The training process of S2Snet.

Since features learned from a single noisy image y cannot be adopted for effective denoising, multiple image pairs have to be generated from y, and therefore, a series of Bernoulli sampling pairs are employed to train the S2Snet network. Such a set of image pairs $\{(\hat{y}_m, \bar{y}_m)\}_{m=1}^{M}$ is defined as below,

$$\hat{y}_m = b_m \cdot y; \bar{y}_m = (1 - b_m) \cdot y \tag{7}$$

where $\cdot$ means the element-wise multiplication and b indicates the mask image obtained after Bernoulli sampling, and its shape is the same as the noisy image y.

With each set of generated image pairs, the self-prediction loss could be defined as below,

$$Loss_{self-prediction} = \min \sum_{m=1}^{M} \|f(\hat{y}_m) - \bar{y}_m\|_{b_m}^2 \tag{8}$$

where $f(\cdot)$ denotes the training network, $\|\cdot\|_b^2 = \|(1-b)\cdot\|_2^2$

To better eliminate the background noise, the background noise attenuation (bna) loss is also introduced, and it is defined as below,

$$Loss_{bna} = \min \frac{\sum_{m=1}^{M} |f(y) - f(\hat{y}_m)|}{M} \tag{9}$$

Given the self-prediction loss and the background noise attenuation loss, the overall network loss function is defined as below,

$$Loss = Loss_{self-prediction} + \alpha \cdot Loss_{bna} \tag{10}$$

where α is weight of background noise attenuation loss.

The main purpose of such a devised loss function is to calculate the loss of partially masked pixels by b_m using the devised self-prediction loss, while with the background noise attenuation loss, the output of the original noisy image and its Bernoulli sampling instances could be

Biomedical Optics EXPRESS

minimized, to moderate the network learning ability of the image background noise. Therefore, self-prediction loss helps S2Snet learn information related to ground truth image x from beginning to end, since training the complementary Bernoulli sampled image pairs $\{(\hat{y}_m, \bar{y}_m)\}$ is very close to that of a Bernoulli sampled image pairs $\hat{y}_m$ and the clean image x. While when those complementary Bernoulli sampled image pairs are used for training; background noise attenuation loss helps degrade the background noise learning. Typically, in the early stage of training, $f(\hat{y}_m)$ generates the image structural details via network learning first, and then learns the relevant noises slowly. $f(y)$ and $f(\hat{y}_m)$ are minimized to degrade the learning ability of the network in the background area in the initial learning process.

3.3. Denoising scheme

A trained neural network with dropout could be modeled as a series of neurons, whose weights follow an independent Bernoulli distribution. In this study, dropout is employed to generate multiple neurons from the trained neural networks, and thus, multiple independent estimators could be generated to reduce the variance of denoised images. With a Bernoulli sampled image y input into every neurons, the multiple denoised images $\tilde{x}_1, \ldots, \tilde{x}_N$ could be generated, while the final denoised result x^* could be generated by averaging over those multiple denoised images,

$$x^* = \frac{1}{N}\sum_{n=1}^{N}\tilde{x}_n = \frac{1}{N}\sum_{n=1}^{N} f\left(b_{M+n}\cdot y\right) \tag{11}$$

In the denoising process, the denoising results can be generated simultaneously with the training process.

4. Experiments

During the training process, the size of the noisy image is cropped to be 256×256 for training purposes, and thus, the optimal parameters could be determined. Experiments are conducted using different sets of OCT noisy images, while the denoising metrics are also calculated simultaneously. Such results are compared to those of the state-of-the-art existing methods in different cases.

4.1. Datasets

Two public OCT image datasets are utilized for denoising experiments in this study. These images are collected by the Bioptigen SDOCT system (Durham, NC, USA) with an axial resolution of $4.5\mu m$ per pixel in tissue [33]. Those datasets are denoted as D1 and D2 for the experiments.

The dataset D1 consists of 18 pairs of noisy and clean OCT images [34,35]. The clean images are obtained by registering and averaging over several B-scans that are acquired at same position. Specifically, 10 high quality noisy and clean image pairs, the size of which is 500×950, are selected for testing. and are normalized all the signals value of images. The clean images are only used to calculate the denoising metrics.

The dataset D2 contains 39 retina OCT images [34], the size of each image is 450×450. There is no corresponding clean image for reference in the D2. All signals value of images in the dataset have also been normalized before training.

4.2. Parameters setting

In the S2Snet, the dropout probability and probability of Bernoulli sampling are both set to 0.5. For the training loss function, α is set to 0.3. The overall training adopts a learning rate of 10^{-4} and consists of a total of 2000 training epochs, with every 200 epochs serving as a test epoch and a denoised image is generated. And then the best denoised image is selected by the best indicator.

Biomedical Optics EXPRESS

The networks were implemented in Python using the PyTorch framework, and all experiments were performed on a workstation equipped with an Intel Xeon W-2145 CPU @3.70GHz and accelerated by an NVIDIA GeForce RTX 3060Ti GPU with 8GB memory.

4.3. Quantitative metrics

In this paper, we choose the following metrics to evaluate denoising performance.

Signal-to-Noise Ratio((PSNR): The PSNR is the main metric that measures the similarity between the denoised image and the reference one, and it can be calculated as follows,

$$PSNR(r, g) = 10log_{10}(255^2/MSE(r, g)) \tag{12}$$

$$MSE(r, g) = \frac{1}{MN}\sum_{i=1}^{M}\sum_{j=1}^{N}(r_{ij} - g_{ij})^2 \tag{13}$$

where r_{ij} and g_{ij} are the pixel values at the corresponding coordinates of the clean and denoised regions, respectively, while M and N are the height and width of the image.

Structural similarity index measurement (SSIM): SSIM is a full reference metric being widely used for image quality evaluation. SSIM is calculated as,

$$SSIM(i, b) = \frac{(2\mu_i\mu_b + C_1)(2\sigma_{ib} + C_2)}{(\mu_i^2 + \mu_b^2 + C_1)(\sigma_i^2 + \sigma_b^2 + C_2)} \tag{14}$$

where μ_b/μ_i and σ_b/σ_i are the mean and standard deviations of a clean/denoised region, respectively, while σ_{ib} denotes the cross-correlation between the clean and the denoised regions. C_1 and C_2 are random positive stabilizing constants.

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.,

$$SNR = 20log(I_{max}/\sigma_b) \tag{15}$$

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.

Contrast to noise ratio(CNR): CNR is the contrast to noise ratio, which is defined as the ratio of peak signal strength to background strength. CNR is the ratio of image contrast to noise. It is an objective index to evaluate image quality. It can be defined as:

$$CNR = \frac{1}{n}\sum_{i=1}^{n} 10\log\left(\frac{|\mu_i - \mu_B|}{\sqrt{\sigma_i^2 + \sigma_B^2}}\right) \tag{16}$$

where μ_i and σ_i^2 denotes the mean and variance of select region, and μ_b, σ_b^2 denotes the mean and variance of background region.

Equivalent Number of Looks(ENL): ENL is a commonly used speckle suppression performance measurement method, which measures smoothness in regions that appear to be homogeneous. Here, it is only in background region, as

$$ENL = \mu_b^2/\sigma_b^2 \tag{17}$$

where μ_b and σ_b denote mean value and standard deviation of the background region, respectively.

Biomedical Optics EXPRESS

4.4. Results of single image with mini size

To begin with, we performed the experiment using the cropped noise image with a size of 256×256. The experimental results are depicted in Fig. 4. In the initial denoised image output, the background section appears noise-free, but there are some blurring artifacts present in the overall structural details. As the training progresses, the structural details of the image gradually improve, resulting in a sequence of subsequent denoised images exhibiting remarkable quality.

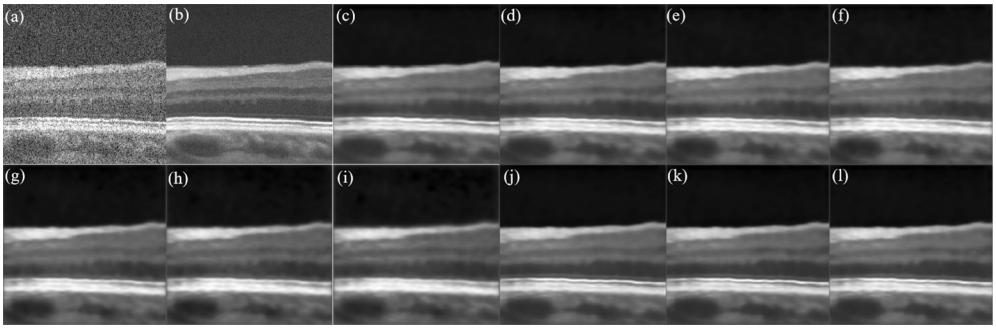

Fig. 4. A single noise OCT image cropped as 256×256 and its clean image and its denoising results. (a) noisy image,(b)clean image,(c)Result of the first test epoch of output, (d)Result of the second test epoch of output, (e)Result of the third test epoch of output,. . . (l)Result of the 10th test epoch of output.

We recorded the loss values for each train epoch and the denoising metrics for each test epoch, as depicted in Fig. 5. Figure 5 (a) illustrates the progression of loss values throughout the training process. It is shown that there exists a rapid decrease in the loss for the first 500 epochs. Although minor fluctuations in the loss value could still be observed in subsequent epochs, it converges steadily during the entire training process, indicating that the method employed is adept at achieving stable training. Figure 5(b) presents the denoising metrics obtained for each epoch of training. As seen, both denoising metrics exhibit a relatively wide range of numerical fluctuations in test epochs due to the inherent randomness presented in the mask processing and dropout. Such a variability is expected and reasonable. In the study, the metric values obtained in the test epochs are recorded and the image with the highest denoising metrics are selected as the final denoised image. It is worth mentioning that the two metrics we utilized are based on a comparison with a clean image. However, since clean images are deemed to be completely free of any noise while the residual noises are always presented in the background even for a clean image, it is important to consider the computed metrics as reference values only.

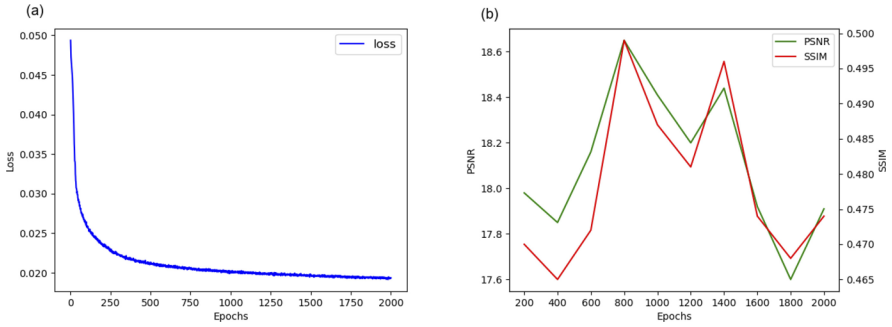

Fig. 5. Numerical results for single image with 256×256. (a)The loss value of each train epoch, (b)The denoising metrics of each test epoch.

Biomedical Optics EXPRESS

4.5. Results comparison with other method in D1

For verifying the effectiveness of the proposed S2Snet, experiments are tested on a public OCT retinal image dataset from [34]. And our method is compared with some state-of-the-art denoising methods: BM3D [10], NWSR [12], B2Unet [27], TSI [13], DRGAN [26], MAP-SNR [24] and Self2Self [28]. 10 OCT retinal images from D1 are processed for comparisons in this study. After denoising 10 different OCT images, we recorded the denoising metrics for each epoch. From these metrics, we selected the image with the best performance as the final denoised result. We also tested various other algorithms to determine the optimal results.

Here, two testing of OCT images are provided to demonstrate visual comparisons of denoising results achieved by different methods. In Fig. 6, the first testing image depicts the denoising outcomes obtained through BM3D, NWSR, TSI, B2Unet, DRGAN, MAP-SNR, and Self2Self. Additionally, a cropped image patch with background noise is included for observation. Based on the observations, it is clear that BM3D exhibits noticeable noise presence. On the other hand, NWSR, TSI, and Self2Self achieve significant noise reduction, although residual noise remains in certain background areas. DRGAN and B2Unet produce denoised images with excellent performance; however, they display some smooth noise in the background. Furthermore, MAP-SNR effectively reduces speckle noise while preserving image structure details and eliminating background noise. Moreover, our proposed method successfully suppresses background noise and preserves structural details in both the complete image and the cropped image patch during the denoising process. Figure 7 (a) shows the loss variation for the first test image during the training process. As seen, the loss value converges with an increasing number of training epochs, demonstrating the stability of our proposed method. Figure 7 (b) depicts the changes of calculated metrics during the testing epochs, and it can be observed that the change trend of PSNR and SSIM remains consistent throughout the entire training process, indicating a high correlation between them. Moreover, these metrics do not continually increase and may decline at times. Therefore, it is necessary to manually select the most effective testing result.

Similar conclusions can be drawn from the second testing image, as depicted in Fig. 8. Furthermore, the change trend of loss value, PSNR and SSIM in Fig. 9 aligns with the previous findings. This reaffirms the convergence of loss values during training and the existence of strong correlation between PSNR and SSIM throughout the entire training process, with an overall trend towards improved metrics.

We utilize 10 OCT images from the D1 as noise images and calculate the average metrics for each method. The results are summarized in Table 1, which illustrates the performances of different methods with the noise images. Hence, it can be observed from Table 1 that S2Snet achieves the highest PSNR and SSIM values among all those denoising methods, which shows that our methods can better remove the noise especially background noise when compared with other methods. Compared with the original Self2Self network, our method shows improvements in PSNR and SSIM by 3.41% and 2.37%, respectively.

S2Snet ranks second for CNR, and in the middle for both SNR and ENL. Such is mainly because B2Unet, DRGAN, and MAP-SNR are trained with a large noisy image dataset, allowing them to denoise effectively. As a result, higher SNR and ENL could be achieved as compared to those single image denoising methods. In contrast, other methods address individual noisy images only, resulting in lower SNR and ENL as compared to those three methods. From the above analysis, we conclude that S2Snet outperforms the other state-of-the-art despeckling methods in terms of PSNR and SSIM, making it a candidate for clinical denoising applications to aid in accurate clinical diagnosis.

4.6. Results comparison with other method in D2

For the noise images of D2, we also choose seven different denoising methods for comparison. Since D2 has no relevant clean image as a reference, we choose Signal-to-Noise Ratio(SNR),

Biomedical Optics EXPRESS

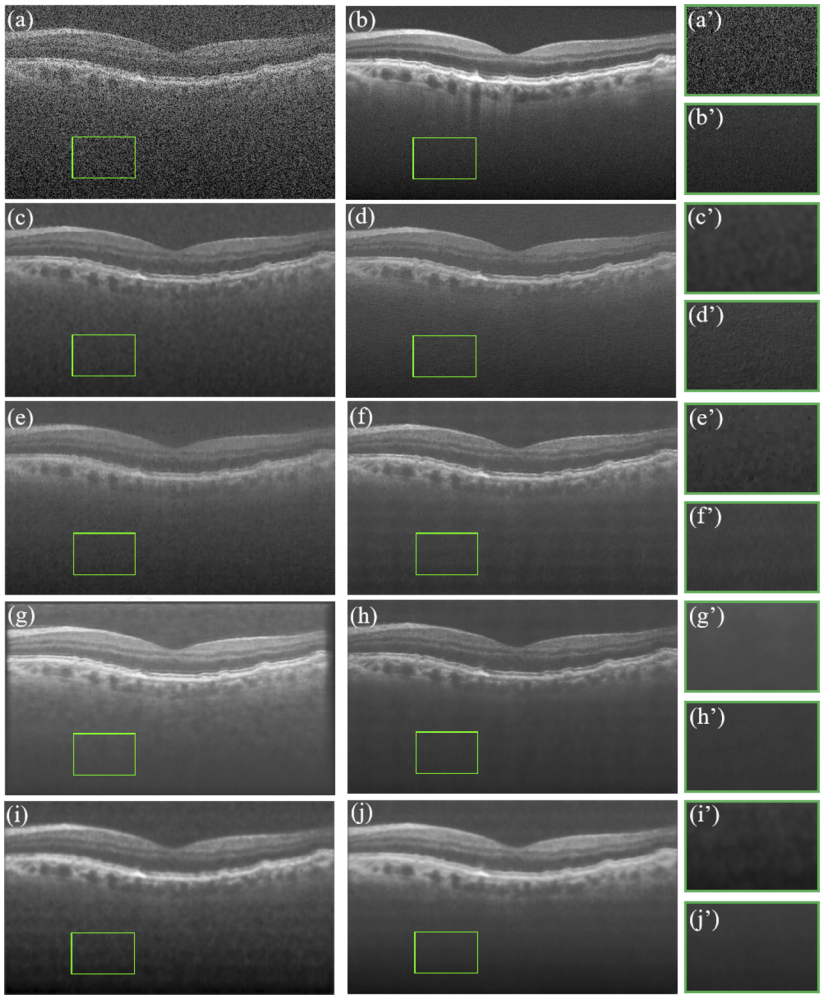

Fig. 6. First testing OCT image selected from D1 and its denoising results. (a)Original, (b)averaged, (c)BM3D, (d)NWSR, (e)TSI, (f)B2Unet, (g)DRGAN, (h)MAP-SNR, (i)Self2Self2, and (j)Our. (a')$\cdots$(j')Enlarged image of the corresponding background area.

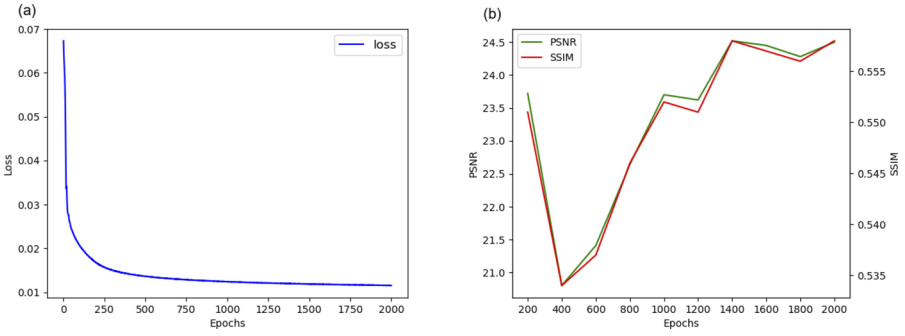

Fig. 7. Numerical results for first testing image. (a)The loss value of each train epoch, (b)The denoising metrics of each test epoch.

Biomedical Optics EXPRESS

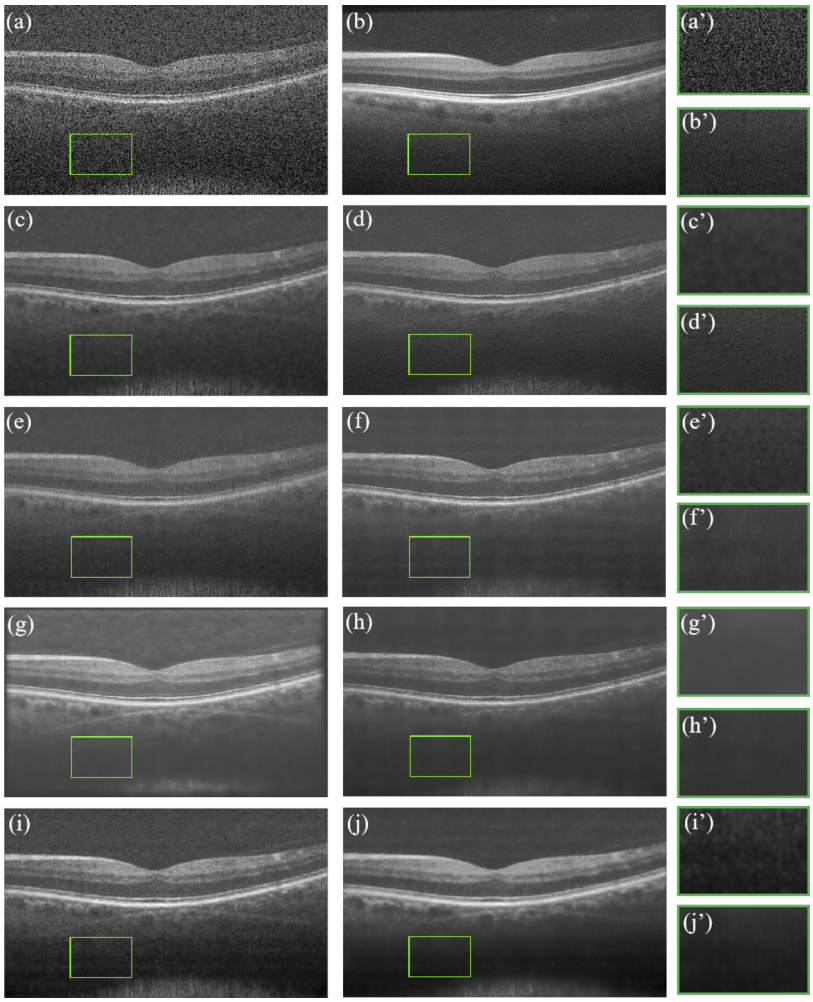

Fig. 8. Second testing OCT image selected from D1 and its denoising results. (a)Original, (b)averaged, (c)BM3D, (d)NWSR, (e)TSI, (f)B2Unet, (g)DRGAN, (h)MAP-SNR, (i)Self2Self2, and (j)Our. (a')$\cdots$(j')Enlarged image of the corresponding background area.

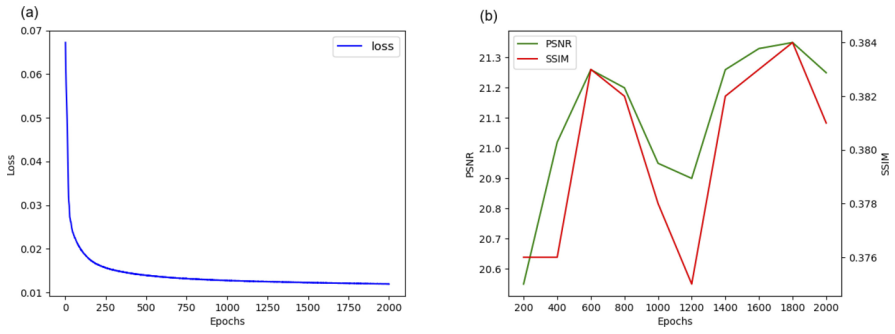

Fig. 9. Numerical results for second testing image. (a)The loss value of each train epoch, (b)The denoising metrics of each test epoch.

Biomedical Optics EXPRESS

Table 1. Quantitative results of the D1 with different methods. The best value is represented in bold.

	PSNR	SSIM	SNR	CNR	ENL
BM3D	21.0777	0.5166	34.7123	3.3472	220.3921
NWSR	20.6912	0.4432	31.4536	3.3357	111.8567
TSI	20.6394	0.4926	33.5246	**3.9108**	235.1185
B2Unet	20.5100	0.4996	**48.3045**	2.8456	**6193.5509**
DRGAN	17.4985	0.4590	37.6299	3.7403	638.7958
MAP-SNR	20.8778	0.5194	39.3228	3.1690	643.2975
Self2Self	23.9815	0.5312	33.0583	3.5047	70.4276
S2Snet	**24.8002**	**0.5438**	36.7366	3.8679	184.1626

Contrast to Noise Ratio(CNR) and Equivalent Number of Looks(ENL) as the denoising metrics, which can be calculated only with the denoised image. In order to calculate these metrics, we need to extract the background and structure regions from all the images in D2. Fig. 10 presents enlarged views of both regions for comparison purposes.

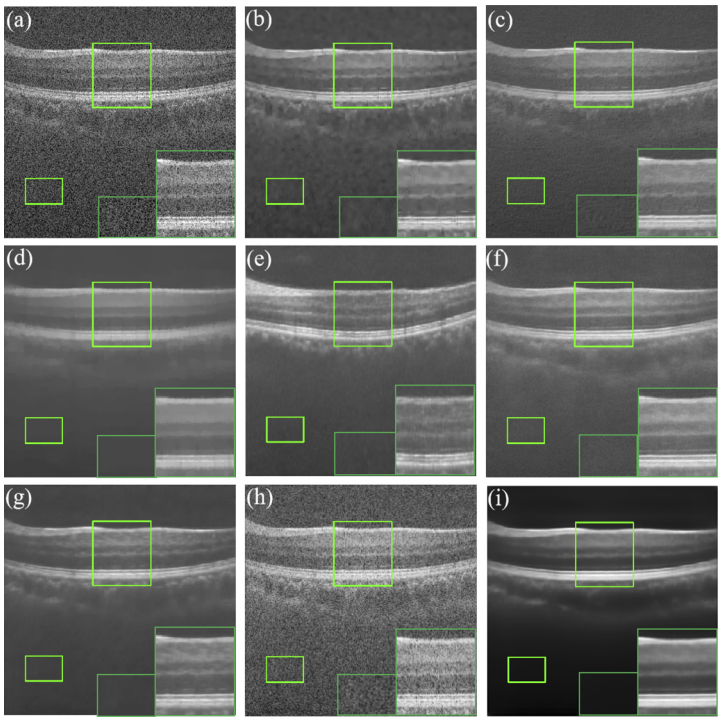

Fig. 10. Testing OCT image selected from D2 and its denoising results. (a)Original, (b)BM3D, (c)NWSR,(d)TSI, (e)B2Unet, (f)DRGAN, (g)MAP-SNR, (h)Self2Self, and (i)Our.

Fig. 10 illustrates the results of different denoising techniques. The image processed by BM3D exhibits unsatisfactory effects, with a few artifacts in the structure area. NWSR demonstrates good denoising effects in the structure area, but some noise remains in the background. Overall denoising effect of TSI is excessively smooth. Conversely, DRGAN and B2Unet, employing unsupervised learning, yield excellent denoising outcomes, yet weak speckle noise still exists

Biomedical Optics EXPRESS

in the background. MAP-SNR exhibits remarkable overall denoising effects, allowing for clear observation of the retinal layer distribution. Denoising effectiveness of Self2Self is very poor, as it fails to remove speckle noise from both the background and structure areas. In our method, the background area not only completely suppresses speckle noise but also darkens the pixels in that region. Meanwhile, the structural area retains its complete features. This approach effectively reduces speckle noise, enhances contrast between the background and structure areas, and highlights the boundary of the retinal layer.

Table 2 presents the quantitative results achieved using different denoising techniques. TSI exhibits the highest SNR, surpassing MAP-SNR by 1.2%. Additionally, our method ranks second in CNR, indicating a significant contrast between the background and structure regions of denoised image, making it easier to highlight the structural content. Due to the darkened background area in our denoised images, the obtained ENL value is relatively low. Thus, both the comparison of denoised images in Fig. 10 and the evaluation of various metrics in Table 2 support the effectiveness of our proposed method for denoising conventional OCT images in clinical processing.

Table 2. Quantitative results of the D2 with different methods. The best value is indicated in bold, while the second best value is indicated by underline.

	SNR	CNR	ENL
BM3D	33.9544	1.2856	175.1801
NWSR	32.6035	1.3459	146.0455
TSI	40.0417	**2.2217**	**1343.8894**
B2Unet	39.5504	1.3117	686.4495
DRGAN	34.4314	1.4646	271.7027
MAP-SNR	41.8373	1.3708	1153.4238
Self2Self	23.4729	1.0848	18.2737
S2Snet	**42.3406**	2.0032	125.0791

4.7. Computational cost study

This section examines the computational costs of different methods to evaluate their potential for practical applications. Although B2Unet, DRGAN, and MAP-SNR exhibit fast processing speeds during the inference stage for using the pre-trained models, they require a substantial number of noisy images for effective long-term training. Hence, the computational cost evaluation considers only BM3D, NWSR, TSI, Self2Self, and S2Snet, of which the time required is for a single image processing, and the denoised image could be obtained directly by inputting a single image. Furthermore, since the computation time also varies depending on the computing devices used (CPUs and GPUs), the time cost of BM3D, NWSR, TSI, Self2Self, and S2Snet are evaluated using CPUs, while the time costs of Self2Self and S2Snet are compared with GPUs. The relevant data is presented in Table 3.

Table 3. Computational cost of different methods with a single image (500×950), where n is the expected training epochs.

Method	BM3D	NWSR	TSI	Self2Self(CPU)	S2Snet(CPU)	Self2Self(GPU)	S2Snet(GPU)
Times(s)	21.842	18.231	12.844	1.116×n	1.465×n	0.040×n	0.062×n

Results Table 3 indicate that BM3D, NWSR, and TSI take approximately 10-20 seconds for CPU processing. However, the denoising performances of these methods are comparable with those of deep learning methods. Furthermore, for both CPU and GPU processing, since

Biomedical Optics EXPRESS

S2Snet method introduces additional loss functions and internal network structure changes, its computation time are 31.3% and 55% higher than those of Self2Self for CPU and GPU, respectively. When utilized for practical applications, however, such increasing computational cost could be easily solved with multi-card training or higher-performance GPUs.

4.8. Segmentation study

Applying image denoising preprocessing before image segmentation has been found to generally enhance segmentation accuracy. Therefore, we employ this approach to preprocess the OCTA vessel segmentation dataset and train the denoised images for segmentation. Additionally, we also utilize the Self2Self method for comparison. Ultimately, we compare and evaluate these trained models.

In this study, the published Retina OCTA Vessel Segmentation (ROSE) dataset provided by [36] was used for experiments. ROSE dataset consists of two subsets, ROSE-1 and ROSE-2. The ROSE dataset provided the superficial vascular complexes (SVC), deep vascular complexes(DVC), and SVC+DVC images for experiments. We only used SVC+DVC images for image segmentation experiments. In SVC+DVC datasets, the training datasets includes 30 pairs of OCTA vessel images and their ground truth images of segmentation, and the verification datasets includes 9 pairs of OCTA vessel images and their ground truth images of segmentation.

During the experiment, the U-net is chosen as the training model for image segmentation. Prior to segmentation, we employ the Self2Self and S2Snet methods to denoise the original image. Both types of denoised images are then utilized as input images for training and testing in the image segmentation process. In order to comprehensively evaluate the segmentation performance, six performance indicators, i.e., area under the ROC Curve (AUC), accuracy (ACC), g-mean score, kappa score, dice coefficient (Dice), and Intersection-over-Union (IOU), are employed for performance assessment in this study [37].

Fig. 11 and Fig. 12 show the results of segmentation by training models processed by different denoising methods. It can be seen from Fig. 11. (b) that our method can effectively remove the subtle noise of the original image, but the micro-vessels of the original image will be affected by the denoising process, as shown in Fig. 11. (e). Because the denoising has an impact on the micro-vessels, the segmentation result cannot effectively restore the thin vessels, but it can restore the thick blood vessels well. As shown in Fig. 11. (c), the denoised image by Self2Self still has some noises and artifacts, and the denoising effect is not very good, while as shown in Fig. 11. (f), the segmented result can retain the coarse blood vessel structure and part of the fine blood vessel structure.

In summary, our denoising method proves effective in removing noise from OCTA images. However, it may adversely affect the segmentation of thin blood vessels, but it can effectively extract the structure of thick blood vessels. On the other hand, the denoising effect of Self2Self method is not particularly significant. But in segmentation results, both thick and thin blood vessels can be segmented, although the extraction of thin blood vessels may not be complete.

To further enhance the evaluation of the segmentation effect after noise removal, we computed the evaluation metrics for the segmentation results of the test dataset, as presented in Table 4. The table reveals that the segmentation results are significantly improved after denoising treatment, emphasizing the necessity of conducting denoising preprocessing before segmentation. Additionally, in comparison to the segmentation indicators following Self2Self denoising, our method outperforms Self2Self in all other metrics except for G-mean. Particularly, in terms of AUC and ACC, our segmentation results are 2.34% and 1.17% higher respectively than those obtained after Self2Self denoising.

Biomedical Optics EXPRESS

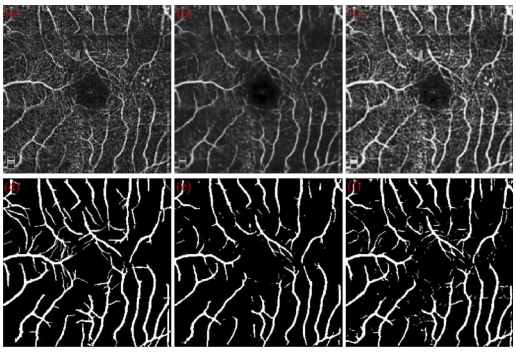

Fig. 11. Different model segmentation results of the first image in the validation dataset. (a)Original image, (b)Denoised image of original image by our method, (c)denoised image of original image by Self2Self, (d)Ground truth image, (e)Segmentation result of the model trained with dataset denoising by our method, (f)Segmentation result of the model trained with dataset denoising by Self2Self.

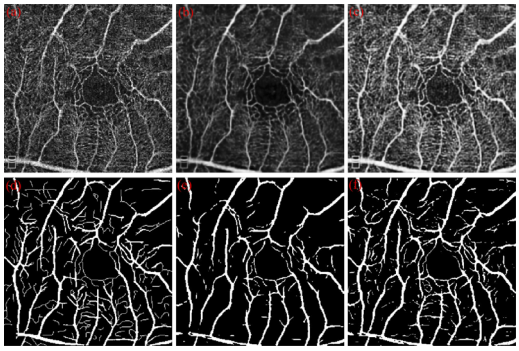

Fig. 12. Different model segmentation results of the second image in the validation dataset. (a)Original image, (b)Denoised image of original image by our method, (c)denoised image of original image by Self2Self, (d)Ground truth image, (e)Segmentation result of the model trained with dataset denoising by our method, (f)Segmentation result of the model trained with dataset denoising by Self2Self.

Table 4. Quantitative results of vessel segmentation after using different denoising methods. The best value is represented in bold.

	AUC	ACC	G-Mean	Kappa	Dice	IOU
original	0.8752	0.8883	0.6733	0.5481	0.6057	0.4359
Self2Self	0.8808	0.8868	**0.7917**	0.6214	0.6904	0.5289
Ours	**0.9014**	**0.8972**	0.7803	**0.6402**	**0.7015**	**0.5421**

4.9. Ablation study

To evaluate the effectiveness of its individual components, the following ablation studies were performed on D1 data set:

without gated convolution layer: replacing all gated convolution layer with Conv layer;

without background noise attenuation loss: Loss function has only self-prediction loss;

without dropout: disabling dropout on all layers during training and test.

Table 5 provides insights into the averaged PSNR, SSIM, SNR, CNR and ENL that are derived from D1. Firstly, the incorporation of a gated convolution layer in the denoising process leads to

Biomedical Optics EXPRESS

significant improvements, indicated by a 6.9% increase in PSNR and a 3.4% increase in SSIM. Secondly, the inclusion of background noise attenuation loss demonstrates its crucial role in our denoising method, resulting in a remarkable enhancement of 30.7% in PSNR and 66.5% in SSIM for the overall approach. Lastly, the utilization of dropout contributes to a 9.1% increase in PSNR and a 4.1% increase in SSIM. While for SNR, S2Snet with all components ranks the first, while for both CNR and ENL, S2Snet without dropout ranks the first. The reason why for such CNR and ENL ranking is that the generated images appear too smooth by the S2Snet without dropout, which lead to higher CNR and ENL, which could be observed in Fig. 13(e).

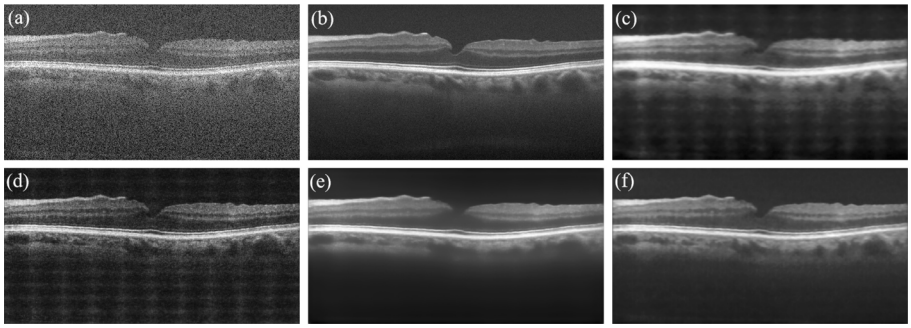

Fig. 13. Visual results for ablation studies. (a) Original, (b) Ground truth image, and denoised images obtained by (c) S2Snet without gated conv, (d) S2Snet without BNA loss, (e) S2Snet without dropout, (f) S2Snet.

Table 5. Results of ablation studies on data set. The best value is represented in bold.

	With Gated Cov	With BNA loss	With dropout	PSNR(dB)	SSIM	SNR	CNR	ENL
S2Snet	✗	✓	✓	23.1990	0.5258	30.9466	3.9268	37.1831
S2Snet	✓	✗	✓	18.9779	0.3267	25.2396	3.2489	26.7884
S2Snet	✓	✓	✗	22.7330	0.5223	36.5994	**4.0390**	**213.4703**
S2Snet	✓	✓	✓	**24.8002**	**0.5438**	**36.7366**	3.8679	184.1626

We also included the relevant denoised images for comparison, as depicted in Fig. 13. In Fig. 13(c), it is known that without the inclusion of gated convolution, there is noticeable regular noise in the background area of the denoised images, indicating that the utilization of BNA loss and dropout for training is ineffective in the absence of gated convolution. Figure 13(d) also demonstrates that denoising is ineffective without BNA loss. Consequently, when comparing Fig. 13(e), it can be observed that the use of both gated convolution and BNA loss in training can effectively eliminate most of the noises, while without dropout results in over smoothed images. Finally, Fig. 13(f) presents the denoised result of S2Snet, highlighting the significance of employing gated convolution, BNA loss, and loss simultaneously during training.

For hyperparameters selection, we take a for BNA loss as an example. We assigned different values to α and measured the relevant performance indicators in experiments, as shown in Table 6. Results indicate that when α=0.3, both PSNR and SSIM ranked first, while SNR ranks second. When a further increases, both PSNR and SNR decreases, while SSIM remains unchanged, CNR shows an increasing trend, yet ENL fluctuates. The reason is that by introducing the BNA loss, it is possible to suppress the image background noise as shown in Figs. 13(d) and 13(f). While as a increases, however, it can enhance the learning capability of the background area. Moreover, with the application of dropout, the denoised images become smoother, resulting in decreased PSNR and SNR, yet increased CNR. Since the focus is mainly on the relevant learning of the background area, SSIM of the generated image stays more or less unchanged.

Biomedical Optics EXPRESS

Table 6. Metrics results obtained using loss functions with different values of α. The best value is represented in bold.

	PSNR(dB)	SSIM	SNR	CNR	ENL
$\alpha = 0.1$	22.68	0.50	34.82	3.57	225.84
$\alpha = 0.2$	23.44	0.53	**36.89**	3.95	**386.53**
$\alpha = 0.3$	**24.80**	**0.54**	36.74	3.87	184.16
$\alpha = 0.4$	22.48	0.53	36.58	4.30	216.01
$\alpha = 0.5$	22.07	0.52	36.07	4.30	265.08
$\alpha = 0.6$	21.87	0.53	35.79	4.45	181.75
$\alpha = 0.7$	20.63	0.52	35.20	4.73	214.84
$\alpha = 0.8$	21.29	0.53	34.93	4.68	140.82
$\alpha = 0.9$	19.75	0.51	34.58	5.33	109.58
$\alpha = 1.0$	19.66	0.51	34.15	**4.85**	85.77

4.10. Discussions

With the above experiments, it is found that our proposed S2Snet scheme outperforms those existing self-supervised denoising methods in different cases. Despite its performance superiority, unfortunately, there still exists a limitation. This is because, since only a single noisy image is employed for training, a substantial amount of training time would be required to achieve satisfactory despeckling performances, and thus, a relatively longer processing time would be required for individual images. Such a processing time may not be an issue for those self-supervised methods, e.g., BUnet and MAP-SNR, since after being trained using a considerable number of noisy samples, those methods could remarkably reduce the inference time by using the pre-trained models.

Although the processing time are relatively longer as compared with those self-supervised methods, our S2Snet scheme still has its own merits as it achieves exceptional denoising results with a single noisy image, which largely alleviates the requirement for its practical application as compared with the other schemes. Furthermore, for practical applications, such a limitation can also be overcome by employing higher-performance GPUs or the multi-card training strategies to process multiple batch images simultaneously. Therefore, it is expected that by leveraging these approaches, our S2Snet scheme can be optimized for faster denoising performance for real-world applications.

5. Conclusion

To tackle the problem of single OCT image denoising without requiring clean images for training, we propose a self-supervised deep learning method called S2Snet. Our approach involves feeding the original OCT noisy image and its Bernoulli sampling images into the S2Snet network. In order to address the denoising problem's focus of reducing prediction variance, dropout is employed during both training and testing stages. The final denoised result is obtained by averaging to mitigate prediction variance.

In the network architecture, we incorporate gated convolution within the encoder's block and introduce a background noise attenuation loss to the training loss function. These enhancements significantly contribute to overall denoising performance. Through extensive experiments, we demonstrate that the S2Snet network outperforms other single image denoising methods in terms of OCT image denoising. The final results show that our method is superior to the single image denoising method and the supervised network that relies on a limited number of training datasets. Notably, compared to the original Self2Self network, the dataset experiment results demonstrate

Biomedical Optics EXPRESS

improvements of 3.41% and 2.37% in PSNR and SSIM respectively. This achievement holds significant implications for OCT single image denoising technology.

Funding. National Natural Science Foundation of China (62220106006); the Guangdong Basic and Applied Basic Research Foundation (2021B1515120013); Northwestern Polytechnical University Postgraduate Practice Innovation Fund (PF2023015); the Singapore Ministry of Health's National Medical Research Council under its Open Fund Individual Research Grant (MOH-OFIRG19may-0009); Ministry of Education - Singapore under its Academic Research Fund Tier 1 (RG35/22) and Academic Research Funding Tier 2 (MOE-T2EP30120-0001).

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, *et al.*, "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, *et al.*, "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, *et al.*, "Speckle reduction in oct using massively-parallel detection and frequency-domain ranging," Opt. Express **14**(11), 4736–4745 (2006).
5. K. Hildebrandt and K. Polthier, "Anisotropic filtering of non–linear surface features," in *Computer Graphics Forum*, vol. 23 (Wiley Online Library), pp. 391–400.
6. A. Buades, B. Coll, and J.-M. Morel, "Non-local means denoising," Image Processing On Line **1**, 208–212 (2011).
7. A. Chambolle, "An algorithm for total variation minimization and applications," J. Math. Imaging Vision **20**(1/2), 89–97 (2004).
8. 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).
9. J.-L. Starck, E. J. Candès, and D. L. Donoho, "The curvelet transform for image denoising," IEEE Trans. on Image Process. **11**(6), 670–684 (2002).
10. K. Dabov, A. Foi, V. Katkovnik, *et al.*, "Bm3d image denoising with shape-adaptive principal component analysis," in SPARS'09-Signal Processing with Adaptive Sparse Structured Representations,.
11. K. Dabov, A. Foi, V. Katkovnik, *et al.*, "Image denoising by sparse 3-d transform-domain collaborative filtering," IEEE Trans. on Image Process. **16**(8), 2080–2095 (2007).
12. A. Abbasi, A. Monadjemi, L. Fang, *et al.*, "Optical coherence tomography retinal image reconstruction via nonlocal weighted sparse representation," J. Biomed. Opt. **23**(03), 1 (2018).
13. X. Wang, X. Yu, X. Liu, *et al.*, "A two-step iteration mechanism for speckle reduction in optical coherence tomography," Biomed. Signal Process. Control. **43**, 86–95 (2018).
14. X. Yu, C. Ge, M. Li, *et al.*, "A noise statistical distribution analysis-based two-step filtering mechanism for optical coherence tomography image despeckling," Laser Phys. Lett. **19**(7), 075601 (2022).
15. V. Jain and S. Seung, "Natural image denoising with convolutional networks," Advances in neural information processing systems **21** (2008).
16. K. Zhang, W. Zuo, Y. Chen, *et al.*, "Beyond a gaussian denoiser: Residual learning of deep cnn for image denoising," IEEE Trans. on Image Process. **26**(7), 3142–3155 (2017).
17. Q. Yang, P. Yan, Y. Zhang, *et al.*, "Low-dose ct image denoising using a generative adversarial network with wasserstein distance and perceptual loss," IEEE Trans. Med. Imaging **37**(6), 1348–1357 (2018).
18. P. Isola, J.-Y. Zhu, T. Zhou, *et al.*, "Image-to-image translation with conditional adversarial networks," in *Proceedings of the IEEE conference on computer vision and pattern recognition*, (2017), pp. 1125–1134.
19. A. L. Maas, A. Y. Hannun, A. Y. Ng, *et al.*, "Rectifier nonlinearities improve neural network acoustic models," in *Proc. icml*, vol. 30 (Atlanta, GA, 2013), p. 3.
20. O. Ronneberger, P. Fischer, and T. Brox, "U-net: Convolutional networks for biomedical image segmentation," in *Medical Image Computing and Computer-Assisted Intervention–MICCAI 2015: 18th International Conference, Munich, Germany, October 5-9, 2015, Proceedings, Part III 18*, (Springer, 2015), pp. 234–241.
21. D. Ulyanov, A. Vedaldi, and S. Victor, "Lempitsky. deep image prior," in Computer Vision and Pattern Recognition (CVPR), vol. 1.
22. Q. Zhou, M. Wen, M. Ding, *et al.*, "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).
23. T. Huang, S. Li, X. Jia, *et al.*, "Neighbor2neighbor: Self-supervised denoising from single noisy images," in *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*, (2021), pp. 14781–14790.
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).

Biomedical Optics EXPRESS

25. J. J. Rico-Jimenez, D. Hu, E. M. Tang, *et al.*, "Real-time oct image denoising using a self-fusion neural network," Biomed. Opt. Express **13**(3), 1398–1409 (2022).
26. Y. Huang, W. Xia, Z. Lu, *et al.*, "Noise-powered disentangled representation for unsupervised speckle reduction of optical coherence tomography images," IEEE Trans. Med. Imaging **40**(10), 2600–2614 (2020).
27. X. Yu, C. Ge, M. Li, *et al.*, "Self-supervised blind2unblind deep learning scheme for oct speckle reductions," Biomed. Opt. Express **14**(6), 2773–2795 (2023).
28. Y. Quan, M. Chen, T. Pang, *et al.*, "Self2self with dropout: Learning self-supervised denoising from single image," in *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*, pp. 1890–1898.
29. J. Yu, Z. Lin, J. Yang, *et al.*, "Free-form image inpainting with gated convolution," in *Proceedings of the IEEE/CVF International Conference on Computer Vision*, pp. 4471–4480.
30. G. Liu, F. A. Reda, K. J. Shih, *et al.*, "Image inpainting for irregular holes using partial convolutions," in *Proceedings of the European conference on computer vision (ECCV)*, pp. 85–100.
31. N. Srivastava, G. Hinton, A. Krizhevsky, *et al.*, "Dropout: a simple way to prevent neural networks from overfitting," The journal of machine learning research **15**, 1929–1958 (2014).
32. Y. Gal and Z. Ghahramani, "Dropout as a bayesian approximation: Representing model uncertainty in deep learning," in *international conference on machine learning*, (PMLR, 2016), pp. 1050–1059.
33. S. Farsiu, S. J. Chiu, R. V. O'Connell, and A.-R. E. D. S. . A. S. D. O. C. T. S. Group, *et al.*, "Quantitative classification of eyes with and without intermediate age-related macular degeneration using optical coherence tomography," Ophthalmology **121**(1), 162–172 (2014).
34. L. Fang, S. Li, R. P. McNabb, *et al.*, "Fast acquisition and reconstruction of optical coherence tomography images via sparse representation," IEEE Trans. Med. Imaging **32**(11), 2034–2049 (2013).
35. L. Fang, S. Li, D. Cunefare, *et al.*, "Segmentation based sparse reconstruction of optical coherence tomography images," IEEE Trans. Med. Imaging **36**(2), 407–421 (2016).
36. Y. Ma, H. Hao, J. Xie, *et al.*, "Rose: a retinal oct-angiography vessel segmentation dataset and new model," IEEE Trans. Med. Imaging **40**(3), 928–939 (2020).
37. X. Yu, C. Ge, M. Z. Aziz, *et al.*, "Cgnet-assisted automatic vessel segmentation for optical coherence tomography angiography," J. Biophotonics **15**(10), e202200067 (2022).