



Unsupervised OCT image despeckling with ground-truth- and repeated-scanning-free features

RENXIONG WU,¹ SHAOYAN HUANG,¹ JUNMING ZHONG,¹
FEI ZHENG,¹ MEIXUAN LI,¹ XIN GE,² JIE ZHONG,³ LINBO LIU,⁴ 
GUANGMING NI,^{1,*}  AND YONG LIU¹

¹*School of Optoelectronic Science and Engineering, University of Electronic Science and Technology of China, Chengdu 610054, China*

²*School of Science, Shenzhen Campus of Sun Yat-sen University, Shenzhen 510275, China*

³*Department of Ophthalmology, Sichuan Provincial People's Hospital, University of Electronic Science and Technology of China, Chengdu 610072, China*

⁴*School of Electrical and Electronic Engineering, Nanyang Technological University, Singapore 639798, Singapore*

*guangmingni@uestc.edu.cn

Abstract: Optical coherence tomography (OCT) can resolve biological three-dimensional tissue structures, but it is inevitably plagued by speckle noise that degrades image quality and obscures biological structure. Recently unsupervised deep learning methods are becoming more popular in OCT despeckling but they still have to use unpaired noisy-clean images or paired noisy-noisy images. To address the above problem, we propose what we believe to be a novel unsupervised deep learning method for OCT despeckling, termed Double-free Net, which eliminates the need for ground truth data and repeated scanning by sub-sampling noisy images and synthesizing noisier images. In comparison to existing unsupervised methods, Double-free Net obtains superior denoising performance when trained on datasets comprising retinal and human tissue images without clean images. The efficacy of Double-free Net in denoising holds significant promise for diagnostic applications in retinal pathologies and enhances the accuracy of retinal layer segmentation. Results demonstrate that Double-free Net outperforms state-of-the-art methods and exhibits strong convenience and adaptability across different OCT images.

© 2024 Optica Publishing Group under the terms of the [Optica Open Access Publishing Agreement](#)

1. Introduction

Optical coherence tomography (OCT), a noninvasive biomedical imaging technology, captures three-dimensional images at micrometer resolution and provides cross-section scans for various biological tissues [1]. In clinical imaging of ophthalmology, cardiology, and dermatology [2–4], OCT has become a preferred technology. However, inherent speckle noise severely causes degradation in image quality and affects OCT clinical diagnosis.

Many despeckling methods have been developed and can be categorized into two groups: hardware-based approaches and software-based approaches. Hardware-based methods modify the setup to acquire uncorrelated speckle patterns between B-scans, such as speckle-modulating OCT [5]. However, these methods face limitations due to low imaging sensitivity and speed. On the contrary, frame averaging, a software-based method can't reduce speckle noise well, especially in the non-ophthalmological OCT imaging fields, and it is impractical for clinical applications due to long time consumption and demanding requirement for sample motion caused by repeated scanning. More software-based despeckling methods have been introduced, including digital filters, nonlocal means (NLM) [6], block matching 3D (BM3D) [7], dual-tree complex

wavelet [8], etc. While these methods can partly remove speckle noise, they always degrade the spatial resolution and lose subtle structure details.

Recently, as the advanced solution in the field, deep learning-based despeckling methods provide the advantage of speckle noise reduction and structure preservation [9–11]. A line-shaped CNN-based method, as demonstrated by DeSpecNet [12], has shown good deep learning performance. Building upon the principles of the generative adversarial network (GAN), several GAN-based methods were proposed to improve the despeckling performance [13–16]. Specifically, Halupka et al. [13] built a GAN utilizing residual block and Chen et al. [14] enhanced the generator of GAN by adding context-encoding, resulting in improved image quality. Sm-Net OCT [15] integrated a setup and a GAN trained with a customized speckle-modulating OCT dataset to remove speckle noise and resolve microstructures [16]. In addition, some researchers have put effort into designing structure loss functions to realize speckle noise reduction and detail preservation [17–19]. Although various despeckling approaches have yielded impressive results, they all share a common constraint. They require clean images as references derived from multi-image averaging. The averaging operation involves repeated scanning at the same location, which can be time-consuming and impractical in certain scenarios where there is uncontrolled motion, such as eyeball motion and other in *vivo* imaging.

To resolve this issue, researchers have proposed semi-supervised and unpaired image denoising methods, allowing the model to learn from a limited set of clean images. The former employs a stepwise training process to obtain initial weights and further refines the network using clean images [20]. The latter is built upon cycleGAN, treating OCT image denoising as a domain adaptation challenge between the noise domain and content domain [21–24]. Moreover, the Noise2Noise (N2N) strategy based on a noisy-noisy paired image acquired by repeating twice was applied in OCT despeckling [25–27]. However, these methods still need either clean images or multiple similar B-scan images.

Very recently, self-supervised methods for OCT despeckling have become a popular choice. Nonlocal-GAN [28] was trained with background noisy patches and content patches cropped from bigger OCT B-scans. However, the speckle patterns in background patches are limited. Rico-Jimenez et al. [29] proposed a self-fusion network to denoise OCT images but three frames were required. Neighbor2Neighbor [30] sub-sampled a noisy image into the image pair to achieve uncorrelated noise removal, motivating more researchers to investigate self-supervised denoising methods. Zhou et al. [31] used the paired images after sub-sampling to train a transformer with probabilistic NLM (PNLM) to suppress speckle noise. Li et al. [32] proposed a patch sampler with a larger receptive field, allowing Neighbor2Neighbor to adapt to OCT image denoising and achieve a good denoising performance. However, their methods are not enough to analyze the denoising performance on different OCT setups and more common tissue sample images. Yu et al. [33] proposed B2Unet based on the Blind2Unblind scheme for denoising OCT images of retinas and various tissue samples but it reduces the resolution of tissue sample images.

To address these challenges, here we proposed a novel unsupervised despeckling network, termed Double-free Net. This network achieves state-of-the-art denoising performance using only a single noisy image, without the requirement of repeated scanning and ground truth. In summary, our main technological contributions are as follows: (1) A novel training strategy is proposed. Double-free Net uses only noisy single-image, eliminating the requirement of repeated scanning and image pairing for unsupervised training. (2) Synthesizing the noisier image based on deep learning to overcome the challenge of estimating the speckle mathematical model of OCT images. (3) Testing on two public retinal OCT datasets and two self-made OCT datasets, including a variety of human ex vivo tissue images and large-field retinal images. Extensive experiments demonstrate the convenience and adaptability of Double-free Net across different samples and setups.

2. Methods and materials

2.1. Theory

Here we supposed that the noisy OCT image is y , and its clean image and noise are x and n . In clinical settings, the OCT signal is normally log-compressed to fit in the limited dynamic display range, thus, speckle noise is changed from multiplicative to additive which is data-independent by logarithmic operation [24,34]. The relationship can be expressed as:

$$y = x + n. \quad (1)$$

The objective of OCT despeckling methods is to reconstruct the clean image x . The OCT despeckling model can be defined as:

$$\hat{x} = M(y), \quad (2)$$

where $M(\bullet)$ is the despeckling model and \hat{x} is its estimator.

Assume there are noisy paired images $g_1(y) = x + n_1$ and $g_2(y) = x + n_2$. n_1 and n_2 are independent and identically distributed (i.i.d) noise. We can construct a noisier image $z = x + n_1 + n_2$ and train a model by minimizing $\mathbb{E} \|F_\omega(z) - g_1(y)\|_2^2$, where F_ω is the network parameterized by ω and \mathbb{E} is the expectation operator. According to reference [35], we can conclude that

$$\begin{aligned} 2\mathbb{E}[g_1(y)|z] &= \mathbb{E}[x|z] + \mathbb{E}[n_1|z] + \mathbb{E}[x|z] + \mathbb{E}[n_1|z] \\ &= \mathbb{E}[x|z] + \mathbb{E}[x + n_1 + n_2|z] \\ &= \mathbb{E}[x|z] + z. \end{aligned} \quad (3)$$

Therefore, the estimate of the clean image can be recovered by $\mathbb{E}[x|z] = 2\mathbb{E}[g_1(y)|z] - z$.

However, it is difficult to construct n_2 due to the lack of an accurate mathematical model describing the i.i.d speckle noise. Here, we synthesize n_2 by virtue of the powerful feature extraction capability of deep learning. Inspired by the Neighbor2Neighbor (NBr2NBr) [30], the noisy paired images $g_1(y)$ and $g_2(y)$ can be obtained by sampling the noisy image y with a neighbor sub-sampler. Then, we can train a model by minimizing $\mathbb{E} \|f_\theta(g_1(y)) - g_2(y)\|_2^2$, where f_θ is the network parameterized by θ . Assume that σ^2 denotes the variance of $g_2(y)$ and the gap between the underlying clean images of paired noisy images $g_1(y)$ and $g_2(y)$ is Δ_{12} , there holds:

$$\begin{aligned} \mathbb{E} \|f_\theta(g_1(y)) - x\|_2^2 &= \mathbb{E} \|f_\theta(g_1(y)) - g_2(y)\|_2^2 - \sigma^2 + \\ &2\Delta_{12}\mathbb{E}(f_\theta(g_1(y)) - x), \end{aligned} \quad (4)$$

Minimizing $\mathbb{E} \|f_\theta(g_1(y)) - x\|_2^2$ is equivalent to minimize $\mathbb{E} \|f_\theta(g_1(y)) - g_2(y)\|_2^2$ when the gap Δ_{12} is sufficiently small. Note that σ^2 is a constant value. In other words, NBr2NBr can extract noise from two noisy images.

Considering $g_1(y)$ and $g_2(y)$ are sampled from the same noisy image, we obtained the i.i.d estimation of speckle noise extracted by NBr2NBr:

$$\hat{n}_2 = g_2(y) - g_2(f_\theta(y)). \quad (5)$$

The noisier image can be synthesized as:

$$z = g_2(y) - g_2(f_\theta(y)) + g_1(y). \quad (6)$$

2.2. Training strategy

Fig. 1 summarizes the pipeline of our proposed Double-free Net. Specifically, the strategy is performed in three steps as follows: First, training Network 1 to extract speckle noise estimation and then training Network 2 using the pairs of noisy images and corresponding synthetic noisier images. Finally, noisy images are fed into well-trained Network 2 and corrected to obtain the denoised images.

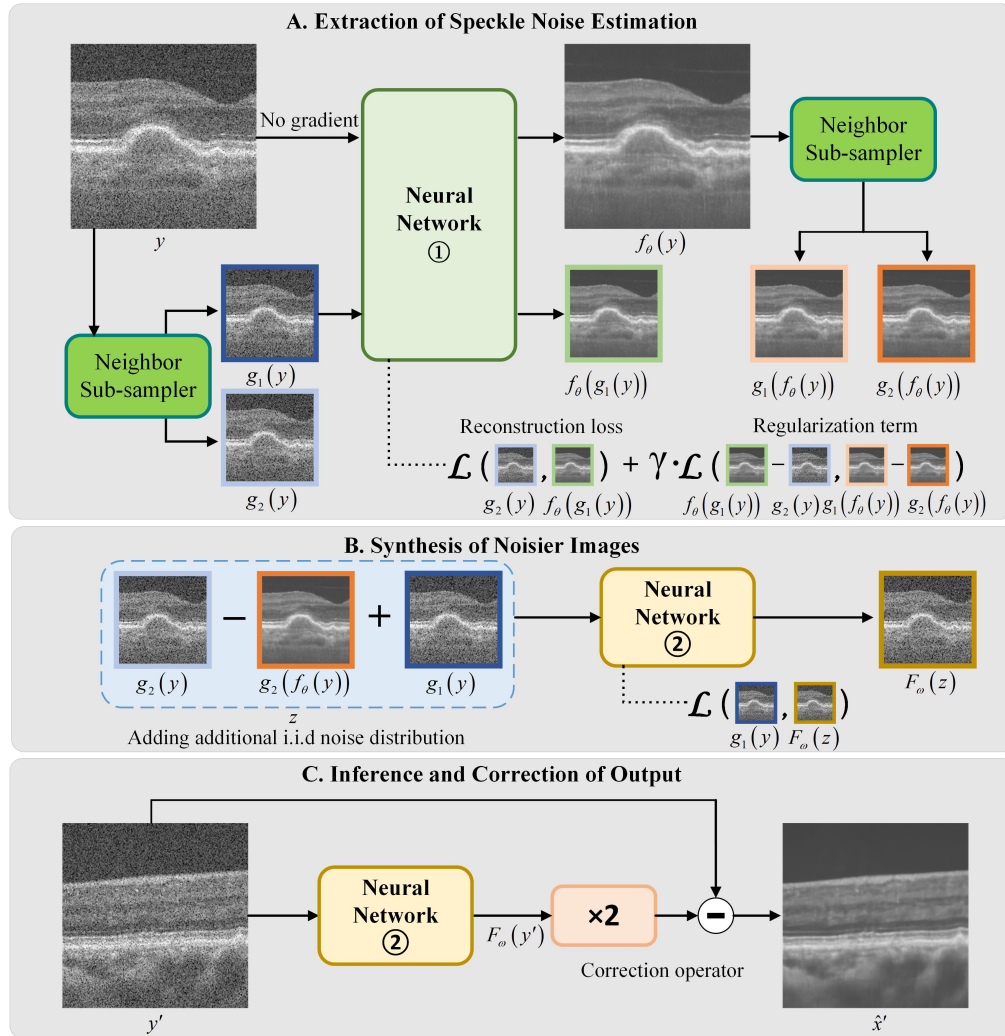


Fig. 1. Schematic of our proposed Double-free Net.

2.2.1. Extraction of speckle noise estimation

According to NBr2NBr, the noisy image y is sampled by the neighbor sub-sampler to two "similar but not identical" images $g_1(y)$ and $g_2(y)$. Network 1 is fed into $g_1(y)$ and output $f_\theta(g_1(y))$. The image y is also input Network 1 but without gradient updating. The output is sampled to $g_1(f_\theta(y))$ and $g_2(f_\theta(y))$ using the same neighbor sub-sampler. The loss function of Network 1 can be written as:

$$\mathcal{L}_A = \|f_\theta(g_1(y)) - g_2(y)\|_2^2 + \gamma \|f_\theta(g_1(y)) - g_2(y) - g_1(f_\theta(y)) + g_2(f_\theta(y))\|_2^2. \quad (7)$$

The second term is the regularization term to optimize the denoising result and γ is the weight to control the strength of the regularization term. The experimental result in Section 3.5 indicates that we can trade off between high metrics value and detail preservation.

2.2.2. Synthesis of noisier images

As Eqs. (5) and (6), we synthesized the noisier image z using sub-sampled images from Network 1. We input the noisier image z into Network 2 and obtained the output $F_\omega(z)$. Network 2 is trained by minimizing the mean square error of the noise image and its noisier image. The loss function is expressed as:

$$\mathcal{L}_B = \|F_\omega(z) - g_1(y)\|_2^2. \quad (8)$$

2.2.3. Inference and correction of output

As Eq. (3), the output of Network 2 is not an estimate of the clean image. We perform a correction operator by doubling the output of Network 2 and subtracting its input during the inference phase. It is theoretically necessary to add extra speckle noise to the noisy image to be predicted, but this would result in an artificially poor view of the noisy image. In this work, we simply predict the noisy image without adding extra noise to obtain better denoising performance. Given the noisy image y' , the denoised image \hat{x}' is obtained by $2F_\omega(y') - y'$

2.3. Network

In this work, we followed our previous work [16] to build a modified U-Net. Both the Network 1 and the Network 2 are implemented with this structure. As shown in Fig. 2, this network is enhanced by replacing the conventional layer with the residual dense block (RDB) and applying RDB at skip-connection between the encoder and decoder. Every RDB consists of three conventional layers followed by Leaky ReLU and a conventional layer as the output layer at the end. The proposed despeckling method can also be implemented by other existing network structures.

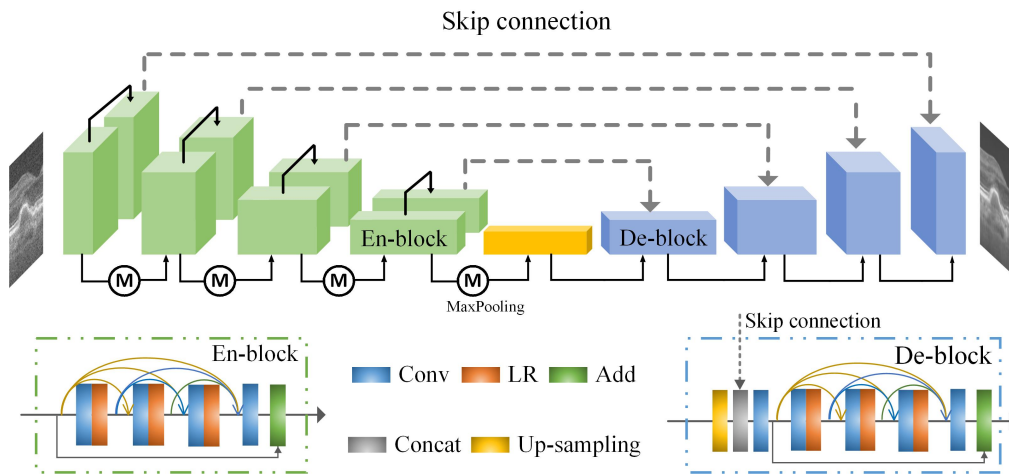


Fig. 2. Structure of the neural network.

2.4. Dataset

In this work, we utilized three retinal OCT image datasets and one high-resolution *ex vivo* OCT image dataset. Table 1 provides detailed information about the four datasets. The first dataset [36] contains 35,400 noisy B-scans from 269 patients with age-related macular degeneration (AMD) and 115 normal subjects. All images with size 1000×512 (width×height) in this dataset were originally used for studying the quantitative classification of eyes with and without intermediate AMD. The data are ground-truth- and repeated-scanning-free. After excluding images of poor

quality, the dataset was divided into training (20,500 images) and testing (13,900 images) sets. Dataset 2 features noisy OCT images captured by high-resolution OCT [37]. This dataset includes various samples, such as polystyrene microparticle calibration samples and ex vivo human stomach, esophagus, and urothelium samples. This human tissue imaging study was approved by the IRB of Nanyang Technological University (IRB-2016-10-015 and IRB-2019-05-050). The total 9852 noisy images with size 1024×512 from 15 samples were divided into a training dataset containing 6897 images and a test dataset containing 2955 images. The third dataset consists of 11,300 wide-field OCT B-scans with size 1789×2345 corresponding to 24×20 mm² from 23 diseased eyes including normal, diabetic macular edema (DME) and AMD. The images in this dataset are acquired using BM-400K BMizar, TowardPi from Sichuan Provincial People's Hospital. This retina imaging study was approved by the Institutional Review Board (IRB) of Sichuan Provincial People's Hospital (IRB-2022-258). The training set comprises 8800 images, while the test set includes 2500 images. Several example images in three datasets are shown in Fig. 3. Dataset 4 [38] is another public dataset that contains paired noisy retina images and clean images acquired by registering and averaging repeated B-scans. For our work, ten pairs from this dataset were exclusively selected for testing purposes. Note that Dataset 1 and Dataset 4 were acquired from the same OCT setup.

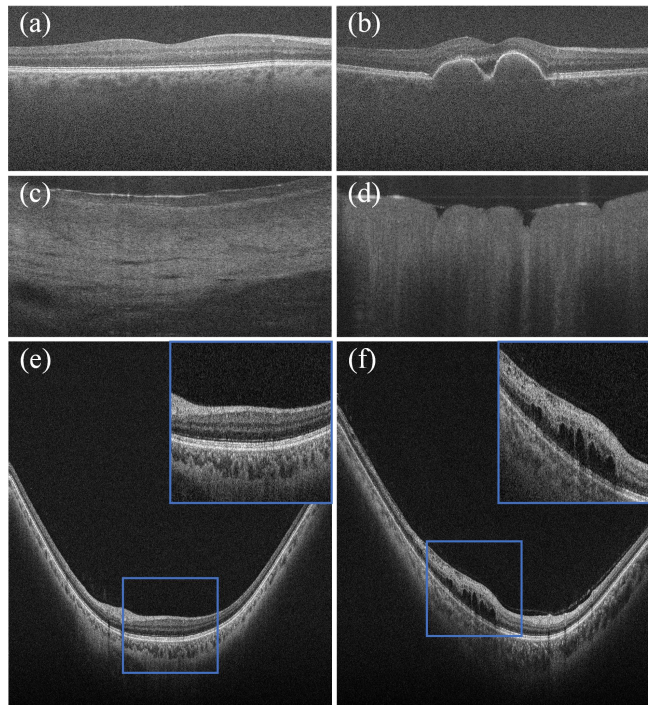


Fig. 3. Example images in datasets. (a) and (b) are images of the normal and AMD in Dataset 1, respectively; (c) and (d) are images of the esophagus and stomach sample in Dataset 2, respectively. (e) and (f) are images of the normal and DME in Dataset 3, respectively.

2.5. Experimental setup

2.5.1. Training detail

We randomly cropped size 512×512 images to be suitable for feeding in the network and performed data augmentation during the training phase. The learning rate of the Adam optimizer

Table 1. Details of the public and customized dataset utilized in this work.

	Dataset 1	Dataset 2	Dataset 3	Dataset 4
Subject	269	15	23	10
Images	24,400	9852	11,300	20
Sample types	normal retina & AMD	microparticle stomach & esophagus & urothelium	normal retina & AMD & DME	normal retina
Repeated-scanning	No	No	No	Yes
Use	training & test	training & test	training & test	only test

in two networks was set as 0.00005. The loss weight γ in Eq. (7) is set to 2 to preserve more detail. We use a batch size of 8 and train Network 1 for 50,000 batches and train Network 2 for 100,000 batches. All networks were trained on an Inter Xeon Sliver 4210R CPU, an NVIDIA Quadro RTX5000 16 GB GPU, and 64 GB of RAM with TensorFlow 2.5.0 or PyTorch 1.8.0 based environment.

2.5.2. Compared methods and quantitative measurement

Here, we compared the proposed method with seven existing methods including two traditional algorithms, BM3D [39], NLM [40], and five deep learning-based methods, Noise2Void (N2V) [41], Blind2Unblind (B2U) [33,42], NBr2NBr [30], DRGAN [21] and MAP-SNR [32]. We have selected the parameters by experiments to achieve the best denoising performance. Deep learning-based methods except DRGAN were trained on our dataset and optimized hyperparameters to obtain their best denoising performance. Due to there being no repeater-scanning and clean images, we didn't train DRGAN but utilized the model weight provided by the authors to compare the performance of the unpaired method when it faces the repeated-free datasets.

Three unsupervised metrics were employed for quantitative performance comparison. Signal-to-noise ratio (SNR) [43] measures the ratio of signal energy to noise energy. Contrast-to-noise (CNR) [43] indicates the contrast between the signal area and the background area. Equivalent number of looks (ENL) [43] measures the smoothness of the homogeneous area in the image. Three signal regions of interest (ROIs) near retinal layers or vascular, and one background ROI located at a homogenous area, were selected for metric calculations. When compared to a clean image, peak-signal-to-noise ratio (PSNR) and structural similarity (SSIM) were used to measure denoising performance and structure preservation performance.

3. Results

3.1. Comparison of denoising methods on the dataset without clean image

Fig. 4 shows a noisy example image from the test data in Dataset 1 and denoised results obtained by different methods. Three regions containing significant retina structures are magnified and presented below each image. Figs. 4(b) and (c) are denoised results of traditional algorithms and Figs. 4(d)-(i) are denoised results of unsupervised deep-learning-based methods. Table 2 shows the quantitative metrics (mean \pm standard deviation) for the different methods on 20 test images from Dataset 1. The results from Fig. 4 and Table 2 highlight the exceptional performance of our proposed Double-free Net in speckle noise suppression and structure detail preservation and enhanced the image to observe retina layer and choroid structure information. As indicated by the yellow arrows, the denoising result of Double-free Net shows the clearest retinal layer boundaries. Double-free Net achieves the highest values in SNR, CNR, and ENL metrics. Besides, MAP-SNR has high metric values over other deep-learning-based methods but falls short of the performance achieved by Double-free Net. BM3D has also high CNR and ENL values but it blurs image

details and introduces block artifacts. N2V and B2U both preserve structure details but have obvious noise residue.

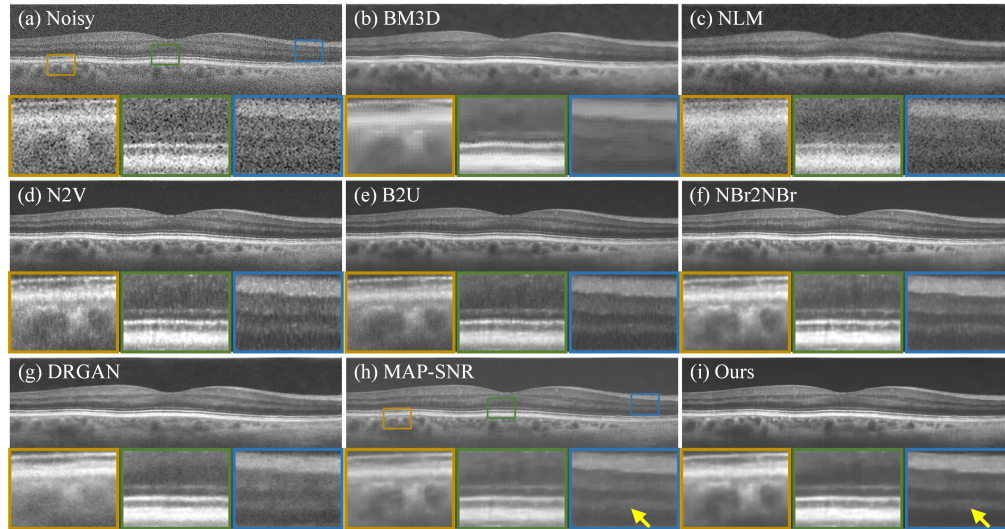


Fig. 4. Denoised results from different methods on Dataset 1. (a) Noisy. (b) BM3D. (c) NLM. (d) N2V. (e) B2U. (f) NBr2NBr. (g) DRGAN. (h) MAP-SNR. (i) Our Double-free Net.

Table 2. Quantitative metrics of different denoising methods on test data of Dataset 1.

	SNR	CNR	ENL
Noisy	17.941±0.294	2.139±0.686	20.147±3.053
BM3D	37.850±3.010	9.769±0.493	448.489±77.829
NLM	23.557±0.250	5.674±0.630	96.336±11.883
N2V	29.279±1.387	6.136±0.388	51.254±10.724
B2U	33.836±0.787	6.893±0.558	131.393±21.600
NBr2NBr	40.156±1.225	8.705±0.620	258.999±48.234
DRGAN	32.513±0.431	8.839±0.704	346.451±73.577
MAP-SNR	43.717±0.787	9.168±0.596	315.696±64.642
Ours	45.559±0.760	9.995±0.768	451.533±99.894

3.2. Comparison of denoising methods on a dataset with clean image

To assess the generalization capabilities of different methods and verify their denoising performance on a dataset with ground truth and repeated scanning, the denoising model trained on Dataset 1 was applied to test Dataset 4. The images in Dataset 4 are not present in Dataset 1. Fig. 5 presents the noisy OCT image and the clean image by averaging multiple images scanned repeatedly, and denoising results from different methods. It can be seen from the green magnified region that Double-free Net effectively removes speckle noise while restoring detailed texture, showing the layer structure of the retina, closely resembling the ground truth. BM3D removes a substantial amount of noise but loses details. Speckle noise in the results of NLM remains significant. N2V and B2U suppress some noise and retain detail, but their denoising performance could be improved. Observing the magnified orange region, Double-free Net strikes a balance

between recovering clean images and preserving details compared to NBr2NBr and MAP-SNR. Table 3 presents quantitative results of denoising methods on Dataset 4. SNR, CNR, ENL, SSIM, and PSNR for Double-free Net achieve the highest scores. These evaluation metrics indicate that Double-free Net has superior generalization ability on datasets with ground truth and repeated scanning. It effectively removes speckle noise, and its denoising results closely approximate ground truth compared to other unsupervised methods.

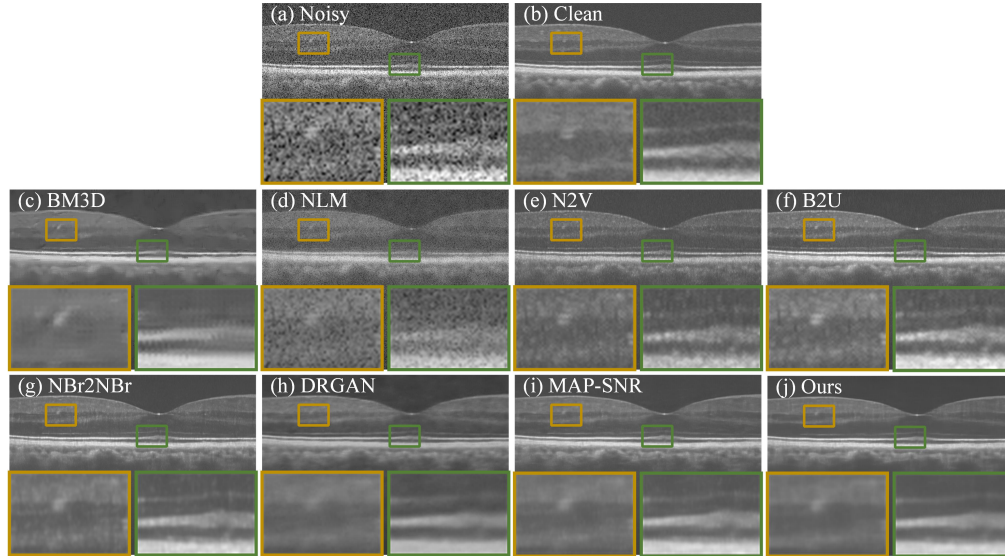


Fig. 5. Denoised results from different methods on Dataset 4 and corresponding averaged clean images. (a) Noisy. (b) Clean. (c) BM3D. (d) NLM. (e) N2V. (f) B2U. (g) NBr2NBr. (h) DRGAN. (i) MAP-SNR. (j) Our Double-free Net.

Table 3. Quantitative metrics of different denoising methods on test data of Dataset 4.

	SNR	CNR	ENL	SSIM	PSNR
Noisy	16.705	1.646	10.322	0.138	18.193
BM3D	32.773	4.489	163.992	0.646	28.967
NLM	21.714	3.249	43.516	0.329	24.477
N2V	33.015	3.521	102.721	0.586	27.465
B2U	32.460	4.147	90.596	0.584	28.602
NBr2NBr	39.640	4.736	142.750	0.688	29.525
DRGAN	33.967	5.413	128.139	0.690	26.743
MAP-SNR	38.798	4.972	190.237	0.695	28.472
Ours	42.512	5.686	254.606	0.697	29.778

3.3. Comparison of denoising methods on datasets of different samples

We performed the proposed Double-free Net on Dataset 2 acquired from a different OCT setup and involving different subjects, to verify its superiority and robustness. Figures 6 and 7 show the noise images and denoising results from different methods for OCT images of human gastric mucosa and esophageal mucosa from Dataset 2. Note that Dataset 2 does not have ground truth

so it is not possible to train DRGAN, only the rest of the methods were compared. Quantitative metrics for noisy images and seven denoised images on 10 images from Dataset 2 are shown in Table 4. BM3D and NLM suppress some noise but blur the vascular boundaries within the gastric mucosa. Double-free Net excels in removing the speckle noise and preserving crucial details. In Fig. 6(h), the small gastric concavity in the magnified orange region and the gastric mucosal blood vessels in the magnified green region are clearly visible. As indicated by the yellow arrow in Figs. 6(g) and (h), the detail of the small gastric concavity appears blurry in the denoised image from MAP-SNR, while Double-free Net retains microstructure information. Additionally, in Figs. 7(g) and (h), the collagen indicated by the yellow arrow is clearly visible after denoising by Double-free Net. Quantitatively, the SNR, CNR, and ENL metrics of Double-free Net are superior to those of other denoising methods. The ENL values of BM3D and NLM are larger than those of other denoising methods, except for Double-free Net, due to the smoother denoising images they produce.

Table 4. Quantitative metrics of different denoising methods on test data of Dataset 2.

	SNR	CNR	ENL
Noisy	19.352±0.251	3.525±0.150	28.539±1.200
BM3D	31.903±1.455	7.076±0.091	110.134±61.340
NLM	25.918±0.375	6.636±0.154	111.121±10.876
N2V	31.711±0.896	5.502±0.136	36.305±1.228
B2U	31.762±5.820	6.087±1.285	66.910±18.342
NBr2NBr	33.711±0.914	6.738±0.174	89.804±9.097
MAP-SNR	32.740±0.892	7.148±0.210	117.917±59.001
Ours	34.925±1.003	7.332±0.173	121.728±23.274

3.4. Analysis of the synthesis and correction operator

To demonstrate the effectiveness of the proposed synthesis and correction operator, we present images of intermediate processes in Fig. 8. Figure 8(a) is the original noisy image and (b) is the synthetic noisier image. The additional noise is derived from the estimated noise distribution extracted by Network 1. (c) is the output of Network 2 and (d) is the result after performing the correction operator (doubling the output and then subtracting the input). It can be seen that the noise level in the output of Network 2 falls between that of the noisy image and the clean image. Thus, the correction operator is necessary and effective.

3.5. Analysis of the loss function weight

The weight γ shown in Eq. (7) was used to control the denoising result of the proposed Double-free Net. Here we conducted a set of experiments to illustrate the impact of different loss weights and gave the visual comparisons and quantitative metrics shown in Fig. 9 and Table 5, respectively. The figure and table demonstrate that with small γ , the denoised result is smoother and has higher ENL but damages some retina structures indicated by the orange box in Fig. 9(b). In contrast, when γ is too large, the denoised result has higher contrast between layer structures but image details are obviously lost. As indicated by the yellow arrows in Fig. 9(c), the external limiting membrane (ELM) and outer plexiform layer (OPL) structures are observed clear and continuous when $\gamma=2$. On the contrary, they are intermittent and indistinguishable when $\gamma=0$ and $\gamma=8$. Therefore, an appropriate value γ is helpful for denoising while preserving details.

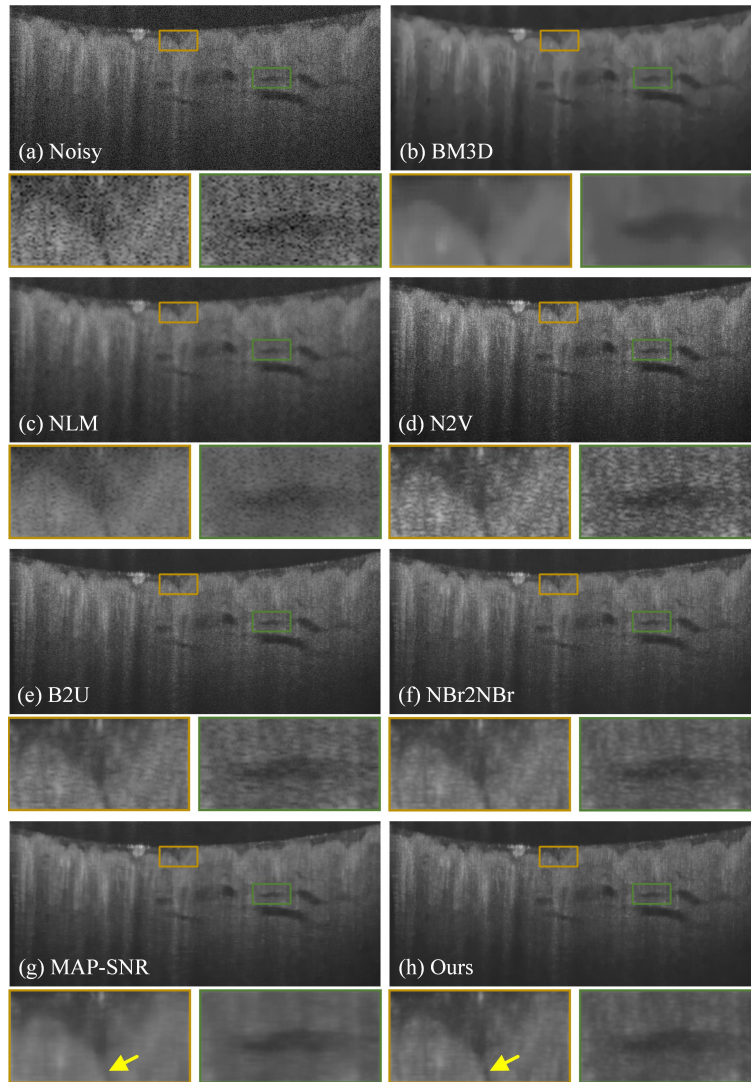


Fig. 6. Test images from representative human gastric mucosa samples on Dataset 2 and corresponding denoised results from different methods. (a) Noisy. (b) BM3D. (c) NLM. (d) N2V. (e) B2U. (f) NBr2NBr. (g) MAP-SNR. (h) Our Double-free Net.

Table 5. Quantitative metrics of different loss weight.

	SNR	CNR	ENL
Noisy	17.264±1.016	1.385±1.327	19.172±5.047
$\gamma=0$	44.944±2.323	8.892±0.277	438.642±113.599
$\gamma=2$	47.181±2.588	8.905±0.359	406.208±136.840
$\gamma=8$	45.317±3.292	8.947±0.282	398.771±113.212

3.6. Application in retinal pathologies

To verify the effectiveness of Double-free Net in facilitating the diagnosis and analysis of retinal pathologies, we tested OCT images comprising drusen or macular edema. These test images

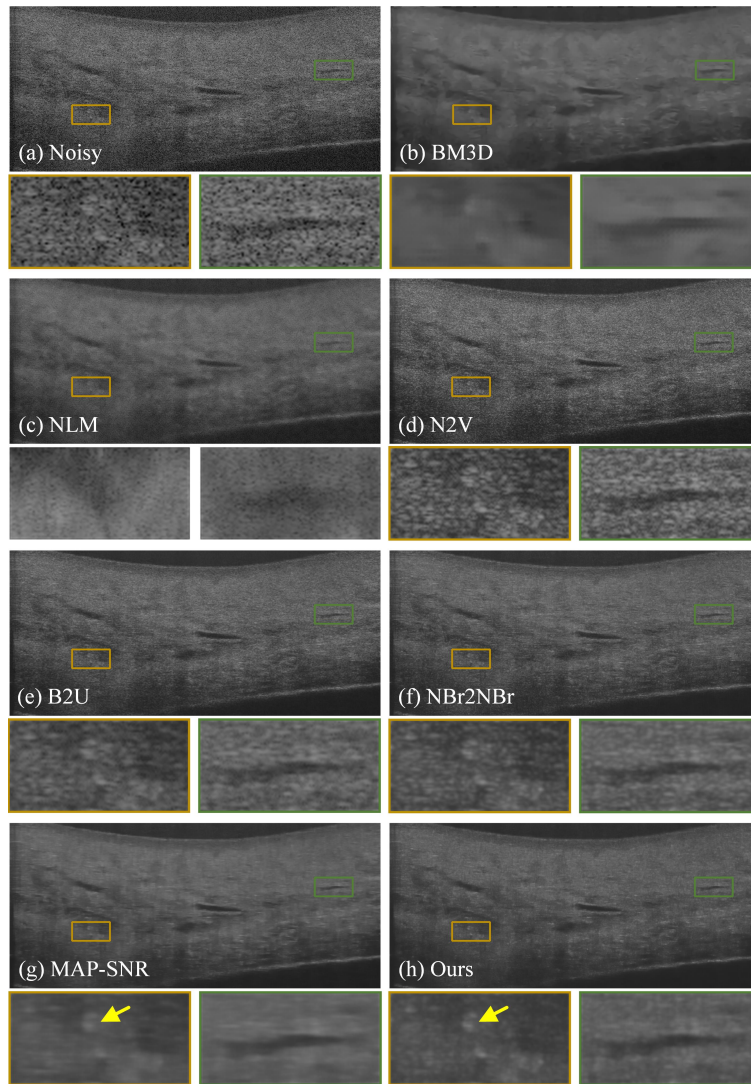


Fig. 7. Test images from representative human esophageal mucosa samples on Dataset 2 and corresponding denoised results from different methods. (a) Noisy. (b) BM3D. (c) NLM. (d) N2V. (e) B2U. (f) NBr2NBr. (g) MAP-SNR. (h) Our Double-free Net.

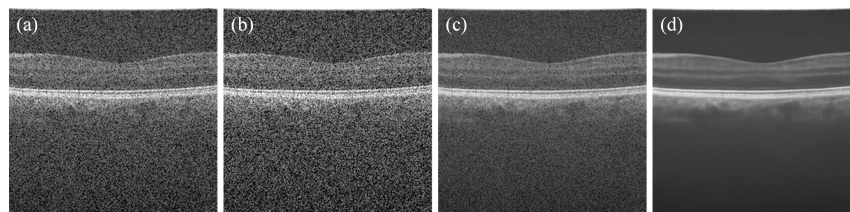


Fig. 8. Results of intermediate processes. (a) is the original noisy image; (b) is the synthetic noisier image; (c) is the output of Network 2; (d) is the denoised image after the correction operator.

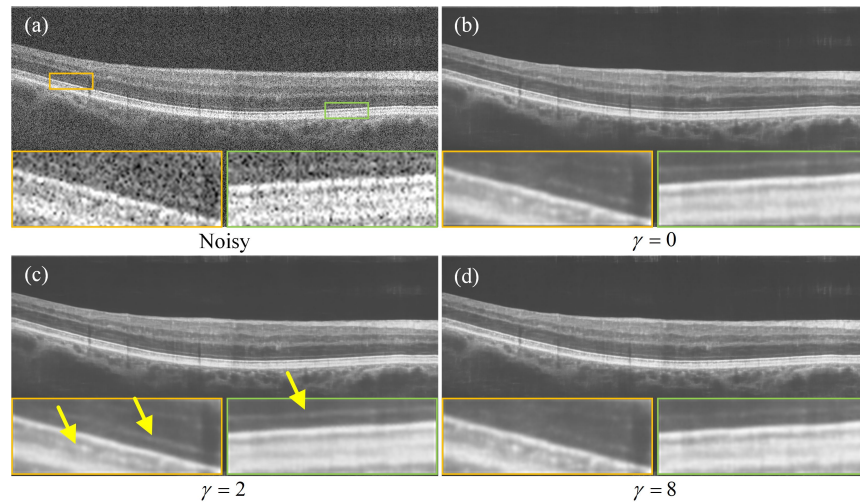


Fig. 9. The denoised results of Double-free Net with different loss weight γ .

are not in any of the training datasets. The original noisy images and corresponding denoised images are shown in Fig. 10. Figure 10(a) is an OCT image from Dataset 1 classified as AMD with edema. After denoising, the small hyperreflective foci near the edema are revealed in Fig. 10(d). These foci serve as biomarkers crucial for early decision-making in retinal diseases, often obscured by speckle noise. Figs. 10(b) and (c) are from Dataset 3. As shown in Figs. 10(e) and (f), Double-free Net removes a large amount of speckle noise, making the edema area more discernible. In particular, the denoising results provide clear boundaries for the edema, facilitating a rapid assessment of its quantity and shape, which is of practical value for clinical diagnosis. Despite the different OCT setups and noise levels in Fig. 10(a) and Figs. 10(b), (c), Double-free Net effectively removes speckle noise and reveals the edge and shape of cystic lesions.

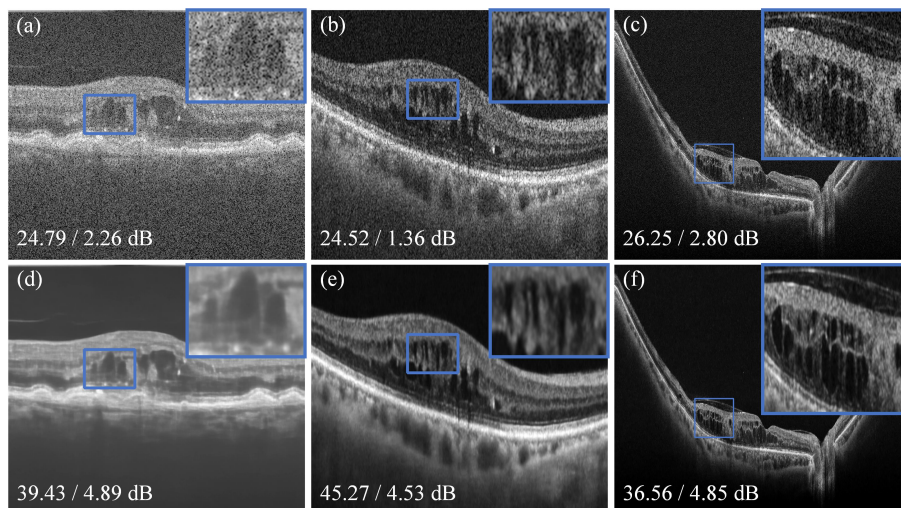


Fig. 10. Denoised results of retinal pathology images using Double-free Net. (a) is a noisy image in Dataset 1, (b) and (c) are noisy images in Dataset 3. (d), (e), and (f) are corresponding denoised images. The evaluation scores at the bottom are SNR and CNR respectively.

3.7. Application in retinal layer segmentation

To further demonstrate the effectiveness of the proposed Double-free Net, we compared the effects of different denoised methods as a pre-processing step for retinal segmentation. Different deep-learning-based denoising methods trained on Dataset 1 and two traditional algorithms were used to denoise noisy images from the public layer segmentation dataset (GOALS) [44]. The dataset, following the competition setup, consists of 100 images each in the training set and test set, with annotations for three layers: the retinal nerve fiber layer (RNFL), ganglion cell-inner plexiform layer (GCIPL), and choroid layer. Since the annotations for the validation set and test set are unavailable, we divided the 100 annotated images of the training set into training and testing in a 7:3 ratio. Nine datasets were collected, where the first dataset comprised original noisy images, and the remaining eight datasets comprised denoised images from eight different denoising methods. The ground truth images for all nine datasets were identical. Nine segmentation models (built with U-Net) were trained individually on the nine datasets to test layer segmentation performance. Fig. 11 shows the visual segmentation results with the noisy image and denoised images. We can see that the proposed Double-free Net as the pre-processing step has the best segmentation performance than other denoised methods. Two quantitative evaluation metrics were calculated, including the Dice coefficient and precision. Table 6 presents the quantitative results for three layers. The segmentation results obtained by Double-free Net exhibit the highest scores of Dice on three layers and the highest scores of precision, except for the GCIPL layer.

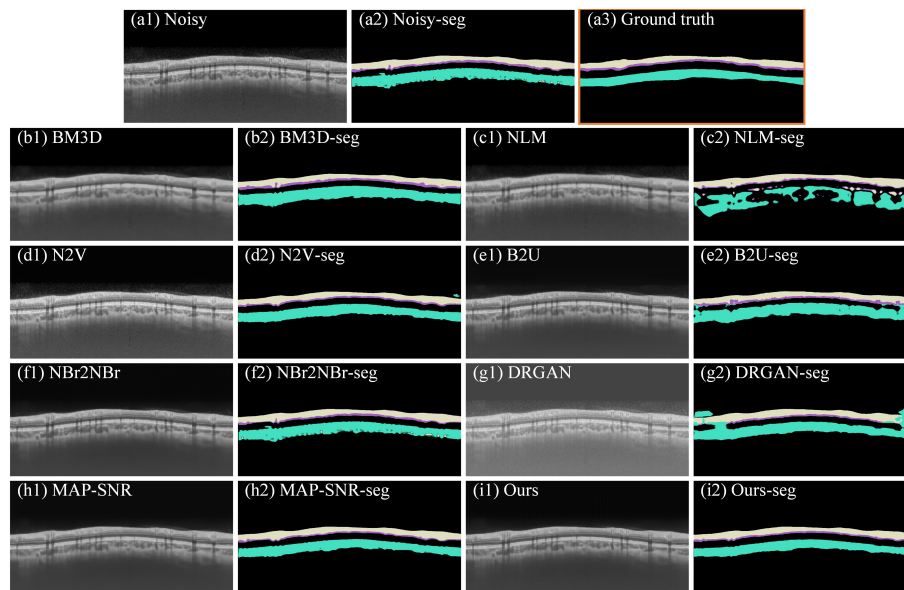


Fig. 11. Segmentation results after denoising by different methods. (a1) - (i1) Noisy OCT image and denoised images by BM3D, NLM, N2V, B2U, NBr2NBr, DRGAN, MAP-SNR, and Our Double-free Net. (a2) - (i2) Segmentation results of corresponding OCT images. (a3) Segmentation ground truth. Three layers from top to bottom are RNFL, GCIPL, and choroid layer.

Table 6. Quantitative metrics of the retinal layer segmentation after denoising by different methods.

	Dice			Precision		
	RNFL	GCIPL	choroid	RNFL	GCIPL	choroid
Noisy	55.2±9.2	81.2±8.2	85.5±6.1	55.8±9.4	74.4±8.0	82.5±6.5
BM3D	61.1±9.0	79.8±5.7	84.4±5.3	60.6±9.1	69.1±7.2	82.2±7.1
NLM	51.2±8.1	82.0±7.4	82.9±6.7	50.1±6.7	76.4±7.2	80.0±7.2
N2V	61.0±11.6	82.5±7.9	84.6±6.8	63.7±13.9	79.3±7.1	83.4±8.6
B2U	59.9±14.3	75.3±6.6	85.2±6.2	62.6±16.5	64.3±7.1	80.5±6.1
NBr2NBr	60.5±9.8	78.4±5.8	81.8±6.3	61.5±10.2	67.7±7.3	81.9±11.6
DRGAN	40.1±11.8	76.8±9.4	75.8±10.5	55.1±12.8	70.0±12.2	74.7±7.8
MAP-SNR	52.5±11.0	82.4±8.5	82.2±9.4	63.6±9.3	77.1±8.7	79.2±6.3
Ours	61.3±13.1	82.6±6.5	85.6±6.2	68.9±14.1	77.3±16.8	83.9±7.7

4. Discussions

In this paper, we demonstrate a novel approach for denoising OCT images without ground-truth images and repeated scanning using unsupervised deep learning. Our method involves two networks in three steps to effectively remove speckle noise in OCT images. The first step is to use the Neighbor2Neighbor strategy and paired noise images to train Network 1 and pre-extract the noise estimate of the OCT image. In the second step, the noise estimates extracted from Network 1, treated as known noise, are added to the noisy image to construct a noisier version. Network 2 is then trained using the Noisier2Noise strategy. The third step is to input the OCT image to be denoised into the well-trained Network 2 and use a correction operation to obtain the denoised OCT image. Neighbor2Neighbor removes part of the speckle noise, but there will be large-grained speckle noise residue. Experimental results, as demonstrated in Figs. 4, 5, and 6, validate this observation. Noisier2Noise requires the use of a known noise model to construct a noisier image during training. However, direct denoising of the OCT image using Noisier2Noise is unfeasible due to the unknown OCT speckle noise model. To address this, noise estimates are pre-extracted utilizing Neighbor2Neighbor to enable training Noisier2Noise. This hybrid strategy combines the strengths of both approaches, efficiently removing speckle noise while preserving detailed structure.

Compared to state-of-the-art deep learning methods, Double-free Net achieves high-quality denoising visually while retaining detailed structural information. In addition, Double-free Net shows efficient speckle noise reduction performance on both normal and pathological retinal images and different biological tissue images collected by different OCT setups and scanning modes. The quantitative evaluation metrics, including SNR, CNR, and ENL, show that Double-free Net attains the highest score on all datasets in this study. Overall, the proposed Double-free Net achieves excellent speckle reduction performance without clean images and repeated scanning, which is comparable to clean images.

OCT image denoising is usually used as a preprocessing step for image recognition and segmentation, and it also provides convenience for doctors to make clinical diagnoses. Following denoising by Double-free Net, retinal images comprising drusen or macular edema reveal clear edges and shapes of cysts that were initially obscured by speckle noise. This enhancement proves valuable in elevating the accuracy of clinical diagnoses. We trained and tested retinal layer segmentation after processing retinal images with different denoising methods. The comprehensive segmentation results show that Double-free Net yields the highest accuracy in segmenting the three retinal layers.

The proposed denoising method is based on the assumption that speckle noise transforms into an additive and data-independent after logarithmic transformation. In practice, speckle noise may not be completely data-independent due to different OCT system hardware and scanning modes. Consequently, training the Noisier2Noise network by adding the extracted noise estimate to the noisy image might have a small amount of structural residues. However, owing to the robustness of deep learning networks and the diversity of training datasets from different OCT setups and scanning samples, these residues are barely visible from the experimental results. In future research, we aim to consider speckle noise as data-dependent and explore novel methods to construct noisier images for tackling OCT image denoising with strong structural correlation. Furthermore, the pre-extraction method of noise estimation, currently exemplified by Neighbor2Neighbor, is not constrained, and we plan to explore other advanced methods such as Blind2Unblind in our ongoing research.

5. Conclusion

In this work, a novel unsupervised learning strategy, named Double-free Net, designed for despeckling OCT images without the need for repeated scanning or ground truth. We compared the denoising performance of Double-free Net with seven state-of-the-art denoising methods and conducted experiments on two public datasets and two self-made datasets. Experimental results demonstrate that Double-free Net excels in removing speckle noise, visually preserving image details, and exhibiting strong adaptability to different sample images collected by different OCT setups. The test result on the dataset with repeated scanning and ground truth indicates that Double-free Net achieves denoising results closest to clean images.

Funding. National Natural Science Foundation of China (61905036); China Postdoctoral Science Foundation (2019M663465, 2021T140090); Medico-Engineering Cooperation Funds from University of Electronic Science and Technology of China (ZYGX2021YGXC019); Fundamental Research Funds for the Central Universities (ZYGX2021J012); Ministry of Education Singapore under its Academic Research Fund Tier 2 (MOE-T2EP30120-0001); Ministry of Education Singapore under its Academic Research Fund Tier 1 (RG35/22); Singapore Ministry of Health's National Medical Research Council under its Open Fund Individual Research Grant (MOH-OFIRG19may-0009); Key Research and Development Project of Health Commission of Sichuan Province (ZH2024-201).

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. T. Klein, W. Wieser, L. Reznicek, *et al.*, "Multi-mhz retinal OCT," *Biomed. Opt. Express* **4**(10), 1890 (2013).
3. T. Gambichler, G. Moussa, M. Sand, *et al.*, "Applications of optical coherence tomography in dermatology," *J. Dermatol. Sci.* **40**(2), 85–94 (2005).
4. M. Paulo, J. Sandoval, V. Lennie, *et al.*, "Combined use of OCT and IVUS in spontaneous coronary artery dissection," *JACC: Cardiovascular Imaging* **6**(7), 830–832 (2013).
5. O. Liba, M. D. Lew, E. D. SoRelle, *et al.*, "Speckle-modulating optical coherence tomography in living mice and humans," *Nat. Commun.* **8**(1), 15845 (2017).
6. J. Aum, J. Kim, and J. Jeong, "Effective speckle noise suppression in optical coherence tomography images using nonlocal means denoising filter with double gaussian anisotropic kernels," *Appl. Opt.* **54**(13), D43–D50 (2015).
7. B. Chong and Y. Zhu, "Speckle reduction in optical coherence tomography images of human finger skin by wavelet modified BM3D filter," *Opt. Commun.* **291**, 461–469 (2013).
8. L. Fang, S. Li, Q. Nie, *et al.*, "Sparsity based denoising of spectral domain optical coherence tomography images," *Biomed. Opt. Express* **3**(5), 927–942 (2012).
9. H. Cheong, S. Krishna Devalla, T. Chuangsuwanich, *et al.*, "OCT-GAN: single step shadow and noise removal from optical coherence tomography images of the human optic nerve head," *Biomed. Opt. Express* **12**(3), 1482–1498 (2021).
10. Y. Zhou, K. Yu, M. Wang, *et al.*, "Speckle noise reduction for OCT images based on image style transfer and conditional GAN," *IEEE J. Biomed. Health Inform.* **26**(1), 139–150 (2021).
11. Z. Dong, G. Liu, G. Ni, *et al.*, "Optical coherence tomography image denoising using a generative adversarial network with speckle modulation," *J. Biophotonics* **13**(4), e201960135 (2020).

12. F. Shi, N. Cai, Y. Gu, *et al.*, “DeSpecNet: a CNN-based method for speckle reduction in retinal optical coherence tomography images,” *Phys. Med. Biol.* **64**(17), 175010 (2019).
13. K. J. Halupka, B. J. Antony, M. H. Lee, *et al.*, “Retinal optical coherence tomography image enhancement via deep learning,” *Biomed. Opt. Express* **9**(12), 6205–6221 (2018).
14. Z. L. Chen, Z. Y. Zeng, H. L. Shen, *et al.*, “DN-GAN: Denoising generative adversarial networks for speckle noise reduction in optical coherence tomography images,” *Biomed. Signal Process. Control* **55**, 101632 (2020).
15. G. Ni, Y. Chen, R. Wu, *et al.*, “Sm-Net OCT: a deep-learning-based speckle-modulating optical coherence tomography,” *Opt. Express* **29**(16), 25511–25523 (2021).
16. G. M. Ni, R. X. Wu, J. M. Zhong, *et al.*, “Hybrid-structure network and network comparative study for deep-learning-based speckle-modulating optical coherence tomography,” *Opt. Express* **30**(11), 18919–18938 (2022).
17. B. Qiu, Z. Huang, X. Liu, *et al.*, “Noise reduction in optical coherence tomography images using a deep neural network with perceptually-sensitive loss function,” *Biomed. Opt. Express* **11**(2), 817–830 (2020).
18. Y. Ma, X. Chen, W. Zhu, *et al.*, “Speckle noise reduction in optical coherence tomography images based on edge-sensitive cGAN,” *Biomed. Opt. Express* **9**(11), 5129–5146 (2018).
19. M. Mehdizadeh, C. MacNish, D. Xiao, *et al.*, “Deep feature loss to denoise OCT images using deep neural networks,” *J. Biomed. Opt.* **26**(04), 046003 (2021).
20. M. Wang, W. Zhu, K. Yu, *et al.*, “Semi-supervised capsule cGAN for speckle noise reduction in retinal OCT images,” *IEEE Trans. Med. Imaging* **40**(4), 1168–1183 (2021).
21. 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 (2021).
22. V. Das, S. Dandapat, and P. K. Bora, “Unsupervised super-resolution of OCT images using generative adversarial network for improved age-related macular degeneration diagnosis,” *IEEE Sens. J.* **20**(15), 8746–8756 (2020).
23. Z. Fu, X. Yu, C. Ge, *et al.*, “ADGAN: An asymmetric despeckling generative adversarial network for unpaired OCT image speckle noise reduction,” in *2021 IEEE 6th Optoelectronics Global Conference (OGC)*, (2021), pp. 212–216.
24. M. Geng, X. Meng, L. Zhu, *et al.*, “Triplet cross-fusion learning for unpaired image denoising in optical coherence tomography,” *IEEE Trans. Med. Imaging* **41**(11), 3357–3372 (2022).
25. B. Qiu, S. Zeng, X. X. Meng, *et al.*, “Comparative study of deep neural networks with unsupervised Noise2Noise strategy for noise reduction of optical coherence tomography images,” *J. Biophotonics* **14**(11), e202100151 (2021).
26. B. Qiu, Y. You, Z. Huang, *et al.*, “N2NSR-OCT: Simultaneous denoising and super-resolution in optical coherence tomography images using semisupervised deep learning,” *J. Biophotonics* **14**(1), e202000282 (2020).
27. Y. Huang, N. Zhang, and Q. Hao, “Real-time noise reduction based on ground truth free deep learning for optical coherence tomography,” *Biomed. Opt. Express* **12**(4), 2027–2040 (2021).
28. K. Liang, X. Liu, S. Chen, *et al.*, “Resolution enhancement and realistic speckle recovery with generative adversarial modeling of micro-optical coherence tomography,” *Biomed. Opt. Express* **11**(12), 7236–7252 (2020).
29. 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).
30. T. Huang, S. J. Li, X. Jia, *et al.*, “Neighbor2Neighbor: A self-supervised framework for deep image denoising,” *IEEE Transactions on Image Processing* **31**, 4023–4038 (2022).
31. 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).
32. 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).
33. 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).
34. H. M. Salinas and D. C. Fernandez, “Comparison of pde-based nonlinear diffusion approaches for image enhancement and denoising in optical coherence tomography,” *IEEE Trans. Med. Imaging* **26**(6), 761–771 (2007).
35. N. Moran, D. Schmidt, Y. Zhong, *et al.*, “Noisier2noise: Learning to denoise from unpaired noisy data,” in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit.*, (2020), pp. 12061–12069.
36. S. Farsiu, S. J. Chiu, R. V. O’Connell, *et al.*, “Quantitative classification of eyes with and without intermediate age-related macular degeneration using optical coherence tomography,” *Ophthalmology* **121**(1), 162–172 (2014).
37. E. Bo, X. Ge, Y. Luo, *et al.*, “Cellular-resolution in vivo tomography in turbid tissue through digital aberration correction,” *Photonix* **1**(1), 9–12 (2020).
38. 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).
39. K. Dabov, A. Foi, V. Katkovnik, *et al.*, “Image denoising by sparse 3-D transform-domain collaborative filtering,” *IEEE Transactions on Image Processing* **16**(8), 2080–2095 (2007).
40. A. Buades, B. Coll, and J.-M. Morel, “Nonlocal image and movie denoising,” *Int. J. Comput. Vis.* **76**(2), 123–139 (2008).
41. A. Krull, T.-O. Buchholz, and F. Jug, “Noise2void-learning denoising from single noisy images,” in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit.*, (2019), pp. 2129–2137.
42. Z. Wang, J. Liu, G. Li, *et al.*, “Blind2unblind: Self-supervised image denoising with visible blind spots,” in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit.*, (2022), pp. 2017–2026.

43. A. Pizurica, L. Jovanov, B. Huysmans, *et al.*, “Multiresolution denoising for optical coherence tomography: a review and evaluation,” *Curr. Med. Imaging* **4**(4), 270–284 (2008).
44. H. Fang, F. Li, H. Fu, *et al.*, “Dataset and evaluation algorithm design for goals challenge,” in *International Workshop on Ophthalmic Medical Image Analysis*, (Springer, 2022), pp. 135–142.