



Simultaneous denoising and super-resolution of optical coherence tomography images based on a generative adversarial network

YONGQIANG HUANG,¹ ZEXIN LU,¹ ZHIMIN SHAO,¹ MAOSONG RAN,¹ JILIU ZHOU,¹ LEYUAN FANG,² AND YI ZHANG^{1,*}

¹*College of Computer Science & College of Software Engineering, Sichuan University, Chengdu 610065, China*

²*College of Electrical and Information Engineering, Hunan University, Changsha, 410082, China*

**yizhang.scu@outlook.com*

Abstract: Optical coherence tomography (OCT) has become a very promising diagnostic method in clinical practice, especially for ophthalmic diseases. However, speckle noise and low sampling rates have intensively reduced the quality of OCT images, which prevents the development of OCT-assisted diagnosis. Therefore, we propose a generative adversarial network-based approach (named SDSR-OCT) to simultaneously denoise and super-resolve OCT images. Moreover, we trained three different super-resolution models with different upscale factors ($2 \times$, $4 \times$ and $8 \times$) to adapt to the corresponding downsampling rates. We also quantitatively and qualitatively compared our proposed method with some well-known algorithms. The experimental results show that our approach can effectively suppress speckle noise and can super-resolve OCT images at different scales.

© 2019 Optical Society of America under the terms of the [OSA Open Access Publishing Agreement](#)

1. Introduction

Optical coherence tomography (OCT) has become a widely used medical diagnosis and treatment technique due to its noninvasive nature, imaging depth capacity and cross-sectional view of tissue structures [1–3], especially in retinal diagnostics [4]. However, two main issues prevent the development of the OCT-based diagnosis. First, OCT images are inevitably corrupted by heavy speckle noise because of the low coherence interferometry imaging modality [5], which intensively degrades the quality of the OCT images and the accuracy of the diagnosis of vision-related diseases [6]. Although many OCT image denoising methods exist, in commercial scanners, noiseless OCT images are most commonly obtained by registering and averaging several B-scan OCT images that are repeatedly acquired at the same position [7]. The second problem comes along with this denoising approach. Due to unconscious body jitter or eye movement during a B-scan, the obtained OCT images that will be used for averaging might not be captured from exactly the same place; consequently, the registration is particularly challenging, some motion artifacts may appear and some key information in the averaged OCT images may be lost [8,9]. Therefore, in a clinic, a low sampling rate is often used to both accelerate the acquisition process and reduce the influence of unconscious movements [10]. To recover high-quality OCT images, it is necessary to propose some effective methods that work well on both OCT image denoising and super-resolution. In this paper, we propose a simultaneous denoising and super-resolution method for OCT image reconstruction. To our knowledge, our method is the first to apply deep learning to OCT image super-resolution.

Several methods have been proposed to address these issues over the last two decades. OCT image denoising methods are mainly divided into hardware-based approaches and software-based approaches [11]. Hardware methods can increase the denoising quality by improving the structure and light source of the imaging system, including frequency

compounding [2,12] and spatial compounding [13–16], which can reduce some kinds of noise, such as the noise of the detector and scanner, but it hardly eliminates speckle or white noise in the system. Software-based methods are digital filters that depend on the local or overall statistics of the image and the model of the speckle noise [17]. These methods include local statistics-based filtering methods [18–20], which are time efficient but do not satisfactorily preserve the details. Diffusion-based methods [21–24] achieve well-denoised results but tend to be over-smooth. Wavelet-based methods [25,26] achieve better performance but also introduce some artifacts and sparse representation and low-rank decomposition-based methods [27–29] use local image patches to denoise OCT images; however, the vectorized patch destroys the structure of the reconstructed OCT image.

OCT image super-resolution has been considered an efficient way to alleviate the low sampling rate problem of OCT images caused by inevitably body jitter and eye movement [8]. However, traditional interpolation approaches may not achieve satisfactory high-resolution (HR) results when the input low-resolution (LR) OCT image was speckled, and it was hard to reconstruct the missed features by the noisy input itself [30]. Therefore, considerable efforts have been made to find a reliable way to reconstruct a noise and LR (denoted as NLR) OCT image to the corresponding clear and HR (denoted as CHR) images. Fang et al. introduced the sparsity-based simultaneous denoising and interpolation (SBSDI) [31] and segmentation-based sparse reconstruction (SSR) [32] methods for the successive reconstruction of retinal OCT images. All studies were based on a sparse shared representation, which needed sophisticated regularizers and thus caused high computational complexity [30] and inflexibility. A more plausible approach was to learn a nonlocal sparse model for every subspace [33–35]. In reality, this method was not suitable for OCT images because usually the LR and HR OCT images were not exactly registered. To solve the above problems, Abbasi et al. proposed the nonlocal weighted sparse representation (NWSR) [8] to reconstruct OCT images. Although this method has achieved good results both on denoising and $2 \times$ interpolation, the recover quality was limited on large-scale super-resolution imagery, such as $4 \times$ and $8 \times$.

Deep learning has played a dominant role in the field of natural image processing over the past few years and has demonstrated the great power to address both high-level and low-level vision tasks, such as image classification [36], image recognition [37], image denoising [38] and image super-resolution [39]. Recently, deep learning-based approaches [40–43] have also been extended to medical image processing and have greatly promoted the development of this field. Several studies have reported OCT image despeckling [17,44,45], enhancement [7,46] and segmentation [47] for OCT images. These methods can provide a promising denoising result but cannot simultaneously tackle interpolation problems. Because the original captured OCT images are intensively speckled and there are no clean and HR references, the super-resolution of OCT images remains challenging.

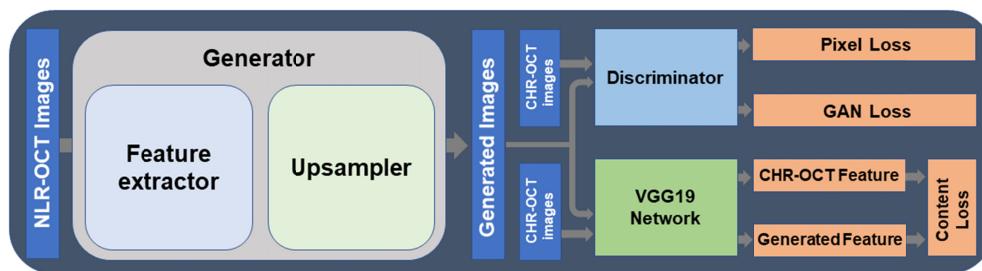


Fig. 1. The general layout of our proposed network.

In this paper, we proposed a generative adversarial network-based approach (named SDSR-OCT) to simultaneously denoise and super-resolve OCT images. To the best of our

knowledge, there is no deep learning-based method for OCT image super-resolution. Figure 1 shows the general layout of our proposed network. Because the denoising and super-resolution of an image is not a global problem, the original 450×900 spectral-domain (SD)-OCT image is cropped into a large set of paired training patches. Each pair contains an NLR patch and its corresponding CHR patch. Afterwards, the NLR patches are fed to the feature extractor to obtain their LR information, and then the upsample is used to upscale the extracted features to the desired spatial resolution to generate the denoised and super-resolved outputs. At the same time, the discriminator of our generative adversarial network (GAN) model takes the generated image and the CHR image as input and makes a judgment on whether the CHR image is more realistic than the generated image. When the model is well-trained, we can achieve the denoised and super-resolved results of the input OCT images.

The remainder of the paper is organized as follows. In the next section, we report the proposed SDSR-OCT method in detail. In section 3, we describe the experimental configurations and results. Finally, the conclusion and future studies are given in section 4.

2. Proposed SDSR-OCT method

2.1 Generative adversarial networks

GANs have been widely used in many image processing applications, such as image denoising and image super-resolution. In the denoising GAN architecture, assuming that x denotes the noise image and y denotes the corresponding clear image, while for the super-resolution GAN model, input x is the LR image and y represents the HR image. The aim of the GAN model is to seek a mapping G of input x and its label y :

$$G : x \rightarrow y \quad (1)$$

Then, a discriminator D is used to judge whether the generated image $G(x)$ is more realistic than the label y . The entire procedure of the GAN model can be described as the following minmax optimization to optimize the generator G and the discriminator D :

$$\min_G \max_D L(G, D) = E_{y \sim P_r} [\log(D(y))] + E_{x \sim P_x} [\log(1 - D(G(x)))] \quad (2)$$

where $E[\cdot]$ represents the expectation operator, P_r and P_x denote the real data distribution and the noisy or low-resolution data distribution, respectively.

Different from other GAN-based super-resolution models that take LR images as inputs and output the HR images, the inputs of the OCT image super-resolution model not only are LR but also are corrupted by the speckle noise. Thus, we propose a GAN-based approach that provides a large capacity for OCT image reconstruction. The general layout of our proposed networks is shown in Fig. 1. Unlike the original GAN [48], which learns a map from the random noise to the output image, the end of our model is to produce an CHR image with its NLR observation. More concretely, the structure of our model can be divided into two parts with opposite goals: the generator G extracts low-resolution features from the input NLR-OCT image x , and then the extracted features are used to reconstruct the CHR-OCT image. The discriminator D based on [49] acts as a judgement that determines whether the label y is more realistic than the image $G(x)$ generated by the generator. On the opposite, during the training process, the generator is optimized to generate more realistic results to fool the discriminator. Simultaneously, the discriminator tries to improve itself by distinguishing whether the label y is more realistic or the generated image $G(x)$ is less realistic.

2.2 Network architecture

2.2.1 Feature extractor

The backbone of our proposed generator is based on the dense deep back-projection network (D-DBPN) [50]. However, the original D-DBPN simply employs two convolution layers to extract features from the LR images, which may be not suitable for our task because our input images not only are low resolution but also are greatly corrupted by speckle noise. Therefore, inspired by the residual in the residual dense block (RRDB) proposed by Wang et al. [51], we adopt this structure including ten cascade dense blocks as our feature extractor, which exploits residual learning and a dense connection to obtain information from the NLR input images. We experimentally adopted more layers in our feature extractor than the original D-DBPN for an improved performance [52,53]. As shown in Fig. 2, every dense block contains five convolution layers, each of which is followed by a leaky rectified linear unit (LReLU) activation function except for the last one.

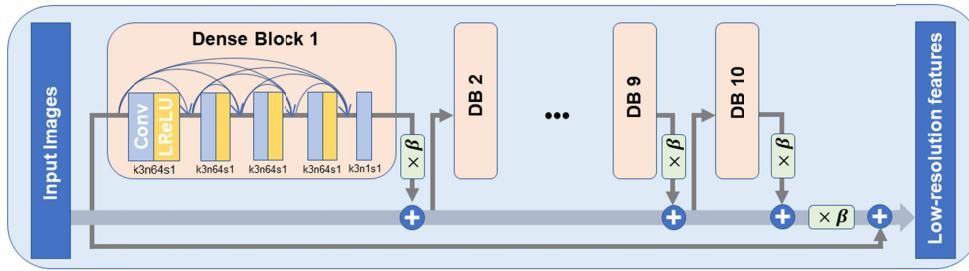


Fig. 2. The network architecture of feature extractor. The same color in the figure represents the same layer or the same block. Conv and LReLU are the Convolution layer and the Leaky ReLU nonlinearity. β is the residual scaling parameter, which is set to 0.2 during all the experiments. Parameter k represents the kernel size, n is the number of filters and s is the size of the stride.

2.2.2 Upsampler

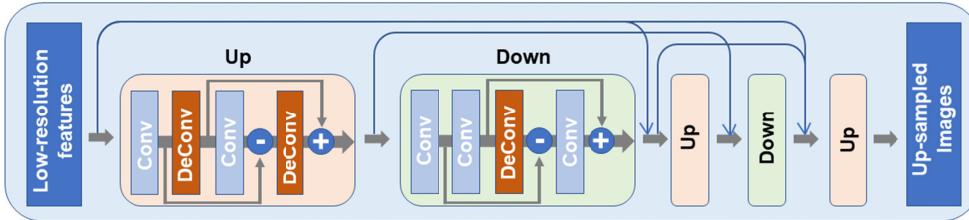


Fig. 3. The network architecture of upsampler.

Many previously proposed convolutional neural network (CNN)-based image super-resolution algorithms use only one upscale operator, such as a bicubic interpolation or single deconvolution layer, to reconstruct the HR image from the LR input, which makes it difficult to learn a good mapping from LR to HR. Borrowing the idea from D-DBPN, we use several deep back-projection layers to alternatively upscale and downscale the features extracted from the feature extractor. Consequently, the error is fed back to the network to obtain better performance.

As shown in Fig. 3, the up block and down block are the same as the original up and down projection units in [50], except that all the up and down blocks of our upsampler begin with a 1×1 convolution layer, rather than parts of them. Three up blocks and two down blocks are included in the upsampler.

2.2.3 Discriminator

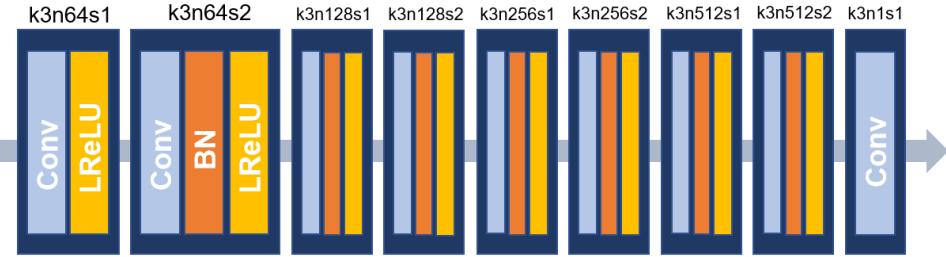


Fig. 4. The discriminator of our proposed network. The same color represents the same layer. The parameter k represents the kernel size of the convolution layers, n is the number of filters, and s represents the stride.

As shown in Fig. 4, the discriminator of our GAN model consists of a sequence of convolution layers, each of which is followed by batch normalization (BN) and leaky ReLU (LReLU) nonlinearity, except for the first and last layers. For the first layer, there is no BN between the convolution layer and its activation, but for the last layer, only the convolution layer exists. The parameter k represents the kernel size of the convolution layers, n is the number of filters, and s represents the stride. Increasing k or n will respectively provide a larger receptive field or stronger representation ability of the network, but the corresponding computational cost will rise simultaneously. In this paper, these parameters are manually set to balance the tradeoff between the performance and computational complexity.

2.3 Objective function

Following the idea of [51], the discriminator of our GAN model determines whether the label y is more realistic than the generated image $G(x)$ as follows:

$$\begin{aligned} D_{real} &= \sigma(D(y) - E[D(G(x))]) \rightarrow 1 \quad \text{if } y \text{ is more realistic than } G(x) \\ D_{fake} &= \sigma(D(G(x)) - E[D(y)]) \rightarrow 0 \quad \text{if } G(x) \text{ is less realistic than } y \end{aligned} \quad (3)$$

where σ is the sigmoid function and $E[\cdot]$ represents the mean of the discriminator output with all images in a mini-batch. Then, we defined the loss function (named GAN loss) as follows:

$$L_{GAN} = \frac{1}{2}(1 - D_{real} + D_{fake}) \quad (4)$$

which is the discriminator loss.

As depicted in Fig. 1, there are two more losses in our model. For the generator, we first use mean square error (MSE) loss to maintain the image details and the information content. We denote the loss as the pixel loss because it performs a pixel-wise error minimization procedure between a processed image $G(x)$ and its label y . Therefore, we can define the pixel loss as follows:

$$L_{Pixel} = \frac{1}{N} \sum_{i=1}^N \|y - G(x)\|^2 \quad (5)$$

However, using the MSE alone tends to oversmooth the denoised image, and it is difficult for image super-resolution because the same LR images may have multiple corresponding HR images [54]. Thus, in addition to the MSE loss, we also add a perceptual loss defined as:

$$L_{Perceptual} = \frac{1}{N} \sum_{i=1}^N \|VGG_{19}(y) - VGG_{19}(G(x))\| \quad (6)$$

where VGG_{19} represents the VGG-19 network [38], which contains 16 convolution layers and three fully connected layers. Here, we only use the first 12 convolution layers to extract the features of the label y and the generated image $G(x)$.

Finally, the loss function of our generator is a combination of L_{Pixel} , $L_{Perceptual}$, and L_{GAN} :

$$L_G = \alpha L_{Pixel} + \beta L_{Perceptual} + \gamma L_{GAN} \quad (7)$$

where α , β , and γ are the weighting coefficients.

3. Experiments

3.1 Data

We perform our experiments on two data sets. The first data set was originally introduced by [31], which contains $28 \times 450 \times 900$ (height \times width) synthetic retinal SD-OCT image pairs generated by subsampling high-resolution images captured from 28 normal or abnormal eyes of 28 subjects. More concretely, in each image pair, there is a noisy SD-OCT image captured by a Biopitgen SD-OCT imaging system, and a clear OCT image which is acquired by registering and averaging several B-scans obtained at the same position. In our experiments, we discarded two pairs of very poorly registered images and used the remaining 26 pairs to conduct our experiments, ten pairs are used for training the models and the rest are used for testing. Since a large number of NLR and CHR OCT image pairs are critical to train our SDSR-OCT models, we experimentally traversed ten 450×900 (height \times width) training pairs with a window of 256×256 and a stride of 8, followed by removing some patches with no signal information to generate total 16000 noisy and clear patch pairs. During the training procedure, the clear patches are regarded as the CHR images and the NLR images are generated by downsampling the noisy patches in the patch pairs with the corresponding scale factor. Based on our experiments, more than 10000 samples are expected to be a sufficient number for training a usable network. In the testing, we use the rest 16 OCT image pairs to quantitatively and qualitatively compare the performance with some state-of-art methods. Moreover, to evaluate the generalization and robustness of our method, one more data set was involved in testing stage. We randomly selected 100 Age-related Macular Degeneration (AMD) OCT images from the second data set introduced in [55] that was from A2A SD-OCT Study and registered at ClinicalTrials.gov. Due to the limitation of GPU memory, the original 512×1000 (height \times width) AMD OCT images were cropped to the size of 448×448 (height \times width) as our test images. Note that, there is no clear reference for the second data set and both data sets were tested with the network model trained by the first data set.

3.2 Implementation details

In the experiment, we use three different scale factors between NLR and CHR images to train three different models that can achieve $2 \times$, $4 \times$ and $8 \times$ super-resolution and its corresponding denoising results.

All the parameters of our feature extractor and the discriminator can be found in Fig. 2 and 4, while the parameters of our upsample are the same as in the original D-DBPN [50]. For the $2 \times$ upsample, we use a kernel size of 6×6 followed by 2 strides, and then the $4 \times$ upsample uses an 8×8 kernel size and a stride of 4. For the $8 \times$ upsample, we employ the convolution layer with a kernel size of 12×12 and a stride of 8. Furthermore, the padding is set to 2 for all the convolution layers.

In the training part, we use the Adam optimizer with $\beta_1=0.9$ and $\beta_2=0.999$ for both the generator and discriminator. Our learning rate is initialized with $1e-4$, and it is decayed by a factor of 0.1 for every 20, 20 and 50 epochs for the $2 \times$, $4 \times$ and $8 \times$ models, respectively. The weighting coefficients of our generator loss are set to $\alpha=0.01$, $\beta=1$ and $\gamma=0.005$. Furthermore, the corresponding batch sizes of our $2 \times$, $4 \times$ and $8 \times$ models are set to 1, 4 and 6, respectively, due to the limitation of the graphics processing unit (GPU) memory. The deep learning framework is Pytorch, and all the experiments are conducted on the NVIDIA GTX 1080Ti GPU.

3.3 Evaluation metrics

For the sake of quantitatively evaluating the denoising and super-resolution performance of our algorithm and compared to other methods, we employ four state-of-art evaluation metrics: peak signal-to-noise ratio (PSNR), edge preservation index (EPI), equivalent number of looks (ENL) and contrast-to-noise ratio (CNR). We compute the PSNR and EPI of the entire image. For the ENL and CNR, we only measure it via several manually selected regions of interest (ROIs). These ROIs can be seen in Figs. 5-7, where the green rectangle (#0) represents the background ROI and four red rectangles (#1~4) indicate the signal ROIs. The signal ROIs are selected at the retinal layers or near the layers since boundaries between retinal layers contain important information to understand the disease severity and the pathogenic processes [56]. We also visually compared our denoising and super-resolution results with that of the compared methods in Figs. 5-10. For better visual comparison, three boundary areas (red rectangles #1-3) are selected and enlarged in Figs. 5-10.

3.3.1 Peak signal-to-noise ratio (PSNR)

The PSNR is the global metric that computes the similarity between the processed image and the reference image. This metric is simply defined by MSE as follows:

$$PSNR = 10 \log_{10} \left(\frac{\max(I)^2}{MSE} \right) \quad (8)$$

where $\max(I)$ represents the theoretical maximum of the pixel in image I .

3.3.2 Edge preservation index (EPI)

The EPI reflects the ability to preserve the edge details of the image after denoising or other processing. Because the edges of testing OCT images mainly exist in the longitudinal direction, the EPI is defined as:

$$EPI = \frac{\sum_i \sum_j |I_d(i+1, j) - I_d(i, j)|}{\sum_i \sum_j |I_n(i+1, j) - I_n(i, j)|} \quad (9)$$

where I_d represents the denoised image and I_n is the noisy image. i and j denote the i -th row and j -th column of the image. For EPI, 1 corresponds to the ideal situation which means perfect edge preservation.

3.3.3 Equivalent number of looks (ENL)

ENL is an index used to measure the smoothness of the homogeneous region of the denoised image, which is widely used in OCT image speckle suppression. This index is calculated it over the background ROI of each test image as follows:

$$ENL = \frac{\mu_b^2}{\sigma_b^2} \quad (10)$$

where μ_b and σ_b represent the mean and standard deviation of selected background ROI in each image, respectively.

3.3.4 Contrast-to-noise ratio (CNR)

The CNR measures the contrast between the signal region and the background region. It is calculated as follows:

$$CNR = \frac{1}{m} \sum_{i=1}^m \left[10 \log_{10} \left(\frac{|\mu_i - \mu_b|}{\sqrt{\sigma_i^2 + \sigma_b^2}} \right) \right] \quad (11)$$

where m is the total number of selected signal ROIs; μ_i and σ_i represent the mean and standard deviation of the i -th selected signal ROI in each image, respectively; μ_b and σ_b represent the mean and standard deviation of the selected background ROI, respectively.

3.4 Experimental results

3.4.1 Comparison to the state-of-the-art methods

To evaluate the performance of our proposed SDSR-OCT algorithm, we quantitatively and qualitatively compare our approach with some state-of-art methods, including the classical Block-matching and 3D filtering (BM3D) + BICUBIC methods (BM3D [57] is for denoising and BICUBIC is for interpolation), newly proposed competitive NWSR [8] and CNN-based image super-resolution method, SRCNN [39]. We tested 16 labeled synthetic retinal OCT images in the data set [58] for both the denoising and super-resolution of different upscale factors ($2 \times$, $4 \times$ and $8 \times$).

For the super-resolution and corresponding denoising problem, we train the model of our proposed method and the SRCNN method from scratch with 50%, 75% and 87.5% data missing. At the same time, we use the BICUBIC interpolate algorithm on the input image and denoise it using the BM3D algorithm to produce the BM3D + BICUBIC results. For the NWSR, we simply use its pre-trained dictionary to perform the experiments.

Figures 5-7 show the visual and quantitative results of two representative cases processed by our proposed approach and the compared methods, from which we can see that our $2 \times$ model performs favorably against the compared methods of both PSNR and EPI metrics. These results suggest that our algorithm can maintain the most edge details as well as keep the content between the ground truth and the processed image. These data can be confirmed by the visual inspection in Fig. 5. The edges are well preserved and reconstructed by SDSR-OCT. NWSR and SRCNN blurred the edges to various degrees. Similar to the $2 \times$ model, the PSNR and EPI results of our $4 \times$ and $8 \times$ models also provide the better performance than all the compared methods. Notably, the visual results of our $4 \times$ and $8 \times$ models considerably outperform the compared methods, as illustrated in Figs. 6-7. The edge details of the compared methods are greatly influenced by the over-smoothness and lose almost all the key information of clinical diagnosis. However, our proposed model can well preserve those edge details and key information, even at large upscale factors, such as $8 \times$.

Table 1 reveals the statistical quantitative results of different methods on the whole first data set. It can be seen that SDSR-OCT achieves best PSNR values in $2 \times$ and $8 \times$ situations. At $4 \times$ upscale factor, the best mean of the PSNR is achieved by the BM3D + BICUBIC due to an oversmoothed output. For the metric EPI, which mainly indicates the ability of edge preservation, SDSR-OCT gains the best score in all the situations, which is coherent to the visual inspection in Figs. 5-7. The ENL mainly calculates the ratio of the mean and standard

deviation of the background area, and the smoother the background is, the higher it is. From Table 1, our results of the ENL lies in the middle level among compared methods, and SRCNN obtains the best mean of the ENL mainly because it has an aggressive background smoothing. Similar to ENL, our results of the CNR also stays at a middle level, while the NWSR and BM3D + BICUBIC achieve better performance in different upscale factors respectively. However, from Figs. 5-7, we observe that both methods have oversmoothed results, which are supposed to the potential reason that why the better results of the ENL and CNR can be achieved by BM3D + BICUBIC, NWSR and SRCNN in certain situations.

Table 1. Quantitative results of the compared methods and our method

		BM3D + BICUBIC	NWSR	SRCNN	OURS
PSNR (mean)	2 ×	27.626	27.898	27.959	28.132
	4 ×	28.029	27.497	27.777	27.889
	8 ×	27.777	27.105	27.812	27.826
EPI (mean)	2 ×	0.751	0.930	0.826	0.953
	4 ×	0.545	0.915	0.735	0.947
	8 ×	0.350	0.883	0.573	0.930
ENL (mean)	2 ×	416.2	763.3	1902.3	537.5
	4 ×	813.3	882.2	2437.4	915.1
	8 ×	2503.3	1098.6	4344.8	1669.5
CNR (mean)	2 ×	4.424	4.693	4.399	4.562
	4 ×	4.950	4.786	4.323	4.451
	8 ×	5.792	4.796	4.448	4.460

We also present the running times of our proposed model and the compared methods in Table 2, from which we can see that the testing time consumption of our approach is much less than that of the BM3D + BICUBIC and the NWSR, but slightly more than that of the SRCNN method due to the network capacity. Combining the above denoising and super-resolution results, we can conclude that our model exhibits fast denoising and good super-resolution performance for OCT images.

Table 2. Average running time of the compared methods and our method

		BM3D + BICUBIC	NWSR	SRCNN	OURS
Running time	2 ×	3.9987 s	35.9375 s	0.0005 s	0.0720 s
	4 ×	3.8488 s	39.0625 s	0.0006 s	0.1183 s
	8 ×	3.9675 s	43.0000 s	0.0006 s	0.0528 s

3.4.2 Analysis for the performance robustness of different data set

To validate the robustness of the proposed model, one more data set from [55] was tested. Due to the lack of reference images, only visual results are demonstrated in Figs. 8-10. In Fig. 8, all the methods can remove most of the noise and our SDSR-OCT still obtains better visual effects. In the magnified ROIs, obvious artifacts appear along the edges in the results of BM3D + BICUBIC. NWSR distorts the structure in ROI #2 and the boundaries of two layers in ROI #3 cannot be well identified due to the blurring effect. SRCNN has better results than NWSR, but it is noticed that the structure details look oversmoothed. In Fig. 9, the advantages of SDSR-OCT are more remarkable. The results of SDSR-OCT still look close to the ones of 2 × model. The artifacts become more serious in the results of BM3D + BICUBIC and mosaic-like artifacts appear in the results of NWSR. In Fig. 10, only our method reconstructs structural information to a certain degree and none of other methods can maintain clinically useful details.

4. Conclusion

In this paper, we present a novel method for the simultaneous denoising and super-resolution of OCT images. Our proposed algorithm uses a GAN to train three different models ($2 \times$, $4 \times$, and $8 \times$ model) in an end-to-end manner. Therefore, this approach is not only suitable for conventional downsampling but also for lower sampling rates, such as when 75% and 87.5% of the data is missing during image acquisition. Additionally, we have quantitatively and qualitatively conducted extensive experiments to compare our proposed method to some well-known denoising and super-resolution approaches. The experimental results show that our proposed method can efficiently denoise and super-resolve SD-OCT images while maintaining the most edge information and image details, especially in the case of a large-scale imagery. However, slight noise still exists in our results, and it is this limitation that we will work hard to address in further research.

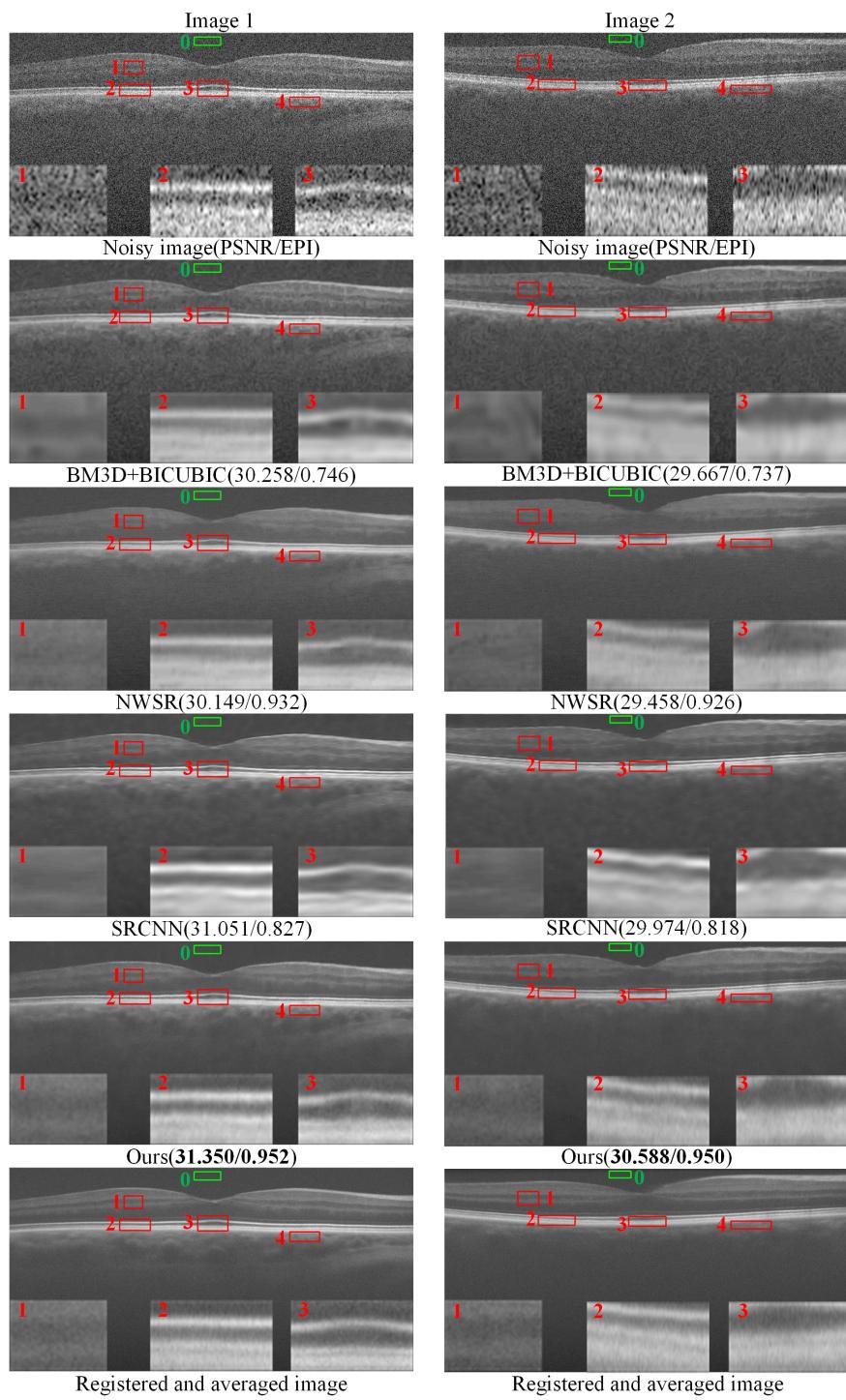


Fig. 5. Visual comparison of the denoising and $2 \times$ super-resolution results of two synthetic subsampled retinal OCT images in the first data set using BM3D + BICUBIC, NWSR, SRCNN and the proposed SDSR-OCT method.

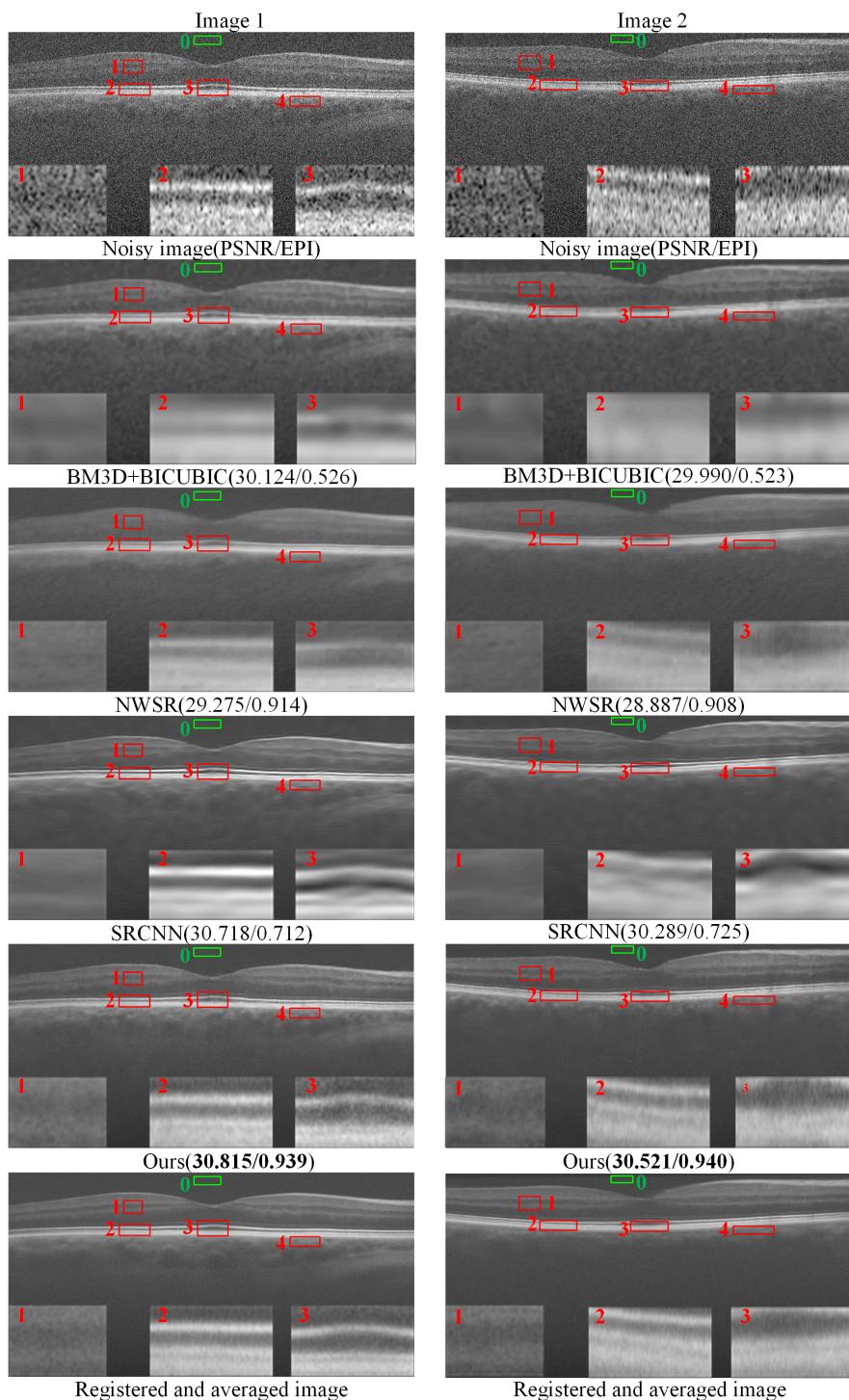


Fig. 6. Visual comparison of the denoising and 4 × super-resolution results of two synthetic subsampled retinal OCT images in the first data set using BM3D + BICUBIC, NWSR, SRCNN, and the proposed SDRS-OCT methods.

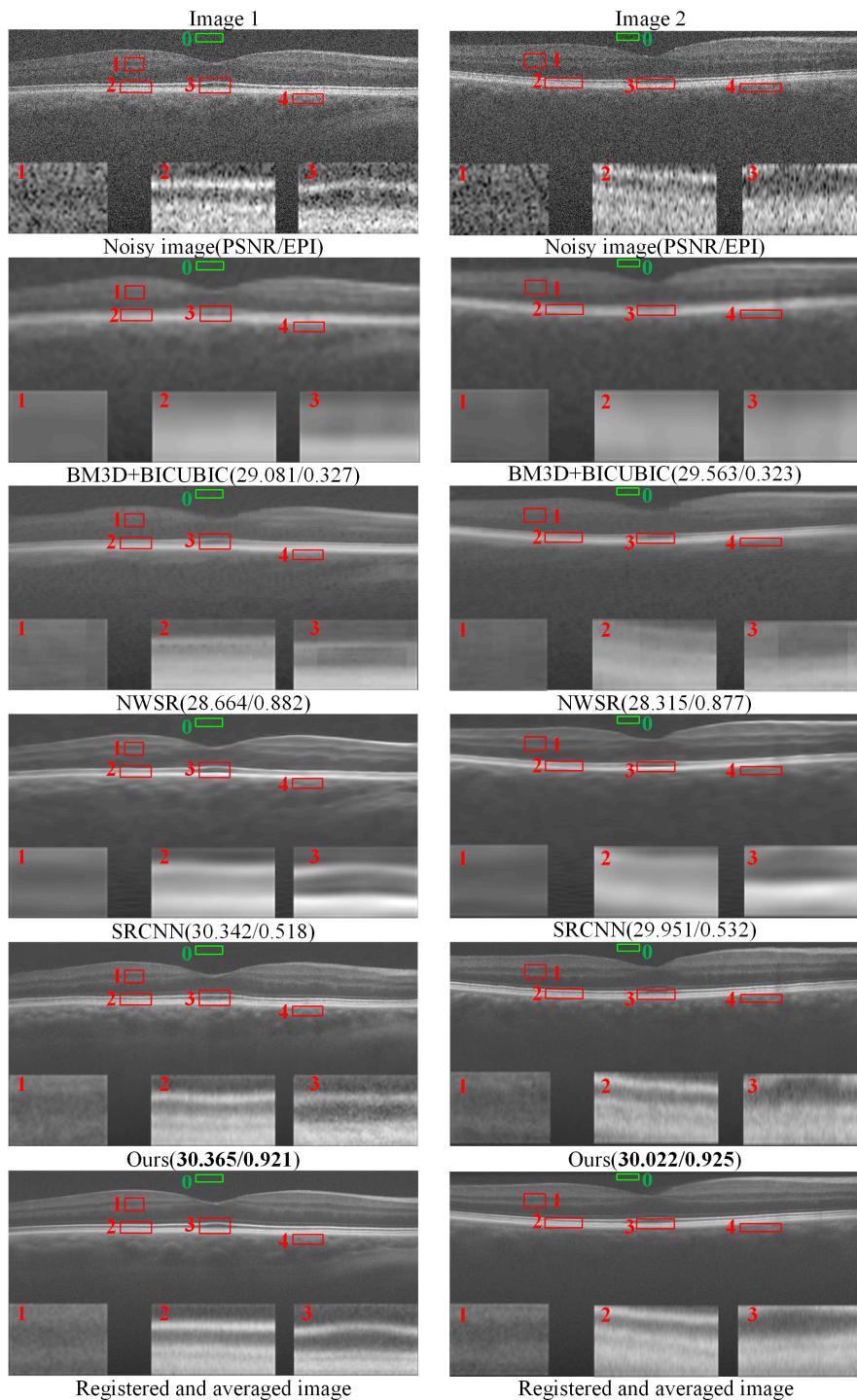


Fig. 7. Visual comparison of the denoising and $8 \times$ super-resolution results of two synthetic subsampled retinal OCT images in the first data set using BM3D + BICUBIC, NWSR, SRCNN, and the proposed SDSR-OCT methods.

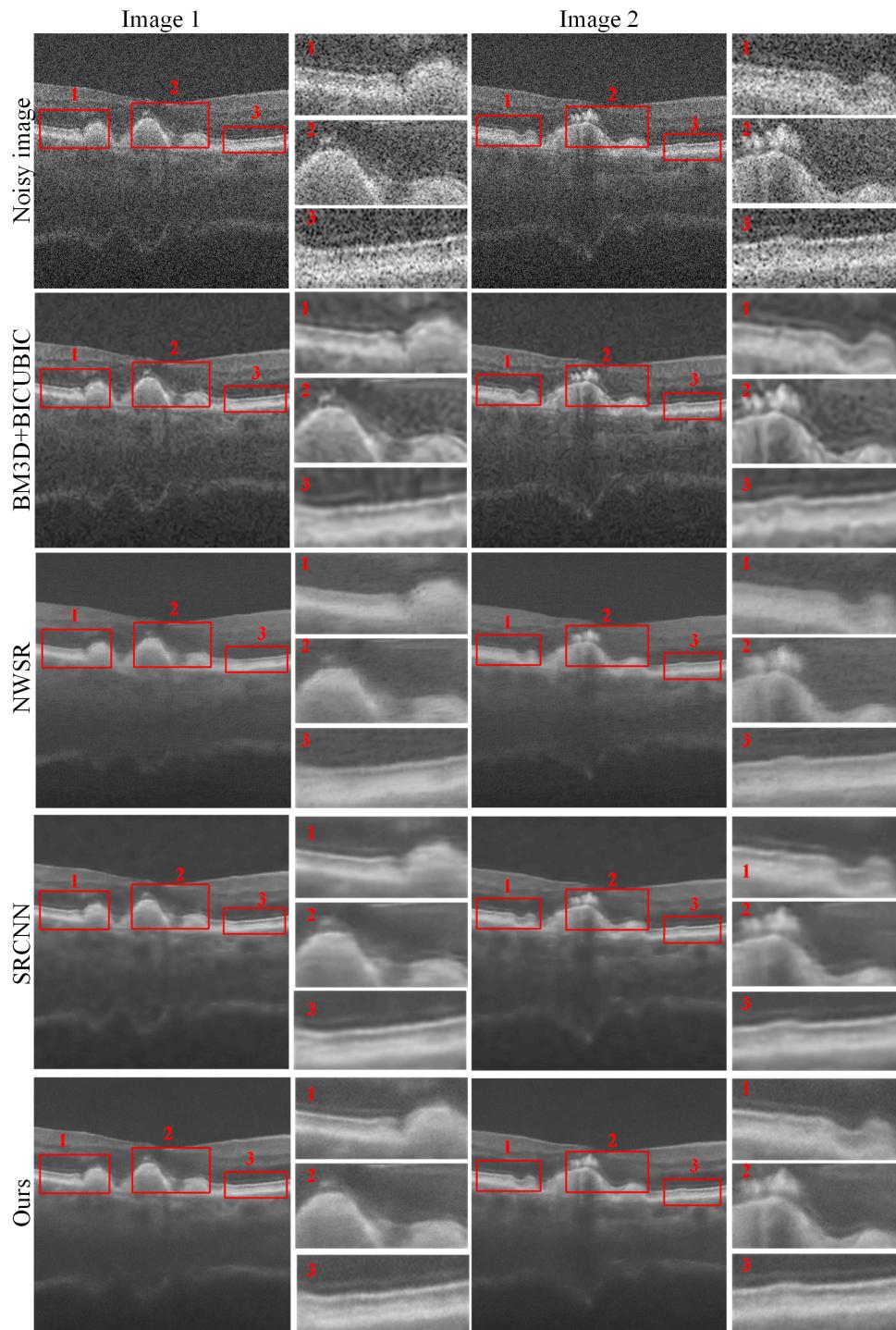


Fig. 8. Visual comparison of the denoising and $2 \times$ super-resolution results of two AMD retinal OCT images in the second data set using BM3D + BICUBIC, NWSR, SRCNN, and the proposed SDRS-OCT methods.

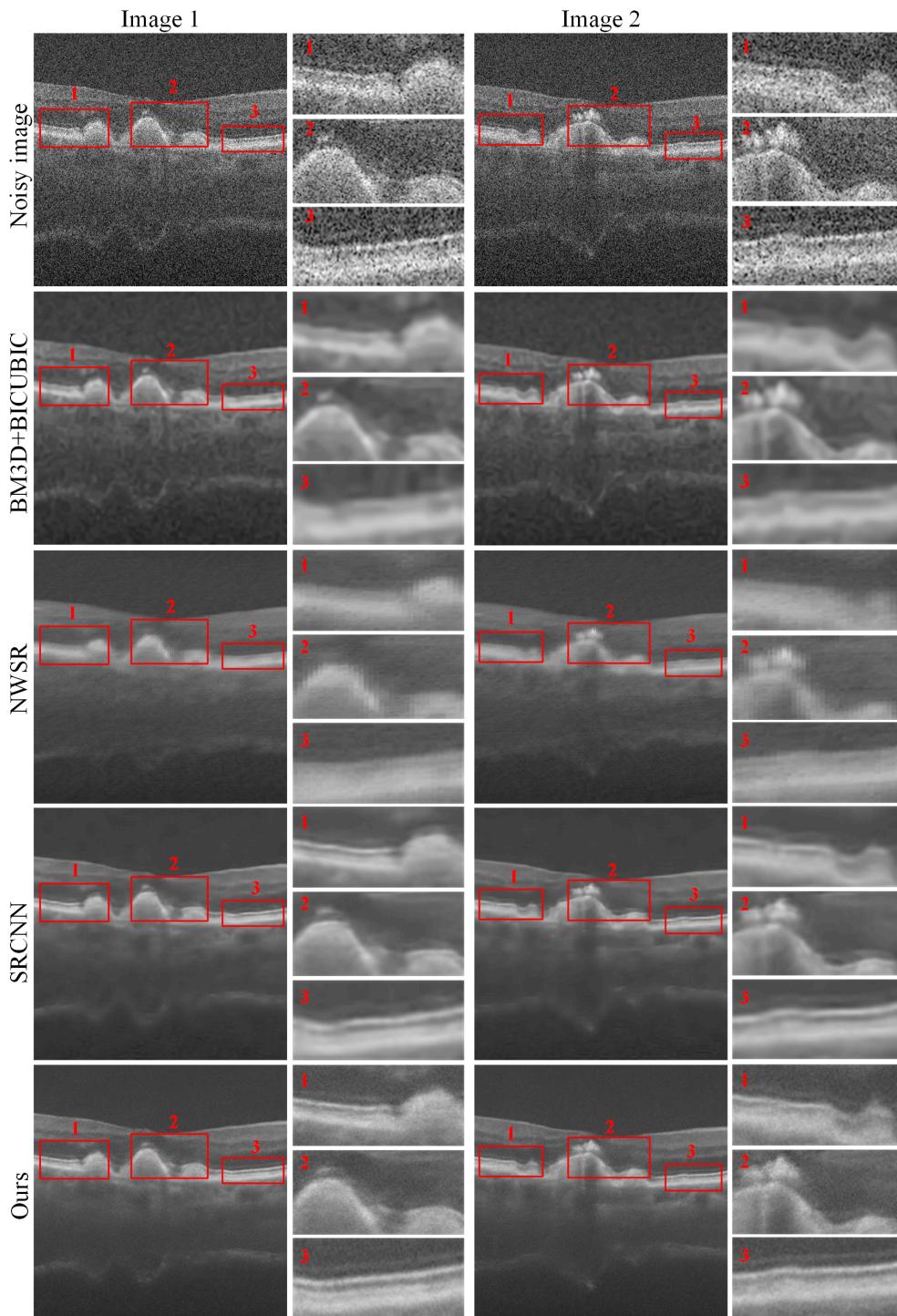


Fig. 9. Visual comparison of the denoising and $4 \times$ super-resolution results of two AMD retinal OCT images in the second data set using BM3D + BICUBIC, NWSR, SRCNN, and the proposed SDSR-OCT methods.

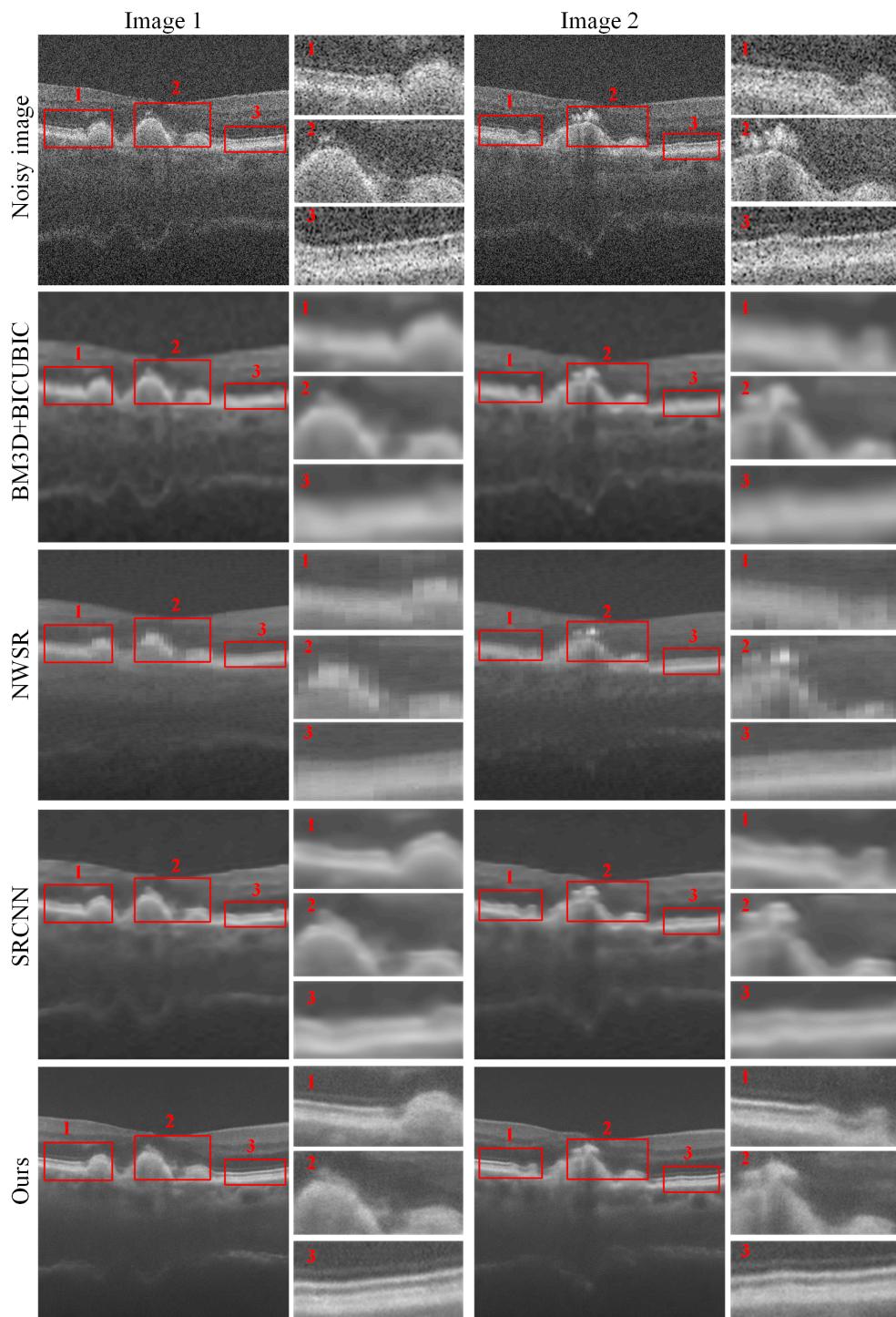


Fig. 10. Visual comparison of the denoising and $8 \times$ super-resolution results of two AMD retinal OCT images in the second data set using BM3D + BICUBIC, NWSR, SRCNN, and the proposed SDSR-OCT methods.

Funding

National Key R&D Program of China (2017YFB0802300); National Natural Science Foundation of China (NSFC) (61671312, 61771192); Sichuan Science and Technology Program (2018HH0070).

References

- W. Drexler, U. Morgner, R. K. Ghanta, F. X. Kärtner, J. S. Schuman, and J. G. Fujimoto, "Ultrahigh-resolution ophthalmic optical coherence tomography," *Nat. Med.* **7**(4), 502–507 (2001).
- J. M. Schmitt, S. H. Xiang, and K. M. Yung, "Speckle in optical coherence tomography," *J. Biomed. Opt.* **4**(1), 95–105 (1999).
- M. Wojtkowski, R. Leitgeb, A. Kowalczyk, T. Bajraszewski, and A. F. Fercher, "In vivo human retinal imaging by Fourier domain optical coherence tomography," *J. Biomed. Opt.* **7**(3), 457–463 (2002).
- W. Drexler and J. G. Fujimoto, "State-of-the-art retinal optical coherence tomography," *Prog. Retin. Eye Res.* **27**(1), 45–88 (2008).
- G. Gong, H. Zhang, and M. Yao, "Speckle noise reduction algorithm with total variation regularization in optical coherence tomography," *Opt. Express* **23**(19), 24699–24712 (2015).
- H. Lv, S. Fu, C. Zhang, and L. Zhai, "Speckle noise reduction of multi-frame optical coherence tomography data using multi-linear principal component analysis," *Opt. Express* **26**(9), 11804–11818 (2018).
- Y. Ma, X. Chen, W. Zhu, X. Cheng, D. Xiang, and F. Shi, "Speckle noise reduction in optical coherence tomography images based on edge-sensitive cGAN," *Biomed. Opt. Express* **9**(11), 5129–5146 (2018).
- A. Abbasi, A. Monadjemi, L. Fang, and H. Rabbani, "Optical coherence tomography retinal image reconstruction via nonlocal weighted sparse representation," *J. Biomed. Opt.* **23**(3), 1–11 (2018).
- A. Boroomand, A. Wong, E. Li, D. S. Cho, B. Ni, and K. Bizheva, "Multi-penalty conditional random field approach to super-resolved reconstruction of optical coherence tomography images," *Biomed. Opt. Express* **4**(10), 2032–2050 (2013).
- M. Young, E. Lebed, Y. Jian, P. J. Mackenzie, M. F. Beg, and M. V. Sarunic, "Real-time high-speed volumetric imaging using compressive sampling optical coherence tomography," *Biomed. Opt. Express* **2**(9), 2690–2697 (2011).
- J. Xu, H. Ou, C. Sun, P.-C. Chui, V. X. Yang, E. Y. Lam, and K. K. Wong, "Wavelet domain compounding for speckle reduction in optical coherence tomography," *J. Biomed. Opt.* **18**(9), 096002 (2013).
- M. Pircher, E. Götzinger, R. Leitgeb, A. F. Fercher, and C. K. Hitzenberger, "Speckle reduction in optical coherence tomography by frequency compounding," *J. Biomed. Opt.* **8**(3), 565–569 (2003).
- T. Bajraszewski, M. Wojtkowski, M. Szkulmowski, A. Szkulmowska, R. Huber, and A. Kowalczyk, "Improved spectral optical coherence tomography using optical frequency comb," *Opt. Express* **16**(6), 4163–4176 (2008).
- A. E. Desjardins, B. J. Vakoc, W. Y. Oh, S. M. Motaghiannezhad, G. J. Tearney, and B. E. Bouma, "Angle-resolved optical coherence tomography with sequential angular selectivity for speckle reduction," *Opt. Express* **15**(10), 6200–6209 (2007).
- B. F. Kennedy, T. R. Hillman, A. Curatolo, and D. D. Sampson, "Speckle reduction in optical coherence tomography by strain compounding," *Opt. Lett.* **35**(14), 2445–2447 (2010).
- T. Klein, R. André, W. Wieser, T. Pfeiffer, and R. Huber, "Joint aperture detection for speckle reduction and increased collection efficiency in ophthalmic MHz OCT," *Biomed. Opt. Express* **4**(4), 619–634 (2013).
- S. Adabi, E. Rashedi, A. Clayton, H. Mohebbi-Kalkhoran, X. W. Chen, S. Conforto, and M. Nasiriavanaki, "Learnable despeckling framework for optical coherence tomography images," *J. Biomed. Opt.* **23**(1), 1–12 (2018).
- M. H. Eybposh, Z. Turani, D. Mehregan, and M. Nasiriavanaki, "Cluster-based filtering framework for speckle reduction in OCT images," *Biomed. Opt. Express* **9**(12), 6359–6373 (2018).
- A. Ozcan, A. Bilenca, A. E. Desjardins, B. E. Bouma, and G. J. Tearney, "Speckle reduction in optical coherence tomography images using digital filtering," *J. Opt. Soc. Am. A* **24**(7), 1901–1910 (2007).
- J. Rogowska and M. E. Brezinski, "Image processing techniques for noise removal, enhancement and segmentation of cartilage OCT images," *Phys. Med. Biol.* **47**(4), 641–655 (2002).
- S. Aja-Fernández and C. Alberola-López, "On the estimation of the coefficient of variation for anisotropic diffusion speckle filtering," *IEEE Trans. Image Process.* **15**(9), 2694–2701 (2006).
- R. Bernardes, C. Maduro, P. Serranho, A. Araújo, S. Barbeiro, and J. Cunha-Vaz, "Improved adaptive complex diffusion despeckling filter," *Opt. Express* **18**(23), 24048–24059 (2010).
- P. Puvanathasan and K. Bizheva, "Interval type-II fuzzy anisotropic diffusion algorithm for speckle noise reduction in optical coherence tomography images," *Opt. Express* **17**(2), 733–746 (2009).
- Y. Yu and S. T. Acton, "Speckle reducing anisotropic diffusion," *IEEE Trans. Image Process.* **11**(11), 1260–1270 (2002).
- M. A. Mayer, A. Borsdorf, M. Wagner, J. Hornegger, C. Y. Mardin, and R. P. Tornow, "Wavelet denoising of multiframe optical coherence tomography data," *Biomed. Opt. Express* **3**(3), 572–589 (2012).
- F. Zaki, Y. Wang, H. Su, X. Yuan, and X. Liu, "Noise adaptive wavelet thresholding for speckle noise removal in optical coherence tomography," *Biomed. Opt. Express* **8**(5), 2720–2731 (2017).

27. H. Chen, S. Fu, H. Wang, H. Lv, and C. Zhang, "Speckle attenuation by adaptive singular value shrinking with generalized likelihood matching in optical coherence tomography," *J. Biomed. Opt.* **23**(3), 1–8 (2018).
28. L. Fang, S. Li, Q. Nie, J. A. Izatt, C. A. Toth, and S. Farsiu, "Sparsity based denoising of spectral domain optical coherence tomography images," *Biomed. Opt. Express* **3**(5), 927–942 (2012).
29. C. Tang, L. Cao, J. Chen, and X. Zheng, "Speckle noise reduction for optical coherence tomography images via non-local weighted group low-rank representation," *Laser Phys. Lett.* **14**(5), 056002 (2017).
30. M. Dinh-Hoan Trinh, M. Luong, F. Dibos, J. M. Rocchisani, Canh-Duong Pham, and T. Q. Nguyen, "Novel example-based method for super-resolution and denoising of medical images," *IEEE Trans. Image Process.* **23**(4), 1882–1895 (2014).
31. L. Fang, S. Li, R. P. McNabb, Q. Nie, A. N. Kuo, C. A. Toth, J. A. Izatt, and S. Farsiu, "Fast acquisition and reconstruction of optical coherence tomography images via sparse representation," *IEEE Trans. Med. Imaging* **32**(11), 2034–2049 (2013).
32. L. Fang, S. Li, D. Cunefare, and S. Farsiu, "Segmentation based sparse reconstruction of optical coherence tomography images," *IEEE Trans. Med. Imaging* **36**(2), 407–421 (2017).
33. W. Dong, L. Zhang, G. Shi, and X. Li, "Nonlocally centralized sparse representation for image restoration," *IEEE Trans. Image Process.* **22**(4), 1620–1630 (2013).
34. W. Dong, L. Zhang, G. Shi, and X. Wu, "Image deblurring and super-resolution by adaptive sparse domain selection and adaptive regularization," *IEEE Trans. Image Process.* **20**(7), 1838–1857 (2011).
35. S. Yang, Z. Liu, M. Wang, F. Sun, and L. Jiao, "Multitask dictionary learning and sparse representation based single-image super-resolution reconstruction," *Neurocomputing* **74**(17), 3193–3203 (2011).
36. A. Krizhevsky, I. Sutskever, and G. E. Hinton, "ImageNet classification with deep convolutional neural networks," in *International Conference on Neural Information Processing Systems* (Curran Associates, 2012), pp.1097–1105.
37. K. He, X. Zhang, S. Ren, and J. Sun, "Deep residual learning for image recognition," in *Proceedings of IEEE Conference on Computer Vision and Pattern Recognition* (IEEE, 2016), pp. 770–778.
38. K. Zhang, W. Zuo, Y. Chen, D. Meng, and L. Zhang, "Beyond a gaussian denoiser: Residual learning of deep cnn for image denoising," *IEEE Trans. Image Process.* **26**(7), 3142–3155 (2017).
39. C. Dong, C. C. Loy, K. He, and X. Tang, "Image super-resolution using deep convolutional networks," *IEEE Trans. Pattern Anal. Mach. Intell.* **38**(2), 295–307 (2016).
40. H. Chen, Y. Zhang, W. Zhang, P. Liao, K. Li, J. Zhou, and G. Wang, "Low-dose CT via convolutional neural network," *Biomed. Opt. Express* **8**(2), 679–694 (2017).
41. Y. Chen, F. Shi, A. G. Christodoulou, Y. Xie, Z. Zhou, and D. Li, "Efficient and Accurate MRI Super-Resolution Using a Generative Adversarial Network and 3D Multi-level Densely Connected Network," in *International Conference on Medical Image Computing and Computer-Assisted Intervention* (Springer, Cham, 2018) pp. 91–99.
42. M. Li, S. Shen, W. Gao, W. Hsu, and J. Cong, "Computed Tomography Image Enhancement using 3D Convolutional Neural Network," in *Deep Learning in Medical Image Analysis and Multimodal Learning for Clinical Decision Support* (Springer, Cham, 2018) pp. 291–299.
43. J. Shi, Q. Liu, C. Wang, Q. Zhang, S. Ying, and H. Xu, "Super-resolution reconstruction of MR image with a novel residual learning network algorithm," *Phys. Med. Biol.* **63**(8), 085011 (2018).
44. S. K. Devalla, G. Subramanian, T. H. Pham, X. Wang, S. Perera, T. A. Tun, T. Aung, L. Schmetterer, A. H. Thiery, and M. J. Girard, "A Deep Learning Approach to Denoise Optical Coherence Tomography Images of the Optic Nerve Head," <https://arxiv.org/abs/1809.10589v1>.
45. X. Fei, J. Zhao, H. Zhao, D. Yun, and Y. Zhang, "Deblurring adaptive optics retinal images using deep convolutional neural networks," *Biomed. Opt. Express* **8**(12), 5675–5687 (2017).
46. K. J. Halupka, B. J. Antony, M. H. Lee, K. A. Lucy, R. S. Rai, H. Ishikawa, G. Wollstein, J. S. Schuman, and R. Garnavi, "Retinal optical coherence tomography image enhancement via deep learning," *Biomed. Opt. Express* **9**(12), 6205–6221 (2018).
47. L. Fang, C. Wang, S. Li, H. Rabbani, X. Chen and Z. Liu, "Attention to Lesion: Lesion-Aware Convolutional Neural Network for Retinal Optical Coherence Tomography Image Classification," *IEEE Trans. Med. Imaging* (to be published).
48. I. Goodfellow, J. Pouget-Abadie, M. Mirza, B. Xu, D. Warde-Farley, S. Ozair, A. Courville, and Y. Bengio, "Generative adversarial nets," in *Advances in Neural Information Processing Systems* (NIPS, 2014), pp. 2672–2680.
49. A. Jolicoeur-Martineau, "The relativistic discriminator: a key element missing from standard GAN," <https://arxiv.org/abs/1807.00734>.
50. M. Haris, G. Shakhnarovich, and N. Ukita, "Deep back-projection networks for super-resolution," in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition* (IEEE, 2018), pp. 1664–1673.
51. X. Wang, K. Yu, S. Wu, J. Gu, Y. Liu, C. Dong, C. C. Loy, Y. Qiao, and X. Tang, "ESRGAN: Enhanced super-resolution generative adversarial networks," <https://arxiv.org/abs/1809.00219>.
52. Y. Zhang, Y. Tian, Y. Kong, B. Zhong, and Y. Fu, "Residual Dense Network for Image Super-Resolution," in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition* (IEEE 2018), pp. 2472–2481.

53. Y. Zhang, K. Li, K. Li, L. Wang, B. Zhong, and Y. Fu, "Image Super-Resolution Using Very Deep Residual Channel Attention Networks," in *Proceedings of the European Conference on Computer Vision* (Springer, 2018) (pp. 294–301).
54. W. S. Lai, J. B. Huang, N. Ahuja, and M. H. Yang, "Fast and accurate image super-resolution with deep laplacian pyramid networks," <https://arxiv.org/abs/1710.01992>.
55. S. Farsiu, S. J. Chiu, R. V. O'Connell, F. A. Folgar, E. Yuan, J. A. Izatt, and C. A. Toth; Age-Related Eye Disease Study 2 Ancillary Spectral Domain Optical Coherence Tomography Study Group, "Quantitative classification of eyes with and without intermediate age-related macular degeneration using optical coherence tomography," *Ophthalmology* **121**(1), 162–172 (2014).
56. L. Fang, D. Cunefare, C. Wang, R. H. Guymer, S. Li, and S. Farsiu, "Automatic segmentation of nine retinal layer boundaries in OCT images of non-exudative AMD patients using deep learning and graph search," *Biomed. Opt. Express* **8**(5), 2732–2744 (2017).
57. K. Dabov, A. Foi, V. Katkovnik, and K. Egiazarian, "Image denoising by sparse 3-D transform-domain collaborative filtering," *IEEE Trans. Image Process.* **16**(8), 2080–2095 (2007).
58. L. Fang, S. Li, R. P. McNabb, Q. Nie, A. N. Kuo, C. A. Toth, J. A. Izatt, and S. Farsiu, "Fast acquisition and reconstruction of optical coherence tomography images via sparse representation," *IEEE Trans. Med. Imaging* **32**(11), 2034–2049 (2013).