Date of publication xxxx 00, 0000, date of current version xxxx 00, 0000.

Digital Object Identifier 10.1109/ACCESS.2020.Doi Number

Axial super-resolution study for optical coherence tomography images via deep learning

ZHUOQUN YUAN, DI YANG, HONGMING PAN, AND YANMEI LIANG

Institute of Modern Optics, Nankai University, Tianjin Key Laboratory of Optoelectronic Sensor and Sensing Network Technology, Tianjin Key Laboratory of Micro-scale Optical Information Science and Technology, Tianjin, 300350, China

Corresponding author: Yanmei Liang (e-mail: ymliang@nankai.edu.cn).

This work was supported in part by the National Natural Science Foundation of China (NSFC) under Grant No. 61875092, in part by the Science and Technology Support Program of Tianjin under Grant No. 17YFZCSY00740, and in part by Beijing-Tianjin-Hebei Basic Research Cooperation Special Program under Grant No. 19JCZDJC65300.

ABSTRACT Optical coherence tomography (OCT) is a noninvasive, high resolution, and real-time imaging technology that has been used in ophthalmology and other medical fields. Limited by the point spread function of OCT system, it is difficult to optimize its spatial resolution only based on hardware. Digital image processing methods, especially deep learning, provide great potential in super-resolving images. In this paper, the matched axial low resolution (LR) and high resolution OCT image pairs from actual OCT imaging are collected to generate the dataset by our home-made spectral domain OCT (SD-OCT) system. Several methods are selected to super-resolve LR OCT images. It is shown from the experimental results that the residual-in-residual dense block network (RRDBNet) trained with different loss functions performs the best super-resolution for OCT images, and it is demonstrated from the preliminary results that deep learning methods have good generalization and robustness between OCT systems. We believe deep learning methods have broad prospects in improving the quality of OCT images.

INDEX TERMS Optical coherence tomography (OCT), axial super resolution, deep learning

I. INTRODUCTION

For medical images, the resolution is one of the critical parameters, which determines what scale of microstructure can be distinguished. As a noninvasive and real-time imaging technology, the resolution of optical coherence tomography (OCT) is 1-2 orders of magnitude higher than those of X-ray computed tomography (X-CT), magnetic resonance imaging (MRI), and ultrasound imaging. Because of its great potential clinical values, the performance of OCT has been developed rapidly in nearly thirty years, especially in the aspect of axial resolution. Various sources, ultra-wide spectrum Ti: Sapphire laser [1], [2], super-luminescent diode (SLD) [3], and supercontinuum (SC) light source [4]-[7], have been thoroughly studied, which made its axial resolution close to 2-3 μm.

Improvement of hardware is expensive and time-consuming. Digital signal or image processing methods provide alternative and relatively cheaper solutions, and some signal processing methods, such as deconvolution [8], [9], spectrum-shaping [10], [11], and spectral estimation [12], have been proposed to optimize the OCT images. Although the qualities of OCT images can be improved to a certain extent, these processing methods cannot upgrade the axial resolution beyond the theoretical limit because there is no prior knowledge.

In 2013, L. Fang *et al.* studied reconstruction of OCT images by sparse representation [13] and they further added a segmentation step in their methods when reconstructing images [14]. The nonlocal weighted sparse representation (NWSR) method [15] was presented to exploit information from noisy and denoised patches' representations to reconstruct images. Unfortunately, their improvement of resolution was not obvious and processing speed was not fast enough to be applied in real time.

With the development of deep learning methods, many kinds of networks have been applied to image recognition [16], [17], image denoising [18], [19], etc. Convolutional neural network (CNN) was firstly used for super-resolution (SR) image reconstruction in 2014 [20], and then different networks were proposed in reconstructing image details [21], especially for natural images and face images [22]-[24]. For medical imaging, such as X-CT [25], [26], MRI [27], [28], and ultrasound imaging [29], [30], some SR methods based on deep learning have also been proposed, which improved their spatial resolution effectively.

In 2019, Y. Huang *et al.* [31] proposed a generative adversarial network-based approach to denoise and super-resolve OCT images simultaneously. In 2020, V. Das *et al.* [32] proposed an unsupervised framework by using the generative adversarial network (GAN) to perform fast and reliable SR image reconstruction without the requirement of aligned low-resolution (LR) - high-resolution (HR) pairs. However, these SR reconstructions for OCT images were all based on existing datasets and assumed a simple and uniform degradation (i.e., bicubic degradation), which is inconsistent with the actual degradation in OCT imaging. In addition, bicubic interpolation is the extension of cubic interpolation for interpolating data points on a two-dimensional regular grid. The bicubic interpolation is carried out based on two correlated dimensional data. For natural images or microscopic images, their resolutions in two dimensional images are the same, or we can say they are correlated, so it is appropriate to perform bicubic interpolation. However, the axial and transversal resolutions of OCT system are independent and its B-scan image is composed of many A-scan signals. If the final destination of super-resolution study is to enhance axial resolution of OCT system, it is inappropriate to do bicubic down-sampling for OCT images in principle.

In this paper, the registered LR-HR OCT image pairs were obtained based on the actual OCT axial resolution degradation to generate the dataset by our home-made spectral domain OCT (SD-OCT) system. Then, residual-in-residual dense block network (RRDBNet) [33] was updated to carry out axial super-resolution reconstruction. RRDBNet was trained by two kinds of loss, namely RRDBNet with mixed loss (GAN-RRDB) and RRDBNet with L1-loss (L1-RRDB), respectively. It was shown that L1-RRDB and GAN-RRDB can reconstruct HR OCT images effectively. In the meantime, the super-resolution result for the image from another OCT system demonstrated the prediction capability of the network, which proved that deep learning has great potential in improving axial resolution of OCT images.

II. METHODS

A. COLLECTION OF LR AND HR OCT IMAGES

A two-dimensional OCT image is considered as the convolution of the original signal *f*(*x,z*) and the point spread function (PSF) *h*(*x,z*) of OCT system in the spatial domain [34].

$$g(x,z) = f(x,z) * h(x,z)\,. \tag{1}$$

Where the symbol '*' indicates the spatial convolution. It is shown that OCT image is mainly degraded by PSF *h*(*x,z*). It has been proven, for a given center wavelength of light source, the axial and transversal PSFs are independent and not affected by each other [8], [34]. Therefore, for degraded OCT images, they can be improved based on two independent dimensions.

The axial PSF *h*(*z*) of OCT system can be easily obtained from the inverse Fourier transform of the power spectral density (PSD) of the light source, and its axial resolution Δz is mainly determined by the coherence length of its light source. Δz is given by the following equation when the spectrum of light source is Gaussian type,

$$\Delta z = \frac{2\ln 2}{\pi}\frac{\lambda_0^2}{\Delta\lambda}\,. \tag{2}$$

Where, λ_0 is the center wavelength of the light source, $\Delta\lambda$ is its 3dB bandwidth or the full width half maximum (FWHM). As shown in Eq. (2), we can randomly adjust the axial resolution by truncating the spectrum of the light source with different Gaussian windows digitally.

The interference signals of the sample arm and the reference arm were firstly collected by a home-made SD-OCT system [35] in our study. After removing the background spectrum in SD-OCT, the interferogram was multiplied by Gaussian windows of two different bandwidths to generate LR and HR signals. Then, these signals were mapped to the wave number domain and Fourier transform was performed to obtain axial LR and HR intensity images, respectively.

The LR and HR images are generated after interference signals acquisition, so they have the same field of view, ensuring the alignment without manual registration.

B. SUPER-RESOLUTION NETWORK

CNN has ultra-strong learning ability and can learn an end-to-end mapping between LR and HR images directly. The network with deep layers and complex structure was selected to reconstruct super-resolved OCT images in our study. The layout of the network is shown in Fig.1. Two kinds of networks with the same generator RRDBNet are trained with different losses, namely RRDBNet with mixed loss (GAN-RRDB) and RRDBNet with L1-loss (L1-RRDB), respectively. GAN-RRDB is composed of the generator and the discriminator. The generator is trained to generate super-resolved images and the discriminator is trained to distinguish super-resolved images from real HR images. In GAN-RRDB, the generator is trained with mixed loss functions between HR and SR images, including adversarial loss, perceptual loss, and L1 loss. GAN-RRDB alternately update the generator and the discriminator until the iterations reach the set value N. As shown by the red dashed lines in Fig. 1, GAN-RRDB is simplified to L1-RRDB when only L1 loss is considered. At this time, only generator needs to be trained, so L1-RRDB is easier to be trained than GAN-RRDB.

We used the generator of the Enhanced Super-Resolution Generative Adversarial Networks (ESRGAN) [33], that is, RRDBNet, which is composed of multi-layer residual network and dense connection. As shown in Fig. 2, RRDBNet can be decomposed into three parts: a single convolution layer, the layer of RRDB blocks, and a reconstruction layer. The layer of RRDB blocks is the residual-in-residual structure, where residual learning is carried out in different levels. It consists of 64 RRDB blocks, each of which has 23 dense blocks with skip connection.

Each dense block has five convolution layers with 3×3 filter kernels and they are used as the basic residual block in the main path to increase the network capacity. Each convolution layer in the dense block is followed by a Leaky ReLU (LReLU) layer except the last one. The reconstruction layer is composed of two convolution layers with 3×3 filter kernels and one LReLU layer. Please note the upper-sampling part is removed because our LR and HR images have the same size.

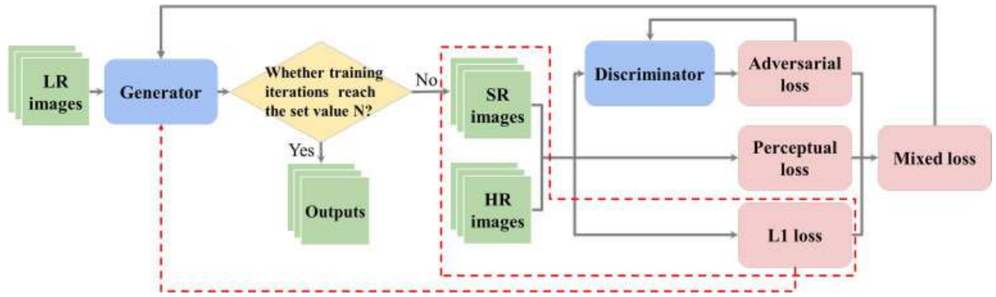

FIGURE 1. Layout of the deep learning network.

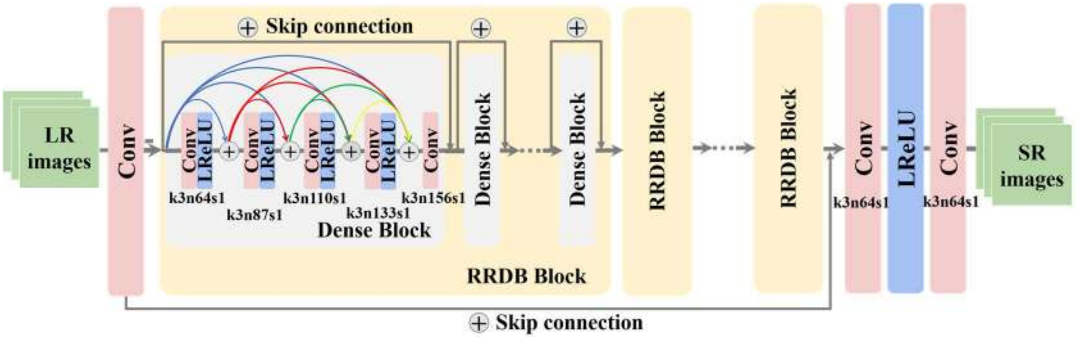

FIGURE 2. Architecture of the generator.

Low level features are obtained after LR images are input in the single convolution layer, and then they are input to the layer of RRDB blocks. Extracted high level features from the layer of RRDB blocks are added with the low level features and input to the reconstruction layer. Finally, the SR images are output.

Following the work of C. Ledig *et al.* [36], a discriminator network is further defined, whose architecture is illustrated in Fig. 3. It contains eight convolution layers with 3×3 filter kernels. Each convolution layer is followed by a batch-normalization (BN) layer and an LReLU layer except the first layer. After the convolution layers are two linear layers and one LReLU layer. Generator outputs are input to the convolution layers. Feature maps are generated from convolution layers and input to two linear layers to distinguish SR images from real HR images. The corresponding kernel size(k), number of feature maps(n), and stride(s) of each convolution layer are given in Figs. 2 and 3, respectively.

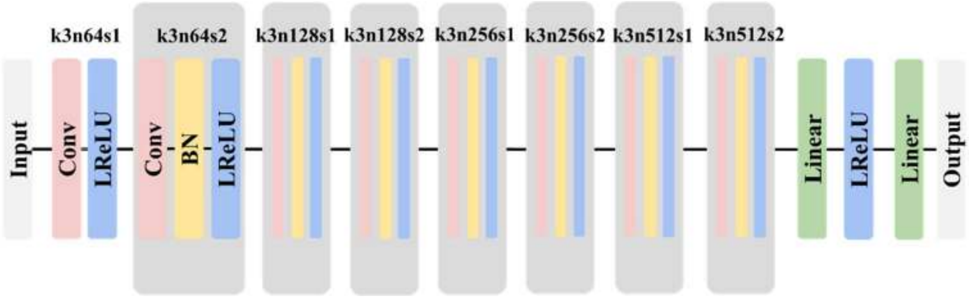

FIGURE 3. Architecture of the discriminator.

The discriminator is used to predict the probability that a real HR image I_{HR} is more realistic than an SR image I_{SR} [41].

$$\begin{array}{l} D_{real} = \sigma(C(I_{HR}) - E_{I_{SR}}[C(I_{SR})]) \longrightarrow 1 \ \text{ if } I_{HR} \text{ is more realistic than } I_{SR} \\ D_{fake} = \sigma(C(I_{SR}) - E_{I_{HR}}[C(I_{HR})]) \longrightarrow 0 \ \text{ if } I_{SR} \text{ is less realistic than } I_{HR} \end{array} . \qquad (3)$$

Where σ is the sigmoid function, $C(x)$ is the output of the non-transformed discriminator, and $E_{I_{SR}}[\cdot]$ and $E_{I_{HR}}[\cdot]$ represent the operation of taking average for all super-resolved data and high-resolution data in the mini-batch, respectively. The adversarial loss for discriminator is further calculated as:

$$L_{Adv}^{D} = -\mathrm{E}_{I_{HR}}[\log(\mathrm{D}_{real})] - \mathrm{E}_{I_{SR}}[\log(1-(\mathrm{D}_{fake}))] . \quad (4)$$

The adversarial loss for generator is in a symmetrical form:

$$L_{Adv}^{G} = -\mathrm{E}_{I_{SR}}[\log(\mathrm{D}_{fake})] - \mathrm{E}_{I_{HR}}[\log(1-(\mathrm{D}_{real}))] . \quad (5)$$

Perceptual loss measures the semantic differences between images by using a pre-trained image classification network. Here, a pre-trained VGG19 network described by K. Simonyan *et al.* [16] is used to extract features of the perceptual domain. VGG19 is a CNN network consisting of 16 convolution layers and 3 fully connection layers. The perceptual loss is defined as:

$$L_{VGG/i,j} = \frac{1}{w_{i,j} h_{i,j}} \sum_{x=1}^{w_{i,j}} \sum_{z=1}^{h_{i,j}} \left(\varphi_{i,j}(I_{SR})_{x,z} - \varphi_{i,j}(I_{HR})_{x,z} \right)^2 . \quad (6)$$

Where I_{SR} and I_{HR} are the intensity of SR and HR images, respectively. $\varphi_{i,j}$ indicates the feature map obtained by the *j-th* convolution before the *i-th* maxpooling layer within the VGG19 network. $w_{i,j}$ and $h_{i,j}$ are width and height of VGG19 feature map, respectively. The feature map $\varphi_{2,2}$ contains low level information [33], so $L_{VGG/2,2}$ defined by $\varphi_{2,2}$ can be used as the perceptual loss.

L1 loss, also known as absolute error loss, is a pixel-wise loss. It is calculated as follows:

$$L_{pixel\_l1}(I_{SR}, I_{HR}) = \frac{1}{hw} \sum_{x,z} \left| I_{SR}(x,z) - I_{HR}(x,z) \right| . \quad (7)$$

Where w and h are width and height of the image, respectively.

Finally, the total mixed loss is given by:

$$L = mL_{pixel\_l1} + nL_{VGG/2,2} + \eta L_{Adv}^{G} . \quad (8)$$

Where m, n and η are weighted parameters to control the trade-off among the three losses.

C. QUANTITATIVE METRICS

Four commonly used quantitative metrics, peak signal-to-noise ratio (PSNR), structural similarity index measure (SSIM), multi-scale-structural similarity index measure (MSSSIM), and the subjective evaluation mean opinion score (MOS) are used to evaluate the SR performance of the proposed algorithms quantitatively.

PSNR is the most common and widely used objective evaluation index, which is based on the error between corresponding pixels of two images, and does not take into account the human visual characteristics.

$$\begin{aligned} MSE &= \frac{1}{hw} \sum_{x=1}^{h} \sum_{z=1}^{w} (I_{SR}(x,z) - I_{HR}(x,z))^2 \\ PSNR &= 10 log_{10} \left(\frac{MAX(I)^2}{MSE} \right) \end{aligned} . \quad (9)$$

Where $MAX(I)$ represents the theoretical maximum of the pixel value in image I.

SSIM attempts to explain the texture change between two images by calculating the similarity from the aspects of luminance, contrast, and structure.

$$SSIM(I_{SR}, I_{HR}) = l(I_{SR}, I_{HR}) \cdot c(I_{SR}, I_{HR}) \cdot s(I_{SR}, I_{HR}) . \quad (10)$$

Where l, c, and s are functions of luminance, contrast, and structure, respectively. They are given as follows:

$$l\left(I_{SR}, I_{HR}\right) = \frac{2\mu_{I_{SR}}\mu_{I_{HR}} + C_1}{\mu_{I_{SR}}^2 + \mu_{I_{HR}}^2 + C_1} , \quad (11)$$

$$c\left(I_{SR}, I_{HR}\right) = \frac{2\sigma_{I_{SR}}\sigma_{I_{HR}} + C_2}{\sigma_{I_{SR}}^2 + \sigma_{I_{HR}}^2 + C_2} , \quad (12)$$

$$s\left(I_{SR}, I_{HR}\right) = \frac{\sigma_{I_{SR} I_{HR}} + C_3}{\sigma_{I_{SR}}\sigma_{I_{HR}} + C_3} . \quad (13)$$

Here $\mu_{I_{SR}}$, $\sigma_{I_{SR}}$, and $\sigma_{I_{SR}I_{HR}}$ are the mean of I_{SR}, the variance of I_{SR}, and the covariance of I_{SR} and I_{HR}. $C_1 = (K_1 L)^2$, $C_1 = (K_2 L)^2$, and $C_3 = C_2/2$. L is the dynamic range of the image. K_1 and K_2 are constants.

SR and HR images are iteratively down sampled with a down-sampling filter by a factor of 2 when calculating MSSSIM. The original image is regarded as scale 1, and the highest scale as scale M, which is obtained after M-1 iterations. At the j-th scale, the contrast function and the structure function are calculated and denoted as $c_j(I_{SR}, I_{HR})$ and $s_j(I_{SR}, I_{HR})$, respectively. The luminance function is computed only at scale M and is denoted as $l_M(I_{SR}, I_{HR})$.The MSSSIM [37] is calculated by combining the measurement at different scales.

$$MSSSIM(I_{SR}, I_{HR}) = l_M(I_{SR}, I_{HR}) \prod_{j=1}^{M} c_j(I_{SR}, I_{HR}) \cdot s_j(I_{SR}, I_{HR}) . \quad (14)$$

MOS is a subjective evaluation method and the observers subjectively score the image quality, which can directly reflect the image visual perception.

III. RESULTS

A. DATASET GENERATION

All images were collected by our home-made SD-OCT system [35]. A super luminescent diode (SLD) (BLM2-D, Superlum) with the center wavelength of 840 nm and the bandwidth of 100 nm is used in the SD-OCT system, whose axial resolution is ~3.4 μm (in air) and transverse resolution is ~13 μm. After truncating the spectrum of the light source with Gaussian windows, the axial resolutions of LR and HR images are ~5 μm and ~18 μm, respectively. Noise in HR images was removed by multi-frame averaging. Different from the previous studies [31], our LR and HR images have the same size, both of which are 2048 × 1000 pixels × pixels (height × width).

100 LR-HR zebrafish OCT images were totally collected, and they were divided into patches with the size of 80 × 80 pixels × pixels to generate the data set. Some patches in deep tissue were deleted because they lack effective information due to the attenuation of biological tissue to light. The data set consisting of 7,000 patches was finally obtained. Finally, training set, verification set, and test set were distributed according to a ratio of 3:1:1.

B. SUPER-RESOLUTION RESULTS

Two kinds of RRDBNet with different loss functions were trained. Both networks were optimized by using Adam algorithm [38], the hyperparameters of which were empirically set as α =0, β_1= 0.9, β_2= 0.99. The training iterations N was set as 150000. The learning rate dropped by step decay. The loss weighted parameters were empirically set as m=0.01, n=1and η=0.005 for training GAN-RRDB. K_1 and K_2 were set to 0.01 and 0.03 when calculating SSIM. The training and test were performed by Pytorch on a server with 64 GB of RAM and an NVIDIA TITAN RTX graphics processing unit (GPU). The loss-iterations curves of L1-RRDB and GAN-RRDB are shown in Fig. 4, in which indicate both models can converge after 150,000 iterations.

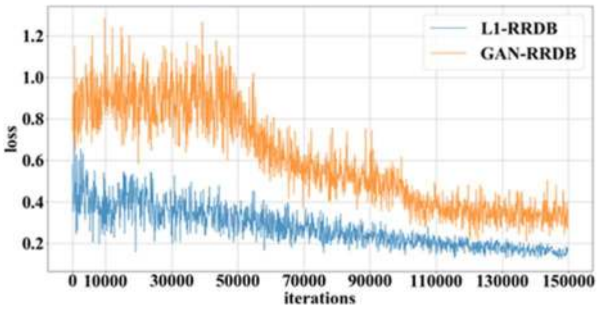

FIGURE 4. The loss-iterations curves of training L1-RRDB and GAN-RRDB.

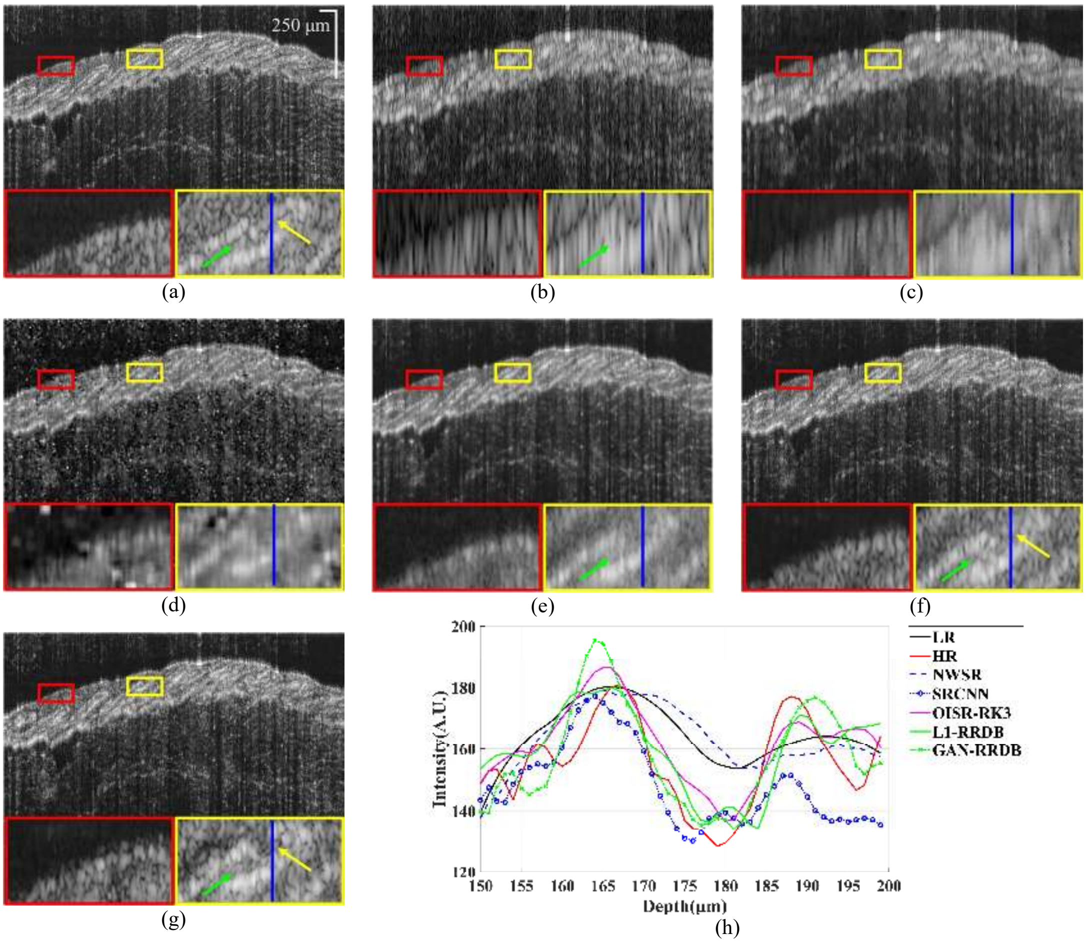

FIGURE 5. SR results of zebrafish OCT images. (a) HR image. (b) LR image. (c)-(g) SR reconstructed images by NWSR, SRCNN, OISR-RK3, L1-RRDB and GAN-RRDB, respectively. (h) A-line profiles of the lateral position pointed out by the blue lines in Figs. 5 (a)-(g). Scale bar in each image is 250 μm.

The SR results are shown in Fig. 5. As a comparison, the SR results of nonlocal weighted sparse representation (NWSR) [15], CNN-based method SRCNN [20], and the ordinary differential equation (ODE) Runge-Kutta (RK) method-inspired design network OISR-RK3 [39] are also shown in Fig. 5. Fig. 5(a) is a zebrafish dorsal HR OCT image. Fig. 5(b) is its corresponding LR image. Fig. 5(c) is the SR image reconstructed by NWSR. Fig. 5(d) is the SR result of SRCNN. Fig. 5(e) is the SR result of OISR-RK3. Figs. 5(f) and 5(g) are the reconstructed SR images of L1-RRDB and GAN-RRDB, respectively. Sub-images at the lower part of Figs. 5(a) - (g) are the magnified views of the regions selected.

As shown by the green arrow in Fig. 5(a), the outline of the scale in zebrafish skin can be clearly seen in the HR image, but cannot be distinguished in the LR image [Fig. 5(b)]. As shown in Fig. 5(c), NWSR cannot reconstruct the outline of the scale of zebrafish. Generally, NWSR need to learn the self-similarity information of the images. Because there is big difference between the LR OCT image and its corresponding HR OCT image with the axial resolutions of 18 μm and 5 μm, respectively, it is difficult to extract self-similarity information in them, which induce NWSR cannot improve the resolution of LR images in our study. SRCNN [Fig. 5(d)] can reconstruct detailed texture of the scale. However, many bright noise spots are introduced in the image reconstruction process. OISR-RK3 [Fig. 5(e)] can recover the texture, but the reconstructed image is not clear enough. Compared with NWSR and SRCNN, GAN-RRDB and L1-RRDB can clearly reconstruct the outline of the scale. A perceptually more realistic SR image by GAN-RRDB [Fig. 5(g)] is obtained than that of L1-RRDB [Fig. 5(f)]. However, as shown by the yellow arrow in Fig. 5(g), it has a few artifacts different from the HR image. Compared Figs. 5(f) and 5(g), better texture details are reconstructed by L1-RRDB, but the general contrast of the SR image is better by GAN-RRDB. In addition, as shown by the red rectangular boxes in Figs. 5(f) and 5(g), noise in the background is somewhat suppressed after image reconstruction.

A-line profiles pointed out by the blue lines in Figs. 5(a)-(g) are shown in Fig. 5(h) to further indicate the quantitative effect of different methods in improving the axial resolution. The narrower the peak width, the better the image axial resolution. The HR signal (the red curve) has the narrowest peak width, and its contrast of intensity is much better than that of the LR signal (the black curve). The LR signal has the widest peak width, and its contrast of intensity is the worst. The peak width and the contrast of intensity of NWSR (the blue dashed curve) are very close to those of LR signal, which demonstrates it is incapable to improve axial resolution. The peak widths of SRCNN (the blue dotted curve), OISR-RK3(the purple curve), L1-RRDB (the green curve), and GAN-RRDB (the green dashed curve) are narrower than that of LR signal, which means they can realize super resolution. Among the four methods, GAN-RRDB has the best contrast of intensity.

The quantitative evaluation parameters based on the test dataset are shown in Table I, where the best result in each metric is shown in bold. Specifically, we asked 21 raters to assign an integral score from 1 (bad quality) to 5 (excellent quality) to the super-resolved images, and their average score was MOS value. There is no doubt that LR images have the worst parameters. LR images are smoothed by NWSR, so SR images by NWSR have better evaluation results than LR images. Although SRCNN can perform SR, the bright noise spots generated by SRCNN may destroy texture details of the image, therefore, its quantitative evaluation results are not as good as NWSR. Compared with NWSR and SRCNN, OISR-RK3, L1-RRDB and GAN-RRDB have the better quantitative evaluation. Based on the four indicators, we think that L1-RRDB and GAN-RRDB have better trade-off between evaluation indicators and visual perception than OISR-RK3, which is consistent with the results of Figs. 5(f) and 5(g).

TABLE I
Quantitative evaluation results of different methods

	LR	**NWSR**	**SRCNN**	**OISR-RK3**	**L1-RRDB**	**GAN-RRDB**
PSNR	24.36	26.60	25.00	**28.58**	27.84	27.05
SSIM	0.3585	0.6206	0.6060	0.6822	**0.6853**	0.6418
MSSSIM	0.6621	0.8071	0.7195	**0.8509**	0.8398	0.8302
MOS	1.86	1.43	2.62	3.33	3.67	**4.05**

Table II lists the running time of different methods. The results illustrate that the deep learning method is much faster than the NWSR method. Among the deep learning algorithms, the running time increases with the increase of the network capacity. Though L1-RRDB and GAN-RRDB have the complex network structures, resulting in the longer running time than SRCNN, their running time is short enough not to affect OCT imaging, which means that it can be applied with OCT signal collecting to achieve real-time SR imaging.

TABLE II
Average running time of different methods

	NWSR	**SRCNN**	**OISR-RK3**	**L1-RRDB**	**GAN-RRDB**
Running Time/s	65.028	0.018	0.170	0.120	0.128

C. ABLATION STUDY

In order to further verify the impact of the hyper parameters (loss weighted parameters) and the basic components in the network during SR training, we conducted ablation experiments on the OCT dataset.

The impact of different loss weighted parameters in the GAN-RRDB was firstly discussed. Fig. 6 shows the SR results of ablation study on loss weighted parameters. Quantitative evaluation results are shown in Table III, where the best result in each metric is shown in bold. m=0 in Eq. (8) means that L1-loss was not considered during training and n=0 means that perceptual loss was not introduced during training.

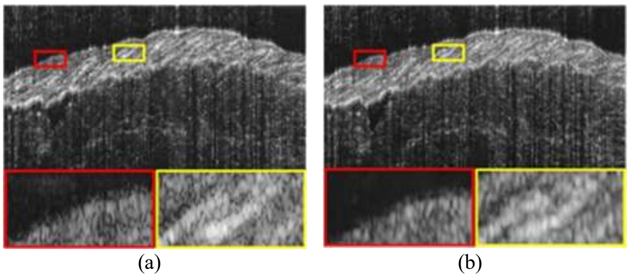

FIGURE 6. SR results of ablation study on loss weighted parameters. (a) SR reconstructed images by GAN-RRDB when m is 0 in Eq. (8). (b) SR reconstructed images by GAN-RRDB when n is 0 in Eq. (8).

TABLE III
Quantitative evaluation results of ablation study on loss weighted parameters

	m=0	*n*=0
PSNR	26.27	**27.91**
SSIM	0.5989	**0.6919**
MSSSIM	0.8227	**0.8355**
MOS	**3.86**	3.43

Eqs (6) and (7) show that L1-loss is set to obtain higher PSNR and perceptual loss is set to better visual perception, which is verified by the quantitative evaluation results of ablation study on the loss weighted parameters. SR images by GAN-RRDB with perceptual loss is better in visual perception while GAN-RRDB with L1 loss has higher PSNR, SSIM, and MSSSIM.

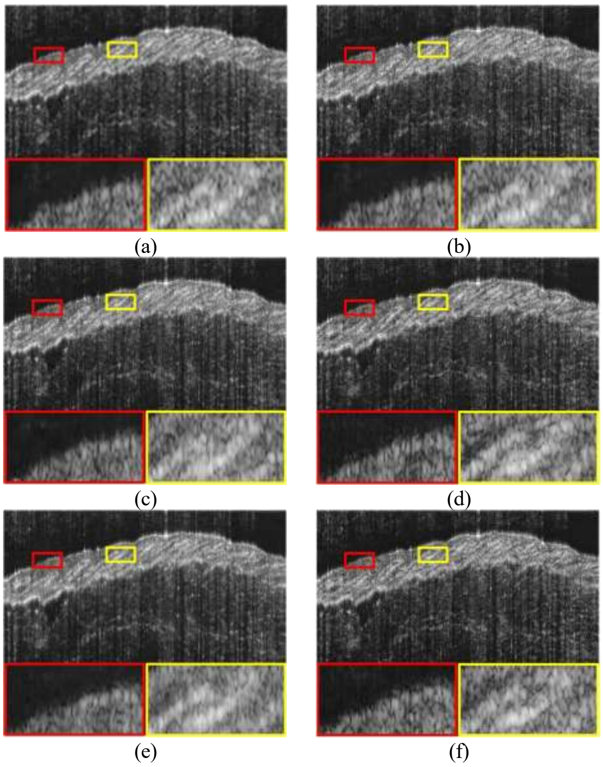

FIGURE 7. SR results of ablation study on different numbers of network components. (a) and (b) are SR images by L1-RRDB and GAN-RRDB with *nr*=48 and *nd*=13, (c) and (d) are SR images by L1-RRDB and GAN-RRDB with *nr*=56 and *nd*=18, and (e) and (f) are SR images by L1-RRDB and GAN-RRDB with *nr*=72 and *nd*=28, respectively.

We further studied the impact of number of network components on image reconstruction by training L1-RRDB and GAN-RRDB with different network capacities. Fig 7 shows SR results of ablation study on number of network components. Figs. 7(a), (c) and (e) are SR images by L1-RRDB with different components numbers. Figs. 7(b), (d) and (f) are SR reconstructed images by GAN-RRDB with different components numbers. We use *nr* for the number of RRDB blocks and *nd* for the number of dense blocks in each RRDB block. Figs. 7(a) and (b) are the results of nr=48 and *nd*=13. Figs. 7(c) and (d) are the results of *nr*=56 and *nd*=18. Figs. 7(e) and (f) are the results of *nr*=72 and *nd*=28.

The quantitative evaluation results of L1-RRDB and GAN-RRDB are shown in Table IV and Table V, respectively.

TABLE IV
Quantitative evaluation results of ablation study on network component number of L1-RRDB

nr	**48**	**56**	**72**
nd	**13**	**18**	**28**
PSNR	27.54	27.71	**28.16**
SSIM	0.699	0.688	**0.701**
MSSSIM	0.837	0.838	**0.852**
MOS	3.52	3.52	**3.81**

TABLE V
Quantitative evaluation results of ablation study on network component number of GAN-RRDB

nr	**48**	**56**	**72**
nd	**13**	**18**	**28**
PSNR	26.19	26.95	**27.09**
SSIM	0.625	0.581	**0.639**
MSSSIM	0.797	0.821	**0.852**
MOS	3.47	**3.86**	**3.86**

Empirically speaking, the larger the network capacity, the stronger the fitting ability of the network. As shown in Fig. 7, Tables IV and V, it can be seen that deeper networks perform better on evaluation indicators and the visual perception. However, deeper networks will increase the cost of training and test. Therefore, the appropriate network capacity should be selected according to the actual situation.

D. EXPERIMENTAL STUDY ON SS-OCT images

By learning the micro-structure map between LR and HR images, deep learning methods can super-resolve LR images and obtain the SR images similar to HR images. Further, we hope the trained networks have the prediction capability, then it will enhance axial resolution of OCT system without increasing cost.

To validate its feasibility, we tested OCT images collected by our home-made SS-OCT system, whose details were described in our previous papers [40]-[42]. Its axial resolution is ~14.6 μm in air based on a swept source (Santec, HSL-20-100-B) with the broadband spectrum of ~87 nm at 1300 nm, and its transversal resolution is ~17 μm.

HR SD-OCT images and SS-OCT images at the same position on the dorsal of the same zebrafish were collected, which are shown in Figs. 8(a) and (b), respectively. The SR results by L1-RRDB and GAN-RRDB of the SS-OCT image are displayed in Figs. 8(c) and (d). Since SS-OCT and SD-OCT systems have different field of view, it was necessary to match their images for comparison. The myotome of the zebrafish indicated by the red arrow in the red enlarged area in Fig. 8 was used as a feature to register images from different systems. The positions of the myotome in the two images [Figs. 8(a) and (b)] were matched, indicating that the images of SS-OCT and SD-OCT were registered.

As can be seen in the yellow magnified views, the multilayer structure of the skin [43] can be clearly shown in the HR SD-OCT image [Fig. 8(a)] while it is indistinguishable in the SS-OCT image [Fig. 8(b)]. Figs. 8(c) and (d) are its SR results of L1-RRDB and GAN-RRDB, respectively. It can be seen that both networks can reconstruct the multilayer structure of the skin. The result of GAN-RRDB [Fig. 8(d)] is clearer than that of L1-RRDB [Fig. 8(c)]. Also, as shown by the yellow arrows, the penetration depth of 1300 nm SS-OCT is deeper than that of 840 nm SD-OCT, and details in the deep area can also be super-resolved by the methods.

Thus, the preliminary experiments on SS-OCT images demonstrated that deep learning methods have good generalization and robustness and can greatly improve the axial resolution of the SS-OCT image. More importantly, as shown in the enlarged yellow rectangles, the multilayer micro-structure of skin in Fig. 8(d) is even clearer than that of Fig. 8(a). The network was trained based on the images from 18 μm to 5 μm. When the axial resolution of the LR image is better than 18 μm, the better SR image is obtained.

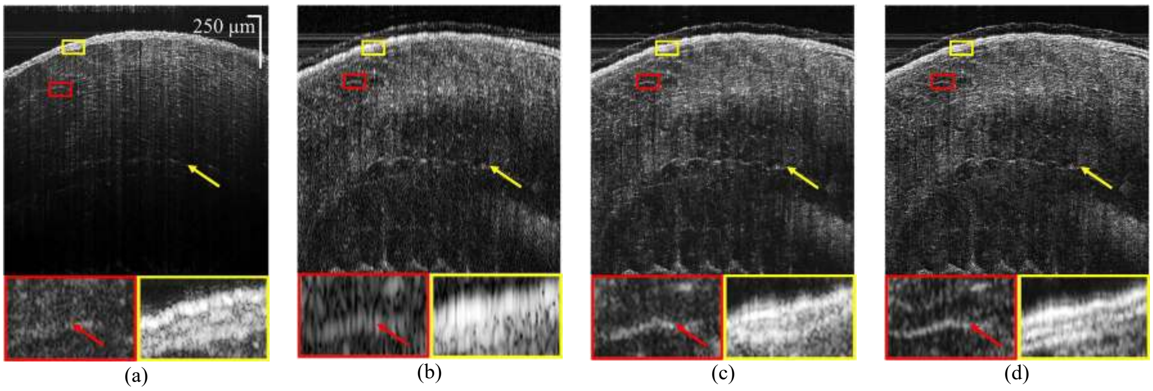

FIGURE 8. SR processing results for a SS-OCT image. (a) is an HR SD-OCT image. (b) is the SS-OCT image. (c)-(d) are the SR results of (b) by L1-RRDB and GAN-RRDB, respectively. Scale bar in each image is 250 μm.

IV. DISCUSSIONS

Based on the actual OCT imaging, we collected HR and LR OCT images in our study. It was proven from visual and quantitative analysis that CNN-based deep learning algorithms can obtain SR images similar to HR images with high signal-to-noise ratio. We primarily confirmed that deep learning can effectively improve OCT axial resolution.

In order to further compare the difference among actual LR OCT images and the bicubic downsampled images, a zebrafish dorsal OCT HR image with ~5 μm axial resolution, its axial downsampled LR images with 4× and 12× bicubic interpolation, and the actual LR image with ~18 μm axial resolution are given in Figs. 9(a) - (d), respectively. For easy to compare, the axial bicubic downsampled LR images are upsampled to the size of the actual LR images. Images in the lower left corner of Figs. 9(a) - (d) are the magnified views of the regions selected.

As shown in Fig. 9(a), rich texture details can be seen in the HR image for the dorsal of zebrafish. The bicubic down-sampled image [Fig. 9(b)] with downscale factor 4 still retains high-frequency information similar to its HR image. In terms of texture details, we find the bicubic downsampled LR image of 12 times [Fig. 9(c)] is closer to the actual LR OCT image [Fig. 9(d)].

The compared result shows that the bicubic downsampled LR image of less than 8× is much better than the degradation from 5 to 18 μm in actual OCT imaging in preserving the sample micro-structure and high-frequency information, which indicates that the actual OCT degradation is much more complicated than the assumed bicubic degradation model.

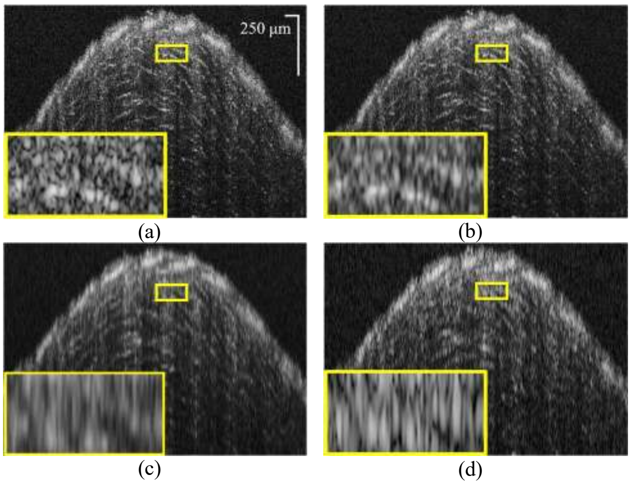

FIGURE 9. Zebrafish OCT images of different axial resolutions. (a) is the HR image. (b) and (c) are axial-downsampled LR images with downscale 4 and 12, respectively. (d) is the actual axial low-resolution image. Scale bar in each image is 250 μm.

Our research proved deep learning can accomplish the axial super-resolution for the actual OCT imaging, and primarily validated its feasibility in improving the resolution of another OCT system. We will further pursue to achieve higher resolution on our available SD-OCT system based on deep learning.

In addition, the axial and transversal PSFs of OCT system are not affected by each other, so the axial resolution and transversal resolution are independent. We only focus on improving the axial resolution of OCT without considering the transversal resolution in this study. In fact, high transversal resolution often leads to the reduction of the depth of focus (DOF). Therefore, our follow-up study will further explore how to achieve high transversal resolution and large DOF simultaneously in OCT system based on deep learning methods.

In our experiments, only four neural networks were trained to perform super-resolution OCT images. We will also test more deep learning networks and algorithms to obtain better SR results.

V. CONCLUSIONS

Based on actual OCT imaging, we collected axial LR and HR OCT images in this paper. Compared with NWSR, we found deep learning methods have better SR effect. Among networks of SRCNN, OISR-RK3, L1-RRDB, and GAN-RRDB, the last two ones have the best SR results. It was proven that deep learning methods have great potential in improving the resolution of OCT images, and have good generalization and robustness.

References

[1] J. Xi, A. Zhang, Z. Liu, W. Liang, L. Y. Lin, S. Yu, and X. D. Li, “Diffractive catheter for ultrahigh-resolution spectral-domain volumetric OCT imaging,” *Opt. Lett.*, vol. 39, no. 7, pp. 2016–2019, Apr. 2014.

[2] W. Yuan, R. Brown, W. Mitzner, L. Yarmus, and X. D. Li, "Super-achromatic monolithic microprobe for ultrahigh-resolution endoscopic optical coherence tomography at 800 nm," *Nat. Commun.*, vol. 8, no. 1, p. 1531, Nov. 2017D.

[3] D. Sen, E. D. SoRelle, O. Liba, R. Dalal, Y. M. Paulus, T.-W. Kim, D. M. Moshfeghi, and A. de la Zerda, "High-resolution contrast-enhanced optical coherence tomography in mice retinae," *J. Biomed. Opt.*, vol. 21, no. 6, p. 066002, Jun. 2016.

[4] W. Yuan, J. Mavadia-Shukla, J. Xi, W. Liang, X. Yu, S. Yu, and X. D. Li, "Optimal operational conditions for supercontinuum-based ultrahigh-resolution endoscopic OCT imaging," *Opt. Lett.*, vol. 41, no. 2, pp. 250-253, Jan. 2016.

[5] Y.-J. You, C. Wang, Y.-L. Lin, A. Zaytsev, P. Xue, and C.-L. Pan, "Ultrahigh-resolution optical coherence tomography at 1.3 µm central wavelength by using a supercontinuum source pumped by noise-like pulses," *Laser Phys. Lett.*, vol. 13, no. 2, p. 025101, Dec. 2015.

[6] X. Yao, Y. Gan, C. C. Marboe, and C. P. Hendon, "Myocardial imaging using ultrahigh-resolution spectral domain optical coherence tomography," *J. Biomed. Opt.*, vol. 21, no. 6, p. 061006, Jun. 2016.

[7] S. P. Chong, T. Zhang, A. Kho, M. T. Bernucci, A. Dubra, and V. J. Srinivasan, "Ultrahigh resolution retinal imaging by visible light OCT with longitudinal achromatization," *Biomed. Opt. Express*, vol. 9, no. 4, pp. 1477-1491, Apr. 2018.

[8] Y. Liu, Y. Liang, G. Mu, and X. Zhu, "Deconvolution methods for image deblurring in optical coherence tomography," *J. Opt. Soc. Amer. A. Opt. Image Sci.*, vol. 26, no. 1, pp. 72-77, Jan. 2009.

[9] S. A. Hojjatoleslami, M. R. N. Avanaki, and A. G. Podoleanu, "Image quality improvement in optical coherence tomography using Lucy–Richardson deconvolution algorithm," *Appl. Opt.*, vol. 52, no. 23, pp. 5663-5670, Aug. 2013.

[10] J. Gong, B. Liu, Y. L. Kim, Y. Liu, X. Li, and V. Backman, "Optimal spectral reshaping for resolution improvement in optical coherence tomography," *Opt. Express*, vol. 14, no. 13, pp. 5909-5915, Jun. 2006.

[11] Y. Chen, J. Fingler, and S. E. Fraser, "Multi-shaping technique reduces sidelobe magnitude in optical coherence tomography," *Biomed. Opt. Express*, vol. 8, no. 11, pp. 5267-5281, Nov. 2017.

[12] X. Liu, S. Chen, D. Cui, X. Yu, and L. Liu, "Spectral estimation optical coherence tomography for axial super-resolution," *Opt. Express*, vol. 23, no. 20, pp. 26521-26532, Oct. 2015.

[13] 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. Imag.*, vol. 32, no. 11, pp. 2034-2049, Nov. 2013.

[14] L. Fang, S. Li, D. Cunefare, and S. Farsiu, "Segmentation based sparse reconstruction of optical coherence tomography images," *IEEE Trans. Med. Imag.*, vol. 36, no. 2, pp. 407-421, Feb. 2017.

[15] A. Abbasi, A. Monadjemi, L. Fang, and H. Rabbani, "Optical coherence tomography retinal image reconstruction via nonlocal weighted sparse representation," *J. Biomed. Opt.*, vol. 23, no. 3, p. 036011, Mar. 2018.

[16] K. Simonyan, and A. Zisserman, "Very deep convolutional networks for large-scale image recognition," in *Proc. Int. Conf. Learn. Representations*, 2015.

[17] K. He, X. Zhang, S. Ren, and J. Sun, "Deep residual learning for image recognition," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jun. 2016, pp. 770–778.

[18] S. Lefkimmiatis, "Non-local color image denoising with convolutional neural networks," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jul. 2017, pp. 3587–3596.

[19] 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*, vol. 9, no. 12, pp. 6205-6221, Dec. 2018.

[20] C. Dong, C. C. Loy, K. He, and X. Tang, "Learning a deep convolutional network for image super-resolution," in *Proc. Eur. Conf. Comput. Vis. (ECCV)*, 2014, pp. 184–199.

[21] Z. Wang, J. Chen, and S. C. H. Hoi, “Deep learning for image super-resolution: A survey,” *IEEE Trans. Pattern Anal. Mach. Intell.*, Mar. 2020.

[22] J. Yamanaka, S. Kuwashima, and T. Kurita, "Fast and accurate image super resolution by deep CNN with skip connection and network in network," in *Proc. Int. Conf. Neural Inf. Process. (NIPS)*, 2017, pp. 217–225.

[23] Y. Zhang, Y. Tian, Y. Kong, B. Zhong, and Y. Fu, "Residual dense network for image super-resolution," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jun. 2018, pp. 2472-2481.

[24] Z. Zhang, Z. Wang, Z. Lin, and H. Qi, "Image super-resolution by neural texture transfer," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jun. 2019, pp. 7982-7991.

[25] C. You, G. Li, Y. Zhang, X. Zhang, H. Shan, S. Ju, Z. Zhao, Z. Zhang, W. Cong, M. W. Vannier, P. K. Saha, and G. Wang, "CT super-resolution GAN constrained by the identical residual and cycle learning ensemble (GAN-CIRCLE)," *IEEE Trans. Med. Imag.*, vol. 39, no. 1, pp. 188–203, Jan. 2020.

[26] J. Chi, Y. Zhang, X. Yu, Y. Wang, and C. Wu, "Computed tomography (CT) image quality enhancement via a uniform framework integrating noise estimation and super-resolution networks," *Sensors*, vol. 19, no. 15, p. 3348, Jul. 2019.

[27] A. S. Chaudhari, Z. Fang, F. Kogan, J. Wood, K. J. Stevens, E. K. Gibbons, J. H. Lee, G. E. Gold, and B. A. Hargreaves, "Super-resolution musculoskeletal MRI using deep learning," *Magn. Reson. Med.*, vol. 80, no. 5, pp. 2139-2154, Feb. 2018.

[28] C.-H. Pham, A. Ducournau, R. Fablet, and F. Rousseau, "Brain MRI super-resolution using deep 3D convolutional networks," in *Proc. Int. Symp. Biomed. Imag.* (*ISBI*), 2017, pp. 197-200.

[29] N. Zhao, Q. Wei, A. Basarab, D. Kouamé, and J.-Y. Tourneret, "Single image super-resolution of medical ultrasound images using a fast algorithm," in *Proc. Int. Symp. Biomed. Imag.* (*ISBI*), Prague, Czech Republic, Apr. 2016, pp 473-476.

[30] R. J. v. Sloun, O. Solomon, M. Bruce, Z. Z. Khaing, Y. C. Eldar, and M. Mischi, "Deep learning for super-resolution vascular ultrasound imaging," in *Proc. IEEE Int. Conf. Acoust. Speech Signal Process.* (*ICASSP*), May. 2019, pp. 1055-1059.

[31] Y. Huang, Z. Lu, Z. Shao, M. Ran, J. Zhou, L. Fang, and Y. Zhang, "Simultaneous denoising and super-resolution of optical coherence tomography images based on a generative adversarial network," *Opt. Express*, vol. 27, no. 9, pp. 12289-12307, Apr. 2019.

[32] 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.*, vol. 20, no. 15, pp. 8746-8756, Aug. 2020.

[33] X. Wang, K. Yu, S. Wu, J. Gu, Y. Liu, C. Dong, Y. Qiao, and C. C. Loy, "ESRGAN: enhanced super-resolution generative adversarial networks," in *Proc. Eur. Conf. Comput. Vis. Workshops (ECCVW)*, Sep. 2018, pp. 1-16.

[34] W. Drexler and J. G. Fujimoto, *Optical Coherence Tomography: Technology and Applications*, 2nd ed. Switzerland: Springer, 2015.

[35] D. Yang, M. Hu, M. Zhang, and Y. Liang, "High-resolution polarization-sensitive optical coherence tomography for zebrafish muscle imaging," *Biomed. Opt. Express*, vol. 11, no. 10, pp. 5618-5632, Oct. 2020.

[36] C. Ledig, L. Theis, F. Huszár, J. Caballero, A. Cunningham, A. Acosta, A. Aitken, A. Tejani, J. Totz, Z. Wang, and W. Shi, "Photo-realistic single image super-resolution using a generative adversarial network," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.* (*CVPR*), Jul. 2017, vol .2, no .3, pp. 4681-4690.

[37] Z. Wang, E. P. Simoncelli, and A. C. Bovik, "Multi-scale structural similarity for image quality assessment," in *Proc. IEEE 37th Conf. Signals Syst. Comput.*, 2003, pp. 1398-1402.

[38] D. P. Kingma and J. L. Ba, "Adam: a method for stochastic optimization," in *Proc. Int. Conf. Learn. Represent.*, 2015, pp. 1-41.

[39] X. He, Z. Mo, P. Wang, Y. Liu, M. Yang, J. Cheng, "ODE-Inspired Network Design for Single Image Super-Resolution." in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.* (*CVPR*), Jun. 2019, pp. 1732–1741.

[40] F. Hou, M. Zhang, Y. Zheng, L. Ding, X. Tang, and Y. Liang, "Detection of laser-induced bulk damage in optical crystals by swept-source optical coherence tomography," *Opt. Express*, vol. 27, no. 3, pp. 3698-3709, Feb. 2019.

[41] Z. Yang, J. Shang, C. Liu, J. Zhang, F. Hou, and Y. Liang, "Intraoperative imaging of oral-maxillofacial lesions using optical coherence tomography," *J. Innov. Opt. Health Sci.*, vol. 13, no. 2, p. 2050010, Feb. 2020.

[42] K. Li, W. Liang, Z. Yang, Y. Liang, and S. Wan, "Robust, accurate depth-resolved attenuation characterization in optical coherence tomography," *Biomed. Opt. Express*, vol. 11, no. 2, pp. 672-687, Feb. 2020.

[43] D. L. Guellec, G. Morvan-Dubois, and J.-Y. Sire, "Skin development in bony fish with particular emphasis on collagen deposition in the dermis of the zebrafish (Danio rerio)," *Int. J. Dev. Biol.*, vol. 48, no. 2, pp. 217-231, 2004