

PAPER • OPEN ACCESS

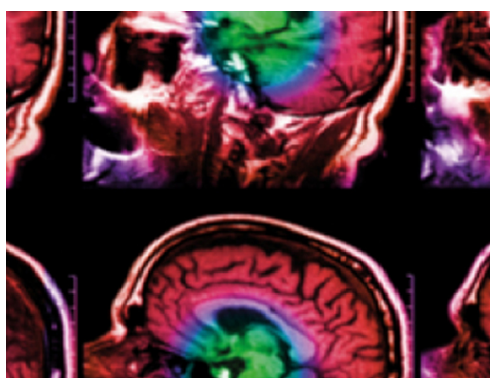
Imaging depth adaptive resolution enhancement for optical coherence tomography via deep neural network with external attention

To cite this article: Shangjie Ren *et al* 2021 *Phys. Med. Biol.* **66** 195006

View the [article online](#) for updates and enhancements.

You may also like

- [Ultrahigh resolution ex vivo ocular imaging using ultrashort acquisition time en face optical coherence tomography](#)
Kate Grieve, Gael Moneron, Arnaud Dubois et al.
- [Multiscale skin imaging in vivo using optical coherence tomography](#)
Xiaojun Yu, Hongying Tang, Chi Hu et al.
- [Complex-based OCT angiography algorithm recovers microvascular information better than amplitude- or phase-based algorithms in phase-stable systems](#)
Jingjiang Xu, Shaozhen Song, Yuandong Li et al.



IPEM | IOP

Series in Physics and Engineering in Medicine and Biology

Your publishing choice in medical physics,
biomedical engineering and related subjects.

Start exploring the collection—download the
first chapter of every title for free.

OPEN ACCESS



CrossMark

RECEIVED

19 May 2021

REVISED

9 August 2021

ACCEPTED FOR PUBLICATION

31 August 2021

PUBLISHED

24 September 2021




Original content from this work may be used under the terms of the [Creative Commons Attribution 4.0 licence](#).

Any further distribution of this work must maintain attribution to the author(s) and the title of the work, journal citation and DOI.



PAPER

Imaging depth adaptive resolution enhancement for optical coherence tomography via deep neural network with external attention

Shangjie Ren^{1,*} , Xiongri Shen¹, Jingjiang Xu², Liang Li^{3,*}, Haixia Qiu⁴, Haibo Jia^{5,6}, Xining Wu⁷, Defu Chen⁸ , Shiyong Zhao⁷, Bo Yu^{3,6}, Ying Gu^{4,9} and Feng Dong¹ 

¹ Tianjin Key Laboratory of Process Measurement and Control, School of Electrical and Information Engineering, Tianjin University, Tianjin, 300072, People's Republic of China

² School of Physics and Optoelectronic Engineering, Foshan University, Foshan, 528000, People's Republic of China

³ College of Intelligence and Computing, Tianjin University, Tianjin, 300072, People's Republic of China

⁴ Department of Laser Medicine, the First Medical Centre, Chinese PLA General Hospital, Beijing, 100853, People's Republic of China

⁵ Department of Cardiology, The 2nd Affiliated Hospital of Harbin Medical University, Harbin, 150081, People's Republic of China

⁶ The Key Laboratory of Myocardial Ischemia, Chinese Ministry of Education, Harbin, 150081, People's Republic of China

⁷ Tianjin Horimed Technology Co., Ltd., Tianjin, 300308, People's Republic of China

⁸ Institute of Engineering Medicine, Beijing Institute of Technology, Beijing, 100081, People's Republic of China

⁹ Precision Laser Medical Diagnosis and Treatment Innovation Unit, Chinese Academy of Medical Sciences, Beijing, 100000, People's Republic of China

* Authors to whom any correspondence should be addressed.

E-mail: rensjie@tju.edu.cn and liangli@tju.edu.cn

Keywords: optical coherence tomography angiography, deep neural network, image super-resolution, external attention, optical coherence tomography

Supplementary material for this article is available [online](#)

Abstract

Optical coherence tomography (OCT) is a promising non-invasive imaging technique that owns many biomedical applications. In this paper, a deep neural network is proposed for enhancing the spatial resolution of OCT *en face* images. Different from the previous reports, the proposed can recover high-resolution *en face* images from low-resolution *en face* images at arbitrary imaging depth. This kind of imaging depth adaptive resolution enhancement is achieved through an external attention mechanism, which takes advantage of morphological similarity between the arbitrary-depth and full-depth *en face* images. Firstly, the deep feature maps are extracted by a feature extraction network from the arbitrary-depth and full-depth *en face* images. Secondly, the morphological similarity between the deep feature maps is extracted and utilized to emphasize the features strongly correlated to the vessel structures by using the external attention network. Finally, the SR image is recovered from the enhanced feature map through an up-sampling network. The proposed network is tested on a clinical skin OCT data set and an open-access retinal OCT dataset. The results show that the proposed external attention mechanism can suppress invalid features and enhance significant features in our tasks. For all tests, the proposed SR network outperformed the traditional image interpolation method, e.g. bi-cubic method, and the state-of-the-art image super-resolution networks, e.g. enhanced deep super-resolution network, residual channel attention network, and second-order attention network. The proposed method may increase the quantitative clinical assessment of micro-vascular diseases which is limited by OCT imaging device resolution.

1. Introduction

As a non-invasive imaging technique, optical coherence tomography (OCT) has gained much attention in the last decade. OCT uses near-infrared light to detect the luminous reflection coefficient of bio-tissues, and

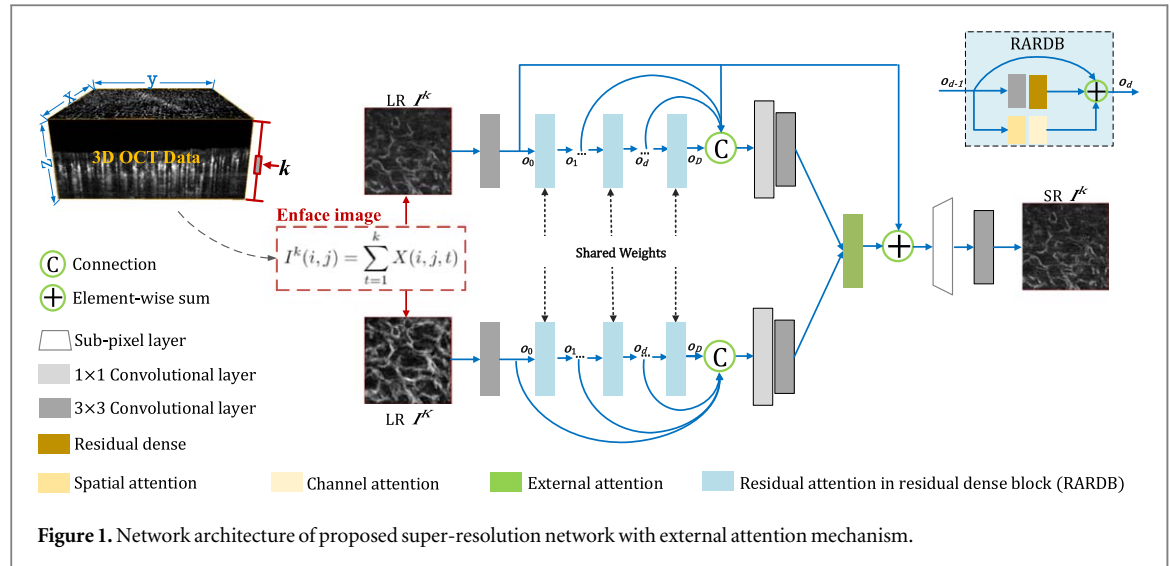
achieved 2–3 mm penetration depth and 10–40 μm spatial resolution, which is efficient for diagnosing (Tao *et al* 2017, Bekkers *et al* 2020), tumor angiogenesis (Kikuchi *et al* 2019, Liu *et al* 2020), and skin disease (Liu and Drexler 2019, Meiburger *et al* 2019), etc.

Resolution of OCT can be mainly divided into axial and lateral resolutions. Axial resolution is defined as the distinguish-ability of two nearest objects along the direction of the incident light, which mainly depends on the wavelength. Lateral resolution is defined as the distinguishability along the direction perpendicular to incident light, which primarily depends on the magnification of the imaging objective (Bizheva *et al* 2017). To improve the spatial resolution of OCT, many methods have been proposed. These methods can be mainly divided into hardware-based and digital-based methods. The hardware-based methods, e.g. adding an axon lens to the sample arm of the interferometer (Ding *et al* 2002) or using liquid-filled polymer lenses in the endoscopic system (Divetia *et al* 2005), improved spatial resolution with additional hardware, thus is cost-intensive and implemented complexly. The digital-based methods, e.g. using the modified Bayesian residual transform (Tan *et al* 2018) or the projection-resolved optical coherence tomography angiography (OCTA) (Wang *et al* 2017), utilized digital signal processing to enhance spatial resolution, thus is low cost and implemented easily.

Image super-resolution (SR) (Yang *et al* 2008) is another efficient method for resolution enhancement. SR method recovers high-resolution (HR) images from low-resolution (LR) ones by using their morphological characteristics, and is widely studied and well-established in the field of computer vision. The traditional SR methods enhance the image resolution by using the intrinsic morphological and structural information through a sparse-coding-based way. They are normally theoretically well-proved, but have limited speed and accuracy. In the last few years, convolutional neural networks (CNN) based super-resolution methods have gained much attention. The related reports showed that the CNN-based methods significantly outperform the traditional neural networks in simulation and natural environments. In 2016, Dong *et al* (2016) applied CNN to super-resolution reconstruction of images to establish an end-to-end mapping relationship between LR and HR images. Later, Kim *et al* (2016) utilized residual and recursive modules in their network to improve SR performance. In 2017, Lim *et al* (2017) proposed an enhanced deep super-resolution (EDSR) network, which enhanced feature flowing in local residual modules of deep neural network by removing batch normalization layers. Ledig *et al* (2017) up-scaled LR features with a sub-pixel layer which has been proved more effective than the traditional transposed convolution layer. In 2018, Zhang *et al* (2018) proposed a residual dense network with continuous and global residual structures to make full use of shallow feature information. Later, they further presented an enhanced deep super-resolution network, named RCAN, to incorporate channel attention in residual dense modules (Zhang *et al* 2018). In 2019, Dai *et al* (2019) proposed a normalization-based second-order attention network named as SAN for SR tasks and achieved state-of-the-art performance.

SR method has also attracted much attention in medical image analysis. In 2018, Wang *et al* (2018) exploited a generative adversarial network-based method to achieve SR across different fluorescence microscopy modalities. Lok *et al* (2021) used a fully convolutional neural network to achieve fast super-resolution ultrasound microvessel imaging. Liu *et al* (2019) proposed a deep learning-based SR pipeline for OCTA, which has a higher signal-to-noise ratio and shows potential in clinical use. Park *et al* (2018) present a semi-supervised deep learning approach to recover HR coherence tomography images from LR counterparts accurately. Shi *et al* (2018) proposed a novel residual learning-based SR algorithm for magnetic resonance imaging, which combines both multi-scale global residual learning and shallow network block-based local residual learning.

The SR methods for OCT image enhancement can be divided into the B-scan image-based and *en face* image-based methods. The B-scan image-based methods aim to enhance the resolution of the B-scan image of OCT. Fang *et al* (2013) proposed sparsity-based simultaneous denoising and interpolation to create high-quality B-scan using fewer A-scans. Das *et al* (2020) proposed a generative adversarial network-based framework to perform fast and reliable SR for OCT B-scan images. Qiu *et al* (2020) exploit U-net to obtain high signal-to-noise ratio and HR B-scan images within a short scanning time from noisy and low-resolution B-scan images. The *en face* image-based methods aim to enhance the resolution of *en face* image of OCT. As a commonly used visualization method in both research and clinical scenarios, *en face* image, which projects a 3D OCT image into 2D along its incident light direction, is helpful for qualitative or quantitative assessments of the retinal micro-vasculature (Jia *et al* 2015, Chu *et al* 2016, Fard *et al* 2018). As a result, the SR results of *en face* images are more intuitionist than these of B-scan images, by considering that the morphological characteristics are more significant in *en face* images than in B-scan images. Zhou *et al* (2020) used a recurrent generative adversarial network to enhance the resolution of images collected from an OCTA device with 8×8 mm receptive field. The results proved that the enhanced images are comparable to the images collected from a device with 3×3 mm receptive field. Gao *et al* (2020) used CNN to improve the image quality of OCTA *en face* image. Compared with raw images, the enhanced images have lower noise intensity, stronger contrast, and better vascular connectivity. Jalili *et al* (2020) exploit a curvelet-based method to combine detailed information of a set of *en face* images to construct high-quality *en face* image. Although there are already some reports on using SR method to enhance OCT *en face* images, they all focus on improving the spatial resolution of *en face* images with a specific imaging



depth. Resolution enchantment for *en face* image with arbitrary imaging depth is helpful for enriching information for both research and clinical studies (Nanji et al 2020).

In this paper, an imaging depth adaptive method is proposed for improving spatial resolution of OCT *en face* images. Different from the previous reports, our method can reconstruct HR *en face* images from LR *en face* images at arbitrary imaging depth by using the full-depth *en face* images (projection along the entire imaging depth) as a guidance. The proposed method is implemented with a deep learning framework with an external attention mechanism. Firstly, two-channel weights shared feature extraction networks are designed to extract the deep feature maps of both arbitrary-depth and full-depth LR *en face* images. Secondly, an external attention map is calculated from these two feature maps according to their morphological similarities. Thirdly, the attention map is used to enhance the feature map of the arbitrary-depth image through a multiplication operation. Finally, the SR image is recovered from the enhanced feature map through an up-sampling network. The proposed method is tested on a clinical skin OCTA data set and a public retinal OCT data set and compared with the traditional image interpolation method, e.g. bi-cubic interpolation, and the state-of-the-art SR networks, e.g. EDSR, RCAN, and SAN.

2. Methodology

2.1. Network overview

As shown in figure 1, to generate an *en face* image, one needs to accumulate a 3D OCT or OCTA image in z -direction. For an *en face* image with imaging depth k , one has

$$I^k(i, j) = \sum_{t=1}^k X(i, j, t), \quad \text{for } 1 \leq k \leq K, \quad (1)$$

where I_k is the *en face* image, X is a $n \times n \times K$ OCT or OCTA image, and K is the total number of pixels along z -direction. Our target is to estimated HR *en face* image I_H^k from LR *en face* image I_L^k at arbitrary imaging depth $1 \leq k \leq K$. During the estimation, the full-depth *en face* image I_L^K is used as a guidance. The proposed method mainly consists of two feature extraction modules with shared weights, one external attention module, and one up-sampling module. The feature extraction module contains two branches with shared weights for extracting deep features from arbitrary- and full-depth *en face* images, I_L^k and I_L^K . The external attention module fuses two deep channel features and emphasizes the vessel structure in the feature maps. The up-sampling module recovers the SR image from deep learned features.

2.2. Feature extraction block

The feature extraction module consists of a convolution layer followed by several sequential residual attention in residual dense blocks (RARDBs) and two additional convolution layers, as shown in figure 1. The entire feature extraction module could be formalized by

$$h^k = \mathcal{F} \circ \mathcal{F}_{1 \times 1} \circ C(o_0, o_1, \dots, o_D), \quad (2)$$

where \mathcal{F} and $\mathcal{F}_{1 \times 1}$ denote convolution operations of convolution layers with 3×3 and 1×1 kernels, respectively, C denotes concatenation operation, and o represents the input/output feature maps of RARDBs.

2.2.1. RARDB

As shown in figure 1, the RARDB block consists of spatial attention, channel attention, convolution layers, and residual dense modules. The spatial attention and channel attention modules aim to emphasize the features which are more helpful for our task. Several sequential residual dense units construct the residual dense modules. The residual dense unit was proposed by Zhang *et al* (2018). Being different from the residual block which has been used in high-performance SR (Kim *et al* 2016, Lim *et al* 2017, Tai *et al* 2017), the residual dense unit has the advantage of making full use of shallow and deep features. The local skip connection is used to reduce the information distillation during the feature refining. Following figure 1, for initial input $o_0 = \mathcal{F}(I_L^k)$, the feature flow in the RARDBs can be formulized as

$$o_{d+1} = o_d + \mathcal{F} \circ \mathcal{G}(o_d) + \mathcal{C} \circ \mathcal{S}(o_d), \quad (3)$$

where \mathcal{G} denotes residual dense module, \mathcal{S} and \mathcal{C} denote the spatial and channel attention units, respectively.

2.2.2. Spatial attention mechanisms

Spatial attention module produces spatial-wise attention by utilizing the inter-spatial relationship of features (Woo *et al* 2018) and is widely used in SR (Chen *et al* 2021), object classification (Zhu *et al* 2021). For a given feature map $x \in \mathbb{R}^{h \times w \times p}$ carrying $h \times w$ -dimensional spatial information and p -dimensional channel information, the spatial attention unit uses the first-order statistics for feature enhancement in spatial space. The spatial attention unit is formulated as

$$y = \mathcal{S}(x) = x \odot \mathcal{S} \circ \mathcal{F}_{1 \times 1} \circ \mathcal{C}(\alpha_s, \beta_s), \quad (4)$$

where \odot denotes element-wise product, \mathcal{S} denotes the sigmoid function, $\alpha_s \in \mathbb{R}^{h \times w}$ and $\beta_s \in \mathbb{R}^{h \times w}$ are averaged and maximized feature maps in spatial space, respectively. Here, α_s and β_s are implemented by maximum pooling and average pooling in spatial dimensions, respectively. They are formulated as

$$\alpha_s(i, j) = \frac{1}{p} \sum_{k=1}^p (x(i, j, k)), \quad (5)$$

$$\beta_s(i, j) = \max (x(i, j, :)). \quad (6)$$

The sigmoid function \mathcal{S} in (4) is mainly used as a gating operator for incorporating nonlinearity and mutual exclusivity into our feature extraction network. It can efficiently emphasize useful features and suppress invalid features.

2.2.3. Channel attention mechanisms

Channel attention module generates channel-wise attention maps, which have been proved efficient on image captioning (Chen *et al* 2017), fusion (Li *et al* 2020), and segmentation (Lee *et al* 2020). The channel attention unit uses the first-order statistics for feature enhancement in channel space. Similar to spatial attention, the channel attention unit is formulated as

$$y = x \odot \mathcal{S}(\mathcal{F}_{1 \times 1}^\dagger \circ \mathcal{F}_{1 \times 1}(\alpha_c) + \mathcal{F}_{1 \times 1}^\dagger \circ \mathcal{F}_{1 \times 1}(\beta_c)), \quad (7)$$

where $\mathcal{F}_{1 \times 1}^\dagger$ is a convolution operation re-scale the feature map from $0.25 \times p$ to p , $\alpha_c \in \mathbb{R}^p$ and $\beta_c \in \mathbb{R}^p$ are averaged and maximized feature maps in channel space, respectively. The α_c and β_c are implemented by maximum pooling and average pooling in channel dimensions, respectively. They are formulated as

$$\alpha_c(k) = \frac{1}{hw} \sum_{i=1}^h \sum_{j=1}^w (x(i, j, k)), \quad (8)$$

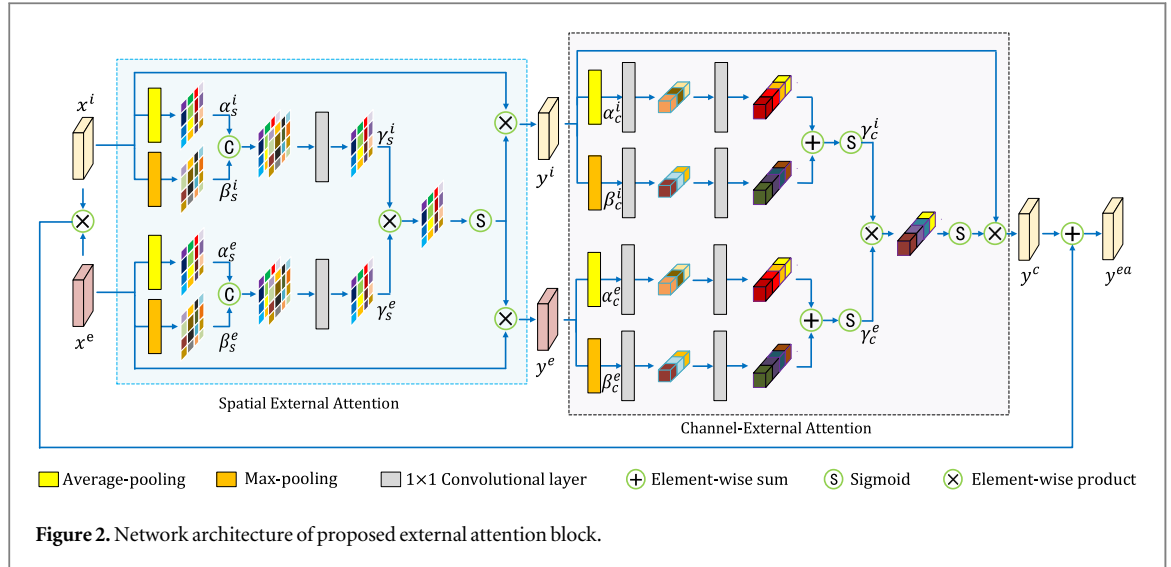
$$\beta_c(k) = \max (\max (x(:, :, k))). \quad (9)$$

2.3. External attention block

The external attention block's primary purpose is to utilize deep feature maps extracted from the full-depth *en face* image to enhance vessel structure information in the deep feature maps extracted from the arbitrary-depth *en face* image. As shown in figure 2, the feature fusion block consists of a spatial-external attention block followed by a channel-external attention block. The entire feature fusion block can be formulated by

$$y^{ea} = x^i \odot x^e + \mathcal{E}_c \circ \mathcal{E}_s(x^i, x^e), \quad (10)$$

where $x^i \in \mathbb{R}^{h \times w \times p}$ and $x^e \in \mathbb{R}^{h \times w \times p}$ are the internal and external feature maps, respectively, $\mathcal{E}_s: \mathbb{R}^{2 \times h \times w \times p} \mapsto \mathbb{R}^{2 \times h \times w \times p}$ and $\mathcal{E}_c: \mathbb{R}^{2 \times h \times w \times p} \mapsto \mathbb{R}^{h \times w \times p}$ denote the spatial-external attention and channel-external attention blocks, respectively. Inspired by spatial attention and channel attention, spatial-external attention and channel-external attention also use the maximum pooling and average pooling operators to enhance the spatial and channel features, which are more efficient on our SR task. Unlike spatial attention and



channel attention, the feature enhancement in spatial-external attention and channel-external attention focused on the structural similarities between the superficial image I_L^K and the full-depth image I_L^K , resulting in an external supervision mechanism for the superficial image SR task.

2.3.1. Spatial-external attention

Figure 2 shows the construction of the spatial-external attention block. The spatial-external attention block mainly consists of two-channel attention units with shared wights, followed by a point element-wise product and sigmoid activation. Firstly, for a given internal deep feature map x^i , two channel-independent feature maps α_s^i and β_s^i are extracted by using average- and max-pooling, respectively. Secondly, these two feature maps are concatenated together and fused into a $h \times w$ feature map γ_s^i through a convolution layer. Thirdly, following the same procedure, the external feature map γ_s^e is generated. Fourthly, γ_s^i and γ_s^e are element-wise produced to extract vessel-dependent attention maps. Finally, the vessel enhanced feature maps y^i and y^e are calculated by multiplying the spatial-wise attention map with x^i and x^e , respectively. The entire spatial-external attention block can be formulated as

$$\begin{bmatrix} y^i \\ y^e \end{bmatrix} = \mathcal{E}_s(x^i, x^e) = \begin{bmatrix} x^i \\ x^e \end{bmatrix} \odot S \circ (\gamma_s^i \odot \gamma_s^e) \quad (11)$$

where sigmoid function S is used to incorporate nonlinearity and mutual exclusivity into our external attention module, and the spatial-wise enhanced deep feature maps $\gamma_s^i \in \mathbb{R}^{h \times w}$ and $\gamma_s^e \in \mathbb{R}^{h \times w}$ are defined as

$$\gamma_s^i = \mathcal{F}_{1 \times 1} \circ C(\alpha_s^i, \beta_s^i), \quad (12)$$

$$\gamma_s^e = \mathcal{F}_{1 \times 1} \circ C(\alpha_s^e, \beta_s^e). \quad (13)$$

with $\{\alpha_s^i, \beta_s^i\}$ and $\{\alpha_s^e, \beta_s^e\}$ calculated by (5) and (6).

2.3.2. Channel-external attention

Figure 2 shows the construction of the channel-external attention block. The channel-external attention block mainly consists of two-channel attention units with shared wights, followed by a point element-wise product and sigmoid activation. Firstly, we use average- and max-pooling to generate channel descriptors α_c and β_c of length p , which contain global spatial information from spatial-external attention output y^i . Secondly, we also introduce a gating mechanism to squeeze and excite α_c and β_c through serially connected convolution layers, merging, and sigmoid operators. Thirdly, the channel-wise attention vector is calculated using element-wise production and sigmoid operation to emphasize correlations between squeezed internal and external feature maps γ_c^i and γ_c^e . Finally, the attention vector multiplies to the original internal feature map y^i to enhance channel-wise vessel information under the information from the full-depth *en face* image. The entire channel-external attention block is defined as

$$y^c = x^i \odot S \circ (\gamma_c^i \odot \gamma_c^e), \quad (14)$$

where channel-wise enhanced deep feature maps $\gamma_c^i \in \mathbb{R}^{p \times 1}$ and $\gamma_c^e \in \mathbb{R}^{p \times 1}$ are defined as

$$\gamma_c^i = \mathcal{F}_{1 \times 1}^\dagger \circ \mathcal{F}_{1 \times 1}(\alpha_c^i) + \mathcal{F}_{1 \times 1}^\dagger \circ \mathcal{F}_{1 \times 1}(\beta_c^i), \quad (15)$$

$$\gamma_c^e = \mathcal{F}_{1 \times 1}^\dagger \circ \mathcal{F}_{1 \times 1}(\alpha_c^e) + \mathcal{F}_{1 \times 1}^\dagger \circ \mathcal{F}_{1 \times 1}(\beta_c^e), \quad (16)$$

with $\{\alpha_c^i, \beta_c^i\}$ and $\{\alpha_c^e, \beta_c^e\}$ calculated with (8) and (9) by replacing x with y^i and y^e , respectively.

2.4. Up-sampling block

In up-sampling module, the shallow feature o_0 and high-level feature y^{ea} are merged by using global skip structure and addition operation. The combination is helpful for enriching both high-frequency and low-frequency features in LR space. Then, the LR feature maps is upsampled to HR output through a sub-pixel convolution layer. Finally, the feature map's channel dimension is reduced to 3 through a 3×3 convolution layer and treated as the output SR image. Two interpolation strategies including sub-pixel layer and deconvolution layer are taken into consideration when we design up-sampling block. Compared with the deconvolution layer, the sub-pixel convolution layer can automatically learn an array of upscaling filters, and was proved more efficient insolving image SR problems (Ledig et al 2017).

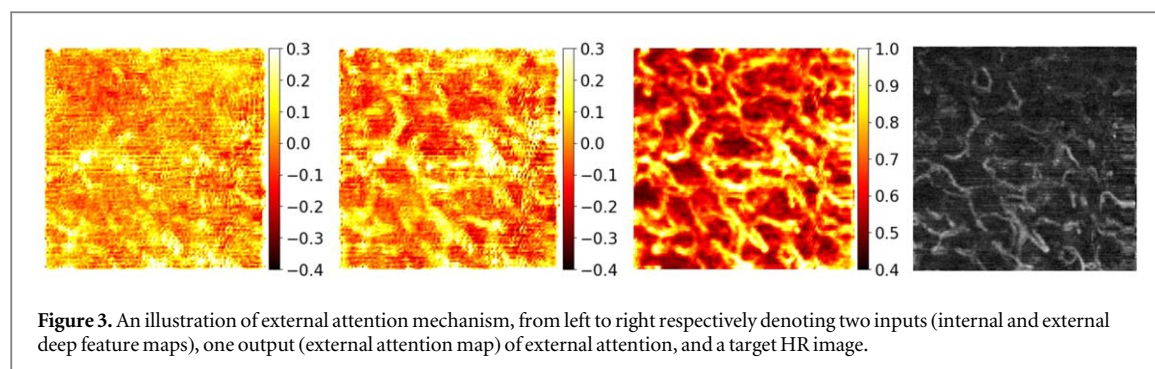
3. Experiments

3.1. Experimental setups

With institutional review board approval and waived patient consent, we retrospectively collected 2400 3D OCTA data from 100 previously laser-treated patients with nevus flammeus. These data are collected at the First Medical Center of PLA General Hospital, Beijing, China. The data acquisition system is developed following the ultrahigh sensitive optical microangiography scanning protocol (An et al 2010). For each B frame, 500 A-lines are collected by driving the X galvo mirror forward and backward using a sawtooth waveform. The B frame data collection frequency is 200 Hz. To observe dynamic flow information, B-scan location is repeated 3 times. The receptive field of the OCTA scanning is 4×4 mm, which took 7.5 s for data acquisition. The eigen-decomposition-based angiography algorithm was adopted to separate the dynamic scattering signals (e.g. flowing components in vascular networks) from the static tissue (e.g. non-moving tissue structural components) by utilizing the statistical properties of time-varying complex OCT signals (Yousefi et al 2011). This algorithm is a robust OCTA algorithm, which can adaptively filter out the bulk tissue motion (e.g. heartbeat and breathing) to provide superior blood flow images for *in vivo* OCTA imaging (Zhang et al 2016). For each 3D OCT data, six imaging depths ($k = 120, 150, 180, 210, 240, 270$) are selected to produce HR *en face* images, leading to 14 400 *en face* images in total. The LR images are generated by two times downsampling the HR images. The entire data set is divided into the training data set and testing data set. The training data set contains 11 520 samples (80% of the entire data set). The testing data set contains 2880 samples (20% of the entire data set).

The proposed method is also evaluated on a public OCT dataset (Li et al 2020), which contains 500 3D OCT of retinal vascular. Each OCT data is further separated into the internal limiting membrane (ILM) layer, the outer plexiform layer (OPL), and the Bruch's Membrane (BM) according to their imaging depth along the axial direction. The full-projection, ILM-OPL, and OPL-BM *en face* images are generated from each 3D OCT data. The full-projection *en face* image is the average projection of the entire 3D OCT volume. The ILM-OPL *en face* image is the average projection of 3D OCT data between the ILM and OPL layers. The OPL-BM *en face* image is the average projection of 3D OCT data between the OPL and BM layers. Different from the skin OCT cases, in the retinal OCT cases, the full-projection *en face* image is used for guiding the SR image reconstruction of LR ILM-OPL and OPL-BM *en face* images. The entire data set is also divided into the training data set (80% of the entire data set) and the testing data set (20% of the entire data set).

The proposed network is implemented on a workstation with Intel Xeon E5-2630 CPU and 64 GB RAM with Tensorflow and computes a unified device architecture environment. The Nvidia GTX 2080Ti GPU is used to accelerate the training process to nearly 12 h each training. The initial learning rate is 10^{-4} and gradually decreases when training the network. The mean square error loss is employed as the loss function of the proposed network. The data set is only divided into a training set and testing set, and the validation set is ignored. The validation set is usually used to select the super-parameters, e.g. number of epochs, and avoid overfitting. The number of epochs used for training our network is 500, which is large enough to guarantee the stability and convergence of the training process. Our training data set contains 11 520 *en face* images. According to the loss curves, overfitting was avoided. The proposed network is compared with the image interpolation method (bi-cubic interpolation), and the state-of-the-art SR networks (EDSR (Lim et al 2017), RCAN (Zhang et al 2018) and SAN (Dai et al 2019)). The comparisons are conducted qualitatively, by visually inspecting the image qualities, and quantitatively, by calculating the peak signal-to-noise ratio (PSNR) and structural similarity (SSIM) between the recovered and target HR images.



3.2. Visualization of external attention mechanism

According to section 2.3, the inputs of the external attention block are the internal and external feature maps output from the feature extraction blocks. The internal feature map y^i is extracted from arbitrary-depth *en face* image I_L^k . The external feature map y^e is extracted from full-depth *en face* image I_L^K . The output of the external attention block is the enhanced feature map y^c . A 2D visualization of the different feature maps is shown in figure 3. The columns, from left to right, respectively show the internal, external, enhanced feature maps and the target HR *en face* image. On the one hand, the blood flow signal is barely observed in the internal feature map, but is relatively obvious in the external feature map. On the other hand, the blood flow signal in the enhanced feature map is significant and strongly correlated with the target HR image, which proved the efficiency of the proposed external attention mechanism.

The feature enhancement results from the proposed external attention mechanism can be explained by the following two aspects. Firstly, the full-depth *en face* image is generated by accumulating all gray values of the 3D OCT image along the depth direction. This kind of accumulation can be considered as a mean value filter, which is helpful for noise suppression. Secondly, the full-depth *en face* image contains all the information of the arbitrary-depth *en face* images. As a result, it will be helpful for guiding the resolution enhancement of the arbitrary-depth *en face* images.

3.3. Results on skin OCTA data set

The EDSR, RCAN, proposed network and its ablation network are trained and tested on the same skin OCTA data set, as it is described in section 3.1. An illustration of the SR image reconstruction results is shown in figure 4, tables 1 and 2. SR results with different imaging depths are evaluated. PSNR and SSIM, displayed with a white number in figure 4, are used for quantitatively evaluating these results. For all tested imaging depth, the proposed network leads to the best results, which owns the highest PSNR and SSIM.

Evolution of quantitative metrics with respect to the imaging depth is shown in figure 5. The quantitative comparison among different methods with different imaging depths is tabulated in tables 1 and 2. As can be seen, the deep learning-based resolution enhancement methods are significantly better than the bi-cubic method. The proposed network significantly outperformed the EDSR, RCAN, and SAN methods, especially in the large imaging depth tests. For half depth SR *en face* image estimation, the proposed method achieved a 40.213 dB PSNR and 0.930 SSIM, which improved upon bi-cubic, EDSR, and RCAN methods (PSNR 1.160, 0.342, 0.162, and 0.165 dB, respectively) (SSIM 0.190, 0.04, 0.02 and 0.03, respectively). For full-depth SR *en face* image estimation, the proposed method achieved a 40.543 dB PSNR and 0.940 SSIM, which improved upon bi-cubic, EDSR, RCAN, and SAN methods (PSNR 1.308, 0.455, 0.218, and 0.196 dB, respectively) (SSIM 0.190, 0.04, 0.02, and 0.02, respectively).

3.4. Quantitative micro-vascular analysis

Firstly, the vessel area map $A[i, j]$, perimeter map $P[i, j]$, and skeleton map $S[i, j]$ are calculated for qualitatively analyzing the skin micro-vascular images. The perimeter map is generated by successively applying threshold segmentation, skeletonization, and Canny edge detection on the original vessel image $I[i, j]$. The columns in figure 6, from left to right, respectively show the original vessel map $I[i, j]$, the vessel area map $A[i, j]$, zoomed vessel area map $A[i, j]$, the perimeter map $P[i, j]$, zoomed perimeter map $P[i, j]$, the skeleton map $S[i, j]$, zoomed skeleton map $S[i, j]$. The rows in figure 6, from top to bottom, respectively show the manually designed feature maps calculated from the original HR, recovered SR images, and LR images. By visually checking these images, the vessel maps from the HR and SR images are quite similar to each other, but significantly different from these from the LR images. The vessel areas are significantly over-estimated from the LR images, but accurately calculated from the estimated SR images. Detailed information of the skeleton feature maps are miss-detected

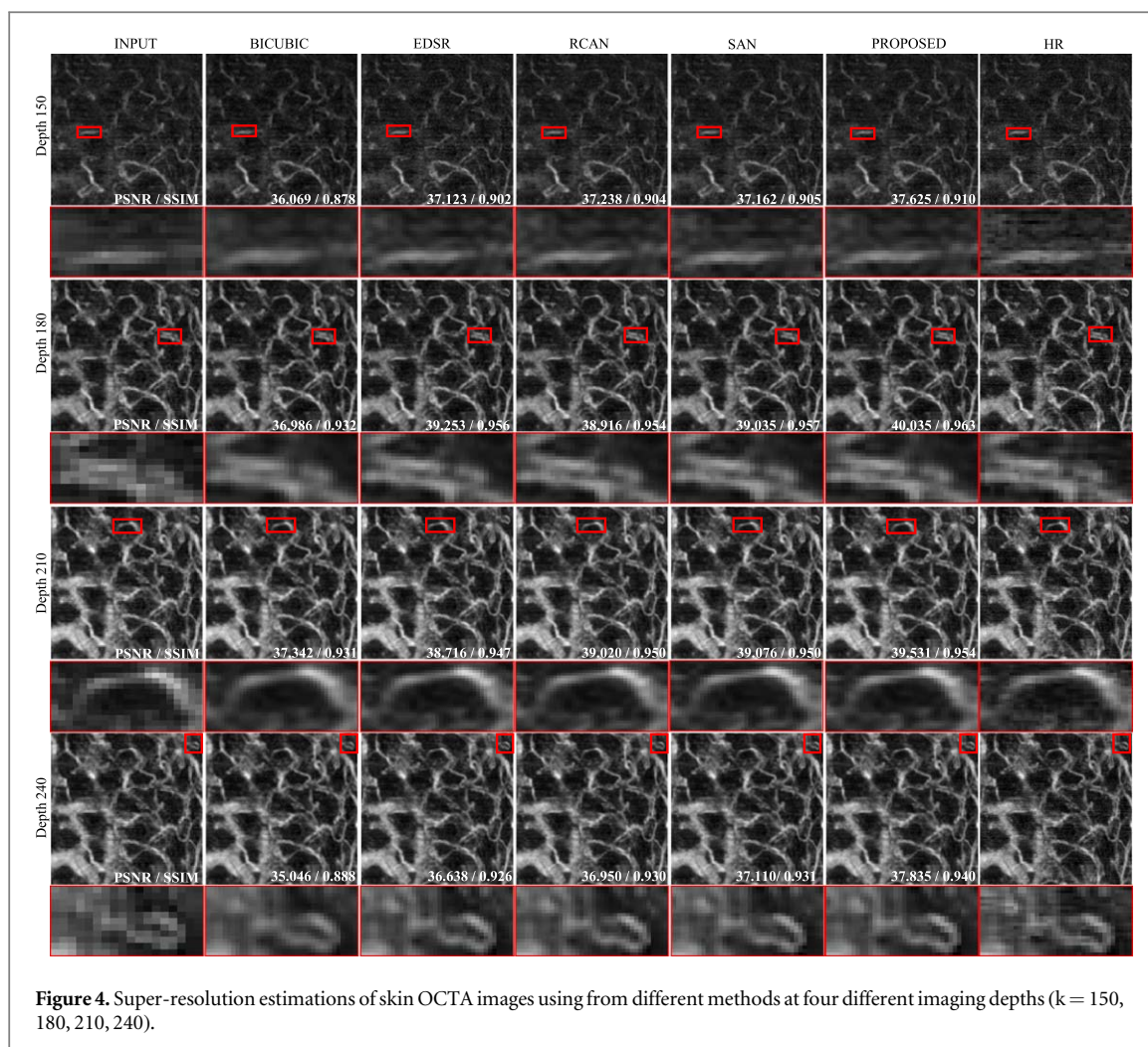
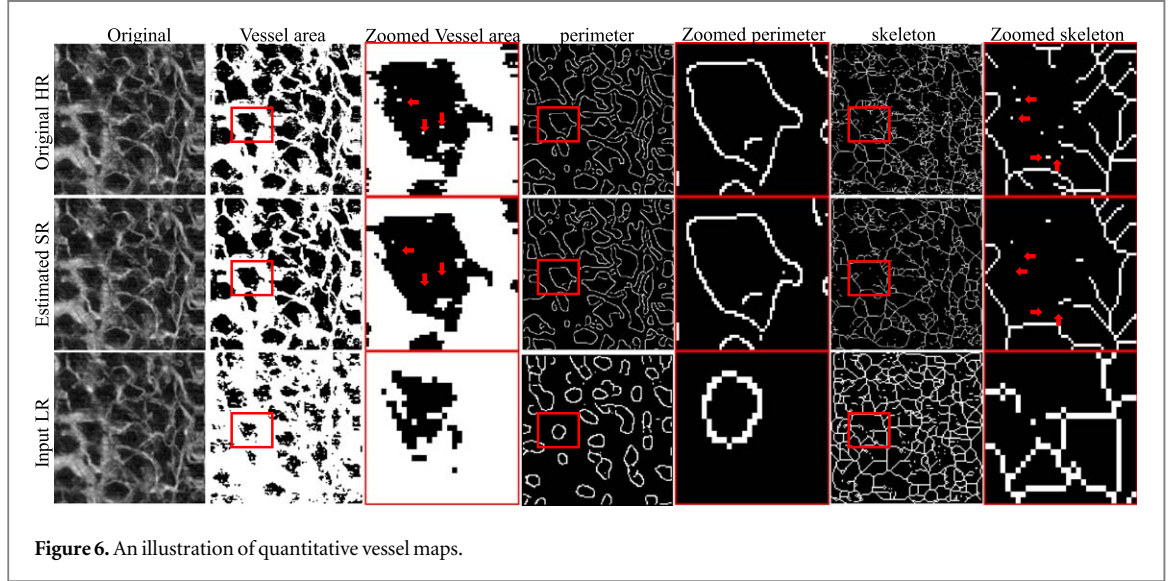
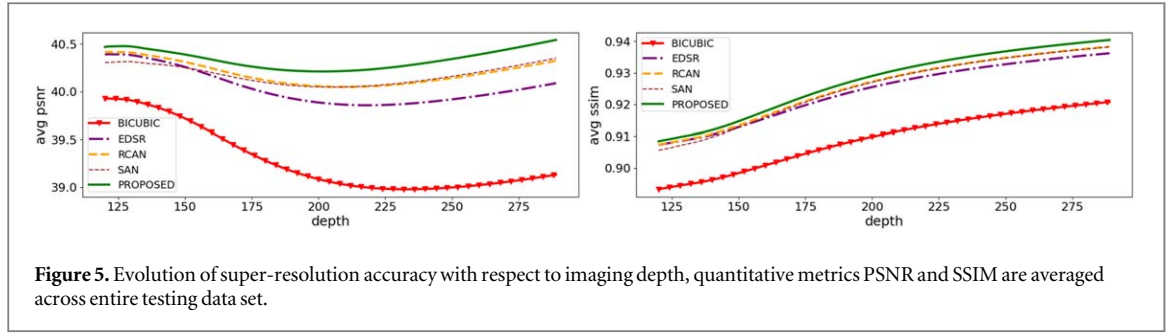


Table 1. PSNR of SR images recovered from different methods at different imaging depths.

Imaging depth	bi-cubic	EDSR	RCAN	SAN	Proposed network
120	39.923	40.389	40.411	40.304	40.469
150	39.717	40.258	40.310	40.252	40.389
180	39.280	39.990	40.118	40.095	40.244
210	39.026	39.862	40.048	40.048	40.216
240	38.981	39.887	40.105	40.120	40.300
270	39.053	39.996	40.229	40.252	40.439

Table 2. SSIM between true HR image and recovered SR image at different imaging depth.

Imaging depth	bi-cubic	EDSR	RCAN	SAN	Proposed network
120	0.893	0.907	0.907	0.905	0.908
150	0.898	0.913	0.913	0.913	0.915
180	0.906	0.921	0.922	0.922	0.926
210	0.911	0.927	0.929	0.929	0.933
240	0.916	0.931	0.933	0.933	0.937
270	0.919	0.935	0.937	0.937	0.942



from the LR image, but successfully recovered from the estimated SR images. Furthermore, compared with the results from original HR images, there are fewer isolated points in the vessel area and skeleton maps from the estimated SR images, which implies that the proposed method can restrain the speckle noise in *en face* images and further improve the accuracy of the following quantitative analysis.

Secondly, three quantitative metrics are developed according to the manually designed feature maps. They are vessel area density, vessel complexity, and vessel perimeter index, which are defined as

$$VAD = \frac{\sum_{i=1}^m \sum_{j=1}^m A[i, j]}{\sum_{i=1}^m \sum_{j=1}^m X[i, j]}, \quad (17)$$

$$VCI = \frac{\left(\sum_{i=1}^m \sum_{j=1}^m P[i, j] \right)^2}{\sum_{i=1}^m \sum_{j=1}^m A[i, j]}, \quad (18)$$

$$VPI = \frac{\sum_{i=1}^m \sum_{j=1}^m P[i, j]}{\sum_{i=1}^m \sum_{j=1}^m X[i, j]}, \quad (19)$$

where pixel value accumulation is conducted at a 25×25 neighbourhood of the testing point $[i, j]$. $A[i, j]$ denotes the non-zero pixels of the vessel area image. $P[i, j]$ denotes the non-zero pixels enclosed by vessel perimeters image. $X[i, j]$ denotes all pixels on the neighbourhood. The box-plots of the quantitative metrics in the entire testing data set are shown in figure 7. The box-plots from SR and HR images are quite similar, but significantly different from those from the LR images, implying that more accurate micro-vascular analysis can be conducted by using the LR OCTA images and the proposed SR network. In other words, the proposed method may reduce the dependence of the clinical micro-vascular analysis accuracy on the resolution of OCT imaging devices.

3.5. SR performance on OCT-500

The performance of the proposed SR network is also evaluated on the published retinal OCT data set. The full-projection *en face* images are used for improving the performance of the SR of the ILM-OPL and OPL-BM *en face*

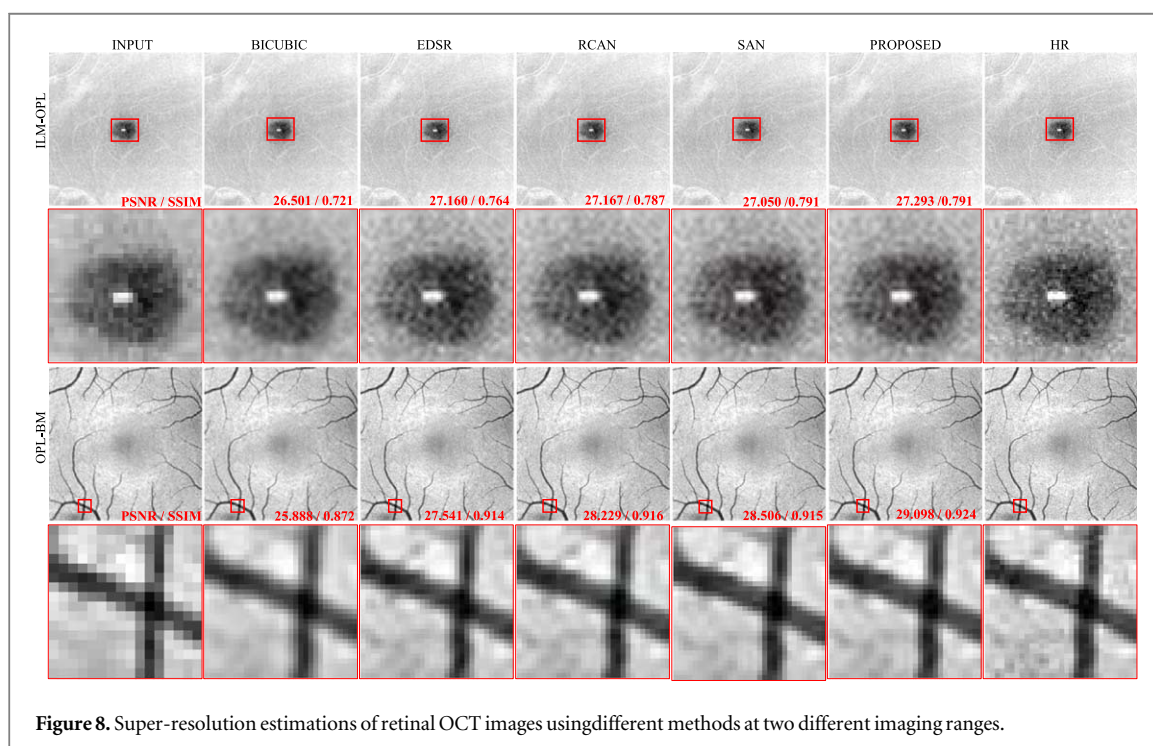
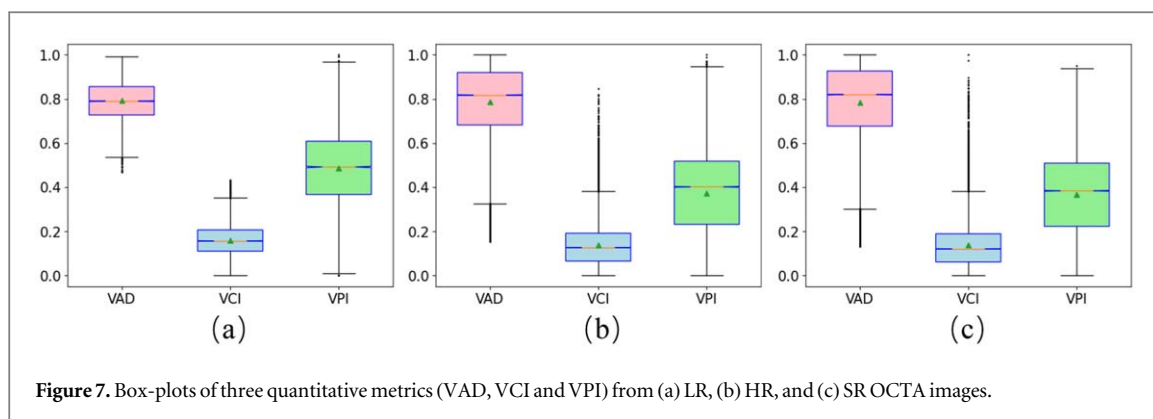


Table 3. PSNR of SR images recovered from different methods at ILM-OPL, OPL-BM projection maps.

Imaging depth	bi-cubic	EDSR	RCAN	SAN	Proposed network
ILM-OPL	32.707	33.494	33.488	33.482	33.568
OPL-BM	29.621	31.056	31.138	31.133	31.224

Table 4. SSIM between true HR image and recovered SR image at ILM-OPL, OPL-BM projection maps.

Imaging depth	bi-cubic	EDSR	RCAN	SAN	Proposed network
ILM-OPL	0.765	0.821	0.821	0.821	0.824
OPL-BM	0.748	0.823	0.822	0.822	0.825

images. In other words, the ILM-OPL and OPL-BM *en face* images are used as I^k , while the full-projection *en face* image is used as I^K in figure 1.

An illustration of the reconstruction results is shown in figure 8. Quantitative analysis of the results on the entire testing set is tabulated in tables 3 and 4. The proposed network significantly outperformed the EDSR, RCAN, and SAN methods, especially in the OPL-BM cases. For ILM-OPL SR *en face* image estimation, the

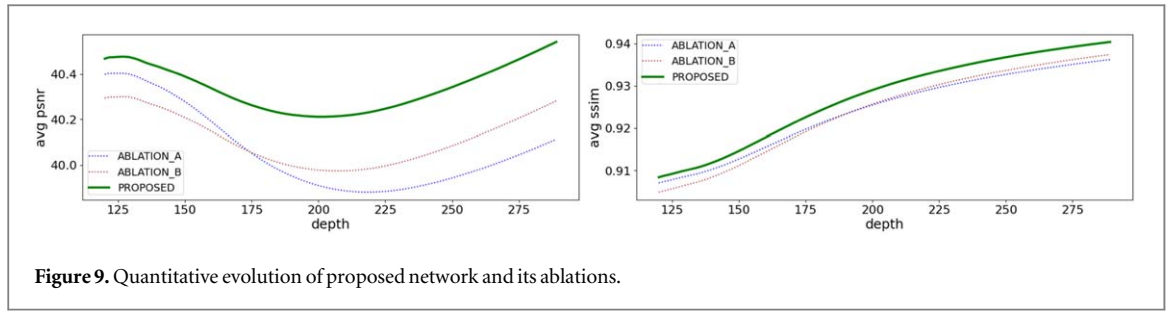


Figure 9. Quantitative evolution of proposed network and its ablations.

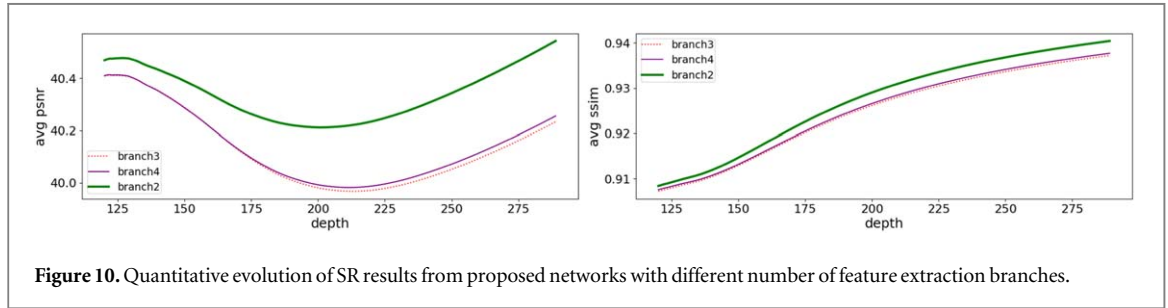


Figure 10. Quantitative evolution of SR results from proposed networks with different number of feature extraction branches.

proposed method achieved a 33.568 dB PSNR and 0.824 SSIM, which improved upon bi-cubic, EDSR, RCAN and SAN methods (PSNR 0.861, 0.074, 0.081, and 0.086 dB, respectively) (SSIM 0.059, 0.003, 0.003 and 0.003, respectively). For OPL-BM SR *en face* image estimation, the proposed method achieved a 31.224 dB PSNR and 0.825 SSIM, which improved upon bi-cubic, EDSR, RCAN and SAN methods (PSNR 1.603, 0.168, 0.086, and 0.089 dB, respectively) (SSIM 0.077, 0.002, 0.003 and 0.003, respectively). Similar to the results from skin OCT tests, the proposed network is also better than its ablation network in the retinal OCT tests, which further proved the effectiveness of the proposed external attention mechanism.

3.6. Ablation study and network structure discussion

A series of ablation studies were conducted for discussing the construction of our network. Firstly, the external attention block, as described in section 2.3, is removed from our network. This network without external attention block is denoted with 'ABLATION_A', and is used to discuss the influence of external attention block on the SR performance. Secondly, the spatial and channel attention units in RARDBs are removed from our network. This network with simplified RARDBs is denoted with 'ABLATION_B', and is used to discuss the influence of spatial and channel attention units on the SR performance. The quantitative evaluation of the performance of the proposed network and its ablations, 'ABLATION_A' and 'ABLATION_B', are shown in figure 9. As can be seen, the PSNR and SSIM from the proposed network are significantly higher than those from its ablations, proved that the external attention block and the spatial and channel attention units are efficient in improving the reconstruction accuracy of our image SR task.

In previous reports, researchers use three or more feature extraction branches to estimate high-quality OCT images (Liu et al 2019, Jalili et al 2020). In our method, only two branches, one for full-depth image I_K and the other for arbitrary-depth image I_k , are employed. To demonstrate the proposed structure is the most effective structure, the proposed two branches network is compared with the three and four branches networks. For the three branches network, two *en face* images (I_K and $I_{(K-10)}$) are used to guide the SR estimation of arbitrary-depth image I_k . For the four branches network, three *en face* images (I_K , $I_{(K-10)}$ and $I_{(K-20)}$) are used to guide the SR estimation of arbitrary-depth image I_k . The quantitative evaluation of the performance of these three networks is shown in figure 10. As can be seen, the PSNR and SSIM from three and four branches strategies are a little lower than those from the two branch strategy, which proved that more feature extraction branches can not improve the SR reconstruction accuracy. Actually, the external attention mechanism is mainly used to emphasize the blood signal contained in both shallow and full-depth *en face* images. During this process, the background signal, including noise, is also amplified. This amplification will be enhanced with the increase of the number of branches, which would reduce the SR reconstruction accuracy.

4. Conclusion

In this paper, a deep neural network is proposed to calculate SR OCTA *en face* images with arbitrary imaging depth. The morphological similarity between the arbitrary-depth and full-depth *en face* images is extracted by deep feature representation and incorporated into the SR estimations using an external attention mechanism. The proposed network is tested on a clinical skin OCTA data set and a public retinal OCT data set. The results show that the proposed external attention mechanism can suppress invalid features and enhances significant features in SR tasks. To further test the quality of the estimated SR images, the estimated SR images are used for quantitative measurements of cutaneous microvessels. The estimated SR images lead to the results basically the same as those from the truth SR images, which implies that the proposed method may improve the clinical quantitative assessment of micro-vascular diseases.

Acknowledgments

This work was supported by the National Natural Science Foundation of China (grant numbers: 81 827 806, 61 835 015, and 62 005 045), CAMS Innovation Fund for Medical Sciences (grant number: 2019-I2M-5-061), and Guangdong Basic and Applied Basic Research Foundation (grant number: 2019A151010805). The authors would also thank Tianjin Horimed Technology Co., Ltd. for their technical support of OCTA.

ORCID iDs

Shangjie Ren  <https://orcid.org/0000-0003-2220-3856>

Defu Chen  <https://orcid.org/0000-0002-7198-1104>

Feng Dong  <https://orcid.org/0000-0002-8478-8928>

References

- An L, Qin J and Wang R K 2010 Ultrahigh sensitive optical microangiography for *in vivo* imaging of microcirculations within human skin tissue beds *Opt. Express* **18** 8220–8
- Bekkers A, Borren N, Ederveen V, Fokkinga E, De Jesus D A, Brea L S, Klein S, van Walsum T, Barbosa-Breda J and Stalmans I 2020 Microvascular damage assessed by optical coherence tomography angiography for glaucoma diagnosis: a systematic review of the most discriminative regions *Acta Ophthalmol.* **97** 537–58
- Bizheva K, Tan B Y, MacLellan B, Hosseinaee Z, Mason E, Hileeto D and Sorbara L 2017 *In-vivo* imaging of the palisades of Vogt and the limbal crypts with sub-micrometer axial resolution optical coherence tomography *Biomed. Opt. Express* **8** 4141–51
- Chen C F, Gong D H, Wang H, Li Z F and Wong K Y K 2021 Learning spatial attention for face super-resolution *IEEE Trans. Image Process.* **30** 1219–31
- Chen L, Zhang H W, Xiao J, Nie L Q, Shao J, Liu W and Chua T S 2017 SCA-CNN: spatial and channel-wise attention in convolutional networks for image captioning *Conference on Computer Vision and Pattern Recognition (CVPR) (Honolulu, HI, JUL 21–26)* 6298–306
- Chu Z D, Lin J, Gao C, Xin C, Zhang Q Q, Chen C L, Roisman L, Gregori G, Rosenfeld P J and Wang R K 2016 Quantitative assessment of the retinal microvasculature using optical coherence tomography angiography *J. Biomed. Opt.* **21** 066008
- Dai T, Cai J R, Zhang Y B, Xia S T and Zhang L 2019 Second-order attention network for single image super-resolution *Conference on Computer Vision and Pattern Recognition (CVPR) (Long Beach, CA, JUN 16–20)* 11057–66
- Das V, Dandapat S and Bora P K 2020 Unsupervised super-resolution of oct images using generative adversarial network for improved age-related macular degeneration diagnosis *IEEE Sens. J.* **20** 8746–56
- Ding Z H, Ren H W, Zhao Y H, Nelson J S and Chen Z P 2002 High-resolution optical coherence tomography over a large depth range with an axicon lens *Opt. Lett.* **27** 243–5
- Divetia A, Hsieh T H, Zhang J, Chen Z P, Bachman M and Li G P 2005 Dynamically focused optical coherence tomography for endoscopic applications *Appl. Phys. Lett.* **86** 103902
- Dong C, Loy C C, He K M and Tang X 2016 Image super-resolution using deep convolutional networks *IEEE Trans. Pattern Anal. Mach. Intell.* **38** 295–307
- Fang L Y, Li S T, McNabb R P, Nie Q, Kuo N, Toth C A, Izatt J A and Farsiu S 2013 Fast acquisition and reconstruction of optical coherence tomography images via sparse representation *IEEE Trans. Med. Imaging* **32** 2034–49
- Fard M, Jalili J, Sahraian A, Khojasteh H, Hejazi M, Ritch R and Subramanian P 2018 Optical coherence tomography angiography in optic disc swelling *Am. J. Ophthalmol.* **191** 116–23
- Gao M, Guo Y K, Hormel T T, Sun J D, Hwang T S and Jia Y L 2020 Reconstruction of high-resolution 6 x 6 mm OCT angiograms using deep learning *Biomed. Opt. Express* **11** 3585–600
- Jalili J, Rabbani H, Dehnavi A M, Kafieh R and Akhlaghi M 2020 Forming optimal projection images from intra-retinal layers using curvelet-based image fusion method *J Medical Signals Sens.* **10** 76–85
- Jia Y L et al 2015 Quantitative optical coherence tomography angiography of vascular abnormalities in the living human eye *Proc. Natl. Acad. Sci. USA* **112** E2395–402
- Kikuchi I, Kase S, Hashimoto Y, Hirooka K and Ishida S 2019 Involvement of circulatory disturbance in optic disk melanocytoma with visual dysfunction *Graefes Arch. Clin. Exp. Ophthalmol.* **257** 835–41
- Kim J, Lee J K and Lee K M 2016 Accurate image super-resolution using very deep convolutional networks *Conference on Computer Vision and Pattern Recognition (CVPR) (Seattle, WA, JUN 27–30)* 1646–54

- Ledig C *et al* 2017 Photo-realistic single image super-resolution using a generative adversarial network *Conference on Computer Vision and Pattern Recognition (CVPR) (Honolulu, HI, JUL 21-26)* [105–14](#)
- Lee H, Park J and Hwang J Y 2020 Channel attention module with multiscale grid average pooling for breast cancer segmentation in an ultrasound image *IEEE T Ultrason. Ferr.* **67** [1344–53](#)
- Li H, Wu X J and Durrani T 2020 NestFuse: an infrared and visible image fusion architecture based on nest connection and spatial/channel attention models *IEEE Trans. Instrum. Meas.* **69** [9645–56](#)
- Li M C, Chen Y, Ji Z X, Xie K R, Yuan S T, Chen Q and Li S 2020 Image projection network: 3D to 2D image segmentation in OCTA images *IEEE Trans. Med. Imaging* **39** [3343–54](#)
- Lim B, Son S, Kim H, Nah S and Lee K M 2017 Enhanced deep residual networks for single image super-resolution *Conference on Computer Vision and Pattern Recognition Workshops (CVPRW) (Honolulu, HI, JUL 21-26)* [1132–40](#)
- Liu M Y and Drexler W 2019 Optical coherence tomography angiography and photoacoustic imaging in dermatology *Photochem. Photobiol. Sci.* **18** [945–62](#)
- Liu X *et al* 2019 A deep learning based pipeline for optical coherence tomography angiography *J. Biophoton.* **12** [e201900008](#)
- Liu Z P, Karp C L, Galor A, AlBayyat G J, Jiang H and Wang J H 2020 Role of optical coherence tomography angiography in the characterization of vascular network patterns of ocular surface squamous neoplasia *Ocul. Surf.* **18** [926–35](#)
- Lok U W *et al* 2021 Fast super-resolution ultrasound microvessel imaging using spatiotemporal data with deep fully convolutional neural network *Phys. Med. Biol.* **66** [075005](#)
- Meiburger K M, Chen Z, Sinz C, Hoover E, Minneman M, Ensher J, Kittler H, Leitgeb R A, Drexler W and Liu M Y 2019 Automatic skin lesion area determination of basal cell carcinoma using optical coherence tomography angiography and a skeletonization approach: Preliminary results *J. Biophoton.* **12** [e201900131](#)
- Nanji A *et al* 2020 Application of corneal optical coherence tomography angiography for assessment of vessel depth in corneal neovascularization *Cornea* **39** [598–604](#)
- Park J, Hwang D, Kim K Y, Kang S K, Kim Y K and Lee J S 2018 Computed tomography super-resolution using deep convolutional neural network *Phys. Med. Biol.* **63** [145011](#)
- Qiu B, You Y, Huang Z, Meng X, Jiang Z, Zhou C, Liu G, Yang K, Ren Q and Lu Y 2020 N2NSR-OCT: Simultaneous denoising and super-resolution in optical coherence tomography images using semi-supervised deep learning *J. Biophoton.* **14** [e202000282](#)
- Shi J, Liu Q P, Wang C F, Zhang Q, Ying S H and Xu H Y 2018 Super-resolution reconstruction of MR image with a novel residual learning network algorithm *Phys. Med. Biol.* **63** [085011](#)
- Tai Y, Yang J and Liu X M 2017 Image super-resolution via deep recursive residual network *Conference on Computer Vision and Pattern Recognition (CVPR) (Honolulu, HI, JUL 21-26)* [2790–8](#)
- Tan B Y, Wong A and Bizheva K 2018 Enhancement of morphological and vascular features in OCT images using a modified Bayesian residual transform *Biomed. Opt. Express* **9** [2394–406](#)
- Tao Y L, Tao L M, Jiang Z X, Liu H T, Liang K, Li M H, Zhu X S, Ren Y L and Cui B J 2017 Parameters of ocular fundus on spectral-domain optical coherence tomography for glaucoma diagnosis *Int. J. Ophthalmol. (Engl. Ed.)* **10** [982–91](#)
- Wang H D, Rivenon Y, Jin Y Y, Wei Z S, Gao R, Gunaydin H, Bentolila L A, Kural C and Ozcan A 2018 Deep learning enables cross-modality super-resolution in fluorescence microscopy *Nat. Methods* **16** [103–10](#)
- Wang J, Zhang M, Hwang T S, Bailey S T, Huang D, Wilson D J and Jia Y L 2017 Reflectance-based projection-resolved optical coherence tomography angiography [Invited] *Biomed. Opt. Express* **8** [1536–48](#)
- Woo S H, Park J, Lee J Y and Kweon I S 2018 CBAM: convolutional block attention module *European Conference on Computer Vision (ECCV) 11211 (Munich, GERMANY, SEP 08-14)* [3–19](#)
- Yang J C, Wright J, Huang T and Ma Y 2008 Image super-resolution as sparse representation of raw image patches *Conference on Computer Vision and Pattern Recognition (Anchorage, AK, JUN 23-28)* [1–8](#)
- Yousefi S, Zhi Z W and Wang R K 2011 Eigendecomposition-based clutter filtering technique for optical micro-angiography *IEEE Trans. Biomed. Eng.* **58** [2316–23](#)
- Zhang Q, Wang J and Wang R K 2016 Highly efficient eigen decomposition based statistical optical microangiography *Quant. Imaging Med. Surg.* **6** [557–63](#)
- Zhang Y L, Li K P, Li K, Wang L C, Zhong B N and Fu Y 2018 Image super-resolution using very deep residual channel attention networks *European Conference on Computer Vision (ECCV) (Munich, GERMANY, SEP 08-14)* [294–310](#)
- Zhang Y L, Tian Y P, Kong Y, Zhong B N and Fu Y 2018 Residual dense network for image super-resolution *Conference on Computer Vision and Pattern Recognition (CVPR) (Salt Lake City, UT, JUN 18-23)* [2472–81](#)
- Zhou T, Yang J L, Zhou K, Fang L Y, Hu Y, Cheng J, Zhao Y T, Chen X P, Gao S H and Liu J 2020 Digital resolution enhancement in low transverse sampling optical coherence tomography angiography using deep learning *arXiv:1910.01344*
- Zhu M H, Jiao L C, Liu F, Yang S Y and Wang J N 2021 Residual spectral-spatial attention network for hyperspectral image classification *IEEE Trans. Geosci. Remote Sens.* **59** [449–62](#)