

Axial super-resolution optical coherence tomography via complex-valued network

LINGYUN WANG,¹ SI CHEN,² LINBO LIU,² XUE YIN,³ GUOHUA SHI,^{4,*}
JIANHUA MO^{1,*}

¹*School of Electronics and Information Engineering, Soochow University, Suzhou, China*

²*School of Electrical and Electronic Engineering, Nanyang Technological University, Singapore*

³*The First Affiliated Hospital of Soochow University, Suzhou, China*

⁴*Jiangsu Key Laboratory of Medical Optics, Suzhou Institute of Biomedical Engineering and Technology, Suzhou, China*

**ghshi_lab@126.com*

**joshuamo@gmail.com*

Abstract:

Optical coherence tomography (OCT) is a fast and non-invasive optical interferometric imaging technique that can provide high-resolution cross-sectional images of biological tissues. OCT's key strength is its depth resolving capability which remains invariant along the imaging depth and is determined by the axial resolution. The axial resolution is inversely proportional to the bandwidth of the OCT light source. Thus, the use of broadband light sources can effectively improve the axial resolution and however leads to an increased cost. In recent years, real-valued deep learning technique has been introduced to obtain super-resolution optical imaging. In this study, we proposed a complex-valued super-resolution network (CVSR-Net) to achieve an axial super-resolution for OCT by fully utilizing the amplitude and phase of OCT signal. The method was evaluated on three OCT datasets. The results show that the CVSR-Net outperforms its real-valued counterpart with a better depth resolving capability. Furthermore, comparisons were made between our network, six prevailing real-valued networks and their complex-valued counterparts. The results demonstrate that the complex-valued network exhibits a better super-resolution performance than its real-valued counterpart and our proposed CVSR-Net achieved the best performance. In addition, the CVSR-Net was tested on out-of-distribution domain datasets and its super-resolution performance is well maintained as compared to that on source domain datasets, indicating a good generalization capability.

Keywords: Optical coherence tomography, Complex-valued network, Super-resolution

1. Introduction

Optical coherence tomography (OCT) is an optical interferometric imaging technique that can provide high-resolution three-dimensional imaging of weak-scattering biological tissues [1]. Over the past decades, many efforts have been made to translate OCT into diagnostic technique in various clinics, such as ophthalmology [2, 3], cardiology [4, 5], and dermatology [6, 7]. The success of OCT is mainly attributed to its good depth-resolving capability without physical depth scanning, which is usually delineated with axial resolution. A higher axial resolution allows for better capturing micro-structure features, which provides more detailed insights into the samples. Therefore, increasing the axial resolution of OCT is continuing to receive research efforts.

The axial resolution of an OCT (Δz) is determined by the center wavelength (λ_c) and bandwidth ($\Delta\lambda$) of the light source as illustrated by the equation: $\Delta z = 2 \ln 2 \lambda_c^2 / \pi \Delta\lambda$. Thus, increasing the spectral bandwidth of OCT light source helps obtain a better axial resolution. For example, Leitgeb et al. developed an ultrahigh resolution (UHR)-OCT with an axial resolution of $\sim 2.5 \mu\text{m}$ in tissue by employing a broad bandwidth Titanium:sapphire laser with

a full-width-at-half-maximum of 120 nm centered at 800 nm [8]. Similarly, Werkmeister et al. used a similar laser to obtain an axial resolution of $\sim 1.2 \mu\text{m}$ on human cornea [9]. However, the use of Titanium:sapphire laser leads to a significantly increased system cost. According to the equation above, the axial resolution is proportional to the central wavelength of OCT light source and consequently can be improved by employing a short-wavelength light source, e.g., visible light [10, 11]. Chong et al. designed a multi-functional fiber-based visible-light (Vis)-OCT for human retinal imaging with an axial resolution of $< 2 \mu\text{m}$ in tissue [12]. Pi et al. presented a fiber-based Vis-OCT with an axial resolution of $\sim 1.2 \mu\text{m}$ which allows for quantifying the thickness of large vessel walls in the chicken embryo [13]. However, Vis-OCT usually suffers from a reduced imaging depth than conventional near-infrared OCT for the higher scattering property of visible light in tissue.

In recent years, deep learning has been extensively applied to realize super-resolution optical imaging [14]. For example, Ozcan et al. demonstrated a superb performance of super-resolution in optical microscopy [15], scanning electron microscopy [16], and fluorescence microscopy [17]. Particularly, for OCT, deep learning has been utilized for developing axial super-resolution OCT. Huang et al. proposed a generative adversarial network (GAN) based on enhanced super-resolution GAN (ESRGAN) [18] to denoise and upscale OCT B-scan images (super-resolution) [19]. The semi-supervised deep learning approach was utilized for a similar purpose by using a dataset of paired noisy images [20]. This approach overcomes the drawback of supervised networks, i.e., the requirement of low quality-high quality pairs. However, down-sampling OCT B-scan image cannot mimic axial resolution degradation caused by the reduced spectral bandwidth of light sources. Hence, the upscaling process will probably not be applicable to generate super-resolution OCT image from low depth resolution image. Residual-in-residual dense block (RRDB) network [18] was employed to obtain axial super-resolution OCT image from the truncated spectra [21]. Liang et al. reported the use of conditional GAN to enhance the axial resolution of OCT image accompanied with a good recovery of realistic speckle [22]. Furthermore, deep learning has also been applied to OCT lateral super-resolution. For example, GAN was utilized for refocusing OCT *en-face* images by mapping of defocused and focused image pairs [23]. Note that OCT image reconstruction is typically obtained by Fourier-transforming of the interference spectra with the resulted image being complex-valued. However, the super-resolution methods based on deep learning above are implemented with OCT image intensities while the phase information is ignored, even though the phase of an image can provide structural information [24, 25]. Theoretically, complex-valued network is superior against its real-valued counterpart by fully utilizing both amplitude and phase [26]. Furthermore, complex-valued network is advantageous in easy optimization, fast learning, and noise-robust memory mechanisms [27-29]. It has been successfully applied to achieve super-resolution for Terahertz imaging [30], hyperspectral imaging [31], and millimeter-wave imaging [32].

In this paper, to the best of our knowledge, we are the first to propose a complex-valued network, namely, complex-valued super-resolution network (CVSR-Net) for realizing axial super-resolution of OCT imaging. CVSR-Net utilizes local residual connections, global residual connections, channel concatenation, and dense connections to extract the complex-valued image features and learn the mapping of low-resolution and high-resolution image pairs. The loss function was specially designed by summing weighted Charbonnier loss and multiscale structural similarity. The CVSR-Net was evaluated on three OCT datasets and spectral truncation was employed to produce low axial resolution image. To show the advantage of complex-valued network, comparisons were conducted among seven super-resolution networks and their real-valued counterparts. The results demonstrate that the complex-valued network can super-resolve images with more realistic speckles and structures than the real-valued network and CVSR-Net achieved the best performance. Next, the CVSR-Net and its real-valued counterpart RVSR-Net were tested on two micro-OCT datasets. The results show that the CVSR-Net could restore axial resolution well from the largely truncated

spectra with better image metrics than RVSR-Net on all the datasets. Last, the generalization ability of the CVSR-Net was evaluated by testing the model trained with the source-domain dataset (*ex vivo* micro-OCT images of swine esophagus) on out-of-distribution domain datasets (*ex vivo* micro-OCT images of swine cornea and *in vivo* swept-source OCT images of human retina). The CVSR-Net appeared to have a good generalization capability.

The main contributions of this work are summarized as follows:

1. We developed a complex-valued deep learning network for the first time to achieve axial super-resolution OCT which is demonstrated to be superior to real-valued deep learning network. The CVSR-Net is able to super-resolve fine structures by extracting features from both amplitude and phase of OCT signals.
2. We demonstrated the superiority of the complex-valued deep learning network against its real-valued counterpart on several prevailing super-resolution networks. The complex-valued networks can super-resolve images with more detailed structures than the corresponding real-valued networks. Our proposed network achieved the best performance among all those evaluated complex-valued and real-valued networks.
3. The good generalization capability of the CVSR-Net was demonstrated by testing the model trained with the source-domain dataset on out-of-distribution dataset (cross tissue/cross system).

2. Methods

2.1 Data acquisition and image pre-processing

Two datasets acquired by a micro-OCT prototype system [33, 34] and one dataset acquired by a home-built 1060nm swept-source OCT system were included in this study, including four *ex vivo* volumetric images of swine esophagus from four swine samples, four *in vivo* volumetric images of swine cornea from four swines, and four *in vivo* volumetric images of human retina images from one human subject. This study was approved by the ethics review board of The First Affiliated Hospital of Soochow University and abided by the tenets of the Declaration of Helsinki (clinical trial registration number: ChiCTR1800014446). The size of each volumetric image from the micro-OCT system and the swept-source OCT system is $1024 \times 1024 \times 512$ and $512 \times 1000 \times 500$ (A-scan \times B-scan \times C-scan), respectively. The pixel size of the image from the micro-OCT system and the swept-source OCT system is $0.85 \times 1.53 \mu\text{m}$ and $10 \times 7.4 \mu\text{m}$ (lateral \times axial), respectively. The OCT image reconstruction comprises the following steps: (1) reference spectrum subtraction, (2) k-linearization (exclusively for micro-OCT), (3) spectrum windowing with a Hanning function, (4) dispersion compensation, and (5) fast Fourier transform (FFT).

To create datasets with low axial resolution, the raw spectrum was truncated using a Hanning window with a size smaller than the spectrum length, e.g., quarter or half of the spectrum length. The truncated spectrum was zero-padded to make up the loss due to the truncation. Then, a complex-valued OCT image was obtained by FFT, called low-resolution (LR) image. The OCT image produced from the non-truncated spectra was taken as high-resolution (HR) image. LR and HR images were used as the input and ground truth (GT) of the deep learning network, respectively, for training, validation and testing. Due to the limitation of GPU memory, the original LR and GT images were randomly cropped to be 300×300 pixels. For each dataset, three volumes were taken for training (900 images) and validation (100 images) and one volume was used for testing (100 images) to ensure that they are independent and the training process is blind to the validation and testing process. The models for different types of tissues were trained separately.

2.2 Complex-valued network architecture

A) Complex-valued convolution layer

To perform the convolution in complex domain, for a complex-valued input image $I = x + iy$ and a complex-valued convolution kernel $W = A + iB$, the complex-valued convolution

(CConv) can be expressed as $W * I = (A * x - B * y) + i(B * x + A * y)$. The matrix notation of this equation is:

$$\begin{bmatrix} \Re(W * I) \\ \Im(W * I) \end{bmatrix} = \begin{bmatrix} A & -B \\ B & A \end{bmatrix} * \begin{bmatrix} x \\ y \end{bmatrix}. \quad (1)$$

This complex-valued convolution layer can be realized by utilizing two real-valued 2D convolution layers in Pytorch.

B) Complex-valued activation

Complex-valued activation can be obtained by splitting the activation function [26, 35], and then conducting the activation by adopting the real-valued activation. Thus, the complex-valued ReLU (CReLU) is given by:

$$\text{CReLU}(z) = \text{ReLU}(\Re(z)) + i\text{ReLU}(\Im(z)), \quad (2)$$

where $\text{ReLU}(\cdot)$ is the real-valued ReLU as follows:

$$\text{ReLU}(x) = \begin{cases} x & \text{if } x \geq 0 \\ 0 & \text{otherwise} \end{cases}. \quad (3)$$

C) Complex-valued super-resolution network architecture

Figure 1 shows the network architecture of our proposed CVSR-Net. The architecture was inspired by Residual Dense Network (RDN) [36, 37] and Residual Network (ResNet) [38]. The CVSR-Net consists of a shallow feature extraction part, a deep feature extraction part, and a final reconstruction part.

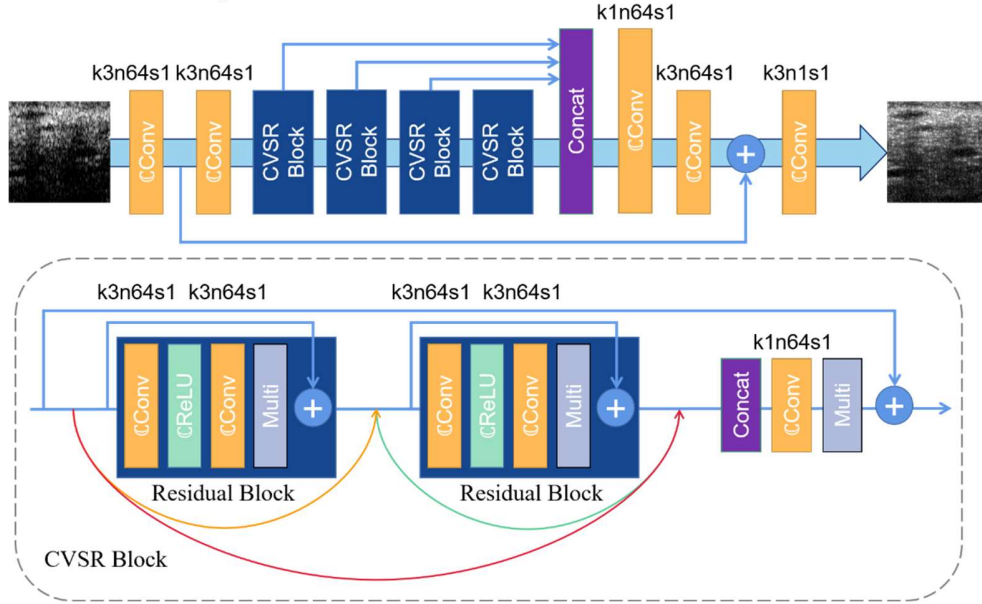


Fig. 1. The network architecture of the CVSR-Net.

In the shallow feature extraction part, two CConv layers with a kernel size of 3, 64 channels and a stride size of 1 learn the low-level features of the image. The features extracted by the first CConv layer are connected to the reconstruction part to achieve a global residual learning. This not only helps transfer the low-level features to the final image, but also avoids learning a complicated mapping of two images [39, 40].

The deep feature extraction comprises four CVSR blocks and a channel concatenation layer. As shown in Fig. 1, the basic unit in the CVSR block is built as the residual block in EDSR [41]. The two residual blocks are utilized to extract the high-level features. The red, orange, and green curves are referred to dense connections as used in DenseNet [36]. Therefore, the

feature-maps in the CVSR block can be fully used and fused by the dense connections. Next, the feature channel is recompressed to 64 by a 1×1 CConv layer.

The multi in Fig. 1 means that the feature-maps are multiplied by a residual scaling factor to make the training process more stable [42]. Before the feature-map is transferred to the next CVSR block, the input feature-map is added with the local residual learning to reduce the degradation and improve the learning ability of the network [43, 44]. The feature-maps produced by four CVSR blocks are concatenated at the end of the deep feature extraction part to fuse the features at various levels.

Finally, the 1×1 CConv layer compresses the channel number of the feature from 256 to 64, and the final image can be generated by two CConv layers with a kernel size of 3. Note that the output of the proposed CVSR-Net is in complex-value domain. To conduct the backpropagation and visualize the image, the amplitude of the output is utilized.

2.3 Quantitative evaluation

To evaluate the performance of our proposed method, three commonly used image quality metrics were employed: peak signal-to-noise ratio (PSNR), structural similarity (SSIM) index, and learned perceptual image patch similarity (LPIPS) [40, 45, 46]. PSNR can reflect the pixel-wise difference between two images and is defined as follows:

$$\text{PSNR} = 10 \log_{10} \left(\frac{\max(I)^2}{\frac{1}{N} \sum_{i=1}^N (I(i) - I_{GT}(i))^2} \right), \quad (4)$$

where I is the processed image, e.g., LR image or SR image, and I_{GT} represents the GT image.

However, PSNR cannot completely depict the image resolution enhancement as only pixel-wise difference is considered, and consequently an individual use of PSNR may not be reliable [40]. Therefore, SSIM was employed to provide a good evaluation of the perceptual similarity. SSIM is calculated by incorporating luminance, contrast, and structure of two images as defined in the following equations:

$$l = \frac{2\mu_I \mu_{I_{GT}} + C_1}{\mu_I^2 + \mu_{I_{GT}}^2 + C_1}, \quad (5)$$

$$c = \frac{2\sigma_I \sigma_{I_{GT}} + C_2}{\sigma_I^2 + \sigma_{I_{GT}}^2 + C_2}, \quad (6)$$

$$s = \frac{\sigma_{II_{GT}} + C_3}{\sigma_I \sigma_{I_{GT}} + C_3}, \quad (7)$$

where l , c , and s represent luminance, contrast, and structures; μ_I and $\mu_{I_{GT}}$ are the mean pixel values of the processed image and the GT image; σ_I and $\sigma_{I_{GT}}$ represent their standard deviations; $\sigma_{II_{GT}}$ is the covariance. The constants- C_1 , C_2 , and C_3 are 0.01, 0.03 and 0.015, respectively. α , β , and γ are set to be 1. Then, SSIM can be obtained according to the following formula:

$$\text{SSIM} = [l]^\alpha [c]^\beta [s]^\gamma = \frac{(2\mu_I \mu_{I_{GT}} + C_1)(2\sigma_{II_{GT}} + C_2)}{(\mu_I^2 + \mu_{I_{GT}}^2 + C_1)(\sigma_I^2 + \sigma_{I_{GT}}^2 + C_2)}, \quad (8)$$

LPIPS is another often-used metric dedicated to evaluating the perceptual similarity of two images. Other than calculating the difference in the pixel level, LPIPS assesses the images based on the high-level features extracted by a trained CNN. Therefore, LPIPS can provide an

image similarity evaluation complementary to PSNR and SSIM. The small LPIPS value represents the high similarity between the two images.

2.4 Loss function

To guide the model optimization, a loss function was designed by combining Charbonnier loss and multiscale structural similarity (MS-SSIM). Charbonnier loss is a variant of L1 loss and can handle outliers better than L1 loss. It has shown an improved performance and convergence as compared to L2 loss [47]. However, Charbonnier loss is a loss calculated from the pixel-wise difference while not considering the structure variation. Consequently, this loss function can maximize PSNR but may lead to an over-smoothing which may smear fine structures [46]. To address the limitation of Charbonnier Loss, MS-SSIM was combined with Charbonnier loss through a linear addition. MS-SSIM is yielded from implementing SSIM at multiple scales, measuring the image details at different resolutions [48, 49]. Moreover, MS-SSIM loss avoids the selection of the standard deviation of the Gaussian filter which may influence the results [49]. Thus, in principle, the new loss function should be able to preserve the pixel intensity well (Charbonnier loss), and simultaneously increase the similarity between two images and retain the texture well (MS-SSIM loss function). The three loss functions above are given by:

$$\mathcal{L}_{\text{Charbonnier}} = \frac{1}{hw} \sum_{i,j} \sqrt{(I_{i,j} - I_{GT,i,j})^2 + \epsilon^2}, \quad (9)$$

$$\mathcal{L}_{\text{MS-SSIM}} = 1 - [l_M]^{\alpha_M} \prod_{j=1}^M [c_j]^{\beta_j} [s_j]^{\gamma_j}, \quad (10)$$

$$\mathcal{L}_{\text{our}} = k_1 \cdot \mathcal{L}_{\text{Charbonnier}} + k_2 \cdot \mathcal{L}_{\text{MS-SSIM}}, \quad (11)$$

where h and w are the height and width of the image; ϵ is set to 0.003 to make Charbonnier loss more robust for outliers; M means the highest scale; α_M , β_j and γ_j are all set to be 1. k_1 and k_2 are empirically set to be 0.16 and 0.84, respectively.

2.5 Implementation details

In this work, three LR datasets were created with various degraded axial resolutions through spectrally truncating the raw data with percentages of 25%, 37.5%, and 50%. In the training process, all the networks were optimized by Adam optimizer with $\beta_1 = 0.9$ and $\beta_2 = 0.999$ [50]. The initial learning rate was set to be 1e-3 for all the layers and decayed by a factor of two every twenty epochs. The batch size was two and all the networks were trained by 100 epochs (45,000 iterations) to guarantee a convergence. Each dataset with different axial resolution was trained separately to realize a mapping of a low axial resolution to its corresponding high axial resolution.

All the experiments were implemented in Python 3.7 with Pytorch 1.13.1 deep learning framework on a server with a 2.4 GHz Intel Xeon Silver 4210R CPU and NVIDIA RTX 3090.

3. Results and discussion

The proposed CVSR-Net trained with the swine esophagus dataset was first evaluated on the *ex vivo* swine esophagus test set from micro-OCT. Three low-resolution image datasets were created by truncating the spectra using a Hanning window function with three truncation percentages (truncated/non-truncated)-25%, 37.5% and 50%, corresponding to Figs. 2(a)-2(c), respectively. It is clearly seen that all the three images above are blurred and exhibit axially enlarged speckles as compared with the GT image (Fig. 2(g)). Moreover, the image degrades more as the truncation percentage is decreased from 50% to 25%. For example, as the truncation percentage decreases, the grainy speckle pattern becomes coarser which can be explained by the broadened axial coherence function due to the spectral truncation. Meanwhile, in the region of interest (ROI, the region framed with rectangle), the flat blood vessel's (denoted by red arrow)

boundary undergoes a larger destroy, resulting a decreased vessel visibility. Figures 2(d)-2(f) describe the SR images produced by the CVSR-Net with Figs. 2(a)-2(c) as inputs, respectively. Overall, for all the three different truncation percentages, the images were recovered well to be close to the GT image. In details, similar speckle patterns and image sharpness were obtained. This can also be illustrated by the depth profile along the dashed red line in Fig. 2(g) as shown in Fig. 3. Obviously, more structure loss occurs as the truncation percentage increases, for example, some neighboring peaks are merged into a broad peak, i.e., in the region denoted by dashed rectangle. The SR image appears to be well overlapped with the GT image for the truncation percentage of 37.5%. This can be further improved on the 50% truncated data. It is also observed that the signal-to-noise (SNR) loss in the deep region resulting from the reduced axial resolution was restored well which is manifested as the increased peaks (SR images). For the truncation percentage of 25%, although peaks are not fully recovered, the SNR is significantly increased, indicating an axial resolution enhancement.

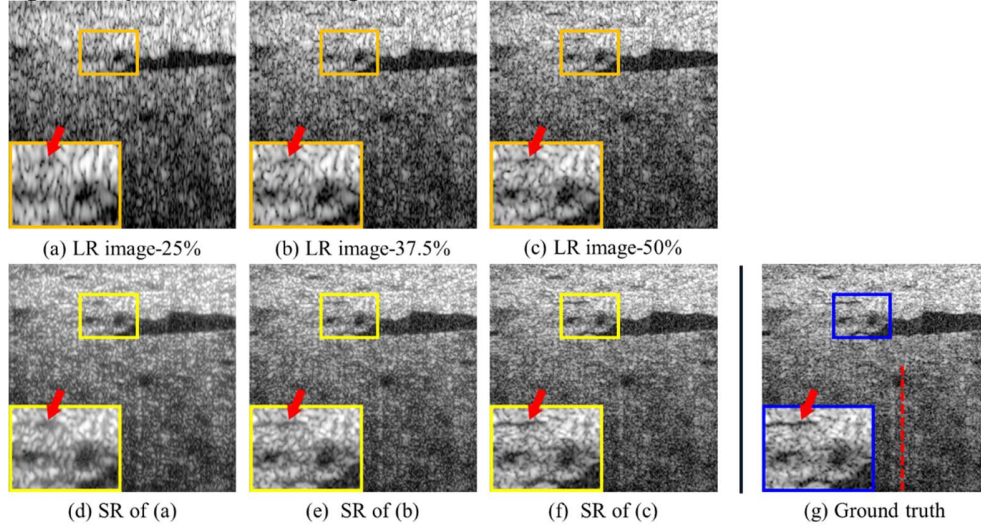


Fig. 2 Super-resolution results of the CVSR-Net on *ex vivo* swine esophagus test set: (a)-(c) are LR images reconstructed from 25%, 37.5% and 50% spectrally truncated data, respectively; (d)-(f) are the corresponding resolution-enhanced images generated by the CVSR-Net; (g) is the Ground truth.

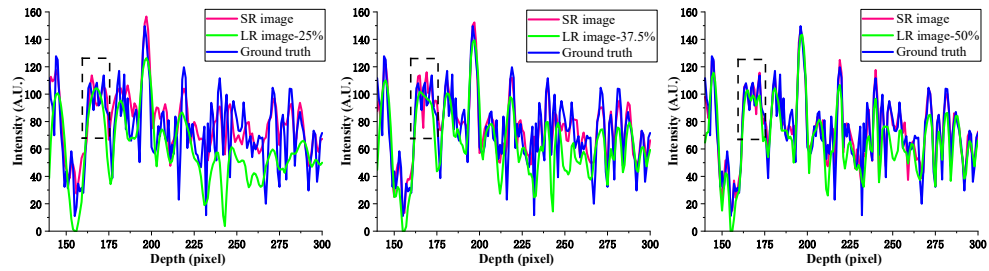


Fig. 3 Depth profiles along the red dashed line in Fig. 2(g) and its corresponding trajectories in Figs. 2(a)-2(c) and 2(d)-2(f).

To quantitatively assess the performance of CVSR-Net, three image metrics, including PSNR, SSIM and LPIPS, were calculated and summarized in Tab. 1. It is found that for all the three truncation percentages, PSNR, SSIM and LPIPS were significantly improved by the CVSR-Net. Moreover, the improvement extent appears to be comparable among all the three LR images, about 3.5-6 for PSNR, 0.2-0.25 for SSIM and 0.2-0.3 for LPIPS. As a result, the CVSR-Net did not change the rank of the three metrics of SR images generated from the three

datasets with different truncation percentages. The image restored from LR image-50% always shows the best image quality, i.e., axial resolution, while the images from LR image-37.5% and LR image-25% are ranked at the second and third, respectively. This is consistent with the perceptual findings from the images in Fig. 2. In addition, it is worth mentioning that the image restored from LR image-50% has a PSNR-26.95, SSIM-0.9098, and LPIPS-0.0948, indicating that the enhanced image is very close to the GT image.

Table 1. Image metrics of the *ex vivo* swine esophagus test set.

Net	PSNR	SSIM	LPIPS
LR Image-25%	16.83	0.3439	0.5893
CVSR-Net	20.47	0.5404	0.4114
LR Image-37.5%	18.77	0.5298	0.4135
CVSR-Net	23.12	0.7648	0.2288
LR Image-50%	20.97	0.6993	0.2779
CVSR-Net	26.95	0.9098	0.0948

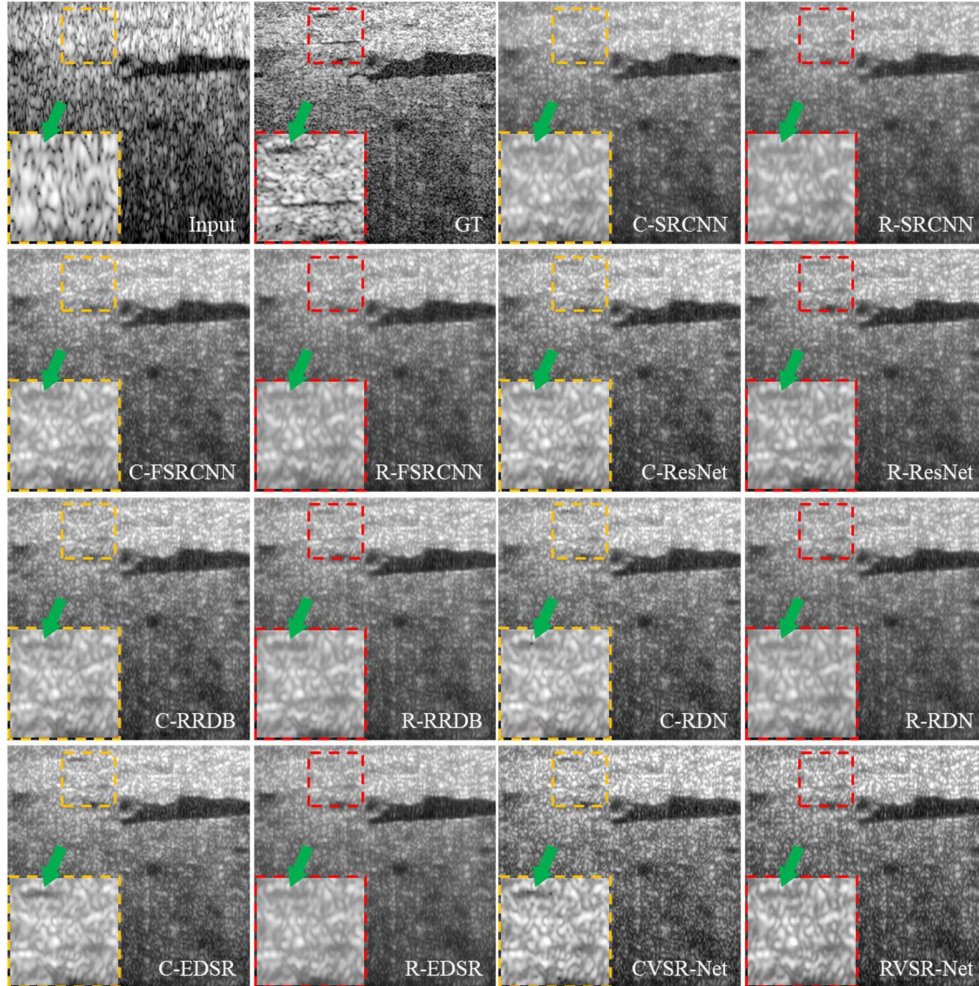


Fig. 4 Comparisons between the complex-valued networks and the real-valued networks. C- and R- represents complex-valued and real-valued, respectively. The input is the image with a 25% spectral truncation.

Prior to further applying the CVSR-Net to OCT imaging of different biological tissues, the superiority of the CVSR-Net against its real-valued counterpart was verified extensively. To do so, a real-valued super-resolution network (RVSR-Net) was designed with the same network architecture as CVSR-Net. To make the comparison between the complex-valued network and the real-valued network fair and generic, six prevailing SR networks, including EDSR, RDN, RRDB network [18], ResNet [51], fast super-resolution convolutional neural network (FSRCNN) [52], super-resolution convolutional neural network (SRCNN) [39] and their complexed counterparts are included. It should be noted that the upsampling layers in the above networks were removed as the size of each image is kept constant in the training and testing processes. In addition, each network and its complex-valued counterpart are trained with the loss function from the original papers on the *ex vivo* swine esophagus dataset and all the training processes were conducted in the same way as mentioned in section 2.5.

Figure 4 describes the comparisons of SR images produced by the complex-valued networks and the real-valued networks. The image reconstructed from 25% truncated spectral data was used as the input of the network. It is obvious that the images produced by both complex-valued networks and real-valued networks show similar speckle pattern and image sharpness. Besides, the SR image looks smoother as compared to the GT image, indicating that the image reconstruction by deep learning network comes with a denoising effect in addition to the resolution enhancement. This may be because the noise distribution and intensity of the LR image and the GT image are different, and the networks fail to learn the noise mapping. The denoising effect makes the morphological structure more distinguishable in the SR image than the GT image. Among all these seven paired networks, the CVSR-Net shows the weakest denoising effect and consequently the generated SR images look most like the GT image. In addition, the CVSR-Net achieved the best performance in reconstructing the blood vessel (pointed out by green arrow) in the ROI. However, it may not be easy to judge whether the complexed-value network performs better than the real-valued network from the images except the CVSR-Net and the C-EDSR.

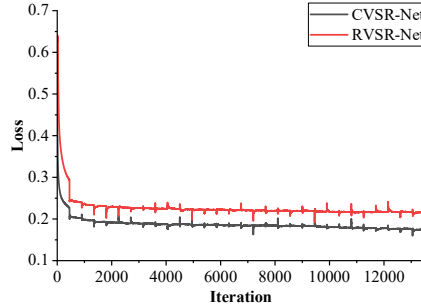


Fig. 5 The loss-iteration functions of training the CVSR-Net and the RVSR-Net.

Therefore, PSNR, SSIM and LPIPS were computed to make an objective and reliable comparison between complexed-value network and real-valued network as depicted in Tab. 2. Each complex-valued network achieves higher PSNR, higher SSIM, and lower LPIPS as compared with its real-valued counterpart. The complexed-value network's superiority is more obvious in SSIM and LPIPS than in PSNR. This may be because the phase of the signal contains structure information of an image, and therefore by incorporating the phase in the learning process, the complexed-value network has a better capability to recover those high-frequency structures. The PSNR of the image by the real-valued network is close to that of the complexed-valued network, indicating that both networks appear to have a comparable denoising capability. In addition, the CVSR-Net yields the highest SSIM and the smallest LPIPS, demonstrating that the CVSR-Net performs better than the other networks in axial resolution enhancement. On the other hand, the complex-valued network also has the advantage of easier optimization over the real-valued network. Figure 5 shows the loss of training before overfitting for the CVSR-Net

and the RVSR-Net. It is obvious that the CVSR-Net converged more quickly and had a smaller loss than the RVSR-Net.

Table 2. Comparisons of image metrics between complex-valued networks and real-valued networks on the *ex vivo* swine esophagus test set.

Net	PSNR	SSIM	LPIPS
LR Image-25%	16.83	0.3439	0.5893
CVSR-Net	20.47	0.5404	0.4114
RVSR-Net	20.07	0.4796	0.4716
C-EDSR	20.75	0.5061	0.4433
R-EDSR	20.29	0.4218	0.5171
C-RDN	20.68	0.4951	0.4699
R-RDN	20.22	0.4233	0.5212
C-RRDB	20.48	0.4680	0.4820
R-RRDB	20.27	0.4297	0.5107
C-ResNet	20.60	0.4760	0.4808
R-ResNet	20.26	0.4287	0.5116
C-FSRCNN	20.49	0.4604	0.5089
R-FSRCNN	20.07	0.3988	0.5400
C-SRCNN	20.37	0.4498	0.4835
R-SRCNN	19.97	0.3862	0.5482

Next, we evaluated the super-resolution performance of both the CVSR-Net and the RVSR-Net on the *in vivo* swine cornea test set from micro-OCT. As can be seen in Fig. 6, the corneal stroma is comprised of collagen bundles (stacked-layer structures) which is delineated clearly in the GT image. When the spectra were truncated by 25%, it is hard to identify the collagen bundles in the resulted image due to the degraded axial resolution. The axial resolution degradation caused by the spectral truncation is well illustrated by the thickening of the edge of the cornea. With the increase of the truncation to 50%, the visibility of the collagen bundles is highly improved and however remains far lower than in the GT image. By using either the CVSR-Net or the RVSR-Net, the collagen bundles can be recovered with a fairly good visibility for all the three truncation percentages. To better visualize the difference of super-resolution performance between the CVSR-Net and the RVSR-Net, a small rectangular region was chosen and zoomed in as framed with various colors (red: input, orange: output of the RVSR-Net, green: output of the CVSR-Net, blue: Ground truth). It is seen that the CVSR-Net outperforms the RVSR-Net. In particular, the output image of the CVSR-Net bears a higher structural similarity to the Ground truth than that of the RVSR-Net. All the findings above were confirmed by the quantitative metrics, PSNR, SSIM and LPIPS as shown in Tab. 3. All the image metrics are significantly improved by the CVSR-Net and the RVSR-Net as compared to the low-resolution image. Meanwhile, the best metrics always occur on the images output from the CVSR-Net for each truncation percentage. For example, for the images reconstructed from the data with 50% spectral truncation, although the SSIM was increased from 0.6925 to 0.8099 by the RVSR-Net, it was further improved to 0.9142 by the CVSR-Net. A similar trend was observed on the LPIPS which underwent a decrease from 0.2796 (LR image) to 0.1457 (RVSR-Net) and further to 0.0616 (CVSR-Net). This indicates that for a truncation percentage of 50%, the CVSR-Net can restore an image that is almost identical to the GT.

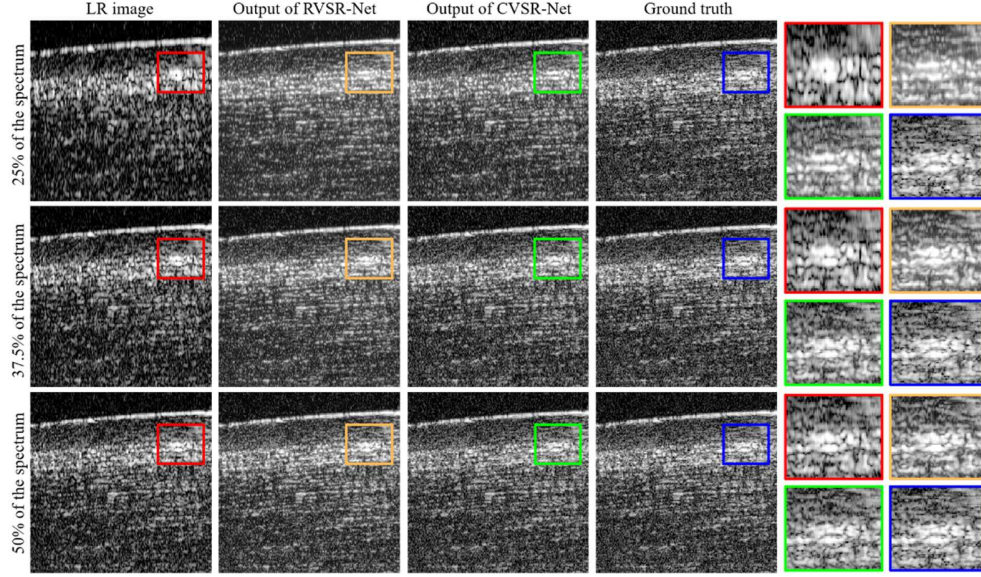


Fig. 6 Super-resolution results of the CVSR-Net and the RVSr-Net on the *in vivo* swine cornea test set.

Table 3. Image metrics of the *in vivo* swine cornea test set.

Net	PSNR	SSIM	LPIPS
LR Image-25%	13.80	0.3281	0.6194
RVSr-Net	16.69	0.4499	0.4489
CVSR-Net	17.19	0.5377	0.3669
LR Image-37.5%	15.72	0.5206	0.4428
RVSr-Net	18.51	0.6525	0.2676
CVSR-Net	19.90	0.7701	0.1623
LR Image-50%	17.87	0.6925	0.2796
RVSr-Net	20.86	0.8099	0.1457
CVSR-Net	24.11	0.9142	0.0616

Last, the CVSR-Net was evaluated from the aspect of the generalization ability. The *ex vivo* swine esophagus dataset (37.5% spectral truncation) was taken as the source domain dataset. The evaluation was conducted with two different out-of-distribution domain test sets as illustrated in Fig. 7. One is the *in vivo* swine cornea test set (37.5% spectral truncation) acquired by the same micro-OCT system as the source domain dataset. The other out-of-distribution domain test set is the *in vivo* human retina test set acquired by 1060nm swept-source OCT system. Figures 7(c) and 7(d) describe the SR images of the cornea generated by the model trained with the source domain dataset and the out-of-distribution domain dataset, respectively. It can be seen that the SR image from the out-of-distribution domain model shows a good recovery of the image as compared to the GT image (Fig. 7(a)) while a slightly reduced sharpness was observed as compared to that from the source domain model. The fine adjacent collagen bundles were discernible. Similar findings can be obtained from the human retina results. Overall, the retinal layer structure and choroid vasculature were resolved well in the SR image by the model trained with the source domain dataset. However, due to the reduced image sharpness, some edge structure is not as clear as in the SR image by model trained with the out-of-distribution domain dataset, such as the upper edge of the inner limiting membrane. Table 4

summarizes the quantitative metrics of the images in Fig. 7. It is observed that, as compared to the model trained with the out-of-distribution domain dataset, SSIM dropped obviously for the model trained with source domain dataset for both two test sets, PSNR decreases for human retinal test set and LPIPS increases for swine cornea test set. This is reasonably acceptable as the metrics change is far less than the improvement against the LR images and the SR performance for out-of-distribution domain is visually close to that for source domain. The results demonstrate that the proposed CVSR-Net has a good generalization capability and can be potentially used to super-resolve unseen OCT data.

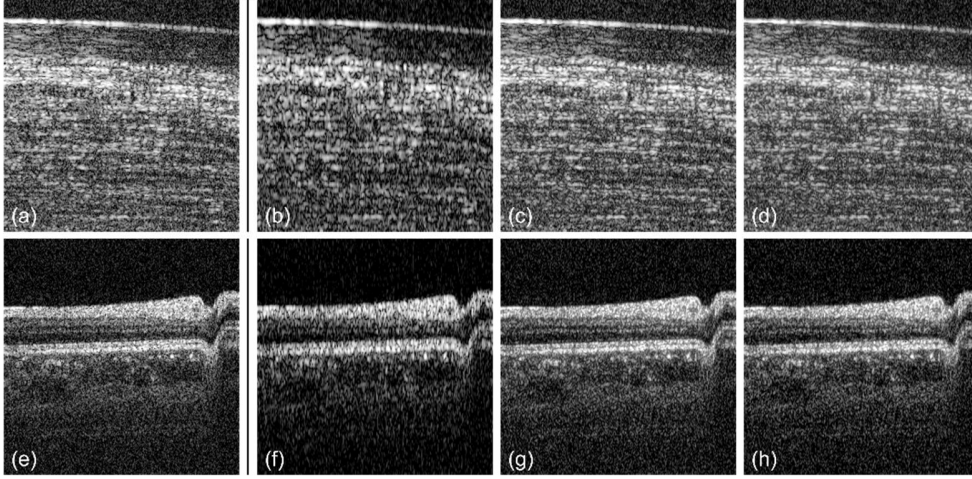


Fig. 7 Generalization performance of the CVSR-Net: (a) and (e) are the GT images of the *in vivo* swine cornea dataset and the *in vivo* human retina dataset respectively; (b) and (f) are the LR images; (c) and (g) are the SR images generated by the model trained with their respective source domain datasets; (d) and (h) are the SR images generated by the model trained with the *ex vivo* swine esophagus dataset. The LR images are created by the 37.5% spectral truncation.

Table 4. Image metrics of the SR images from both source-domain and out-of-distribution domain.

Tissue	Image	PSNR	SSIM	LPIPS
Swine Cornea	LR Image	15.72	0.5206	0.4428
	SR Image: source-domain	19.90	0.7701	0.1623
	SR Image: out-of-distribution domain	19.88	0.7451	0.2019
Human Retina	LR Image	17.68	0.3824	0.5274
	SR Image: source-domain	21.51	0.7243	0.2160
	SR Image: out-of-distribution domain	20.93	0.6883	0.2130

4. Conclusion

In this paper, we proposed a complex-valued super-resolution network (CVSR-Net) to achieve axial super-resolution for OCT imaging. The CVSR-Net's performance was evaluated on three micro-OCT datasets of various biological tissues. The results demonstrated that the CVSR-Net has a good capability to recover the axial resolution and the speckles from spectrally truncated data. Moreover, the CVSR-Net was also proven to be superior to its counterpart-RVSR-Net. This can be accounted for by the fact that the CVSR-Net can resolve more detailed structures of images by fully utilizing not only the amplitude but also the phase of OCT signals while the real-valued network only uses the amplitude. In addition, the superior performance of the complex-valued network against the real-valued network was verified on other prevailing super-resolution deep learning networks, including EDSR, RDN, RRDB, ResNet, FSRCNN, and SRCNN. Among all the complex-valued and real-valued networks, the CVSR-Net yielded

the best super-resolution performance. Last, the CVSR-Net was proven to have a good generalization ability through comparing the performance of the models trained with source domain datasets and out-of-distribution domain datasets. Thus, it is reasonable to conclude that the CVSR-Net has a good potential to aid in developing axial super-resolution OCT imaging and may be able to help reduce the OCT system cost by using narrow-band light source while retaining the axial resolution.

Funding. National key R&D program of China (2022YFC2404204); National natural science foundation of China (81401451); Suzhou new clinical diagnosis and treatment technology and public health project (SKY2022040); Ministry of Health's National Medical Research Council Open Fund - Individual Research Grant (MOH-OFIRG19may-0009); Ministry of Education Academic Research Fund Tier 1 (RG35/22); Ministry of Education Academic Research Fund Tier 2 (MOE-T2EP30120-0001); China-Singapore International Joint Research Institute (203-A022001).

Declaration of competing interest. The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Data availability. Data underlying the results presented in this paper are not publicly available at this time but may be obtained from the authors upon reasonable request.

References

1. D. Huang, E. A. Swanson, C. P. Lin, J. S. Schuman, W. G. Stinson, W. Chang, M. R. Hee, T. Flotte, K. Gregory, and C. A. Puliafito, "Optical coherence tomography," *Science* **254**, 1178-1181 (1991).
2. T. C. Chen, B. Cense, M. C. Pierce, N. Nassif, B. H. Park, S. H. Yun, B. R. White, B. E. Bouma, G. J. Tearney, and J. F. De Boer, "Spectral domain optical coherence tomography: ultra-high speed, ultra-high resolution ophthalmic imaging," *Archives of Ophthalmology* **123**, 1715-1720 (2005).
3. C. A. Puliafito, M. R. Hee, C. P. Lin, E. Reichel, J. S. Schuman, J. S. Duker, J. A. Izatt, E. A. Swanson, and J. G. Fujimoto, "Imaging of macular diseases with optical coherence tomography," *Ophthalmology* **102**, 217-229 (1995).
4. F. Alfonso, M. Paulo, N. Gonzalo, J. Dutary, P. Jimenez-Quevedo, V. Lennie, J. Escaned, C. Bañuelos, R. Hernandez, and C. Macaya, "Diagnosis of spontaneous coronary artery dissection by optical coherence tomography," *Journal of the American College of Cardiology* **59**, 1073-1079 (2012).
5. T. Kume, T. Akasaka, T. Kawamoto, Y. Ogasawara, N. Watanabe, E. Toyota, Y. Neishi, R. Sukmawan, Y. Sadahira, and K. Yoshida, "Assessment of coronary arterial thrombus by optical coherence tomography," *The American Journal of Cardiology* **97**, 1713-1717 (2006).
6. T. Gambichler, G. Moussa, M. Sand, D. Sand, P. Altmeyer, and K. Hoffmann, "Applications of optical coherence tomography in dermatology," *Journal of Dermatological Science* **40**, 85-94 (2005).
7. J. Welzel, E. Lankenau, R. Birngruber, and R. Engelhardt, "Optical coherence tomography of the human skin," *Journal of the American Academy of Dermatology* **37**, 958-963 (1997).
8. R. Leitgeb, W. Drexler, A. Unterhuber, B. Hermann, T. Bajraszewski, T. Le, A. Stingl, and A. Fercher, "Ultrahigh resolution Fourier domain optical coherence tomography," *Optics Express* **12**, 2156-2165 (2004).
9. R. M. Werkmeister, S. Sapeta, D. Schmidl, G. Garhöfer, G. Schmidinger, V. A. Dos Santos, G. C. Aschinger, I. Baumgartner, N. Pircher, and F. Schwarzhans, "Ultrahigh-resolution OCT imaging of the human cornea," *Biomedical Optics Express* **8**, 1221-1239 (2017).
10. B. Povazay, A. A. Apolonski, A. Unterhuber, B. Hermann, K. K. Bizheva, H. Sattmann, P. S. J. Russell, F. Krausz, A. F. Fercher, and W. Drexler, "Visible light optical coherence tomography," in *Coherence Domain Optical Methods in Biomedical Science and Clinical Applications VI*, (2002), 90-94.
11. X. Shu, L. Beckmann, and H. F. Zhang, "Visible-light optical coherence tomography: a review," *Journal of Biomedical Optics* **22**, 121707-121707 (2017).
12. S. P. Chong, M. Bernucci, H. Radhakrishnan, and V. J. Srinivasan, "Structural and functional human retinal imaging with a fiber-based visible light OCT ophthalmoscope," *Biomedical Optics Express* **8**, 323-337 (2017).
13. S. Pi, A. Camino, M. Zhang, W. Cepurna, G. Liu, D. Huang, J. Morrison, and Y. Jia, "Angiographic and structural imaging using high axial resolution fiber-based visible-light OCT," *Biomedical Optics Express* **8**, 4595-4608 (2017).
14. Y. Li, B. Sixou, and F. Peyrin, "A review of the deep learning methods for medical images super resolution problems," *IRBM* **42**, 120-133 (2021).
15. Y. Rivenson, Z. Göröcs, H. Günaydin, Y. Zhang, H. Wang, and A. Ozcan, "Deep learning microscopy," *Optica* **4**, 1437-1443 (2017).
16. K. de Haan, Z. S. Ballard, Y. Rivenson, Y. Wu, and A. Ozcan, "Resolution enhancement in scanning electron microscopy using deep learning," *Scientific Reports* **9**, 1-7 (2019).

17. H. Wang, Y. Rivenson, Y. Jin, Z. Wei, R. Gao, H. Günaydin, L. A. Bentolila, C. Kural, and A. Ozcan, "Deep learning enables cross-modality super-resolution in fluorescence microscopy," *Nature Methods* **16**, 103-110 (2019).
18. X. Wang, K. Yu, S. Wu, J. Gu, Y. Liu, C. Dong, Y. Qiao, and C. Change Loy, "EsrGAN: Enhanced super-resolution generative adversarial networks," in *Proceedings of the European Conference on Computer Vision Workshops*, (2018), 0-0.
19. 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 generative adversarial network," *Optics Express* **27**, 12289-12307 (2019).
20. B. Qiu, Y. You, Z. Huang, X. Meng, Z. Jiang, C. Zhou, G. Liu, K. Yang, Q. Ren, and Y. Lu, "N2NSR - OCT: Simultaneous denoising and super-resolution in optical coherence tomography images using semisupervised deep learning," *Journal of Biophotonics* **14**, e202000282 (2021).
21. Z. Yuan, D. Yang, H. Pan, and Y. Liang, "Axial super-resolution study for optical coherence tomography images via deep learning," *IEEE Access* **8**, 204941-204950 (2020).
22. K. Liang, X. Liu, S. Chen, J. Xie, W. Q. Lee, L. Liu, and H. K. Lee, "Resolution enhancement and realistic speckle recovery with generative adversarial modeling of micro-optical coherence tomography," *Biomedical Optics Express* **11**, 7236-7252 (2020).
23. Z. Yuan, D. Yang, Z. Yang, J. Zhao, and Y. Liang, "Digital refocusing based on deep learning in optical coherence tomography," *Biomedical Optics Express* **13**, 3005-3020 (2022).
24. T. Huang, J. Burnett, and A. Deczky, "The importance of phase in image processing filters," *IEEE Transactions on Acoustics, Speech, and Signal Processing* **23**, 529-542 (1975).
25. F. A. Wichmann, D. I. Braun, and K. R. Gegenfurtner, "Phase noise and the classification of natural images," *Vision Research* **46**, 1520-1529 (2006).
26. C. Trabelsi, O. Bilaniuk, Y. Zhang, D. Serdyuk, S. Subramanian, J. F. Santos, S. Mehri, N. Rostamzadeh, Y. Bengio, and C. J. Pal, "Deep Complex Networks," in *International Conference on Learning Representations*, (2017).
27. T. Nitta, "On the critical points of the complex-valued neural network," in *Proceedings of the 9th International Conference on Neural Information Processing. ICONIP'02.*, (2002), 1099-1103.
28. I. Danihelka, G. Wayne, B. Uria, N. Kalchbrenner, and A. Graves, "Associative long short-term memory," in *International Conference on Machine Learning*, (2016), 1986-1994.
29. S. Wisdom, T. Powers, J. Hershey, J. Le Roux, and L. Atlas, "Full-capacity unitary recurrent neural networks," *Advances in Neural Information Processing Systems* **29**, (2016).
30. Y. Wang, F. Qi, and J. Wang, "Terahertz image super-resolution based on a complex convolutional neural network," *Optics Letters* **46**, 3123-3126 (2021).
31. N. Aburaed, M. Q. Alkhatib, S. Marshall, J. Zabalza, and H. Al Ahmad, "Complex-valued neural network for hyperspectral single image super resolution," in *Hyperspectral Imaging and Applications II*, (2023), 102-109.
32. R. Sharma, J. Zhang, R. Kumar, B. Deka, V. Fusco, and O. Yurduseven, "3-D Super-Resolution of Coded Aperture Millimeter-Wave Images Using Complex-Valued Convolutional Neural Network," *IEEE Sensors Journal* **22**, 20921-20936 (2022).
33. S. Chen, X. Liu, N. Wang, Q. Ding, X. Wang, X. Ge, E. Bo, X. Yu, H. Yu, and C. Xu, "Contrast of nuclei in stratified squamous epithelium in optical coherence tomography images at 800 nm," *Journal of Biophotonics* **12**, e201900073 (2019).
34. L. Liu, J. A. Gardecki, S. K. Nadkarni, J. D. Toussaint, Y. Yagi, B. E. Bouma, and G. J. Tearney, "Imaging the subcellular structure of human coronary atherosclerosis using 34-optical coherence tomography," *Nature Medicine* **17**, 1010-1014 (2011).
35. C. Lee, H. Hasegawa, and S. Gao, "Complex-valued neural networks: A comprehensive survey," *IEEE/CAA Journal of Automatica Sinica* **9**, 1406-1426 (2022).
36. G. Huang, Z. Liu, L. Van Der Maaten, and K. Q. Weinberger, "Densely connected convolutional networks," in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, (2017), 4700-4708.
37. Y. Zhang, Y. Tian, Y. Kong, B. Zhong, and Y. Fu, "Residual dense network for image super-resolution," in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, (2018), 2472-2481.
38. K. He, X. Zhang, S. Ren, and J. Sun, "Deep residual learning for image recognition," in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, (2016), 770-778.
39. C. Dong, C. C. Loy, K. He, and X. Tang, "Learning a deep convolutional network for image super-resolution," in *Proceedings of the European Conference on Computer Vision*, (2014), 184-199.
40. Z. Wang, J. Chen, and S. C. Hoi, "Deep learning for image super-resolution: A survey," *IEEE Transactions on Pattern Analysis and Machine Intelligence* **43**, 3365-3387 (2020).
41. B. Lim, S. Son, H. Kim, S. Nah, and K. Mu Lee, "Enhanced deep residual networks for single image super-resolution," in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition Workshops*, (2017), 136-144.
42. C. Szegedy, S. Ioffe, V. Vanhoucke, and A. Alemi, "Inception-v4, inception-resnet and the impact of residual connections on learning," in *Proceedings of the AAAI Conference on Artificial Intelligence*, (2017).

43. X. Mao, C. Shen, and Y.-B. Yang, "Image restoration using very deep convolutional encoder-decoder networks with symmetric skip connections," *Advances in Neural Information Processing Systems* **29**, (2016).
44. J. Li, F. Fang, K. Mei, and G. Zhang, "Multi-scale residual network for image super-resolution," in *Proceedings of the European Conference on Computer Vision*, (2018), 517-532.
45. R. Zhang, P. Isola, A. A. Efros, E. Shechtman, and O. Wang, "The unreasonable effectiveness of deep features as a perceptual metric," in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, (2018), 586-595.
46. Z. Wang, A. C. Bovik, H. R. Sheikh, and E. P. Simoncelli, "Image quality assessment: from error visibility to structural similarity," *IEEE Transactions on Image Processing* **13**, 600-612 (2004).
47. W.-S. Lai, J.-B. Huang, N. Ahuja, and M.-H. Yang, "Deep laplacian pyramid networks for fast and accurate super-resolution," in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, (2017), 624-632.
48. Z. Wang, E. P. Simoncelli, and A. C. Bovik, "Multiscale structural similarity for image quality assessment," in *The Thirty-Seventh Asilomar Conference on Signals, Systems & Computers*, (2003), 1398-1402.
49. H. Zhao, O. Gallo, I. Frosio, and J. Kautz, "Loss functions for image restoration with neural networks," *IEEE Transactions on Computational Imaging* **3**, 47-57 (2016).
50. D. P. Kingma and J. Ba, "Adam: A method for stochastic optimization," *arXiv preprint arXiv:1412.6980* (2014).
51. C. Ledig, L. Theis, F. Huszár, J. Caballero, A. Cunningham, A. Acosta, A. Aitken, A. Tejani, J. Totz, and Z. Wang, "Photo-realistic single image super-resolution using a generative adversarial network," in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, (2017), 4681-4690.
52. C. Dong, C. C. Loy, and X. Tang, "Accelerating the super-resolution convolutional neural network," in *Proceedings of the European Conference on Computer Vision*, (2016), 391-407.