



Comparative study of deep learning models for optical coherence tomography angiography

ZHE JIANG,^{1,2,3} ZHIYU HUANG,^{1,2,3} BIN QIU,^{1,2,3} XIANGXI MENG,^{1,4}  YUNFEI YOU,^{1,2,3} XI LIU,¹ GANGJUN LIU,^{2,3} CHUANGQING ZHOU,³ KUN YANG,⁵ ANDREAS MAIER,⁶ QIUSHI REN,^{1,2,3} AND YANYE LU^{1,2,3,6,*} 

¹Department of Biomedical Engineering, College of Engineering, Peking University, No. 5 Yihe Yuan Road, Haidian District, Beijing 100871, China

²Institute of Biomedical Engineering, Peking University Shenzhen Graduate School, No. 2199 Lishui Road, Nanshan District, Shenzhen 518055, China

³Institute of Biomedical Engineering, Shenzhen Bay Laboratory 5F, No. 9 Duxue Road, Nanshan District, Shenzhen 518071, China

⁴Key Laboratory of Carcinogenesis and Translational Research (Ministry of Education), Department of Nuclear Medicine, Peking University Cancer Hospital & Institute, Beijing 100142, China

⁵College of Quality and Technical Supervision, Hebei University, No. 2666 Qiyidong Road, Baoding, 071000, China

⁶Pattern Recognition Lab, Department of Computer Science, Friedrich-Alexander-University Erlangen-Nuremberg, Martensstrasse 3, 91058 Erlangen, Germany

*yanye.lu@pku.edu.cn

Abstract: Optical coherence tomography angiography (OCTA) is a promising imaging modality for microvasculature studies. Meanwhile, deep learning has achieved rapid development in image-to-image translation tasks. Some studies have proposed applying deep learning models to OCTA reconstruction and have obtained preliminary results. However, current studies are mostly limited to a few specific deep neural networks. In this paper, we conducted a comparative study to investigate OCTA reconstruction using deep learning models. Four representative network architectures including single-path models, U-shaped models, generative adversarial network (GAN)-based models and multi-path models were investigated on a dataset of OCTA images acquired from rat brains. Three potential solutions were also investigated to study the feasibility of improving performance. The results showed that U-shaped models and multi-path models are two suitable architectures for OCTA reconstruction. Furthermore, merging phase information should be the potential improving direction in further research.

© 2020 Optical Society of America under the terms of the [OSA Open Access Publishing Agreement](#)

1. Introduction

Optical coherence tomography (OCT), a powerful imaging modality in biomedical studies and healthcare, allows for the non-invasive acquisition of depth-resolved biological tissue images by measuring the echoes of backscattered light [1,2]. Depending on its excellent ability to resolve the axial information, OCT is successfully applied to neurology, ophthalmology, dermatology, and cardiology for structural imaging [3–5]. Meanwhile, with the development of light source and detection techniques, OCT is also expanded for functional imaging, for instance, optical coherence tomography angiography (OCTA) [6–8].

Different from the traditional fluorescein angiography, OCTA, a new OCT-based imaging modality, could noninvasively obtain high-quality angiograms without using any contrast agents. The objective of OCTA is to extract the variation of OCT signals caused by the movement of red blood cells to provide microvasculature contrast against the static retinal tissue. Hence, various OCTA algorithms have been invented to calculate the differences of OCT signals obtained at the

same location within a short time sequence. Generally, the OCTA algorithms can be classified into three categories, i.e., intensity-signal-based algorithms, phase-signal-based algorithms, and complex-signal-based algorithms. phase-based algorithms, such as Doppler variance [9] and phase variance [6], merely use the phase part of the OCT signal to map the microvasculature. However, most of the phased-based algorithms are sensitive to phase noise, therefore elimination of motion artifacts is required. Unlike phase-based algorithms, intensity-based algorithms, such as speckle variance, correlation mapping [10], power intensity differential [11] and split-spectrum amplitude decorrelation angiography [12] use different statistics (variance, correlation, squared difference, and decorrelation) of OCT signal intensity to compute the blood flow information. Hence, the influence of the phase noise and motion artifacts can be relieved. However, the algorithms cannot apply to the scenarios where the flow induces the Doppler variation only in the phase part of the OCT signal. In contrast, the complex-signal-based algorithms [13,14] use both the intensity and phase information in reconstruction. Some representative algorithms, such as the split-spectrum amplitude and phase-gradient angiography (SSAPGA) [15] have significantly improved the OCTA imaging quality. In summary, the three traditional categories of OCTA algorithms calculate the flow intensity through different methods to measure OCT signal changes across temporally consecutive cross-sectional images (B-scans) taken at the same location. However, due to the limitations of analytical methods, the traditional OCTA algorithms can only utilize a fraction of information in the OCT signal variation.

In recent years, deep learning (DL) has achieved phenomenal success. As a representative category of DL methods, the convolutional neural network (CNN) [16–20] has greatly promoted the progress of various computer vision tasks and has become a popular option in visual and perception-based tasks. In the field of ophthalmology, various CNN models have been developed for disease classification, object segmentation and image enhancement. In order to utilize the DL's great capability to mine the underlying connection between data, some DL-based solutions have been proposed to be alternatives to traditional analytic OCTA algorithms. Lee et al. [21] proposed an attempt to use a U-shaped auto-encoder network to generate a retinal flow network from the clinical data, but the result was not quite satisfactory and the structural details of small vessels were difficult to be distinguished from OCT noise. Meanwhile, Liu et al. [22] also proposed a DL-based pipeline for OCTA reconstruction and obtained promising results in *in-vivo* studies. Although the results have demonstrated the superiority of the DL-based pipeline to the traditional OCTA algorithms, they only adopted one single-path network modified from the DnCNN, a network originally designed for image denoising [18], as the DL-model to verify their pipeline. In fact, with the rapid development of technology, various new CNN structures have been proposed for image-to-image translation tasks (e.g., image denoising, super-resolution, and image synthesis). Most of these models can be extended to OCTA reconstruction. Therefore, in order to fully utilize the current research achievements of DL and explore the future direction of DL-based OCTA methods, it is necessary to scrutinize the current network architectures for image translation tasks and investigate their suitability for OCTA reconstruction.

In our previous work [22], we demonstrated the DL-based pipeline for OCTA reconstruction. Building on this foundation, more comprehensive investigations have been conducted, and four representative network architectures were investigated, i.e., single-path model, U-shaped model, generative adversarial network (GAN)-based model and multi-path model. The effectiveness of each architecture was investigated using an *in vivo* animal dataset. The OCTA reconstruction results were quantitatively evaluated to reflect the performance of comparative studies. Three potential solutions (i.e., loss function optimization, data augmentation and merging phase information) were also investigated to study the feasibility of improving the performance.

2. Methods

2.1. Conventional OCTA reconstruction

A typical OCT signal at lateral location x^l , axial location z^l and time t in a B-scan can be written as:

$$S(x^l, z^l, t) = A(x^l, z^l, t) \cdot e^{i \cdot \Phi(x^l, z^l, t)}, \quad (1)$$

where A and Φ are the signal amplitude and phase component, respectively. As mentioned above, OCTA algorithms are mainly categorized into three categories. Among them, complex-signal-based OCTA algorithms are widely used in the field due to their comprehensive utilization of amplitude and phase information of the OCT signal. The SSAPGA algorithm, as a representative complex-signal-based OCTA algorithm, maps the microvasculature with a simple correction of phase artifacts via phase gradient method, demonstrating superior performance over other conventional OCTA algorithms. The SSAPGA method can be described as Eqs. (2) and (3).

$$\text{SSAPG}(x^l, z^l, t) = 1 - \frac{1}{R^l-1} \frac{1}{M} \left| \sum_{m=1}^M \sum_{r=1}^{R^l-1} \frac{2A_m(x^l, z^l, t+(r-1) \cdot \Delta t) \cdot A_m(x^l, z^l, t+r \cdot \Delta t) \cdot e^{j \cdot \rho \cdot \text{PG}_m(x^l, z^l, t+(r-1) \cdot \Delta t)}}{[A_m(x^l, z^l, t+(r-1) \cdot \Delta t)]^2 + [A_m(x^l, z^l, t+r \cdot \Delta t)]^2} \right| \quad (2)$$

$$\text{PG}(x^l, z^l, t + (r-1) \cdot \Delta t) = \frac{d(\Delta \Phi(x^l, z^l, t + (r-1) \cdot \Delta t))}{dz^l} \quad (3)$$

In the equations, R^l , M and ρ are the number of repetitions at the same B-scan location, the number of narrow split spectrum bands and the weight parameter that controls the contribution from phase gradient contrast, respectively.

2.2. Deep learning-based OCTA pipeline

In this work, the DL-based OCTA pipeline proposed in [22] was employed in the comparative study. As shown in Fig. 1, the pipeline deals with OCTA reconstruction as an end-to-end image translation task and consists of three phases, i.e., training data preparation, model learning and OCTA predicting.

In the training data preparation, a learning set of the instance-label pairs $\{(x^n, y^n)\}_{n=1}^N$ of ground truth image $y \in \mathbb{R}^{H \times W}$ and their corresponding multi-channel OCT structural image $x \in \mathbb{R}^{C \times H \times W}$ were constituted for DL-based OCTA reconstruction, where n is the index of the image pair; C , H and W are the numbers of input channels, rows and columns of the image. Each cross-sectional image pair (x, y) in the learning set corresponds to a slow-axis location. At each slow-axis location, 48 consecutive B-scans was firstly registered by a rigid registration algorithm [23] and further calculated using the SSAPGA algorithm to generate a label angiogram y with high signal-to-noise ratio (SNR); meanwhile, registered consecutive B-scans (OCT structure images) were randomly selected as the multi-channel OCT structural image x .

In the model learning phase, the learning set was first split into a training set $\{(x^p, y^p)\}_{p=1}^P$ and a validation set $\{(x^q, y^q)\}_{q=1}^Q$. Then, four representative DL models were selected to investigate the performance of each model, finding out the most promising DL architectures for OCTA reconstruction. The training of each models consisted of forward propagation, loss function calculation and back propagation steps. That is, the input x was sent to the respective neural network to output the predicted image $z \in \mathbb{R}^{H \times W}$; then the loss between the predicted image z and the ground truth image y was calculated; finally, the back-propagation procedure passed the loss value back to the network to compute the gradient, and updated the layer weights. Meanwhile, the validation set was used to monitor the model training process with the peak signal-to-noise

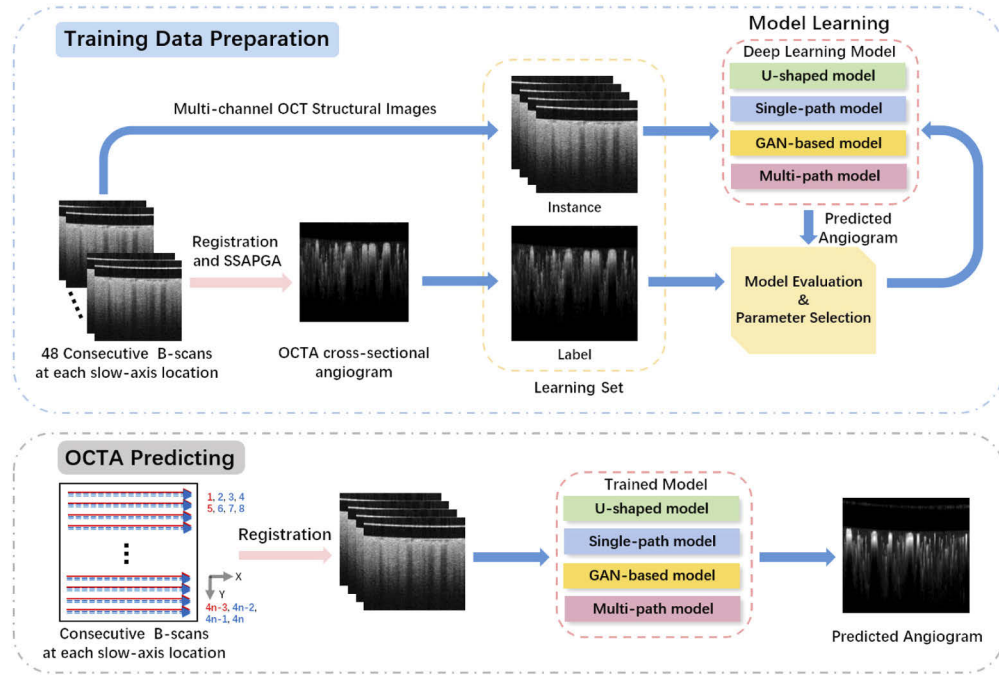


Fig. 1. A schematic diagram of deep learning-based optical coherence tomography angiography pipeline.

ratio (PSNR) as the quantitative metric. The PSNR can be defined as:

$$\text{PSNR} = 10 \cdot \lg \left(\frac{\text{MAX}_y^2}{\text{MSE}} \right), \quad (4)$$

where, MAX_y is the maximum value of ground truth image and the MSE is the mean square error between the ground truth y and predicted image z . The expression of MSE can be formulated as:

$$\text{MSE} = \frac{\sum_{H,W} (z - y)^2}{H \times W}. \quad (5)$$

As for the OCTA predicting, the consecutive B-scans at each slow-axis location are extracted and registered; then the angiogram can be reconstructed from the B-scans using the model trained in the learning stage of the process.

2.3. Image translation and typical network architectures

Image translation, also known as image-to-image translation, is an important field in computer vision. The goal is to establish a mapping from an image in the source domain to a corresponding image in the target domain through learning. The field includes various type of problems such as super-resolution [19,24–26], noise reduction [18,27] and image synthesis [28–32]. Encoder-decoder network-based models [18,19,24,25] and GAN-based models [28,29,31,33] are two mainstream types of implementation in image-to-image translation. For encoder-decoder network-based models, the goal is to train a single CNN model with a particular structure to convert one input image into the target image. According to the different architectures of the CNN, the encoder-decoder network-based models can be further divided into single-path models, U-shaped models, and multi-path models. On the other hand, Pix2Pix GAN [28] and

CycleGAN [29], which work on datasets consisting of paired images and unpaired images, respectively, represent two fundamental frameworks for GAN-based models. Considering that the cross-sectional image pairs in our study are spatially aligned, the Pix2Pix GAN is more suitable for OCTA reconstruction than CycleGAN.

In this work, we investigated four representative CNN models, which are DnCNN for single-path models, U-Net [17] for U-shaped models, residual dense network (RDN) [19] for multi-path models and Pix2Pix GAN for GAN-based models. Preliminary experiments based on the 1/10 of the full dataset were performed for parameter tuning. Important hyperparameters, such as learning rate and batch size, were investigated, respectively. Early stopping was employed in the training phase, ensuring that the DL models can obtain the optimal model parameters in the training phase.

2.3.1. Single-path model

As a representative DL architecture, single-path models offer a simple yet effective way to implement image translation. The network structures of single-patch models are line-shaped and without any skip connections, as in SRCNN [24] and VDSR [25] for super-resolution and DnCNN for noise reduction. In this work, we employed a modified DnCNN as the representative single-path model and investigated it in the OCTA pipeline. The network structure is shown in Fig. 2. This network included 20 convolutional layers. The first layer consisted of 64 filters of size $4 \times 3 \times 3$ to handle the four input OCT structural images and utilized the rectified linear units (ReLU) [34] as the activation function. Each layer in layers 2 ~ 19, which were used to extract the features gradually, included 64 filters of size $64 \times 3 \times 3$, batch normalization (BN) [35] and ReLU. With a single filter of size $64 \times 3 \times 3$, the last layer yielded the predicted angiogram reconstructed from the four structural images. The network parameters were denoted as Θ for this end-to-end system. For a given training set $\mathcal{P} = \{(x^p, y^p) | p = 1, 2, \dots, P\}$, the model was trained by minimizing the mean squared error (MSE) between the ground truth images y^p and reconstructed images $z = F(x^p; \Theta)$. The loss function is characterized by:

$$L(\Theta) = \frac{1}{P} \sum_{p=1}^P \|F(x^p; \Theta) - y^p\|_2^2. \quad (6)$$

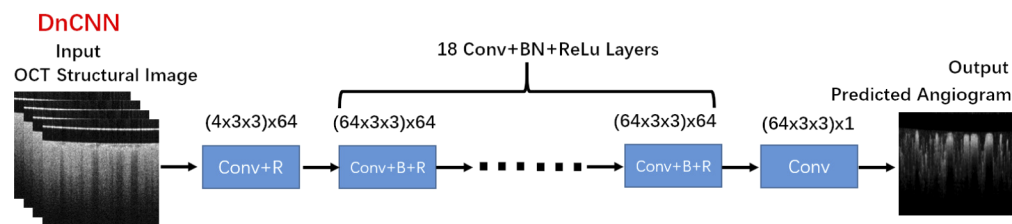


Fig. 2. The structure of the DnCNN for OCTA reconstruction. (Conv: convolutional layer; R: ReLU layer; B: BN layer.)

As for the training details, we initialized the weights by the method in [36] and used Adam optimization algorithm [37] to minimize the loss function. The learning rate, batch size, and epoch were set to 1×10^{-4} , 32 and 50, respectively.

2.3.2. U-shaped model

As one of the most important DL architectures, U-shaped models refer to a category of networks with symmetrical structures [17,38,39]. Among these models, U-Net, which was first proposed by Ronneberger et al. has achieved promising performance in image segmentation tasks. Since

then, various U-shaped models were proposed to further improve segmentation performance and extend the application scope. Depending on the high efficiency in training and the multi-level feature concatenation, U-shaped models are also widely used for image reconstruction in medical imaging, such as X-ray computed tomography (CT) and photoacoustic imaging (PAT). Hence, U-Net was selected as the representative of U-shaped models in the OCTA pipeline to study whether the U-Net can capture the variation among the structural images to reconstruct the angiograms.

The network structure is shown in Fig. 3, which was divided into two parts, i.e., contracting encoder and expanding decoder. In the encoder part, the down-sampling procedure was performed hierarchically through five stages. Each down-sampling stage contained two 3×3 convolutions, followed by BN and Leaky ReLU [36], and a 2×2 max-pooling with a stride of 2. In the first stage, 32 feature maps were extracted from the 4-channel input. Then, the number of feature maps was doubled in each stage. In the decoder part, the features extracted by the encoder part were up-sampled sequentially. Each up-sampling stage contained one 2×2 deconvolution with a stride of 2 and two 3×3 convolutions followed by BN and Leaky ReLU. After each deconvolution, the feature maps from the symmetric layer in the encoder path were stacked as extra channels through skip connections. In this process, the feature maps were halved in each stage of the decoder. In the last layer, a 3×3 convolution was used to reconstruct the angiogram.

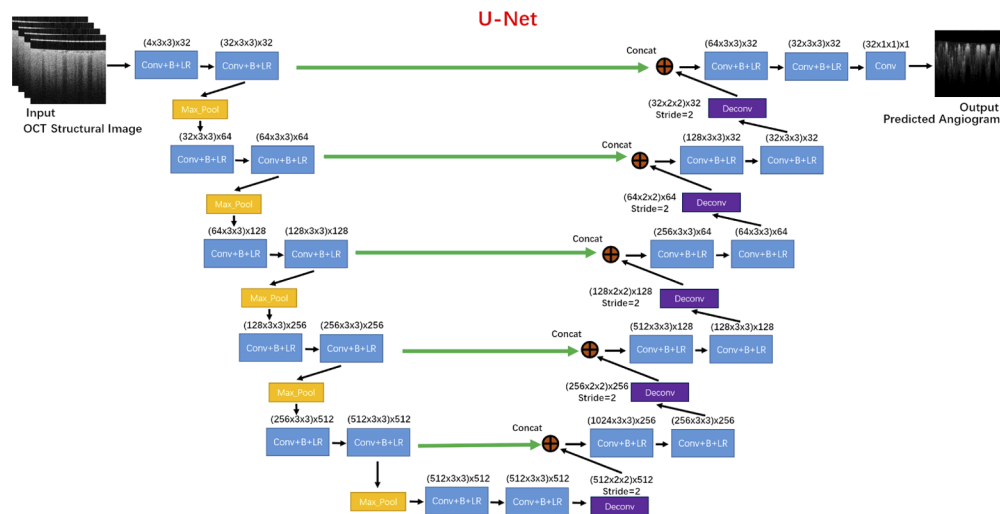


Fig. 3. The structure of the U-Net for OCTA reconstruction. (Conv: convolutional layer; LR: Leaky ReLU layer; B: BN layer; Max-Pool: max-pooling layer; Deconv: deconvolution layer.)

To train the network, we initialized the weights using the method in [36] and adopted the MSE loss function with Adam optimizer. The learning rate, batch size, and epoch were set to 1×10^{-4} , 32 and 50, respectively.

2.3.3. Multi-path model

As a kind of effective DL architectures, multi-path models refer to the very deep line-shaped asymmetric networks with skip connections. ResNet [16] and DenseNet [40] are two representative networks of multi-path models, which use residual connection and dense connection as the key to design the network structure, respectively. Figure 4 demonstrates the principles of the residual connection and dense connection, where v_l and $T_l(*)$ represents the output and non-linear transformation of l^{th} convolutional layer, respectively. It can be seen that the residual connection

bypasses the middle convolution layers as the identity mapping:

$$v_l = T(v_{l-1}) + v_{l-3}. \quad (7)$$

This identity mapping is helpful for the backpropagation of the gradient flow, therefore makes it possible to train a very deep network. On the other hand, the dense connection is the connections from any layer to all subsequent layers. That is, for each convolutional layer, the input is a concatenation of all the output of previous layers:

$$v_l = T([v_0, v_1, \dots, v_l]). \quad (8)$$

This concatenation of feature maps reduces feature redundancy and further improves efficiency and performance.

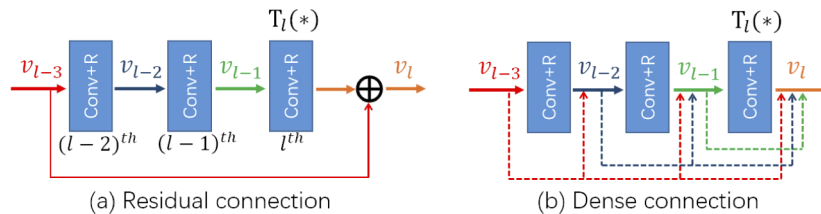


Fig. 4. The schematic diagram of Residual connection and Dense connection. (Conv: convolutional layer; R: ReLU layer.)

Owing to the advanced characteristics, the residual connection and dense connection are widely used in image translation tasks [41,42]. As the network combining through both residual connection and dense connection, the RDN has achieved significant performance enhancement in super-resolution tasks. Hence, a modified RDN was selected as the representative multi-path model in this study.

The network structure of the RDN is shown in Fig. 5, where the residual dense block (RDB) served as the basic module. There were four parts in the RDN structure. The first two layers, which belong to the shallow feature extraction net, were used to extract shallow features from input to the global residual path and RDBs. Then, the hierarchical features were obtained through 20 successive RDBs, and the feature maps from all the RDBs were concatenated to form the dense hierarchical features. Each RDB consisted of seven convolutional layers with block residual connection and dense connection to get high-quality local features. Afterward, two convolution layers were used to fuse the dense hierarchical features for global residual learning. Finally, the last convolution layer reconstructed the angiogram.

To train the network, we used the method in [36] to initialize the weights and adopted the MSE loss function with the Adam optimizer. The learning rate, batch size, and epoch were set to 1×10^{-4} , 32 and 50, respectively.

2.3.4. GAN-based model

As an important family of DL networks, GANs [43] have been vigorously studied in recent years for a wide variety of problems. Typically, in the GAN system, a generator network and a discriminator network are coupled and trained simultaneously. The generator is trained to learn a mapping from a random noise vector, and output a realistic ("fake") image which the discriminator cannot distinguish from "real" image. Meanwhile, the discriminator is trained to discriminate between "real" and "fake" images. Among various GAN structures [28,29,44], Pix2Pix GAN, which is based on the idea of conditional GAN (cGAN) [43], is an outstanding architecture for image translation tasks with datasets of paired images. Therefore, Pix2Pix GAN was chosen for our investigations.

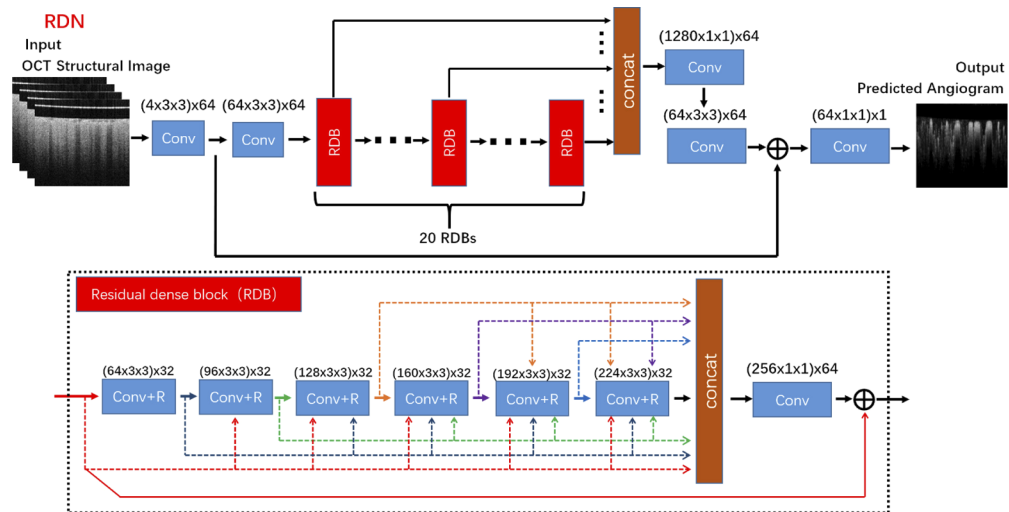


Fig. 5. The structure of the RDN for OCTA reconstruction. (Conv: convolutional layer; R: ReLU layer.)

The structure and training procedure of the Pix2Pix GAN are shown in Fig. 6. We employed the same U-Net model in Fig. 3 as the generator and a Patch GAN [28] as the discriminator. There were five 4×4 convolutional layers in the Patch GAN. Except for the last layer, all the convolution layers were with a stride of 2 and used Leaky ReLU as the activation functions. In the layers 2 ~ 4, instance normalization [28] was added between convolution and ReLU. The number of feature maps was doubled in each layer. In such a GAN system, the goal is to learn an effective mapping of the generator to predict the angiograms. Here, the training set takes the form of $\mathcal{P} = \{(x^p, y^p) | p = 1, 2, \dots, P\}$, and the mappings of generator and discriminator were defined as G and D , respectively. Then, the loss function could be expressed as the combination

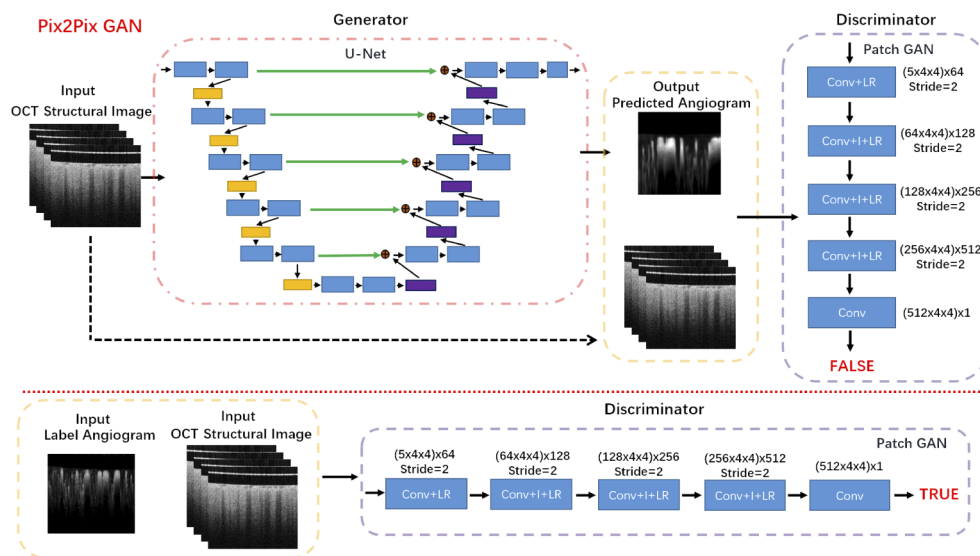


Fig. 6. The structure of the Pix2Pix GAN for OCTA reconstruction. (Conv: convolutional layer; LR: Leaky ReLU layer; I: instance normalization layer.)

of adversarial loss [43] and global L_2 loss:

$$L_{\text{P2P}}(G, D) = \mathbf{E}_{x,y}[\log D(x, y)] + \mathbf{E}_x [\log(1 - D(x, G(x)))] + \lambda \mathbf{E}_{x,y} [\|y - G(x)\|_2], \quad (9)$$

where \mathbf{E} is a statistical expectation and λ is the weight term to balance the adversarial loss and global L_2 loss. The adversarial loss promotes the network to achieve sharp results and the global L_2 loss ensures the correctness of low-frequency information in predicted image $z = G(x)$.

In this study, the generator and discriminator were trained through an alternating iterative scheme to optimize the overall objective of Eq. (9). Method [43] was used to initialize both the generator and the discriminator. Adam optimizers with learning rate 1×10^{-4} and 2×10^{-4} were used for generator and discriminator, respectively. Meanwhile, the λ was 1000 to adjust the weight of global L_2 loss. Moreover, the batch size was set to 1 to meet the architectural characteristic of the Pix2Pix GAN.

3. Experimental setup

3.1. Spectral-domain OCT system

A customized spectral-domain OCT system [22] was used to acquire the OCT B-scan images for OCTA. The system used a wideband super luminescent diode with a central wavelength of 845 nm and a full width at half maximum bandwidth of 30 nm as the light source, and adopted a fast line scan CCD attached to a high-speed spectrometer with a 28 kHz line scan rate as the detector. The axial resolution and lateral resolution of the system were 10 μm and 12 μm , respectively.

3.2. OCTA dataset and experimental protocol

In this study, an OCTA dataset of brain tissue from four Sprague Dawley rats was acquired. Firstly, a 4 mm \times 4 mm bone window was prepared through a craniotomy for each rat. Then, six data volumes were obtained from the four rats by the OCT system. For each volumetric scan, the field of view (FOV) and imaging depth were 2.5 mm \times 2.5 mm and 1 mm, respectively. Two of the rats were scanned twice using different FOVs, and the interval between the two scans was no less than one day. A total of 300 slow-axis locations were sampled as the data volume. In each slow-axis location, 48 consecutive B-scans with pixel size 1024 \times 300 were captured. Following the above-mentioned pipeline, 1800 cross-sectional image pairs were calculated from six volumes to form the OCTA dataset and each image pair includes four randomly selected structural OCT images and one label angiogram.

The training dataset and test dataset were carefully designed so as to avoid using data acquired on the same animal in both datasets. 1500 image pairs (from five volumes acquired from three rats) as the training set and the remaining 300 image pairs (from one volume of the remaining rat) as the test set. Based on the same dataset, comparative studies were carried out for the quantitative analysis and evaluation of the four representative network models. All the networks were implemented via Pytorch (<https://pytorch.org/>) on NVIDIA GPUs. DnCNN, U-Net, RDN, and Pix2Pix GAN were selected as the respective representative network for the studies. The detailed network structures and parameter settings are elaborated in Section 2.3. Meanwhile, SSAPGA as the reference method was also implemented for comparison. In addition, considering that 2-input and 3-input imaging protocols are alternatively used for OCTA in practical applications to increase the imaging speed, 2-channel and 3-channel models are further trained and discussed.

3.3. Evaluation metrics

For all the networks, PSNR, structural similarity (SSIM) [45] and the Pearson correlation coefficient (R) were used as quantitative evaluation metrics. All three metrics were calculated between the ground truth image y and the predicted image z . PSNR measures image distortion and noise level between two images. The larger the PSNR, the better the quality of predicted

image z is. The calculation of PSNR can be found in Eqs. (4) and (5). SSIM is a metric to evaluate the structural consistency between the target image and the reference image by fusing the information of brightness, contrast, and structure. The value of SSIM is between 0 and 1, and can be denoted as follow:

$$\text{SSIM}(y, z) = \frac{(2\mu_y\mu_z + C_1)(2\sigma_{zy} + C_2)}{(\mu_y^2 + \mu_z^2 + C_1)(\sigma_y^2 + \sigma_z^2 + C_2)}, \quad (10)$$

where μ_y and δ_y terms are the mean value and the standard deviation of ground truth y , respectively; μ_z and δ_z terms characterize the mean value and the standard deviation of predicted image z , respectively; δ_{zy} represents the cross-covariance between y and z . The C_1 and C_2 terms are constants used to avoid numerical instability of the numerator and denominator. R reflects the degree of linear correlation between two images, which is between 0 and 1. The Formula for R is:

$$R(y, z) = \frac{\sigma_{yz}}{\sigma_z\sigma_y}, \quad (11)$$

where, δ_{yz} is the covariance between y and z . For both SSIM and R , a larger value represents a better result.

4. Result and discussion

Representative reconstructed angiograms from four consecutive B-scans via SSAPGA and four DL models are shown in Fig. 7. It can be seen that the DL-based methods were able to predict enhanced blood flow signals with less speckle noise compared with the traditional 4-input SSAPGA algorithm. Although compared with the ground truth image, the predicted angiograms were prone to overemphasize local smoothing due to relatively limited input information (4 input vs 48 input), some DL methods, such as RDN, still successfully predicted the blood flow signal while preserving a certain degree of high-frequency details.

Table 1 summarizes the quantitative performance of the five OCTA algorithms with different input protocols on the cross-sectional angiograms. Each evaluation index was made by averaging the corresponding results in the test dataset. As shown in the table, all the DL-based models demonstrate better PSNR, SSIM, and R compared to the traditional SSAPGA, indicating that the deep neural networks have effectively utilized their modeling capability to mine more intrinsic connections from OCT signals for OCTA reconstruction. The overall performance of the four DL-based models can be ranked in the following order: RDN > U-Net > Pix2Pix GAN > DnCNN. One of the major explanations is the use of skip-connection, which was adopted in RDN, Pix2Pix GAN and U-Net, was absent in the DnCNN. As mentioned in section 2.3, skip-connections can improve the vanishing-gradient problem in the training procedure of very deep neural networks and is benefactive to the hierarchical feature fusion. This is important for the OCTA reconstruction because such a property is beneficial for DL-based models to learn multi-level representations under different receptive fields or different scales to capture tiny changes of the OCT signals in angiogram reconstruction. For U-Net and Pix2Pix GAN, the skip-connection is used in concatenating the low-level features and symmetrical high-level features, which belong to the encoding path and decoding path, respectively. For RDN, the residual connection and the dense connection, as two variants of the skip-connection, are used together to facilitates the sufficient application of hierarchical features along the feedforward path in the network. All these connection architecture designs in U-Net, Pix2Pix GAN and RDN help models extract more informative features from the training images with less mid/high-frequency information loss to yield the reconstructed OCTA images. However, compared with U-Net and Pix2Pix, the RDN effectively combined the residual connection and dense connection, which should further ensure the model increase the network depth without sacrifice the trainability.

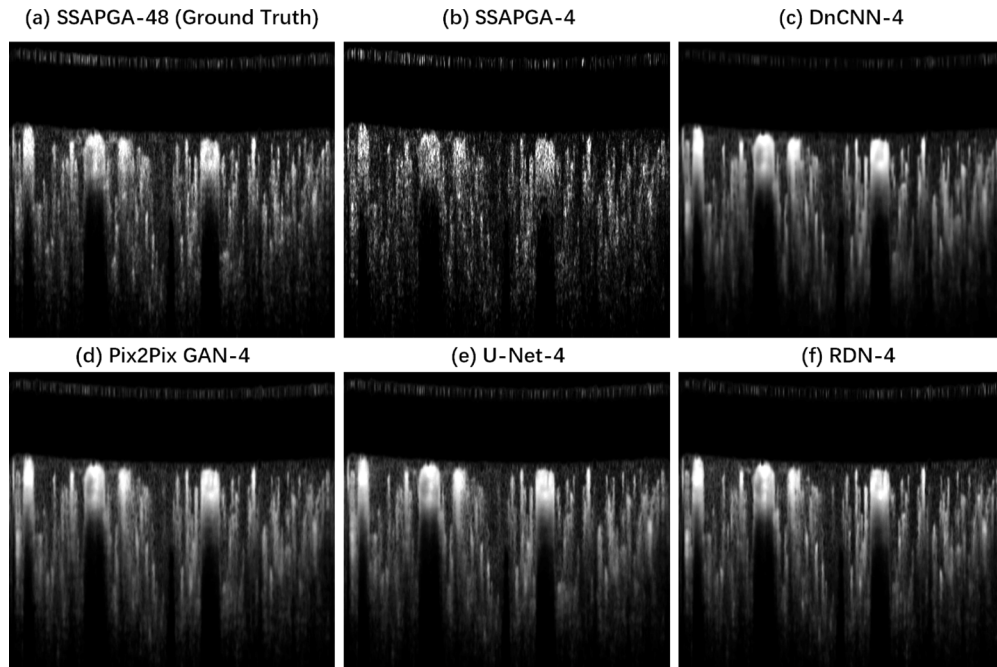


Fig. 7. Reconstructed cross-sectional angiograms via different algorithms; for visual comparison, all the results shown in figure is the same region-of-interest (ROI) (337×371) extracted from the original cross-sectional angiogram; (a) ground truth image obtained from SSAPGA algorithm with 48 consecutive B-scans as input; (b)-(f) reconstructed angiograms through corresponding OCTA algorithms with 4 consecutive B-scans as input.

Table 1. The average PSNR(dB)/SSIM/R results of the cross-sectional angiograms with different algorithms and protocols (The best results are highlighted).

	Scale	SSAPGA	DnCNN	U-Net	Pix2Pix	RDN
PSNR	4 inputs	39.20±0.96	43.76±0.82	44.39±0.91	44.20±0.91	44.40 ± 0.91
	3 inputs	38.26±0.75	42.91±0.76	43.57±0.95	43.59±0.91	43.73 ± 0.89
	2 inputs	35.89±0.81	41.80±0.81	42.16±0.98	42.44±0.89	42.53 ± 0.92
SSIM	4 inputs	0.971±0.006	0.982±0.003	0.986 ± 0.003	0.985±0.004	0.986 ± 0.003
	3 inputs	0.965±0.006	0.979±0.004	0.981±0.003	0.982±0.004	0.984 ± 0.003
	2 inputs	0.949±0.008	0.976±0.004	0.977±0.004	0.976±0.005	0.978 ± 0.004
R	4 inputs	0.931±0.013	0.974±0.007	0.975 ± 0.007	0.974±0.008	0.975 ± 0.006
	3 inputs	0.915±0.010	0.969±0.008	0.970±0.008	0.970±0.008	0.972 ± 0.007
	2 inputs	0.868±0.014	0.959±0.012	0.959±0.011	0.961±0.011	0.962 ± 0.011

Hence, depending on both the efficient feature fusion and the strong expression capability, RDN has achieved the best performance in this study.

A volumetric OCT data can be rendered in 2D slices along different dimensions, such as the cross-sectional visualization as mentioned before, or as *enface* representations. In angiogram applications, the scanned results are conventionally converted to *enface* maximum intensity projection (MIP) to facilitate vasculature visualization. Therefore, the performance of different methods was also compared in a MIP *enface* view in this study. Table 2 shows the calculated evaluation indices obtained by the five OCTA algorithms with different input protocols on the MIP *enface* angiogram.

Table 2. The PSNR(dB)/SSIM/R results of the MIP *enface* angiogram with different algorithms and protocols (The best results are highlighted).

	Scale	SSAPGA	DnCNN	U-Net	Pix2Pix	RDN
PSNR	4 inputs	14.62	17.31	20.00	20.43	22.14
	3 inputs	14.16	20.16	19.27	19.72	20.89
	2 inputs	13.27	17.44	20.82	19.68	20.53
SSIM	4 inputs	0.527	0.702	0.714	0.714	0.712
	3 inputs	0.500	0.677	0.690	0.685	0.686
	2 inputs	0.432	0.605	0.625	0.616	0.633
R	4 inputs	0.837	0.928	0.929	0.929	0.929
	3 inputs	0.815	0.918	0.922	0.919	0.920
	2 inputs	0.763	0.898	0.899	0.895	0.901

It can be seen that every DL-based algorithm outperforms SSAPGA. Across all the reconstruction results, U-Net and RDN have achieved the top two overall performance. On the other hand, although the Pix2Pix GAN outperformed the U-Net under the 4-input protocol, it failed to keep high effectiveness across all the protocols. This may due to the inherent characteristic of the GAN which makes it difficult to train. Adopting special stable training techniques for GAN should be one potential solution to solve the problem, however it is not the case in this study. In order to keep relative fairness of the comparison, we adopted the uniform training strategy for all the DL-based models and found that it is hard to keep the balance between the generator and the discriminator in Pix2Pix GAN to achieve superior performance. Consequently, under the normal training strategy, the performance of Pix2Pix GAN has not met expectations.

Figure 8 shows the 4-input reconstructed results of the MIP *enface* angiogram using different algorithms. It can be seen that DL-based models have demonstrated obvious advantages over 4-input SSAPGA in noise suppression and microvessel reconstruction. Meanwhile, the local smoothing problem existing in Fig. 8(c) of DnCNN has been somewhat alleviated in the reconstructed results of the Pix2Pix GAN, U-Net, and RDN. When we further compared Figs. 8(d) ~ 8(f), it can be found that RDN-based angiogram contains more high-frequency details than the angiograms obtained by Pix2Pix GAN and U-Net. However, it also suffered from some wavy artifacts, which are unobvious in the Figs. 8(d) and 8(e). The wavy artifacts were essentially caused by the strong sample motion during the acquisition process of 48 consecutive B-scans at each slow-axis location. Hence, the trouble influenced the performance of all the DL-based models. The reason why the RDN was more sensitive to the sample motion than the others should be attributed to the extremely high expression capability of the network, which needs more balanced training data to reduce the wavy artifacts. On the other hand, U-Net, depending on its moderate model size and high efficiency in computation and training, has an intrinsic advantage to learn a relatively generalized model from limited training data. As a result, U-Net

and U-Net-based Pix2Pix GAN suffered less from the wavy artifacts, which is also reflected in the higher SSIM values in the Table 2.

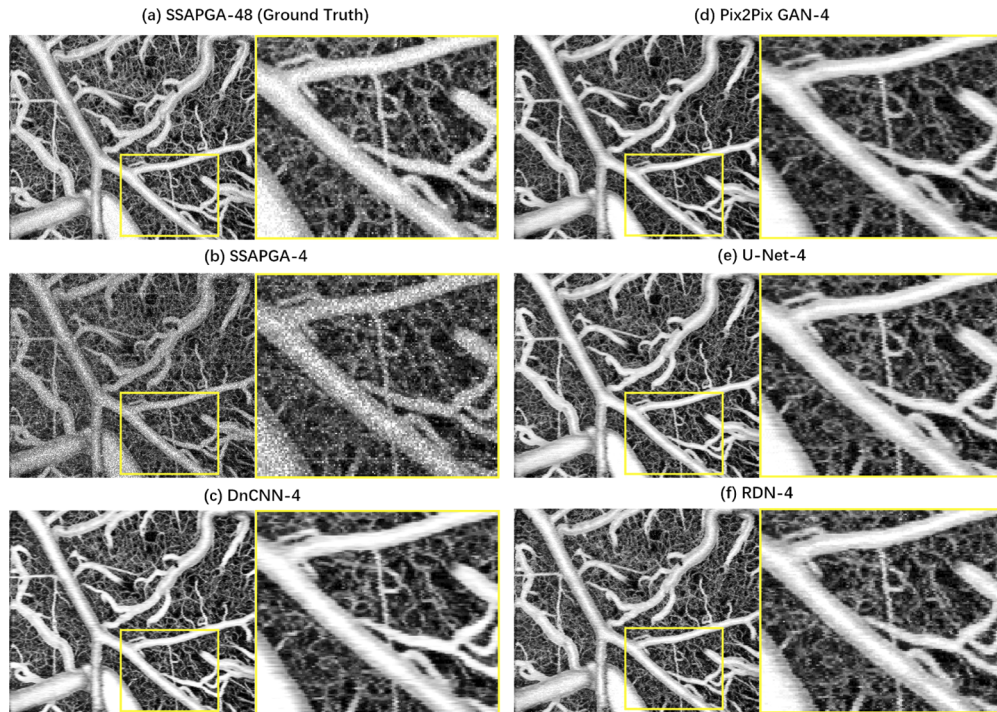


Fig. 8. Reconstructed MIP *enface* angiograms through different algorithms; for visual comparison, an ROIs are labeled by the rectangular box, and the 2.5-fold magnified images of the ROIs are also provided in the right of the original image; (a) ground truth image obtained via SSAPGA algorithm with 48 consecutive B-scans as input; (b)-(f) reconstructed angiograms via corresponding OCTA algorithms with 4 consecutive B-scans as input.

Based on the discussion above, we believe that U-shaped models and multi-path models should be two suitable categories of architectures for OCTA reconstruction tasks. U-shaped model is more suitable for the tasks with small data size and limited computational power. On the other hand, multi-path models should play a full part under sufficient training data and computational power.

Finally, an additional experiment was conducted to investigate the robustness of the DL-based OCTA reconstruction to noise. We assumed the noise is additive white Gaussian noise (AWGN), which is the most common noise existing in OCT images. Noisy B-scans, which were generated by adding AWGN with a certain noise level to the test set, were tested by the trained 4-input RDN model for demonstration. In the experiment, four noise levels, i.e., $\sigma \in [4, 6, 8, 10]$, were respectively used to generate the noisy inputs for OCTA reconstruction. The MIP *enface* angiograms of the results are shown in the Fig. 9. It can be seen that, the trained DL-model demonstrated certain robustness to the AWGN. With the noise level $\sigma = 4$ and $\sigma = 6$, the reconstructed MIP *enface* angiograms successfully preserved the primary vascular composition without generating obvious false blood flow signal. Although there was some degradation of the reconstruction quality, the results are acceptable with such high noise levels of $\sigma = 8$ and $\sigma = 10$. In fact, σ of $[0, 2]$ is hampered to OCT [46], and severely limit OCTA performance due its intrinsic high-frequency sensitivity. Table 3 summarized the corresponding quantitative

evaluation. The quantitative results verify the visual perception of Fig. 9 and indicate that the well-trained DL-based OCTA models are robustness to AWGN.

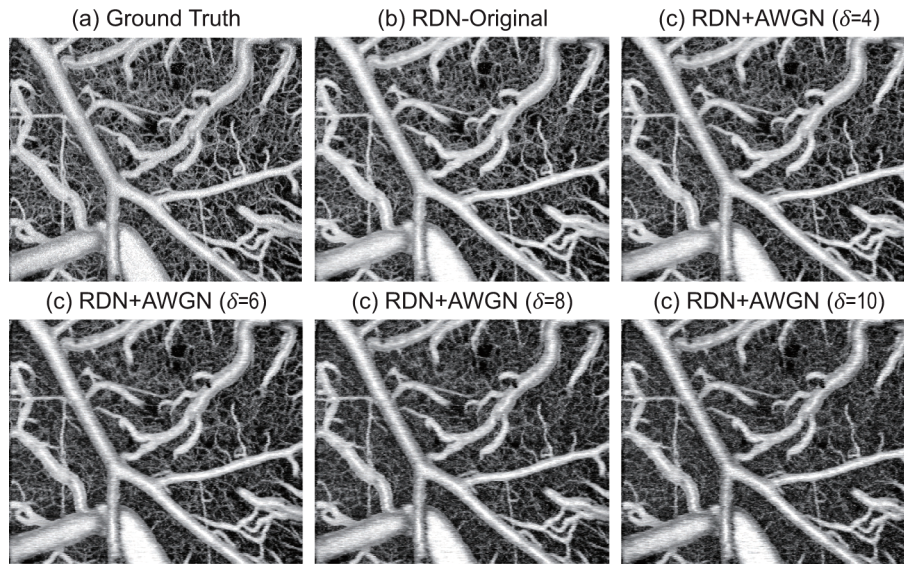


Fig. 9. Reconstructed MIP *enface* angiograms under the AWGN with different noise levels; (a) ground truth image obtained via SSAPGA algorithm with 48 consecutive B-scans as input; (b) original reconstructed angiogram of RDN without AWGN; (c)-(f) reconstructed angiograms of RDN with different AWGN.

Table 3. The PSNR(dB)/SSIM/R results of the MIP *enface* angiograms reconstructed by the 4-input RDN under different noise levels.

	RDN	RDN + AWGN			
	Without Noise	$\sigma = 4$	$\sigma = 6$	$\sigma = 8$	$\sigma = 10$
PSNR	22.1354	21.6584	21.5274	20.9981	20.4828
SSIM	0.7118	0.6910	0.6514	0.5924	0.5368
R	0.9287	0.9231	0.9104	0.8908	0.8685

5. Improvement exploration

Some potential solutions were pilot explored for further improving the DL-based OCTA reconstruction. Considering the training efficiency of the network, U-Net was chosen as the basic model for the investigation. Three schemes were investigated via the basic model, i.e., L_1 loss replacement, data-augmentation, and phase information merging.

For the first scheme, the L_1 loss was used to replace the original MSE loss in the U-Net. Referred to Eq. (6), the L_1 loss can be defined as:

$$L(\Theta) = \frac{1}{P} \sum_{p=1}^P |F(x^p; \Theta) - y^p|. \quad (12)$$

As for the second scheme, random data-augmentation was implemented while inputting training data into the network. That is, for each cross-sectional image pair, during the reading process, 180° clockwise rotate and horizontal flip were randomly conducted on the original image pair.

As for the third scheme, the cross-sectional image pairs were reorganized by adding three OCT phase images into the image pairs. The OCT phase images, which correspond to the four OCT structural images [15], were calculated through the phase gradient angiography method.

All three pilot studies adopted the same training strategy and parameter setting with the basic U-Net model in Section 2.3. Only the 4-input image protocol was tested in these experiments. The results of the MIP *enface* angiograms are shown in the Table 4.

Table 4. The PSNR(dB)/SSIM/R results of the MIP *enface* angiogram with different improvement schemes (The best results are highlighted).

	U-Net	U-Net + L_1	U-Net + Augmentation	U-Net + Phase
PSNR	20.00	20.27	20.34	21.30
SSIM	0.714	0.717	0.722	0.750
R	0.929	0.930	0.931	0.938

A plethora of information can be deduced from Table 4. First, the L_1 loss replacement has improved the reconstruction quality of OCTA. It may be attributed to the advantage of L_1 loss in preserving the low-frequency information. However, the improvement is relatively limited. Hence, designing a more suitable loss function for OCTA tasks is a direction that remains

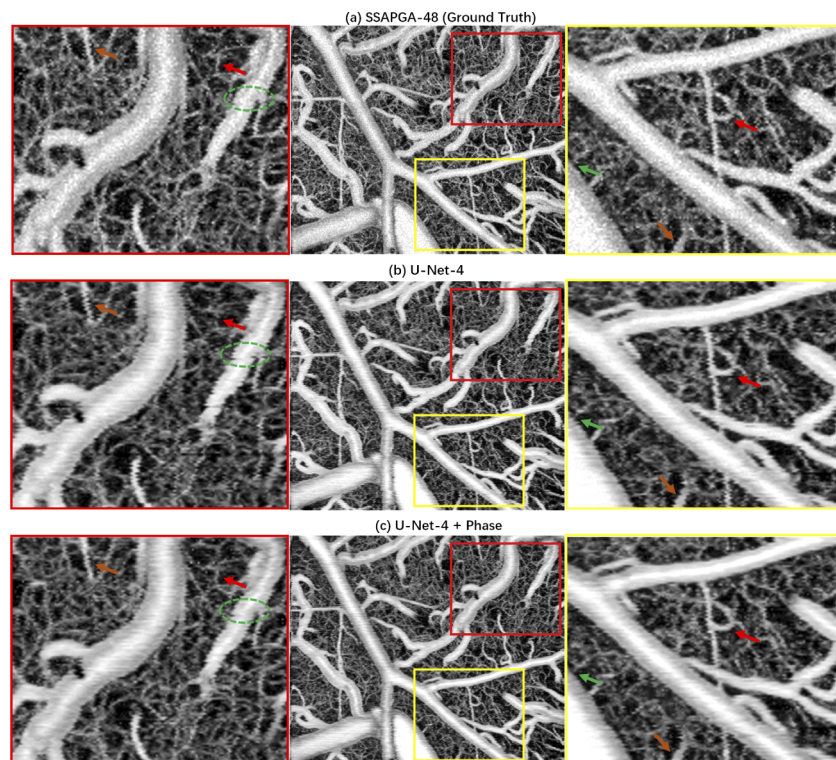


Fig. 10. Reconstructed MIP *enface* angiograms via different algorithms; for visual comparison, two ROIs are labeled by rectangular boxes, and the corresponding 2.5-fold magnified images of the ROIs are also provided in both sides of the figure; (a) ground truth image obtained from SSAPGA algorithm with 48 consecutive B-scans as input; (b) reconstructed angiograms through basic U-Net model; (c) reconstructed angiograms through U-Net model with phase information fusion.

to be explored in further studies. Secondly, the effectiveness of data-augmentation has been demonstrated. One possible explanation for this is that data-augmentation help mitigate the overfitting problem of the U-Net to a certain extent, resulting in better generalization for the trained model. Therefore, it is meaningful to build a larger dataset with proper data-augmentation techniques for OCTA tasks. Finally, similar to the traditional analytic OCTA algorithms, merging phase information also exerts a positive impact on the performance of the DL models. Figure 10 shows the reconstructed results of the MIP *enface* angiogram through phase information merging. Compared to the basic U-Net model, the microvessel pointed by red and brown arrows present higher reconstructive precision. Meanwhile, the vessel areas marked by the green arrow and dotted circle demonstrate better connectivity of edge and less wavy artifacts. The corresponding quantitative metrics of the two ROIs are also provided in the Table 5. It can be seen that all the metrics of the phase merging scheme have obtained obvious improvement relative to the corresponding metrics of the basic U-Net model. Hence, given the great modeling capability of the DL models, merging phase information into the training process to comprehensively use OCT signals for accurate OCTA reconstruction is a new impetus for DL-based OCTA tasks.

Table 5. The PSNR(dB)/SSIM/R results of the ROIs from MIP *enface* angiogram with merging phase information (The increase relative to basic U-Net model is also provided for each metrics).

	U-Net			U-Net + Phase		
	PSNR	SSIM	R	PSNR	SSIM	R
ROI-yellow	20.06	0.750	0.938	21.14 (+5.37%)	0.784 (+4.64%)	0.947 (+0.93%)
ROI-red	20.41	0.719	0.911	21.57 (+5.64%)	0.768 (+6.70%)	0.927 (+1.76%)

6. Conclusion

In this study, we conducted a comparative study on the DL-based OCTA reconstruction algorithms. Four representative models from the field of image translation were investigated using the DL-based OCTA pipeline. From the results, we found that U-shaped models and multi-path models were two suitable architectures for OCTA reconstruction. As an extension of the study, preliminary explorations of three potential solutions for further improvement of the DL-based OCTA tasks were conducted. The results showed that merging phase information should be the potential improving direction in the further research of DL-based OCTA reconstruction algorithms.

Funding

National Natural Science Foundation of China (61875123, 81421004); National Key Instrumentation Development Project of China (2013YQ030651); Natural Science Foundation of Hebei Province (H2019201378).

Acknowledgments

We acknowledge the support of the NVIDIA Corporation with the donation of the Titan Xp GPU used for this research.

Disclosures

The authors declare no conflicts of interest.

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**(5035), 1178–1181 (1991).

2. W. Drexler, U. Morgner, F. Kärtner, C. Pitris, S. Boppart, X. Li, E. Ippen, and J. Fujimoto, "In vivo ultrahigh-resolution optical coherence tomography," *Opt. Lett.* **24**(17), 1221–1223 (1999).
3. H. G. Bezerra, M. A. Costa, G. Guagliumi, A. M. Rollins, and D. I. Simon, "Intracoronary optical coherence tomography: a comprehensive review: clinical and research applications," *JACC: Cardiovasc. Interv.* **2**, 1035–1046 (2009).
4. H.-P. Hammes, Y. Feng, F. Pfister, and M. Brownlee, "Diabetic retinopathy: targeting vasoregression," *Diabetes* **60**(1), 9–16 (2011).
5. E. C. Sattler, R. Käßler, and J. Welzel, "Optical coherence tomography in dermatology," *J. Biomed. Opt.* **18**(6), 061224 (2013).
6. D. Y. Kim, J. Fingler, J. S. Werner, D. M. Schwartz, S. E. Fraser, and R. J. Zawadzki, "In vivo volumetric imaging of human retinal circulation with phase-variance optical coherence tomography," *Biomed. Opt. Express* **2**(6), 1504–1513 (2011).
7. J. Fingler, R. J. Zawadzki, J. S. Werner, D. Schwartz, and S. E. Fraser, "Volumetric microvascular imaging of human retina using optical coherence tomography with a novel motion contrast technique," *Opt. Express* **17**(24), 22190–22200 (2009).
8. A. Maier, S. Steidl, V. Christlein, and J. Hornegger, *Medical Imaging Systems: An Introductory Guide*, vol. 11111 (Springer, 2018).
9. L. F. Yu and Z. P. Chen, "Doppler variance imaging for three-dimensional retina and choroid angiography," *J. Biomed. Opt.* **15**(1), 016029 (2010).
10. J. Enfield, E. Jonathan, and M. Leahy, "In vivo imaging of the microcirculation of the volar forearm using correlation mapping optical coherence tomography (cmOCT)," *Biomed. Opt. Express* **2**(5), 1184–1193 (2011).
11. C. Blatter, J. Weingast, A. Alex, B. Grajciar, W. Wieser, W. Drexler, R. Huber, and R. A. Leitgeb, "In situ structural and microangiographic assessment of human skin lesions with high-speed OCT," *Biomed. Opt. Express* **3**(10), 2636–2646 (2012).
12. Y. L. Jia, O. Tan, J. Tokayer, B. Potsaid, Y. M. Wang, J. J. Liu, M. F. Kraus, H. Subhash, J. G. Fujimoto, J. Hornegger, and D. Huang, "Split-spectrum amplitude-decorrelation angiography with optical coherence tomography," *Opt. Express* **20**(4), 4710–4725 (2012).
13. L. An, J. Qin, and R. K. Wang, "Ultrahigh sensitive optical microangiography for in vivo imaging of microcirculations within human skin tissue beds," *Opt. Express* **18**(8), 8220–8228 (2010).
14. V. J. Srinivasan, J. Y. Jiang, M. A. Yaseen, H. Radhakrishnan, W. C. Wu, S. Barry, A. E. Cable, and D. A. Boas, "Rapid volumetric angiography of cortical microvasculature with optical coherence tomography," *Opt. Lett.* **35**(1), 43–45 (2010).
15. G. Liu, Y. Jia, A. D. Pechauer, R. Chandwani, and D. Huang, "Split-spectrum phase-gradient optical coherence tomography angiography," *Biomed. Opt. Express* **7**(8), 2943–2954 (2016).
16. 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* (Nevada, USA, 2016), pp. 770–778.
17. O. Ronneberger, P. Fischer, and T. Brox, "U-net: Convolutional networks for biomedical image segmentation," in *IEEE Conference on Computer Vision and Pattern Recognition* (Springer, 2015), pp. 234–241.
18. K. Zhang, W. Zuo, Y. Chen, D. Meng, and L. Zhang, "Beyond a Gaussian Denoiser: Residual Learning of Deep CNN for Image Denoising," *IEEE Trans. on Image Process.* **26**(7), 3142–3155 (2017).
19. Y. L. Zhang, Y. P. Tian, Y. Kong, B. N. Zhong, and Y. Fu IEEE, "Residual Dense Network for Image Super-Resolution," in *2018 IEEE/CVF Conference on Computer Vision and Pattern Recognition* (IEEE, 2018), pp. 2472–2481.
20. A. Maier, C. Syben, T. Lasser, and C. Riess, "A gentle introduction to deep learning in medical image processing," *Z. Med. Phys.* **29**(2), 86–101 (2019).
21. C. S. Lee, A. J. Tyring, Y. Wu, S. Xiao, A. S. Rokem, N. P. DeRuyter, Q. Q. Zhang, A. Tufail, R. K. Wang, and A. Y. Lee, "Generating retinal flow maps from structural optical coherence tomography with artificial intelligence," *Sci. Rep.* **9**(1), 5694 (2019).
22. X. Liu, Z. Y. Huang, Z. Z. Wang, C. Y. Wen, Z. Jiang, Z. K. Yu, J. F. Liu, G. J. Liu, X. L. Huang, A. Maier, Q. S. Ren, and Y. Y. Lu, "A deep learning based pipeline for optical coherence tomography angiography," *J. Biophotonics* **12**(10), 10 (2019).
23. D. G. Lowe, "Object recognition from local scale-invariant features," in *Proc. Seventh IEEE Int. Conf. on Comput. Vis.*, vol.2, (1999), pp. 1150–1157.
24. C. Dong, C. C. Loy, K. M. He, and X. O. Tang, "Image Super-Resolution Using Deep Convolutional Networks," *IEEE Trans. Pattern Anal. Mach. Intell.* **38**(2), 295–307 (2016).
25. J. Kim, J. K. Lee, and K. M. Lee IEEE, "Accurate Image Super-Resolution Using Very Deep Convolutional Networks," in *2016 IEEE Conference on Computer Vision and Pattern Recognition* (IEEE, 2016), pp. 1646–1654.
26. Z. Jiang, Z. K. Yu, S. X. Feng, Z. Y. Huang, Y. H. Peng, J. X. Guo, Q. S. Ren, and Y. Y. Lu, "A super-resolution method-based pipeline for fundus fluorescein angiography imaging," *Biomed. Eng. Online* **17**(1), 125 (2018).
27. B. Qiu, Z. Huang, X. Liu, X. Meng, Y. You, G. Liu, K. Yang, A. Maier, Q. Ren, and Y. Lu, "Noise reduction in optical coherence tomography images using a deep neural network with perceptually-sensitive loss function," *Biomed. Opt. Express* **11**(2), 817–830 (2020).
28. P. Isola, J. Y. Zhu, T. H. Zhou, and A. A. Efros IEEE, "Image-to-Image Translation with Conditional Adversarial Networks," in *30th IEEE Conference on Computer Vision and Pattern Recognition*, (IEEE, 2017), pp. 5967–5976.

29. J. Y. Zhu, T. Park, P. Isola, and A. A. Efros IEEE, "Unpaired Image-to-Image Translation using Cycle-Consistent Adversarial Networks," in *2017 IEEE International Conference on Computer Vision*, (IEEE, 2017), pp. 2242–2251.
30. A. Ben-Cohen, E. Klang, S. P. Raskin, S. Soffer, S. Ben-Haim, E. Konen, M. M. Amitai, and H. Greenspan, "Cross-modality synthesis from CT to PET using FCN and GAN networks for improved automated lesion detection," *Eng. Appl. Artif. Intell.* **78**, 186–194 (2019).
31. Z. K. Yu, Q. Xiang, J. H. Meng, C. X. Kou, Q. S. Ren, and Y. Y. Lu, "Retinal image synthesis from multiple-landmarks input with generative adversarial networks," *Biomed. Eng. Online* **18**(1), 62 (2019).
32. Y. Y. Lu, M. Kowarschik, X. L. Huang, Y. Xia, J. H. Choi, S. Q. Chen, S. Y. Hu, Q. S. Ren, R. Fahrig, J. Hornegger, and A. Maier, "A learning-based material decomposition pipeline for multi-energy x-ray imaging," *Med. Phys.* **46**(2), 689–703 (2019).
33. X. Yi and P. Babyn, "Sharpness-Aware Low-Dose CT Denoising Using Conditional Generative Adversarial Network," *J. Digit. Imaging* **31**(5), 655–669 (2018).
34. A. Krizhevsky, I. Sutskever, and G. E. Hinton, "Imagenet classification with deep convolutional neural networks," in *Advances in Neural Information Processing Systems Nevada, USA, 2012*, pp. 1097–1105.
35. S. Ioffe and C. Szegedy, "Batch normalization: Accelerating deep network training by reducing internal covariate shift," arXiv preprint arXiv:1502.03167 (2015).
36. K. M. He, X. Y. Zhang, S. Q. Ren, and J. Sun IEEE, "Delving Deep into Rectifiers: Surpassing Human-Level Performance on ImageNet Classification," in *2015 IEEE International Conference on Computer Vision* (IEEE, 2015), pp. 1026–1034.
37. D. P. Kingma and J. Ba, "Adam: A method for stochastic optimization," arXiv preprint arXiv:1412.6980 (2014).
38. X. J. Mao, C. H. Shen, and Y. B. Yang, "Image Restoration Using Very Deep Convolutional Encoder-Decoder Networks with Symmetric Skip Connections," in *Advances in Neural Information Processing Systems, vol. 29* D. D. Lee, M. Sugiyama, U. V. Luxburg, I. Guyon, and R. Garnett, eds. (Neural Information Processing Systems, 2016).
39. V. Badrinarayanan, A. Kendall, and R. Cipolla, "SegNet: A Deep Convolutional Encoder-Decoder Architecture for Image Segmentation," *IEEE Trans. Pattern Anal. Mach. Intell.* **39**(12), 2481–2495 (2017).
40. G. Huang, Z. Liu, L. van der Maaten, and K. Q. Weinberger Ieee, "Densely Connected Convolutional Networks," in *30th IEEE Conference on Computer Vision and Pattern Recognition* (IEEE, 2017), pp. 2261–2269.
41. M. Haris, G. Shakhnarovich, and N. Ukita IEEE, "Deep Back-Projection Networks For Super-Resolution," in *2018 IEEE/Cvf Conference on Computer Vision and Pattern Recognition* (IEEE, 2018), pp. 1664–1673.
42. Y. Tai, J. Yang, X. M. Liu, and C. Y. Xu IEEE, "MemNet: A Persistent Memory Network for Image Restoration," in *2017 IEEE International Conference on Computer Vision* (IEEE, 2017), pp. 4549–4557.
43. I. J. Goodfellow, J. Pouget-Abadie, M. Mirza, B. Xu, D. Warde-Farley, S. Ozair, A. Courville, and Y. Bengio, "Generative Adversarial Nets," in *Advances in Neural Information Processing Systems, Vol. 27* Z. Ghahramani, M. Welling, C. Cortes, N. D. Lawrence, and K. Q. Weinberger, eds. (Neural Information Processing Systems, 2014).
44. A. Radford, L. Metz, and S. Chintala, "Unsupervised representation learning with deep convolutional generative adversarial networks," arXiv preprint arXiv:1511.06434 (2015).
45. Z. Wang, A. C. Bovik, H. R. Sheikh, and E. P. Simoncelli, "Image quality assessment: from error visibility to structural similarity," *IEEE Trans. on Image Process.* **13**(4), 600–612 (2004).
46. S. K. Devalla, G. Subramanian, T. H. Pham, X. Wang, S. Perera, T. A. Tun, T. Aung, L. Schmetterer, A. H. Thiéry, and M. J. Girard, "A deep learning approach to denoise optical coherence tomography images of the optic nerve head," *Sci. Rep.* **9**(1), 14454–13 (2019).