A Fast Optical Coherence Tomography Angiography Image Acquisition and Reconstruction Pipeline for Skin Application

JINPENG LIAO,[1] SHUFAN YANG,[2, 3] TIANYU ZHANG,[1] CHUNHUI LI,[1, *] ZHIHONG HUANG[1]

[1]*School of Science and Engineering, University of Dundee, DD1 4HN, Scotland (United Kingdom)*
[2]*Engineering and Built Environment, Edinburgh Napier University, Edinburgh (United Kingdom)*
[3]*Research Department of Orthopaedics and Musculoskeletal Science, University College London (United Kingdom)*
**c.li@dundee.ac.uk*

Abstract: Traditional high-quality OCTA images require multi-repeated scans (e.g., 4-8 repeats) in the same position, which causes patient uncomfortable. We propose a deep-learning-based pipeline that can extract high-quality OCTA images from only two-repeat OCT scans. The performance of the proposed Image Reconstruction U-Net (IRU-Net) outperforms state-of-the-art UNet vision transformer and UNet in OCTA image reconstruction from a two-repeat OCT signal. The results demonstrated a mean peak-signal-to-noise ratio increased from 15.7 to 24.2; the mean structural similarity index measure improved from 0.28 to 0.59; while OCT data acquisition time was reduced from 21 seconds to 3.5 seconds (reduced by 83%).

1. Introduction

Optical coherence tomography (OCT) is a non-invasive, label-free, real-time, in vivo imaging technique [1]. With broadband infrared lasers (e.g., 1300±100 nm), the OCT can provide theoretically 2-10 μm axial resolution tomographic structural images with depth information up to 2-3 mm in biological tissue [2]. In the past two decades, OCT has developed well in ophthalmology [3], dermatology [4] and intravascular imaging [5]. In the application of skin disease diagnosis and monitoring, OCT structural imaging has been used in non-melanoma skin cancer [6], basal cell carcinomas [7] and actinic keratosis [8]. Besides traditional structural images, OCT can be extended with different functions, among which the OCT-Angiography (OCTA) imaging attracted the most attention because skin vasculature is altered in diseased skin [9]. OCTA can pick out moving red blood cells from the relatively static tissue from the temporal change of the sequence scan and suppress the static signal. Hence, OCTA imaging assists in identifying diseased or healthy skin areas by assessing vasculature rather than relying on surface appearance [10].

The conventional OCT data processing pipelines can be used to extract OCTA images by suppressing the static signals from tissues. Those algorithms were based on the phase signal [11], intensity signal [12–14] and complex signal [15]. However, the noise ratio of the processed OCTA image of those algorithms was highly dependent on the number of repeated OCT scans in the same positions (i.e., the number of repeat (NR) scans can improve the image signal-to-noise ratio (SNR) by a factor $\sqrt{NR}$) [16]. Hence, more repeat B-scans and more scanning times (e.g., 8-12 repeat scans in 14-21 seconds for a 200k swept rate OCT system) were required to obtain a high-quality skin OCTA image. However, in non-invasive multi-repeat OCTA scans (e.g., six repeat scans), the unpredictable movement of the patient and complex distribution of reflection angles can lead to an SNR gap in skin OCTA images [17]. It is essential to keep repeated scan requirements to a minimum to reduce those negative influences. Hence, there is a trade-off between OCT scanning time and OCTA image quality in skin applications in the conventional processing flow.

Artificial neural networks (ANN) are widely used for OCT image processing, such as lesion segmentation [18], detection [19], classification [20], and image denoising [21]. OCTA image reconstruction research communities are starting to use ANN as a way of reducing SNR in reconstructed OCTA images. For instance, the Denoising Convolution Neural Network (DnCNN) has been proven to be able to reconstruct the OCTA image using a two-repeat OCT scan method. [22]. However, their method has not been able to reduce the artifacts caused by intensity non-uniformity between angiogram slides. Another work using the U-Net for OCT reconstruction on retinal blood flow maps showed very clear vascular structure images [23]. Furthermore, a residual-based network was proposed to reconstruct high-quality retinal OCTA images [24]. Tavakkoli et al. [25] also demonstrated that generative adversarial network can produce high-quality retinal OCTA images. However, their proposed works were focusing on the retinal OCTA images in the field of ophthalmology; hence, rather than solely concentrating on image reconstruction performance, the CNN model has to relearn the different signatures present in skin OCTA images for dermatology applications. While optimising neural network structure can improve the quality of image reconstruction, loss function optimisation has been ignored in the OCTA image reconstruction publications. The loss functions commonly used in the published OCTA image reconstruction method were mean square error and mean absolute error. Those two loss functions are not generalisable to find details in high-frequency textures that are widely presented in OCTA images.

To better reconstruct the skin OCTA image while reducing the repeat numbers of the OCT scan, we propose a novel deep-learning-based pipeline for OCTA image reconstruction. In this study, we developed an encoder-decoder architecture network, called image reconstruction U-Net (IRU-Net), to reconstruct skin OCTA images by using the fast two-repeat OCT scan method. In a traditional two-repeat OCT scan, the angiography images were seriously degraded by high-level shot noise, and the contrast of the vessel signal was low due to an insufficient number of repeats and movement from the *in vivo* human skin (i.e., SNR is 1.6 in two-repeat scans and 2.7 in six-repeat scans.).

This study demonstrates the performance of the IRU-Net in the OCTA image reconstruction task compared with a series of state-of-the-art networks (e.g., UNet [26], UNet-ViT [27] and SRResNet [28]). We also investigate the influence of the VGG19-based content loss on the network optimization under different control weights and output layers settings and the way of optimising training parameters. Based on the quantitative results, the proposed method demonstrated the capability of reducing OCT scanning time, while maintaining the quality of the OCTA image reconstruction. Finally, the counterpart skin OCTA datasets used in this study and trained networks will be published to allow further investigations.

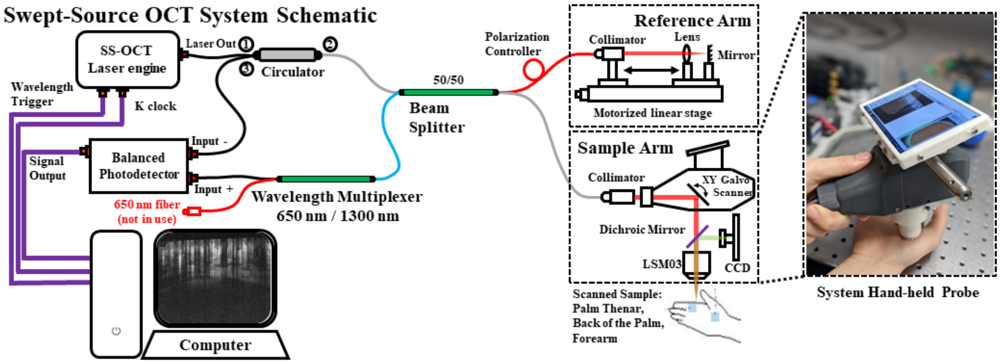

Fig. 1. The swept-source optical coherence tomography system schematic. The scan positions were palm thenar, back of the palm and forearm. The scanning was based on the hand-held probe shown in the right figure. The scanning probe was fixed and stable during the data acquisition. (LSM03, Thorlabs Inc.; CCD: charge-coupled device).

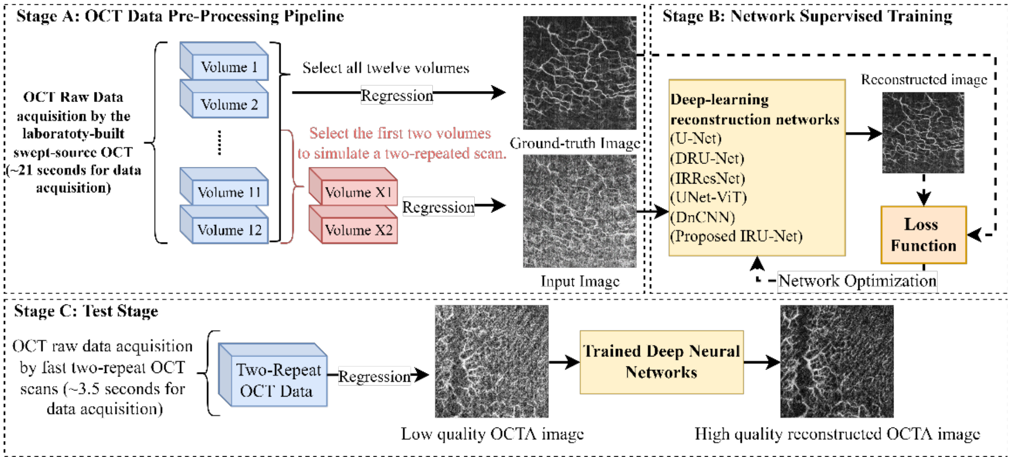
Stage A: OCT Data Pre-Processing Pipeline
OCT Raw Data acquisition by the laboratoty-built swept-source OCT (~21 seconds for data acquisition)
Volume 1
Volume 2
Volume 11
Volume 12
Select all twelve volumes
Regression
Ground-truth Image
Select the first two volumes to simulate a two-repeated scan.
Volume X1
Volume X2
Regression
Input Image
Stage B: Network Supervised Training
Deep-learning reconstruction networks
(U-Net)
(DRU-Net)
(IRResNet)
(UNet-ViT)
(DnCNN)
(Proposed IRU-Net)
Reconstructed image
Loss Function
Network Optimization
Stage C: Test Stage
OCT raw data acquisition by fast two-repeat OCT scans (~3.5 seconds for data acquisition)
Two-Repeat OCT Data
Regression
Low quality OCTA image
Trained Deep Neural Networks
High quality reconstructed OCTA image

Fig. 2. (Stage A) The OCTA data pre-process pipeline to generate the neural network train datasets and validation datasets. One volume of OCT data with a shape of 600 × 600 × 360. The input image is generated based on the two volumes of OCT scan data, and the ground-truth image with higher SNR is generated based on twelve volumes of OCT scan data. (Stage B) The simple demonstration of the supervised training pipeline for the neural networks. The implementation details of the loss functions were different in the network training. (Stage C) The testing stage of the neural networks. The high-quality OCTA image is generated based on the fast two-repeat OCT scan data. IRResNet is the modified version of the SRResNet, which remove the upsample pixel-shuffler layer.

2. Data Preparation

2.1 System Setup and Participants

The imaging system used in this study was a laboratory-built, portable, and non-invasive swept-source OCT (SSOCT) system with a hand-held probe. Fig. 1 is a demonstration of the SSOCT system setup. The system and its hand-held probe have been detailed and described in our previous publication [29]. In brief, the light source in this system is a swept laser source (SL132120, Thorlabs Inc.) with a central wavelength of 1310 ± 100 nm, and a swept rate of 200kHz. The focus length of the sample arm lens is 35 mm (LSM03, Thorlabs Inc.). This system has a lateral resolution of 19.68 μm and a theoretical axial resolution of 7.4 μm in air. For the determination of ground truth, we used OCTA images that were generated from 12-repeat OCT scans (~21 seconds of data acquisition time) as reference images [30]. These OCTA images, generated from 12-repeat OCT scans, represent high SNR (i.e., in this study, the high-quality OCTA images have an SNR greater than 3.0) and exhibit good vascular connectivity and architecture. One OCT scanned volume consists of 600 B-scans; one B-scan consists of 600 A-lines, and one A-line contains ~2.0 mm depth information. The field of view was set as 5.16 mm × 5.16 mm. The scanning positions are the forearm, palm thenar, and back of the palm. Each position of the single participant was scanned by the SSOCT device three times, and the images with the least motion artefacts and highest image quality were selected.

The data collection of the volunteers was approved by the School of Science and Engineering Research Ethics Committee of University of Dundee (Approval Number: UOD-SSREC-PGR-2022-003), which also conformed to the tenets of the Declaration of Helsinki. There are seven health participants from the age range of 20 to 35 (two females and five males). All participants had to give their informed consent before entering the lab for the data collection, and the data collected in this article had obtained the informed consent of the participants. The collected data were anonymised, and the participant's identification was removed.

2.2 Data Pre-Processing

In total, 21 high-quality raw OCT data were obtained after the data collection. Fig. 2 (Stage A) was the pre-processing pipeline for the single OCT raw data. The OCT raw data with twelve repeat scans are used to generate the ground-truth images. The corresponding high noise and low SNR input images were generated by selected first two repeat scans from the raw OCT data. Considering that the skin structure contains complicated multi-surface layers, the regression algorithm was also applied to the input images to previously subtract part of the static signals. A fast Fourier transformation (FFT)-based non-rigid B-spline transformation was used to reduce speckle noise during the OCT scanning and suppress participants' motion-induced artefacts. Then, the complex-signal-based eigen-decomposition (ED)-OCTA algorithm was utilized to extract the OCTA image [31]. Compared with the phase-compensation technique, the results from the ED-OCTA algorithm were shown to be less sensitive to tissue motion and have better performance in static tissue suppression [32]. Finally, the selected depth of the enface OCTA images was between 0.2mm and 1.2mm in depth axis (z-axis) to ensure most of the vascular signals were included. In our data collected in this study, the enface OCTA images shallower than 0.2 mm predominantly represent epidermis images without vascular signals, and images with a depth exceeding 1.2 mm exhibit weak vascular signals. Based on considerations related to training efficiency and neural network performance mentioned in this study, we decided to use a selection depth ranging from 0.2 mm to 1.2 mm. Equation (1) and Equation (2) are the descriptions of the ED-OCTA algorithm.

$$E \wedge E^H = \sum_{i=1}^{N} \lambda_B(i) e_B(i) e_B^H(i) \tag{1}$$

where H is the Hermitian transpose operation. N is the number of repeated scans. $E = [e_B(1), e_B(2), \dots, e_B(N)]$ is the $N \times N$ unitary matrix of eigenvectors, $\wedge = [\lambda_B(1), \lambda_B(2), \dots, \lambda_B(N)]$ is the $N \times N$ diagonal matrix of eigenvalues. The eigenvalues $\wedge$ are sorted in descending order. And the static signal components are the main contribution of the first K-th eigenvectors. Thus, the extraction of the moving signals (e.g., blood signals) by the ED-OCTA algorithm can be written as the (2).

$$X_m = \left[I - \sum_{i=1}^{K} e_B(i) e_B^H(i) \right] X \tag{2}$$

where the X is the tissue signal from the OCT data. I is the identity matrix. The value of K depends on the number of repeated scans. For the generating of ground-truth images, the K is set as 7 when the number of repeated OCT scans is 12 since K=7 can provide the best quality of OCTA images in this study. In terms of input images, the K is set as 1 when the number of repeated OCT scans is 2. $e_B(i)$ is the $1 \times N$ unitary matrix of eigenvectors. H is the Hermitian transpose operation. X_m represents the moving signals from the OCT signals after subtracting the static signals. After the pre-processing by the pipeline in Fig. 2 (Stage A), 1784 pairs of enface OCTA images were generated from twenty-one raw OCT files. 77% of images (1384 pairs of images) from 17 raw OCT files were used as training datasets. The remaining 23% of images (400 pairs of images) from the other 4 raw OCT files were used as validation datasets.

3. Image Reconstruction Methods

3.1 Image Reconstruction Methods

Inspired by the comparative result in [33], the encoder-decoder architecture was utilized in the IRU-Net. Based on the encoder-decoder architecture [26], we proposed the IRU-Net to reconstruct the high-quality OCTA image from the two-repeat B-scan OCTA image. Fig. 3 is the IRU-Net architecture, and Fig. 2 (Stage B) is the supervised training pipeline for the IRU-Net. To improve the performance of the image reconstruction, the residual learning strategy [34] and densely connected method [35] were used to stabilize the network training and strengthen the extracted features sharing ability between the shallow layer and the deep layer. Based on the residual dense block (RDB) [36], we replaced the receptive field from 1 × 1 to the 3 × 3 in the last layer of RDB to stabilise the training and called modified RDB (mRDB), which is also depicted in Fig. 3. In mRDB, the batch normalization layer was removed to increase the performance and reduce the computational cost in the image reconstruction task [37,38]. The output from the mRDB and the concatenate layer were used to provide the extracted feature from the encoder to better reconstruct the OCTA images.

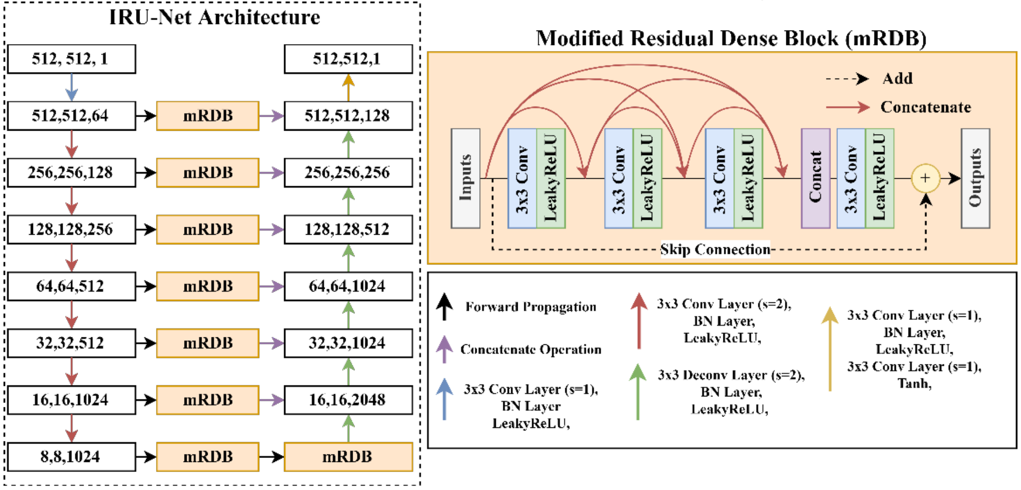

Fig. 3. The architecture of the IRU-Net. The setting of the filter size in the convolution layers of each mRDB was equal to the channel number of the input tensor. In the network, all convolution neural layers were not included with bias, and each convolution neural layer was applied the kernel initializer with a mean was 0, and a standard deviation was 0.02. (In the modified residual dense block (mRDB), the filter size of the convolution layer (blue blocks) was equal to the number of the feature channels of the input tensor. The kernel size was set as 3 × 3, and the strides was set as 1 × 1, and the padding was set as 'same' padding. The red arrow was the concatenate operation in the feature channel dimension. The black dotted arrow was the skip connection for the element-wise summation operation.

3.2 Loss Function

In the image reconstruction task, the mean absolute error (MAE, or L_1) loss and mean square error (MSE, or L_2) loss were the most used loss functions to optimize the reconstructed images pixel-by-pixel. Equation (3) and Equation (4) are the L_1 loss function and L_2 loss function, respectively.

$$L_1(y,\hat{y}) = \frac{1}{n}\sum_{i=1}^{n}|y_i - \hat{y}_i| \tag{3}$$

$$L_2(y,\hat{y}) = \frac{1}{n}\sum_{i=1}^{n}(y_i - \hat{y}_i)^2 \tag{4}$$

where y is the ground-truth image, $\hat{y}$ is the network reconstructed image, n is the total number of the pixel in the y and $\hat{y}$, and i in the y_i and $\hat{y}_i$ means the $No. i$ pixel. Although many studies showed that the L_2 loss will cause the reconstructed image to become blurry [38,39], in this study, the L_2 loss was utilized because it can stabilize the network training and the network convergence was better. Moreover, the experiment results in Fig. 6 show that L_2 loss can reduce the noise in the reconstructed OCTA image, and the result in [22] also supports this advantage. The content loss has been used to enhance the high-frequency detail of the reconstructed image and reduce the negative influence from the L_2 loss [40]. The content loss had performed outstandingly in the image super-resolution [38] and reconstruction [41] by optimizing the network based on the extracted feature maps from the pre-trained network. The $L_{content}$ loss function is shown in Equation (5).

$$L_{content}(y,\hat{y}) = \frac{1}{C_i H_j W_k} \sum_{h=1}^{H_j} \sum_{w=1}^{W_k} \sum_{c=1}^{C_i} \left(G\left(y_{h,w,c}\right) - G\left(\hat{y}_{h,w,c}\right)\right)^2 \qquad (5)$$

C, H, W are the channel, height and weight of the reconstructed image and the ground-truth image. i, j, k are the number of pixels in each dimension of the image. G is the pre-trained network output layer to extract the image features from the input tensor. Inspired by [28,38], the low computing cost ImageNet pre-trained VGG19 network [42] was used as G in Equation (5) to provide an extracted feature map. Fig. 4 is the demonstration of the skin OCTA image feature maps extracted by a pre-trained VGG19 network. Considering the computing cost and limited memory in the graphics card, the ResNet-32 and ResNet-50, which have larger weights, are not used as G for content loss.

In the content loss, the convolution neural layer output before the ReLU activation layer was used to reduce the influence of the activation function [38]. The $'Block5\_Conv4'$ in the VGG19 network was used as the default output layer as G in (5) to calculate the $L_{content}$. Finally, the combined object function (L_c) for the IRU-Net is shown in (6):

$$L_c(y,\hat{y}) = \alpha * L_2(y,\hat{y}) + \beta * L_{content}(y,\hat{y}) \qquad (6)$$

where α and β were the parameters to control the weight of loss function.

4. Experiments

4.1 Implementation Details

Before the network training, the input images and ground-truth images were resized from 600×600 to 512×512 and then normalized to between 0 and 1. The Gaussian noise with σ=0.4 was applied to the input images in training datasets to enhance the generalization of the trained network and simulate the shot noise generated from the balance photon detector. The default setup of the applied loss function Equation (6) was α=1 and β=0.01, and the output layer in (5) was $'block5\_conv4'$. Adam [43] was used as the optimizer to update the IRU-Net, and it was set as $learning\ rate = 1 \times 10^{-4}$, $beta1 = 0.8$ and $beta2 = 0.999$. The learning rate was decayed by a factor of 0.95 every 1×10^4 training steps. The batch size was set as 4, and the training epoch was set as 400. The early stopping was used to stop the network training when the calculated validation loss did not improve for 20 consecutive epochs. The validation loss is calculated at the end of each training epoch. The network training was under an NVIDIA RTX 3090 graphic card.

4.2 Comparison with State-of-the-Art Neural Networks

To evaluate the performance of the IRU-Net, we first compared our IRU-Net with several public networks, including the DnCNN [22], U-Net [33], DRU-Net [44], UNet-ViT [27] and SRResNet [28]. Fig. 2 (Stage C) demonstrates the test pipeline. The architecture of the SRResNet was proposed for the image super-resolution task, and it also performed well in medical image denoising [45]. Hence, we modified the SRResNet by removing the pixel-shuffler blocks, and the modified version of the network was called the image reconstruction residual network (IRResNet). There were 12 convolution to transformer blocks in UNet-ViT, and the number of heads in the transformer block was set as 6. The architectures and loss function of the other compared networks were the same as the public. To provide a fair comparison and ensure the compared networks were well trained, the implementation details of those networks were: batch size was set as 8; the epoch was set as 400; the early stopping was used when the loss of the network was not updated under the 20 epochs. The optimizers for the networks were Adam [43].

4.3 Performance of the Loss Function

The VGG19-based content loss has been proven can enhance the image quality of the reconstrued image in perception and visualization [38]. However, there was still a lack of studies to provide the optimal setting of the content loss in the OCTA image reconstruction task. Therefore, in this section, the aim was to investigate the performance of the proposed loss function under the two different settings: 1) Different control weight settings of the content loss (i.e., β in (6)); 2) Different VGG19 network output layer settings of the content loss (i.e., G in (5)). The default implementation details mentioned above were set as the baseline group for comparison. In each experiment, only one parameter (i.e., the output layer setting or control weight setting) of content loss was changed to reduce the influence of other parameters (e.g., batch size and learning rate).

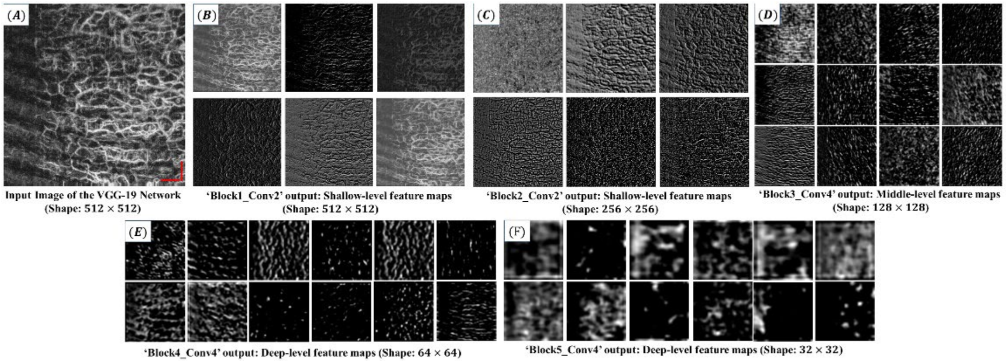
(A)
(B)
(C)
(D)
Input Image of the VGG-19 Network (Shape: 512 × 512)
'Block1_Conv2' output: Shallow-level feature maps (Shape: 512 × 512)
'Block2_Conv2' output: Shallow-level feature maps (Shape: 256 × 256)
'Block3_Conv4' output: Middle-level feature maps (Shape: 128 × 128)
(E)
(F)
'Block4_Conv4' output: Deep-level feature maps (Shape: 64 × 64)
'Block5_Conv4' output: Deep-level feature maps (Shape: 32 × 32)

Fig. 4. The extracted feature maps of the OCTA image by VGG19 network. (A) The input OCTA image for the VGG-19 network; (B) The output from the VGG-19 'block1_conv2', the shape was $512 \times 512 \times 64$; (C) The output from the VGG-19 'block2_conv2', the shape was $256 \times 256 \times 128$; (D) The output from the VGG-19 'block3_conv4', the shape was $128 \times 128 \times 256$; (E) The output from the VGG-19 'block4_conv4', the shape was $64 \times 64 \times 512$; (F) The output from the VGG-19 'block5_conv4', the shape was $32 \times 32 \times 512$. Scale: 645 µm.

To investigate the performance of the loss function in the different settings of the control weight, the parameters β in Equation (6) were changed, as Table I. shows. The L_c^4 was the baseline group. The other implementation details were the same. A further experiment was designed to investigate the content loss performance in different output layer settings. The different setup of the output layer in $L_{content}$ in Equation (5) was shown in Table II. The aim was to investigate which setting of the content loss output layers can provide the best performance for the OCTA image reconstruction. The $L_{content}^5$ was the baseline group. The other implementation details were the same.

Table I. The different Settings of the Control Weights in the Loss Function

Control Weights Setting	Loss Function
$\boldsymbol{\alpha = 1; \beta = 0}$	$L_c^1(y,\hat{y}) = L_2(y,\hat{y})$
$\boldsymbol{\alpha = 1; \beta = 1 \times 10^{0}}$	$L_c^2(y,\hat{y}) = L_2(y,\hat{y}) + 1 * L_{content}(y,\hat{y})$
$\boldsymbol{\alpha = 1; \beta = 1 \times 10^{-1}}$	$L_c^2(y,\hat{y}) = L_2(y,\hat{y}) + 0.1 * L_{content}(y,\hat{y})$
$\boldsymbol{\alpha = 1; \beta = 1 \times 10^{-2}}$	$L_c^4(y,\hat{y}) = L_2(y,\hat{y}) + 0.01 * L_{content}(y,\hat{y})$
$\boldsymbol{\alpha = 1; \beta = 1 \times 10^{-3}}$	$L_c^5(y,\hat{y}) = L_2(y,\hat{y}) + 0.001 * L_{content}(y,\hat{y})$

Table II. The Difference Setup of the Content Loss Output Layers

Symbol	Output Layer Setting
$\boldsymbol{L_{content}^1}$	$'block1\_conv2'$
$\boldsymbol{L_{content}^2}$	$'block2\_conv2'$
$\boldsymbol{L_{content}^3}$	$'block3\_conv4'$
$\boldsymbol{L_{content}^4}$	$'block4\_conv4'$
$\boldsymbol{L_{content}^5}$	$'block5\_conv4'$
$\boldsymbol{L_{content}^6}$	$'block4\_conv4' + 'block5\_conv4'$

$L_{content}^{7}$	$'block1\_conv2' + 'block2\_conv2' + 'block3\_conv4' + 'block4\_conv4' + 'block5\_conv4'$

4.4 Evaluation Metrics

The metrics method was necessary to quantitatively evaluate the reconstructed image result and evaluate the performance of the network for parameter tuning. In this study, the evaluation metrics were peak-signal-to-noise ratio (PSNR) and structural similarity index measure (SSIM) [46]. The equation of the PSNR is in (7), and the equation of the SSIM is shown in (8).

$$PSNR = 10\, log_{10}\left(\frac{I_{max}^{2}}{MSE\left(I,\hat{I}\right)}\right) \tag{7}$$

where I is the ground-truth image, $\hat{I}$ is the reconstructed image from the network. I_{max} is the maximum value in the images. MSE is the mean square error in (4).

$$SSIM\left(I,\hat{I}\right) = C_l\left(I,\hat{I}\right)^{\alpha} C_c\left(I,\hat{I}\right)^{\beta} C_s\left(I,\hat{I}\right)^{\gamma} \tag{8}$$

where I is the ground-truth image, $\hat{I}$ is the reconstructed image from the network. $\alpha > 0, \beta > 0, \gamma > 0$, and they are the parameters to adjust the weights of C_l, C_c and C_s. $C_l\left(I,\hat{I}\right)$, $C_c\left(I,\hat{I}\right)$ and $C_s\left(I,\hat{I}\right)$ are the comparison of luminance, contrast, and structure between the I and $\hat{I}$.

5. Results

5.1 Comparison with State-of-the-Art Networks

Fig. 5 shows the quantitative comparison of the results for different types of networks. Fig. 6 is the visual results of the reconstructed enface OCTA image from different networks. The visual results were from two different raw OCT files in the validation datasets. The yellow and orange arrows in Fig. 6 show that the reconstructed result from our proposed method (G1/G2) can present the best vascular texture details (PSNR: 23.979; SSIM: 0.503) than the other public methods (B1/B2, C1/C2, D1/D2, E1/E2). Furthermore, compared with the (A1/ A2) ground-truth images, the results from the encoder-decoder architecture networks (E-G) have a higher contrast and less noise (PSNR>23). Based on the quantitative result, compared with the state-of-the-art networks, the results from the IRU-Net (PSNR: 24.23 ± 0.83; mean SSIM:0.59 ± 0.09) had the best performance in OCTA image reconstruction, and the standard deviation was the smallest, showing that our method has good generalization in the image reconstruction.

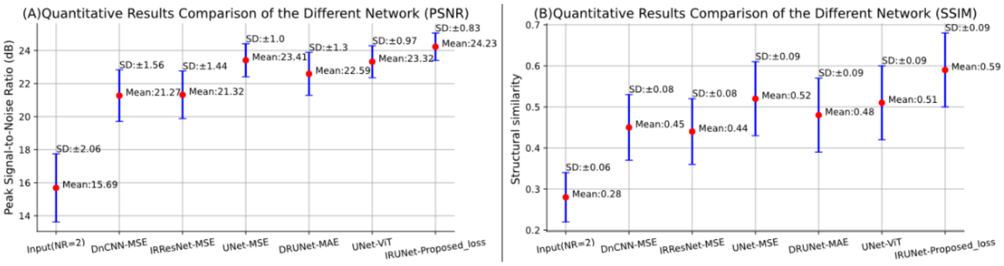

Fig. 5. The quantitative results comparison of the different networks. (SD: standard deviation).

5.2 Comparison with the Different Control Weights

Fig. 7. is the performance of the IRU-Net under training in the different settings of the control weights in the proposed loss function. Fig. 8 is the visual results of the reconstructed enface OCTA image from different settings of the loss function. Based on the result in Fig. 7, the IRU-Net had the best performance (PSNR: 24.23 ± 0.83; SSIM: 0.59 ± 0.09) when $\beta = 1 \times 10^{-2}$ in the content loss. Nevertheless, the performance of the IRU-Net with the proposed loss function will be worse than L_2 loss-only (PSNR: 23.56 ± 0.95; SSIM: 0.56 ± 0.09) when the setting of the β was larger than the 1×10^{-2}. However, when the β setting was too small (i.e., $\beta \leq 1 \times 10^{-3}$), the performance of IRU-Net (PSNR: 23.5 ± 1.18; SSIM: 0.56 ± 0.09) will be worse than proposed setting (β=0.01). Based on the visual result in Fig. 8, compared with the ground truth image (Fig. 8 (A)), the reconstruction enface OCTA images in Fig. 8 (B-F) have less noise and higher contrast in visual observation. The quantitative result in (E) has the best performance (PSNR: 24; SSIM: 0.53).

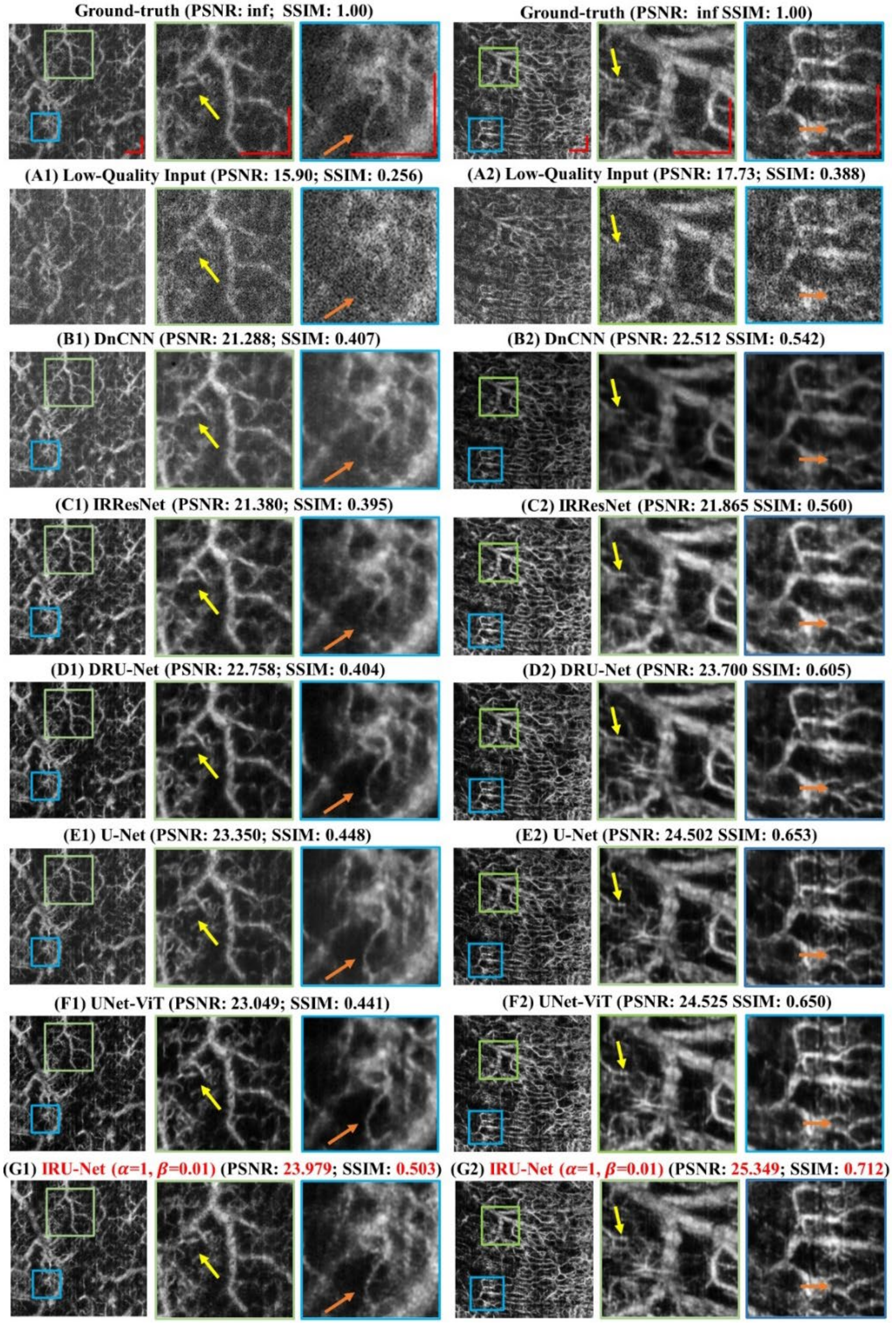
Ground-truth (PSNR: inf; SSIM: 1.00)
Ground-truth (PSNR: inf SSIM: 1.00)
(A1) Low-Quality Input (PSNR: 15.90; SSIM: 0.256)
(A2) Low-Quality Input (PSNR: 17.73; SSIM: 0.388)
(B1) DnCNN (PSNR: 21.288; SSIM: 0.407)
(B2) DnCNN (PSNR: 22.512 SSIM: 0.542)
(C1) IRResNet (PSNR: 21.380; SSIM: 0.395)
(C2) IRResNet (PSNR: 21.865 SSIM: 0.560)
(D1) DRU-Net (PSNR: 22.758; SSIM: 0.404)
(D2) DRU-Net (PSNR: 23.700 SSIM: 0.605)
(E1) U-Net (PSNR: 23.350; SSIM: 0.448)
(E2) U-Net (PSNR: 24.502 SSIM: 0.653)
(F1) UNet-ViT (PSNR: 23.049; SSIM: 0.441)
(F2) UNet-ViT (PSNR: 24.525 SSIM: 0.650)
(G1) IRU-Net (α=1, β=0.01) (PSNR: 23.979; SSIM: 0.503)
(G2) IRU-Net (α=1, β=0.01) (PSNR: 25.349; SSIM: 0.712)

Fig. 6. Comparison results of the high SNR ground-truth image with corresponding reconstruction results from the different networks. Ground-truth image from the validation set (Based on the independent raw OCT datasets); (A1, A2) Low-quality two-repeat OCTA input image; (B1, B2) DnCNN; (C1, C2) IRResNet; (D1, D2) DRU-Net; (E1, E2) U-Net; (F1, F2) UNet-ViT; (G1, G2) IRU-Net (ours); The scale bar was the 645µm.

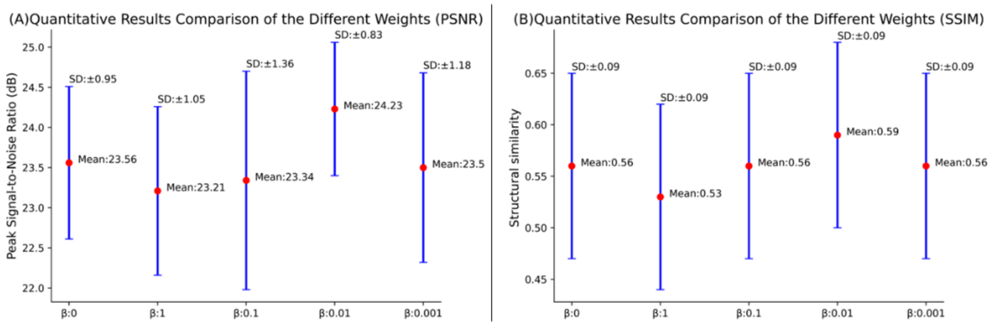

Fig. 7. The quantitative results comparison of the different settings of the loss function control weights. (SD: standard deviation of the results).

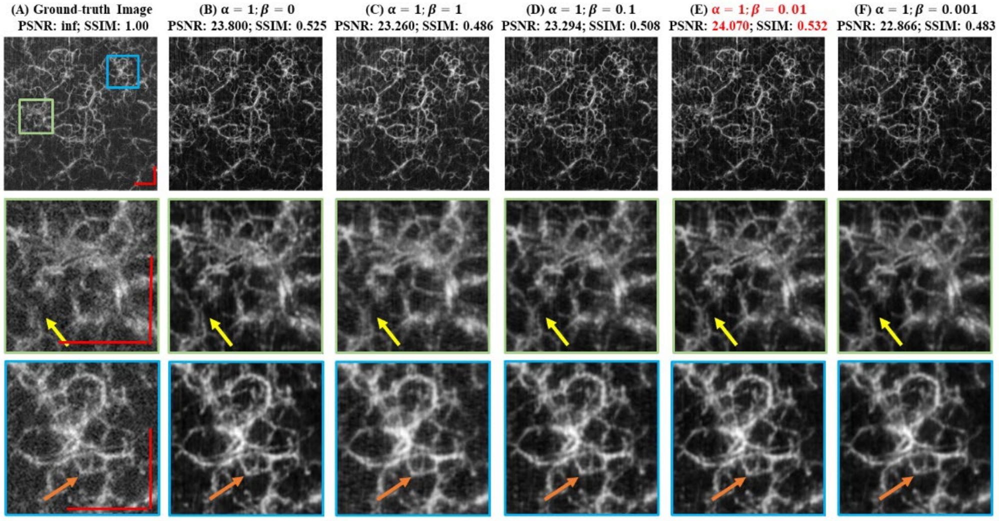

Fig. 8. Comparison results of the high SNR ground-truth image with corresponding reconstruction results from the different loss function settings of the control weights. (A) Ground-truth image from the validation set (Based on the independent raw OCT datasets); (B) α=1 and β=0; (C) α=1 and β=1; (D) α=1 and β=0.1; (E) α=1 and β=0.01 (compared group); (F) α=1 and β=0.001; The red scale bar was the 645 µm. The α and β were the weights for in Equation (6).

5.3 Comparison with Different Output Layers

Fig. 9 is the quantitative results comparison of the content loss under the different selections of the output layers. Fig. 10 is the visual comparison results of the reconstructed enface OCTA image from different settings of the output layers selection. Based on the VGG19 architecture and extracted feature maps in Fig. 4, we define $L_{content}^{1}$ and $L_{content}^{2}$ are the content loss based on the shallow-level feature maps, and $L_{content}^{3}$ is the content loss based on the middle-level feature maps, and $L_{content}^{4}$ and $L_{content}^{5}$ are the content loss based on the deep-level feature maps. Moreover, $L_{content}^{6}$ and $L_{content}^{7}$ used the combined output from the different output layers. In Fig. 9, the $L_{content}^{5}$ (PSNR: 24.23 ± 0.83; SSIM:0.59 ± 0.09) had the best performance for the IRU-Net. However, the performance of the IRU-Net results based on the middle-level feature maps (i.e., $L_{content}^{3}$) was seriously degraded (PSNR: 21.1 ± 0.93; SSIM: 0.4 ± 0.08). In Fig. 9, from $L_{content}^{1}$ to $L_{content}^{5}$, the content loss based on deep-level features maps (PSNR: 25.09; SSIM: 0.448) and shallow-level features maps (PSNR: 24.05; SSIM: 0.416) have a better performance than the content loss based on the middle-level features maps. Moreover, the results from $L_{content}^{6}$ and $L_{content}^{7}$ showed that the more combination output from the different layers, the performance of the content loss would be worse (compared with $L_{content}^{5}$ result, PSNR<25). In Fig. 10, compared with the ground-truth image in (A), the results (b, c, e-g) show the acceptable performance of micro-vascular texture details reconstruction. However, the artefacts in Fig. 10 (D)(H) were obvious, while the results in Fig. 10 (B)(F) had less noise and better image quality.

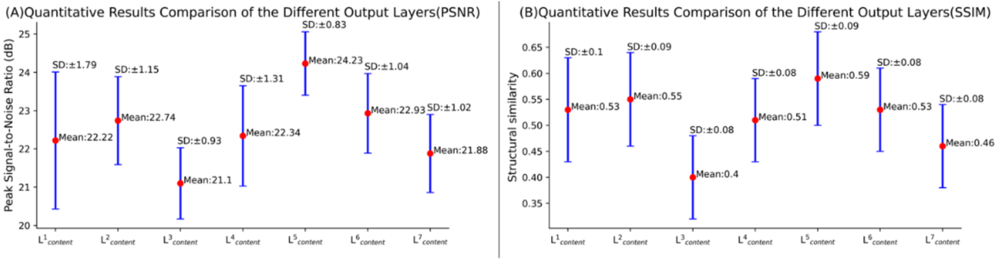

Fig. 9. The quantitative results comparison of the different selections of the output layers. (SD: standard deviation of the results).

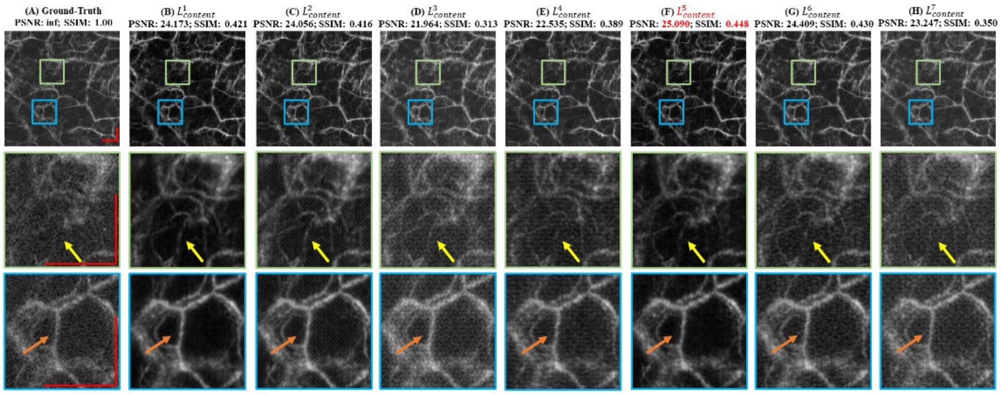
(A) Ground-Truth
PSNR: inf; SSIM: 1.00
(B) $L^1_{content}$
PSNR: 24.173; SSIM: 0.421
(C) $L^2_{content}$
PSNR: 24.056; SSIM: 0.416
(D) $L^3_{content}$
PSNR: 21.964; SSIM: 0.313
(E) $L^4_{content}$
PSNR: 22.535; SSIM: 0.389
(F) $L^5_{content}$
PSNR: 25.090; SSIM: 0.448
(G) $L^6_{content}$
PSNR: 24.409; SSIM: 0.430
(H) $L^7_{content}$
PSNR: 23.247; SSIM: 0.350

Fig. 10. Comparison results of the high SNR ground-truth image with corresponding reconstruction results from the different output layers setting in content loss. (A) Ground-truth image from the validation set (Based on the independent raw OCT dataset); (B) $\boldsymbol{L^1_{content}}$; (C) $\boldsymbol{L^2_{content}}$; (D) $\boldsymbol{L^3_{content}}$; (E) $\boldsymbol{L^4_{content}}$; (F) $\boldsymbol{L^5_{content}}$ (compared group); (G) $\boldsymbol{L^6_{content}}$; (H) $\boldsymbol{L^7_{content}}$. The red scale bar was 645 µm. The definition of $\boldsymbol{L^X_{content}}$ and the counterpart settings are in Table II.

6. Conclusion and Discussion

In this work, we proposed an IRU-Net to achieve a fast OCT-angiography scan (~3.5 seconds) while maintaining the field of view and image quality of vascular texture. We introduced the dense connection blocks to IRU-Net to improve the quality of the reconstructed OCTA images. We also investigate a VGG19-based content loss and provide the optimal setup for OCTA image reconstruction tasks. The main contribution of our work is to provide a fast deep-learning-based scan pipeline for wide field of view (6mm × 6mm) OCTA scan in skin applications. Under this scan pipeline, the motion artefact from the patients can be prevented while the field of view and image resolution are moderated for the clinical scan. Furthermore, the OCT-signal processing speed is reduced by 83% because only two-repeated scan (3.5 seconds for OCTA scan) is used in this scan pipeline, and it also decreases the size of data acquired from the SSOCT device, reducing the pre-processing time (from 3mins to 1min).

Based on the error bar in Fig. 5, the encoder-decoder network (e.g., U-Net, IRU-Net) has better performance than the end-to-end architecture network (e.g., DnCNN). That might be because of the significant difference in vascular structure in the enface OCTA images. However, the mean ± standard deviation SSIM results were low in all results. We hypothesize that it might be because the contrast of the reconstructed OCTA image was changed by the batch normalization layer. Furthermore, the reconstructed OCTA images have a lower noise level which might be because of the utilises of the $L2$ loss [22].

In Fig. 8, the difference from (B) to (F) was slightly and hard to classify in the aspect of visualized results, and the error bar in Fig. 7 shows that the proposed implementation details of the loss function can provide the optimal reconstructed results. In Fig. 9 and Fig. 10, the content loss based on the deep-level ($\boldsymbol{L_{content}^{5}}$) and shallow-level ($\boldsymbol{L_{content}^{1}}$) features maps had better performance than the other groups ($\boldsymbol{L_{content}^{3}}$, $\boldsymbol{L_{content}^{4}}$, $\boldsymbol{L_{content}^{6}}$, $\boldsymbol{L_{content}^{7}}$). In Fig. 10 yellow arrow, result in (B) can provide more vascular details than (F). However, compared with the ground-truth image (A), the result in (F) is more similar to the ground-truth image (A) than the (B) in visual and quantitative aspects. It is hard to clarify if the more vascular details provided in Fig. 10 (B) are true or generated by the neural network. We authorize that situation might be because of the content loss based on the shallow-level feature maps (i.e., $L_{content}^{1}$ and $L_{content}^{2}$ in Fig. 4) is aimed to optimize more vascular texture details in the reconstructed images. Hence, based on the experiment observation, to ensure that the reconstructed results from the IRU-Net are more closely related to the ground-truth images, and the consideration of the stability and efficiency of network training, we proposed that the implementation details of loss should be the same as the default setting.

Based on the IRU-Net architecture, we tried to increase the network depth and applied more residual connections or dense connections to improve reconstruction image quality. However, the network deeper than IRU-Net had worse performance than before. We also investigate the full densely connection network proposed by [38], called SRDenseNet, but the result (SSIM: 0.391 (mean) ± 0.063 (std); PSNR: 20.25 (mean) ± 1.32 (std)) is worse than the IRResNet after fine-tuning details with the different loss function (L_1 and L_2) and training strategy (early stopping and reducing learning rate with epoch).

Our study has limitations. First, the network architecture was based on the convolution neural layer, which cannot provide long-term information and limit the receptive field during the feature extraction. UNet-ViT introduced a transformer block to U-Net, but the performance is not better than IRU-Net. Hence, it is essential to investigate a better network architecture for image reconstruction, while lightweight and high performance. Secondly, in the network training strategy, we also investigated the performance of the IRU-Net under unsupervised training. The loss function used in unsupervised training was combined between the mean-squared-error loss and adversarial loss (i.e., generative adversarial network (GAN) [47] and relativistic average standard GAN (RaSGAN) [48]). Inspired by [38], the weight for adversarial loss was set as 0.001, and the weight for L_2 loss was 1. However, the unsupervised training strategy cannot provide higher performance and reduce the stabilization of the training. In the RaSGAN train, the PSNR is 22.35 ± 1.30, and the SSIM is 0.525 ± 0.08. In the standard-GAN, the PSNR is 22.04 ± 1.04, and the SSIM is 0.512 ± 0.08 (result in mean ± standard deviation format). We also investigated the different weights of adversarial loss (from 0.1 to 0.001) to stabilize the network training, but the result was not better than the IRU-Net result based on supervised training. It is necessary to investigate further training strategies for OCTA image reconstruction. Thirdly, the experiment results in this study were based on the data collected from health participants; hence, we concern it will be a series of potential risks when applying on diseased subjects: 1) low-quality of the collected data due to high motion artefacts from patients; 2) further investigation of the IRU-Net performance on diseased skin OCTA images is essential in future clinical studies. Fourthly, the size of data used in this study is 1784 pairs of images. We are concerned that the limited data size will lead to an over-fitting problem of the neural networks examined in this study. In the future, we plan to increase the size of the OCTA data that should help improve the robustness and generalizability of our proposed method.

Our proposed pipeline has achieved a well comparative result in image reconstruction for low-quality OCTA images acquired by a fast two-repeated OCTA scan in skin application. The fast OCTA scan can prevent motion artifacts from patients, and the processing time of the fast scan data is reduced by 60%. In the future, we will introduce this fast OCTA scan pipeline to oral and retinal scans to achieve a high-quality OCTA scan with low motion artifacts and fast OCT-signal processing.

Data Availability.

The datasets generated during and/or analysed during the current study are not publicly available due to participant privacy and violation of informed consent.

Author contributions statement

Conceptualization, J.L.; methodology, J.L..; software, J.L.; formal analysis, J.L.; investigation, J.L.; resources, J.L. and T.Z..; data curation, J.L. and T.Z.; writing—original draft preparation, J.L..; writing—review and editing, C.L., S.Y., Z.H., J.L.; visualization, J.L.; supervision, C.L. and Z.H.; All authors have read and agreed to the published version of the manuscript.

Additional Information

Competing Interests. The author(s) declare no competing interests.
Code accessible from URL: https://github.com/LiaoJinpeng/IRU-Net-Training-

References

1. S. Aumann, S. Donner, J. Fischer, and F. Müller, "Optical Coherence Tomography (OCT): Principle and Technical Realization," in *High Resolution Imaging in Microscopy and Ophthalmology: New Frontiers in Biomedical Optics*, J. F. Bille, ed. (Springer International Publishing, 2019), pp. 59–85.

2. W. Fujimoto James G. and Drexler, "Introduction to OCT," in *Optical Coherence Tomography: Technology and Applications*, J. G. Drexler Wolfgang and Fujimoto, ed. (Springer International Publishing, 2015), pp. 3–64.

3. W. Drexler and J. G. Fujimoto, "State-of-the-art retinal optical coherence tomography," Prog Retin Eye Res **27**, 45–88 (2008).

4. M. Ulrich, L. Themstrup, N. de Carvalho, M. Manfredi, C. Grana, S. Ciardo, R. Kästle, J. Holmes, R. Whitehead, and G. B. E. Jemec, "Dynamic optical coherence tomography in dermatology," Dermatology **232**, 298–311 (2016).

5. K. S. Rathod, S. M. Hamshere, D. A. Jones, and A. Mathur, "Intravascular ultrasound versus optical coherence tomography for coronary artery imaging–apples and oranges?," Interventional Cardiology Review **10**, 8 (2015).

6. A. B. E. Attia, S. Y. Chuah, D. Razansky, C. J. H. Ho, P. Malempati, U. S. Dinish, R. Bi, C. Y. Fu, S. J. Ford, and J. S.-S. Lee, "Noninvasive real-time characterization of non-melanoma skin cancers with handheld optoacoustic probes," Photoacoustics **7**, 20–26 (2017).

7. C. Wahrlich, S. A. Alawi, S. Batz, J. W. Fluhr, J. Lademann, and M. Ulrich, "Assessment of a scoring system for Basal Cell Carcinoma with multi-beam optical coherence tomography," Journal of the European Academy of Dermatology and Venereology **29**, 1562–1569 (2015).

8. K. B. E. Friis, L. Themstrup, and G. B. E. Jemec, "Optical coherence tomography in the diagnosis of actinic keratosis—A systematic review," Photodiagnosis Photodyn Ther **18**, 98–104 (2017).

9. A. J. Deegan, F. Talebi-Liasi, S. Song, Y. Li, J. Xu, S. Men, M. M. Shinohara, M. E. Flowers, S. J. Lee, and R. K. Wang, "Optical coherence tomography angiography of normal skin and inflammatory dermatologic conditions," Lasers Surg Med **50**, 183–193 (2018).

10. B. Zabihian, Z. Chen, E. Rank, C. Sinz, M. Bonesi, H. Sattmann, J. R. Ensher, M. P. Minneman, E. E. Hoover, and J. Weingast, "Comprehensive vascular imaging using optical coherence tomography-based angiography and photoacoustic tomography," J Biomed Opt **21**, 096011 (2016).

11. J. Fingler, D. Schwartz, C. Yang, and S. E. Fraser, "Mobility and transverse flow visualization using phase variance contrast with spectral domain optical coherence tomography," Opt Express **15**, 12636–12653 (2007).

12. E. Jonathan, J. Enfield, and M. J. Leahy, "Correlation mapping method for generating microcirculation morphology from optical coherence tomography (OCT) intensity images," J Biophotonics **4**, 583–587 (2011).

13. Y. Jia, O. Tan, J. Tokayer, B. Potsaid, Y. Wang, J. J. Liu, M. F. Kraus, H. Subhash, J. G. Fujimoto, and J. Hornegger, "Split-spectrum amplitude-decorrelation angiography with optical coherence tomography," Opt Express **20**, 4710–4725 (2012).

14. A. Mariampillai, B. A. Standish, E. H. Moriyama, M. Khurana, N. R. Munce, M. K. K. Leung, J. Jiang, A. Cable, B. C. Wilson, and I. A. Vitkin, "Speckle variance detection of microvasculature using swept-source optical coherence tomography," Opt Lett **33**, 1530–1532 (2008).

15. J. Xu, S. Song, Y. Li, and R. K. Wang, "Complex-based OCT angiography algorithm recovers microvascular information better than amplitude-or phase-based algorithms in phase-stable systems," Phys Med Biol **63**, 015023 (2017).

16. B. Baumann, C. W. Merkle, R. A. Leitgeb, M. Augustin, A. Wartak, M. Pircher, and C. K. Hitzenberger, "Signal averaging improves signal-to-noise in OCT images: But which approach works best, and when?," Biomed. Opt. Express **10**, 5755–5775 (2019).

17. Y. Ji, K. Zhou, S. H. Ibbotson, R. K. Wang, C. Li, and Z. Huang, "A novel automatic 3D stitching algorithm for optical coherence tomography angiography and its application in dermatology," J Biophotonics **14**, e202100152 (2021).

18. Y. Giarratano, E. Bianchi, C. Gray, A. Morris, T. MacGillivray, B. Dhillon, and M. O. Bernabeu, "Automated segmentation of optical coherence tomography angiography images: benchmark data and clinically relevant metrics," Transl Vis Sci Technol **9**, 5 (2020).

19. D. Lu, M. Heisler, S. Lee, G. Ding, M. v Sarunic, and M. F. Beg, "Retinal fluid segmentation and detection in optical coherence tomography images using fully convolutional neural network," arXiv preprint arXiv:1710.04778 (2017).

20. T. Marvdashti, L. Duan, S. Z. Aasi, J. Y. Tang, and A. K. E. Bowden, "Classification of basal cell carcinoma in human skin using machine learning and quantitative features captured by polarization sensitive optical coherence tomography," Biomed Opt Express **7**, 3721–3735 (2016).

21. M. Wang, W. Zhu, K. Yu, Z. Chen, F. Shi, Y. Zhou, Y. Ma, Y. Peng, D. Bao, and S. Feng, "Semi-supervised capsule cGAN for speckle noise reduction in retinal OCT images," IEEE Trans Med Imaging **40**, 1168–1183 (2021).

22. X. Liu, Z. Huang, Z. Wang, C. Wen, Z. Jiang, Z. Yu, J. Liu, G. Liu, X. Huang, A. Maier, Qiushu Ren, and Yanye Lu, "A deep learning based pipeline for optical coherence tomography angiography," J Biophotonics **12**, e201900008 (2019).

23. C. S. Lee, A. J. Tyring, Y. Wu, S. Xiao, A. S. Rokem, N. P. DeRuyter, 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**, 1–11 (2019).

24. M. Gao, Y. Guo, T. T. Hormel, J. Sun, T. S. Hwang, and Y. Jia, "Reconstruction of high-resolution 6× 6-mm OCT angiograms using deep learning," Biomed Opt Express **11**, 3585–3600 (2020).

25. A. Tavakkoli, S. A. Kamran, K. F. Hossain, and S. L. Zuckerbrod, "A novel deep learning conditional generative adversarial network for producing angiography images from retinal fundus photographs," Sci Rep **10**, 1–15 (2020).

26. O. Ronneberger, P. Fischer, and T. Brox, "U-net: Convolutional networks for biomedical image segmentation," in *International Conference on Medical Image Computing and Computer-Assisted Intervention* (Springer, 2015), pp. 234–241.

27. D. Torbunov, Y. Huang, H. Yu, J. Huang, S. Yoo, M. Lin, B. Viren, and Y. Ren, "UVCGAN: UNet Vision Transformer cycle-consistent GAN for unpaired image-to-image translation," arXiv preprint arXiv:2203.02557 (2022).

28. 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), pp. 4681–4690.

29. Y. Ji, K. Zhou, S. H. Ibbotson, R. K. Wang, C. Li, and Z. Huang, "A novel automatic 3D stitching algorithm for optical coherence tomography angiography and its application in dermatology," J Biophotonics **14**, e202100152 (2021).

30. R. K. Wang, A. Zhang, W. J. Choi, Q. Zhang, C. Chen, A. Miller, G. Gregori, and P. J. Rosenfeld, "Wide-field optical coherence tomography angiography enabled by two repeated measurements of B-scans," Opt Lett **41**, 2330–2333 (2016).

31. Q. Zhang, J. Wang, and R. K. Wang, "Highly efficient eigen decomposition based statistical optical microangiography," Quant Imaging Med Surg **6**, 557 (2016).

32. S. Yousefi, Z. Zhi, and R. K. Wang, "Eigendecomposition-based clutter filtering technique for optical microangiography," IEEE Trans Biomed Eng **58**, 2316–2323 (2011).

33. Z. Jiang, Z. Huang, B. Qiu, X. Meng, Y. You, X. Liu, G. Liu, C. Zhou, K. Yang, and A. Maier, "Comparative study of deep learning models for optical coherence tomography angiography," Biomed Opt Express **11**, 1580–1597 (2020).

34. 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), pp. 770–778.

35. 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), pp. 4700–4708.

36. Y. Zhang, Y. Tian, Y. Kong, B. Zhong, and Y. Fu, "Residual dense network for image restoration," IEEE Trans Pattern Anal Mach Intell (2020).

37. S. Nah, T. Hyun Kim, and K. Mu Lee, "Deep multi-scale convolutional neural network for dynamic scene deblurring," in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition* (2017), pp. 3883–3891.

38. 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 (ECCV)* (2018), p. 0.

39. Y. Ma, X. Chen, W. Zhu, X. Cheng, D. Xiang, and F. Shi, "Speckle noise reduction in optical coherence tomography images based on edge-sensitive cGAN," Biomed Opt Express **9**, 5129–5146 (2018).

40. L. A. Gatys, A. S. Ecker, and M. Bethge, "A neural algorithm of artistic style," arXiv preprint arXiv:1508.06576 (2015).

41. G. Yang, S. Yu, H. Dong, G. Slabaugh, P. L. Dragotti, X. Ye, F. Liu, S. Arridge, J. Keegan, and Y. Guo, "DAGAN: deep de-aliasing generative adversarial networks for fast compressed sensing MRI reconstruction," IEEE Trans Med Imaging **37**, 1310–1321 (2017).

42. K. Simonyan and A. Zisserman, "Very Deep Convolutional Networks for Large-Scale Image Recognition," (2014).

43. D. P. Kingma and J. Ba, "Adam: A method for stochastic optimization," arXiv preprint arXiv:1412.6980 (2014).

44. K. Zhang, Y. Li, W. Zuo, L. Zhang, L. van Gool, and R. Timofte, "Plug-and-play image restoration with deep denoiser prior," IEEE Trans Pattern Anal Mach Intell (2021).

45. Z. Dong, G. Liu, G. Ni, J. Jerwick, L. Duan, and C. Zhou, "Optical coherence tomography image denoising using a generative adversarial network with speckle modulation," J Biophotonics **13**, e201960135 (2020).

46. A. Hore and D. Ziou, "Image quality metrics: PSNR vs. SSIM," in *2010 20th International Conference on Pattern Recognition* (IEEE, 2010), pp. 2366–2369.

47. I. Goodfellow, J. Pouget-Abadie, M. Mirza, B. Xu, D. Warde-Farley, S. Ozair, A. Courville, and Y. Bengio, "Generative adversarial nets," Adv Neural Inf Process Syst **27**, (2014).

48. A. Jolicoeur-Martineau, "The relativistic discriminator: a key element missing from standard GAN," arXiv preprint arXiv:1807.00734 (2018).