

Can deep learning improve the automatic segmentation of deep foveal avascular zone in optical coherence tomography angiography?

MENGLIN GUO,^{1,#} MEI ZHAO,^{2,#} ALLEN MY CHEONG² CORVI FEDERICO,³ YONGJIN ZHOU,^{1,*} ANDREW KC LAM,^{2,*}

¹ School of Biomedical Engineering, Health Science Center, Shenzhen University, Shenzhen 518055, China

² Centre for Myopia Research, School of Optometry, Faculty of Health and Social Sciences, The Hong Kong Polytechnic University, Hong Kong, China

³ Eye Clinic, Department of Biomedical and Clinical Science “Luigi Sacco”, Sacco Hospital, University of Milan, via G.B Grassi 74, Milan, 20157, Italy

Both authors contribute equally to this work

*Correspondence to:

Andrew KC Lam

E-mail: andrew.kc.lam@polyu.edu.hk

or Yongjin Zhou

E-mail: yjzhou@szu.edu.cn

Abstract: Optical coherence tomography angiography (OCTA) is extensively used for visualizing retinal vasculature, including the foveal avascular zone (FAZ). Assessment of the FAZ is critical in the diagnosis and management of various retinal diseases. Accurately segmenting the FAZ in the deep retinal layer (dFAZ) is very challenging due to unclear capillary terminals. In this study, a customized encoder-decoder deep learning network was used for dFAZ segmentation. Three-fold cross-validation was performed on a total of 80 subjects (63 healthy subjects and 17 diabetic retinopathy subjects). The proposed method obtained an average Dice of 0.86 and an average Hausdorff distance of 23.35, suggesting the dFAZ was accurately segmented. The proposed method is expected to realize good clinical application value by providing an objective and faster and spatially-quantitative preparation of dFAZ-related investigations.

Keywords: Deep learning; automatic segmentation; optical coherence tomography angiography; deep foveal avascular zone.

1. Introduction

The fovea, which supports the highest visual acuity, is a depression of the retinal surface approximately 1 mm in diameter. The center of the fovea, which is termed the fovea centralis or foveola, has the highest cone density in the photoreceptor layer. It is generally centered within a small zone devoid of retinal blood vessels, known as the foveal avascular zone (FAZ) (J. M. Provis et al., 2013). The FAZ is surrounded by interconnected retinal capillary networks (Tick et al., 2011) and plays a critical role in the development of the foveal pit (Springer & Hendrickson, 2004). Classical histology studies suggest that there are two capillary networks (Jan M. Provis, 2001): the superficial network, which is formed in the nerve fiber layer and the ganglion cell layer, and the deep network lying in between the inner nuclear

layer and the outer plexiform layer. The deep capillary network is denser and more complex than the superficial (Hayreh, 2015). Both fluorescein angiography (FA) (Marmor & Ravin, 2011; Mendis et al., 2010) and optical coherence tomography angiography (OCTA) (Kuehlewein et al., 2015; Mammo et al., 2015; Tan et al., 2015) support the visualization of the retinal capillary networks. However, FA requires invasive intravenous dye administration by clinicians and the imaging of the deep retinal capillary network is poor (Spaide et al., 2015a). OCTA is a more convenient, non-invasive technique to visualize both the superficial and deep retinal capillary networks (Spaide et al., 2015a).

OCTA has been widely used to visualize the retinal vasculature of various retinal vascular pathologies, including but not limited to diabetic retinopathy (Dimitrova & Chihara, 2019; Enders et al., 2020; Matsunaga et al., 2015; Samara et al., 2017; Takase et al., 2015; Yoon et al., 2017), retinal vein occlusion (Coscas et al., 2016; Werner et al., 2019) macular telangiectasia type 2 (Spaide et al., 2015b; Villegas & Kovach, 2017), non-infectious posterior uveitis (Waizel et al., 2018), choroidal melanoma (Y. F. Li et al., 2017), and X-linked retinoschisis (Romano et al., 2019). The size of the FAZ correlates with the condition of the vascular circulation surrounding the fovea, whereby an enlarged FAZ commonly indicates an ischemic condition in the fovea and may indicate visual acuity (Romano et al., 2019; Samara et al., 2017). Interestingly, pathological changes in the circulation tend to be more prominent in the deep (dFAZ) than the superficial vascular layer (sFAZ). Dimitrova et al (Dimitrova & Chihara, 2019) suggested that deep-vascular-layer alterations were most frequently present in patients with diabetic macular edema. Tang et al (F. Y. Tang et al., 2020; Fang Yao Tang et al., 2017) reported a more pronounced association between the severity of diabetic retinopathy (DR) and dFAZ (F. Y. Tang et al., 2020) than with sFAZ (Fang Yao Tang et al., 2017). Similarly, the deep capillary plexus appeared to be more severely affected in retinal vein occlusion (Coscas et al., 2016), macular telangiectasia type 2 (Villegas & Kovach, 2017), and choroidal melanoma (Y. F. Li et al., 2017). The dFAZ is also a common parameter reported in myopia-related hemodynamic studies and it seems to be more sensitive to axial elongation than the superficial FAZ (sFAZ). Sung et al. (Park & Kang, 2018) found a significant association between the axial length and dFAZ, but no such correlation was observed between the axial length and superficial FAZ (sFAZ). Cheng et al (Cheng et al., 2019) also reported a stronger association between axial length and dFAZ than with sFAZ. With the extensive application of OCTA technique, the dFAZ has become increasingly important in both clinical and research studies.

Accurate segmentation of the dFAZ in the OCTA images is the prerequisite for precise quantification of the dFAZ. However, quantifying dFAZ is quite challenging, since its boundary is usually less well-defined than sFAZ due to various artifacts, such as that due to projection. There are many OCTA devices in the market, some of which have built-in software for automatic segmentation and quantification of the superficial and deep FAZ (e.g. AngioVue) (Chun et al., 2019; Falavarjani et al., 2018; Magrath et al., 2017; Pilotto et al., 2018), but others only quantify the superficial plexuses and provide the sFAZ parameters (e.g. Heidelberg, Cirrus) (Lin et al., 2020; Magrath et al., 2017; Wylegala et al., 2019). In addition to relying on the built-in software, some researchers and clinicians have employed customized automatic segmentation programs (Diaz et al., 2019; Lu et al., 2018; F. Y. Tang et al., 2020; Fang Yao Tang et al., 2017) or conducted manual delineation by using public images processing tools, such as ImageJ (Corvi et al., 2018) and AngiOCTool (Lupidi et al., 2016). Several studies have concluded that the built-in automatic segmentation software was less reliable than manual segmentation (Lin et al., 2020; Linderman et al., 2017). The built-in software might mislead due to the noise signals inside the FAZ and recognize the noise as capillaries, resulting in an inaccurate FAZ detection (Lin et al., 2020). Conventional image processing techniques employed by customized programs, include the denoising

filters, region-growing approach, and morphologic operators (Diaz et al., 2019; Lu et al., 2018; F. Y. Tang et al., 2020; Fang Yao Tang et al., 2017). However, the major limitation shared by most non-deep learning techniques is the empirical parameter settings, which are highly dependent on the designer’s prior knowledge and are sensitive to the position, size, and shape of the FAZ. These methods could have excellent performance on a small dataset but might encounter more difficulties when applied to larger and more complex datasets. Manual segmentation is reliable, but time-consuming and might be subject to greater inter-observer variance. Shahlae et al (Shahlaee et al., 2016) showed that the interobserver agreement of manual segmentation was lower for dFAZ than sFAZ (intraclass correlation coefficient of 0.84 vs 0.95).

As conventional image processing techniques might not be sensitive enough in detecting the dFAZ boundary, while manual segmentation is very time-consuming, this study proposed the use of a deep learning solution to address this issue. The deep learning technique displayed excellent ability to determine the intricate structures of high-dimensional data and extracts their features and, therefore, could provide accurate and objective results. In addition, it can maintain excellent performance when generalizing to new datasets (LeCun et al., 2015). In a recent study (Guo et al., 2019), the deep learning method achieved good performance in accurate automatic segmentation and quantification of sFAZ, demonstrating the great potential of applying deep learning in automatic segmentation in dFAZ. In this paper, an encoder-decoder network constructed for the segmentation of dFAZ is described. A boundary alignment module (BAM) was implemented at the bottom of the decoder network to extract global information and align boundaries. Thereafter, the boundary supervision modules (BSMs) were employed as decoders to refine the segmentation results scale-by-scale with deep and boundary supervision.

2. Method

2.1 Data

Cirrus HD-OCT 5000 with AngioPlex (Carl Zeiss Meditec, Inc., Dublin, California) was used to acquire the 3×3 mm OCTA images centered on the macula. The A-scan rate of Cirrus HD-OCT 5000 is 68,000 scans per second, and the central wavelength of the light source is 840 nm. The axial and transverse resolutions in the tissue are 5 and 15 $\mu\text{m}/\text{pixel}$, respectively. A total of 45 OCTA images were collected from random eyes of 45 healthy subjects (12 high myopes with a spherical equivalent $\leq -6\text{D}$ and 33 non-high myopes with a spherical equivalent $> -6\text{D}$) at the Optometry Clinic of The Hong Kong Polytechnic University (HK PolyU).

One experienced expert in dFAZ delineation and two researchers contributed to the annotation of the ground truth (GT). The GT of dFAZ for both training and testing sets was generated manually by the experienced expert using the “polygon selections” and “fill” functions of the ImageJ (National Institutes of Health, Bethesda, MD). Following this, the two researchers reviewed the GT together and modified the boundary if necessary.

In order to improve the stability of the proposed method, a portion of 3×3 mm OCTA images from the public image dataset OCTAGON (D íz M, 2018) was included in the experiments and the details were summarized as follows: 36 images from both eyes of 18 healthy subjects and 17 images from 17 DR subjects. The data distribution of different datasets was shown in Table 1.

Table 1. Data distribution of different datasets

	Class	Subject	Image
HK Polytechnic University dataset	Healthy	45	45

	DR	0	0
OCTAGON dataset	Healthy	18	36
	DR	17	17
Total		80	98

In order to implement the boundary supervision, the boundary ground truth (BGT) was defined as:

$$\text{BGT} = \text{GT} - \text{Erode}(\text{GT}, \text{SE}) \quad (1)$$

where *Erode* denotes the morphologic erode operation and SE denotes the structural element of 9×9 kernel size.

2.2 Image Pre-processing

All OCTA images were resized to 704×704 pixel for the subsequent training and testing. The size of raw images from HK PolyU were 736×718 pixel, so they were cropped the image to 704×704 pixel with the image center maintained in the same location as the raw image, whilst for the OCTAGON dataset, the bilinear interpolation was used to upsample the 320×320 pixel raw images to 704×704 pixel in both horizontal and vertical directions. For the HK PolyU dataset, cropping was applied to preserve local details of the image, while for the OCTAGON dataset, bilinear interpolation achieved the balance of upsampling performance and running time, compared with nearest neighbor interpolation and trilinear interpolation. Examples of raw OCTA images from two datasets and their preprocessing results are shown in Fig. 1. All OCTA images were normalized by the equation (2) to substantially improve the final generalization error of the proposed network and accelerate the training. The normalization formula is defined as:

$$\hat{X} = (X - X_\mu) / X_\sigma \quad (2)$$

where \hat{X} denotes the result of the normalization, X denotes the image, X_μ denotes the mean pixel value in the image and X_σ denotes the standard deviation of the pixel values in the image.

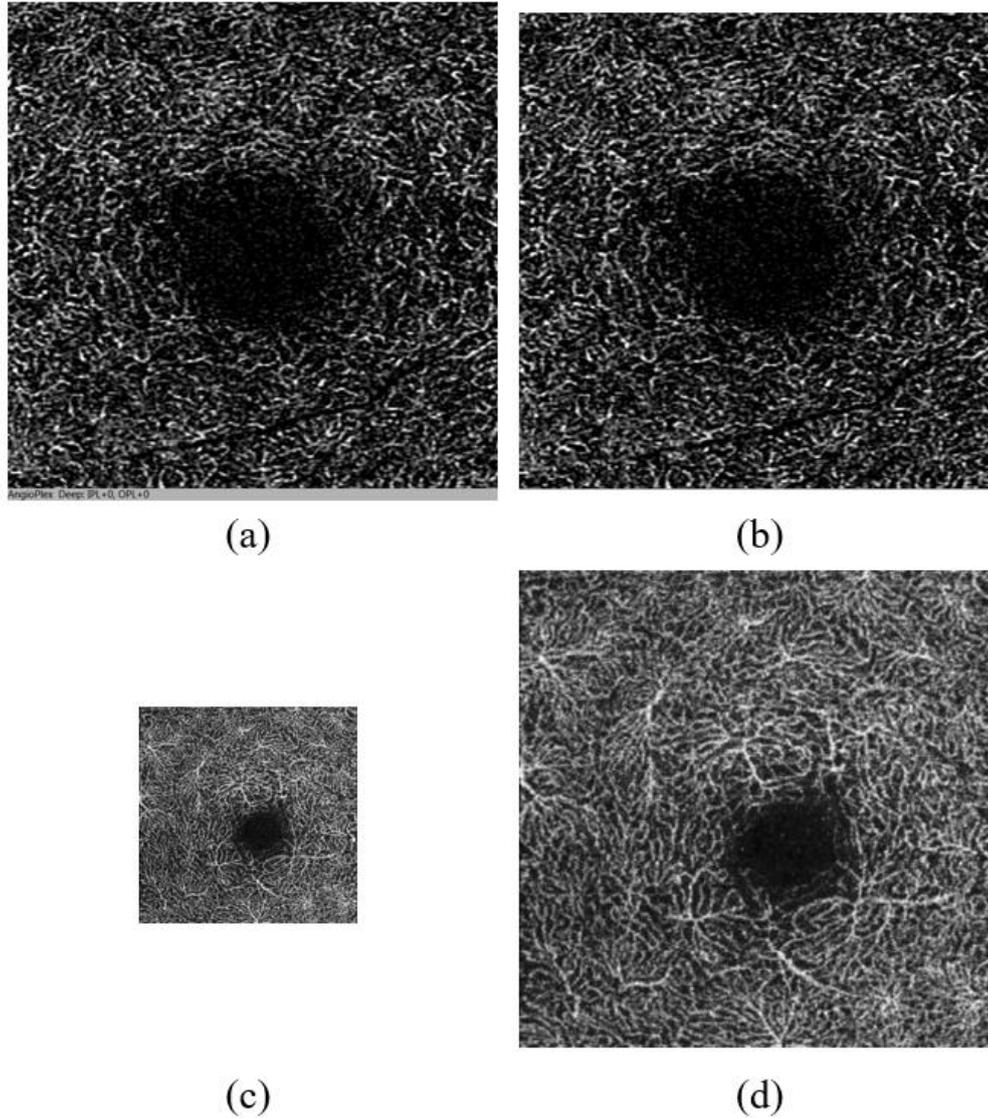


Fig. 1. Raw optical coherence tomography angiography (OCTA) images and their preprocessing results. (a) A raw OCTA image example in The Hong Kong Polytechnic University dataset. (b) The preprocessing result corresponding to (a). (c) A raw OCTA image example in the OCTAGON dataset. (d) The preprocessing result corresponding to (c).

The randomized data augmentations (a random combination of horizontal flipping, vertical flipping, 90° rotation, 180° rotation, and 270° rotation) were applied to improve the generalization performance of the proposed network model. In order to save local storage space, data augmentation was applied before network training. After the image was input to the CPU, each data augmentation method was randomly selected to be applied or not. This process only transformed the image data stored in the CPU, and did not transform the image in the local storage space. The total number of images obtained after applying data augmentation was 3136.

2.3 Network Architecture

The proposed network consists of two processing components: the encoders extract the multi-scale features from the input image and the decoders refine the segmentation result scale-by-scale as shown in Fig. 2.

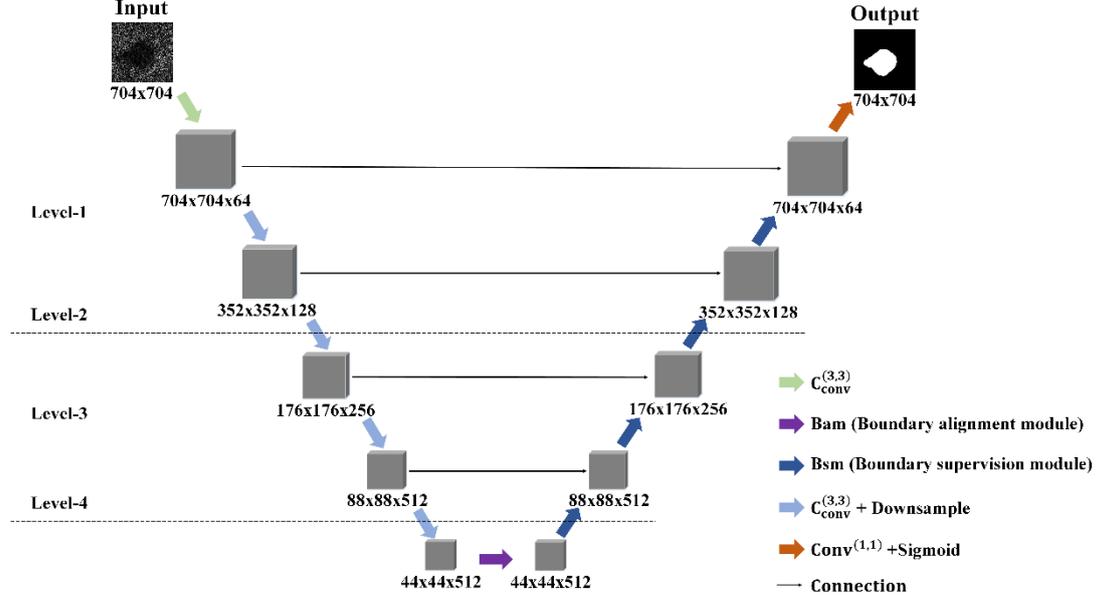


Fig. 2. Graphical representation of the proposed network. The proposed network includes the encoders (left) and the decoders (right). The output of each layer is a three-dimensional feature map of the size ($h \times w \times d$), where h and w are the height and width of the feature map, respectively, and d is the feature dimension.

In the encoder, the basic convolutional group C_{conv} , which consists of convolutional layers, batch normalization (Ioffe & Szegedy, 2015) layers, and activation layers (i.e. ReLu layer or Sigmoid layer), extracts the features step by step and is expressed as:

$$c_{conv}(x) = Act(Bn(Conv(x))) \quad (3)$$

$$C_{conv}(x) = c_{conv}(c_{conv}(x)) \quad (4)$$

where $Conv$ denotes the convolutional layer, Bn denotes the batch normalization layer and Act denotes the activation layer.

The pooling layer was applied to provide feature maps with a smaller size for the extraction of global information. Multi-level encoders were implemented to extract the multi-scale features.

The decoder combined the features extracted by the encoder to reconstruct the segmentation results in the same size as the input image. A boundary alignment module (BAM), as shown in Fig. 3, was implemented at the bottom of the decoder network to extract the global information combined with the encoder. In the BAM, the feature map was initialized using a basic convolutional group with a convolution kernel 1×1 pixel in size to integrate the information from the encoder and decoder network. Two $SConv^{(7,7)}$ paths were used to extract features from the feature maps, and two global feature maps with the same abstract degree were obtained. The subsequent addition operation weighted the important semantic information on the feature maps, while reducing the noise interference. These operations improved the ability of the receptive field to extract global features, while saving computational resources. Finally, the refined boundary was provided through the residual connections. The BAM (Bam) is expressed as:

$$SConv^{(7,7)}(x) = Conv^{(1,7)}(Conv^{(7,1)}(x)) \quad (5)$$

$$DSCConv^{(7,7)}(x) = SConv^{(7,7)}(x) + SConv^{(7,7)}(x) \quad (6)$$

$$c_{DSCConv}^{(7,7)}(x) = Act(Bn(DSCConv^{(7,7)}(x))) \quad (7)$$

$$Bam(x) = x + c_{DSCConv}^{(7,7)}(c_{DSCConv}^{(7,7)}(C_{conv}^{(1,1)}(x))) \quad (8)$$

where $SConv^{(7,7)}$ denotes the asymmetric convolutions with a convolution kernel size of 7×7 and $DSCConv^{(7,7)}$ denotes the additive application of two asymmetric convolutions.

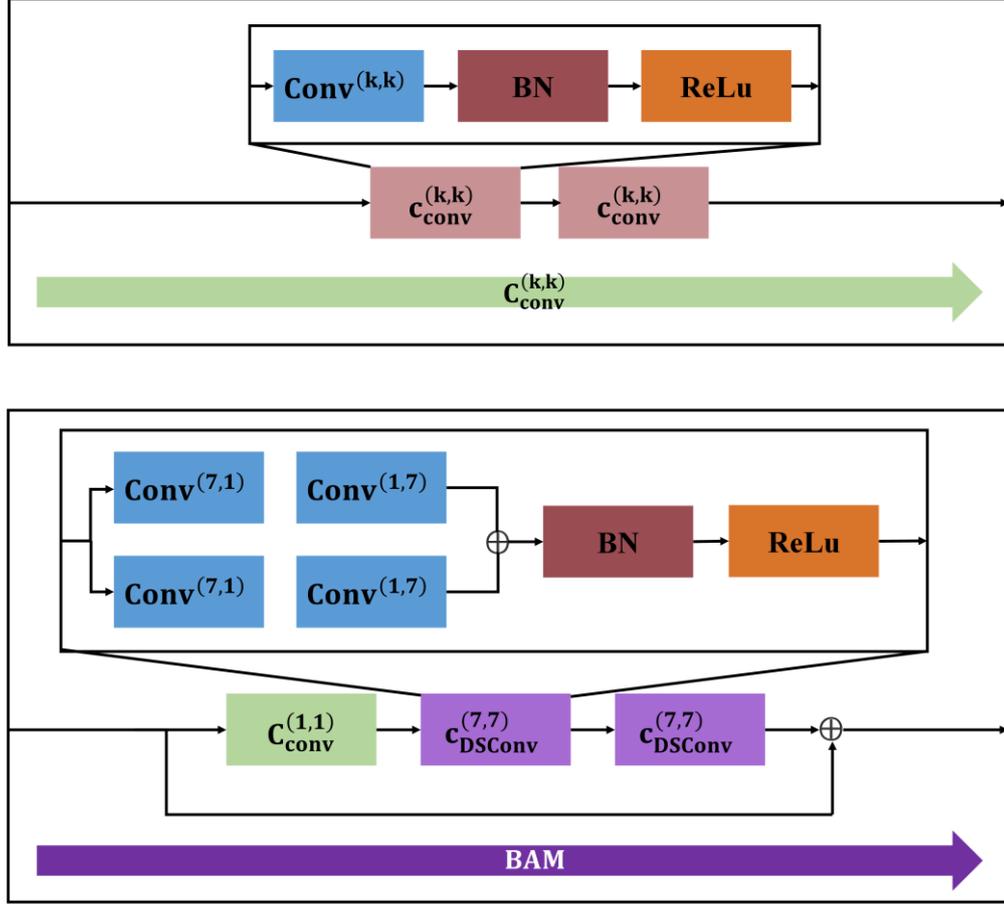


Fig. 3. Graphical representation of the boundary alignment module (BAM).

As with the decoder, the boundary supervision module (BSM), as shown in Fig. 4, used the feature maps of the skip connection in the same scale-level encoder x_s to connect with the feature maps of the previous scale-level decoder x_p . Thus, it performed convolution to extract the global feature maps, which alleviates the information loss that could be caused by applying the pooling layer in the encoder. The branch used for boundary supervision was designed to extract the boundary features and output the boundary segmentation results to calculate the loss and refine the boundary. The convolution of the concatenated global and boundary feature maps and outputting the segmentation results to calculate for the loss for deep supervision contributed to optimizing the segmentation results of the target subject, while refining the target boundary. Multi-scale BSMs were implemented to optimize the segmentation results scale-by-scale with deep supervision and boundary supervision, and to obtain the segmentation result with a fine boundary. The BSM (Bsm) can be expressed as:

$$Gf(x_s, x_p) = C_{conv}^{(3,3)}(C_{conv}^{(3,3)}(x_s) \oplus Up(x_p)) \quad (9)$$

$$Bf(x_s, x_p) = C_{conv}^{(3,3)}(Gf(x_s, x_p)) \quad (10)$$

$$B_{out}(x_s, x_p) = Act(Conv^{(1,1)}(Bf(x_s, x_p))) \quad (11)$$

$$Bsm(x_s, x_p) = C_{conv}^{(3,3)}(Gf(x_s, x_p) \oplus Bf(x_s, x_p)) \quad (12)$$

$$S_{out}(x_s, x_p) = Act(Conv^{(1,1)}(Bsm(x_s, x_p))) \quad (13)$$

where Gf denotes the global feature maps, Bf denotes the boundary feature maps, B_{out} denotes the boundary segmentation result, S_{out} denotes the segmentation result, Up denotes the upsample layer and \oplus denotes the concatenation operation.

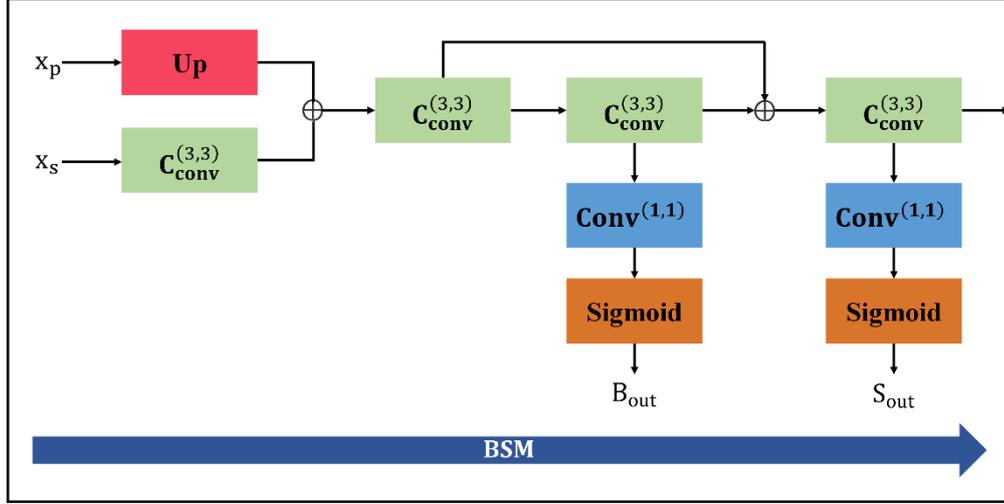


Fig. 4. Graphical representation of the boundary supervision module (BSM).

2.4 Loss Function

Dice loss ($DLoss$) was calculated when the segmentation result (r) was compared against the GT (g), and it is given by:

$$DLoss(r, g) = 1 - (2 * \sum(r * g)) / (\sum r + \sum g) \quad (14)$$

For deep supervision and boundary supervision, the loss function of the network L was designed as follows:

$$L = \lambda_1 * DLoss_S^1 + \lambda_2 * (DLoss_S^2 + DLoss_B^1 + DLoss_B^2) + \lambda_3 * (DLoss_S^3 + DLoss_B^3) + \lambda_4 * (DLoss_S^4 + DLoss_B^4) \quad (15)$$

where $DLoss_y^x$ denotes the dice loss of the x -level segmentation result when y is S or the boundary segmentation result when y is B.

The output results of deep and boundary supervision were used to calculate the loss to optimize the segmentation results and their boundaries. Although the high-level segmentation results participated in the loss calculation, such as 4-level and 3-level, these feature maps extracted from these high-level networks were only able to provide the rough location and boundary of dFAZ due to their small scale. With the decrease of network-level and the increase of the scale of the feature maps, more location information and boundary details of dFAZ were extracted. Therefore, as the network level decreased, λ was set to increase gradually. The ultimate goal was to make the final segmentation result (1-level segmentation result) of the network in order to achieve the best performance, and other output results only played the role of assisting the supervision of network learning. As a result, λ_1 was set to the maximum value. $DLoss_B^1$ was used to supervise the network learning of the boundary of the final segmentation result, and it played an auxiliary role in optimization and supervision. The contribution of $DLoss_B^1$ was considered to be less than that of $DLoss_S^1$, which was similar to that of $DLoss_S^2$ and $DLoss_B^2$. Therefore, the values of $\lambda_1, \lambda_2, \lambda_3$ and λ_4 were set to 7, 3, 2 and 1, respectively, to control the loss weights in the implementation.

3. EXPERIMENTS

3.1 Cross Validation

The proposed method was evaluated using all the 98 images without data augmentations based on cross validation. Generally, statistics take more than 30 samples as large samples. Therefore, in order to ensure

that there are enough images (the number of images is more than 30) in each fold to train and test the network, we used three-fold stratified cross-validation to validate the proposed method. Eighty subjects (45 subjects from the HK PolyU + 35 subjects from the OCTAGON) were evenly distributed into three groups (27 subjects +27 subjects +26 subjects). In each group, the proportion of DR / healthy subjects, as well as the proportion of the HK PolyU / the OCTAGON dataset were the same. At each fold, one group was used as the test set, without repetition, and the other two were used as the training sets. Finally, the segmentation results of the test set of each fold were obtained together as the final segmentation results of the proposed method.

3.2 Network Training

The Adam optimization method (Kingma & Ba, 2014) was applied to minimize the loss of the proposed network and guarantee an efficient calculation. To ensure that the hardware had sufficient memory space, each batch contained only one sample. The learning rate and decay rate were set to 0.0001 and 0.99, respectively. All the weights were initialized using He initialization (He et al., 2015) and the variance of the forward pass output was 1. The network was trained until the accuracy of the test set did not increase within 30 epochs. In the three-fold stratified cross-validation, the network achieved the best generalization performances in the 104-th, 145-th, and 135-th training epochs.

Pytorch (A. Paszke, 2017) and a GPU NVIDIA GeForce GTX 1080TI equipped on an Intel Xeon E5-2650 2.30 GHz machine with a Linux Ubuntu 14.04 operating system were used to train the proposed network.

3.3 Performance evaluation

Dice coefficient (Dice, 1945), precision and recall were used to evaluate the performance of the proposed method. The equation of these metrics are as follows:

$$\text{Dice} = 2TP / (FP + 2TP + FN) \quad (16)$$

$$\text{precision} = TP / (TP + FP) \quad (17)$$

$$\text{recall} = TP / (TP + FN) \quad (18)$$

where TP denotes true positives (correctly predicted dFAZ pixels), FP denotes false positives (incorrectly predicted dFAZ pixels), and FN denotes false negatives (incorrectly rejected dFAZ pixels).

In addition, Hausdorff distance was calculated to evaluate the performance of the proposed method in boundary segmentation. Hausdorff distance H was defined as:

$$H(A, B) = \max(h(A, B), h(B, A)) \quad (19)$$

$$h(A, B) = \max_{a \in A} \min_{b \in B} \|a - b\| \quad (20)$$

where A and B denote two sets of points in a given Euclidean space. $\|\cdot\|$ denotes the distance paradigm (e.g. L2 or Euclidean distance) form between point set A and point set B .

Hausdorff distance was used to measure the degree of maximum mismatch between two point sets. When it was used to measure the degree of dissimilarity between the segmentation result boundary and the GT boundary, the larger the Hausdorff distance was, the higher the dissimilarity between the segmentation result boundary and the GT boundary, and the lower the accuracy of boundary segmentation.

4. RESULTS AND ANALYSIS

The performance of the proposed method was compared with other deep-learning networks. Table 2 shows the comparison of the segmentation performance in terms of the mean Dice coefficient, precision,

recall and Hausdorff distance between our method and other methods in all OCTA images (HK PolyU dataset and OCTAGON dataset). The mean Dice coefficient of our method with BAM and BSM (0.86) was higher than those of the other methods (0.79 in FCN (Long et al., 2015), 0.78 in SegNet (Badrinarayanan et al., 2017), 0.84 in U-Net (Ronneberger et al., 2015), 0.85 in OCNet (Yuan & Wang, 2018), 0.81 in DFANet (H. Li et al., 2019)). This indicates a superior performance and better accuracy in the dFAZ segmentation of the proposed method. In addition, BAM and BSM with our method achieved the highest precision and recall of 0.94 and 0.82, respectively. We used SPSS20.0 to analyze Dice of the proposed method and Dice of other methods. We used paired sample t test to test whether Dice of the proposed method was significantly different from that of other methods. The statistical results were shown in Table 3. Table 3 showed that the proposed method was significantly different from other methods in Dice. This proves that the proposed method achieved a superior performance compared to other methods. The Hausdorff distance of the proposed method with BAM and BSM (23.35) was lower than those of the other methods (284.93 in FCN, 282.54 in SegNet, 165.58 in U-Net, 65.99 in OCNet, 120.54 in DFANet). This elaborated the proposed method had the best performance in boundary segmentation compared with other methods.

Table 2. Comparison of The Mean (Standard Deviation) of Dice Coefficient, Precision, Recall and Hausdorff distance of The Proposed Method Against Other Methods

	Dice	Precision	Recall	Hausdorff distance
FCN32	0.79(0.15)	0.94(0.12)	0.72(0.19)	284.93(154.14)
SegNet	0.78(0.14)	0.92(0.13)	0.69(0.18)	282.54(146.85)
U-Net	0.84(0.12)	0.93(0.11)	0.79(0.15)	165.58(124.33)
OCNet	0.85(0.12)	0.92(0.14)	0.81(0.15)	65.35(119.74)
DFANet	0.81(0.12)	0.93(0.14)	0.74(0.15)	120.54(147.76)
Proposed method with BAM	0.84(0.11)	0.96(0.09)	0.76(0.14)	57.55(105.44)
Proposed method with BSM	0.85(0.12)	0.95(0.10)	0.78(0.14)	31.46(68.31)
Proposed method with BAM, BSM	0.86(0.10)	0.95(0.10)	0.81(0.13)	23.35(53.86)

Table 3. Paired Sample T-Test Results of Dice of The Proposed Method and Dice of Other Methods

	p value
FCN32	0.000(<0.05)
SegNet	0.000(<0.05)
U-Net	0.000(<0.05)
OCNet	0.005(<0.05)
DFANet	0.000(<0.05)

A comparison of the segmentation results of the proposed method and other methods is shown in Fig. 5. The proposed method provided a more accurate and smoother boundary while accurately locating the target subject. In contrast, different degrees of glitch appeared at the segmentation boundary of other methods.

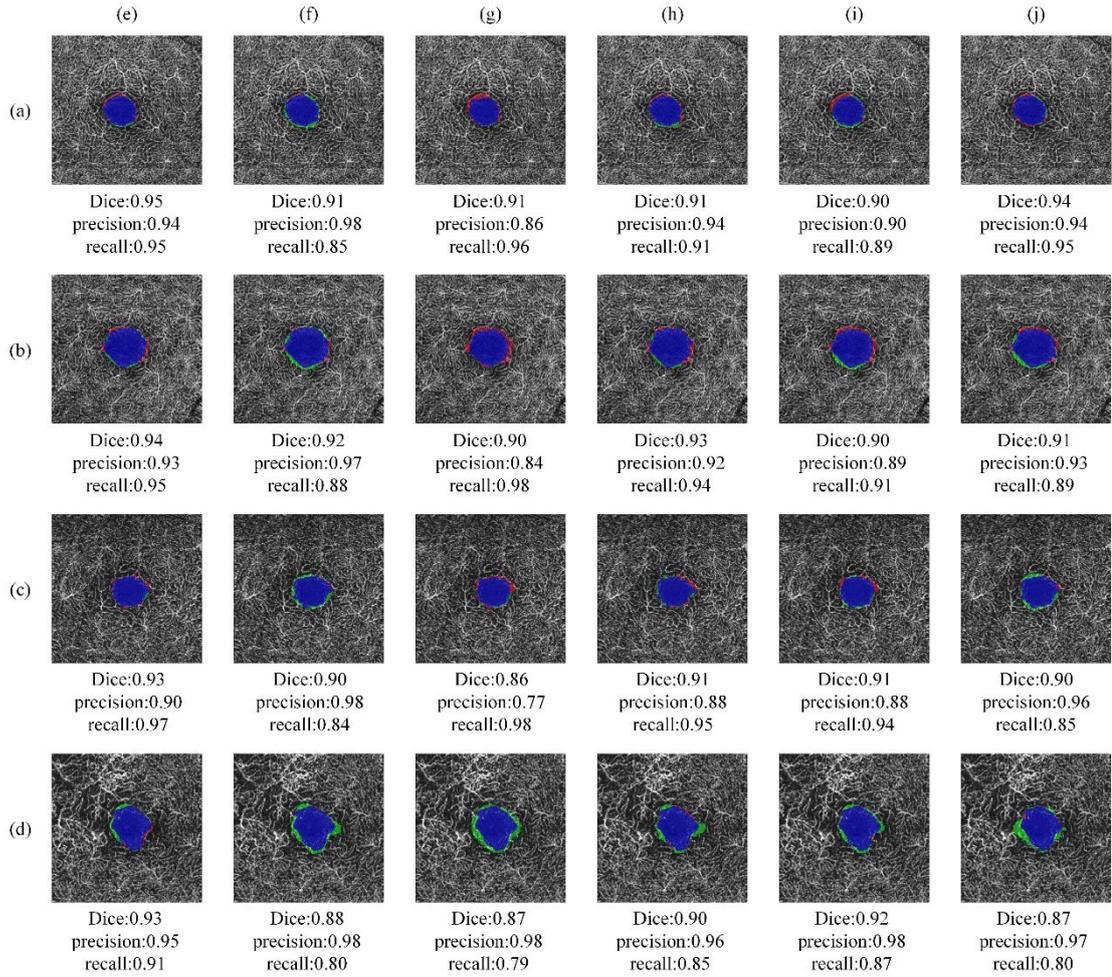


Fig. 5. Segmentation results of the proposed method and other methods against the ground truth. The green area represents true positives, the blue area represents false negatives and the red area represents false positives. (a) and (b) rows were examples of healthy subjects, while (c) and (d) rows were examples of DR subjects. (e) column: Segmentation result of the proposed method. (f) column: Segmentation result of FCN32. (g) column: Segmentation result of SegNet. (h) column: Segmentation result of U-Net. (i) column: Segmentation result of OCNet. (j) column: Segmentation result of DFANet. The respective values of Dice, precision and recall were placed at the bottom of the image.

Table 4 shows the performance of the proposed method on individual datasets. The OCTAGON dataset was added for training and validation to expand the amount and diversity of the data, and further, to improve the stability and generality of the proposed method. The proposed method achieved a mean Dice Coefficient of 0.89 on the HK PolyU dataset which was slightly better than that of the OCTAGON dataset (0.84).

Table 4. Mean (Standard Deviation) of Dice Coefficient, Precision and Recall of The Proposed Method on Different Datasets

	Dice	Precision	Recall
HK Polytechnic University dataset	0.89(0.06)	0.97(0.09)	0.84(0.08)
OCTAGON dataset	0.84(0.13)	0.91(0.13)	0.81(0.14)

The dFAZ of the diabetic retinopathy (DR) subjects was usually severely destroyed, with hardly distinguishable boundary in the noisy OCTA image as shown in Fig. 6. Therefore, led to a smaller Dice coefficient on DR comparing with healthy subjects, as shown in Table 5.

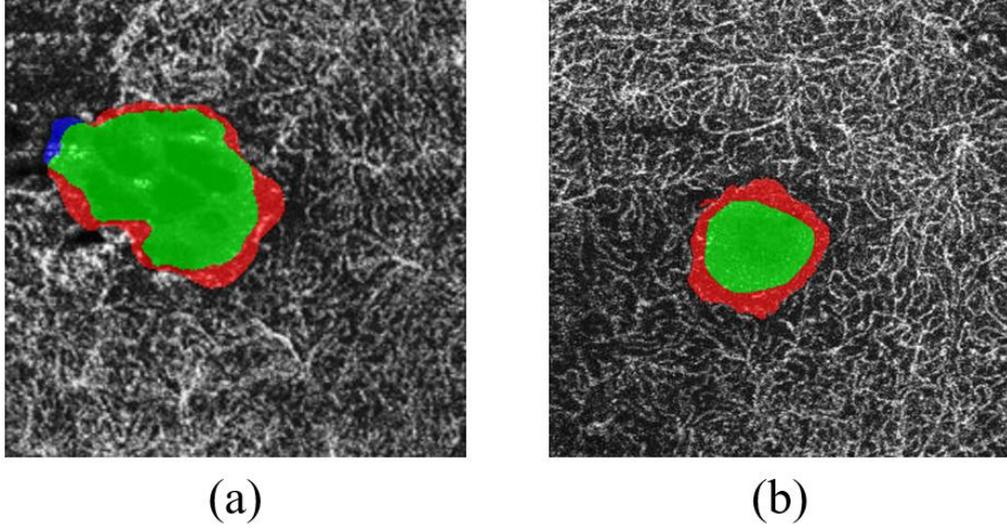


Fig. 6. Segmentation results of the proposed method on two diabetic retinopathy subjects (a and b) from the OCTAGON dataset against the ground truth. The green area represents true positives, the blue area represents false negatives and the red area represents false positives.

Table 5. Mean (Standard Deviation) of Dice Coefficient, Precision and Recall of The Proposed Method and Other Methods on Healthy Subjects and DR Subjects

	Healthy subjects			DR subjects		
	Dice	Precision	Recall	Dice	Precision	Recall
FCN32	0.82(0.12)	0.95(0.07)	0.75(0.18)	0.65(0.19)	0.90(0.22)	0.56(0.19)
SegNet	0.80(0.11)	0.94(0.09)	0.71(0.16)	0.67(0.23)	0.84(0.23)	0.60(0.26)
U-Net	0.86(0.07)	0.95(0.07)	0.81(0.12)	0.72(0.21)	0.88(0.21)	0.68(0.20)
OCNet	0.87(0.06)	0.93(0.09)	0.84(0.10)	0.72(0.22)	0.87(0.25)	0.66(0.25)
DFANet	0.83(0.06)	0.95(0.09)	0.75(0.12)	0.71(0.22)	0.87(0.25)	0.64(0.24)
Proposed method with BAM, BSM	0.89(0.05)	0.96(0.05)	0.83(0.09)	0.74(0.18)	0.88(0.19)	0.70(0.21)

5. DISCUSSION

Accurate segmentation of the dFAZ is of great clinical importance and represents a major research interest. Quantifying dFAZ remains very challenging due to the less well-defined boundary, which is sensitive to the projection artifacts and signal attenuation. Current solutions for dFAZ segmentation, including built-in software, customized programs using conventional images processing techniques, and manual delineation may not be satisfactory. Therefore, this study proposed a deep-learning solution. Instead of directly applying the classic deep learning networks for image segmentation, such as FCN, SegNet, U-Net, OCNet and DFANet, BAM and BSM were implemented to further improve the performance. This proposed method obtained a mean Dice coefficient of 0.86 in three-fold stratified cross-validation, which indicates a superior performance in dFAZ segmentation while allowing the

independence of empirical parameters that were manually set. This makes the proposed method less sensitive to the position, size, and shape of dFAZ.

The proposed method has been compared with three classic networks (FCN, SegNet, and U-Net) and two newly proposed methods (OCNet and DFANet). U-Net exhibits strong generalization ability and excellent performance in medical image segmentation. A mean Dice coefficient of 0.84 was obtained from our datasets by applying U-Net directly, which was better than 0.79 in FCN, 0.78 in SegNet, and 0.81 in DFANet. Though OCNet was the only one slightly outperformed U-Net, it was worse than the network we proposed in terms of Dice coefficient.

As shown in Fig. 5, the proposed method accurately located the dFAZ and provided a naturally smooth dFAZ boundary. This good performance may be attributable to the implementation of the BAM. The asymmetric convolutions in BAM, inspired by Inception V3 (Szegedy et al., 2016), obtained rich global information from the feature maps output of the encoder with a significantly improved receptive field. Although the stack of convolutions of a small convolution kernel theoretically provided the same receptive field as the convolution with a large convolution kernel, it actually caused the attenuation of the edge receptive field, subsequently causing the effective receptive field to be smaller than the theoretical receptive field. However, the direct use of a convolution with a large convolution kernel would inevitably increase the computational cost. The asymmetric convolution was thus a compromise with a reasonable computational cost and an improved effective receptive field. The residual branch was designed to refine the boundary of the segmentation result. The combination of the asymmetric convolution and the residual branch significantly improved the effective receptive field and provided rich global information for refining the boundary of the segmentation result.

The second explanation for the good performance might be the BSM applied in the current method. In BSM, the basic convolutional group was applied to boost the ability of the skip connection structure to extract features. The designed boundary branch outputted the boundary result and calculated the boundary loss, providing not only boundary features, but also performed boundary supervision. As the layers deepen, low-scale deep networks may struggle to obtain adequate training. In order to achieve full training on all scale-level networks, the segmentation result to calculate the loss in the BSM of each scale-level was outputted in a deep supervision manner, thus providing the proposed network with better ability to extract features.

The proposed method achieved a mean Dice coefficient of 0.88 and 0.76 in healthy subjects and DR subjects, respectively. Despite the smaller number of OCTA images of the DR subjects in this study and the irregular nature of dFAZ in these images, the output indicated that the proposed method could provide accurate dFAZ segmentation and quantification. Further studies on evaluating the robustness of this method by recruiting DR patients with different severities would be needed to establish a more customized network structure.

Deep learning is an emerging technology with potential applications in optometry and ophthalmology for improving diagnosis and management of eye diseases. This is the first study using deep learning techniques to deal with the challenges of dFAZ segmentation and quantification. The proposed method provides an objective, repeatable, and reliable tool for dFAZ segmentation and quantification. It is hoped the future application of this method could both save clinician time and subsequently boost the investigations in monitoring dFAZ changes in various ocular conditions.

6. CONCLUSION

In this paper, a method using a customized encoder-decoder network was proposed to automatically segment dFAZ in OCTA images. The application of BAM and BSM improved the accuracy in locating

the dFAZ and produced output segmentation results with smoother boundaries (on a dataset from a total of 80 subjects, an average Dice of 0.86 and an average Hausdorff distance of 23.35 is achieved). The proposed method is expected to be able to equip clinicians with an automated and more generalized method and be helpful to handle more dFAZ-related research with OCTA data.

Funding

This research did not receive any specific grant from funding agencies in the public, commercial, or not-for-profit sectors.

Disclosures

The authors declare no conflicts of interest.

References:

- A. Paszke, S. G., S. Chintala and G. Chanan. (2017). Pytorch [Computer software].
- Badrinarayanan, V., Kendall, A., & Cipolla, R. (2017). Segnet: A deep convolutional encoder-decoder architecture for image segmentation. *IEEE transactions on pattern analysis and machine intelligence*, 39(12), 2481-2495.
doi:10.1109/TPAMI.2016.2644615
- Cheng, D., Chen, Q., Wu, Y. F., Yu, X. T., Shen, M. X., Zhuang, X. R., . . . Shen, L. J. (2019). Deep perifoveal vessel density as an indicator of capillary loss in high myopia. *Eye*, 33(12), 1961-1968. doi:10.1038/s41433-019-0573-1
- Chun, L. Y., Silas, M. R., Dimitroyannis, R. C., Ho, K., & Skondra, D. (2019). Differences in macular capillary parameters between healthy black and white subjects with Optical Coherence Tomography Angiography (OCTA). *Plos One*, 14(10).
doi:10.1371/journal.pone.0223142
- Corvi, F., Pellegrini, M., Erba, S., Cozzi, M., Staurengi, G., & Giani, A. (2018). Reproducibility of Vessel Density, Fractal Dimension, and Foveal Avascular Zone Using 7 Different Optical Coherence Tomography Angiography Devices REPLY. *Am J Ophthalmol*, 192, 253-255. doi:DOI 10.1016/j.ajo.2018.05.019
- Coscas, F., Glacet-Bernard, A., Miere, A., Caillaux, V., Uzzan, J., Lupidi, M., . . . Souied, E. H. (2016). Optical Coherence Tomography Angiography in Retinal Vein Occlusion: Evaluation of Superficial and Deep Capillary Plexa. *Am J Ophthalmol*, 161, 160-171. doi:10.1016/j.ajo.2015.10.008
- D íaz M, N. J., Ortega M, Penedo M, Gómez-Ulla F. . (2018). *OCTAGON (dataset)*.
- Diaz, M., Novo, J., Cutrin, P., Gomez-Ulla, F., Penedo, M. G., & Ortega, M. (2019). Automatic segmentation of the foveal avascular zone in ophthalmological OCT-A images. *Plos One*, 14(2). doi:10.1371/journal.pone.0212364
- Dice, L. R. (1945). Measures of the amount of ecologic association between species. *Ecology*, 26(3), 297-302.
doi:10.2307/1932409
- Dimitrova, G., & Chihara, E. (2019). Implication of Deep-Vascular-Layer Alteration Detected by Optical Coherence Tomography Angiography for the Pathogenesis of Diabetic Retinopathy. *Ophthalmologica*, 241(4), 179-182.
doi:10.1159/000495624

- Enders, C., Baeuerle, F., Lang, G., Dreyhaupt, J., Lang, G., Loidl, M., & Werner, J. (2020). Comparison between Findings in Optical Coherence Tomography Angiography and in Fluorescein Angiography in Patients with Diabetic Retinopathy. *Ophthalmologica*, 243(1), 21-26. doi:10.1159/000499114
- Falavarjani, K. G., Shenazandi, H., Naseri, D., Anvari, P., Kazemi, P., Aghamohammadi, F., . . . Alemzadeh, S. A. (2018). Foveal Avascular Zone and Vessel Density in Healthy Subjects: An Optical Coherence Tomography Angiography Study. *J Ophthal Vis Res*, 13(3), 260-265. doi:10.4103/jovr.jovr_173_17
- Guo, M., Zhao, M., Cheong, A. M. Y., Dai, H., Lam, A. K. C., & Zhou, Y. (2019). Automatic quantification of superficial foveal avascular zone in optical coherence tomography angiography implemented with deep learning. *Vis Comput Ind Biomed Art*, 2(1), 21. doi:10.1186/s42492-019-0031-8
- Hayreh, S. S. (2015). The Retinal Capillaries. In *Ocular Vascular Occlusive Disorders* (pp. 153-164). Cham: Springer International Publishing.
- He, K., Zhang, X., Ren, S., & Sun, J. (2015). *Delving deep into rectifiers: Surpassing human-level performance on imagenet classification*. Paper presented at the Proceedings of the IEEE international conference on computer vision.
- Ioffe, S., & Szegedy, C. (2015). Batch normalization: Accelerating deep network training by reducing internal covariate shift. *arXiv preprint arXiv:1502.03167*.
- Kingma, D. P., & Ba, J. (2014). Adam: A method for stochastic optimization. *arXiv preprint arXiv:1412.6980*.
- Kuehlewein, L., Tepelus, T. C., An, L., Durbin, M. K., Srinivas, S., & Sadda, S. R. (2015). Noninvasive Visualization and Analysis of the Human Parafoveal Capillary Network Using Swept Source OCT Optical Microangiography. *Invest Ophth Vis Sci*, 56(6), 3984-3988. doi:10.1167/iovs.15-16510
- LeCun, Y., Bengio, Y., & Hinton, G. (2015). Deep learning. *Nature*, 521(7553), 436-444. doi:10.1038/nature14539
- Li, H., Xiong, P., Fan, H., & Sun, J. (2019). *Dfanet: Deep feature aggregation for real-time semantic segmentation*. Paper presented at the Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition.
- Li, Y. F., Say, E. A. T., Ferenczy, S., Agni, M., & Shields, C. L. (2017). Altered Parafoveal Microvasculature in Treatment-Naive Choroidal Melanoma Eyes Detected by Optical Coherence Tomography Angiography. *Retina-J Ret Vit Dis*, 37(1), 32-40. doi:10.1097/iae.0000000000001242
- Lin, A. D., Fang, D. Q., Li, C. L., Cheung, C. Y., & Chen, H. Y. (2020). Reliability of foveal avascular zone metrics automatically measured by Cirrus optical coherence tomography angiography in healthy subjects. *Int Ophthalmol*, 40(3), 763-773. doi:10.1007/s10792-019-01238-x
- Linderman, R., Salmon, A. E., Strampe, M., Russillo, M., Khan, J., & Carroll, J. (2017). Assessing the accuracy of foveal avascular zone measurements using optical coherence tomography angiography: segmentation and scaling (vol 6, 16, 2017). *Transl Vis Sci Techn*, 6(4). doi:10.1167/tvst.6.4.17
- Long, J., Shelhamer, E., & Darrell, T. (2015). *Fully convolutional networks for semantic segmentation*. Paper presented at the Proceedings of the IEEE conference on computer vision and pattern recognition.

- Lu, Y. S., Simonett, J. M., Wang, J., Zhang, M., Hwang, T., Hagag, A. M., . . . Jia, Y. L. (2018). Evaluation of Automatically Quantified Foveal Avascular Zone Metrics for Diagnosis of Diabetic Retinopathy Using Optical Coherence Tomography Angiography. *Invest Ophthalmol Vis Sci*, 59(6), 2212-2221. doi:10.1167/iovs.17-23498
- Lupidi, M., Coscas, F., Cagini, C., Fiore, T., Spaccini, E., Fruttini, D., & Coscas, G. (2016). Automated Quantitative Analysis of Retinal Microvasculature in Normal Eyes on Optical Coherence Tomography Angiography. *Am J Ophthalmol*, 169, 9-23. doi:10.1016/j.ajo.2016.06.008
- Magrath, G. N., Say, E. A. T., Sioufi, K., Ferenczy, S., Samara, W. A., & Shields, C. L. (2017). Variability in Foveal Avascular Zone and Capillary Density Using Optical Coherence Tomography Angiography Machines in Healthy Eyes. *Retina-J Ret Vit Dis*, 37(11), 2102-2111. doi:10.1097/IAE.0000000000001458
- Mammo, Z., Balaratnasingam, C., Yu, P. L., Xu, J., Heisler, M., Mackenzie, P., . . . Yu, D. Y. (2015). Quantitative Noninvasive Angiography of the Fovea Centralis Using Speckle Variance Optical Coherence Tomography. *Invest Ophthalmol Vis Sci*, 56(9), 5074-5086. doi:10.1167/iovs.15-16773
- Marmor, M. F., & Ravin, J. G. (2011). Fluorescein Angiography Insight and Serendipity a Half Century Ago. *Arch Ophthalmol-Chic*, 129(7), 943-948. doi:10.1001/archophthalmol.2011.160
- Matsunaga, D. R., Jack, J. Y., De Koo, L. O., Ameri, H., Puliafito, C. A., & Kashani, A. H. (2015). Optical coherence tomography angiography of diabetic retinopathy in human subjects. *Ophthalmic Surgery, Lasers and Imaging Retina*, 46(8), 796-805. doi:10.3928/23258160-20150909-03
- Mendis, K. R., Balaratnasingam, C., Yu, P., Barry, C. J., McAllister, I. L., Cringle, S. J., & Yu, D. Y. (2010). Correlation of Histologic and Clinical Images to Determine the Diagnostic Value of Fluorescein Angiography for Studying Retinal Capillary Detail. *Invest Ophthalmol Vis Sci*, 51(11), 5864-5869. doi:10.1167/iovs.10-5333
- Park, S. W., & Kang, S. J. (2018). Association between Optic Nerve Head Deformation and Retinal Microvasculature in High Myopia. *Invest Ophthalmol Vis Sci*, 59(9). doi:10.1016/j.ajo.2018.01.033
- Pilotto, E., Frizziero, L., Crepaldi, A., Della Dora, E., Deganello, D., Longhin, E., . . . Midena, E. (2018). Repeatability and Reproducibility of Foveal Avascular Zone Area Measurement on Normal Eyes by Different Optical Coherence Tomography Angiography Instruments. *Ophthalmic Res*, 59(4), 206-211. doi:10.1159/000485463
- Provis, J. M. (2001). Development of the Primate Retinal Vasculature. *Progress in Retinal and Eye Research*, 20(6), 799-821. doi:10.1016/s1350-9462(01)00012-x
- Provis, J. M., Dubis, A. M., Maddess, T., & Carroll, J. (2013). Adaptation of the central retina for high acuity vision: cones, the fovea and the avascular zone. *Prog Retin Eye Res*, 35, 63-81. doi:10.1016/j.preteyeres.2013.01.005
- Romano, F., Arrigo, A., Ch'ng, S. W., Parodi, M. B., Manitto, M. P., Martina, E., . . . Stanga, P. E. (2019). Capillary Network Alterations in X-Linked Retinoschisis Imaged on Optical Coherence Tomography Angiography. *Retina-J Ret Vit Dis*, 39(9), 1761-1767. doi:10.1097/iae.0000000000002222
- Ronneberger, O., Fischer, P., & Brox, T. (2015). *U-net: Convolutional networks for biomedical image segmentation*. Paper presented at the International Conference on Medical image computing and computer-assisted intervention.

- Samara, W. A., Shahlaee, A., Adam, M. K., Khan, A., Chiang, A., Maguire, J. I., . . . Ho, A. C. (2017). Quantification of Diabetic Macular Ischemia Using Optical Coherence Tomography Angiography and Its Relationship with Visual Acuity. *Ophthalmology*, *124*(2), 235-244. doi:10.1016/j.ophtha.2016.10.008
- Shahlaee, A., Pefkianaki, M., Hsu, J., & Ho, A. C. (2016). Measurement of Foveal Avascular Zone Dimensions and its Reliability in Healthy Eyes Using Optical Coherence Tomography Angiography. *Am J Ophthalmol*, *161*, 50-55. doi:10.1016/j.ajo.2015.09.026
- Spaide, R. F., Klancnik, J., & Cooney, M. J. (2015a). Retinal Vascular Layers Imaged by Fluorescein Angiography and Optical Coherence Tomography Angiography. *Jama Ophthalmol*, *133*(1), 45-50. doi:10.1001/jamaophthalmol.2014.3616
- Spaide, R. F., Klancnik, J. M., & Cooney, M. J. (2015b). Retinal Vascular Layers in Macular Telangiectasia Type 2 Imaged by Optical Coherence Tomographic Angiography. *Jama Ophthalmol*, *133*(1), 66-73. doi:10.1001/jamaophthalmol.2014.3950
- Springer, A. D., & Hendrickson, A. E. (2004). Development of the primate area of high acuity. 1. Use of finite element analysis models to identify mechanical variables affecting pit formation. *Visual Neurosci*, *21*(1), 53-62. doi:10.1017/S0952523804041057
- Szegedy, C., Vanhoucke, V., Ioffe, S., Shlens, J., & Wojna, Z. (2016). *Rethinking the inception architecture for computer vision*. Paper presented at the Proceedings of the IEEE conference on computer vision and pattern recognition.
- Takase, N., Nozaki, M., Kato, A., Ozeki, H., Yoshida, M., & Ogura, Y. (2015). Enlargement of foveal avascular zone in diabetic eyes evaluated by en face optical coherence tomography angiography. *Retina*, *35*(11), 2377-2383. doi:10.1097/IAE.0000000000000849
- Tan, P. E. Z., Balaratnasingam, C., Xu, J., Mammo, Z., Han, S. X., Mackenzie, P., . . . Yu, D. Y. (2015). Quantitative Comparison of Retinal Capillary Images Derived By Speckle Variance Optical Coherence Tomography With Histology. *Invest Ophthalm Vis Sci*, *56*(6), 3989-3996. doi:10.1167/iovs.14-15879
- Tang, F. Y., Chan, E. O., Sun, Z. H., Wong, R., Lok, J., Szeto, S., . . . Cheung, C. Y. (2020). Clinically relevant factors associated with quantitative optical coherence tomography angiography metrics in deep capillary plexus in patients with diabetes. *Eye Vision*, *7*(1). doi:10.1186/s40662-019-0173-y
- Tang, F. Y., Ng, D. S., Lam, A., Luk, F., Wong, R., Chan, C., . . . Cheung, C. Y. (2017). Determinants of Quantitative Optical Coherence Tomography Angiography Metrics in Patients with Diabetes. *Scientific Reports*, *7*(1), 2575. doi:10.1038/s41598-017-02767-0
- Tick, S., Rossant, F., Ghorbel, I., Gaudric, A., Sahel, J. A., Chaumet-Riffaud, P., & Paques, M. (2011). Foveal Shape and Structure in a Normal Population. *Invest Ophthalm Vis Sci*, *52*(8), 5105-5110. doi:10.1167/iovs.10-7005
- Villegas, V. M., & Kovach, J. L. (2017). Optical Coherence Tomography Angiography of Macular Telangiectasia Type 2 with Associated Subretinal Neovascular Membrane. *Case Rep Ophthalmol*. doi:10.1155/2017/8186134
- Waizel, M., Todorova, M. G., Terrada, C., LeHoang, P., Massamba, N., & Bodaghi, B. (2018). Superficial and deep retinal foveal avascular zone OCTA findings of non-infectious anterior and posterior uveitis. *Graef Arch Clin Exp*, *256*(10), 1977-1984. doi:10.1007/s00417-018-4057-y

- Werner, J. U., Bohm, F., Lang, G. E., Dreyhaupt, J., Lang, G. K., & Enders, C. (2019). Comparison of foveal avascular zone between optical coherence tomography angiography and fluorescein angiography in patients with retinal vein occlusion. *Plos One*, *14*(6). doi:10.1371/journal.pone.0217849
- Wylegala, A., Wang, L. F., Zhang, S. J., Liu, Z. Q., Teper, S., & Wylegala, E. (2019). Comparison of foveal avascular zone and retinal vascular density in healthy Chinese and Caucasian adults. *Acta Ophthalmol*. doi:10.1111/aos.14316
- Yoon, Y. S., Woo, J. E., & Woo, J. M. (2017). Optical Coherence Tomography Angiography According to Severity of Diabetic Retinopathy. *J Kor Ophthalmol Soc*, *58*(6), 653-662. doi:10.3341/jkos.2017.58.6.653
- Yuan, Y., & Wang, J. (2018). Ocnet: Object context network for scene parsing. *arXiv preprint arXiv:1809.00916*.

*Credit Author Statement

YONGJIN ZHOU and ANDREW KC LAM conceived of the presented idea. MENGLIN GUO and MEI ZHAO developed the theory and performed the computations. ALLEN MY CHEONG and CORVI FEDERICO verified the analytical methods. All authors discussed the results and contributed to the final manuscript