The Journal of Supercomputing (2025) 81:535
https://doi.org/10.1007/s11227-025-06985-6

SMFDNet: spatial and multi-frequency domain network for OCT angiography retinal vessel segmentation

Sien Li[1] · Fei Ma[1] · Fen Yan[2] · Jing Meng[1] · Yanfei Guo[1] · Hongjuan Liu[1] · Ronghua Cheng[1]

Accepted: 25 January 2025

Abstract

Optical coherence tomography angiography (OCTA) is a non-invasive imaging technique, and automatic segmentation of retinal vessels is crucial for understanding ocular diseases and making informed clinical decisions. However, the automatic segmentation of retinal vessels in OCTA images is particularly challenging due to several inherent issues. Retinal vessels often exhibit low contrast against the surrounding tissue, making it difficult to distinguish them clearly. Additionally, the complex and irregular branching structures of retinal vessels, along with the presence of noise and artefacts in OCTA images, further complicate the segmentation task. To address these challenges, we propose a novel method, the spatial and multi-frequency domain-based segmentation network (SMFDNet), specifically designed for vessel segmentation in OCTA fundus images. This network effectively combines spatial and multi-frequency domain features to enhance the segmentation accuracy of retinal vessels. To demonstrate the superiority of our proposed network, we conduct experiments on the Retinal Vessels Images in OCTA (REVIO), Retinal OCT-Angiography Vessel Segmentation (ROSE) and Optical Coherence Tomography Angiography-500 (OCTA-500) datasets. The extensive experimental results show that our approach consistently outperforms state-of-the-art methods, particularly in handling low contrast and complex vessel structures. The codes are available at https://kyanbis.github.io/SMFDNet/.

Keywords Segmentation · Deep learning · OCT angiography · Retinal image segmentation

✉ Fei Ma
mafei0603@163.com

[1] School of Computer Science, Qufu Normal University, Rizhao, China

[2] Ultrasound Medicine Department Qufu People's Hospital, Qufu, China

Published online: 24 February 2025

1 Introduction

Optical coherence tomography angiography (OCTA) is a non-invasive imaging technique that uses optical principles and data processing techniques to provide high-resolution vascular structure in biological tissues without the contrast agents. With the continuous advancement of medical imaging technology [1], the application of OCTA in clinical medicine has gained widespread attention [2], especially in the fields of ophthalmology, cardiovascular science and neuroscience.

Figure 1 illustrates the REVIO [3], ROSE [4] and OCTA-500 [5] datasets. OCTA has high resolution, high contrast and high sensitivity to clearly visualize the distribution and structure of the microvascular network. OCTA allows observation of microvascular detail down to the capillary level, allowing quantitative assessment of retinal microvascular and morphovascular outcomes.

Recently, OCTA image segmentation still faces several challenges. First, OCTA images tend to suffer from low signal-to-noise ratios and image artefacts, such as motion and projection artefacts. These problems can adversely affect the accuracy of segmentation. Secondly, the complex and intricate nature of retinal blood vessels, such as variations in vessel diameter, branching patterns and vessel tortuosity, presents difficulties for automatic segmentation tasks. In addition, OCTA images from different patients and different imaging devices can vary in image contrast and brightness [6, 7], which will be more difficult for the segmentation in OCTA images.

Recently, there exist many classical deep learning [8, 9] networks such as UNet [9], convolutional neural network (CNN) [10] and recurrent neural network (RNN) [11]. These methods can automatically extract discriminative features and

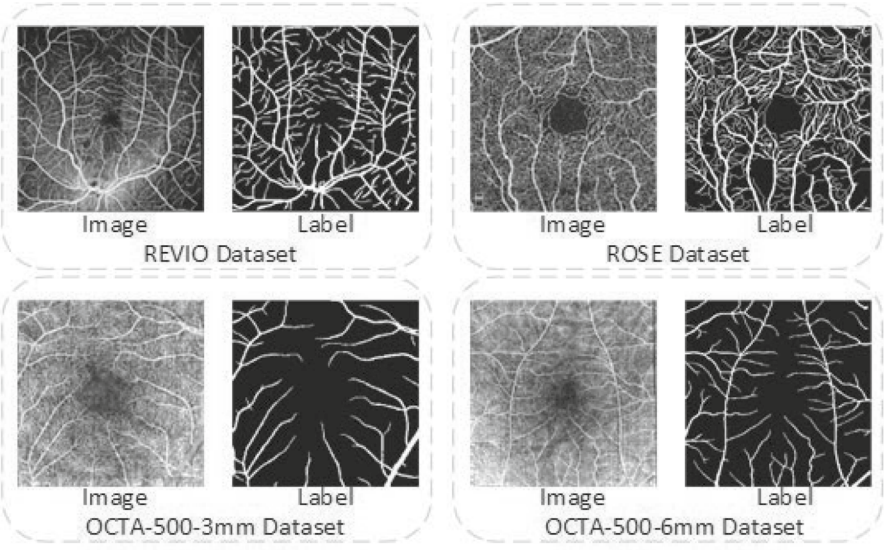

Fig. 1 REVIO, ROSE, OCTA-500-3 mm and OCTA-500-6 mm datasets corresponding to OCTA images and ground truth

adapt to changes in image appearance and structure. Deep learning frameworks can be conducted by the end-to-end way [12] and improve the generalization of models.

OCTA [13–15] imaging technology can produce images with a low signal-to-noise ratio. In addition, the appearances of different vessels at different depths and the underlying pathology increase the challenge of achieving accurate segmentation, especially for densely connected capillaries, which can easily lead to segmentation discontinuities.

OCTA images typically show complex structures and multilayered features, including different layers of the retina, the retinal nerve fibre layer, etc. Spatial domain [16, 17] can be used to segment the above layers and obtain effective feature vectors.

OCTA images have rich spectral information in the frequency domain, reflecting different optical properties and reflectivity. By analysing OCTA images in the frequency domain [18–20], the spectral features of tissues can be extracted to achieve segmentation and identification of different tissues and rich feature vectors can be obtained, making segmentation more accurate. By extracting the microvascular structures of the different OCTA depth layers, their corresponding surface projections can be obtained and their respective variations can be analysed.

The automatic vessel detection [21, 22] in OCTA images is valuable for early diagnosis of vascular diseases affecting the retinal circulation and assessment of disease progression. Despite the great success of many deep learning-based medical segmentation in colour images, automatic vascular segmentation in OCTA images has rarely been explored and remained a challenging task.

1.1 Motivation

Optical coherence tomography angiography (OCTA) is a non-invasive imaging technique used to visualize microvascular structures, especially in retinal and choroidal layers. Many existing methods have difficulty dealing with background noise, which can obscure subtle vascular structures and reduce segmentation accuracy. To address these challenges, we propose a novel method, the spatial and multi-frequency domain-based segmentation network (SMFDNet), specifically designed for vessel segmentation in OCTA fundus images. In the spatial domain, we focus on processing the image at the pixel level, working to improve the accuracy of tasks such as image quality and edge detection. And in the frequency domain, we analyse the frequency components of the signal by transforming it into the frequency domain using techniques such as wavelet transform. An intuitive solution is to fuse frequency domain features with the spatial domain to obtain different high and low-frequency features of the image to enhance the global feature information.

The contributions of this work can be summarized as follows:

(1) We propose a novel spatial domain and multi-frequency domain network (SMFDNet) for vessel segmentation in OCTA fundus images.

(2) In this work, the features are extracted and fused by using spatial and frequency domains. The regional location and frequency characteristics of the target are detected in the spatial and frequency domains, respectively.
(3) The proposed segmentation network is extensively evaluated quantitatively on three OCTA fundus datasets, including REVIO, ROSE and OCTA-500. The extensive experiments show that the proposed approach outperforms existing methods on three OCTA fundus datasets.

The remainder of this paper is organized as follows: The information of the existing OCTA datasets is described in Sect. 3. Section 4 introduces the details of the proposed framework. The experimental results are detailed in Sect. 5. Section 6 concludes the work and the future work.

2 Related work

In this section, we will review and discuss the most relevant work on blood vessel segmentation. Recently, there exist plenty of impressive works for retinal vessel segmentation in colour fundus images. The OCTA imaging modality allows the visualization of micro-vessels and has been proved an advantageous tool for observing the retinal vasculature. Vessel segmentation in OCTA presents a number of challenges due to noise and poor contrast.

Thanks to the inheritance and development of previous vascular segmentation methods for colour fundus images, vascular segmentation in OCTA images has been developing rapidly in recent years.

Both spatial and frequency domain methods have been widely studied. Spatial domain methods operate directly on the image pixel space and use local neighbourhood information to represent the image content. Frequency domain methods, which transform the image into different frequencies using discrete wavelet transform techniques, are capable of identifying target edges and locations at different levels of detail.

Deep learning methods use multilayered artificial neural networks to learn effective representations from raw data (such as images, video, or text) and have shown great potential and effectiveness in handling complex pattern recognition tasks.

Convolution-Based Networks: Convolutional neural network (CNN) [10]-based methods are used for better retinal vessel detection. Fully convolutional network (FCN) [23] can be used for semantic segmentation that classifies images at the pixel level. FCN achieves end-to-end learning of the input image by replacing the fully connected layer of the CNN with a convolutional layer.

Enhanced UNet Architectures: Cao et al. introduced an enhanced UNet architecture named GPC-Net, which achieves superior results in OCTA vascular segmentation tasks. The proposed network leverages an innovative lossless dimensionality compression method to effectively utilize 3D information, and integrates Group Normalization, PReLU activation and channel attention mechanisms to deliver outstanding performance metrics [24]. Ronneberger et al. applied the UNet [9] to retinal vessel segmentation in fundus images of pathological conditions. Gu et al.

designed a context encoder network (CENet) [25] consisting of dense atrous convolution and residual multi-kernel pooling modules for retinal vessel image segmentation. UNet++ [26] adds skip connection between each encoder and decoder layer to allow direct feature propagation. These skip connections help improve the network to capture detailed information and solve the gradient vanishing problem. ResAt-UNet [28] is an enhanced UNet network by fusing ResNet34 and the Attention module, while retaining skip connection [27] and upsampling structure for accurate segmentation. This combination improves the network to capture details of images. DenseUNet+ [29] applies dense concatenation, which improves feature utilization and reduces overfitting, thus improving the accuracy of segmentation.

Hybrid and Transformer-Based Models: Wang et al. presented a novel multi-scale Transformer global attention network [30]. Xie et al. proposed a simple, efficient and powerful semantic segmentation framework named as SegFormer, which contains a positional-encoding-free, hierarchical transformer encoder and a lightweight All-MLP (multilayer perceptron) decoder [31]. Chen et al. proposed a retinal OCT angiography vessel segmentation transformer. This model captures global retinal vessel information and reduces the expensive calculation in both time and space perspectives [32]. Tan et al. proposed a multi-head dynamic token aggregation attention-based dynamic token aggregation transformer for retinal OCTA vessel segmentation [33].

However, there are few deep learning methods for vascular segmentation in OCTA fundus images.

3 DataSet

In this section, we describe in detail the information on the existing fundus images datasets. Table 1 lists the information on the common fundus databases. Most existing retinal datasets have image-level annotations. Several datasets also have pixel-level annotations for lesions.

(1) **ONHSD** The ONHSD (Optic Nerve Head Surface Distance dataset) [34] dataset contains a total of 100 images with a resolution of 640×480 pixels.

Table 1 Information of representative fundus datasets

Dataset	Images	Tasks	Image size	Modality	Year
ONHSD	100	Diagnosis	640×480	RGB (24 bits)	2004
DRIVE	40	Segmentation	584×565	RGB (24 bits)	2004
IDRID	516	Detection and Segmentation	512×512	RGB (24 bits)	2017
OCTA-500	500	Segmentation	400×400	Grey(OCTA)	2020
ROSE	117	Segmentation	304×304	Grey(OCTA)	2021
REVIO	50	Segmentation	288×288	Grey(OCTA)	2024

(2) **DRIVE** The DRIVE (Digital Retinal Images for Vessel Extraction) [35] dataset is a retinal vessel segmentation dataset. It consists of a total of 40 colour fundus images in JPEG format, including seven cases with abnormal pathology. The images were taken with a Canon CR5 non-dilated 3CCD camera with a field of view (FOV) of 45 degrees. The resolution of each image was 584×565 pixels with 8 bits per colour channel. These 40 images were divided equally into 20 images for the training set and 20 images for the test set.

(3) **IDRID** The IDRID (Indian Diabetic Retinopathy Image Dataset) [36] dataset consists of 516 images acquired with the Kowa VX-10 alpha digital fundus camera with a 50 degree field of view (FOV), all located near the macula. The resolution of the images is 2848×4288 pixels and the dataset consists of 81 colour fundus images with DR signatures, with precise pixel-level annotations of DR-related abnormalities (e.g. microaneurysms (MA), soft exudates (SE), hard exudates (EX) and haemorrhages (HE)) provided as binary masks for performance evaluation of individual lesion segmentation techniques.

(4) **ROSE** The ROSE (Retinal OCT-Angiography vessel SEgmentation) [4] dataset is an open-source OCTA fundus vessel segmentation dataset consisting of two sub-datasets, i.e. ROSE-1 and ROSE-2. The ROSE-1 dataset consisted of 117 OCTA images from 39 subjects. These images consisted of 26 patients with Alzheimer's disease and the remaining 13 were healthy subjects. All OCTA scans were performed with an RTVue XR Avanti SD-OCT system (Optovue, equipped with AngioVue software, with an image resolution of 304×304 pixels. The scan area was $3 \times 3\ mm^2$ centred on the fovea, with a ring of 0.6 mm−2.5 mm diameter around the centre of the fovea.

(5) **OCTA-500** The OCTA-500 (Optical Coherence Tomography Angiography-500) [5] dataset contains a total of 500 images with data acquired using a 70 kHz spectral domain OCT system with a centre wavelength of 840 nm (RTVue XR, Optovue, CA) and an image resolution of 400×400 pixels. The images are classified into OCTA-500-3 mm and OCTA-500-6 mm according to the type of field of view.

(6) **REVIO** The Retinal Vessels Images in OCTA (REVIO) [3] dataset was imaged using swept-source optical coherence tomography (SS-OCT) with a 12 mm × 12 mm single scan centred on the fovea. The REVIO dataset is captured from 50 people. OCTA images showed blood vessels in the retinal and choroidal layers. The inner limiting membrane (ILM), nerve fibre layer (NFL), ganglion cell layer (GCL), inner plexiform layer (IPL), inner nuclear layer (INL), outer plexiform layer (OPL), retinal pigment epithelium (RPE) and Bruch's membrane (BM) were used to define the border of the retinal vascular layer.

4 Methodology

4.1 Network architecture

Figure 2 shows the novel SMFDNet architecture for retinal vessel segmentation in OCTA images. Our proposed network mainly consists of a spatial domain module (SDM), frequency domain module (FDM), dual-domain fusion module (DDFM) and decoder module.

The spatial domain module uses skip connections to preserve the original feature vectors. The frequency domain module uses discrete wavelet transform to obtain low-frequency feature vectors and high-frequency feature vectors. To capture multi-frequency domain information and to compensate for the loss of spatial and frequency domain features, we fused the spatial and frequency domain features, and spatial and frequency domain feature vectors. In the decoder module, the fused spatial domain and multi-frequency domain feature vectors are fed to the decoder module to recover the original resolution.

4.2 Frequency domain module

We intend to use a reversible downsampling operation to preserve the details of the original image, which allows multi-frequency analysis of the feature vectors, while reducing the information loss by using a discrete wavelet transform operation. The designed frequency domain module consists of four main components, discrete wavelet

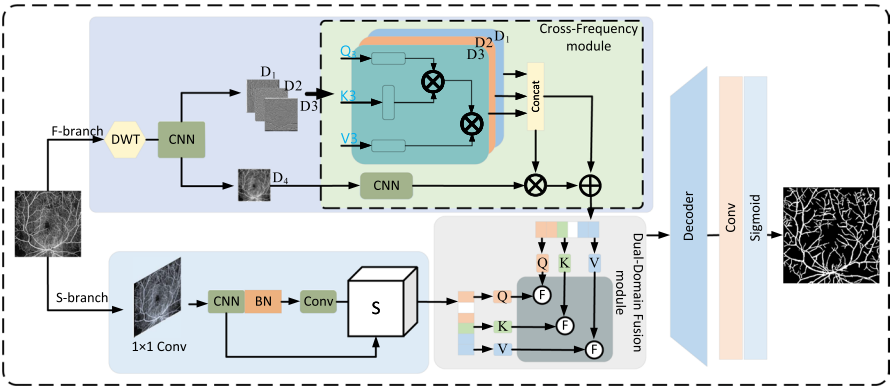

Fig. 2 Architecture of the proposed SMFDNet network. The F-branch is the frequency domain-based perceptual module. Q_3 is the query vector obtained by projecting D_3 in the spatial domain. K_3 represents the key vector obtained by projecting D_3 in the spatial domain. V_3 denotes the value vector obtained by projecting D_3 in the spatial domain. The S-branch is a directed module based on the spatial domain. It is fed to the decoding module to obtain the final fine segmentation result. CNN refers to the feature vector coding process including convolutional layers, RELU function, which is used a total of 8 times. The detailed parameters of the convolutional layer are 3x3 kernel size, padding=1 and stride=1

transform, inverse discrete wavelet transform, convolutional layer and encoder network. The discrete wavelet transform is used to obtain four frequency domain feature vectors, while the inverse discrete wavelet transform is used to convert the frequency domain feature vectors into feature vectors. The encoder network is used to convolve the input high- and low-frequency feature vectors and further extract the frequency domain features.

First, the input features are subjected to a discrete wavelet transform to obtain high- and low-frequency feature vectors in the frequency domain, which are transformed to $D1$, $D2$, $D3$ and $D4$.

$$D_1 = f_{LL}/2. \tag{1}$$

$$D_2 = f_{LH}/2. \tag{2}$$

$$D_3 = f_{HL}/2. \tag{3}$$

$$D_4 = f_{HH}/2. \tag{4}$$

f_{LL} is the low-frequency features obtained after discrete wavelet transform. f_{LH}, f_{HL}, f_{HH} is the high-frequency features obtained after discrete wavelet transform. D_2, D_3 and D_4 retain high-frequency information of vessel texture on fine-grained feature, while D_1 represents all low-frequency information on coarse-grained feature. The low-frequency component can maintain the original image feature information. Two discrete wavelet transforms are applied to obtain high-frequency and low-frequency feature vectors of deep vessel information. After the encoder of the network, the image feature vectors are further extracted and perceived.

$$(I_1) = \text{Encoder}(D_1), \tag{5}$$

$$(I_2, I_3, I_4) = \text{Encoder}(D_2, D_3, D_4), \tag{6}$$

where Encoder$(\cdot)$ represents the encoder module. I_1 denotes the encoder low-frequency feature vector. I_2, I_3, I_4 denotes the encoerd high-frequency feature vector. The encoder adopts four convolutional layers, each of which is a 3×3 convolutional kernel. After the convolutional layers, the Batch Normalization (BN) layer and the Rectified Linear Unit(ReLU) layer are connected.

Then, after two inverse discrete wavelet transforms, the high-frequency information on the horizontal, vertical and diagonal lines is fused together, which can preserved the small feature vectors of the blood vessels.

$$L_1 = IDWT(I_1), \tag{7}$$

$$L_2 = IDWT(I_2, I_3, I_4), \tag{8}$$

where $IDWT(\cdot)$ denotes inverse discrete wavelet transform. L_1 represents the feature vector of low-frequency information subjected to inverse discrete wavelet transform. L_2 represents the feature vector of high-frequency information subjected to inverse discrete wavelet transform. To obtain effective features, we construct a module by fusing the feature vectors L_1 and L_2 as follows.

$$L_3 = \text{Cross}(L_1, L_2). \tag{9}$$

Cross$(\cdot)$ is a function that fuses low- and high-frequency feature vectors. L_3 is the feature vector after the fusion of L_1 and L_2.

Finally, to extract the normalized vectors after the fusion of low and high frequencies, we design the following operations:

$$L_4 = \text{Reshape}(\text{Conv}(ReLU(BN(L_3)))), \tag{10}$$

where Reshape$(\cdot)$ is the reshape operation. Conv$(\cdot)$ denotes convolutional operation. $RELU(\cdot)$ represents Rectified Linear Unit. $BN(\cdot)$ represents batch normalization.

4.3 Spatial domain module

The spatial domain module is to extract the texture and edge features in the blood vessels from the image and fuse the feature information from different locations to capture the global spatial information, which makes the model have better perceive and capture the local regions of the blood vessel features. First, we perform 1×1 convolution on the input image to obtain the feature vectors. The formula is as follows:

$$W_1 = \text{Conv}(f), \tag{11}$$

Conv$(\cdot)$ denotes convolution operation, and f is the input feature.

Secondly, we reduce the feature vector loss by a skip connection operation, which is able to preserve the detailed features in the input data.

$$W_2 = \text{Conv}(BN(RELU(\text{Conv}(W_1)))), \tag{12}$$

$$W_3 = (1 + f) \times W_2, \tag{13}$$

where $RELU(\cdot)$ is the activation function, Conv$(\cdot)$ is the convolution operation and $BN(\cdot)$ is the batch normalization. W_2 is the encoded feature vector.

4.4 Dual-domain fusion module

By taking full advantage of both the spatial and frequency domains, the model is able to capture more information from the input data and further recover the continuum details of small vessels, we design the dual-domain fusion module. Figure 3 shows the dual-domain fusion module. The spatial domain contains the local information and geometric structure of the image, and the frequency domain reflects the

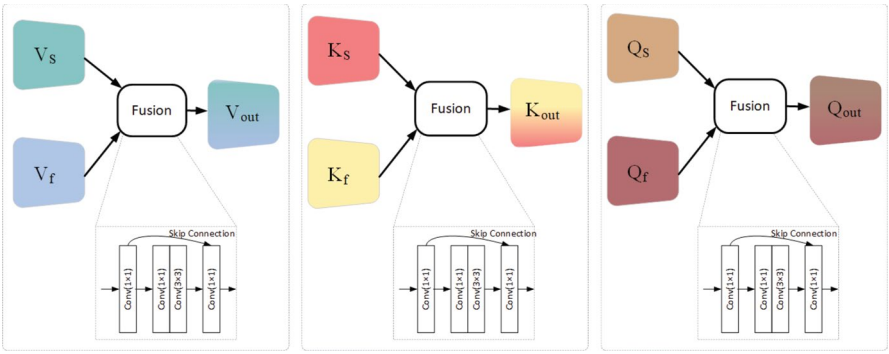

Fig. 3 Architecture of dual-domain fusion module

texture information of the image. It is of great significance for vessel segmentation by fusing two modules.

In order to fully integrate the feature vectors from the spatial and frequency domain modules, the spatial domain feature vector W_3 and the frequency domain feature vector L_4 are first subjected to the fully connected operation.

We use a fully connected layer to project W_3 in the spatial domain into Q_s, K_s, V_s vectors, and L_4 in the frequency domain into Q_f, K_f, V_f. To obtain Q and K in the feature vectors, we use softmax operation to compute Q_s, K_s in the spatial domain and Q_f, K_f in the frequency domain, with the following equations.

$$F_{\text{score}} = \text{Softmax}(Q_s + Q_f, K_s + K_f), \tag{14}$$

where $\text{Softmax}(\cdot)$ is represents Soft Maximum function. Q_s is the query vector obtained by projecting W_3 in the spatial domain. K_s represents the key vector obtained by projecting W_3 in the spatial domain. Q_f denotes the query vector obtained from the L_4 projection in the frequency domain. K_f represents the key vector obtained from the L_4 projection in the frequency domain. F_{score} is the output of $\text{Softmax}(\cdot)$ operation.

We operate the resulting fusion-score using the dot product with the following equation.

$$\text{Fusion} = (V_s + V_f) \times F_{\text{score}}, \tag{15}$$

where V_s denotes the value vector obtained by projecting W_3 in the spatial domain. V_f represents the value vector obtained from the L_4 projection in the frequency domain.

4.5 Decoder module

We use upsampling convolution with 3 convolutional layers, each with a convolution kernel size of 3×3, connecting a batch normalization layer and a RELU function layer. An upsampling operation is performed on the input feature vector, which can obtain the vessel segmentation result as follows.

$$\text{Out} = \text{Sigmoid}(\text{Decoder}(F_{\text{score}})). \tag{16}$$

5 Experimental results

5.1 Datasets and evaluation metrics

We perform extensive experiments on the REVIO, ROSE and OCTA-500 datasets to evaluate our proposed SMFDNet.

REVIO. The REVIO dataset contains 50 OCTA images with a resolution of 288×288 pixels.

ROSE. The ROSE [4] dataset consists of 117 OCTA images, an image resolution of 304×304 pixels.

OCTA-500. The OCTA-500 [5] dataset contains a total of 500 images, which are divided into two subsets according to the field of view type: OCTA-500-6 mm and OCTA-500-3 mm. The OCTA-500-6 mm dataset contains 300 images with an image resolution of 400×400 pixels. The OCTA-500-3 mm dataset contains 200 images with an image resolution of 304×304 pixels.

Metrics for Vessel Segmentation. We utilize the following metrics to comprehensively and objectively evaluate the segmentation performance of all methods.

①Sensitivity=TP/(TP+FN).

②Specificity=TN/(TN+FP).

③Dice Coefficient=2×TP/(FP+FN+2×TP).

④Accuracy=(TP+TN)/(TP+TN+FP+FN).

⑤Jaccard=TP/(TP+FN+FP).

TP stands for True Positive which indicates the number of positive pixels classified accurately, *FP* denotes False Positive which indicates the number of positive pixels classified accurately. *TN* stands for True Negative which indicates the number of negative pixels classified accurately. *FN* presents False Negative which is the number of negative pixels classified accurately.

Experimental Setting. In experiments, we set a learning rate and a weight decay of Adam to 0.001 and 0.0001, respectively. All algorithms are trained and tested with two NVIDIA RTX2080-Ti GPUs. The epochs of training model were set to 350.

5.2 Performance comparison and analysis

In this section, we conduct comprehensive experiments on the REVIO, ROSE and OCTA-500 datasets for the OCTA microvessel segmentation, and compare our approach with the existing state-of-the-art segmentation methods to demonstrate the advantages of our method for OCTA microvessel segmentation.

Compared methods: To verify the superiority of our method, we compared our method with other state-of-the-art segmentation methods on the REVIO,

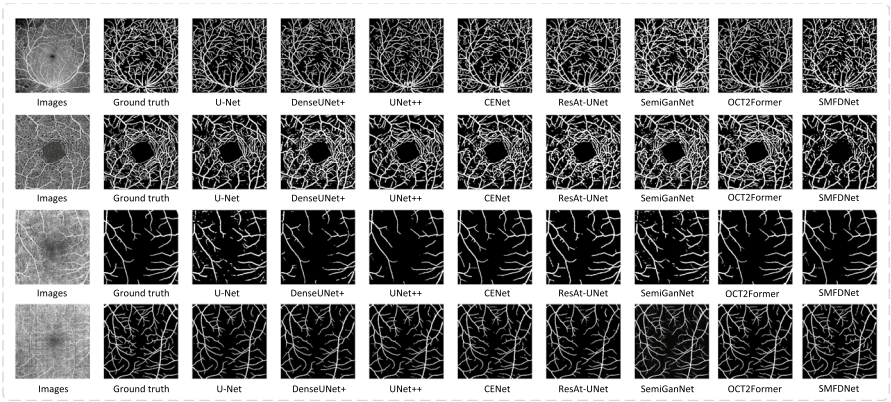

Fig. 4 The REVIO, ROSE, OCTA-500-3 mm and OCTA-500-6 mm datasets from top to bottom. The results of our method and other segmentation methods on four datasets

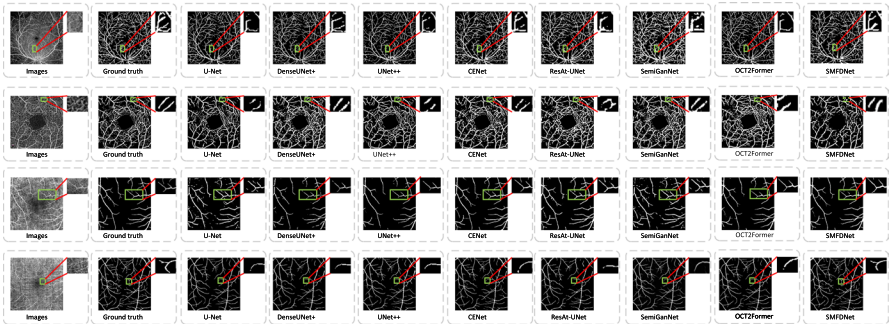

Fig. 5 Vessel segmentation results of different methods on the REVIO, ROSE, OCTA-500-3 mm and OCTA-500-6 mm datasets. The results of our method and other segmentation methods on the four datasets. From left to right: original images, manual annotations, vessel segmentation results obtained by UNet, DenseUNet+, UNet++, CENet, ResAt-UNet, SemiGanNet, OCT2Former and the proposed method (SMFDNet), respectively

ROSE and OCTA-500 datasets. The compared methods includes UNet, ResAt-UNet, CENet, DenseUNet+, UNet++ and SemiGanNet.

5.3 Results

The respective vessel segmentation results obtained by the proposed method and other selected segmentation networks are shown in Figs. 4 and 5. From the visual inspection, it is clear that our method identifies more complete and thinner vessels. It can be seen that other methods are less able to preserve fine vessels in areas of low contrast. In contrast, the proposed SMFDNet produced more

Table 2 Results on the REVIO dataset (%)

Model	Sensitivity	Specificity	Dice	Accuracy	Jaccard
UNet	74.56	95.35	77.11	91.11	62.77
UNet++	70.32	95.77	74.96	90.56	59.96
SemiGanNet	66.71	96.93	73.68	91.43	62.45
CENet	79.73	94.65	79.29	91.65	66.01
ResAt-UNet	81.67	94.36	79.94	91.86	67.05
DenseUNet+	74.74	95.53	77.56	91.32	63.38
OCT2former	78.14	93.14	75.9	90.06	61.2
SMFDNet (ours)	**84.67**	**96.99**	**86.9**	**94.21**	**77.42**

Best results are in bold

Table 3 Results on the ROSE dataset (%)

Model	Sensitivity	Specificity	Dice	Accuracy	Jaccard
UNet	63.01	95.41	69.46	88.68	53.35
UNet++	67.92	95.32	72.98	89.65	57.67
SemiGanNet	67.88	98.28	75.84	94.1	61.31
CENet	76.36	93.83	76.22	90.17	62.11
ResAt-UNet	71.94	92.2	71.05	87.98	55.41
DenseUNet+	68.63	94.94	72.68	89.45	57.26
OCT2former	71.97	91.29	69.79	87.28	55.66
SMFDNet (ours)	**76.70**	**98.46**	**77.68**	**94.46**	**63.08**

Best results are in bold

intuitive information and was able to detect thinner vessels more completely. Therefore, in order to better evaluate the performance of the proposed method, we provide quantitative results in the following subsection.

Performance on the REVIO dataset: Table 2 shows the segmentation results obtained by our method and the state-of-the-art methods. The proposed network outperforms other comparative methods. In particular, specificity and dice are improved by 4.94% and 7.61% compared to CENet. The method results in the extraction of small capillaries with good continuity and integrity, while the other methods have a relatively low capillary response.

Performance on ROSE dataset: we evaluated the vessel segmentation performance of all methods. Table 3 shows the segmentation performance of different methods on the images. Our method achieves the best segmentation performance on Dice, F1-score and Jaccard.

Performance on the OCTA-500-3 mm dataset: Table 4 demonstrates the segmentation results of the different methods on the OCTA-500-3 mm dataset. Overall, our method achieves the best performance on almost all metrics. Nevertheless, using

Table 4 Results on the OCTA-500-3 mm dataset (%)

Model	Sensitivity	Specificity	Dice	Accuracy	Jaccard
UNet	79.68	95.44	65.19	94.36	58.54
UNet++	68.08	98.47	78.12	97.5	64.59
SemiGanNet	77.57	98.13	81.93	97.6	69.45
CENet	91.46	97.32	80.08	96.93	67.02
ResAt-UNet	78.25	98.08	81.71	97.69	69.24
DenseUNet+	90.75	98.35	76.88	97.35	62.97
OCT2former	91.16	98.4	85.51	97.92	74.97
SMFDNet (ours)	**92.03**	**98.96**	**88.13**	**98.39**	**78.84**

Best results are in bold

Table 5 Results on the OCTA-500-6 mm dataset (%)

Model	Sensitivity	Specificity	Dice	Accuracy	Jaccard
UNet	78.58	98.52	82.26	96.81	69.08
UNet++	78.9	98.79	82.26	96.84	70.23
SemiGanNet	89.38	98.68	81.65	96.82	69.12
CENet	78.24	98.99	82.99	97.08	71.09
ResAt-UNet	72.31	98.29	76.11	95.88	61.76
DenseUNet+	72.53	98.94	78.85	96.5	65.53
OCT2former	80.91	98.67	83.07	97.05	71.38
SMFDNet (ours)	**89.65**	**99.01**	**85.84**	**97.5**	**75.31**

Best results are in bold

our spatial and frequency domain network, the proposed method is able to correctly identify most vessels.

Performance on OCTA-500-6 mm dataset: Table 5 shows the segmentation results of the different segmentation methods on the OCTA-500-6 mm dataset. The proposed SMFDNet achieves the best performance except Specificity. The method achieves a score of 99.01%, which is higher than SemiGanNet by 0.33%. This indicates that SemiGanNet in this paper is weaker than SMFDNet for identifying negative pixels.

5.4 K-fold cross-validation

We used k-fold cross-validation on SMFDNet. We split the REVIO dataset with 5 folds and then trained our SMFDNet by each fold, respectively. Thus, we obtain 5

trained models. The results are in Table 6. Figure 6 shows the results of K-fold operations on the REVIO dataset using SMFDNet method.

5.5 Ablation study

The SMFDNet vessel segmentation method proposed in this paper consists of spatial and frequency domain branches. To validate the effectiveness of these components, we performed ablation experiments on the REVIO dataset, where we evaluated how each component affects the model. The results show that it improves the performance of model by fusing the spatial and frequency domains in Table 7.

For the REVIO dataset, the performance of frequency branching is shown in Table 7, SMFDNet-S is the backbone only with frequency domain. One can see that the Sensitivity, Specificity, Dice, Accuracy, F1-Score and Jaccard are lower than SMFDNet by 8.86%, 4.63%, 18.16%, 7.96%, 18.16% and 11.9%, respectively. And the segmented vessels were discontinuous.

Table 7 shows the results of the spatial domain. SMFDNet-F is the model only with spatial domain, and we can see that the Sensitivity, Specificity, Dice, Accuracy, F1-Score and Jaccard are lower than SMFDNet by 13.62%, 4.63%, 21.59%, 9.04%, 21.59% and 15.7%, respectively.

We used floating point operations per second (FLOPS) and parameters (params) to quantify the ablation experiments. FLOPS is the number of floating point operations that can be performed per second. Params is the number of parameters in the model. Training time is the time from the start of model training to the end of training.

$$\text{Total FLOPs} = \sum_{l=1}^{L} \text{FLOPs}_l, \tag{17}$$

where Total FLOPs denotes total floating point operations. GFLOPs = 10^9FLOPs. FLOPs_l is the FLOPs in layer L.

$$\text{Total parameters} = \sum_{l=1}^{L} P_l, \tag{18}$$

where Total parameters denotes total number of parameters. P_l is the number of parameters in the l'th layer.

Table 8 shows the quantitative results of the ablation experiments. SMFDNet-S is the model with the frequency domain, while SMFDNet-F is the model with the spatial domain. The total number of parameters of SMFDNet-S is 15.5M, and the total number of parameters of SMFDNet-F is 14.7M. The number of floating point operations per second of SMFDNet-S is 16.8 GFLOPS, which is less than that of SMFDNet 0.4 GFLOPS. SMFDNet-S can perform 16.8 GFLOPS per second, which is 0.4 GFLOPS less than that of SMFDNet. The training time of SMFDNet-S is 1787s, which is 11 s less than that of SMFDNet (Table 8).

Table 6 Results of K-fold cross validation on the REVIO dataset using SMFDNet (%)

Fold	Sensitivity	Specificity	Dice	Accuracy	Jaccard
Fold1	82.5	97.34	87.68	95.41	70.67
Fold2	83.05	98.6	88.16	94.99	81.72
Fold3	85.9	97.77	89.63	95.97	85.66
Fold4	83.77	96.56	84.69	92.46	76.57
Fold5	86.23	95.61	83.02	93.18	78.77
Mean	84.29	97.17	86.36	94.02	78.67

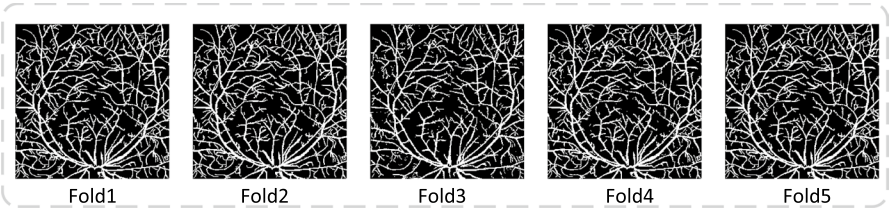

Fig. 6 The results by our SMFDNet on the REVIO dataset with 5-fold validation, respectively. From left to right: Fold1 image, Fold2 image, Fold3 image, Fold4 image, Fold5 image, respectively

Table 7 Results on the ablation (%)

Model	Sensitivity	Specificity	Dice	Accuracy	Jaccard
SMFDNet-S	75.81	92.36	68.74	86.25	65.52
SMFDNet-F	71.05	92.37	65.31	85.17	61.72
SMFDNet	**84.67**	**96.99**	**86.9**	**94.21**	**77.42**

Best results are in bold

Table 8 The summary of required GFLOPS, Params and Second for model

Model	GLOPS	Params(M)	Second(s)
SMFDNet-S	16.8	15.2	1787
SMFDNet-F	16.1	14.7	1631
SMFDNet	17.2	15.5	1798

6 Conclusion

In this paper, we propose a novel OCTA vessel segmentation framework, SMFDNet, which can obtain low and high-frequency feature vectors in the frequency domain and spatial domain. By leveraging these complementary features, SMFDNet improves the clarity and accuracy of vessel detection, especially in areas with

small or complex vessel networks, outperforming other state-of-the-art methods on datasets such as REVIO, ROSE and OCTA-500. However, future research could explore improving the model's computational efficiency for real-time applications, especially for large-scale clinical deployment. Additionally, further enhancements could focus on extending its applicability to other imaging modalities or more complex pathologies, making it more versatile in different clinical settings.

Acknowledgements The authors would like to thank the editors and anonymous reviewers for their constructive comments and suggestions.

Author Contributions Conception and design were performed by S.L. and F.M.; (II) administrative support was provided by F. M. and J.M.; (III) provision of study materials or patients was distributed by J.M. and Y.F.; (IV) collection and assembly of data were carried out by F. M.; (V) data analysis and interpretation were conducted by S.L., Y.G., H.L. and R.C.; (VI) manuscript writing was done by all authors; (VII) final approval of manuscript was approved by all authors.

Funding This work was supported by the Natural Science Foundation of Shandong Province (No. ZR2020MF105), Guangdong Provincial Key Laboratory of Biomedical Optical Imaging Technology (No. 2020B121201010), the Natural National Science Foundation of China (Nos. 62175156 and 61675134) and Qufu Normal University Foundation for High Level Research (No. 116-607001).

Data Availability No datasets were generated or analysed during the current study.

Declarations

Ethical approval The authors are accountable for all aspects of the work in ensuring that questions related to the accuracy or integrity of any part of the work are appropriately investigated and resolved. The study was conducted in accordance with the Declaration of Helsinki (as revised in 2013).

Conflict of interest: The authors declare there is no conflict of interest.

References

1. Zhou SK, Greenspan H, Davatzikos C, Duncan SJ, Ginneken BV, Madabhushi A, Prince JL, Rueckert D, Summers MR (2021) A review of deep learning in medical imaging: Imaging traits, technology trends, case studies with progress highlights, and future promises. Proc IEEE 109(5):820–838. https://doi.org/10.1109/JPROC.2021.3054390
2. Li MC, Chen YR, Ji ZX, Xie KR, Yuan ST, Chen Q, Li S (2020) Image projection network: 3D to 2D image segmentation in OCTA images. IEEE Trans Med Imaging 39(11):3343–3354. https://doi.org/10.1109/tmi.2020.2992244
3. Dong BJ, Wang XD, Xie Q, Du F, Gao L, Wu QY, Cao GH, Dai CS (2022) A multi-branch convolutional neural network for screening and staging of diabetic retinopathy based on wide-field optical coherence tomography angiography. IRBM 43(6):614–620. https://doi.org/10.1016/j.irbm.2022.04.004
4. Ma YH, Hao HY, Xie JY, Fu HZ, Zhang J, Yang JL, Wang Z, Liu J, Zheng YL, Zhao Y, Rose T (2021) A retinal OCT-angiography vessel segmentation dataset and new model. IEEE Trans Med Imaging 40(3):928–939. https://doi.org/10.1109/tmi.2020.3042802
5. Li MC, Huang K, Xu QZ, Yang JD, Zhang YH, Ji ZX, Xie KR, Yuan ST, Liu QH, Chen Q (2024) OCTA-500: a retinal dataset for optical coherence tomography angiography study. Med Image Anal 93:103092–103092. https://doi.org/10.1016/j.media.2024.103092
6. Lns I, Wang JC, Cui Y, Katz R, Vingopoulos F, Staurenghi G, Vavvas DG, Miller JW, Miller JB (2021) Retinal applications of swept source optical coherence tomography (OCT) and optical

coherence tomography angiography (OCTA). Prog Retin Eye Res 84:100951–100951. https://doi.org/10.1016/j.preteyeres.2021.100951
7. Carlo TE et al (2015) A review of optical coherence tomography angiography (OCTA). Int J Retina Vitreous 1(1):5–5. https://doi.org/10.1186/s40942-015-0005-8
8. Minaee S, Boykov Y, Porikli F, Plaza A, Kehtarnavaz N, Terzopoulos D (2022) Image segmentation using deep learning: a survey. IEEE Trans Pattern Anal Mach Intell 44(7):3523–3542. https://doi.org/10.1109/TPAMI.2021.3059968
9. Ronneberger O, Fischer P, Brox T (2015) Convolutional Networks for Biomedical Image Segmentation. Medical Image U-Net Comput Comput Assisted Interv 9351:234–241. https://doi.org/10.1007/978-3-319-24574-4_28
10. Zhang T, Deng L, Huang T, Chanussot J, Vivone G (2023) A triple-double convolutional neural network for panchromatic sharpening. IEEE Trans Neural Networks Learning Syst 34(11):9088–9101. https://doi.org/10.1109/tnnls.2022.3155655
11. Wang Y, Wu H, Zhang J, Gao Z, Wang J, Yu PS, Long M (2023) PredRNN: A Recurrent Neural Network for Spatiotemporal Predictive Learning. IEEE Trans Pattern Anal Mach Intell 45(2):2208–2225. https://doi.org/10.1109/TPAMI.2022.3165153
12. Hassini K, Khalis S, Habibi O, Chemmakha M, Lazaar M (2024) An end-to-end learning approach for enhancing intrusion detection in Industrial-Internet of Things. Knowledge-based Syst. https://doi.org/10.1016/j.knosys.2024.111785
13. Juneja M, Minhas JS, Singla N, Thakur S, Thakur N, Jindal P (2022) Fused framework for glaucoma diagnosis using Optical Coherence Tomography (OCT) images. Expert Syst Appl 201:117202. https://doi.org/10.1016/j.eswa.2022.117202
14. Bayhaqi YA, Hamidi A, Canbaz F, Navarini AA, Cattin PC, Zam A (2022) Deep-learning-based fast optical coherence tomography (OCT) image denoising for smart laser osteotomy. IEEE Trans Med Imaging 41(10):2615–2628. https://doi.org/10.1109/TMI.2022.3168793
15. He X, Fang L, Tan M, Chen X (2022) Intra- and inter-slice contrastive learning for point supervised OCT fluid segmentation. IEEE Trans Image Process 31(1):1870–1881. https://doi.org/10.1109/TIP.2022.3148814
16. Dong W, Du Y, Xu J, Dong F, Ren S (2022) Spatially adaptive blind deconvolution methods for optical coherence tomography. Comput Biol Med 147:105650. https://doi.org/10.1016/j.compbiomed.2022.105650
17. Garca G, Colomer A, Naranjo V (2021) Glaucoma detection from raw Sd-Oct volumes: a novel approach focused on spatial dependencies. Comput Methods Programs Biomed 200:105855. https://doi.org/10.1016/j.cmpb.2020.105855
18. Xu YX, Han K, Xu C, Tang Y, Xu C, Wang Y (2021) Learning Frequency Domain Approximation for Binary Neural Networks. Computing Research Repository, 25553–25565
19. Cheng H, Yang S, Zhou TY, Guo L, Wen B (2023) Frequency guidance matters in few-shot learning. IEEE Int Conf Comput Vision. https://doi.org/10.1109/iccv51070.2023.01085
20. Amato A, Nadin F, Borghesan F (2020) Widefield optical coherence tomography angiography in diabetic retinopathy. J Diabetes Res. https://doi.org/10.1155/2020/8855709
21. Mistelbauer G, Morar A, Schernthaner R, Strassl A, Fleischmann D, Moldoveanu F, Grller ME (2021) Semi-automatic vessel detection for challenging cases of peripheral arterial disease. Comput Biol Med 133:104344. https://doi.org/10.1016/j.compbiomed.2021.104344
22. Su R, Sluijs MVD, Cornelissen SAP, Lycklama G, Hofmeijer J, Majoie CBL, Doormaal PJ, Es AC, Ruijters D, Niessen WJ, Lugt A, Walsum T (2022) Spatio-temporal deep learning for automatic detection of intracranial vessel perforation in digital subtraction angiography during endovascular thrombectomy. Med Image Anal 77:102377. https://doi.org/10.1016/j.media.2022.102377
23. Long J, Shelhamer E, Darrell T (2015) Fully convolutional networks for semantic segmentation. IEEE Trans Pattern Analy Machine Intell. https://doi.org/10.1109/TPAMI.2016.2572683
24. Cao G, Peng Z, Zhou Z, Wu Y, Zhang Y, Yan R (2024) Multi-task OCTA image segmentation with innovative dimension compression. Pattern Recogn. https://doi.org/10.1016/j.patcog.2024.111123
25. Gu Z, Cheng J, Fu H, Zhou K, Hao H, Zhao Y, Zhang T, Gao S, Liu J (2019) CE-Net: context encoder network for 2D medical image segmentation. IEEE Trans Med Imaging 38(10):2281–2292. https://doi.org/10.1109/TMI.2019.2903562
26. Zhou Z, Siddiquee MMR, Tajbakhsh N, Liang J (2018) Unet plus plus: a nested U-net architecture for medical image segmentation. Lect Notes Comput Sci 11045:3–11. https://doi.org/10.1007/978-3-030-00889-5_1

27. Liu X, Wang S, Zhang Y, Liu D, Hu W (2021) Automatic fluid segmentation in retinal optical coherence tomography images using attention based deep learning. Neurocomputing 452:576–591. https://doi.org/10.1016/j.neucom.2020.07.143
28. Fan Z, Liu Y, Xia M, Hou J, Yan F, Zang Q (2023) ResAt-UNet: a U-Shaped network using ResNet and attention module for image segmentation of urban buildings. IEEE J Select Topics Appl Earth Observ Remote Sens 16:2094–2111. https://doi.org/10.1109/jstars.2023.3238720
29. Etiner H, Metlek S (2023) DenseUNet+: a novel hybrid segmentation approach based on multi-modality images for brain tumor segmentation. J King Saud Univ Comput Inf Sci 35(8):101663–101663. https://doi.org/10.1016/j.jksuci.2023.101663
30. Wang M et al (2022) MsTGANet: automatic drusen segmentation from retinal OCT images. IEEE Trans Med Imaging 41(2):394–406. https://doi.org/10.1109/tmi.2021.3112716
31. Xie E, Wang W, Yu Z, Anandkumar A, Luo Alvarez Jp (2021) SegFormer: simple and efficient design for semantic segmentation with transformers. Comput Res Repository. https://doi.org/10.48550/arxiv.2105.15203
32. Chen Y, Lu Q, Yu X, Luo E, Adeli Y, Wang L, Lu Yuille A, Zhou Y (2021) Transunet: Transformers make strong encoders for medical image segmentation, arXiv preprint arXiv:2102.04306, https://doi.org/10.48550/arxiv.2102.04306
33. Tan X, Chen X, Meng Q, Shi F, Xiang D, Chen Z, Pan L, Zhu W (2023) OCT2Former: a retinal OCT-angiography vessel segmentation transformer. Comput Methods Programs Biomed. https://doi.org/10.1016/j.cmpb.2023.107454
34. Lowell J, Hunter A, Steel D, Basu A, Ryder R, Fletcher E, Kennedy L (2004) Optic nerve head segmentation. IEEE Trans Med Imaging 23:256–264. https://doi.org/10.1109/TMI.2003.823261
35. Staal J, Abrmoff MD, Niemeijer M, Viergever MA, Ginneken BV (2004) Ridge-based vessel segmentation in color images of the retina. IEEE Trans Med Imaging 23:501–509. https://doi.org/10.1109/TMI.2004.825627
36. Porwal P, Pachade S, Kamble R, Kokare M, Deshmukh G, Sahasrabuddhe V, Meriaudeau F (2018) Indian diabetic retinopathy image dataset (Idrid): a database for diabetic retinopathy. Screen Res Data 3:25. https://doi.org/10.3390/data3030025