RESEARCH ARTICLE

SFNet: Spatial and Frequency Domain Networks for Wide-Field OCT Angiography Retinal Vessel Segmentation

Sien Li[1] | Fei Ma[1] | Fen Yan[2] | Xiwei Dong[3] | Yanfei Guo[1] | Jing Meng[1] | Hongjuan Liu[1]

[1]School of Computer Science, Qufu Normal University, Rizhao, Shandong, China | [2]Ultrasound Medicine Department Qufu People's Hospital, Qufu, Shandong, China | [3]School of Computer and Big Data Science, Jiujiang University, Jiujiang, Jiangxi, China

Correspondence: Fei Ma (mafei0603@163.com)

ABSTRACT

Automatic segmentation of blood vessels in fundus images is important to assist ophthalmologists in diagnosis. However, automatic segmentation for Optical Coherence Tomography Angiography (OCTA) blood vessels has not been fully investigated due to various difficulties, such as vessel complexity. In addition, there are only a few publicly available OCTA image data sets for training and validating segmentation algorithms. To address these issues, we constructed a wild-field retinal OCTA segmentation data set, the Retinal Vessels Images in OCTA (REVIO) dataset. Second, we propose a new retinal vessel segmentation network based on spatial and frequency domain networks (SFNet). The proposed model are tested on three benchmark data sets including REVIO, ROSE and OCTA-500. The experimental results show superior performance on segmentation tasks compared to the representative methods.

1 | Introduction

Optical Coherence Tomography Angiography (OCTA) [1] is a high-resolution imaging technique used to visualise the vascular structure and blood flow in the eye. With OCTA technology, the structure and blood flow of the ocular microvascular network, including the retina and choroid, can be visualised in real time [2]. Figure 1 shows the detailed information on blood flow, which demonstrates the comparison between REVIO, ROSE [3], OCTA-500-3 mm [4] and OCTA-500-6 mm [4] FOVs. For the REVIO data set, we can clearly observe that samples have a wider field of view, while the ROSE, OCTA-500-3 mm and OCTA-500-6 mm data sets do not have a sufficient field of view. The wider FOVs can provide the rich information. Therefore, it makes effective to investigate the retinal diseases with a wider field of view.

The retinal vascular is an important structure of the retina that can help clinicians to accurately identify and localise areas of ocular pathology and contribute to a better understanding of the extent of disease and the effectiveness of treatment. Therefore, It is necessary to develop the automatic segmentation system [5-7] for the retinal vascular.

In recent years, neural network-based methods for blood vessel segmentation have produced remarkable results. In addition, deep-learning based methods are known as general-purpose learning methods [8,9], where a single model can be effectively used for different medical imaging modalities. To address this problem, deep-learning methods have been introduced to vessel segmentation to improve the accuracy and robustness of segmentation by automatically learning features. The framework for the segmentation task requires convolutional coding and decoding units, where the coding unit is used to encode the input image multiplied by numerous low-dimensional feature maps. The decoding unit is used to generate segmentation maps of exactly the same dimension as the original input image.
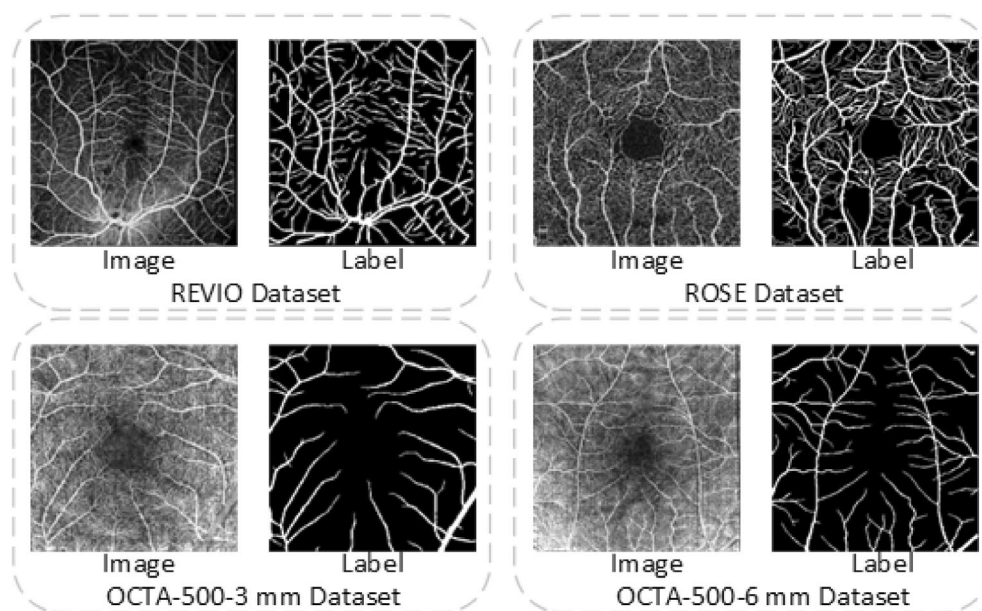
FIGURE 1 | Samples and ground-truths from the REVIO, ROSE, OCTA-500-3 mm and OCTA-500-6 mm data sets.

Spatial domain [10,11] is analysed and processed in image processing by distributing the image in space. In the spatial domain, image quality can be enhanced and features can be extracted by manipulating the raw pixel values of an image. The advantage of the spatial domain is its ability to effectively capture the spatial structure and topological relationships of the vasculature in an image. The blood vessels in the retina have different scales, and the spatial domain is used to pool the vascular features at different scales to improve the model's ability to perceive the blood vessels at different scales.

Convolutional operations in the spatial domain can effectively extract local features in the image, which is crucial for capturing the fine structure and texture information in the retinal vessels. The distribution, connection and arrangement of retinal vessel networks have certain spatial relationships, and the use of the spatial domain can learn these complex spatial relationships, which helps to segment the vessels more accurately. In the spatial domain convolution operation, we use different sizes of convolution kernels to achieve scale invariance [12], which makes the model effective at recognising different scales of vascular networks for different scales of vascular structures. The convolution operation in the spatial domain has the nature of parameter sharing, which helps to reduce the number of parameters in the model and prevent over-fitting, especially in the case of more limited medical image data.

In addition, the frequency [13] and spatial domains [14] are used in medical images for localisation tasks. In the segmentation task, each pixel is assigned a class label with a desired boundary related to the target vessel contour. To identify these target vessel boundaries, we need to highlight the relevant pixels.

The frequency domain represents the frequency distribution of a signal [15]. Frequency domain analysis is often used to study the different frequency components in a signal. In the frequency domain, a signal can be transformed into amplitude and phase information for a series of frequency components using the Discrete Wavelet Transform [16,17] or other transform techniques. The discrete wavelet transform is one of the most commonly used transform methods. Its purpose is to transform a signal from the time domain to the frequency domain so that we can analyse the different frequency components contained in the signal. The network can capture the detailed information of the vessel edges through the high frequency component, while the low frequency component preserves the overall structure and morphology of the vessel, making the segmentation results more accurate. By using the discrete wavelet transform, we are able to highlight the specific frequency components of the blood vessels.

These segmentation methods usually use Convolutional Neural Networks (CNN) [18] to extract the features of an image and then perform classification or regression through a fully connected layer to achieve the segmentation of blood vessels. Amongst them, U-Net [19] is a common network structure that implements the mapping from input image to output segmentation result through an encoder-decoder structure and skip connections. In addition, there are some other network architectures that are also used for blood vessel segmentation, such as CENet [20], helping clinicians to better understand and assess vascular changes in ocular diseases.

The contributions of this work can be summarised as follows:

1. It is the first time to construct a wide-field OCTA fundus images for the vessel segmentation. We built a publicly available Retinal Vessels Images in OCTA (REVIO) data set with accurate manual annotation.

2. We propose a novel Spatial Domain and Frequency Domain Network (SFNet) for OCTA vessel segmentation to detect small differences in vessels. This model consists of a directional attention-based spatial domain and a frequency domain perceived attention-based frequency domain.

3. We conduct a comprehensive experiment on OCTA retinal vessels data sets. Comparative analysis shows that the proposed SFNet performs accurate vessel segmentation for different types of retinal images.

2 | Related Work

Retinal vessel segmentation is a research hot-spot in the field of medical image processing and computer vision. Medical technology continues to evolve, and vessel segmentation plays an increasingly important role in diagnosing and treating diseases. This section introduces some work with blood vessel segmentation.

In recent years, deep-learning based methods have made significant progress in the field of medical image segmentation. In particular, many deep neural networks have been modified and applied for blood vessel segmentation. Vaswani et al. proposed transformer, a model that introduces self-attention, a mechanism by which information at different locations in a sequence can be considered simultaneously without relying on a fixed structure, as in the case of Recurrent Neural Networks (RNN) [21] or Convolutional Neural Networks (CNN) [22-25].

Ronneberger et al. designed the U-Net [19] for the retinal vessel segmentation in the pathological fundus images. U-Net uses an encoder-decoder structure to efficiently capture contextual information in the image and produce fine-grained segmentation results. The overall architecture of U-Net helps to fuse information between low-level and high-level features to extract more contextual and detailed information for the segmentation task. Zhou et al. presented UNet++ [26] to improve the U-Net architecture by introducing the concepts of Multi-scale Information Aggregation Module (MSAM) and path connected. UNet++ reduces the semantic features between encoder and decoder through dense-skipping connections. UNet++ employs a new mechanism (Attention Gate, AG) to attend to scale-differentiated targets. Gu et al. proposed a Context Encoder Network (CENet) [20], which consists of dense atrous convolution and residual multi-kernel pooling modules for retinal vessel image segmentation. Wang et al. presented a new multi-scale transformer global attention network (SemiGanNet) [27] and achieved promising results. Li et al. [28] presented an image projection network that can achieve 3D-to-2D image segmentation in OCTA images. The main feature of ResNet

is the introduction of the concept of Residual Learning. This structure is designed to allow the network layer to skip one or more layers to pass the original information of the input, while using the residual units to learn the residuals instead of learning the original mapping relationships directly. DenseUNet+ [29] is characterised by having both the encoding-decoding structure of U-Net and the densely connected nature of DenseNet.

We found excellent performance of transformer [30-32] in processing sequence data. It also demonstrates the potential for segmenting medical images at the pixel level. Transformers can learn the concept of hierarchical objects by downsampling feature maps appropriately, allowing for sampling at different scales. But transformers cannot capture global information well. Therefore, there is an urgent need to design a network that simultaneously grasps global and local information. We attempt to design new approach to achieve the connection between global and local information.

3 | Data set

In this paper, a new widerfield retinal vascular dataset is constructed using the Optical Coherence Tomography Angiography (OCTA) for the segmentation task. In this section, we describe the newly collected REVIO data set and the existing OCTA dataset. Table 1 lists the information of commonly used fundus databases.

3.1 | Existing Data Sets

1. STARE: The STARE (Structured Analysis of the Retina) [33] data set is a publicly available retinal vessel segmentation data set. The STARE data set consists of 20 fundus images with a resolution of 605×700 pixels. These images include a variety of lesions such as macular degeneration, hypertensive retinopathy and diabetic retinopathy. Each image provides hand-labelled vessel segmentation maps for training and evaluation of vessel segmentation algorithms and is one of the most commonly used fundus vessel segmentation data sets.

2. IDRID: The IDRID (Indian Diabetic Retinopathy Image Dataset) [34] data set consists of 516 images acquired with the Kowa VX-10 alpha digital fundus camera with a 50°

TABLE 1 | Information of representative fundus data sets.

Dataset	Images	Tasks	Image size	Modality	Year
STARE	400	Segmentation	605×700	RGB (24 bits)	2000
DRIVE	40	Segmentation	584×565	RGB (24 bits)	2004
HRF	45	Segmentation	2336×3504	RGB (24 bits)	2013
CHASE_DB1	28	Diagnosis	960×999	RGB (24 bits)	2016
IDRID	516	Detection and segmentation	512×512	RGB (24 bits)	2017
OCTA-500	500	Segmentation	400×400	Gray (OCTA)	2020
ROSE	117	Segmentation	304×304	Gray (OCTA)	2021
REVIO(ours)	150	Segmentation	288×288	Gray (OCTA)	2024

field of view (FOV), all located near the macula. The resolution of the images is 2848×4288 pixels and the data set consist of 81 colour fundus images with DR signatures.

3. DRIVE: The DRIVE (Digital Retinal Images for Vessel Extraction) [35] data set 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 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.

4. HRF: The HRF (High-Resolution Fundus) data set [36] contains 45 images and is a specialised database for automated retinal image segmentation. Each image has a corresponding binarised ground truth vessel segmentation map.

5. CHASE_DB1: The CHASE_DB1 (the Child Heart and Health Study in England) [37] data set included 28 digital fundus colour images obtained from both eyes of 14 children. fundus imaging included two digital images (30°, 1280×960 pixels) centered on the optic disc of each child's eye recorded in subdued lighting using the Nidek NM-200D handheld fundus camera (Nidek Co Ltd. Hiroishi, Japan). Typically, the first 20 images were used as training data and the remaining 8 images were used for testing. The size of each image is 999×960 pixels.

6. ROSE: The ROSE (Retinal Optical Coherence Tomography Angiography Vessel Segmentation) [3] data set was divided into two parts, ROSE-1 and ROSE-2. The ROSE-1 set consisted of a total of 117 OCTA images. All OCTA scans were captured by an RTVue XR Avanti SD-OCT system (Optovue, USA) equipped with AngioVue software with an image resolution of 304×304 pixels. The ROSE-2 subset contains a total of 112 OCTA images. All images in this data set are frontal angiograms of the SVC in a 3×3 mm² area of the central concavity.

7. OCTA-500: The OCTA-500(Optical Coherence Tomography Angiography-500) [4] data set contains 2D voxel data of 500 eyes in both OCT and OCTA modalities. There are two subsets based on field of view type: OCTA-500-6 and OCTA-500-3 mm. OCTA-500-6 mm data set includes 300 subjects who receive a field of view of 6×6 mm. Image size is 400×400 pixels. OCTA-500-3 mm data set includes 200 subjects who receive a field of view of 3×3 mm. Image size is 304×304 pixels. The OCT/OCTA images are from the same device, which is a commercial 70 kHz spectral domain OCT system with a center wavelength of 840 nm (RTVue-XR, Optovue, CA).

3.2 | OCTA System

In this section, we introduce the basic theory of the SS-OCT imaging technology and construct a SS-OCT platform.

3.2.1 | Design of SS-OCT

Figure 2 shows the photographs of the overall SS-OCTA system [38,39]. We choose a light source with a wavelength of 1.3 nm. In addition, we want to obtain a higher resolution and imaging depth, so we choose a swept source from santec (HSL-20-100-B, santec, Japan), the centre wavelength of the light source is 1310 nm, the output power is 16 mW, the sweep rate is 100 kHz, the bandwidth is 80 nm and the instantaneous linewidth is 0.14 nm. In the choice of interferometer, we must consider the bandwidth of the optical fibre, the bandwidth of the optical fibre must be greater than the bandwidth of the light source has to avoid the loss of bandwidth caused by the transmission process, thus affecting the imaging resolution of the system, so with the light source, choose the central wavelength of 1310 nm, the bandwidth of the optical fibre is greater than 80 nm. The interferometer system is based on a fibre optic Mach-Zehnder interferometer structure.

The electrical signals from the balanced detector are fed to a high-speed data acquisition card (ATS9350, AlazarTech, Canada) with a real-time sampling rate of 500 MS/s per channel and a signal-to-noise ratio of up to 60.03 db. In the SS-OCT system, the role of the reference arm is to provide a reference light signal to interpolate and amplify the weak signal returned from the sample arm. The light from the reference arm fibre is converted to quasi-parallel light by a fibre collimator (PAF2A-18CFC/APC, Thorlabs, USA), passes through a slit and through a achromatic doublet lens (AC254-50-C, Thorlabs, USA) where it is focused on a mirror and then reflected back to the reference arm fibre in the original manner. The interference information is incident on the (+, −) ends of a balanced detector (PDB480C-AC,
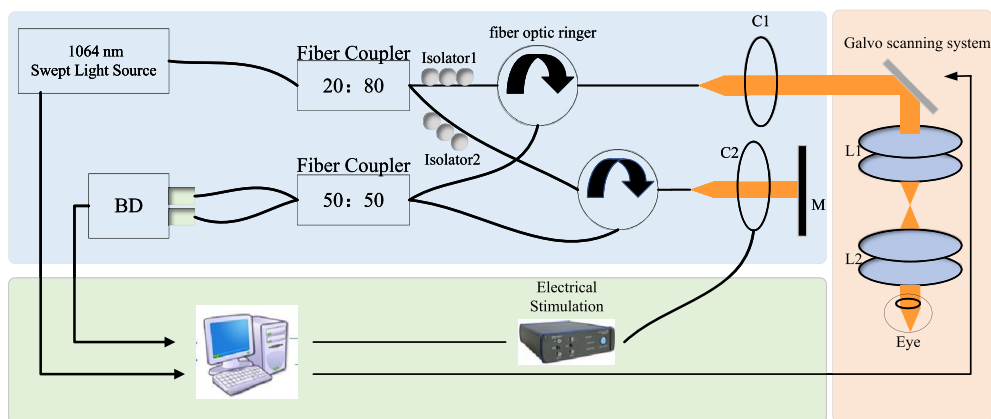
FIGURE 2 | System diagram of swept source optical coherence tomography (SS-OCT).

Thorlabs, USA) with a bandwidth of 30 kHz −1.6 GHz for photo-electric conversion.

3.2.2 | Image Construction of SS-OCT

In ophthalmic applications, en-face OCT has gradually become a useful tool for evaluating OCT angiography [40], which is an advantage of high detection sensitivity. OCT angiography utilises several repeated B-scans to record contrast of the flowing blood cells.

The OCT uses the principle of interference to obtain structural information from the tissue. The system performs multiple repetitive A-scans (i.e., depth-directed scans) at the same location to record the reflected signals from the tissue at different times. Multiple A-scans form a B-scan, or cross-sectional image, in which OCTA analyses the signal changes by repeatedly acquiring multiple B-scans to reconstruct the vascular structure in three dimensions.

3.3 | The New REVIO Data Set

3.3.1 | Data Set Statistic

REVIO data set contains images of 50 people of different ages, with 288×288 pixels, acquired with an SS-OCT system. As shown in Figure 3, there are some samples of the REVIO data set and ground truths. We constructed a database of retinal images for segmentation of blood vessels by optical coherence tomography angiography (OCTA), namely *RE*tinal *V*essels *I*mages in *O*CTA (*REVIO*) data set, which can be viewed at URL https://kyanbis.github.io/REVIO. As shown in Figure 3, there were 150 retinal vessel images from normal subjects and 150 ground truth images. The resolution of each image is 288×288 pixels.

3.3.2 | Data Labeling

OCTA images reveal the intricate structure of the retinal capillary network by capturing blood flow signals. Ground truth annotation is performed manually, as shown in Figure 1. For privacy reasons, personal details such as name and address are hidden.

4 | Method

In this section, we describe in detail our proposed SFNet, including its architecture, the present model for retinal vessel segmentation in OCTA images. Figure 4 shows the details of SFNet.

SFNet extracts retinal vascular structures using several different feature vectors containing both spatial and frequency domains. Spatial features of retinal vessels are extracted using a branch containing informative features in the spatial domain, and frequency domain features of retinal vessels are extracted using a branch containing informative features in the frequency domain, in order to obtain all the spatial domain features and frequency domain features while maintaining a high level of inference speed.

4.1 | Directional Attention-Based Spatial Domain

The spatial branch is a spatial domain-based neural network responsible for learning contextual semantics as well as fine structure and texture information in the vasculature. The shallow network contains a convolutional layer, a batch normalization (BN) layer and a rectified linear unit (ReLU) activation function to preserve all spatial detail, allowing the model to effectively detect and segment all different vascular structures. The images are fed into a shallow network for coding to obtain a feature vector f of retinal vessels. The shallow network operates with different sized convolution kernels in convolution layers, allowing the model to capture the relative position, structure and distribution of blood vessels to each other. We use the skip connections to reduce gradient vanishing, preserve the original information of the vessel feature vectors and enhance the spatial information of the retinal vessel feature vectors. Dot-wise multiplication of the feature vectors in the 1×1 convolutional layer and shallow network improves the richness of the vessel feature vectors, preserves the main information of the image
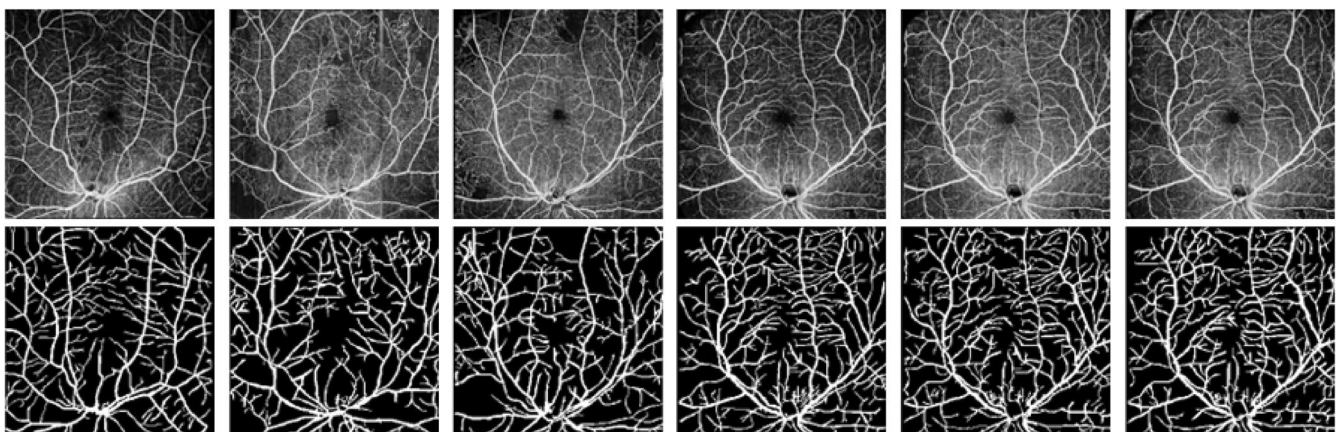
FIGURE 3 | The OCTA images and ground truths in the REVIO data set. Top Rows: The original OCTA fundus images. Bottom Rows: The ground truths of fundus images.
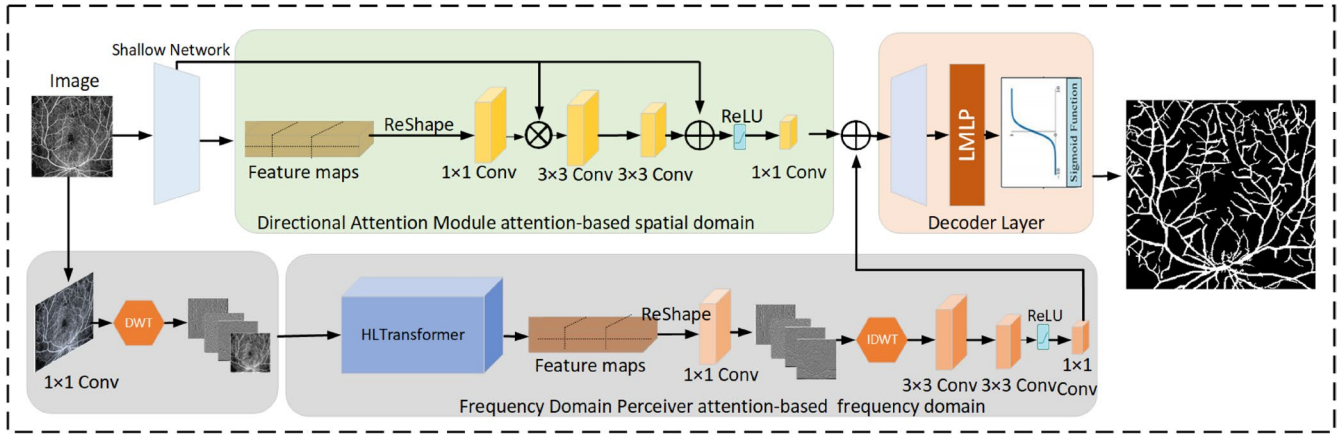
FIGURE 4 | Architecture of the proposed SFNet network. The upper branch is a directional attention module based on the spatial domain. The lower branch is the frequency domain-based perceiver attention module. It is fed to the decoding module to obtain the final fine segmentation result.

and efficiently extracts the local features of the image. The formula is as follows:

$$S_1 = U_{Sam}\left(P_{Max}\left(ReLU\left(BN\left(S_{Enc}\left(f_{i,j}\right)\right)\right)\right)\right), \quad (1)$$

where $U_{Sam}(\cdot)$ denotes upsampling function. $P_{Max}(\cdot)$ denotes max-pooling function. $RELU(\cdot)$ is the nonlinear activation function. $BN(\cdot)$ represents batch normalisation. $S_{Enc}(\cdot)$ represents shallow neural network. $f_{i,j}$ is the input feature from the original image. $i \in \{0, 1, \cdots, H-1\}$, $j \in \{0, 1, \cdots, W-1\}$. H is the height of the input feature vector. W is the width of the input feature vector.

$$S_2 = (1 + (U_{Sam}(S_1)) \otimes (S_{2Enc}). \quad (2)$$

$U_{Sam}(\cdot)$ denotes upsampling function. $\otimes$ represents multiplication operation. and S_{2Enc} denotes the second convolution function during encoding.

$$S_3 = \phi(S_1) \oplus \phi(S_2), \quad (3)$$

where $\phi(\cdot)$ is the reshape operation. $\oplus$ denotes the addition of elements.

Deeper structures are not conducive to closer attention to shallow features, which are important for vessel segmentation at the centreline level.

4.2 | Attention-Based Frequency Domain Perceiver

To further recover the continuous details of small vessels, fine-stage adaptive vessel segmentation prediction results are introduced. The frequency branch is a deep neural network responsible for learning the frequency domain features of an image. The OCTA image is fed into a 1×1 convolution kernel, which aims at obtaining a rich feature vector, preserving the rich overall features and obtaining a profile of the overall blood vessels.

In addition, we intend to preserve the details of the original image features while reducing the information loss with an operation that allows up and down sampling and have considered the Discrete Wavelet Transform (DWT). DWT is a basic time-frequency analysis method which decomposes the signal into different frequency sub-bands to solve the problem of misidentification of retinal blood vessels. The DWT has the property of multi-scale analysis, which can provide detail information at the same time, and the detail coefficients at different scales can be obtained. By decomposing the signal, the DWT distributes the signal capability in different frequency sub-bands, which helps to extract the important information of the signal. Inverse Discrete Wavelet Transform (IDWT) is an inverse transformation of a signal or image that has been subjected to a DWT, where the wavelet coefficients are recombined to recover the original signal or image. The IDWT can be used to reconstruct a signal or image from these wavelet coefficients while selectively removing the noise or unimportant details. The wavelet transform is performed on the high frequency sub-bands to obtain four discrete feature vectors, which are low frequency (image contour) feature vectors, high frequency feature vectors in horizontal, vertical and diagonal directions to obtain more detailed feature vectors, feeds into the HLTransFormer (short for High and Low frequency Transformer) module to obtain detailed features of the image. Figure 5 shows the details of the HLTransFormer. The feature vectors containing more image details are obtained, and the inverse discrete wavelet transform is applied to the four feature vectors to obtain the feature vector. The formula is as follows:

$$F_{h,w} = \sum_{i=0}^{H-1} \sum_{j=0}^{W-1} DWT(f_{i,j}). \quad (4)$$

$DWT(\cdot)$ stands for Discrete Wavelet Transform. $f_{i,j}$ is the input feature from the original image. H is the height of the input feature vector. W is the width of the input feature vector. $i, h \in \{0, 1, \cdots, H-1\}$. $j, w \in \{0, 1, \cdots, W-1\}$.

$$L_1 = IDWT\left(Conv\left(D_{Enc}\left(F_{h,w}\right)\right)\right). \quad (5)$$

$IDWT(\cdot)$ stands for Inverse Discrete Wavelet Transform. $Conv(\cdot)$ denotes convolutional operation. $D_{Enc}(\cdot)$ represents deep neural
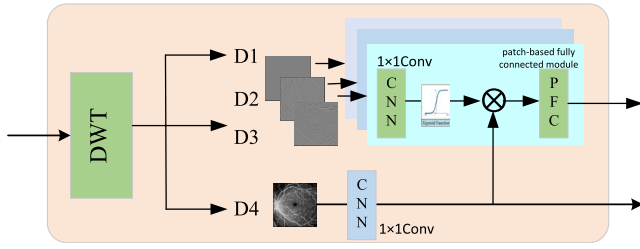
FIGURE 5 | Details of HLTransFormer module. DWT denotes Discrete Wavelet Transform. D1 is the horizontal feature vector in the high frequency information. D2 is the vertical feature vector in the high frequency information. D3 is the diagonal feature vector in the high frequency information. D4 is the low frequency (image contour) feature vector. CNN denotes convolutional operation. PFC denotes the patch-based fully connected layers.

network. The transformer introduces the mechanism of self-attention, which models the ability to efficiently deal with the dependencies between different positions in the input sequence and is able to capture long distance dependencies when dealing with sequential data for a better understanding of the context and the relationships in the sequence. We obtained the feature vector by the discrete wavelet operation. Through the proposed transformer mechanism, we re-position encoded and embedded the image feature vectors so that the more detailed features are preserved, and each feature vector is associated with the positional encoding, which makes the identification of the subtle blood vessels more accurate, and improves the segmentation capability of the model. We use Layer Normalisation to accelerate the training convergence speed of the neural network, to improve the robustness and generalisation ability of the model to the input data and to reduce the overfitting to the training data.

4.3 | Decoding Module

We use an inverse convolution operation to increase the size of the feature map. And a skip connection operation is used to enable the model to capture deeper global and local information. This eliminates the need for additional special modules, further accelerates inference and reduces memory costs, and preserves both spatial and frequency domain features. We intend that it can be used to map hidden representations from the context, allowing for multi-level feature abstraction and representation learning, which helps the model to capture different levels of abstract features of the data and to map abstract representations to comprehensible outputs, and we find that the use of the Multi-Layer Perceptron (MLP) [42,43], module, which enables the network to better adapt to complex data, can extract and learn the abstract features in the data by extracting and learning them layer-by-layer for multiple hidden layers, but it will require a lot of time and resources. Therefore, we improve the MLP module, we use a convolutional layer, on which the operation is performed, and the convolutional kernel in the convolutional neural network can be shared in multiple locations, which can reduce the number of parameters in the network, and thus shorten the training time, which we define as the Light Multi-Layer Perceptron (LMLP) in the process of decoding the operation, U (F) is subjected to fully connected operations with LMLP. Figure 6 shows detail of LMLP. In addition, each neuron performs a non-linear transformation of its input through
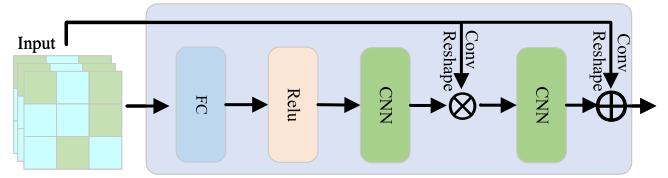
FIGURE 6 | Details of LMLP module. FC denotes fully connected layer. Relu denotes Rectified Linear Units. CNN denotes 1×1 convolutional operation.

an activation function to improve the model's ability to segment retinal vessels. In our task, instead of sharing the main network parameters, we use two separate branches to learn the spatial and frequency domain features, respectively. In the SFNet architecture, we use frequency domain to learn vascular texture features.

$$O = M(L_1, S_3). \tag{6}$$

$M(\cdot)$ denotes decoding function. L_1 represents the output of the frequency domain module. S_3 denotes the output result of the spatial domain module. O is the segmentation result.

5 | Experiment

5.1 | Data Sets and Evaluation Metrics

We perform extensive experiments on the REVIO, ROSE and OCTA-500 data set to validate SFNet. We describe the former three data sets as follows.

REVIO. The REVIO data set contains 150 OCTA images with a resolution of 288×288 pixels. With the advice and experience of an ophthalmologist, the ground truths are labeled manually. The REVIO data set was imaged using SS-OCT with a 12×12 mm single scan centered on the fovea. To be privacy, we hide personal information, such as name and age.

ROSE. The ROSE [3] data set consists of 117 OCTA images, an image resolution of 304×304 pixels. The scan area was 3×3 mm centered on the fovea, within an annular zone of 0.6–2.5 mm diameter around the foveal center.

OCTA-500. The OCTA-500 [4] data set contains a total of 500 images, which are divided into two subsets according to the field of view type: OCTA-500-6 and OCTA-500-3 mm. The OCTA-500-6 mm data set contains 300 images with an image resolution of 400×400 pixels. The OCTA-500-3 mm data set contains 200 images with an image resolution of 304×304 pixels. The OCT/OCTA images are from the same device, which is a commercial 70 kHz spectral domain OCT system with a center wavelength of 840 nm (RTVUE-XR, Optovue, CA).

5.2 | Evaluation Metrics

In order to comprehensively and objectively evaluate the segmentation performance of the proposed method, the following metrics were calculated and compared.

1. Sensitivity $= TP/(TP + FN)$.
2. Specificity $= TN/(TN + FP)$.
3. Accuracy $= (TP + TN)/(TP + TN + FP + FN)$.
4. Dice Coefficient $= 2 \times TP/(FP + FN + 2 \times TP)$.
5. Jaccard $= TP/(TP + FN + FP)$.

TP stands for True Positive which indicates the prediction is correct and the actual value is positive. FP stands for False Positive which indicates the prediction is wrong and the actual value is positive. TN stands for True Negative which indicates the prediction is correct and the actual value is negative. FN stands for False Negative which indicates the prediction is wrong and the actual value is negative.

5.3 | Performance Comparison and Analysis

We have evaluated the proposed method on our REVIO data set and compared it to existing state-of-the-art segmentation methods to demonstrate the superiority of our SFNet in the segmentation of OCTA microvasculature.

5.3.1 | Comparison Methods

In order to verify the superiority of our method, we compared our method with other state-of-the-art segmentation methods on the REVIO, ROSE and OCTA-500 data sets, including six deep-learning approaches: U-Net [19], UNet++ [26], SemiGanNet [27], CENet [20], ResAt-UNet [41] and DenseUNet+ [29]. The parameters were tuned to achieve segmentation results of all vessels in REVIO, ROSE and OCTA-500 data sets. For deep-learning methods, all hyper-parameters were manually adjusted to achieve the best performances.

5.3.2 | Subjective Comparisons

Figure 7 shows the respective vessel segmentation results produced by the proposed method and the other three state-of-the-art segmentation networks. We can see that U-Net [19] achieves relatively low performance, due to its over-segmentation at regions with high density. CENet [20] and ResAt-UNet [41] achieve better performance than U-Net. However, they are not able to preserve fine capillaries well in terms of producing weak vessel responses. In contrast, the proposed method yields more visually informative results. The benefit of the proposed method for segmentation can be observed from the representative regions (red patches). It is clear from visual inspection that our method has identified more complete vessels.

In contrast, all methods achieve very similar segmentation performance in OCTA-500-6 mm. Therefore, to better evaluate the performance of the proposed method, we provide quantitative results in the following subsections.

5.3.3 | Performance on the REVIO Data Set

We will first evaluate the vessel segmentation performance on each of the plexus layers of the REVIO data set. Table 2 quantifies the segmentation performance on REVIO data set of the different approaches. Overall, our method achieves the best performance in terms of almost all the metrics. The proposed method is able to correctly identify the majority of vessels using our architecture. It significantly outperforms the other methods on the REVIO data set, with an increase in Sensitivity and Dice of approximately 8% and 4%, respectively, and an increase in specificity of approximately 0.21% compared to CENet.
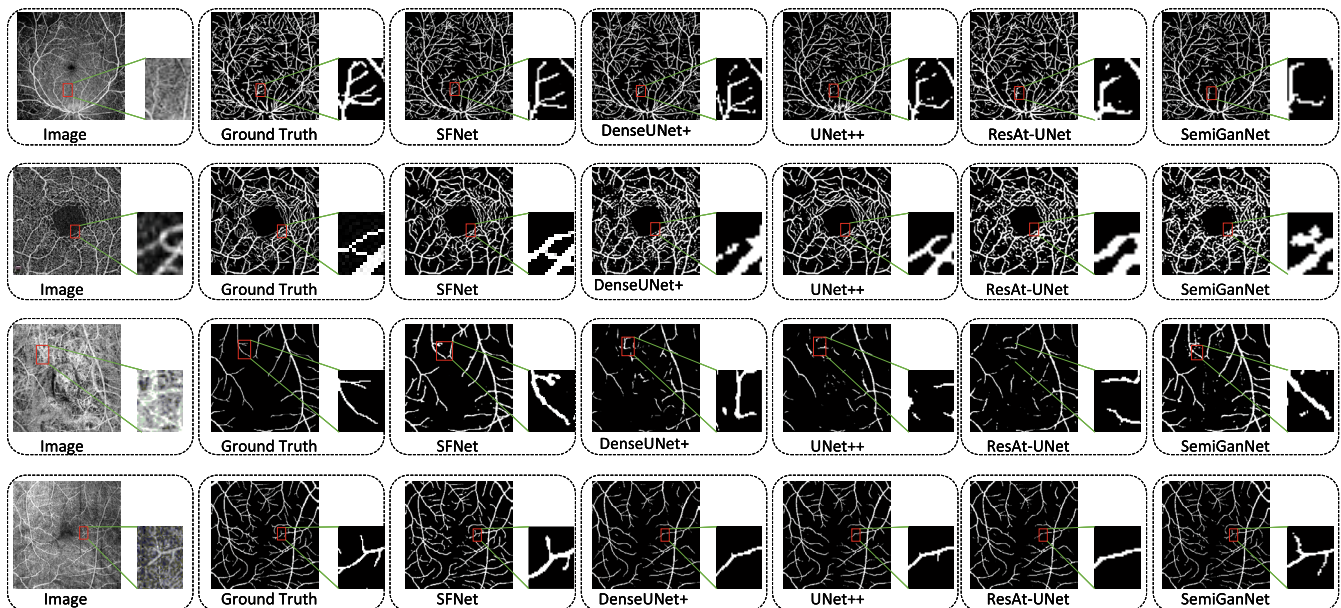
FIGURE 7 | Vessel segmentation results of different methods on REVIO, ROSE, OCTA-500-3 mm and OCTA-500-6 mm data sets. From left to right: En face angiograms (original images), manual annotations, vessel segmentation results obtained by SFNet, DenseUNet+, UNet++, ResAt-UNet and SemiGanNet, respectively.

5.3.4 | Performance on the ROSE Data Set

Table 3 demonstrates the segmentation results achieved by our method and the state-of-the-art methods. Although the improvement on AUC is not significant when compared with CENet and SemiGanNet, the proposed network outperformed all the other compared approaches. In particular, it significantly outperforms other methods by a large margin. These performance improvements are consistent with the segmentation results shown in the middle row of Figure 7, where the proposed method successfully extracts small capillaries from macula regions with promising continuity and integrity, while other methods produce relatively lower capillary responses. On the ROSE data set, it significantly outperforms the other methods, in particular with an increase in Sensitivity of about 10% and a decrease in Specificity of about 5% compared to SemiGanNet.

5.3.5 | Performance on the OCTA-500-3 mm and OCTA-500-6 mm Data Set

Tables 4, and 5 demonstrate the segmentation results achieved by our method and the state-of-the-art methods. In OCTA-500-3 and OCTA-500-6 mm data sets, our method uses a spatial domain-based directional perception attention module and a frequency domain-based frequency domain perception module to segment the blood vessels and obtain the final segmentation results. In OCTA-500-6 mm data set, the segmentation metrics are approximate compared to the state-of-the-art methods, both of which are able to identify the vessel

endings completely. On the OCTA-500-3 mm data set, there is a decrease in sensitivity of approximately 1% and an increase in Dice of approximately 2% compared to CENet. On the OCTA-500-6 mm dataset, there is a decrease in sensitivity of approximately 4% and an increase in Dice of approximately 2% compared to SemiGanNet.

5.4 | Experiments Setup

The experiment was conducted on windows 10 and GeForce RTX 2080Ti with 11GB. The learning rate was set to 0.001. The epochs of training model were set to 350.

5.5 | Ablation Experiment

Spatial domain directional attention-based spatial domain is used to extract the texture structure of the blood vessel branches by using the spatial domain of the image for the image features. Frequency domain perceptual attention-based frequency domain is used to extract the fine structure at the end of the blood vessel by using the frequency domain of the image for the frequency domain features of the image. In this subsection, we investigate the effectiveness of the proposed directional attention-based spatial domain and frequency domain perceiver attention-based frequency domain through two experiments. Then, the structure of the SFNet is further analysed by further ablation experiments. The experiments are tested on the REVIO data set. Table 6 shows ablation experiments on the REVIO data set.

TABLE 2 | Results on the REVIO data set (%). Best results are in bold.

Model	Sensitivity	Specificity	Dice	Accuracy	Jaccard
U-Net	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
SFNet (ours)	**88.26**	94.86	**83.85**	**93.67**	**75.66**

TABLE 3 | Results on the ROSE data set (%). Best results are in bold.

Model	Sensitivity	Specificity	Dice	Accuracy	Jaccard
U-Net	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
SFNet (ours)	**78.59**	93.71	**76.7**	90.49	**63.71**

TABLE 4 | Results on the OCTA-500-3 mm data set (%). Best results are in bold.

Model	Sensitivity	Specificity	Dice	Accuracy	Jaccard
U-Net	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
SFNet (ours)	88.26	97.64	**82.85**	97.42	**70.83**

TABLE 5 | Results on the OCTA-500-6 mm data set (%). Best results are in bold.

Model	Sensitivity	Specificity	Dice	Accuracy	Jaccard
U-Net	78.58	98.52	81.43	96.81	69.08
UNet++	78.9	98.79	82.26	96.84	70.23
SemiGanNet	**89.38**	98.68	81.65	81.65	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
SFNet (ours)	85.31	98.02	**83.28**	**96.86**	**71.43**

TABLE 6 | Ablation on the REVIO data set (%). Best results are in bold.

Model	Sensitivity	Specificity	Dice	Accuracy	Jaccard
SFNet/LF	78.17	89.76	71.17	87.61	56.37
SFNet/LS	87.06	91.1	78.59	90.83	66.87
SFNet (ours)	**88.26**	**94.86**	**83.85**	**93.67**	**75.66**

5.5.1 | Effective of Directional Attention-Based Spatial Domain

We train another segmentation model by removing the frequency domain perceiver attention-based frequency domain from the proposed SFNet, and we denote this new model as SFNet/LF. Compared with the SFNet performance, there is a significant performance degradation. Table 6 shows ablation experiments on the REVIO data set, Sensitivity reduced by approximately 10%, Specificity reduced by approximately 5% and Dice reduced by approximately 12%. Specifically, after removing the frequency domain perceiver attention-based frequency domain, the metrics decrease, respectively. The phenomenon indicates that directional attention-based spatial domain, as a key module in SFNet, plays a key role in achieving its good segmentation performance, and also verifies that directional attention-based spatial domain is useful for extracting the image spatial domain features. The spatial domain is able to locate the position of a blood vessel, identify the shape, size and position of the vessel, and perform vessel segmentation.

5.5.2 | Effective of Frequency Domain Perceiver Attention-Based Frequency Domain

We learn another segmentation model by removing the directional attention-based spatial domain, and we notate this new model as SFNet/LS. Comparing with SFNet performance, there is a significant performance degradation. Specifically, the metrics drop after removing the Directional attention-based spatial domain, respectively. Table 6 shows ablation experiments on the REVIO data set, Sensitivity reduced by approximately 1%, Specificity reduced by approximately 3% and Dice reduced by approximately 5%. The experimental results indicate that directional attention-based spatial domain in SFNet plays a positive role in achieving its good segmentation performance, and also verifies that directional attention-based spatial domain is useful for extracting image frequency domain features. The frequency domain allows the decomposition of feature vectors into feature vectors of different frequencies, which allows for the selective removal of noise and the completion of vessel segmentation in the frequency domain.

TABLE 7 | The summary of required Gflops, params and time for model.

Model	GFLOPS	Params (M)	Time (s)
SFNet/LF	9	807	804
SFNet/LS	11	1514	836
SFNet (ours)	17	2197	982

We introduce Floating Point Operations (FLOPs), model parameters and training time as validation metrics for the ablation experiments. FLops measures the total number of floating point calculations performed by the model. The number of parameters is the total number of parameters to be learned in the model, including weights and bias terms. Training Time is the time from the start of model training to the end of training.

$$\text{Total FLOPs} = \sum_{l=1}^{L} FLOPs_l, \tag{7}$$

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

$$\text{Total parameters} = \sum_{l=1}^{L} P_l, \tag{8}$$

where total parameters denotes total number of parameters. P_l is the number of parameters in the lth layer.

Table 7 shows that the number of branching parameters in the spatial domain is 807 M, the number of branching parameters in the frequency domain is 1514 M, and the number of SFNet parameters is 2197 M. The number of SFNet parameters increases by 1380 M for the spatial domain branches and 683 M for the frequency domain branches. The flops of SFNet increases by 8 GFLOPS over the spatial domain branch and 6 GFLOPS over the frequency domain branch. The training time of the SFNet model increases by 178 s over the spatial domain branch and 146 s over the frequency domain branch. We can clearly see that as the number of parameters increases, the flops and time of the model increase.

6 | Conclusion

In this paper, we constructed a wild-field retinal OCTA segmentation dataset, the Retinal Vessels Images in OCTA (REVIO) data set. We proposed a novel segmentation network based on spatial domain and frequency domain network (SFNet). The experimental results show that the proposed SFNet exhibits better performance in segmentation tasks with the same number of network parameters compared to existing approaches. Future studies will use detailed information about the retinal vasculature to help doctors detect lesions such as vascular abnormalities, microangiomas or capillary occlusions at an early stage, allowing more precise quantification of the extent and progression of the lesion, and thus timely therapeutic intervention.

Ethics Statement

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).

Conflicts of Interest

The authors declare no conflicts of interest.

Data Availability Statement

The data that support the findings of this study are available from the corresponding author upon reasonable request.

References

1. L. Bottazzi, A. Montesel, and G. Querques, "Optical Coherence Tomography Angiography and Corresponding Histology," *JAMA Ophthalmology* 141, no. 4 (2023): 403–404.

2. X. Li, R. Bala, and V. Monga, "Robust Deep 3D Blood Vessel Segmentation Using Structural Priors," *IEEE Transactions on Image Processing* 31 (2022): 1271–1284.

3. Y. Ma, H. Hao, J. Xie, et al., "ROSE: A Retinal OCT-Angiography Vessel Segmentation Dataset and New Model," *IEEE Transactions on Medical Imaging* 40, no. 3 (2021): 928–939.

4. M. Li, K. Huang, Q. Xu, et al., "OCTA-500: A Retinal Dataset for Optical Coherence Tomography Angiography Study," *Medical Image Analysis* 93 (2024): 103092.

5. S. Pang, A. Du, M. A. Orgun, et al., "Beyond CNNs: Exploiting Further Inherent Symmetries in Medical Image Segmentation," *IEEE Transactions on Cybernetics* 53, no. 11 (2023): 6776–6787.

6. H. Wu, J. Pan, Z. Li, Z. Wen, and J. Qin, "Automated Skin Lesion Segmentation via an Adaptive Dual Attention Module," *IEEE Transactions on Medical Imaging* 40, no. 1 (2021): 357–370.

7. F. Ma, S. Li, S. Wang, et al., "C Deep-Learning Segmentation Method for Optical Coherence Tomography Angiography in Ophthalmology," *Journal of Biophotonics* 17 (2023): e202300321.

8. M. Abdar, F. Pourpanah, S. Hussain, et al., "A Review of Uncertainty Quantification in Deep Learning: Techniques, Applications and Challenges," *Information Fusion* 76 (2021): 243–297.

9. X. Chen, X. Wang, K. Zhang, et al., "Recent Advances and Clinical Applications of Deep Learning in Medical Image Analysis," *Medical Image Analysis* 79 (2022): 102444.

10. S. Roger, M. Cobos, C. Botella-Mascarell, and G. Fodor, "Fast Channel Estimation in the Transformed Spatial Domain for Analog Millimeter Wave Systems," *IEEE Transactions on Wireless Communications* 20, no. 9 (2021): 5926–5941.

11. B. He, F. Zhang, H. Zhang, and R. Han, "A Hybrid Frequency-Spatial Domain Model for Sparse Image Reconstruction in Scanning Transmission Electron Microscopy," *2021 IEEE/CVF International Conference on Computer Vision* (*ICCV*) (Montreal, QC, Canada: 2021), 2662–2671.

12. L. Wang, Y. Wang, L. Wang, Y. Zhan, Y. Wang, and H. Lu, "Can Scale-Consistent Monocular Depth be Learned in a Self-Supervised Scale-Invariant Manner?," *IEEE/CVF International Conference on Computer Vision* (*ICCV*) (Montreal, QC, Canada: 2021), 12716–12727.

13. C. Wang, J. Wu, A. Fang, Z. Zhu, P. Wang, and H. Chen, "An Efficient Frequency Domain Fusion Network of Infrared and Visible Images," *Engineering Applications of Artificial Intelligence* 133 (2024): 108013.

14. B. Pu, J. Liu, Y. Kang, J. Chen, and P. S. Yu, "MVSTT: A Multiview Spatial-Temporal Transformer Network for Traffic-Flow Forecasting," *IEEE Transactions on Cybernetics* 54, no. 3 (2024): 1582–1595.

15. W. Bai, T. Li, Y. Long, and C. L. P. Chen, "Event-Triggered Multigradient Recursive Reinforcement Learning Tracking Control for Multiagent Systems," *IEEE Transactions on Neural Networks and Learning Systems* 34, no. 1 (2023): 366–379.

16. J. Ahn, S. Jeong, C. Cho, and H. Ryoo, "DC Series Arc Fault Detector With Differential DWT and Variable Threshold Method for Photovoltaic Systems," *IEEE Transactions on Industrial Informatics* 20 (2024): 1–9.

17. G. Wang, Z. Wang, K. Gu, L. Li, Z. Xia, and L. Wu, "Blind Quality Metric of DIBR-Synthesized Images in the Discrete Wavelet Transform Domain," *IEEE Transactions on Image Processing* 29 (2020): 1802–1814.

18. S. Minaee, Y. Boykov, F. Porikli, A. Plaza, N. Kehtarnavaz, and D. Terzopoulos, "Image Segmentation Using Deep Learning: A Survey," *IEEE Transactions on Pattern Analysis and Machine Intelligence* 44, no. 7 (2020): 3523–3542.

19. O. Ronneberger, P. Fischer, and T. Brox, "U-Net: Convolutional Networks for Biomedical Image Segmentation," *Medical Image Computing and Computer-Assisted Intervention* 9351 (2015): 234–241.

20. H. Tao, C. Xie, J. Wang, and Z. Xin, "CENet: A Channel-Enhanced Spatiotemporal Network With Sufficient Supervision Information for Recognizing Industrial Smoke Emissions," *IEEE Internet of Things Journal* 9, no. 19 (2022): 18749–18759.

21. J. Guo, Q. Zhang, Y. Zhao, H. Shi, Y. Jiang, and J. Sun, "RNN-Test: Towards Adversarial Testing for Recurrent Neural Network Systems," *IEEE Transactions on Software Engineering* 48, no. 10 (2022): 4167–4180.

22. Z. W. Gu, J. Cheng, H. Z. Fu, et al., "CE-Net: Context Encoder Network for 2D Medical Image Segmentation," *IEEE Transactions on Medical Imaging* 38 (2019): 2281–2292.

23. Z. Li, F. Liu, W. Yang, S. Peng, and J. Zhou, "A Survey of Convolutional Neural Networks: Analysis Applications, and Prospects," *IEEE Transactions on Neural Networks and Learning Systems* 33, no. 12 (2022): 6999–7019.

24. X. Deng and P. L. Dragotti, "Deep Convolutional Neural Network for Multi-Modal Image Restoration and Fusion," *IEEE Transactions on Pattern Analysis and Machine Intelligence* 43, no. 10 (2021): 3333–3348.

25. M. Li, Y. Chen, Z. Ji, et al., "Image Projection Network: 3D to 2D Image Segmentation in OCTA Images," *IEEE Transactions on Medical Imaging* 42, no. 1 (2023): 329.

26. Z. Zhou, M. M. R. Siddiquee, N. Tajbakhsh, and J. Liang, "UNet++: Redesigning Skip Connections to Exploit Multiscale Features in Image Segmentation," *IEEE Transactions on Medical Imaging* 39, no. 6 (2020): 1856–1867.

27. M. Wang, W. F. Zhu, F. Shi, et al., "MsTGANet: Automatic Drusen Segmentation From Retinal OCT Images," *IEEE Transactions on Medical Imaging* 41, no. 2 (2022): 394–406.

28. M. C. Li, Y. R. Chen, Z. X. Ji, et al., "Image Projection Network: 3D to 2D Image Segmentation in OCTA Images," *IEEE Transactions on Medical Imaging* 39, no. 11 (2022): 3343–3354.

29. H. Çetiner and S. Metlek, "DenseUNet+: A Novel Hybrid Segmentation Approach Based on Multi-Modality Images for Brain Tumor Segmentation," *Journal of King Saud University, Computer and Information Sciences* 35, no. 8 (2023): 101663.

30. K. Han, Y. Wang, H. Chen, et al., "A Survey on Vision Transformer," *IEEE Transactions on Pattern Analysis and Machine Intelligence* 45, no. 1 (2023): 87–110.

31. H. Wu, B. Xiao, N. Codella, et al., "CvT: Introducing Convolutions to Vision Transformers," *2021 IEEE/CVF International Conference on Computer Vision* (*ICCV*) (Montreal, QC, Canada: 2021), 22–31.

32. H. Touvron, M. Cord, A. Sablayrolles, G. Synnaeve, and H. Jégou, "Going Deeper With Image Transformers," *2021 IEEE/CVF International Conference on Computer Vision* (*ICCV*) (Montreal, QC, Canada: 2021), 32–42.

33. A. D. Hoover, V. Kouznetsova, and M. Goldbaum, "Locating Blood Vessels in Retinal Images by Piecewise Threshold Probing of a Matched Filter Response," *IEEE Transactions on Medical Imaging* 19, no. 3 (2000): 203–210.

34. P. Prasanna, P. Samiksha, K. Ravi, et al., "Indian Diabetic Retinopathy Image Dataset (Idrid): A Database for Diabetic Retinopathy Screening Research," *Data* 3, no. 3 (2018): 25.

35. J. Staal, M. D. Abramoff, M. Niemeijer, M. A. Viergever, and B. V. Ginneken, "Ridge-Based Vessel Segmentation in Color Images of the Retina," *IEEE Transactions on Medical Imaging* 23, no. 4 (2004): 501–509.

36. A. Budai, R. Bock, A. Maier, J. Hornegger, and G. Michelson, "Robust Vessel Segmentation in Fundus Images," *International Journal of Biomedical Imaging* 2013 (2013): 1–11.

37. C. G. Owen, A. R. Rudnicka, C. M. Nightingale, et al., "DeepVessel: Retinal Vessel Segmentation via Deep Learning and Conditional Random Field," *Arteriosclerosis, Thrombosis, and Vascular Biology* 31 (2011): 1933–1938.

38. R. Kong, W. Wu, H. Zhao, et al., "Imaging Depth Extension of OCT by Optical Clearing Method In Vitro Rabbit Eye," *Optics in Health Care and Biomedical Optics IX* (2019).

39. I. Lains, J. C. Wang, Y. Cui, et al., "Retinal Applications of Swept Source Optical Coherence Tomography (OCT) and Optical Coherence Tomography Angiography (OCTA)," *Progress in Retinal and Eye Research* 84 (2021): 100951.

40. L. Hung-Chang, L. Miao-Hui, T. Ching-Heng, W. Yi-Min, and S. Chia-Wei, "Intraoperative Application of Optical Coherence Tomography for Lung Tumor," *Journal of Biophotonics* 16 (2023): e202200344.

41. Z. Fan, Y. Liu, M. Xia, J. Hou, F. Yan, and Q. Zang, "ResAt-UNet: A U-Shaped Network Using ResNet and Attention Module for Image Segmentation of Urban Buildings," *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing* 16 (2023): 2094–2111.

42. C. Reiser, S. Peng, Y. Liao, and A. Geiger, "KiloNeRF: Speeding Up Neural Radiance Fields With Thousands of Tiny MLPs," *2021 IEEE/CVF International Conference on Computer Vision* (*ICCV*) (Montreal, QC, Canada: 2021), 14315–14325.

43. X. Ding, H. Chen, X. Zhang, J. Han, and G. Ding, "RepMLPNet: Hierarchical Vision MLP With re-Parameterized Locality," *2022 IEEE/CVF Conference on Computer Vision and Pattern Recognition* (*CVPR*) (New Orleans, LA, USA: 2022), 568–577.