RESEARCH ARTICLE

Deep-learning segmentation method for optical coherence tomography angiography in ophthalmology

Fei Ma[1] | Sien Li[1] | Shengbo Wang[1] | Yanfei Guo[1] | Fei Wu[2] |
Jing Meng[1] | Cuixia Dai[3]

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

[2]School of Automation, Nanjing University of Posts and Telecommunications, Jiangsu, China

[3]College Science, Shanghai Institute of Technology, Shanghai, China

Correspondence
Fei Ma, School of Computer Science, Qufu Normal University, Shandong, China.
Email: mafei0603@163.com

Cuixia Dai, College Science, Shanghai Institute of Technology, Shanghai, China.
Email: sdadai7412@163.com

Abstract

Purpose: The optic disc and the macular are two major anatomical structures in the human eye. Optic discs are associated with the optic nerve. Macular mainly involves degeneration and impaired function of the macular region. Reliable optic disc and macular segmentation are necessary for the automated screening of retinal diseases.


Methods: A swept-source OCTA system was designed to capture OCTA images of human eyes. To address these segmentation tasks, first, we constructed a new *Optic Disc* and *Macula* in fundus *Image* with optical coherence tomography angiography (OCTA) dataset (*ODMI*). Second, we proposed a *Coarse* and *Fine Attention*-Based *Net*work (*CFANet*).

Results: The five metrics of our methods on ODMI are 98.91%, 98.47%, 89.77%, 98.49%, and 89.77%, respectively.

Conclusions: Experimental results show that our CFANet has achieved good performance on segmentation for the optic disc and macula in OCTA.

KEYWORDS
deep learning network, medical image segmentation, OCTA fundus database
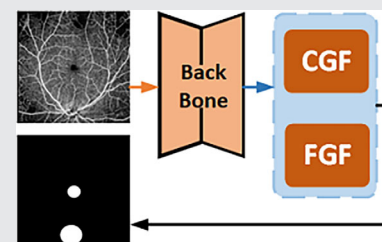
1 | INTRODUCTION

The optic disc [1–3] and macular segmentation technologies not only help ophthalmologists diagnose and screen retinal diseases but also help researchers study the pathogenesis and genetic factors of eye diseases. Figure 1A shows the structure of human retina [4], which includes some tissues, such as optic disc and macula. Optic disc segmentation can be utilized to calculate the disc geometrical parameters automatically [5], and help the doctors have a wide view in detail for some retinal diseases, such as glaucoma. Medical image segmentation is a key technique in the field of computer vision and it plays a crucial role in many applications [6]. Recently, the Convolutional Neural Network (CNN) [7] and Recurrent Neural Network (RNN) [8] have made significant breakthroughs in understanding vision and language, making it
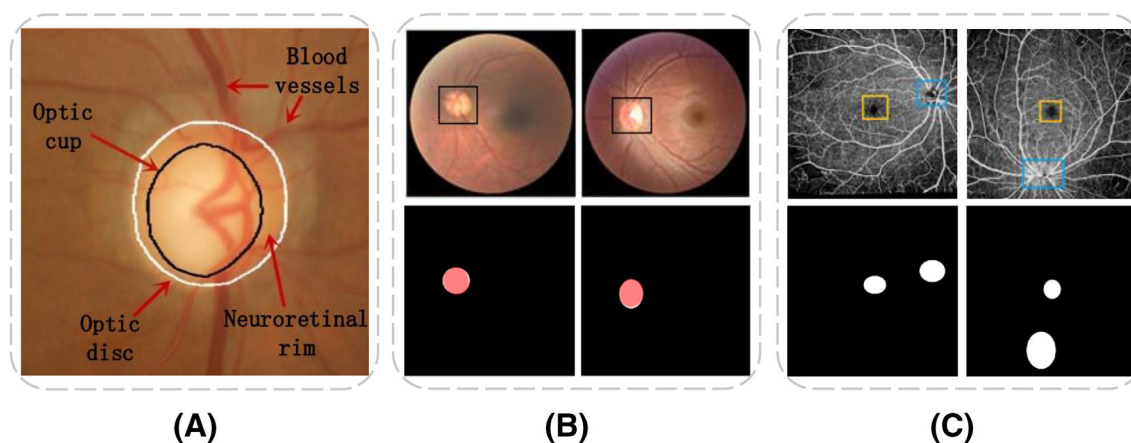
FIGURE 1 The retinal structure and illustration of the optic disc and macular segmentation. (A) Different tissues of retina. (B) The optic discs and segmentation in color fundus images. (C) The original images with OCTA and ground truths for optic disc and macula.

possible to handle more complex pixel-level cross-modal prediction tasks [9, 10]. Deep-learning methods have achieved great success in medical image segmentation [11, 12]. However, methods of medical image segmentation still face many challenges [13], such as boundary blurring and low contrast. Therefore, this paper aims to design a new image segmentation method based on attention mechanism [14] to handle these problems.

In recent years, both Full Convolutional Networks (FCN) [15] and UNet [16] have achieved impressive results in image segmentation. Both of these architectures benefit from an encoder–decoder structure, where skip connections [17] allow the fusion of semantic features on the encoder and fine-grained features in the decoder.

The main contributions of this study are summarized as follows:

1. We construct a new *Optic Disc and Macula in fundus Image* with optical coherence tomography angiography (OCTA) dataset (*ODMI*), which will be helpful for further researches.
2. We propose a novel framework, namely a Coarse and Fine Attention-based Network (CFANet). CFANet can integrate the coarse and fine features to segment the optic disc and macula.
3. We design the Coarse-Grained Feature extraction module (CGF), which can extract the coarse features in retinal images. We propose a Fine-Grained Feature extraction module (FGF), which can extract the fine features in retinal images.
4. Experimental results show that our framework is effective and achieves the best results on the ODMI dataset.

The remainder of this paper is organized as follows. Section 2 reviews relevant literature and introduces existing methods for image segmentation. We describe the details of our proposed CFANet in Section 4. Section 3 introduces our OCTA dataset ODMI. In Section 5, extensive experiments are conducted on the ODMI dataset. Finally, Section 6 draws the conclusion of this study.

2 | RELATED WORK

In this section, we review the common imaging technologies, and describe the related works of optic disc and macular segmentation.

2.1 | Popular imaging technologies

2.1.1 | Review of common imaging technologies

Today, there exist several popular imaging technologies [18] for people exam in Table 1, including x-rays, nuclear magnetic resonance (NMR), ultrasound, positron emission tomography (PET), magnetic resonance imaging (MRI), and OCT. Table 1 shows the detailed information on these imaging technologies.

An x-ray device uses the x-rays beams (electromagnetic radiation) through the soft tissue to capture photos of dense material inside the body, such as the bone, chest, abdomen, and other parts of people's body. MRI machine can capture the inside of the body with outstanding clarity, which is helpful for medical professionals to diagnose the injuries of organs. Diagnostic

TABLE 1   The common imaging technology.

Imaging modality	Resolution	Damage	Speed	Tasks	Cost
X-ray	~2–3 mm	Destructive	Fast	Lung, brain	High
Nuclear magnetic resonance (NMR)	0.1–1 mm	Destructive	Slow	Cardiovascular, orthopedic	High
Ultrasound	~500 μm	None	Fast	Cardiovascular, urinary system	Low
Positron emission tomography (PET)	−4 to 5 mm	Little	Slow	Cerebrovascular, brain	High
Magnetic resonance imaging (MRI)	~2–3 mm	Destructive	Slow	Organs and skeletal muscles	High
OCT	~1–15 μm	None	Fast	Ophthalmology	Low

ultrasound device utilizes the sound waves to generate images of structures inside of the human body. PET takes photos with a radioactive tracer that is injected into the human body.

2.1.2 | Optical coherence tomography

Recently, optical coherence tomography (OCT) is developed to generate photos of human organs in vivo [19], which is a non-invasive diagnostic technique. OCT utilizes a concept of interferometry to generate scans of the retina, whose resolution can reach within 15 μm. OCT was introduced since 1991 and has been extended for many applications outside of ophthalmology. Today, OCT has popularly been used as an ophthalmic diagnostic tool.

OCT contains the following types of technologies, that is, Time Domain OCT (TD-OCT), Spectral Domain OCT (SD-OCT), and Swept Source OCT (SS-OCT). Time domain OCT systems has been widely used to acquire images at a rate of 40 000 axial scans per second with an axial resolution within 15μm. TD-OCT can capture the improved image quality and minimize motion artifacts.

SD-OCT technology gets approximately 20 000–40 000 A-scans per second. The high scan rate can diminish the motion artifacts, which can capture high-quality photos. SD-OCT systems can improve the photos by averaging multiple B-scans at the same location of organ. Most TD-OCT systems get around 15 microns resolution, while the newer SD-OCT systems can achieve 3-micron resolution. SD-OCT utilizes a broad range and continuous light source, such as super luminescent diode laser. SS-OCT systems often utilize the light source with 1000-nm wavelength and scan speeds above 100 kHz. SS-OCT offers the advantages in visualization of the choroid and sclera in fundus.

As mentioned above, OCT technologies have been used to capture non-invasive OCT angiography (OCTA) images. OCTA is a powerful tool to visualize the microvasculature plexus of macula and optic nerve head.

2.2 | Overview of the optic disc and macula

The optic disc and the macula are two tissues of the eye related to vision in Figure 1A. Doctors can measure the optic disc, peripapillary retinal nerve fiber layer, and macula parameters by OCTA, which can help people with healthy controls and early treatment. The optic disc is an elevation on the medial aspect of the retina. The macula helps us understand specific details like objects and colors. The optic disc and the macula change with age [20], which poses a challenge for segmentation. There exist some associations between the optic disc and macula, which can be described by their spatial location. The optic disc is a circular area at the base of the eye located on the retina, and the macula is a central area of the retina located closer to the optic disc. Their morphology and position may vary considerably in different fundus images, which will be difficult to segment the optic disc and the macula. Therefore, there is an urgent need to design more advanced methods with higher accuracy and robustness.

Examination of the optic disc and the macula is of great importance for ophthalmic diagnosis. Morphological and color changes in the optic disc can reflect a number of ocular diseases, such as glaucoma and retinopathy. Changes in the macular region can be associated with diseases such as age-related macular degeneration.

2.3 | Methods for optic disc and macular segmentation

CNN-based segmentation methods are currently considered a powerful solution for the optic disc and macular segmentation tasks. The existing image segmentation methods are divided into two categories, namely general segmentation methods and specific segmentation methods. The general segmentation approaches contain the typical neural networks for multiple medical images, such as CNN, especially variants of FCN and UNet, which have shown excellent performance in medical

image segmentation. Specifically, UNet is a classical image segmentation network, which was proposed by Ronneberger et al. in 2015. It is widely used for medical image segmentation tasks. UNet employs an encoder–decoder structure to efficiently capture contextual information in images and generate fine-grained segmentation results [21, 22]. The overall architecture of UNet is similar to a symmetric U-shape with skip connections. This helps to fuse information between low-level and high-level features [23–25]. More contextual and detailed information are extracted for the segmentation task. With skip connections between the encoder and decoder, UNet is able to retain both global contextual information and local detail information. And it performs well in tasks such as medical image segmentation [26]. It has achieved excellent performance on many medical image segmentation datasets. It has also become a benchmark model for medical image segmentation.

UNet++ [27] is designed with a series of nested dense skip connections, reducing the semantic gap between the encoder and decoder. UNet++ is designed with a new Attention Gate (AG) mechanism. The model can focus on scale-different targets and suppress irrelevant features. UNet is also designed with a ResNet [17, 28] skip connections scheme and a weighted attention mechanism, which can improve the performance of retinal vessel segmentation. KiU-Net [29] is designed with a new architecture using under-complete and overcomplete features to improve the segmentation of small anatomical structures. DoubleU-Net [30] is a strong baseline for medical image segmentation using two UNet sequences and employing a spatial pyramid pool module. FANet combines the previous epoch mask unified with the current epoch feature map during training. It is worth noting that all these methods are still based on CNN, therefore they lack the ability to establish remote dependencies and global contextual connectivity. Although some studies attempt to model remote dependencies for convolution, they still encounter significant limitations when modeling contextual dependencies.

DeepLab [31] includes a series of deep-learning-based image segmentation methods. It employs cavity convolution and multi-scale contextual information to capture features. It improves the accuracy and detail retention of segmentation. Techniques such as spatial pyramid pooling [32–34] and deformable convolution are further introduced. There also exist some challenges and disadvantages in multi-scale and deformable convolution operations.

In summary, CNN-based optic disc segmentation methods are more advantageous than those traditional methods. However, to improve the efficiency of the multitask model by exploiting the association of the optic disc and the macula, we design a novel CFANet to mine the relationships between fine features and coarse features.

3 | DATASETS

3.1 | Existing datasets

In this paper, we construct a new *Optic Disc* and *Macula* in fundus *Image* with optical coherence tomography angiography (OCTA) dataset (*ODMI*) for segmentation tasks. In this section, we describe the information on the newly collected ODMI dataset and the existing DR datasets in detail. Table 2 lists the information on the common fundus databases. Most existing retinal datasets have image-level annotations. Several DR datasets also have pixel-level annotations for lesions.

1. STARE: The STARE (Structured Analysis of the Retina) dataset [35] was proposed in 2000 for vascular segmentation of fundus images, with a total of 400 images, and 20 hand-labeled vascular segmentation images (700 × 605) were selected for comparison of experimental results.

2. ORIGA: The IMED team presented ORIGA [36] presented in 2010 from the team for glaucoma detection and segmentation of the optic disc and cup. The dataset contains 650 color fundus images with 168 glaucomatous and 482 non-glaucomatous images.

3. HRF: The HRF (Hyperreflective foci) dataset [37] presented in 2013 is used to describe the high reflection points or circular lesions in the retinal layer visible through Optical coherence tomography, and it is used for diabetes Macular edema, retinal vascular occlusion and age-related macular disease. There are 45 images in total, including 15 healthy images, 15 images with diabetes and 15 images with glaucoma.

4. SCES: The SCES dataset [38] was used for the diagnosis of glaucoma and contains 1676 images, 46 of which are labeled with glaucoma in 2015. This group was to judge the associated risk factors for glaucoma of participants in Singapore. The participants were got by several examinations, including slitlamp ophthalmic and dilated retinal examinations.

5. RC-RGB-MA: The RC-RGB-MA dataset presented in 2018 [39] is used to diagnose retinal microaneurysms. It contains 250 images with image resolution of 2594 × 1944. The retinal microaneurysms were annotated by using a specialized annotation tool.

6. IDRiD: The IDRiD (Indian diabetic retinopathy) dataset [40] is used to diagnose the extent of diabetic retinopathy and macular edema in 2018. The IDRiD dataset contains 81 labeled DR images and 164 images

TABLE 2 Information of representative fundus datasets.

Dataset	Images	Tasks	Image size	Modality	Year
STARE	400	Segmentation	605×700	RGB (24 bits)	2000
ORIGA	650	Detection and segmentation	3072×2048	RGB (24 bits)	2010
HRF	45	Diagnosis	3504×2336	RGB (24 bits)	2013
SCES	1676	Diagnosis	3072×2048	RGB (24 bits)	2015
RC-RGB-MA	250	Diagnosis	2594×1944	RGB (24 bits)	2018
IDRiD	516	Segmentation	4288×2848	RGB (24 bits)	2018
DDR	136 736	Diagnosis	304×304	RGB (24 bits)	2019
CLOUD	112	Diagnosis	3456×5184	Gray (OCT)	2019
LAG	11 760	Diagnosis	3456×5184	RGB (24 bits)	2020
ODMI (ours)	50	Segmentation	288×288	Gray (OCTA)	2023

without signs. The annotation of DR image is pixel level. DR signs include microaneurysms (MA), soft exudates (SE), hard exudates (EX) and hemorrhages (HE), which are binary masks of lesions.

7. DDR: The DDR (general-purpose Diabetic Retinopathy) dataset [41] presented in 2019 is used to diagnose DR disease and fundus exudates. The DDR dataset contains 136 736 DR images. And 757 DR images have the annotations of fundus exudate lesions.

8. CLOUD: The CLOUD dataset [42] is a set of Optical Coherence Tomography of the Anterior Segment images (AS-OCT) for automatic detection. The dataset includes 112 AS-OCT images that were captured from 16 patients. In particular, the images were obtained by an OCT Cirrus 500 scanner model of Carl Zeiss Meditec.

9. LAG: The LAG (large-scale attention based fundus image) dataset [43] is used for glaucoma detection in 2020. It contains 11 760 images and 2392 glaucoma images. Fundus images were taken by three cameras, such as Topcon, Canon and Carl Zeiss. The resolution of image is 3456×5184 pixels.

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 | System diagram of SS-OCT

The functional principle of OCT imaging technology is light interference [44]. The interferometer developed by Michelson sent a beam of light splitting into two paths by a beam splitter. In OCT system, the light from a low-coherence source is sent into two modules, that is, sample arm and reference arm. Back-reflected light from two arms is redirected toward fiber coupler. The light can be shaped by different optical components, such as lenses. The light is back-reflected by a mirror in the reference arm and returns to the interference system along the same path. However, it is in the opposite direction. In the sample arm, the light propagates as the similar to that in the reference arm. The only difference is that the beam is back-scattered by the sample.

Figure 2 shows the schematic diagram of SS-OCT system (More details of SS-OCT system can be viewed at our website: https://kyanbis.github.io/ODMI). The light from the swept light source passes through the beam splitter and enters reference and sample arms, respectively. The light power from the swept source was split 90% into the sample arm and 10% into the reference arm by a fiber coupler. In the sample arm, the light from swept source component reaches retina through a fiber coupler and collimation lens. In the reference arm, the light passes through a collimator and a focusing lens and a mirror reflects the light beam. The light from the reference and sample arms generates interference in a fiber coupler with a ratio of 50:50, and then was redirected toward a balance detector. Electronic signals from the detector were input into the computer.

3.2.2 | Design of SS-OCT

Figure 3 shows the photographs of the overall SS-OCTA system based on Figure 2. Some optical components are set in the work surface. The swept source is the key component of OCT system, which is 1.0 μm High speed scanning laser and has better penetration to detect deeper fundus than 1.5 μm. The swept source component in Figure 3 is HSL-20-100-B of santec in Japan. Scan rates with 100 kHz are available from the robust

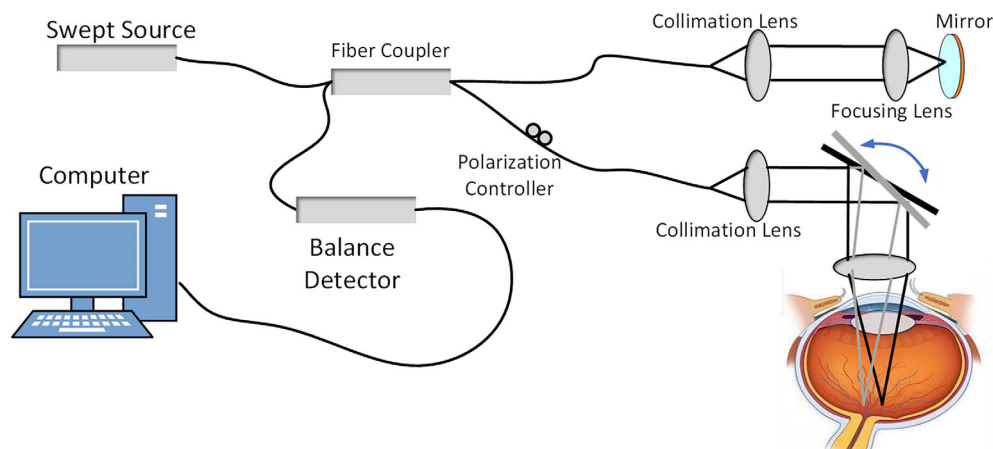**FIGURE 2** System diagram of swept source optical coherence tomography (SS-OCT).
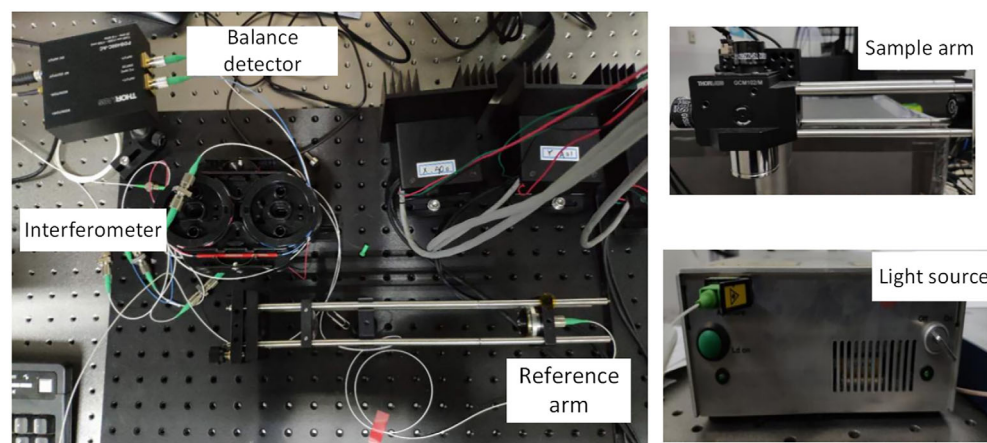

FIGURE 3 Physical diagram of SS-OCT system.

laser that may be configured with start trigger and k-clock. The longitudinal resolution is 9.4 μm, and the maximum imaging depth is 9.6 mm.

The interferometer utilizes the Mach–Zehnder configuration in this paper, which is TW1300R5A2 of Thorlabs in the United States. The sample arm of SS-OCT system consists of a fiber collimator, a scanning galvanometer, and a scanning objective. The fiber collimator is a Thorlabs PAF2A-18-C with an achromatic lens, focus of 15 mm, and a beam expansion diameter of 3.25 mm. The light from sample arm and reference arm passes through the fiber coupler at a ratio 50:50 in a balanced detector (PDB480C-AC, Thorlabs, USA), which has a bandwidth of 30 kHz to 1.6 GHz.

3.2.3 | Image construction of SS-OCT

In ophthalmic applications, en face OCT has gradually become a useful tool for evaluating OCT angiography [45, 46], which is an advantage of high detection sensitivity. OCT angiography utilizes several repeated B-scans to record contrast of the flowing blood cells.

Figure 4 depicts the flowchart of processing OCTA images. The OCT signal recorded by the detector during a complete travel of the reference mirror is called an A-scan. The OCTA system obtains multiple A-scan at each position of the beam in Figure 4A. Therefore, lots of consecutive A-scans can be forming an OCT image, called a B-scan in Figure 4B. A B-scan can be viewed as a number of T-scans (i.e., en face) repeating successive scans at depth. En face is keeping the axial coordinates fixed (reference mirror at rest) scanning in Figure 4C. In other words, the advantage of the en face OCT system is that it allows the collection of constant depth images in real time.

3.3 | The new ODMI dataset

3.3.1 | Dataset statistic

We constructed a retinal image database for segmentation of the human optic disc and macular with optical coherence tomography angiography (OCTA), namely *Optic Disc and Macula in fundus Image* (*ODMI*), which

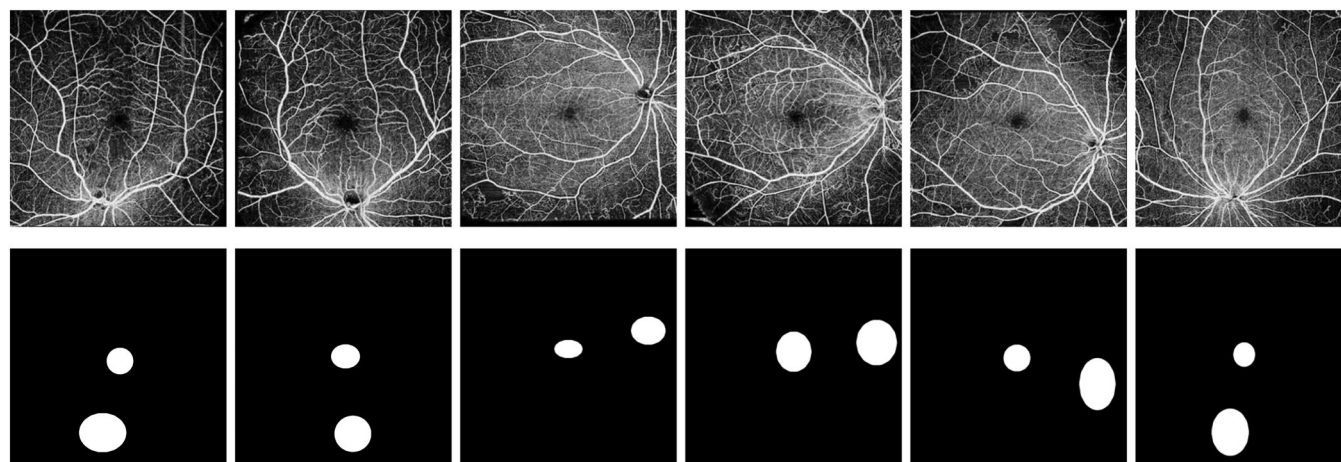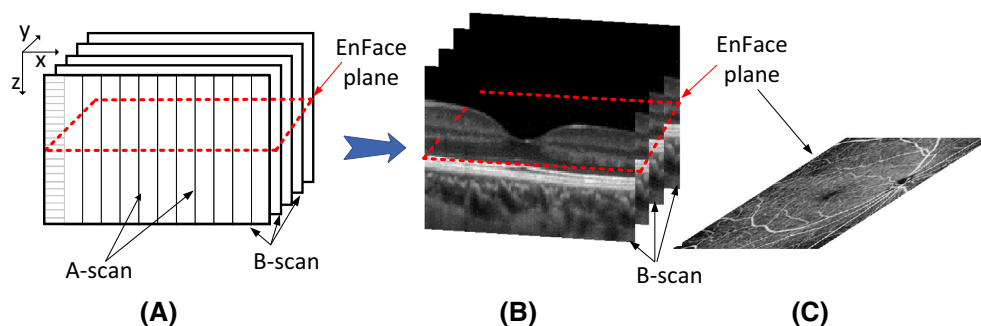**FIGURE 4**   Scheme of generating OCT angiography images.

FIGURE 5   The OCTA images and ground truths in the ODMI dataset.

can be viewed at our website https://kyanbis.github.io/ODMI. We also make the ground truth [47] for the optic disc and macular segmentation. As shown in Figure 5, there exist 50 normal subjects optic disc and macular images and 50 ground truths in total. The resolution of each image is 288×288 pixels.

3.3.2 | Data labeling

The optic disc is the area where the retinal nerve fibers converge, and the macula is a collection of optic cells located in the superficial layer of the retina indicating the nerve fibers. With the advice and experience of an ophthalmologist, the ground truths are labeled manually. The ground truths of the optic disc and macula in the color fundus image are consistent with those in the OCTA image in Figure 1. To by privacy, we hide personal information, such as name and age.

4 | METHODS

In this section, we first give a brief overview of our proposed approach, and then describe two components,

including a Coarse-Grained Feature extraction module (CGF) and a Fine-Grained Feature extraction module (FGF).

4.1 | Overview

Given a retinal OCTA image with an optic disc and macular, the proposed network is presented to output segmentation image of the optic disc and macular. Figure 6 shows the overall architecture, which is comprised of four components: input, backbone, two modules, and segmentation output. To segment the optic disc and the macula of fundus images with OCTA, we propose a novel framework, namely a Coarse and Fine Attention-based Network (CFANet), which includes two key modules, that is, a Coarse-Grained Feature extraction module (CGF) and a Fine-Grained Feature extraction module (FGF). CFANet can integrate the coarse and fine features to segment the optic disc and the macula in an end-to-end manner [48].

First, we utilize the classical network as backbone, such as CENet [49] and UNet, to extract global features. The features are fed into the refinement module to obtain the refined maps, and then fuse the global features and
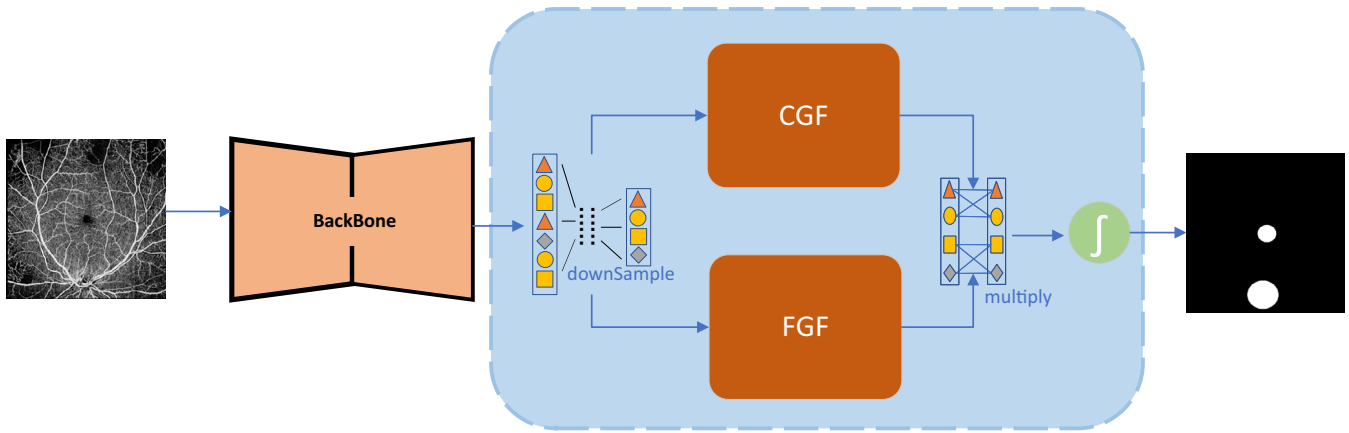
FIGURE 6 The overall architecture of our CFANet. It mainly consists of the backbone and Coarse and Fine Attention-Based Network.

the refinement features. The segmentation results of optic disc and the macula are obtained by CFANet. To obtain a segmentation of the optic disc and the macula in OCTA images, we design a new attention module into the backbone. Skip connections in the CENet only fuse features of the same spatial resolution at a single scale [50, 51]. Due to the lack of multi-scale information interaction and constraints between adjacent layers. This limits the ability of multi-scale feature extraction. Therefore, we propose the CFANet to effectively learn feature information between coarse-grained features and fine-grained features. The features of the fundus of different scales are captured step by step from low to high levels. The attention subnetwork has two modules, that is, a Coarse-Grained Feature extraction module (CGF) and a Fine-Grained Feature extraction module (FGF). Each module contains a 4×4 convolution kernel, batch normalization (BN), a nonlinear function (ReLU), and a residual module. The 4×4 convolutional kernels allow the neural network to obtain effective features. BN converts the data distribution of each layer into a normal distribution, which makes the network converge faster. ReLU is a parameter-free nonlinear activation function and allows the network to reduce gradient disappearance. The residual blocks can mitigate the gradient explosion of the neural network.

4.2 | Coarse-Grained Feature extraction module

The proposed CGF module is illustrated in the top branch of Figure 7. CGF focuses on capturing the global features from maps, and can effectively learn the features of the optic disc and the macula by increasing the receptive field. The backbone outputs feature maps with C ×

H × W. First, the size of feature maps becomes $4 \times C \times H \times W$ with two upsampling operations. This can increase the receptive field of the optic disc and the macular. CGF can identify the area more accurately. Finally, A_m denotes the Coarse-Grained Feature, which is written as follows:

$$A_m = \phi(\sigma(Conv(ReLU(BN(D\_Sam(U\_Sam(f)))))))), \quad (1)$$

where $\phi(\cdot)$ is the reshape operation. $\sigma(\cdot)$ denotes Sigmoid function. $ReLU(\cdot)$ is the nonlinear activation function. $BN(\cdot)$ represents batch normalization. $D\_Sam(\cdot)$ represents downsampling. $U\_Sam(\cdot)$ is upsampling operation. $Conv(\cdot)$ denotes convolutional operation. f is the input feature from the backbone network.

4.3 | Fine-Grained Feature extraction module

The bottom branch in Figure 7 shows the proposed FGF module, which focuses on capturing the detailed features from fundus. Through the skip connection in ResNet, the detailed features of the region are preserved and the edge part of the region is enhanced. The bottom branch uses the ResNet network to generate the feature $A_d \in R^{C \times HW}$. The formula is written as follows:

$$A_d = \phi(Conv(f)), \quad (2)$$

where $\phi(\cdot)$ is the reshape operation. $Conv(\cdot)$ denotes convolutional operation. f is the input features generated from the backbone.

With Equations (1) and (2), we have obtained the Coarse-Grained Feature A_m and the Fine-Grained Feature A_d, respectively. To better obtain the global feature,
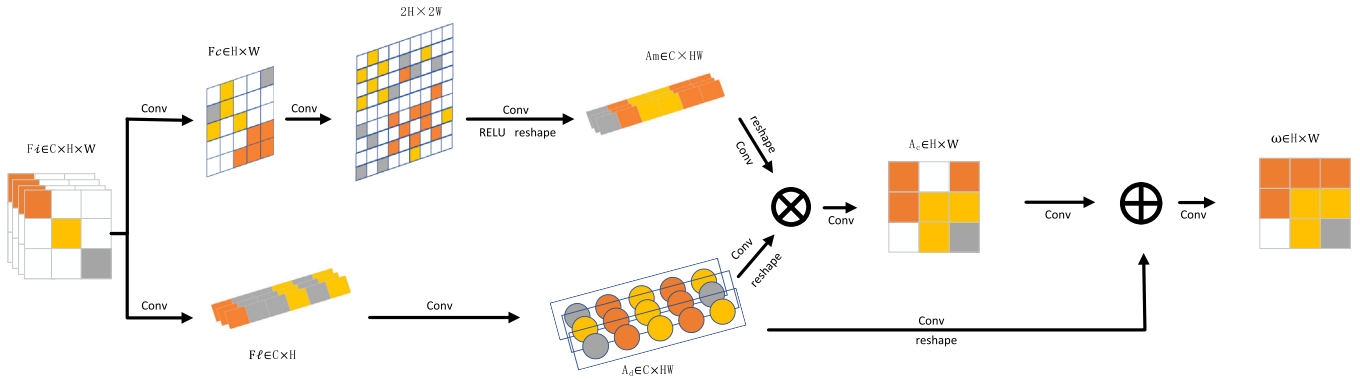
FIGURE 7 Details of CGF and FGF attention modules.

we mine the relationships of A_m and A_d to concatenate the feature A_c, which is computed as follows:

$$A_c = A_m \bigotimes A_d, \quad (3)$$

where $\bigotimes$ denotes element multiplication. A_c is the enhanced feature computed by A_m and A_d. The formula is as follows:

$$w = A_c \bigoplus A_d, \quad (4)$$

where w is the point-wise attention weight of the same shape of the input, and its value ranges from 0 to 1. $\bigoplus$ denotes the addition of elements. The range of values indicates the feature importance of each spatial point in each input feature map. By the above operations, target regions are accurately identified while non-target regions are suppressed.

5 | EXPERIMENTS

In this section, we first give the experimental setting. Then we report the results on the new ODMI dataset.

5.1 | Data augmentation

There exist 50 original images and 50 ground truths in the ODMI dataset. We expand samples by using 5 types of augmentation operations, such as adjust the brightness, saturation, contrast of the image, horizontal flipped and rotation operations [52] for original photos. After the data augmentation, we performed the necessary pre-processing on each dataset, including normalization.

5.2 | Data partition

To ensure the reliability and repeatability of the ODMI dataset, we partition the dataset into the training set and testing set randomly [52], which can improve the generalization of the model. All experiments are conducted by using fivefold cross validation. Forty original samples are selected as the training set. With the data augmentation, there are 240 augmented samples for training in total. Ten original sample pairs are selected as the test set. Then we report the five metrics of the test set performance.

5.3 | Experimental 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 300.

5.4 | Evaluation metrics

To evaluate the segmentation performance of all methods, we utilize the following evaluation metrics: F1-score, dice coefficient, accuracy, sensitivity, and specificity.

5.4.1 | F1-score

F1-score is also known as the f-score. F1-score is a weighted average of precision and recall. It is a common evaluation criterion and is often used to evaluate the quality of classification models. Based on accuracy and recall. The formula is as follows:

$$\text{F1-score} = \frac{\text{Precision} \times \text{Recall}}{\text{Precision} + \text{Recall}}. \quad (5)$$

T A B L E 3   Results on the ODMI dataset (%).

Model	Sensitivity	Specificity	Dice	Accuracy	F1-score
UNet	94.21	**98.69**	88.84	98.38	88.84
UNet+ CFANet (ours)	96.61	96.84	80.52	96.83	80.52
ResNet34UNet	97.09	97.87	85.19	97.8	85.19
ResNet34UNet+ CFANet (ours)	94.98	97.95	85.29	97.72	85.29
DenseUNet	89.19	98.74	86.37	98.13	86.37
DenseUnet+ CFANet (ours)	91.05	98.74	87.66	98.22	87.66
SemiMsTGANet	85.27	98.21	70.81	97.62	70.81
SemiMsTGANet+ CFANet (ours)	79.53	98.97	75.03	98.14	75.03
CENet	94.37	98.33	86.82	98.11	86.82
CENet+ CFANet (ours)	**98.91**	98.47	**89.77**	**98.49**	**89.77**

Note: Best results are in bold.
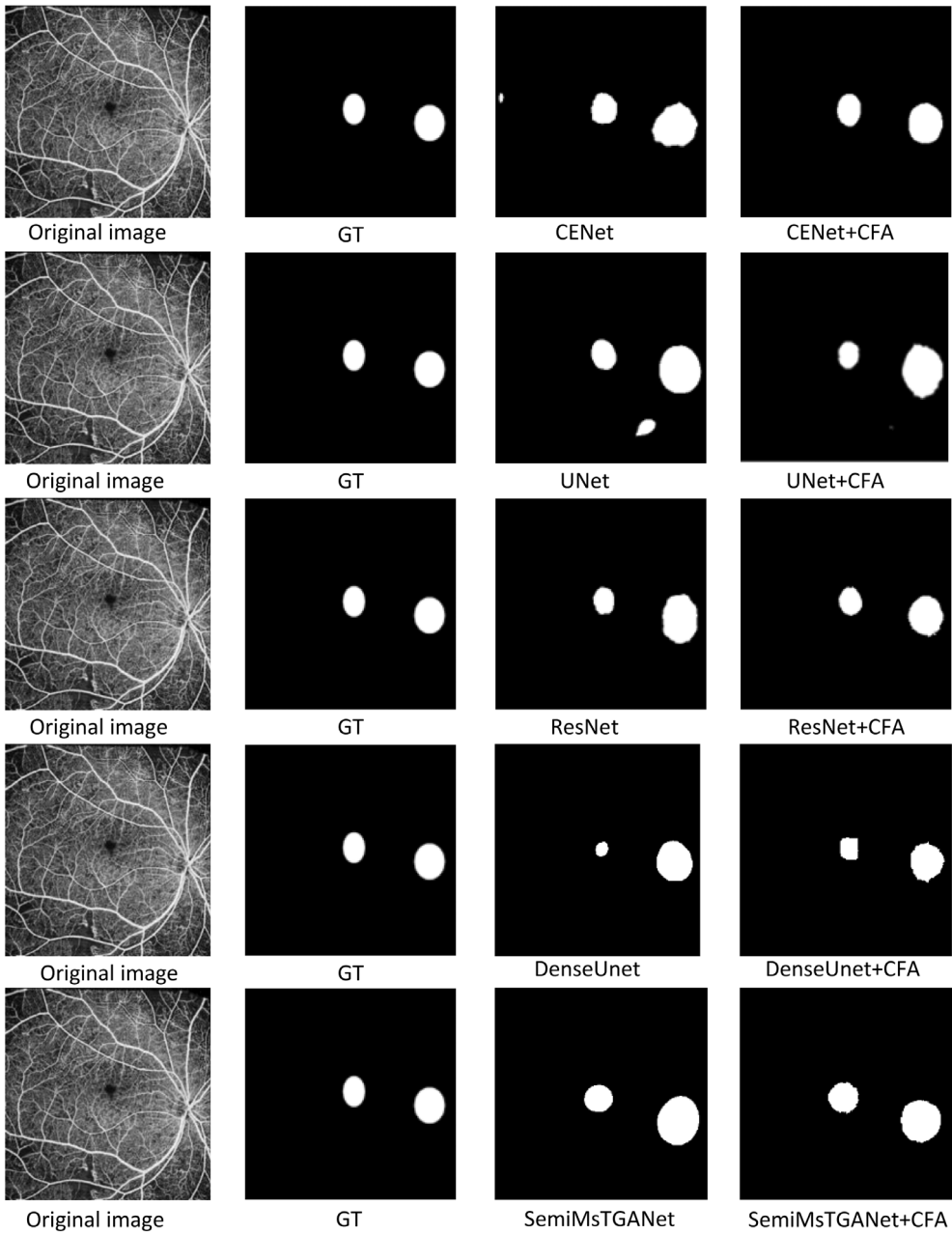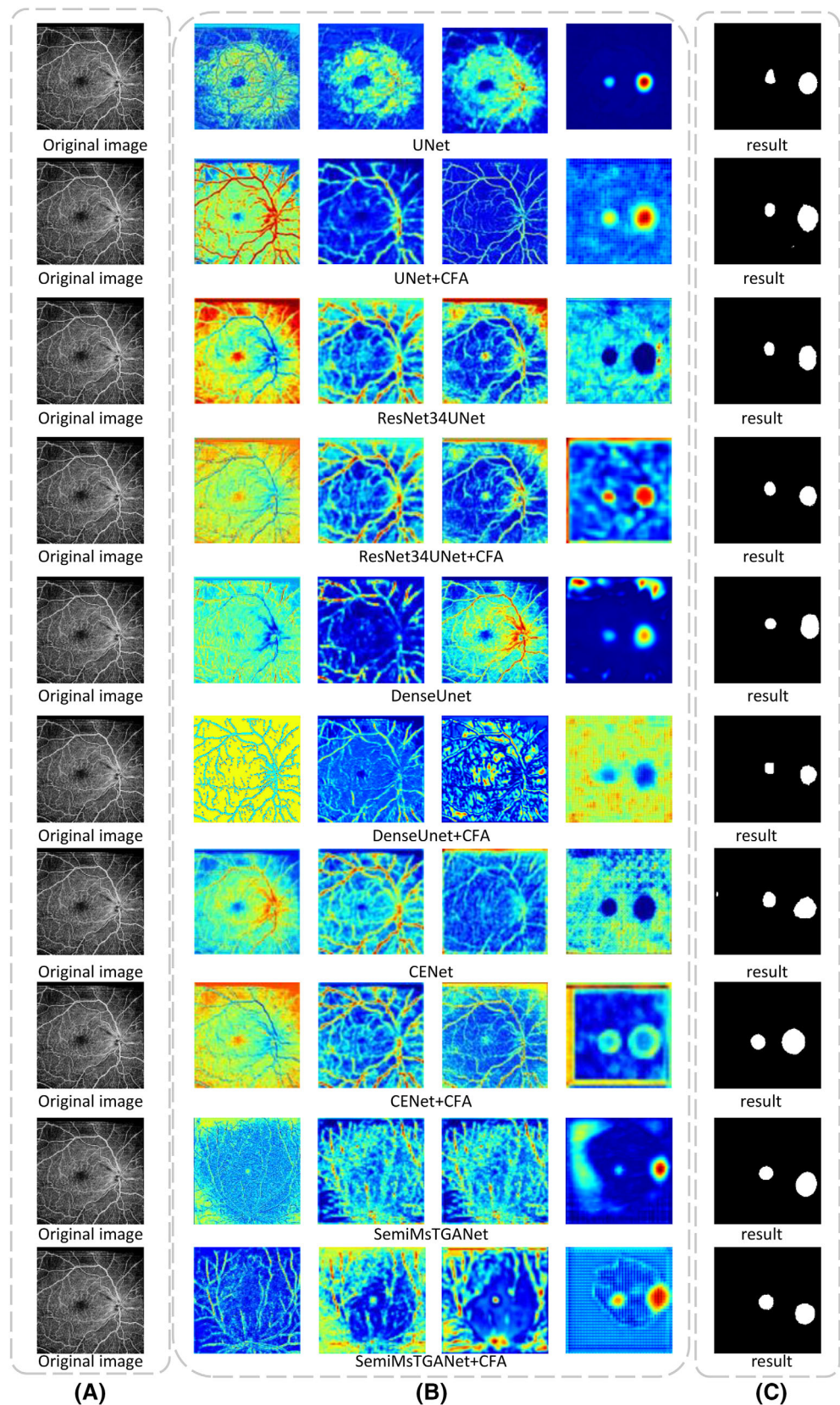

F I G U R E 8   Visualization of the optic disc and macular segmentation and the ground truth.

FIGURE 9 Heatmap visualization.

(A) (B) (C)

F1-score maintains a balance between accuracy and recall. We use this comparison metric if the categories are not evenly distributed, as precision and recall may give misleading results.

5.4.2 | Dice coefficient

Dice coefficient is a statistic used to evaluate the similarity of two samples. It is commonly used in image

segmentation and medical image processing. It measures the degree of overlap of two collections. The formula is as follows:

$$\text{Dice coefficient} = 2 \times \frac{TP}{(FP + FN + 2 \times TP)}. \qquad (6)$$

5.4.3 | Further metrics

To further evaluate the performance of the proposed method, we utilize the evaluation metrics as follows.

$$\text{Specificity} = \frac{TN}{TN + FP}, \qquad (7)$$

$$\text{Sensitivity} = \frac{TP}{TP + FN}, \qquad (8)$$

$$\text{Accuracy} = \frac{TP + TN}{TP + FP + FN + TN}, \qquad (9)$$

where True positive (TP) is the number of true pixels when the actual observation is positive. False positive (FP) is the number of incorrect pixels when the actual observation is positive. True negative (TN) is the number of actual true pixels when the model predicted as negative. False negative (FN) is that the actual and predicted pixels are negative.

5.5 | Comparison with other methods

We compared the proposed method with three competing networks on the ODMI dataset, including UNet, ResNet34UNet [53], SemiMsTGANet [54], and CENet. In experiments, we utilize three typical networks as backbone. First, samples in training set are fed into the backbone. Then we can obtain the global feature maps. Finally, the segmentation results can be output by CFANet.

Table 3 depicts the results of five metrics of the comparative methods, including sensitivity, specificity, dice coefficient, accuracy, and F1-score. One can see that our CFANet achieves the highest rate of segmentation on the ODMI dataset. Specifically, sensitivity, specificity, Dice, accuracy and F1-score of our CFANet with CENet are 98.91%, 98.47%, 89.77%, 98.49%, and 89.77%, respectively.

Figure 8 shows the visualization of the optic disc and macular segmentation of different methods on the ODMI dataset.

5.6 | Visualization of heatmaps of CFANet

A heat map is a visualization by the representing a relative intensity of feature maps. Pixels in blue color denote cool while those in red mean hot. We first conduct experiments to observe the visualization of attention maps. Figure 9 shows the heat maps of different layers in CFANet to better understand the effects of attention. We select several OCTA images from the ODMI dataset. The model with the addition of the CFANet generates better results than the model without the CFANet. For the model without the CFANet, some irrelevant regions may be identified. Specifically, for the segmentation of some irrelevant regions in the fifth row of Figure 9, the overall effect is somewhat worse than those in the sixth row. It can be seen that the attention module of CFANet can suppress irrelevant regions. More accurate interpretation of segmentation results are useful for clinical treatment.

6 | CONCLUSION

In this paper, a SS-OCTA system was designed to capture OCTA images of human eyes. We introduce the new ODMI dataset, which contains 50 OCTA fundus images. Images in the ODMI dataset have rich fundus details and contextual information. In addition, we propose a new framework CFANet for the optic disc and macular segmentation. Our CFANet integrates CGF and FGF to retain more spatial details for the optic disc and macular segmentation. In future work, we will collect more OCTA fundus images, and further improve the network structure to have more accurate segmentation for the optic disk and macula.

CONFLICT OF INTEREST STATEMENT
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.

ORCID

Fei Ma https://orcid.org/0000-0002-5472-4763

REFERENCES

[1] Z. F. Hua, C. Jun, W. X. Yan, W. K. W. Damon, L. Jiang, C. C. Xiao, *IEEE Trans. Med. Imaging* **2018**, *37*, 1597.

[2] L. Annan, H. N. Zhi, C. Jun, S. Y. Feng, W. K. W. Damon, C. Y. Shui, L. Jiang, *Neurocomputing* **2018**, *275*, 350.

[3] U. R. Zaka, S. N. Syed, M. K. Tariq, A. Muhammad, A. K. Muhammad, M. A. Khan, M. A. Khalil, *Expert Syst. Appl.* **2019**, *120*, 461.

[4] Y. H. Ma, H. Y. Hao, J. Y. Xie, H. Z. Fu, J. Zhang, J. L. Yang, Z. Wang, J. Liu, Y. L. Zheng, Y. T. Zhao, *IEEE Trans. Med. Imaging* **2021**, *40*, 928.

[5] S. J. Wang, L. Q. Yu, X. Yang, C. W. Fu, P. A. Heng, *IEEE Trans. Med. Imaging* **2019**, *38*, 2485.

[6] M. Shervin, R. B. Yu, P. Fatih, P. Antonio, K. Nasser, T. Demetri, *IEEE Tran. Pattern Anal. Mach. Intell.* **2022**, *44*, 3523.

[7] J. W. Xiao, L. W. Wei, J. L. Zhen, Q. W. Yi, H. H. Yu, J. H. Zhan, *Chin. Conf. Pattern Recognit.* **2018**, *11258*, 52.

[8] S. Alex, *Phys D* **2020**, *404*, 404.

[9] F. W. Jian, L. H. Xiao, *IEEE Trans. Pattern Anal. Mach. Intell.* **2022**, *44*, 3421.

[10] T. S. Heng, C. L. Lu, Y. Yang, X. Xing, H. Zi, S. S. Fu, C. H. Ri, *IEEE Trans. Knowl. Data Eng.* **2021**, *33*, 3351.

[11] H. G. Chen, X. H. He, L. B. Qing, Y. Y. Wu, C. Ren, R. E. Sheriff, C. Zhu, *Inf. Fusion* **2022**, *79*, 124.

[12] P. J. Wang, B. Bayram, E. Sertel, *Earth Sci. Rev.* **2022**, *232*, 232.

[13] E. Göçeri, *Comput. Sci.* **2019**, *11*, 1.

[14] D. Bahdanau, K. Cho, Y. Bengio, in Int. Conf. on Learning Representations, **2015**, abs/1409.0473.

[15] J. Long, E. Shelhamer, T. Darrell, *IEEE Trans. Pattern Anal. Mach. Intell.* **2017**, *39*, 640.

[16] O. Ronneberger, P. Fischer, T. Brox, *Med. Image Comput. Comput. Assist. Interv.* **2015**, *9351*, 234.

[17] H. Zhu, M. R. Ma, W. P. Ma, L. C. Jiao, S. K. Hong, J. C. Shen, B. Hou, *Inf. Fusion* **2021**, *70*, 72.

[18] E. Bercovich, M. C. Javitt, *Rambam Maimonides Med. J.* **2018**, *9*, e0034.

[19] R. Chopra, S. K. Wagner, P. A. Keane, *Eye* **2021**, *35*, 236.

[20] S. Sreng, N. Maneerat, K. Hamamoto, K. Y. Win, *Appl. Sci.* **2020**, *10*, 4916.

[21] S. X. Zheng, J. C. Lu, H. S. Zhao, X. T. Zhu, Z. K. Luo, Y. B. Wang, Y. W. Fu, J. F. Feng, T. Xiang, P. H. S. Torr, L. Zhang, *Comput. Vis. Pattern Recognit.* **2021**, *33*, 6881.

[22] Z. L. Zhong, Z. Q. Lin, R. Bidart, X. D. Hu, I. B. Daya, Z. F. Li, W. S. Zheng, A. Wong, *Comput. Vis. Pattern Recognit.* **2020**, *32*, 13062.

[23] X. L. Peng, R. F. Zhong, Z. Li, Q. Y. Li, *IEEE Trans. Geosci. Remote Sens.* **2021**, *59*, 7296.

[24] J. Liu, M. G. Gong, A. K. Qin, P. Z. Zhang, *IEEE Trans. Neural Netw. Learn. Syst.* **2018**, *29*, 545.

[25] S. Mittal, M. Tatarchenko, T. Brox, *IEEE Trans. Pattern Anal. Mach. Intell.* **2021**, *43*, 1369.

[26] K. Nguyen, C. Fookes, S. Sridharan, *Pattern Recognit.* **2020**, *105*, 107358.

[27] Z. W. Zhou, M. M. R. Siddiquee, N. Tajbakhsh, J. M. Liang, *IEEE Trans. Med. Imaging* **2020**, *39*, 1856.

[28] K. M. He, X. Y. Zhang, S. Q. Ren, J. Sun, in *IEEE Conf. on Computer Vision and Pattern Recognition*, IEEE, New York **2016**, p. 770.

[29] J. M. J. Valanarasu, V. A. Sindagi, I. Hacihaliloglu, M. V. Patel, *IEEE Trans. Med. Imaging* **2022**, *41*, 965.

[30] P. Wang, A. C. S. Chung, in *ICCV Workshops*, Springer, Berlin **2020**, p. 338.

[31] L. C. Chen, G. Papandreou, I. Kokkinos, K. Murphy, A. L. Yuille, *IEEE Trans. Pattern Anal. Mach. Intell.* **2018**, *40*, 834.

[32] K. M. He, X. Y. Zhang, S. Q. Ren, J. Sun, *IEEE Trans. Pattern Anal. Mach. Intell.* **2015**, *37*, 1904.

[33] M. Hofinger, S. R. Bulò, L. Porzi, A. Knapitsch, T. Pock, P. Kontschieder, in *European Conf. on Computer Vision*, Springer, Berlin **2020**, p. 770.

[34] S. L. Feng, H. M. Zhao, F. Shi, X. Cheng, M. Wang, Y. H. Ma, D. H. Xiang, W. F. Zhu, X. J. Chen, *IEEE Trans. Med. Imaging* **2020**, *39*, 3008.

[35] A. D. Hoover, V. Kouznetsova, M. Goldbaum, *J. Am. Med. Inf. Assoc.* **2000**, *19*, 203.

[36] Z. Zhang, F. S. Yin, J. Liu, W. K. Wong, N. M. Tan, B. H. Lee, J. Cheng, T. Y. Wong, in *IEEE EMBC*, IEEE, New York **2010**, p. 3065.

[37] A. Budai, R. Bock, A. Maier, J. Hornegger, G. Michelson, *Int. J. Biomed. Imaging* **2013**, *2013*, 1.

[38] A. Baskaran, R. C. Foo, C. Y. Cheng, A. K. Narayanaswamy, Y. F. Zheng, R. Y. Wu, S. M. Saw, P. J. Foster, T. Y. Wong, T. Aung, *JAMA Ophthal.* **2015**, *133*, 874.

[39] B. Dashtbozorg, J. Zhang, F. Huang, B. M. T. Romeny, *IEEE Trans.* **2018**, *27*, 3300.

[40] P. Prasanna, P. Samiksha, K. Ravi, K. Manesh, D. Girish, S. Vivek, M. Fabrice, *Int. Conf. Data Technol. Appl.* **2018**, *3*, 25.

[41] T. Li, Y. Gao, K. Wang, S. Guo, H. Liu, H. Kang, *Inf. Sci.* **2019**, *501*, 511.

[42] P. Cabaleiro, J. de Moura, J. Novo, P. Charlón, M. Ortega, *Sensors* **2019**, *19*, 5087.

[43] I. Li, M. Xu, H. R. Liu, Y. Li, X. F. Wang, L. Jiang, Z. L. Wang, X. Fan, N. L. Wang, *IEEE Trans. Med. Imaging* **2020**, *39*, 413.

[44] L. Hung-Chang, L. Miao-Hui, T. Ching-Heng, W. Yi-Min, S. Chia-Wei, *J. Biophoton.* **2023**, *16*, e202200344.

[45] R. A. Leitgeb, *Biomed. Opt. Express* **2019**, *10*, 2177.

[46] J. Xu, X. Yuan, Y. Huang, J. Qin, G. Lan, H. Qiu, B. Yu, H. Jia, H. Tan, S. Zhao, Z. Feng, L. An, X. Wei, *J. Biophoton.* **2023**, *16*, e202200366.

[47] R. Guidotti, *Artif. Intell.* **2021**, *291*, 103428.

[48] M. Goutay, F. A. Aoudia, J. Hoydis, J. N. Gorce, in *Global Communications Conference*, IEEE, New York **2021**, p. 1.

[49] Z. W. Gu, J. Cheng, H. Z. Fu, K. Zhou, H. Y. Hao, Y. T. Zhao, T. Y. Zhang, S. H. Gao, J. Liu, *IEEE Trans. Med. Imaging* **2019**, *38*, 2281.

[50] H. C. Li, A. Li, M. H. Wang, *Comput. Biol. Med.* **2019**, *108*, 150.

[51] Y. D. Zhang, H. Y. Liu, Q. Huu, *Med. Image Comput. Comput. Assist. Interv.* **2021**, *12901*, 14.

[52] D. Feng, S. C. Haas, L. Rosenbaum, H. Hertlein, C. Gläser, F. Timm, W. Wiesbeck, K. Dietmayer, *IEEE Trans. Intell. Transp. Syst.* **2021**, *22*, 1341.

[53] L. Gao, H. S. Su, C. Y. Wang, K. Liu, S. H. Chen, *Remote Sens.* **2022**, *14*, 3003.

[54] M. Wang, W. F. Zhu, F. Shi, J. Z. Su, H. Y. Chen, Y. Zhou, Y. Y. Peng, Z. Y. Chen, X. J. Chen, *IEEE Trans. Med. Imaging* **2022**, *41*, 394.