Joint Segmentation of Multi-Class Hyper-Reflective Foci in Retinal Optical Coherence Tomography Images

Chenpu Yao, Meng Wang, Weifang Zhu, Haifan Huang, Fei Shi, Zhongyue Chen, Lianyu Wang, Tingting Wang, Yi Zhou, Yuanyuan Peng, Liangjiu Zhu, Haoyu Chen, and Xinjian Chen, *Senior Member, IEEE*

***Abstract*—Hyper-reflective foci (HRF) refers to the spot-shaped, block-shaped areas with characteristics of high local contrast and high reflectivity, which is mostly observed in retinal optical coherence tomography (OCT) images of patients with fundus diseases. HRF mainly appears hard exudates (HE) and microglia (MG) clinically. Accurate segmentation of HE and MG is essential to alleviate the harm in retinal diseases. However, it is still a challenge to segment HE and MG simultaneously due to similar pathological features, various shapes and location distribution, blurred boundaries, and small morphology dimensions. To tackle these problems, in this paper, we propose a novel global information fusion and dual decoder collaboration-based network (GD-Net), which can segment HE and MG in OCT images jointly. Specifically, to suppress the interference of similar pathological features, a novel global information fusion (GIF) module is proposed, which can aggregate the global semantic information efficiently. To further improve the segmentation performance, we design a dual decoder collaborative workspace (DDCW) to comprehensively utilize the semantic correlation between HE and MG while enhancing the mutual influence on them by feedback alternately. To further optimize GD-Net, we explore a joint loss function which integrates pixel-level with image-level. The dataset of this study comes from patients diagnosed with diabetic macular edema at the department of ophthalmology, University Medical Center Groningen, The Netherlands. Experimental results show that our proposed method performs better than other state-of-the-art methods, which suggests the effectiveness of the proposed method and provides research ideas for medical applications.**

***Index Terms*—Dual decoder collaborative workspace, global information fusion module, hyper-reflective foci (HRF), joint segmentation.**

Manuscript received April 15, 2021; revised July 13, 2021 and September 9, 2021; accepted September 18, 2021. Date of publication September 27, 2021; date of current version March 21, 2022. This work was supported in part by the National Key R&D Program of China under Grant 2018YFA0701700 and in part by the National Nature Science Foundation of China under Grants U20A20170 and 61622114. *(Corresponding author: Xinjian Chen.)*

Chenpu Yao, Meng Wang, Weifang Zhu, Fei Shi, Zhongyue Chen, Lianyu Wang, Tingting Wang, Yi Zhou, Yuanyuan Peng, and Liangjiu Zhu are with the School of Electronics and Information Engineering and Medical Image Processing, Analysis and Visualization Lab, Soochow University, China.

Haifan Huang and Haoyu Chen are with the Shantou University and the Chinese University of Hong Kong, Joint Shantou International Eye Center, China.

Xinjian Chen is with the School of Electronics and Information Engineering and the State Key Laboratory of Radiation Medicine and Protection, Soochow University, Jiangsu 215006, China (e-mail: xjchen@suda.edu.cn).

Digital Object Identifier 10.1109/TBME.2021.3115552

I. Introduction

OPTICAL coherence tomography (OCT) is a high-resolution, non-contact, non-invasive biological tissue imaging technology that performs tomographic imaging of the anterior and posterior segments of the eye with micron resolution [1]–[4]. It generates cross-sectional images of ocular biological tissues including the macula and optic disc, which can reveal the pathological changes in the fundus. Therefore, OCT is widely used in the diagnosis and prognosis planning of ophthalmic diseases [5], [6].

Hyper-reflective foci (HRF) refers to the spot-shaped and block-shaped area with high local contrast and high reflectivity in retinal OCT images. These small HRFs are mostly observed in OCT images of patients with fundus diseases such as diabetic macular edema (DME), age-related macular degeneration (AMD) and retinal vein occlusion (RVO) [7], that is to say, the existence of HRF is related to the severity of various retinal diseases [8], [9]. Several studies show that HRF mainly manifests as hard exudates (HE) and microglia (MG) [7], [10]–[15].

HE usually shows as block-shaped HRF in retinal OCT images, which generally occurs in retinal vascular diseases, such as diabetic retinopathy (DR), RVO, hypertensive retinopathy and other retinal diseases. Previous studies have found that visual acuity has a significant correlation with HE [16], [17].

MG originates from hematopoietic stem cells and is mainly distributed in the inner retina [18], which is not visible on fundus images but can be seen on OCT images as small spot-shaped HRF. MG may be activated, proliferated, migrated and polarized in the pathogenesis of DR. These changes may play an important role in the occurrence and development of DR retinal neurodegeneration and microcirculation regulation disorders [19], [20], [21]. At the same time, MG is also a clinical indicator of the response to dexamethasone implant treatment in DME patients [12].

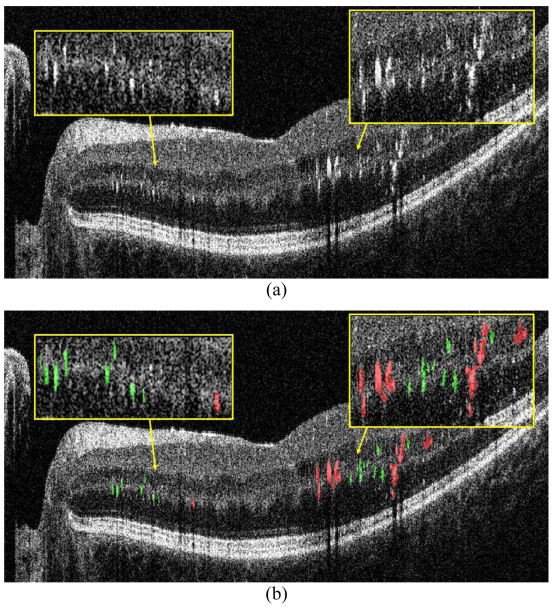

Fig. 1. Example of HE and MG in an OCT B-scan. (a) Unlabeled OCT B-scan image with HE and MG, (b) Labeled OCT B-scan image with HE and MG, the red marking is HE, and the green marking is MG.

A previous study showed that activated MG can be distinguished from HE by its diameter, reflectivity and shady states [18]. HE mainly refers to HRF with an area larger than $1.6 \times 10^3 \mu m^2$ and accompanied by a shadow area underneath. MG mainly refers to HRF with an area larger than $4.0 \times 10^2 \ \mu m^2$ and no shadow area. The highly reflective bright spot with an area less than $4.0 \times 10^2 \ \mu m^2$ is considered to be noise generated during photographing [12], [22]. An example of HE and MG is shown in Fig. 1.

Previous clinical studies have used manual calculation of selected OCT slices to study the HRF area [23]–[25], which is usually time-consuming and leads to subjective bias in the analysis. Therefore, the accurate and automatic joint segmentation of HE and MG is of great significance to ophthalmologic clinical diagnosis.

There are some previous works which focused on the automatic evaluation of HRF [8], [26]–[30]. Mokhtari *et al.* [31] presented an automatic detection method using morphological component analysis to detect HRF. Xie *et al.* [32] adopted an adaptive threshold method to segment HRF in frequency domain optical coherence tomography (SD-OCT) images. Okuwobi *et al.* [33] proposed a method for segmenting and quantifying HRF in SD-OCT images based on Grow-cut algorithm. Recently, in [34], Okuwobi *et al.* segmented HRF by extracting the retinal ROI area and extracting the extreme value area from the connected area of the component tree at the same time, followed by the combination of the two to get the result area. These above methods were all based on traditional algorithms with unsatisfactory segmentation accuracy of HE and MG.

Convolutional neural networks (CNNs) have achieved state-of-the-art performance for automated medical image segmentation in recent years. There are some previous CNNs-based methods have been proposed for HE detection and segmentation in fundus images [35]–[37] and OCT images [38]–[40]. Schlegl *et al.* [38] proposed a method of adding a residual module in U-Net to segment HRF in OCT images. Katona *et al.* [39] used the deep neural networks (DNNs) for quantifying HRF and AMD related biomarkers recognition in retinal OCT images. In addition, Varga el.al [40] adopted several existing networks such as deep rectifier neural networks (DRNs), fully convolutional neural networks (FCN) to automatically segment HRF in OCT images. However, their work mainly focuses on the preprocessing of OCT images rather than the joint segmentation of HE and MG.

In this paper, we propose a novel global information fusion and dual decoder collaboration based network (GD-Net) for the joint segmentation of HE and MG in retinal OCT images. In which, a novel global information fusion (GIF) module is proposed and embedded into the top layer of encoder, while a novel dual decoder collaborative workspace (DDCW) with two decoder branches is designed for the multi-class joint segmentation task. Furthermore, a joint loss function is designed to guide the optimization of GD-Net training. As far as we know, this is the first study on joint segmentation of multi-class HRF. Comprehensives experiments have been conducted to evaluate the proposed GD-Net, and the results show that our proposed method achieves better performance than other state-of-the-art methods. The main contributions of this paper are as follows:

- We propose a novel global information fusion and dual decoder collaboration based network (GD-Net) for HE and MG joint segmentation task.
- To suppress the interference of similar pathological features between HE and MG, a global information fusion (GIF) module is designed.
- A novel dual decoder collaborative workspace (DDCW) is explored, which can deal with problems caused by morphological characteristics with the coordination of the proposed joint loss function.
- Comprehensive experiments are conducted to evaluate the performance of the proposed method, the experiments results show that compared with other excellent CNN-based method, our proposed GD-Net achieves the best performance.

II. Methods

A. Overview

Fig. 2 demonstrates the overview structure of the proposed GD-Net, which consists of three main parts: feature encoder, GIF module and DDCW.

The GIF module is inserted at the top of the encoder to capture global context information, while DDCW replaces the traditional single decoder structure with dual decoders to give spaces for feature fusion and continuous information feedback.

B. Dual Decoder Collaborative Workspace

We use U-Net [41] as the baseline, however, the traditional framework of U-Net always has unsatisfactory segmentation when objects have small size and blurred boundaries. Inspired

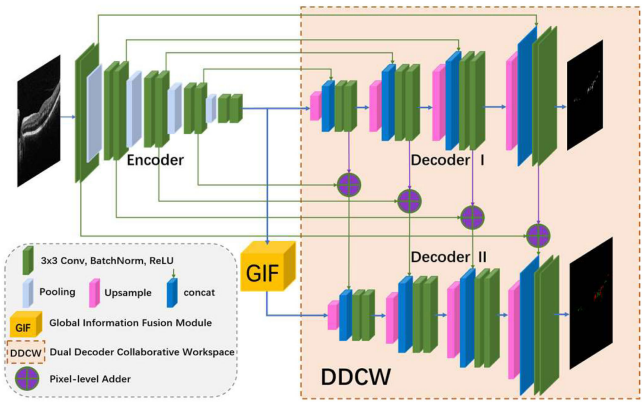

Fig. 2. The structure of our proposed GD-Net.

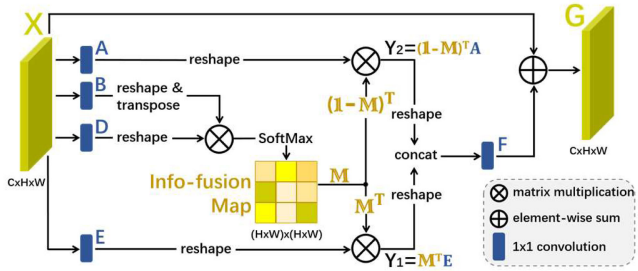

Fig. 3. The structure of Global Information Fusion (GIF) module.

by this problem, we propose a dual decoder collaborative workspace (DDCW), which has two decoder branches.

In the workspace, Decoder II is designed as the main decoder path while Decoder I is explored to play an auxiliary role, whose output is the entire HRF. In Decoder I, the regular skip-connection is adopted, which is similar to U-Net. HE and MG have a variety of shapes and location distribution in morphology, resulting in difficulties in the fine segmentation.

To solve this problem, the pixel-level adder is introduced, which completes the addition between the output of the corresponding layer of Decoder I and the features from the encoder part. The results of adder, which contain rich semantic information, are fed into the concatenate operation in Decoder II. Therefore, the Decoder II that used to segment HE and MG can optimize its performance continuously under the feedback of the Decoder I. In addition, with the coordination of the joint loss function, two decoder branches interact with each other to optimize performance efficiently.

C. Global Information Fusion Module

As has been discussed in the introduction, the similar pathological features between HE and MG can interfere the joint segmentation task. How to solve this and integrate information effectively is a problem worth exploring. To deal with it, a novel global information fusion (GIF) module is proposed. Fig. 3 shows the structure of the GIF module.

As illustrated in Fig. 3, the input feature map is denoted as $X \in \mathrm{R}^{\mathrm{C} \times \mathrm{H} \times \mathrm{W}}$ and fed into a convolution layer to generate two new feature maps B and D, respectively, where $\{B, D\} \in \mathrm{R}^{\mathrm{C} \times \mathrm{H} \times \mathrm{W}}$, and the C, H, W represents the channels, height, width, respectively. Then we reshape them to $\mathrm{R}^{\mathrm{C} \times \mathrm{N}}$, where C and $N = H \times W$ represents the number of feature maps and number of pixels in each feature map, respectively. We further generate the Info-fusion Map $M \in \mathrm{R}^{\mathrm{N} \times \mathrm{N}}$ after performing matrix multiplication between the transpose of B and D followed by softmax operation. Specifically, each position m_{ij} of M is denoted as:

$$m_{ij} = \frac{e^{B_j \bullet D_i}}{\sum_{j=1}^{N} e^{B_j \bullet D_i}} \tag{1}$$

where m_{ij} measures the impact of the i^{th} position on the j^{th} position. After integrating the information of each location in turn, M covers the global features. The larger value of m_{ij} contributes greater correlation between the two positions. We refer M as the map positively correlated with the feature. The more similar feature representations of the two position contributes to greater correlation between them. Therefore, *1-M* is the map negatively correlated with the feature, where *1* is a matrix with the same size as M and all values of 1. M extracts the information belonging to the same category, while *1-M* focuses on the information of different classes.

At the same time, we feed feature X into a convolution layer to generate another two new feature maps A and E, respectively, where $\{A, E\} \in \mathrm{R}^{\mathrm{C} \times \mathrm{H} \times \mathrm{W}}$ and reshape them to $\mathrm{R}^{\mathrm{C} \times \mathrm{N}}$. After performing matrix multiplication between the transpose of M and E, and the transpose of *1-M* and A, the results denoted as Y_1 and Y_2 are obtained, which contain positive and negative correlation features, respectively.

$$Y_1 = \sum_{j=1}^{N} (m_{ij} \bullet E_j) \tag{2}$$

$$Y_2 = \sum_{j=1}^{N} [(1 - m_{ij}) \bullet A_j] \tag{3}$$

According to the correlation between the same categories, Y_1 further reinforces the interdependencies between features. Meanwhile, Y_2 highlights the differences in different categories and captures the missing information from the negative correlation map. After reshaping, concatenation operation is done on Y_1 and Y_2. The result is fed into a convolutional layer to generate a new feature map F.

$$F_i = Conv\left[Concat\left(Y_1, Y_2\right)\right] \tag{4}$$

It can be inferred from Eq. (4) that through the feature fusion of two dimensions, positive correlation and negative correlation, we can infer the semantic relevance between pixels under each correlation. Therefore, the obtained information suppresses the interference of similar pathological features on segmentation performance. Finally, we multiply F by a scale parameter β and perform an element-wise summation operation with feature X to obtain the final output $G \in \mathrm{R}^{\mathrm{C} \times \mathrm{H} \times \mathrm{W}}$. The procedure of the GIF module can be represented as follows:

$$G_i = \beta \bullet F_i + X_i \tag{5}$$

where β is learnable weight and initialized as 0.

From Eq. (5), the resulting feature G at each position is a weighted sum of the features across all positions and original features, which has a global contextual view. The output of the GIF module is shown in Eq. (5), which is used as the input of Decoder II.

D. Joint Loss Function

A main challenge in joint segmentation is class distribution imbalance. In order to optimize our model further, we employ a joint loss function to guide the optimization of the proposed GD-Net.

For the HRF segmentation in Decoder I, the binary cross-entropy loss and the dice loss function [42] are adopted to jointly optimize the branch:

$$L_{HRF} = L_{BCE} + L_{Dice} \tag{6}$$

$$L_{BCE} = -\sum\nolimits_{h,w} \left[(1-y) \bullet \log\left(1-\hat{y}\right)^{h,w} + y \bullet \log\left(\hat{y}\right)\right] \tag{7}$$

$$L_{Dice} = 1 - \frac{2\left(y \bullet \hat{y}\right)}{sum\left(y\right) + sum\left(\hat{y}\right)} \tag{8}$$

where y and $\hat{y}$ represent the ground truth and the corresponding HRF segmentation results, (w, h) is the coordinates of the pixel.

For the joint segmentation of HE and MG, to reduce the interference of background information and improve the accuracy of lesions segmentation, multi-category cross entropy loss and foreground Dice loss function are jointly adopted to optimize the other branch:

$$L_{Joint} = L_{MCE} + L_{F-Dice} \tag{9}$$

$$L_{MCE} = -\sum\nolimits_{C=1}^{N} \sum\nolimits_{h,w} \left[(1-v_c) \bullet \log\left(1-\hat{v}_c\right)^{h,w} + v_c \bullet \log\left(\hat{v}_c\right)\right] \tag{10}$$

$$L_{F-Dice} = 1 - \frac{1}{N} \sum\nolimits_{C=1}^{N} \frac{2\left(v_c \bullet \hat{v}_c\right)}{sum\left(v_c\right) + sum\left(\hat{v}_c\right)} \tag{11}$$

where C indicates the class being evaluated and N is the total number of classes segmented, which is 2 in this paper. v_1 and $\hat{v}_1$ represent the ground truth and the segmentation results for HE, and v_2 and $\hat{v}_2$ represent the ground truth and the segmentation results for MG.

The total joint loss function is as follows:

$$L_{Total} = \alpha \bullet L_{Joint} + (1-\alpha) \bullet L_{HRF} \tag{12}$$

where α is the loss weight of HE and MG joint segmentation task with a value less than 1. The value of α is finally set to 0.4 in this paper, and subsequent experiments will also verify it.

From Eq. (7) and (10), it can be seen that the cross-entropy loss function is mainly based on pixel-level pixel classification to guide the network to focus on the local information of the image. As shown in Eq. (8) and (11), the Dice loss function is mainly used to guide network optimization based on image-level errors and pay more attention to the global features of the lesions. It should be noted that the proportion of HE and MG lesions in OCT images is very small. In order to strengthen the supervision of the target area segmentation and reduce the interference of background information, the background Dice loss is ignored when training the model. The foreground Dice guidance of the lesion area segmentation is adopted, while the background segmentation loss is mainly guided by the pixel-level cross-entropy loss. Therefore, the total joint loss function Eq. (12) can not only ensure the accuracy and continuity of the lesions segmentation, but also reduce the influence of the background and reducing false positives.

TABLE I
DATASET USED FOR FOUR-FOLD CROSS VALIDATION IN THIS STUDY

	Training		Testing	
	Patient ID	Images	Patient ID	Images
Fold 1	1, 3, 4, 5, 6, 7	152	2	50
Fold 2	2, 4, 6, 7	151	1, 3, 5	51
Fold 3	1, 2, 3, 5, 6	152	4, 7	50
Fold 4	1, 2, 3, 4, 5, 7	151	6	51

III. EXPERIMENTS AND RESULTS

A. Dataset

Seven retinal OCT scans from seven patients diagnosed with DME at the department of ophthalmology, University Medical Center Groningen, The Netherlands were included in this study. OCT files with poor quality evaluated by Canon OCT built-in software with a score of Image Quality < 4 were excluded. All patients underwent examination with the Canon OCT-HS100 device (version 4.4.0.13) using the Macula 3D mode. The OCT scanning area was $10 \times 10 \times 2$ mm^3 centered on the fovea, corresponding to $128 \times 1024 \times 1176$ pixels. This study adhered to the tenets of Declaration of Helsinki and was approved by the Medical Ethics Review Board of University Medical Center Groningen, The Netherlands (METc number: METc2019/599). Informed consents were obtained from each patient.

Manual labeling of HRF B-scans was performed by a senior ophthalmologist using ITK-SNAP software (version 3.4.0) [43]. Two types of HRF, HE and activated MG, were identified and labelled separately with different colors. Specifically, HE was defined as particles larger than 40 μm with back shadowing and reflectivity similar to the retinal pigment epithelium-Bruch complex24. Activated MG was defined as particles 20 to 40 μm in diameter with similar reflectivity to the nerve fiber layer without back shadowing on OCT. Signals smaller than 20 μm were regarded as noise and excluded.

Based on above data rules, 202 valid B-scan images are selected to establish the joint segmentation dataset. Meanwhile, a four-fold cross validation strategy is performed to objectively validate the performance of the proposed method. The training set and testing set of four-fold are shown in the Table I, which are divided according to patients.

For each fold, 100 epochs are trained. We resize the images from 1024×1176 to 1024×1024. Considering that the lesions are small and have blurred boundaries and uneven distribution (some lesions are located at the border of the image), we did not use any affine transformation and Gaussian noise addition for data augmentation, but only used flipping.

TABLE II
SEGMENTATION RESULTS OF HE, MG AND THEIR MEAN VALUES IN THE CONTRAST EXPERIMENTS OF DIFFERENT NETWORKS (MEAN ± STANDARD DEVIATION)

	DSC (%)		IoU (%)		Recall (%)		Precision (%)	
	HE	Mean	HE	Mean	HE	Mean	HE	Mean
	MG		MG		MG		MG	
CE-Net	69.84±3.32	57.31±1.35	55.21±3.07	42.59±1.15	70.34±4.80	57.50±2.20	73.15±4.06	63.24±4.26
	44.77±4.23		29.98±3.73		44.67±4.34		53.34±12.36	
DeepLab	67.08±2.36	55.25±2.86	52.19±1.76	40.58±2.54	65.79±2.41	57.06±3.42	72.12±4.18	59.32±4.88
	43.42±6.38		28.78±5.13		48.32±5.28		46.51±13.50	
PSPNet	68.71±2.73	54.25±2.30	53.42±2.51	39.66±1.89	67.71±1.23	52.41±1.44	73.51±4.14	62.29±4.17
	39.80±4.44		25.90±3.58		37.11±2.72		51.08±11.93	
CPFNet	70.48±4.04	57.52±2.51	56.05±3.95	42.94±2.29	69.55±3.27	56.81±3.23	75.84±4.55	64.47±4.11
	44.57±5.49		29.82±4.68		44.07±5.66		53.10±10.11	
U-Net	72.31±2.93	60.43±1.40	58.50±2.77	45.88±1.05	**72.58±5.01**	61.15±3.55	77.10±3.06	66.37±6.22
	48.55±4.89		33.27±4.22		49.72±4.58		55.20±12.38	
U-Net++	72.14±3.66	61.25±2.70	58.35±3.40	46.72±2.49	70.93±1.94	61.50±2.45	77.41±4.47	65.61±4.73
	50.36±6.50		35.09±5.88		52.07±5.81		53.82±9.44	
Att U-Net	71.32±4.58	60.58±1.70	57.76±3.69	46.21±1.17	69.44±4.56	60.35±2.99	79.34±4.62	66.65±3.13
	49.84±4.65		34.67±4.41		51.25±4.36		53.96±9.48	
GD-Net	**73.47±3.33**	**62.76±2.28**	**59.88±2.99**	**48.32±2.03**	71.72±3.35	**62.72±0.95**	**79.73±2.15**	**67.52±4.05**
	52.06±5.94		**36.76±5.50**		**53.72±2.74**		**55.32±9.08**	

B. Evaluation Metrics

In order to evaluate the performance of the proposed GD-Net objectively, the metrics shown in Eq. (13), (14), (15), (16) are adopted.

$$DSC = \frac{2TP}{2TP + FP + FN} \tag{13}$$

$$IoU = \frac{TP}{TP + FP + FN} \tag{14}$$

$$Recall = \frac{TP}{TP + FN} \tag{15}$$

$$Precision = \frac{TP}{TP + FP} \tag{16}$$

where TP, FP, TN and FN represent true positive, false positive, true negative and false negative, respectively. Dice similarity coefficient (DSC) can be used to compare the similarity between ground truth and results [5], [44]. Intersection over union (IoU), also known as Jaccard index, is the main measure of overlap between ground truth and segmentation results [45]. Moreover, DSC and IoU are the main evaluation metrics in experiments. In addition, since automatic joint segmentation of lesions is used to assist doctors in diagnosing and analyzing retinal diseases, recall and precision are also included in our comprehensive experiments [46].

C. Implementation Details

The implementation of the proposed GD-Net is based on the public platform PyTorch and NVIDIA GeForce 3090 GPU with 24GB memory. In the joint training process, the Poly strategy is used to set the initial learning rate to 1e-4. The learning rate of the entire joint learning stage is 5e-4, and the optimizer is Adam. The Batch size is set as 2. Moreover, the code of the proposed GD-Net will be released in: https://github.com/cpyao20/GD-Net.

D. Comparison to Other State-of-the-Art Deep Learning Based Networks

In this section, we compare our proposed GD-Net with other state-of-the-art segmentation networks, such as CE-Net [47], DeepLab [48], PSPNet [49], CPFNet [50], U-Net [41], U-Net++ [51] and Att U-Net [52]. We calculate the metrics of HE and MG and their mean values respectively and the comparison results for all methods are listed in Table II.

It can be seen from Table II that the proposed GD-Net outperforms almost all state-of-the-art methods. Although U-Net performs well when compared with several other networks, the main evaluation metrics, DSC and IoU of GD-Net, are improved by 1.60%, 2.36% on HE and 7.23%, 10.49% on MG respectively. Compared with PSPNet, which achieves the worst results with 68.71% and 39.80% for DSC, the metrics of our method achieve a remarkable improvement and reach 73.47% and 52.06% respectively. Moreover, compared with U-Net++, which achieves the second best Recall on MG, our proposed GD-Net is improved by 3.17%, which indicates that our method can detect small targets precisely.

It also can be seen from Table II that our proposed method achieves best segmentation results on mean values of all metrics, especially for DSC and IoU. Besides, compared with the excellent U-Net, the evaluation metrics of the GD-Net are increased by 2.47%, 3.45%, 2.57%, and 1.73% in DSC, IoU, Recall and Precision, respectively.

An example of segmentation results with different methods is shown in Fig. 4 and from which it can be seen that our proposed method achieves best segmentation results. Compared with other state-of-the-art methods, our proposed method can not only accurately segment the lesion with small sizes, but also maintain good regional continuity in segmenting large targets. These results further demonstrate the effectiveness of the proposed method.

Meanwhile, to further investigate the statistical significance of the performance improvement by the proposed GD-Net over other state-of-the-art segmentation networks, we conduct T-test

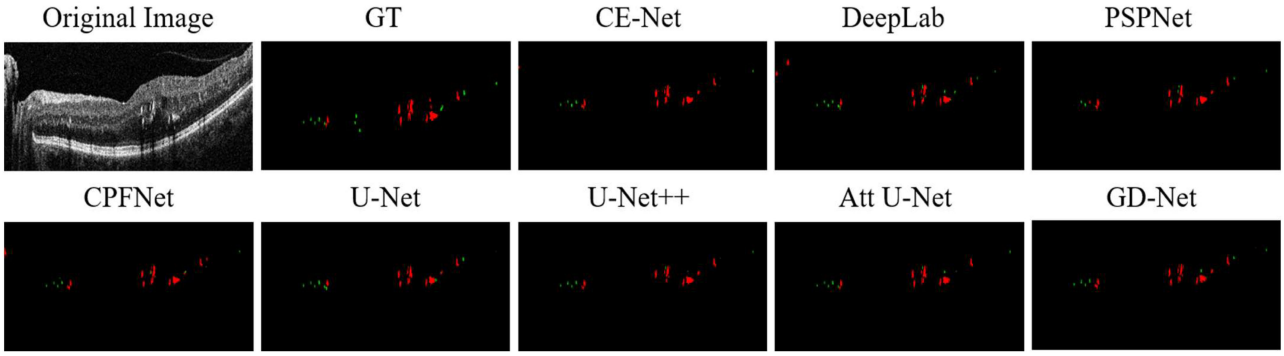

Fig. 4. Segmentation examples of HE and MG with different networks. The red and green regions represent HE and MG, respectively.

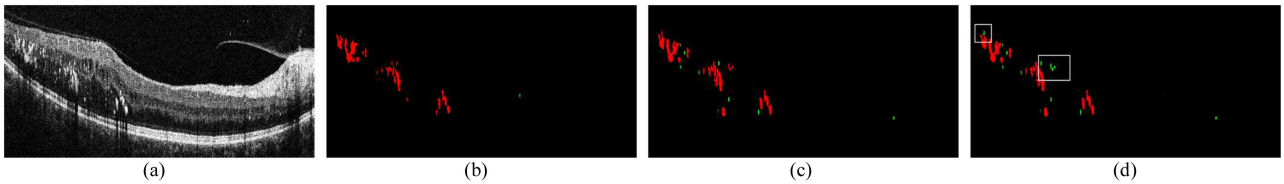

Fig. 5. Segmentation example of U-Net and threshold method. (a) OCT B-scan image. (b) The ground truth. (c) U-Net's multi-class segmentation result (HE, MG). (d) Threshold method's segmentation result (HE, MG).

TABLE III
STATISTICAL ANALYSIS (P-VALUE) OF THE PROPOSED GD-NET VERSE OTHER STATE-OF-THE-ART SEGMENTATION NETWORKS

	Mean_DSC	Mean_IoU
GD-Net - CE-Net	0.005	0.003
GD-Net - DeepLab	0.002	0.002
GD-Net - PSPNet	0.001	0.001
GD-Net - CPFNet	0.002	0.003
GD-Net - U-Net	0.017	0.021
GD-Net - U-Net++	0.008	0.011
GD-Net - Att U-Net	0.035	0.024

and the p-values are listed in Table III. As can be observed in Table III, all of the DSC and IoU improvement on mean values are statistically significant with p-values less than 0.05, which further demonstrates the effectiveness of the proposed GD-Net.

E. Comparison of Threshold Method

Threshold has been used in many previous researches on image segmentation and HRF detection [32], [53], [54]. In this section, therefore, we compare the convolutional neural network method with the traditional threshold method. We perform experiments on U-Net, which have achieved the second best performance in comparison experiments.

We use a U-Net to segment the entire HRF, and then differentiate it by applying a threshold on dimensions of the segmented objects. After several trials, we set the optimal threshold, 63 pixels, as the basis for segmentation. If the size is larger than 63 pixels, it will be classified as HE, while those with a size smaller than 63 pixels are classified as MG. The results are compared with those obtained by using U-Net directly for multi-class segmentation, which is shown in Table IV and Fig. 5.

It can be seen from Table IV that the calculation results of U-Net on all metrics perform better, which indicates the better effectiveness and practicability of CNN method. From the comparison of Fig. 5(c) and (d), we can clearly see the difference between the results of U-Net and the threshold method. The areas marked by the white box in (d) are the areas that have been erroneously segmented.

F. Comparison of Model Variations

The U-Net is used as the baseline of this experiment. We conduct ablation experiments and calculate the metrics of HE and MG and their mean values respectively.

Ablation experiment for GIF module: As shown in Table V, an ablation experiment is conducted to evaluate the proposed GIF module. Compared with the baseline model, our Baseline+GIF achieves improvement in main evaluation metrics of MG (1.44% for DSC, 1.74% for IoU and 5.71% for Recall), which benefits from the fact that GIF module can aggregate the high-level comprehensive semantic information and capture the missing information from the negative correlation map to improve the performance on small targets. Moreover, Baseline+GIF achieves remarkable performance on Recall, which proves the effectiveness of the proposed module.

Ablation experiment for DDCW: We also conduct experiment to demonstrate the effectiveness of our proposed DDCW. It can be seen from Table V that, compared with the baseline, Baseline+DDCW achieves higher metrics in all four evaluation metrics except the Precision of MG and mean values. Especially, compared with the Baseline, the main evaluation metrics, DSC and IoU of Baseline+DDCW, are improved by 2.22% and 3.51% and reaches 61.77% and 47.49% on mean values, respectively.

TABLE IV
SEGMENTATION RESULTS OF HE, MG AND THEIR MEAN VALUES OF U-NET AND THRESHOLD METHOD (MEAN ± STANDARD DEVIATION)

	DSC (%)		IoU (%)		Recall (%)		Precision (%)	
	HE	Mean	HE	Mean	HE	Mean	HE	Mean
	MG		MG		MG		MG	
U-Net	**72.31±2.93**	**60.43±1.40**	**58.50±2.77**	**45.88±1.05**	**72.58±5.01**	**61.15±3.55**	**77.10±3.06**	**66.37±6.22**
	48.55±4.89		**33.27±4.22**		**49.72±4.58**		**55.20±12.38**	
Threshold method	66.07±4.24	53.54±1.97	50.96±3.98	38.88±1.94	62.45±6.44	50.84±6.23	75.94±4.42	65.02±7.65
	41.01±3.22		26.81±2.83		39.23±8.07		54.10±14.10	

TABLE V
SEGMENTATION RESULTS OF HE, MG AND THEIR MEAN VALUES IN THE ABLATION EXPERIMENTS OF MODEL VARIATIONS (MEAN ± STANDARD DEVIATION)

	DSC (%)		IoU (%)		Recall (%)		Precision (%)	
	HE	Mean	HE	Mean	HE	Mean	HE	Mean
	MG		MG		MG		MG	
Baseline	72.31±2.93	60.43±1.40	58.50±2.77	45.88±1.05	72.58±5.01	61.15±3.55	77.10±3.06	66.37±6.22
	48.55±4.89		33.27±4.22		49.72±4.58		55.20±12.38	
Baseline+GIF	71.24±2.70	60.25±1.92	57.05±2.32	45.45±1.68	**72.97±5.69**	62.77±4.18	73.12±3.78	62.65±5.28
	49.25±4.80		33.85±4.29		52.56±5.20		52.19±12.06	
Baseline+DDCW	73.31±3.45	61.77±2.15	59.89±3.10	47.49±1.87	72.64±2.44	61.96±2.09	77.58±3.17	66.27±4.33
	50.24±5.34		35.09±4.94		51.28±2.22		54.95±11.00	
GD-Net (α = 0.2)	73.21±3.52	62.28±2.08	59.52±3.25	47.74±1.78	71.33±4.30	63.25±3.60	79.05±3.30	66.19±4.54
	51.35±5.53		35.97±5.08		**55.17±4.71**		53.32±11.14	
GD-Net (α = 0.4)	**73.47±3.33**	**62.76±2.28**	**59.89±2.99**	**48.32±2.03**	71.72±3.35	62.72±0.95	**79.73±2.15**	**67.52±4.05**
	52.06±5.94		**36.76±5.50**		53.72±2.74		**55.32±9.08**	
GD-Net (α = 0.5)	73.19±2.77	62.19±2.07	59.31±2.76	47.57±2.00	72.51±3.87	**63.59±2.57**	77.68±3.21	65.67±4.78
	51.20±5.92		35.83±5.67		54.68±4.48		53.66±12.44	
GD-Net (α = 0.7)	73.14±2.12	62.12±2.73	59.41±1.87	47.61±2.52	71.79±1.75	62.45±3.28	78.57±1.45	66.93±5.39
	51.10±6.48		35.80±5.75		53.10±6.60		55.27±11.79	

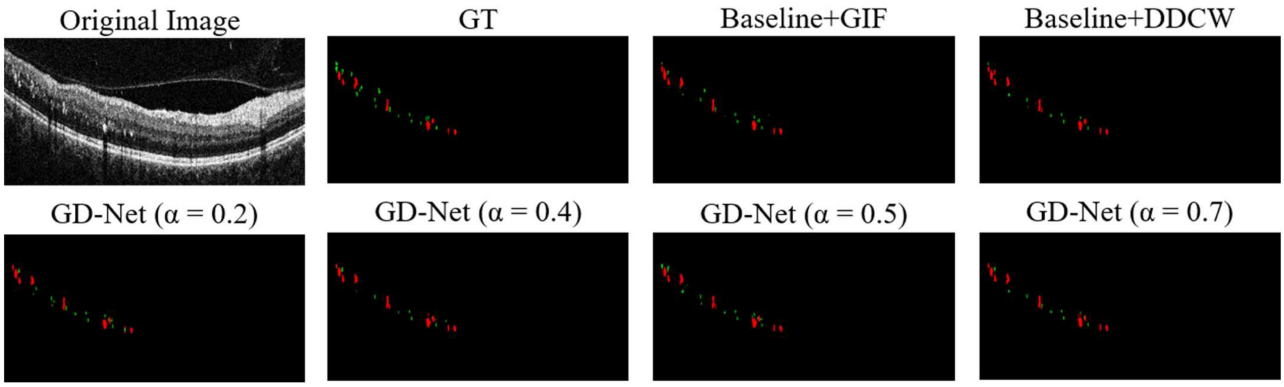

Fig. 6. Segmentation examples of HE and MG with model variations. The red and green regions represent HE and MG, respectively.

Ablation experiment for different parameter α: We conduct experiments on GD-Net (Baseline+GIF+DDCW) with different parameter α, where α is the loss weight of multi-class joint segmentation task with a value less than 1, as shown in Eq. (12). Plenty of experiments are conducted and we choose four sets of data with good performance, as shown in Table V. GD-Net achieves the best performance on DSC, IoU and Precision when $\alpha = 0.4$. Although GD-Net performs well when $\alpha = 0.2$, 0.5, 0.7, compared with the second best result, Mean DSC and Mean IoU are improved by 0.77% and 1.21% respectively when $\alpha = 0.4$. These results indicate that the suitable parameter can keep better balance with the proportion of pixel-level and image-level, which can improve the performance of the network.

An example of segmentation results with models variations is shown in Fig. 6. It can be seen from it that the proposed GD-Net can preserve the structure of HE and MG more completely and present more details, which effectively improves the accuracy of segmentation and reduces false positives. It is also obvious that GD-Net is closer to GT when $\alpha = 0.4$, which meets the medical accuracy requirements for HE and MG segmentation tasks.

G. Inter- and Intra- observer Error

To further demonstrate the research significance of this study and the reliability of proposed algorithm, we conduct experiments to report the inter- or intra- observer error and compare it to the results of the algorithm.

Ground truth in original dataset is abbreviated as GT. In addition, we invite another senior ophthalmologist to manually mark the HRF B-scans twice, which are named as GT_Dr.2_1 and GT_Dr.2_2 respectively. The inter-observer error is verified on

TABLE VI
THE RESULTS OF PROPOSED ALGORITHM AND INTER- AND INTRA- OBSERVER ERROR (MEAN ± STANDARD DEVIATION)

	DSC (%)		IoU (%)		Recall (%)		Precision (%)	
	HE	Mean	HE	Mean	HE	Mean	HE	Mean
	MG		MG		MG		MG	
Proposed Algorithm --- GT	73.47±3.33	62.76±2.28	59.88±2.99	48.32±2.03	71.72±3.35	62.72±0.95	79.73±2.15	67.52±4.05
	52.06±5.94		36.76±5.50		53.72±2.74		55.32±9.08	
Inter: GT --- GT_Dr.2_1	63.26±2.22	61.04±2.32	52.88±2.53	49.93±2.64	69.64±5.38	67.61±6.32	79.68±4.70	77.63±4.43
	58.82±2.66		46.97±3.09		65.59±7.37		75.58±4.43	
Intra: GT_Dr.2_1 --- GT_Dr.2_2	66.81±2.83	64.99±2.54	58.57±3.46	56.15±3.07	70.80±4.92	69.10±5.81	84.82±4.56	83.17±4.29
	63.16±2.52		53.73±2.99		67.40±6.82		81.52±4.07	

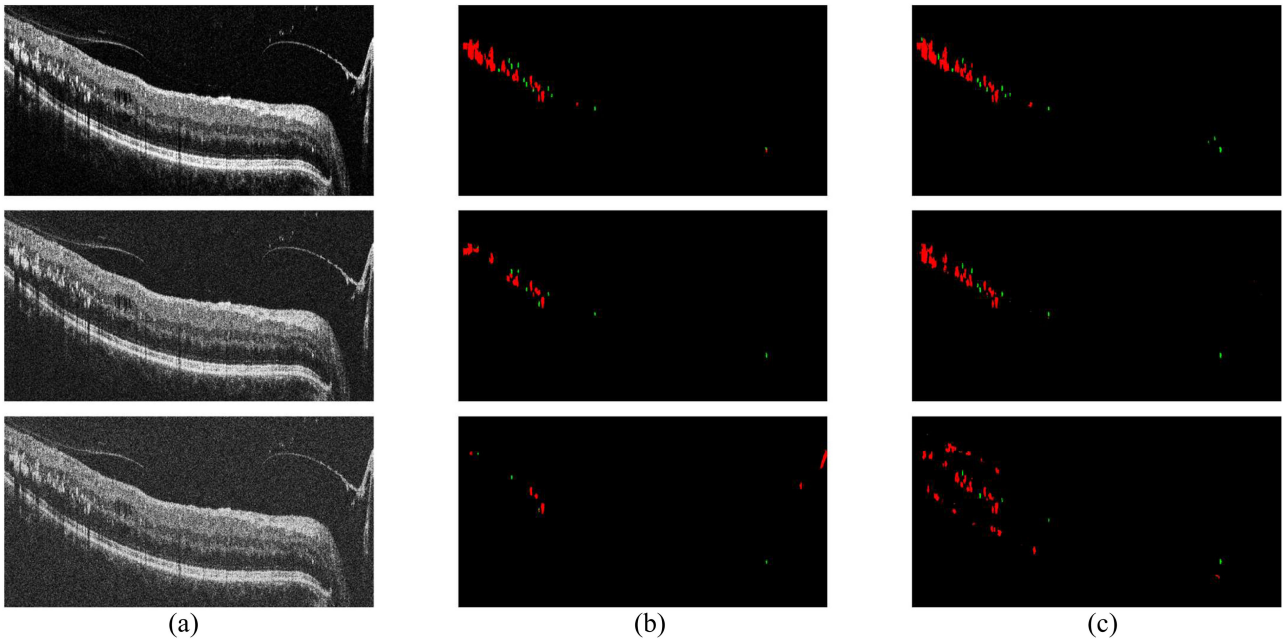

Fig. 7. Col. (a), from top to bottom are original B-scan, the B-scan added 30 dB and 60 dB Gaussian noise respectively. Col. (b), the segmentation results of the proposed network with different levels of noise. Col. (c), the segmentation results of U-Net with different levels of noise.

GT and GT_Dr.2_1, while the intra-observer error is verified on GT_Dr.2_1 and GT_Dr.2_2. The results are shown in Table VI.

Table VI shows the comparison between the three sets of data. Our analysis is based on the main evaluation metrics, DSC. The inter-observer error is 61.04%, which is lower than the proposed algorithm in mean values and indicates that our method has quite good reliability. The intra-observer error is slightly higher than the proposed algorithm, which is 64.99%. This result also shows that this segmentation task is fairly challenging. It is especially worth noting that the proposed algorithm has achieved excellent results on HE, but the performance on MG is slightly worse than the ophthalmologist's manual marking and needs to be improved.

H. Comparison of Different Levels of Noise

The OCT B-scans used for the network training were collected by Canon OCT-HS100 with Macula 3D mode and they were not preprocessed with noise reduction. As mentioned above, considering that the lesions are small and have blurred boundaries and uneven distribution, we do not use Gaussian noise addition for data augmentation in the training stage. However, to report the level of persistence of the proposed algorithm, the discussion about the effect of noise is necessary. In the test stage, we add 30 dB and 60 dB Gaussian noise to OCT B-scans respectively to evaluate the proposed network and U-Net, which has achieved the second best performance in comparison experiments. An example of segmentation results is shown in Fig. 7.

It can be seen from Fig. 7 that adding different levels of noise makes the performance of the proposed algorithm and U-Net worse. After analysis, we believe that the extra semantic information of the lesions may be interfered by noise.

IV. CONCLUSION AND DISCUSSIONS

In this paper, we propose a novel GD-Net combined two newly proposed modules of global information fusion (GIF) and dual decoder collaborative workspace (DDCW) for joint segmentation of HE and MG in retinal OCT images, which mainly focuses on solving two problems: (1) the interferences caused by similar pathological features between HE and MG. (2) the difficulties in joint segmentation resulted from a variety of shapes and location distribution, blurred boundaries, small size in morphology. To the best of our knowledge, it

is the first work to explore the joint segmentation of HE and MG in retinal OCT images. Comprehensive experiments have been conducted to evaluate the effectiveness of the proposed method. The experimental results show that compared with other state-of-the-art algorithms, the segmentation performance of our proposed GD-Net has been improved obviously.

There are still some limitations on this study: (1) All comparison algorithms and the proposed GD-Net has been trained and evaluated based on small-scale dataset. (2) No noise has been added to the OCT B-scans used for the network training. Although the proposed GD-Net has achieved better performance, we believe that if more quality data can be collected and different levels of noise can be supported, the performance of the proposed network will be further improved. Therefore, improving the level of persistence of the algorithm to the level of speckle noise will definitely be the focus of our follow-up work. It is also one of our future works to collect more data, with further evaluation and extension of our proposed module and network.

REFERENCES

[1] J. Hu, "The clinical application of optical coherence tomography in macular diseases," *Chin. J. Practical Ophthalmol.*, vol. 17, no. 02, pp. 66–68, 1999.

[2] X. Liu and J. Huang, "The application of optical coherence tomography in clinical ophthalmology in my country," *Ophthalmol. CHN*, vol. 13, no. 4, pp. 196–199, 2004.

[3] X. Liu and J. Huang, "The application of optical coherence tomography in clinical ophthalmology in my country," *Ophthalmol. CHN*, vol. 13, no. 5, pp. 260–263, 2004.

[4] J. M. Schmitt *et al.*, "Speckle in optical coherence tomography," *J. Biomed. Opt.*, vol. 4, no. 1, pp. 95–105, 1999.

[5] D. Xiang *et al.*, "Automatic segmentation of retinal layer in OCT images with choroidal neovascularization," *IEEE Trans. Image Process.*, vol. 27. no. 12, pp. 5880–5891, Dec. 2018.

[6] S. Zhu *et al.*, "Choroid neovascularization growth prediction with treatment based on reaction-diffusion model in 3D OCT images," *IEEE J. Biomed. Health Inform.*, vol. 21, no. 6, pp. 1667–1674, Nov. 2017.

[7] S. Vujosevic *et al.*, "Hyperreflective intraretinal spots in diabetics without and with nonproliferative diabetic retinopathy: An in vivo study using spectral domain OCT," *J. Diabetes Res.*, vol. 2013, 2013, Art. no. 491835.

[8] J. G. Christenbury *et al.*, "Progression of intermediate age-related macular degeneration with proliferation and inner retinal migration of hyperreflective foci," *Ophthalmology*, vol. 120, no. 5, pp. 1038–1045, 2013.

[9] K. Sleiman *et al.*, "Optical coherence tomography predictors of risk for progression to non-neovascular atrophic age-related macular degeneration," *Ophthalmology*, vol. 124, no. 12, pp. 1764–1777, 2017.

[10] G. Panozzo *et al.*, "An optical coherence tomography-based grading of diabetic maculopathy proposed by an international expert panel: The european school for advanced studies in ophthalmology classification," *Eur. J. Ophthalmol.*, vol. 30, no. 1, pp. 8–18, 2019.

[11] Y. R. Chung *et al.*, "Role of inflammation in classification of diabetic macular edema by optical coherence tomography," *J. Diabetes Res.*, vol. 2019, p. 8164250, 2019, Art. no. 8164250.

[12] S. Vujosevic *et al.*, "Hyperreflective retinal spots in normal and diabetic eyes: B-scan and en face spectral domain optical coherence tomography evaluation," *Retina*, vol. 37, no. 6, pp. 1092–1103, 2016.

[13] H. W. Van Dijk *et al.*, "Selective loss of inner retinal layer thickness in type 1 diabetic patients with minimal diabetic retinopathy," *Invest. Ophthalmol. Vis. Sci.*, vol. 50, no. 7, pp. 3404–3409, 2009.

[14] H. W. Van Dijk *et al.*, "Decreased retinal ganglion cell layer thickness in patients with type 1 diabetes," *Invest. Ophthalmol. Vis. Sci.*, vol. 51, no. 7, pp. 3660–3665, 2010.

[15] H. W. Van Dijk *et al.*, "Early neurodegeneration in the retina of type 2 diabetic patients," *Invest. Ophthalmol. Vis. Sci.*, vol. 53, no. 6, pp. 2715–2719, 2012.

[16] B. Miljanovic *et al.*, "A prospective study of serum lipids and risk of diabetic macular edema in type 1 diabetes," *Diabetes*, vol. 53, no. 11, p. 2883, 2004.

[17] T. Otani and S. Kishi, "Tomographic findings of foveal hard exudates in diabetic macular edema," *Amer. J. Ophthalmol.*, vol. 131, no. 1, pp. 50–54, 2001.

[18] L. Chen *et al.*, "Distribution, markers, and functions of retinal microglia," *Ocular Immunol. Inflammation*, vol. 10, no. 1, pp. 27–39, 2002.

[19] S. F. Abcouwer, "Müller cell-microglia cross talk drives neuroinflammation in diabetic retinopathy," *Diabetes*, vol. 66, no. 2, pp. 261–263, 2017.

[20] F. S. Sorrentino *et al.*, "The importance of glial cells in the homeostasis of the retinal microenvironment and their pivotal role in the course of diabetic retinopathy," *Life Sci.*, vol. 162, pp. 54–59, 2016.

[21] A. I. Arroba and Á. M. Valverde, "Modulation of microglia in the retina: New insights into diabetic retinopathy," *Acta Diabetologica*, vol. 54, no. 6, pp. 527–533, 2017.

[22] V. Schreur *et al.*, "Retinal hyperreflective foci in type 1 diabetes mellitus," *Retina*, vol. 40, no. 8, pp. 1565–1573, 2020.

[23] H. H. Seong *et al.*, "Association between hyperreflective dots on spectral-domain optical coherence tomography in macular edema and response to treatment," *Invest. Ophthalmol. Vis. Sci.*, vol. 58, no. 13, pp. 5958–5967, 2017.

[24] A. Fonollosa *et al.*, "Predictive capacity of baseline hyperreflective dots on the intravitreal dexamethasone implant (Ozurdex(R)) outcomes in diabetic macular edema: A multicenter study," *Graefe's Arch. Clin. Exp. Ophthalmol.*, vol. 257, no. 1, pp. 2381–2390, 2019.

[25] H. Lee *et al.*, "Association between soluble CD14 in the aqueous humor and hyperreflective foci on optical coherence tomography in patients with diabetic macular edema," *Invest. Ophthalmol. Vis. Sci.*, vol. 59, no. 2, pp. 715–721, 2018.

[26] G. Coscas *et al.*, "Hyperreflective dots: A new spectral-domain optical coherence tomography entity for follow-up and prognosis in exudative age-related macular degeneration," *Ophthalmologica*, vol. 229, no. 1, pp. 32–37, 2013.

[27] H. Lee *et al.*, "Correlation between optical coherence tomographic hyperreflective fociand visual outcomes after anti-VEGF treatment in neovascular age-related macular degeneration and polypoidal choroidal vasculopathy," *Retina*, vol. 36, no. 3, pp. 465–475, 2016.

[28] E. Korot *et al.*, "Algorithm for the measure of vitreous hyperreflective foci in optical coherence tomographic scans of patients with diabetic macular edema," *JAMA Ophthalmol.*, vol. 134, no. 1, pp. 15–20, 2016.

[29] O. Segal *et al.*, "Prognostic value of hyperreflective foci in neovascular age-related macular degeneration treated with bevacizumab," *Retina*, vol. 36, no. 11, pp. 2175–2182, 2016.

[30] P. M. Battaglia *et al.*, "Hyperreflective foci number correlates with choroidal neovascularization activity in angioid streaks," *Invest. Ophthalmol. Vis. Sci.*, vol. 59, no. 8, pp. 3314–3319, 2018.

[31] M. Mokhtari, Z. Ghasemi Kamasi, and H. Rabbani, "Automatic detection of hyperreflective foci in optical coherence tomography B-scans using morphological component analysis," in *Proc. 39th Annu. Int. Conf. IEEE Eng. Med. Biol. Soc.*, 2017, pp. 1497–1500.

[32] S. Xie *et al.*, "Segmentation of bright speckles in 3D SD-OCT diabetic retinal images based on self-adaption threshold," *Comput. Eng. Appl.*, vol. 54, no. 22, pp. 211–215, 2018.

[33] I. P. Okuwobi *et al.*, "Automated segmentation of hyperreflective foci in spectral domain optical coherence tomography with diabetic retinopathy," *J. Med. Imag.*, vol. 5, no. 1, pp. 1–16, 2018.

[34] I. P. Okuwobi *et al.*, "Automated quantification of hyperreflective foci in SD-OCT with diabetic retinopathy," *IEEE J. Biomed. Health Inform.*, vol. 24, no. 4, pp. 1125–1136, Apr. 2020.

[35] J. H. Tan *et al.*, "Automated segmentation of exudates, haemorrhages, microaneurysms using single convolutional neural network," *Inf. Sci.*, vol. 420, pp. 66–76, 2017.

[36] A. Benzamin and C. Chakraborty, "Detection of hard exudates in retinal fundus images using deep learning," in *Proc. Joint 7th Int. Conf. Inform., Electron. Vis. (ICIEV) 2018 2nd Int. Conf. Imag., Vis. Pattern Recognit.*, 2018, pp. 465–469.

[37] M. Mateen *et al.*, "Exudate detection for diabetic retinopathy using pretrained convolutional neural networks," *Complexity*, vol. 2020, no. 1, pp. 1–11, 2020.

[38] T. Schlegl *et al.*, "Fully automated segmentation of hyperreflective foci in optical coherence tomography images," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, 2018, Art. no. 1131308.

[39] M. Katona *et al.*, "Automatic detection and characterization of biomarkers in OCT images," in *Proc. Int. Conf. Image Anal. Recognit.*, 2018, pp. 706–714.
[40] L. Varga *et al.*, "Automatic segmentation of hyperreflective foci in OCT images," *Comput. Methods Programs Biomed.*, vol. 178, pp. 91–103, 2019.
[41] O. Ronneberger *et al.*, "U-net: Cssonvolutional networks for biomedical image segmentation," in *Proc. Int. Conf. Med. image Comput. Comput.-Assist. Intervention*, 2015, pp. 234–241.
[42] F. Milletari *et al.*, "V-net: Fully convolutional neural networks for volumetric medical image segmentation," in *Proc. Fourth Int. Conf. 3D Vis.*, 2016, pp. 565–571.
[43] P. A. Yushkevich and G. Gerig, "ITK-SNAP: An intractive medical image segmentation tool to meet the need for expert-guided segmentation of complex medical images," *IEEE Pulse*, vol. 8, no. 4, pp. 54–57, Jul./Aug. 2017.
[44] W. Ju *et al.*, "Random walk and graph cut for cosegmentation of lung tumor on PET-CT images," *IEEE Trans. Image Process.*, vol. 24, no. 12, pp. 5854–5867, 2015.
[45] J. Long, Evan Shelhamer, and Trevor Darrell, "Fully convolutional networks for semantic segmentation," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, 2015, pp. 3431–3440.
[46] H. H. Chang *et al.*, "Performance measure characterization for evaluation neuroimage segmentation algorithms," *Neuroimage*, vol. 47, no. 1, pp. 122–135, 2009.
[47] Z. Gu *et al.*, "CE-net: Context encoder network for 2D medical image segmentation," *IEEE Trans. Med. Imag.*, vol. 38, no. 10, pp. 2281–2292, Oct. 2019.
[48] L. C. Chen *et al.*, "DeepLab: Semantic image segmentation with deep convolutional nets, atrous convolution, and fully connected CRFs," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 40, no. 4, pp. 834–848, Apr. 2018.
[49] H. Zhao *et al.*, "Pyramid scene parsing network," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, pp. 2881–2890, 2017.
[50] S. Feng *et al.*, "CPFNet: Context pyramid fusion network for medical image segmentation," *IEEE Trans. Med. Imag.*, vol. 39, no. 10, pp. 3008–3018, Oct. 2020.
[51] Z. Zhou *et al.*, "UNet++: A nested U-Net architecture for medical image segmentation," in *Proc. Deep Learn. Med. Image Anal. (DLMIA) Multimodal Learn. Clin. Decis. Support*, 2018, pp. 1–13.
[52] O. Oktay *et al.*, "Attention U-Net: Learning where to look for the pancreas," in *Proc. Int. Conf. Med. image with Deep Learning (MIDL)*, 2018, pp. 1–10.
[53] J. Lammer *et al.*, "Detection and analysis of hard exudates by polarization-sensitive optical coherence tomography in patients with diabetic maculopathy," *Invest. Ophthalmol. Vis. Sci.*, vol. 55, no. 3, pp. 1564–1571, 2014.
[54] T. Walter *et al.*, "A contribution of image processing to the diagnosis of diabetic retinopathy: Detection of exudates in color fundus images of the human retina," *IEEE Trans. Med. Imag.*, vol. 21, no. 10, pp. 1236–1243, Oct. 2002.