

# RC-Net: A region-level context network for hyperreflective dots segmentation in retinal OCT images

Bo Zhang<sup>a</sup>, Hui Zhao<sup>b</sup>, Mingwei Si<sup>b</sup>, Wenxuan Cui<sup>b</sup>, Yuanfeng Zhou<sup>c</sup>, Shujun Fu<sup>a,\*</sup>, Hong Wang<sup>b,\*</sup>

<sup>a</sup> School of Mathematics, Shandong University, Jinan, 250100, China

<sup>b</sup> Department of Ophthalmology, Qilu Hospital of Shandong University, Jinan, 250001, China

<sup>c</sup> School of Software, Shandong University, Jinan, 250101, China

## ARTICLE INFO

Dataset link: <https://github.com/a1486364635/RCNet>

### Keywords:

Hyperreflective dots segmentation  
Diabetic retinopathy  
Deep learning  
Ocular biomechanics

## ABSTRACT

Due to chronic exposure to hyperglycemia, the ocular biomechanical homeostasis of diabetics is disturbed, which may cause multiple ocular diseases, such as diabetic retinopathy (DR). In retinal optical coherence tomography (OCT) images of patients with DR, hyperreflective dots (HRDs) are spot-shaped areas with high reflectivity and high local contrast. Quantitative analyses of HRDs are the key to make appropriate decisions on diagnosis and treatment. The small scale, similarity to background and interference of other foci bring great challenges to accurate and automatic segmentation of HRDs. To address above challenges, we propose a novel region-level context network (RC-Net), which introduces region-level position attention module (RPAM) and region-level channel attention module (RCAM) for HRDs segmentation. Four RPAMs are embedded behind ordinary convolution blocks to extract rich contextual dependencies, which help to cluster pixels in same class and reveal differences between HRDs and background. To further improve feature representation, a RCAM is embedded into the top of the encoder structure to adaptively utilize the semantic dependencies between different channels. We evaluate the proposed RC-Net on a dataset, which consists of 450 2D OCT B-scans. The mean and standard deviation of dice similarity coefficient ( $DSC$ ), intersection over union ( $IoU$ ), recall ( $R$ ) and precision ( $P$ ) are  $75.29 \pm 0.42\%$ ,  $62.27 \pm 0.54\%$ ,  $78.36 \pm 1.79\%$  and  $75.34 \pm 0.52\%$ , respectively. The experiments show that the proposed RC-Net achieves superior segmentation performance of HRDs segmentation to state-of-the-art methods.

## 1. Introduction

Diabetic retinopathy (DR) is one of the most common complications of diabetes mellitus (DM). Chronic exposure to hyperglycemia increases vascular permeability and disturbs the biomechanical homeostasis in retina, which may result in a series of ocular diseases [1]. DR generally has no obvious symptoms in its early stage, but the best treatment time has been missed [2] when patients have symptoms such as visual deterioration. Optical coherence tomography (OCT) is one of the important auxiliary tools for ophthalmologists due to its high resolution, non-invasive characteristics and high-speed imaging acquisition [3–5]. With the help of OCT, we can image the histological structure of retinal tissues, which is of great significance in diagnosing retinal diseases. Numerous algorithms are designed for the segmentation of retinal layers and lesions in OCT images [6–9]. Nawaz *et al.* [6] reviewed extensive

OCT image segmentation methods including traditional machine learning and deep learning methods.

In the OCT images of DR patients, some spot-shaped areas with high reflectivity similar to nerve fiber layer (NFL) without back shadowing can be observed, as shown in Fig. 1. In 2009, Coscas first named these spot-shaped areas as hyperreflective dots (HRDs) [10]. Different names are available in the literature, including “hyperreflective foci (HRF)”, “hyperreflective spots (HRS)” and “hyperreflective material”. What these names represent depends on the definition in the literature. The diameter of HRDs is between 20  $\mu\text{m}$  and 40  $\mu\text{m}$  [11]. HRDs are scattered throughout all retinal layers but mainly located in outer retina layers around fluid accumulation in the intraretinal cystoid spaces [12]. Currently, the origin of HRDs is not clear. Their size and behavior indicate that they may represent cells that can move through the retina [13,14], such as microglia, which may assume amoeboid characteris-

\* Corresponding authors.

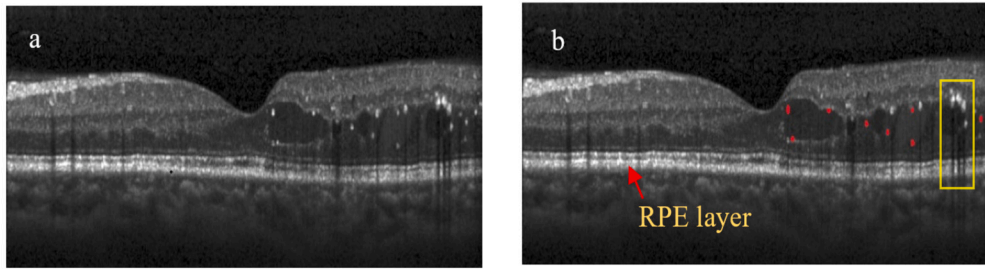
E-mail addresses: [1486364635@qq.com](mailto:1486364635@qq.com) (B. Zhang), [shujunfu@163.com](mailto:shujunfu@163.com) (S. Fu), [dr.wanghong@163.com](mailto:dr.wanghong@163.com) (H. Wang).

<https://doi.org/10.1016/j.optlaseng.2023.107872>

Received 22 February 2023; Received in revised form 27 September 2023; Accepted 28 September 2023

Available online 10 October 2023

0143-8166/© 2023 Elsevier Ltd. All rights reserved.



**Fig. 1.** Examples of HRDs in an OCT image. (a) original OCT image with HRDs, (b) corresponding ground truth of HRDs labeled with red marks. Examples of hard exudates (HEs) with back shadowing are marked by yellow box.

tics [15], and macropages, which are inherently mobile [16]. These cells exhibit high reflectivity in OCT images, resulting in spot-shaped areas with high reflectivity, namely HRDs, appearing. The location and number of HRDs can be used to predict the treatment outcome of DR [17,18]. The more HRDs in OCT images, the poorer treatment outcome of anti-vascular endothelial growth factor (anti-VEGF) [19–21]. If there are a large number of HRDs in retinal OCT images, ophthalmologists may not choose to use anti-VEGF therapy, which is quite expensive. Therefore, HRDs can be used as a biomarker. Previous clinical studies mark HRDs by hand [22,23], which is time-consuming and may cause subjective bias. Computer-aided detection methods can identify HRDs and analyze the evolution of HRDs in time. Due to high-speed imaging acquisition of OCT device, HRDs segmentation is not affected by dynamic intraocular pressure and blood flow. Therefore, the automatic and accurate segmentation of HRDs is of great significance to clinical diagnosis and treatment.

At present, there are not many researches on HRDs segmentation, and existing segmentation methods are divided into two categories: traditional methods and deep learning methods. Traditional methods design hand-crafted features based on various assumptions. Based on morphological component analysis (MCA) technique [24], Mokhtari *et al.* propose an automatic segmentation method of HRDs [25]. The OCT B-scan image is decomposed into texture part and cartoon part by using curvelet transform and Daubechies wavelet basis. In texture part reconstructed by Daubechies wavelet dictionary, HRDs can be easily segmented by an optimal threshold. Okuwobi *et al.* [26] propose an automatic HRF segmentation method based on fuzzy c-means clustering [27] and component tree [28]. According to the high reflectivity of the retinal nerve fiber layer (RNFL) and retinal pigment epithelium layer (RPE), the region of interest (ROI) is obtained by fuzzy c-means clustering. Meanwhile, HRF are segmented by extracting the extremal regions obtained from a component tree. Midena *et al.* [29] propose a semi-automatic method to count the number of HRDs by using the open-source available ImageJ software. Zhang *et al.* [30] propose an automatic segmentation of HRDs via focal priors and visual saliency. First, the ROI and the reflectivity estimation of the RPE layer is obtained by using graph search [31]. Then, a local contrast measure is used to calculate the multiscale local contrast saliency map. Finally, HRDs are segmented by an adaptive threshold. Varga *et al.* [32] compare six traditional machine learning methods and two deep learning methods. The results demonstrate that the full convolutional neural (FCN) network without pooling operation is superior to the traditional machine learning methods. Generally, traditional methods limited by multiple assumptions have poor generality.

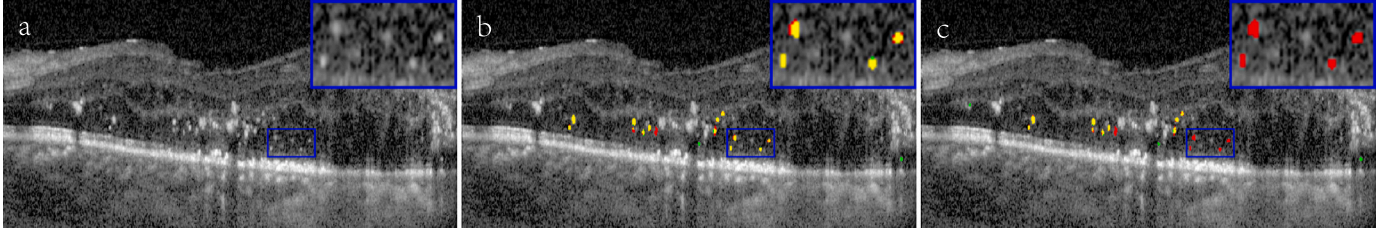
Recently, deep convolutional neural networks (CNNs) have achieved remarkable progress in medical image segmentation. Several CNNs-based methods have been proposed for HRDs segmentation [33–35]. Yao *et al.* [33] propose a self-adaptive network (SANet) for HRF segmentation. In this work, dual residual modules (DRM) are used instead of the ordinary convolution blocks to simplify the learning process and enhance the gradient propagation. Besides, position self-attention module [36,37] based on deformable convolution [38,39] is added at the

top of the encoder structure, which enables the network to integrate local features and global dependencies adaptively. Huang *et al.* [34] use multi-scale convolution modules (MSCM) to replace the ordinary convolution blocks and add channel attention module [40] in the encoder structure. Yao *et al.* [35] propose a joint segmentation method to segment hard exudates (HEs) and HRDs simultaneously.

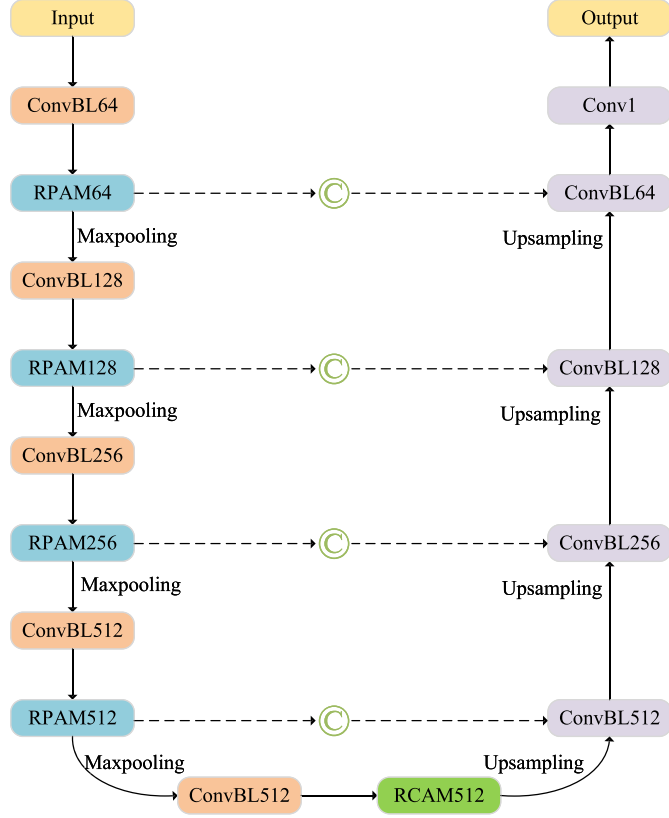
Although several researches are reported, the existing approaches do not achieve optimal performance. The major challenges of HRDs segmentation are summarized as below. First, the sizes of HRDs are extremely small in OCT images, even containing only several pixels. In other words, background pixels are usually dominant in OCT images. Second, some HRDs have similar reflectivity to the background. Third, other hyperreflective foci, such as HEs, may interfere with the HRDs segmentation. HEs are block-shaped areas with high reflectivity and high local contrast and the diameter of HEs is greater than 40  $\mu\text{m}$ . HRDs can be distinguished from HEs by its diameter and back shadowing, as shown in Fig. 1. Intuitively, the network can segment HRDs if it is endowed with a strong capability to capture features of small targets from complex background.

In recent years, U-Net [41] has become the most widely used backbone method for medical image segmentation because of its excellent performance. For HRDs segmentation task, U-Net tends to under-segment HRDs, as shown in Fig. 2. Because convolution operations lead to local receptive fields, convolution operations are applied repeatedly to obtain the large receptive fields, which causes the features with small size and low contrast to be washed out by consecutive multiplications [42]. Self-attention [36,37,43] can effectively capture the global dependencies by calculating the similarity between any two pixels of the input features. Appending self-attention to convolutional network can significantly improve the ability learning semantic features. However, pixel-level similarity in attention matrix is prone to be disturbed by outliers and ignore the region-level information. Intuitively, the local region can provide more useful information than a single pixel. In this paper, we change the way of calculating self-attention by calculating the similarity between any two regions instead of the similarity between any two pixels. We propose a novel network, called region-level context network (RC-Net) for HRDs segmentation in OCT images. To improve the ability to represent small targets, we introduce four region-level position attention modules (RPAMs) in the encoder structure. By introducing the RPAM, the network is able to analyze the input image in a global view, which is beneficial for distinguishing between HRDs and background. A region-level channel attention module (RCAM) is embedded into the top of the encoder structure, which can adaptively aggregate the features of each channel by weighted sum of all channels. The overall pipeline of the proposed RC-Net is illustrated in Fig. 3. Our main contributions can be summarized as follows:

- We propose a novel region-level context network (RC-Net) with region-level attention mechanism to extract more powerful feature maps for HRDs segmentation in OCT images.
- A RPAM is proposed to learn region-level spatial long-range context dependencies.



**Fig. 2.** A sample of HRDs segmentation predictions. (a) original OCT image, (b) Our method (c) U-Net. Red regions represent false negative (FN), green regions represent false positive (FP) and yellow regions represent true positive (TP).



**Fig. 3.** Architecture of RC-Net.

- A RCAM is designed to aggregate high-level semantic features in channel dimensions.
- Compared with other CNN-based methods, the proposed RC-Net achieves state-of-the-art results.

## 2. Methods

In this section, we first introduce a general framework of our network and then present the two region-level attention modules which model long-range dependencies in spatial and channel dimension respectively. Finally we describe the loss function in detail.

### 2.1. Overview of the region-level context network

The overall structure of the proposed RC-Net is illustrated in Fig. 3, which is an encoder-decoder architecture and consists of three main parts: feature encoder with region-level position attention module (RPAM), region-level channel attention module (RCAM) and feature decoder. Due to its excellent performance for medical image segmentation, U-Net is the used backbone architecture in this study. In Fig. 3, the ConvBL, RPAM and RCAM represent the convolutional block, region-

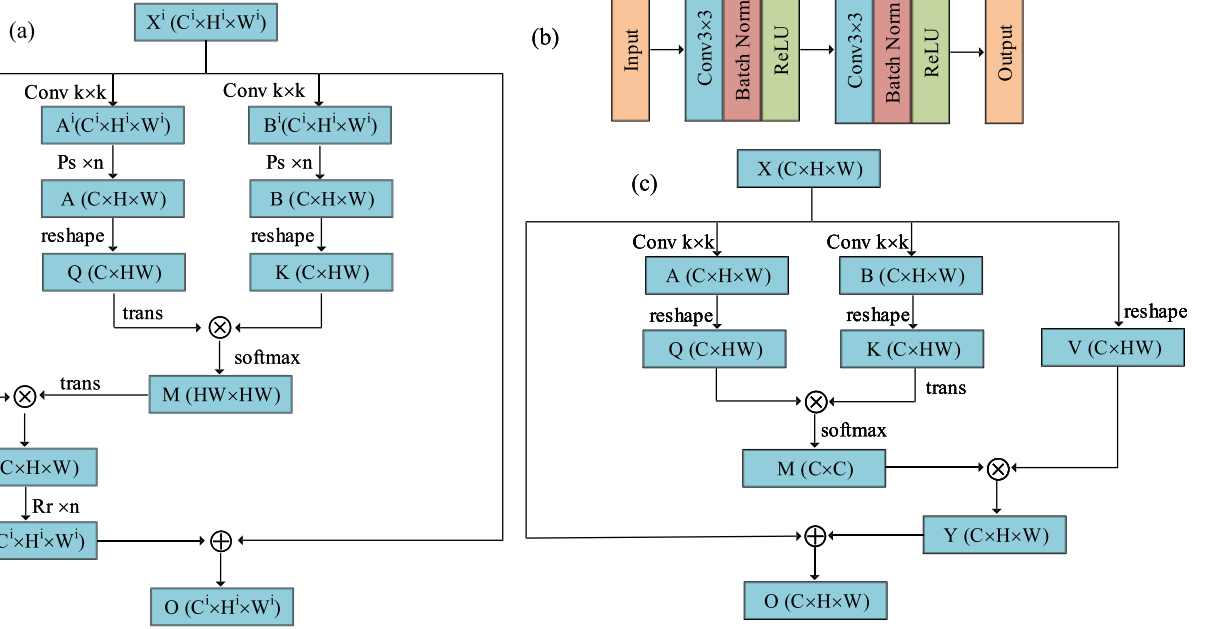
level position attention module and region-level channel attention module, respectively, whose detailed structures are shown in Fig. 4. The notation such as ConvBL64 indicates the block has an output with 64 channels. The  $\oplus$  in the architecture represents the concatenate operation. The Conv1 denotes a  $1 \times 1$  convolution layer which has an output with 1 channel.

In the encoder structure, a RPAM is embedded behind each ordinary convolution block to capture the long-range dependencies, which greatly improves the learning ability of the network. The RCAM is inserted at the top of the encoder to fuse the global features in different channels.

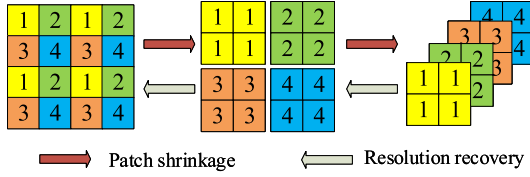
### 2.2. Region-level position attention module

Recently, self-attention [36,37] has been proved to improve feature representation, as it can effectively capture the global dependencies of the input features. The first step of self-attention is to generate an attention matrix which calculates the similarity between any two pixels of the input feature map. Then, the enhanced feature map can be obtained by performing a matrix multiplication between the attention matrix and the input feature map. Finally, an element-wise sum operation is performed on the input feature map and the enhanced feature map to obtain the final feature map. However, pixel-level similarity in attention matrix is prone to be disturbed by outliers and ignore the region-level information. Inspired by this, we proposed an improved self-attention module, called region-level position attention module (RPAM). As shown in Fig. 4(a), given a input feature  $X^i \in \mathbb{R}^{C^i \times H^i \times W^i}$ , we feed it into a  $k \times k$  depth-wise separable convolution layer to generate two new feature maps  $A^i$  and  $B^i$ , respectively, where  $A^i, B^i \in \mathbb{R}^{C^i \times H^i \times W^i}$ . Each pixel in  $A^i$  and  $B^i$  is fused with the information of the surrounding  $k \times k$  region. The attention matrix calculated by  $A^i$  and  $B^i$  contains the similarity between any two  $k \times k$  regions. Due to the computational complexity of self-attention is quadratic to the resolution of  $A^i$  and  $B^i$ , using  $A^i$  and  $B^i$  directly will consume a lot of memory. As shown in Fig. 5, patch shrinkage stacks adjacent features into different channels instead of spatial locations, similar to the patch merging in Swin transformer [44]. The patch shrinkage operation concatenates the features of each group of  $2 \times 2$  neighboring patches. After a patch shrinkage operation, the number of rows and columns of the feature map is halved, the number of channels is quadrupled and the total number of pixels in the feature map remains unchanged. To reduce computation cost and information loss, the resolution of  $A^i$  and  $B^i$  is reduced by patch shrinkage. Due to the limitation of GPU memory, in this paper, we adjust the resolution of  $A^i$  and  $B^i$  to be the same as the resolution of the input feature map of the ConvBL512 block. The RPAM64, RPAM128, RPAM256 and RPAM512 require patch shrinkage operations to be performed 4, 3, 2 and 1 times, respectively.

Here we just take RPAM512 as an example and RPAM64, RPAM128 and RPAM256 can be processed in a similar way. For RPAM512, as illustrated in Fig. 4(a),  $X^i \in \mathbb{R}^{C^i \times H^i \times W^i}$  denotes the input feature map. First, in order to obtain the similarity between any two regions in the input feature map  $X^i$ , we feed  $X^i$  into a  $k \times k$  depthwise separable convolution layer to generate two new feature maps  $A^i$  and  $B^i$ , in which



**Fig. 4.** Architecture of (a) RPAM, (b) ConvBL, (c) RCAM. Conv  $k \times k$  represents a  $k \times k$  depthwise separable convolution. Ps  $\times n$  means to perform patch shrinkage operation  $n$  times. Rr  $\times n$  means to perform resolution recovery operation  $n$  times.



**Fig. 5.** Illustration of patch shrinkage and resolution recovery.

each pixel incorporates  $k \times k$  local region information. Then, we perform patch shrinkage operation once on  $X^i$ ,  $A^i$  and  $B^i$  to reduce the computation cost. After that, we reshape  $A$  and  $B$  to  $Q \in \mathbb{R}^{C \times N}$  and  $K \in \mathbb{R}^{C \times N}$ , respectively, where  $N = H \times W$ . The matrix multiplication between the transpose of  $Q$  and  $K$  and softmax are performed to calculate the region-level similarity weight  $M \in \mathbb{R}^{N \times N}$ .

$$M = \delta(Q^T K) \quad (1)$$

where,  $\delta$  denotes the softmax activation function. Meanwhile, we reshape  $V^i$  to  $V \in \mathbb{R}^{C \times N}$  and perform a matrix multiplication between  $V$  and the transpose of  $M$  to obtain the enhanced feature map  $Y$ . Then we perform resolution recovery operation once on  $Y$ . The resolution recovery operation is the inverse of patch shrinkage operations as shown in Fig. 5. Finally, Then, we perform an element-wise sum between  $Y^o$  and the input feature map  $X^i$ .

### 2.3. Region-level channel attention module

The structure of the region-level channel attention module can be found in Fig. 4(c). Different from the channel attention module in [37], we feed feature map  $X$  into a  $k \times k$  depthwise separable convolution layer to generate two feature map  $A \in \mathbb{R}^{C \times H \times W}$  and  $B \in \mathbb{R}^{C \times H \times W}$ , in which each pixel incorporates the information of the surrounding  $k \times k$  neighborhood. Then, we reshape  $A$  and  $B$  to  $Q$  and  $K$  and the channel attention map capturing similarities between any two local regions in different channels is obtained by performing a matrix multiplication between  $Q$  and the transpose of  $K$ . The remaining operations are similar to region-level position attention module except the resolution recovery.

### 2.4. Loss function

Since the sizes of HRDs are extremely small in OCT images, even containing only several pixels, we use a joint loss  $L_{total}$  consisting of binary cross-entropy (BCE) loss  $L_{BCE}$  and Dice loss  $L_{Dice}$  to optimize our model. The joint loss is as follows:

$$L_{total} = L_{BCE} + \lambda L_{Dice} \quad (2)$$

$$L_{BCE} = - \sum_{i=1}^N [(1 - y_i) \times \log(1 - \hat{y}_i) + y_i \times \log(\hat{y}_i)] \quad (3)$$

$$L_{Dice} = 1 - \frac{2(y \times \hat{y})}{\text{sum}(y) + \text{sum}(\hat{y})} \quad (4)$$

where  $y$  and  $\hat{y}$  represent the ground truth and the results predicted by networks, respectively.  $N$  is the number of pixels in the input image. The parameter  $\lambda$  is a factor to balance binary cross-entropy loss and Dice loss. In our experiments,  $\lambda$  is set to 1.

## 3. Experiments and results

In this section, we describe the datasets, the evaluation metrics and the implementation details used for HRDs segmentation in OCT images. To show the effectiveness of our proposed RC-Net, we compare RC-Net with four state-of-the-art networks on our datasets. We also perform comprehensive ablation studies to demonstrate each component of RC-Net.

### 3.1. Data

In this study, 450 clinical OCT B-scans from 50 patients are included. These patients are diagnosed with different levels of diabetic retinopathy (DR) severity by a Heidelberg Eye Explore, Heidelberg Engineering GmbH (Heidelberg, Germany, version 1.10.4.0) at the Department of Ophthalmology, Qilu Hospital of Shandong University. The device scanning area is  $8.8 \times 8.8 \times 2 \text{ mm}^3$  centered on the fovea. We crop each OCT B-scan to  $256 \times 512$  pixels. This study is approved by the Medical Ethics Committee of Qilu Hospital, China, and adheres to the tenets of the Declaration of Helsinki.



**Table 1**Results of five-fold cross validation (mean  $\pm$  standard deviation).

metric	Fold 1	Fold 2	Fold 3	Fold 4	Fold 5
DSC (%)	75.54 $\pm$ 0.45	74.40 $\pm$ 0.79	75.84 $\pm$ 0.64	75.11 $\pm$ 0.82	74.87 $\pm$ 0.64

**Table 2**Segmentation performance comparison with state-of-the-art method. (mean  $\pm$  standard deviation).

Metrics	KiU-Net [45]	LSPM [46]	SANet [33]	MSCM [34]	Ours
parameter	<b>1.2M</b>	156.8M	195.7M	44.4M	56.2M
memory	24033M	11241M	23187M	12529M	<b>9877M</b>
DSC (%)	72.77 $\pm$ 2.04	72.49 $\pm$ 1.07	71.06 $\pm$ 0.98	73.27 $\pm$ 1.43	<b>75.29<math>\pm</math>0.42</b>
IoU (%)	59.04 $\pm$ 1.73	59.02 $\pm$ 1.16	57.23 $\pm$ 1.50	59.61 $\pm$ 1.98	<b>62.27<math>\pm</math>0.54</b>
P (%)	76.24 $\pm$ 4.86	73.68 $\pm$ 0.58	73.65 $\pm$ 3.40	72.93 $\pm$ 5.83	<b>78.36<math>\pm</math>1.79</b>
R (%)	73.12 $\pm$ 6.74	75.17 $\pm$ 1.57	72.50 $\pm$ 2.30	<b>77.44<math>\pm</math>3.91</b>	75.34 $\pm$ 0.52
$P_d$ (%)	75.92 $\pm$ 6.67	78.67 $\pm$ 2.43	74.86 $\pm$ 4.18	<b>79.17<math>\pm</math>2.32</b>	78.18 $\pm$ 1.28
$P_{fa}$ (%)	21.24 $\pm$ 3.01	24.45 $\pm$ 2.67	25.33 $\pm$ 6.42	23.72 $\pm$ 6.25	<b>20.23<math>\pm</math>2.87</b>

Manual labeling of OCT B-scans is performed by a senior ophthalmologist with more than 3 years of clinical experience using Microsoft Paint software and two other senior ophthalmologists carefully review the annotation results. Meanwhile, a five-fold cross validation strategy is performed to objectively validate the performance of our method. The results of five-fold cross validation are shown in the Table 1. For each fold, 150 epochs are trained. When compared with other state-of-the-art method, the dataset is randomly divided into 360 B-scans for training, 45 B-scans for validating and 45 B-scans for testing.

### 3.2. Evaluation metrics

In order to evaluate the performance of the proposed RC-Net, four metrics including dice similarity coefficient (DSC), intersection over union (IoU), recall (R) and precision (P) are adopted, which are defined as follows:

$$DSC = \frac{2TP}{2TP + FP + FN} \quad (5)$$

$$IoU = \frac{TP}{TP + FP + FN} \quad (6)$$

$$P = \frac{TP}{TP + FP} \quad (7)$$

$$R = \frac{TP}{TP + FN} \quad (8)$$

where  $TP$ ,  $FP$  and  $FN$  represent true positive, false positive and false negative, respectively.

To further demonstrate the effectiveness of the proposed RC-Net, we also adopt the detection accuracy  $P_d$  and the probability of false alarm  $P_{fa}$ :

$$P_d = \frac{nc}{nt} \quad (9)$$

$$P_{fa} = \frac{ni}{nd} \quad (10)$$

where  $nc$  denotes the total number of correctly detected HRDs in the test set,  $ni$  denotes the total number of incorrectly detected HRDs in the test set,  $nt$  is the total number of true HRDs in the test set,  $nd$  is the total number of detected HRDs in the test set. In this paper, a detected HRD is regarded as the correctly detected HRD when the overlap between the true HRD and the detected HRD exceeds 50%.

### 3.3. Implementation details

The implementation of the proposed RC-Net is based on the PyTorch and NVIDIA GeForce 3090 GPU with 24 GB memory. We use the poly learning rate policy, where  $lr = baselr \times (1 - \frac{iter}{total\_iter})^{0.9}$ , the initial learning rate  $baselr$  is set to 0.0001. Besides, the optimizer is Adam and the

batch size is set to 8. The network is trained for a total of 150 epochs. Finally, the training weights with the lowest validation loss are employed to evaluate the segmentation performance on the test dataset.

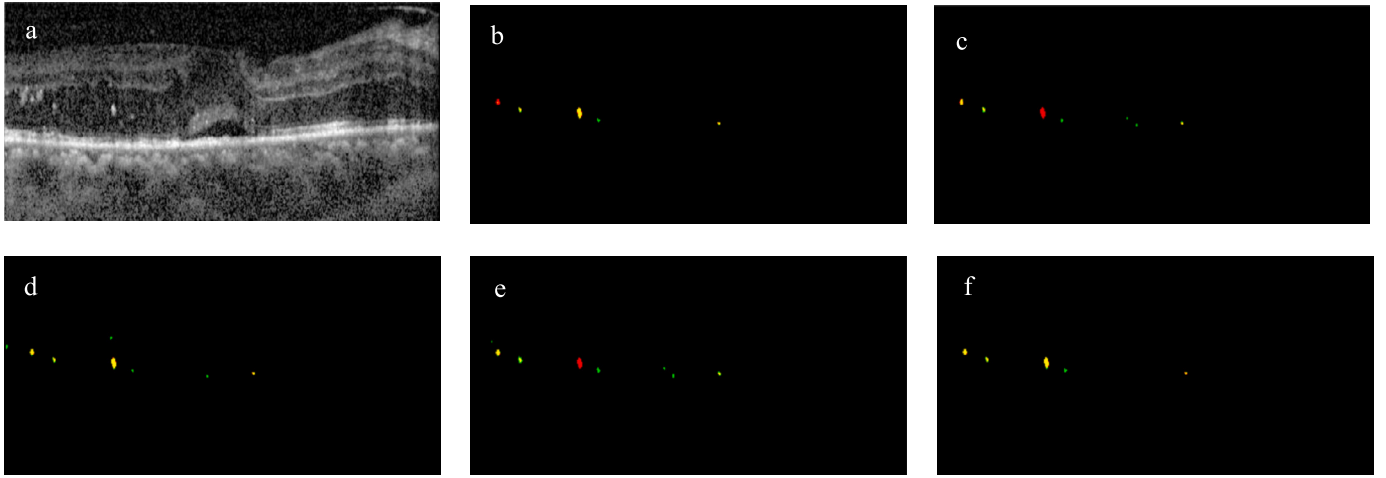
### 3.4. Comparison with other state-of-the-art deep learning based networks

We compare our proposed RC-Net with four state-of-the-art deep learning-based methods: KiU-Net [45], LSPM [46], SANet [33], MSCM [34]. KiU-Net is proposed for dealing with small anatomical structures with blurred noisy boundaries. LSPM is a method to segment small infrared targets which are similar with HRDs. SANet and MSCM are proposed for HRDs segmentation. The four compared methods are implemented with the same hyperparameters that are described in their original papers. We implement these experiments three times for different random folds of training, validating and testing data and report the mean metrics with the variance. Table 2 shows the segmentation results of different methods. KiU-Net is an over-complete architecture which includes an upsampling layer after every convolution layer. Therefore, KiU-Net requires a lot of memory. KiU-Net only have 3 layers of convolution blocks in the encoder and each convolution block consists of a 2D convolution layer. So, the number of parameters in KiU-Net is significantly fewer, only 1.2M. In our RC-Net, depthwise separable convolution is used in RPAMs and RCAM, which does not increase many parameters and computational complexity. Patch shrinkage operation is also used to reduce memory usage. When the batch size is set to 8, the memory usage of RC-Net is 9877M. We can observe that the proposed RC-Net obtains the best segmentation results in terms of  $DSC$ ,  $IoU$ ,  $P$  and  $P_{fa}$  and achieves comparable performance on  $R$  and  $P_d$ . Compared with MSCM, which achieves the second best performance with 73.27% for DSC, our method achieves a remarkable improvement and reach 75.29%. Moreover, our method obtains the largest  $P$  value and the lowest  $P_{fa}$  value, which demonstrates that the introduced RPAM and RCAM are helpful to distinguish between HRDs and background. Fig. 6 and Fig. 7 illustrate the visual comparison of HRDs segmentation using different methods. Obviously, our RC-Net achieves the best segmentation performance. We can observe that our RC-Net is able to capture even very small HRDs precisely. Besides, the HRDs which has similar reflectivity to background can also be segmented accurately.

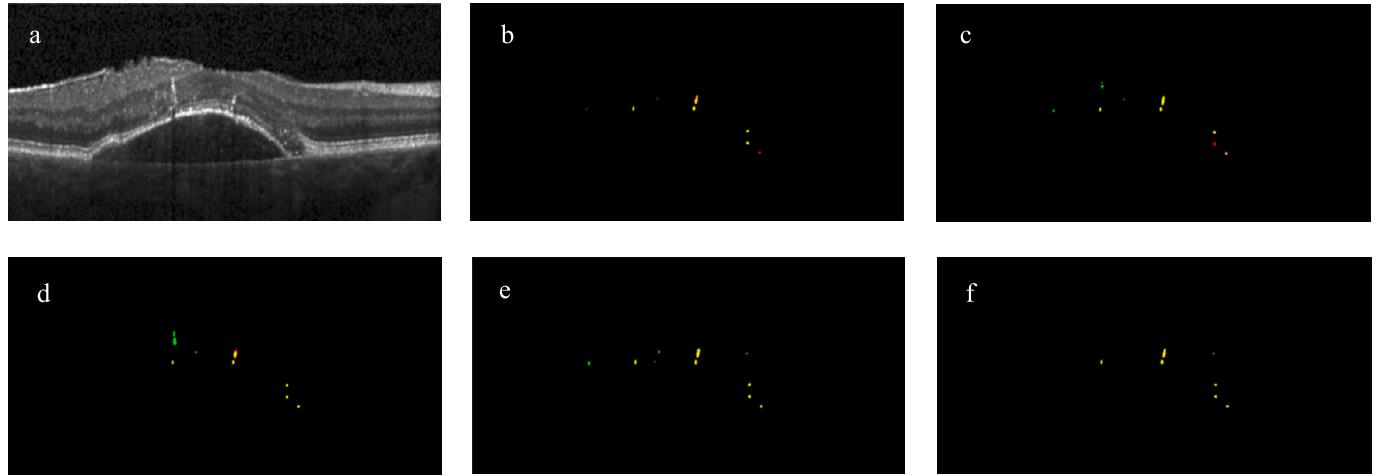
### 3.5. Ablation study

#### 3.5.1. Effect of the kernel size of the depthwise separable convolution

To investigate the effect of the kernel size  $k$  of the depthwise separable convolution on HRDs segmentation, we conduct extensive experiments with  $k$  ranging from 1 to 7 in a step of 2, and the segmentation results evaluated with  $DSC$ ,  $IoU$ ,  $P$  and  $R$  are shown in Table 3. We can observe that with the increase of  $k$ , the value of  $DSC$  increases first and then decreases, and the largest value for  $DSC$  is achieved when



**Fig. 6.** Visual comparison of different segmentation methods. (a) original OCT B-scan, (b) KiU-Net, (c) LSPM, (d) SANet, (e) MSCM, (f) our RC-Net. Red regions represent false negative (FN), green regions represent false positive (FP), and yellow regions represent true positive (TP).



**Fig. 7.** Visual comparison of different segmentation methods. (a) original OCT B-scan, (b) KiU-Net, (c) LSPM, (d) SANet, (e) MSCM, (f) our RC-Net. Red regions represent false negative (FN), green regions represent false positive (FP), and yellow regions represent true positive (TP).

**Table 3**

Segmentation performance comparison with different kernel size  $k$ . (mean  $\pm$  standard deviation).

Kernel size	DSC (%)	IoU (%)	P (%)	R (%)
1 $\times$ 1	73.43 $\pm$ 1.81	60.15 $\pm$ 2.31	73.49 $\pm$ 1.18	77.04 $\pm$ 0.37
3 $\times$ 3	<b>75.29<math>\pm</math>0.42</b>	<b>62.27<math>\pm</math>0.54</b>	<b>78.36<math>\pm</math>1.79</b>	75.34 $\pm$ 0.52
5 $\times$ 5	74.71 $\pm$ 0.64	61.55 $\pm$ 0.60	75.36 $\pm$ 0.62	77.19 $\pm$ 1.45
7 $\times$ 7	74.48 $\pm$ 1.08	60.92 $\pm$ 1.39	71.94 $\pm$ 2.28	<b>80.25<math>\pm</math>0.06</b>

$k = 3$ . Due to the small size of HRDs, large kernel size, that is, large local region will incorporate more background feature which causes the segmentation performance to reduce. Therefore, it is appropriate to set the kernel size  $k$  of the depthwise separable convolution to 3.

### 3.5.2. Effect of the region-level attention

In our network, the region-level position attention modules (RPAMs) are introduced in the encoder structure to extract global context and the region-level channel attention module (RCAM) is embedded into the top of the encoder structure to adaptively aggregate the features of different channels. To explore the effect of RPAMs and RCAM, we conduct ablation experiments with the following configurations:

- BaseNet: The U-Net.
- BaseNet+RPAMs: The U-Net combined with four RPAMs in the encoder structure.

- BaseNet+RCAM: The U-Net combined with a RCAM embedded into the top of the encoder.
- BaseNet+RPAMs+RCAM: The proposed RC-Net.

HRDs segmentation results with different configurations are reported in Table 4. We can observe that the performance gradually improves with addition of each module to the basenet. Compared with BaseNet, BaseNet+RPAMs and BaseNet+RCAM achieve much better segmentation results and the values of  $DSC$  obtained using BaseNet+RPAMs and BaseNet+RCAM increase by 0.68 and 0.32, respectively. Although the values of  $P$  of BaseNet+RPAMs and BaseNet+RCAM decrease slightly, the values of  $R$  of BaseNet+RPAMs and BaseNet+RCAM increase significantly, indicating that the introduced attention modules are helpful to segmentation small targets. As expected, our proposed RC-Net achieves the best segmentation performance, which demonstrates the effectiveness of RPAM and RCAM.

**Table 4**Segmentation performance comparison with different configurations. (mean  $\pm$  standard deviation).

Metrics	BaseNet	BaseNet+RPAMs	BaseNet+RCAM	RC-Net
<i>DSC</i> (%)	73.61 $\pm$ 0.33	74.29 $\pm$ 0.37	73.93 $\pm$ 0.62	<b>75.29<math>\pm</math>0.42</b>
<i>IoU</i> (%)	60.57 $\pm$ 0.39	60.99 $\pm$ 0.03	60.87 $\pm$ 0.78	<b>62.27<math>\pm</math>0.54</b>
<i>P</i> (%)	76.95 $\pm$ 1.46	74.75 $\pm$ 1.57	72.96 $\pm$ 1.48	<b>78.36<math>\pm</math>1.79</b>
<i>R</i> (%)	73.61 $\pm$ 0.23	<b>78.10<math>\pm</math>1.48</b>	77.74 $\pm$ 1.13	75.34 $\pm$ 0.52

**Table 5**Segmentation performance comparison with different configurations. (mean  $\pm$  standard deviation).

Metrics	<i>n</i> = 1	<i>n</i> = 2	<i>n</i> = 3	<i>n</i> = 4
<i>DSC</i> (%)	73.67 $\pm$ 0.73	73.85 $\pm$ 0.87	74.51 $\pm$ 0.66	<b>75.29<math>\pm</math>0.42</b>

**Table 6**Segmentation performance comparison with different configurations. (mean  $\pm$  standard deviation).

Metrics	mRC-Net	RC-Net	IRC-Net
<i>DSC</i> (%)	73.23 $\pm$ 0.41	<b>75.29<math>\pm</math>0.42</b>	73.73 $\pm$ 0.59

### 3.5.3. Effect of the number of RPAMs

To investigate the effect of the number *n* of introduced RPAMs, we conduct ablation experiments with the following configurations:

- BaseNet+RPAM512
- BaseNet+RPAM256+RPAM512
- BaseNet+RPAM128+RPAM256+RPAM512
- RC-Net

The BaseNet represents U-Net+RCAM. We conduct extensive experiments with *n* ranging from 1 to 4 and the segmentation results of different number of introduced RPAMs are shown in Table 5. We can observe that the value of *DSC* gradually increases with the increase of *n* and the largest value for *DSC* is achieved when *n* = 4. From Table 5, it can be seen that the RC-Net can learn more feature information that is helpful for HRDs segmentation.

### 3.6. Selection of model size

To investigate the effect of the network depth, we compare the segmentation results of mRC-Net, RC-Net and IRC-Net. These network depths are 3, 4, and 5, respectively. The network depth refers to the number of maxpooling in the encoder phase. The encoder of mRC-Net include ConvBL64, RPAM64, ConvBL128, RPAM128, ConvBL256, RPAM256, ConvBL512, RCAM256. The encoder of IRC-Net include ConvBL 64, RPAM64, ConvBL128, RPAM128, ConvBL256, RPAM256, ConvBL512, RPAM512, ConvBL1024, RPAM1024, ConvBL1024, RCAM1024. From Table 6, we can see that RC-Net obtains the best performance. The depth of mRC-net is shallow, so that mRC-net cannot learn rich semantic features. For IRC-Net, due to too many maxpooling times, small target information in deep semantic features is lost. In this paper, the network depth is set to 4. The network structure is shown in Fig. 3.

## 4. Conclusions and discussions

Biomechanical homeostasis of eyes is associated with ocular health. Due to chronic exposure to hyperglycemia, the retinal biomechanical homeostasis of patients with diabetic retinopathy (DR) is disturbed, which causes visual impairment. Hyperreflective dots (HRDs) segmentation is a prerequisite step for precise diagnosis and treatment of DR. In this study, we propose a novel region-level context network (RC-Net) for HRDs segmentation to facilitate automatic HRDs segmentation

closer to clinical applications. The embedded RPAMs are useful in modeling dependencies between distant regions. By introducing the RPAMs, our RC-Net can capture the dependencies between HRDs and non-HRDs regions and improve HRDs segmentation performance. Besides, we further improve HRDs segmentation performance by introducing a region-level channel attention module (RCAM) to fuse high-level features in different channels. Extensive experiments demonstrate that the proposed RC-Net achieves superior and competitive segmentation performance to state-of-the-art methods.

At present, the *DSC* and *P<sub>d</sub>* of our model only achieve 75.29% and 78.18%, respectively. There is still a certain gap between our model and clinical application. In the future, we plan to narrow the gap from the following aspects:

- (1) Expand our dataset. If we collect more quality data, the performance of our model will be further improved.
- (2) Improve our model. We can use more advanced models to improve segmentation performance.
- (3) Segment multiple biomarkers. Ophthalmologists need to observe multiple biomarkers for diagnosis and treatment. We can design a model to simultaneously segment multiple biomarkers, such as HRDs, hard exudates, and fluid accumulation.

After precise segmentation, we can obtain some quantitative indicators of HRDs, such as quantity and location. With the help of quantitative indicators, ophthalmologists can diagnose more accurately and provide personalized treatment plans for patients, which accelerates the restoration of biomechanical homeostasis in ocular disorders.

### CRedit authorship contribution statement

**Bo Zhang:** Conceptualization, Formal analysis, Methodology, Software, Validation, Writing – original draft. **Hui Zhao:** Data curation, Investigation, Resources. **Mingwei Si:** Data curation, Investigation, Resources. **Wenxuan Cui:** Data curation, Resources. **Yuanfeng Zhou:** Resources. **Shujun Fu:** Writing – review & editing. **Hong Wang:** Writing – review & editing.

### Declaration of competing interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

### Data availability

Data underlying the results presented in this study are not publicly available at this time but may be obtained from the authors upon reasonable request.

The code of the proposed RC-Net is available at <https://github.com/a1486364635/RCNet>.

### Acknowledgements

We would like to thank Qilu hospital for providing the SD-OCT data.

### Funding

National Natural Science Foundation of China (12071263, 11971269, 61671276); Natural Science Foundation of Shandong Province (ZR2019MF045).

## References

- [1] Cheng Y, Ren T, Wang N. Biomechanical homeostasis in ocular diseases: a mini-review. *Front Public Health* 2023;11:1106728.
- [2] Wojtkowski M, Bajraszewski T, Kowalczyk PTA. Real-time in vivo imaging by high-speed spectral optical coherence tomography. *Opt Lett* 2003;28:1745–7.
- [3] Huang D, Swanson EA, Lin CP, Schuman JS, Stinson WG, Chang W, et al. Optical coherence tomography. *Science* 1991;254:1178–81.
- [4] Coscas G, Benedetto UD, Coscas F, Calzi C, Vismara S, Roudot-Thoraval F, et al. Hyperreflective dots: a new spectral-domain optical coherence tomography entity for follow-up and prognosis in exudative age-related macular degeneration. *Ophthalmologica* 2013;229:32–7.
- [5] Wang J, Li W, Chen Y, Fang W, Kong W, He Y, et al. Weakly supervised anomaly segmentation in retinal OCT images using an adversarial learning approach. *Biomed Opt Express* 2021;12:4713–29.
- [6] Nawaz M, Uvaliyev A, Bibi K, Wei H, Abaxi SMD, Masood A, et al. Unravelling the complexity of optical coherence tomography image segmentation using machine and deep learning techniques: a review. *Comput Med Imaging Graph* 2023;108:102269.
- [7] Zhang Y, Li Z, Nan N, Wang X. Transegnet: hybrid cnn-vision transformers encoder for retina segmentation of optical coherence tomography. *Life* 2023;13:976.
- [8] Paringer R, Mukhin A, Ilyasova N, Demin N. Neural network application for semantic segmentation of fundus. *Comput Optics* 2022;46:596–602.
- [9] Mittal P, Bhatnagar C. Effectual accuracy of OCT image retinal segmentation with the aid of speckle noise reduction and boundary edge detection strategy. *J Microsc* 2023;289:164–79.
- [10] Coscas G, Coscas F, Vismara S, Zourdani A, Calzi L. Clinical features and natural history of AMD: 1. age-related maculopathy. In: *Optical coherence tomography in age-related macular degeneration*. Springer; 2009. p. 171–94.
- [11] Lee H, Ji B, Chung H, Kim HC. Correlation between optical coherence tomographic hyperreflective foci and visual outcomes after anti-VEGF treatment in neovascular age-related macular degeneration and polypoidal choroidal vasculopathy. *Retina* 2016;36:465–75.
- [12] Hwang HS, Chae JB, Kim JY, Kim DY. Association between hyperreflective dots on spectral-domain optical coherence tomography in macular edema and response to treatment. *Investig Ophthalmol Vis Sci* 2017;58:5958–67.
- [13] Christenbury JG, Folgar FA, O'Connell RV, Chiu SJ, Farsiu S, Toth CA. Progression of intermediate age-related macular degeneration with proliferation and inner retinal migration of hyperreflective foci. *Ophthalmology* 2013;120:1038–45.
- [14] Curcio CA, Zanzottera EC, Ach T, Balaratnasingam C, Freund KB. Activated retinal pigment epithelium, an optical coherence tomography biomarker for progression in age-related macular degeneration. *Investig Ophthalmol Vis Sci* 2017;58: BIO211–BIO226.
- [15] Zhang Y, Zhao L, Wang X, Ma W, Lazere A, Qian H, et al. Repopulating retinal microglia restore endogenous organization and function under CX3CL1-CX3CR1 regulation. *Sci Adv* 2018;4:eaap8492.
- [16] Omri S, Behar-Cohen F, De Kozak Y, Sennlaub F, Verissimo LM, Jonet L, et al. Microglia/macrophages migrate through retinal epithelium barrier by a transcellular route in diabetic retinopathy: role of PKC $\zeta$  in the Goto Kakizaki rat model. *Am J Pathol* 2011;179:942–53.
- [17] Turgut B, Yildirim H. The causes of hyperreflective dots in optical coherence tomography excluding diabetic macular edema and retinal venous occlusion. *Open Ophthalmol J* 2015;9:36–40.
- [18] Hanumunthadu D, Matet A, Rasheed MA, Goud A, Vuppurabina KK, Chhablani J. Evaluation of choroidal hyperreflective dots in acute and chronic central serous chorioretinopathy. *Indian J Ophthalmol* 2019;67:1850–4.
- [19] Murakami T, Suzuma K, Uji A, Yoshitake S, Dodo Y, Fujimoto M, et al. Association between characteristics of foveal cystoid spaces and short-term responsiveness to ranibizumab for diabetic macular edema. *Jpn J Ophthalmol* 2018;62:292–301.
- [20] Kang JW, Chung H, Chan Kim H. Correlation of optical coherence tomographic hyperreflective foci with visual outcomes in different patterns of diabetic macular edema. *Retina* 2016;36:1630–9.
- [21] Ashraf M, Souka A, Adelman R. Predicting outcomes to anti-vascular endothelial growth factor VEGF therapy in diabetic macular oedema: a review of the literature. *Br J Ophthalmol* 2016;100:1596–604.
- [22] Fonollosa A, Zarranz-Ventura J, Valverde A, Becerra E, Bernal-Morales C, Pastor-Idoate S, et al. Predictive capacity of baseline hyperreflective dots on the intravitreal dexamethasone implant (Ozurdex®) outcomes in diabetic macular edema: a multicenter study. *Graef Arch Clin Exp Ophthalmol* 2019;257:2381–90.
- [23] Lee H, Jang H, Choi YA, Kim HC, Chung H. Association between soluble CD14 in the aqueous humor and hyperreflective foci on optical coherence tomography in patients with diabetic macular edema. *Investig Ophthalmol Vis Sci* 2018;59:715–21.
- [24] Fadili JM, Starck JL, Elad M, Donoho D. MCALab: reproducible research in signal and image decomposition and inpainting. *Comput Sci Eng* 2010;12:44–63.
- [25] Mokhtari M, Kamasi ZG, Rabbani H. Automatic detection of hyperreflective foci in optical coherence tomography B-scans using morphological component analysis. In: 39th annual international conference of the IEEE engineering in medicine and biology society (EMBC); 2017. p. 1497–500.
- [26] Okuwobi IP, Ji Z, Fan W, Yuan S, Bekalo L, Chen Q. Automated quantification of hyperreflective foci in SD-OCT with diabetic retinopathy. *IEEE J Biomed Health Inform* 2019;24:1125–36.
- [27] Lei T, Jia X, Zhang Y, He L, Meng H, Nandi AK. Significantly fast and robust fuzzy c-means clustering algorithm based on morphological reconstruction and membership filtering. *IEEE Trans Fuzzy Syst* 2018;26:3027–41.
- [28] Carlinet E, Géraud T. A comparative review of component tree computation algorithms. *IEEE Trans Image Process* 2014;23:3885–95.
- [29] Midena E, Torresin T, Velotta E, Pilotto E, Parozzani R, Frizziero L. OCT hyperreflective retinal foci in diabetic retinopathy: a semi-automatic detection comparative study. *Front Immunol* 2021;12:1–8.
- [30] Zhang B, Ma L, Zhao H, Hao Y, Fu S, Wang H, et al. Automatic segmentation of hyperreflective dots via focal priors and visual saliency. *Med Phys* 2022;49:7025–37.
- [31] Chiu SJ, Li XT, Nicholas P, Toth CA, Izatt JA, Farsiu S. Automatic segmentation of seven retinal layers in SDOCT images congruent with expert manual segmentation. *Opt Express* 2010;18:19413–28.
- [32] Varga L, Kovács A, Grósz T, Thury G, Hadarits F, Dégi R, et al. Automatic segmentation of hyperreflective foci in OCT images. *Comput Methods Programs Biomed* 2019;178:91–103.
- [33] Yao C, Zhu W, Wang M, Zhu L, Huang H, Chen H, et al. SANet: a self-adaptive network for hyperreflective foci segmentation in retinal OCT images. In: *Medical imaging 2021: image processing*; 2021. p. 809–15.
- [34] Huang H, Zhu L, Zhu W, Lin T, Los LJ, Yao C, et al. Algorithm for detection and quantification of hyperreflective dots on optical coherence tomography in diabetic macular edema. *Front Med* 2021;8:1–11.
- [35] Yao C, Wang M, Zhu W, Huang H, Shi F, Chen Z, et al. Joint segmentation of multi-class hyper-reflective foci in retinal optical coherence tomography images. *IEEE Trans Biomed Eng* 2021;69:1349–58.
- [36] Wang X, Girshick R, Gupta A, He K. Non-local neural networks. In: *Proceedings of the IEEE conference on computer vision and pattern recognition*; 2018. p. 7794–803.
- [37] Fu J, Liu J, Tian H, Li Y, Bao Y, Fang Z, et al. Dual attention network for scene segmentation. In: *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*; 2019. p. 3146–54.
- [38] Dai J, Qi H, Xiong Y, Li Y, Zhang G, Hu H, et al. Deformable convolutional networks. In: *Proceedings of the IEEE international conference on computer vision*; 2017. p. 764–73.
- [39] Zhu X, Hu H, Lin S, Dai J. Deformable convnets v2: more deformable, better results. In: *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*; 2019. p. 9308–16.
- [40] Hu J, Shen L, Sun G. Squeeze-and-excitation networks. In: *Proceedings of the IEEE conference on computer vision and pattern recognition*; 2018. p. 7132–41.
- [41] Ronneberger O, Fischer P, Brox T. U-net: convolutional networks for biomedical image segmentation. In: *International conference on medical image computing and computer-assisted intervention*. Springer; 2015. p. 234–41.
- [42] Zhang Y, Liu H, Hu Q. Transfuse: fusing transformers and cnns for medical image segmentation. In: *International conference on medical image computing and computer-assisted intervention*. Springer; 2021. p. 14–24.
- [43] Vaswani A, Shazeer N, Parmar N, Uszkoreit J, Jones L, Gomez AN, et al. Attention is all you need. *Adv Neural Inf Process Syst* 2017;30.
- [44] Liu Z, Lin Y, Cao Y, Hu H, Wei Y, Zhang Z, et al. Swin transformer: hierarchical vision transformer using shifted windows. In: *Proceedings of the IEEE/CVF international conference on computer vision*; 2021. p. 10012–22.
- [45] Valanarasu JMJ, Sindagi VA, Hachililoglu I, Patel VM. Kiu-net: overcomplete convolutional architectures for biomedical image and volumetric segmentation. *IEEE Trans Med Imaging* 2022;41:965–76.
- [46] Huang L, Dai S, Huang T, Huang X, Wang H. Infrared small target segmentation with multiscale feature representation. *Infrared Phys Technol* 2021;116:103755.