Neural Computing and Applications (2023) 35:16129–16142
https://doi.org/10.1007/s00521-021-06554-x

S.I.: AI-BASED E-DIAGNOSIS

Joint optic disc and cup segmentation based on multi-scale feature analysis and attention pyramid architecture for glaucoma screening

Guangmin Sun[1] · Zhongxiang Zhang[1] · Junjie Zhang[1] · Meilong Zhu[1] · Xiao-rong Zhu[2,3] · Jin-Kui Yang[2,3] · Yu Li[1]

Received: 30 June 2021 / Accepted: 14 September 2021 / Published online: 30 September 2021

Abstract

Automatic segmentation of optic disc (OD) and optic cup (OC) is an essential task for analysing colour fundus images. In clinical practice, accurate OD and OC segmentation assist ophthalmologists in diagnosing glaucoma. In this paper, we propose a unified convolutional neural network, named ResFPN-Net, which learns the boundary feature and the inner relation between OD and OC for automatic segmentation. The proposed ResFPN-Net is mainly composed of multi-scale feature extractor, multi-scale segmentation transition and attention pyramid architecture. The multi-scale feature extractor achieved the feature encoding of fundus images and captured the boundary representations. The multi-scale segmentation transition is employed to retain the features of different scales. Moreover, an attention pyramid architecture is proposed to learn rich representations and the mutual connection in the OD and OC. To verify the effectiveness of the proposed method, we conducted extensive experiments on two public datasets. On the Drishti-GS database, we achieved a Dice coefficient of 97.59%, 89.87%, the accuracy of 99.21%, 98.77%, and the Averaged Hausdorff distance of 0.099, 0.882 on the OD and OC segmentation, respectively. We achieved a Dice coefficient of 96.41%, 83.91%, the accuracy of 99.30%, 99.24%, and the Averaged Hausdorff distance of 0.166, 1.210 on the RIM-ONE database for OD and OC segmentation, respectively. Comprehensive results show that the proposed method outperforms other competitive OD and OC segmentation methods and appears more adaptable in cross-dataset scenarios. The introduced multi-scale loss function achieved significantly lower training loss and higher accuracy compared with other loss functions. Furthermore, the proposed method is further validated in OC to OD ratio calculation task and achieved the best MAE of 0.0499 and 0.0630 on the Drishti-GS and RIM-ONE datasets, respectively. Finally, we evaluated the effectiveness of the glaucoma screening on Drishti-GS and RIM-ONE datasets, achieving the AUC of 0.8947 and 0.7964. These results proved that the proposed ResFPN-Net is effective in analysing fundus images for glaucoma screening and can be applied in other relative biomedical image segmentation applications.

Keywords Convolutional neural network · Deep learning · Glaucoma Screening · Joint OD and OC segmentation

1 Introduction

Glaucoma is the second leading cause of blindness in the world (after cataracts) and the first irreversible cause of blindness [26]. It is estimated that glaucoma will affect over 111.8 million people by 2040 [40]. As a chronic disease, glaucoma affects the physiological structure of patients' eyes, causing the thinning of ganglion cells with internal plexiform layer (GCIPL), the increase of cup-disc ratio, and the narrowing of optic disc rim [15]. Normally, no evident symptoms appear in the early stage of glaucoma, which causes numerous patients diagnosed with glaucoma in the late stage when the damage to visual

Guangmin Sun and Zhongxiang Zhang contributed equally to this paper.

✉ Yu Li
yuli@bjut.edu.cn

Guangmin Sun
gmsun@bjut.edu.cn

Zhongxiang Zhang
zzx@emails.bjut.edu.cn

[1] Faculty of Information Technology, Beijing University of Technology, Beijing 100124, China

[2] Beijing Tongren Hospital, Beijing 100730, China

[3] Beijing Institute of Diabetes Research, Beijing 100730, China

function is irreversible. Therefore, early screening is essential for the treatment of glaucoma and prevents the loss of vision.

Currently, colour fundus images and optical coherence tomography (OCT) are the most broadly implemented imaging techniques in the early screening of glaucoma. Compared with OCT, colour fundus image is less expensive and more frequently used for detecting glaucoma. The optic cup (OC) to optic disc (OD) ratio (CDR) of fundus images is an important indicator in the screening and diagnosis of glaucoma [9]. As shown in Fig. 1, the CDR of healthy eyes is generally between 0.3 to 0.4. When the value of CDR reaches 0.65, it is clinically considered to be glaucoma. Manually checking OD and OC is a time-consuming task, and it normally takes a professional ophthalmologist about 8 minutes on average to completely segment the OD and OC in a fundus image [21]. Hence, developing automatic algorithms to segment OD and OC from fundus images is significant for lightening the burden of ophthalmologists and promoting large-scale screenings of glaucoma.

Most of the early segmentation methods of OD and OC are based on hand-crafted features (e.g. colour, gradient and texture features), which include adaptive threshold-based method [2, 27], regional growth method [28] and segmentation method based on Wavelet transform [6]. However, these hand-crafted features are easily affected by the physiological structure of the fundus images.

In recent years, deep learning has achieved excellent performance in tasks such as image classification [16], object detection [30], and image segmentation [24]. A large number of OD and OC segmentation methods based on deep learning have been proposed [12, 34, 36]. Due to the uncertainty of the boundary of the OD and OC in the fundus image, the accurate segmentation of OD and OC is still a challenging task. Most of the existing methods divide the segmentation of OD and OC into two stages or only conduct OD segmentation, which overlooks the inner connection between OD and OC. Moreover, most methods only use a single scale to process the image, which cannot fully capture the detailed features of the OD and OC, especially edge information.

In this paper, we propose a convolutional neural network, named ResFPN-Net, for joint OD and OC segmentation. The main contribution of our work can be summarized as follows:

(1) A segmentation network for joint OD and OC segmentation: Through multi-scale loss supervision, the network can accurately segment the OD and OC from fundus images by fully taking advantage of the internal relationship between OD and OC.
(2) A multi-scale feature extractor: It takes images of different scales as input and merges information from various feature maps, which can adequately express the feature information of the fundus image and preserve the edge features.
(3) An attention pyramid structure: This structure combines attention mechanism with feature pyramid architecture to enhance the representation of OD and OC in the fundus image, which improves the segmentation performance of the network.

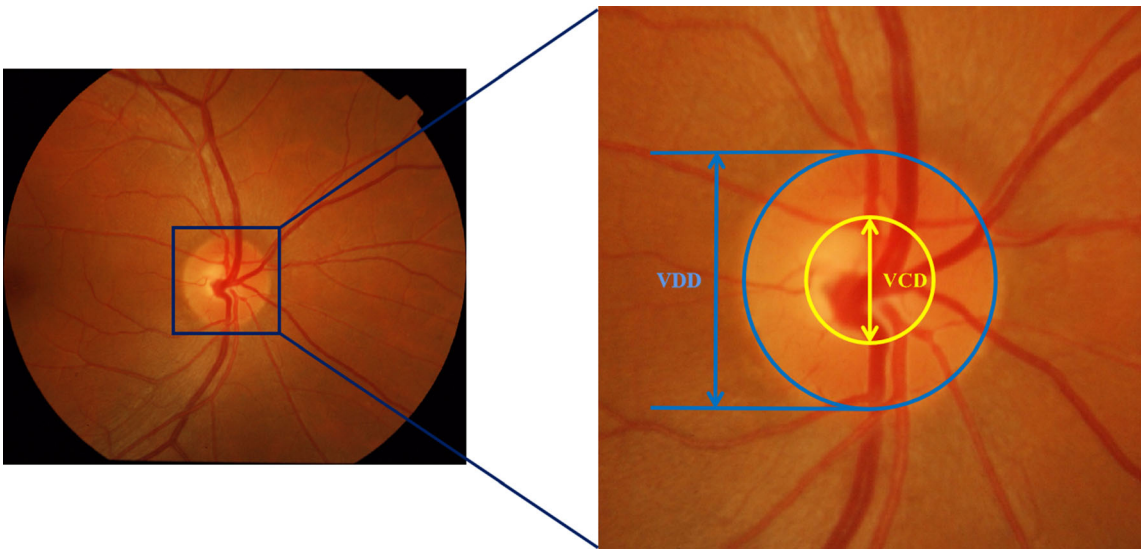

Fig. 1 Structure of the optic disc and optic cup in a fundus image. The region denoted with a blue circle is the optic disc (OD); the region denoted with a yellow circle is the optic cup (OC). The vertical cup-to-disc ratio (CDR) is calculated by the ratio of vertical cup diameter (VCD) to vertical disc diameter (VDD) (color figure online)

2 Related works

In the early stage, most research on OD and OC segmentation is based on hand-craft features. These features mainly include colour, texture, contrast, and gradient information. Abdel-Ghafar et al. [1] proposed a threshold-based segmentation method to segment the OD. This method utilizes the Sobel operator to enhance the fundus image; subsequently, the image is processed by the local threshold and applies Hough transform to get the OD region. Osareh et al. [29] proposed an OD location method based on colour channels. Juneja et al. [39] applied fuzzy C-means clustering method to segment the OD and OC, and the Canny operator is employed for post-processing. In the segmentation method of OD and OC, edge detection algorithms such as the Sobel operator and the Canny operator can improve the accuracy of segmentation. Different from the edge detection operators, the pixel classification-based method transforms the edge detection problem into the pixel segmentation problem and achieves satisfactory results. Jun Cheng et al. [8] proposed a superpixel classification to segment OD and OC and applied histograms and centre-surround statistics to divide each superpixel into disc region and non-disc region. In [42], a method based on deformation is proposed to locate the OD and OC. In addition, template-based methods [20] and reconstruction-based learning method [41] are also widely used in OD and OC segmentation. However, these methods heavily rely on hand-crafted features, which largely affects their performance.

Recently, deep learning has made great achievements in natural image segmentation and medical image segmentation, such as Mask-RCNN [13], U-Net [31]. Many OD and OC segmentation methods based on deep learning have also emerged. In [34], a modified U-Net architecture is proposed to segment the OD and OC, which achieves the lowest possible prediction time compared with traditional convolutional networks. In [18], an end-to-end convolutional neural network, named JointRCNN, is proposed to segment OD and OC, which applied the atrous convolution to boost the performance of segmentation results. However, these methods separate OD and OC segmentation separately. Gu et al. [12] proposed a CE-Net to capture more advanced information and retain spatial information for segmenting OD. Motivated by conventional U-Net architecture, Baid et al. [5] proposed a ResUnet Architecture to segment OD. Al-Bander et al. [33] used VGG as the backbone and transfer learning to solve the problem of OD segmentation. However, based on these methods, only the optic disc region is segmented. Therefore, they ignored the intimate relationship between the OD and OC. Subsequently, the Stack-U-Net [35] was further proposed, which takes U-Net as the backbone and assists the thought training network of iterative refinement. In [43], using ResNet-34 as an encoding layer, a modified U-Net architecture was proposed for the segmentation of OD and OC. Al-Bander et al. [3] proposed a new segmentation network that utilized DenseNet incorporated with a fully convolutional network. Fu et al. [10] used polar transformation to flat the image based on OD centre and applied interpolation to enlarge the cup region. However, the transformation of polar coordinates causes the edges of the OD to be not smooth.

3 Methodology

Inspired by RetinaNet [23], we proposed the ResFPN-Net, as shown in Fig. 2. The framework has four components: multi-scale feature extractor, multi-scale segmentation transition, attention pyramid architecture, and multi-scale loss supervision. The multi-scale feature extractor receives various scale fundus images as input. The multi-scale segmentation transition is used to achieve multi-level feature maps fusion and preserve feature maps of different scales. And then, the feature maps are transmitted into an attention pyramid structure to capture the inner connection within OD and OC. Finally, the segmentation result of the OD and OC is achieved. The entire network is trained by multi-scale loss supervision. The following sub-sections will introduce the details of this architecture.

3.1 Multi-scale extractor

The extraction of OD and OC edge information in fundus images can improve segmentation accuracy. However, in the fundus image, the boundary information of OD and OC is usually not clear, so it is difficult to retain the details based on a single scale. In general, the convolution with a large receptive field is suitable for large objects, while the convolution with a small receptive field can capture detailed information. Therefore, we take the multi-scale fundus image as input to construct various receptive fields and completely learn the edge features. As shown in Fig. 2, we modify the ResNet [14] as our feature extractor. ResNet is an efficient residual network for image classification. Specifically, all fundus images are resized into 512×512, 256×256, 128×128, and 64×64 pixels. We initially applied convolution with a kernel size of 7×7 on the fundus images with a size of 512×512 pixels. Batch normalization (BN) and ReLU activation function are applied to derive the feature map, denoted as s_2. Then we construct different convolution layers to receive multi-scale fundus images, whose kernel size is 3×3, the channel number is 64, 128, and 256, respectively. And,

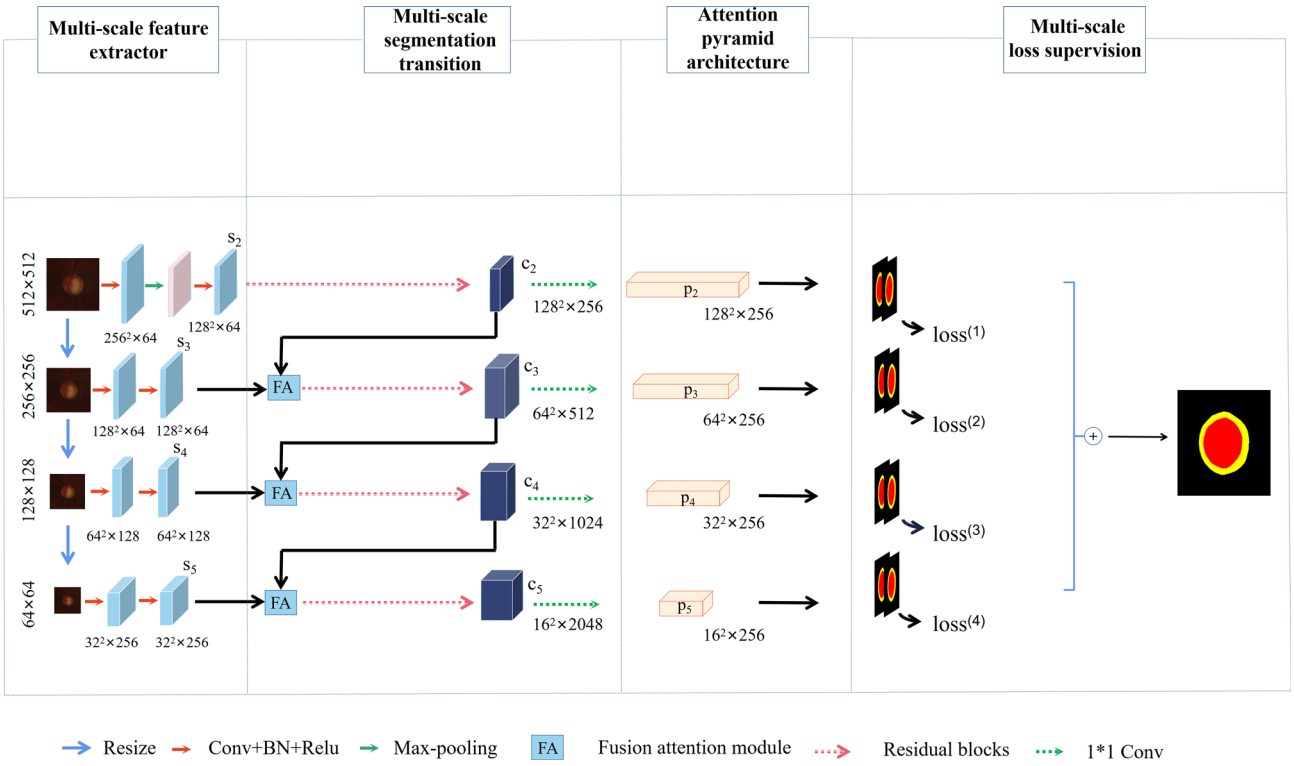

Fig. 2 Overview of our proposed ResFPN-Net. The input to the network consists of multi-scale fundus images. Firstly, the multi-scale fundus images generate the intermediate feature: c_2, c_3, c_4, c_5. Then, the intermediate features are input into the attention pyramid structure for the fusion of different features. Finally, the OD and OC segmentation result is obtained through training of the network

each convolution is followed by a rectified linear unit (ReLU). Finally, the feature map derived from the fundus images of the other three scales is denoted as s_3, s_4, s_5.

3.2 Multi-scale segmentation transition

The encoder–decoder structure is generally employed in many frameworks for image segmentation. In this paper, our segmentation architecture is also based on this structure. In an encoder–decoder structure, the encoder is used to compress and encode the feature information of the image; the decoder is deployed to restore the encoded information. However, some segmentation methods [4, 45] based on encoder-decoder structure do not fully preserve multi-scale feature information. In our segmentation task, the multi-scale input is integrated into the decoder layer to broaden the network width of the decoder path.

To transfer the detailed feature and the multi-scale information to the decoder. We generate a set of feature maps produced by different multi-scale feature maps as information transitions between encoder and decoder. Specifically, the feature map s_2 is fed to a residual block, which consists of a set of convolution and downsamples operations. The feature map derived from the residual block is denoted as c_2. However, there are significant feature gaps between the features extracted from multi-size fundus images. Directly merging these features can weaken the representation of the multi-scale image. In this paper, we proposed a fusion attention module to alleviate gaps among these feature maps, as shown in Fig. 3. Firstly, we merge two feature maps by channel-wise concatenation followed by convolution layer and BN. This procedure can be formulated as follows.

$$V = Conv(concat(c_{i-1}, s_i)), (2 < i \leqslant 5)\,. \tag{1}$$

Then, we collect global contextual information by global average pooling. We apply 1×1 convolution operation and Softmax activate function to derive the attention matrix based on global context information. And the attention matrix is multiplied with V to get the fusion feature map. Finally, the fusion feature map is forwarded to the corresponding residual block. Following the above illustration, multi-level features used to build by fusion attention module and residual blocks are denoted as $\{c_2, c_3, c_4, c_5\}$, which correspond channels are {256, 512, 1024, 2028}, as shown in Fig. 2.

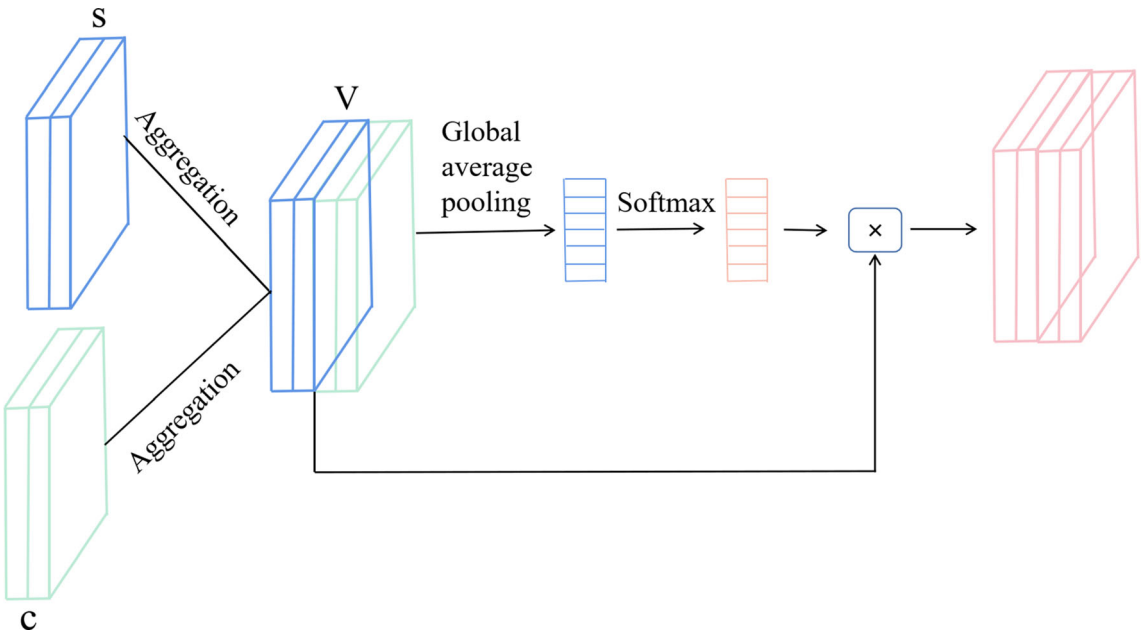

Fig. 3 Illustration of the fusion attention module

3.3 Attention pyramid architecture

We collect four feature maps of different scales through the multi-scale segmentation transition: $\{c_2, c_3, c_4, c_5\}$. Then, we utilize Feature Pyramid Network (FPN) [22] to explore features at different scales. The FPN was originally employed in the object detection task to solve the problem of multi-scale object detection. It adds different feature maps through Top-down pathway and lateral connections to aggregate multi-scale features. However, there are significant differences in these four feature maps. Specifically, the feature maps in the deeper layer are spatially coarser but have more semantic information. In contrast, the feature maps in the lower layer contain rich location information but fewer semantic features. We believe that this simple addition method will weaken the expression of some features and cannot fully learn the close relation between OD and OC. More importantly, fundus vessels in the OD and OC region make it difficult to segment the OD and the OC accurately.

In this paper, we propose an attention pyramid mechanism that concatenates multi-scale features to solve the above problems. In this architecture, an attention module integrates the high-level feature map and the low-level feature map, which bridges the gaps between the deeper feature map and the lower feature map. On the other hand, each region of the input image is given different weights to extract more critical information and help the model distinguish between the target region and the background. Specifically, feature maps obtained by the multi-scale transition: $\{c_2, c_3, c_4, c_5\}$ are fed to the corresponding convolution layer of the pyramid network. Subsequently, the attention module concatenates high-level features with low-level features to achieve feature fusion, as shown in Fig. 4.

Our attention module is based on CBAM [32] and is shown in Fig. 5, where p_i, p_j represents the feature maps from diverse convolution layer. We first feed the p_j with bilinear interpolation and add it to p_i to produce the intermediate feature map f. Then, an Adaptive Average Pooling, Adaptive Max Pooling and 3×3 kernel convolution layer followed by ReLu and Sigmoid activate function to generate two new feature maps $S \in R^{C\times H\times W}$ and $L \in R^{C\times H\times W}$, where C indicates the number of channels, and H and W is the height and width of the feature map. Finally, these two new feature maps are added together to receive the final feature map $O \in R^{C\times H\times W}$.

3.4 Loss function

The OD and OC segmentation is formulated as a multi-label problem in our task. In the original fundus image, the proportion of the background region is more significant than that of OD and OC. The performance of the network is affected by the imbalance of categories in the training process. Therefore, we use focal loss [23] as the loss function for multi-class segmentation, which balances the proportion of the target region and background region by adding weights to the corresponding loss of the sample.

To adequately train the network, we introduced the sub-output layers to construct multi-scale loss. The advantage of the multi-scale loss is that it prevents the gradient from

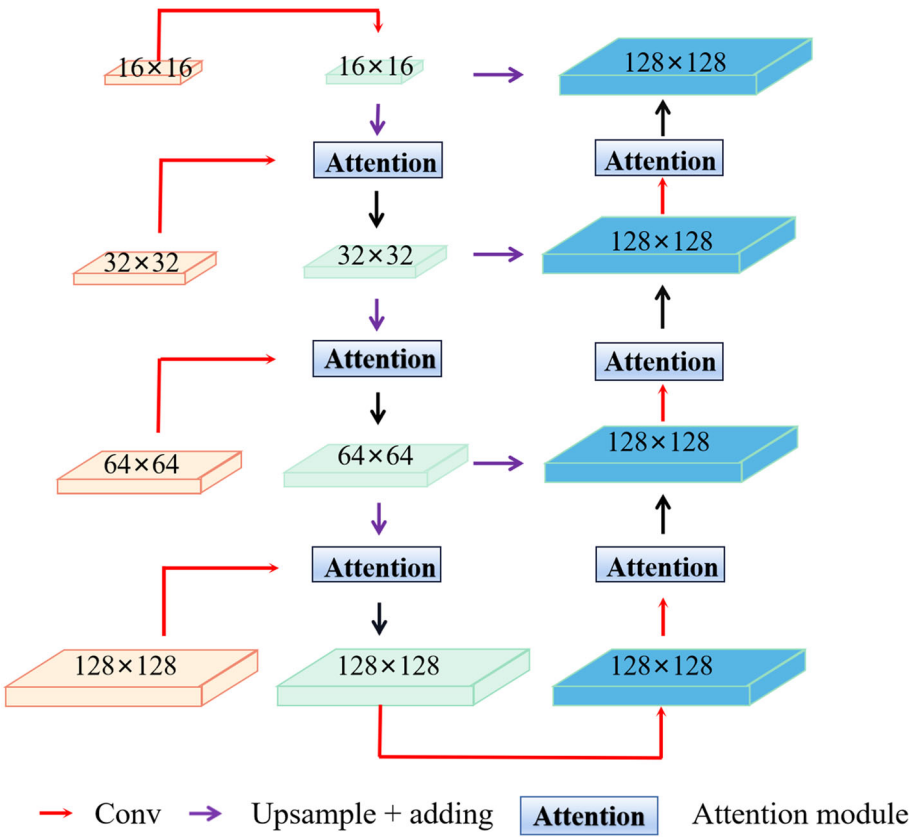

Fig. 4 Structure of the attention pyramid architecture. The structure consists of a feature pyramid and an attention module. The feature pyramid retains multi-scale feature information. The attention module fuses different feature maps and highlights the feature representations

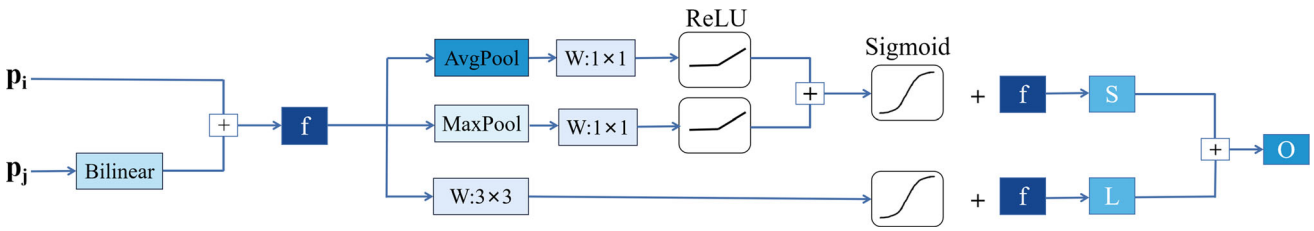

Fig. 5 Illustration of the attention module. The attention module first applies the bilinear interpolation on feature map p_j and adds with the feature map p_i to generate the intermediate feature f. Then, convolution, average pooling, and max pooling followed by nonlinear operation are applied to produce the final feature map O

disappearing during training. In sub-output, the segmentation loss between the mask and the fundus image is formulated by Eq. (2).

$$L_{sub}(P_t) = -\alpha(1-P_t)^{\gamma}\log(P_t), \tag{2}$$

where P_t is the probability of truth class in the network, and α is an equilibrium variable to balance the number of positive and negative samples. γ is a hyperparameter used to focus the model on samples that are difficult to classify during training.

Besides, we integrate sub-outputs to calculate the fusion loss (L_{fusion}). There are four sub-outputs in our task, denoted as O_1, O_2, O_3, O_4, and the fusion of four sub-outputs O can be formulated as:

$$O = O_1 + O_2 + O_3 + O_4\,. \tag{3}$$

L_{fusion} is defined as follows:

$$L_{fusion}(O) = -\beta(1-O)^{\gamma}\log(O)\,. \tag{4}$$

Finally, the multi-scale loss function of the segmentation network is formulated as:

$$L = \sum_{i=1}^{N} L_{sub}^{(i)}(O_i) + L_{fusion}(O)\,, \tag{5}$$

where N represents the number of sub-outputs.

4 Experiments and results

4.1 Datasets and evaluation method

Experiments are conducted on two public datasets. The first dataset is the Drishti-GS dataset [37], collected by Aravind Eye Hospital, Madurai, India. It contains 101 colour fundus images, which are divided into a training set and a testing set. The training set contains 50 images with ground truth for OD and OC segmentation. The remaining 51 images are used for the testing.

The second database is RIM-ONE [11]. It contains 159 fundus images, including 85 images from healthy eyes as well as 74 images from eyes with glaucoma at different stages. RIM-ONE database provides pixel-level segmentation of OD and OC labelled by two ophthalmologists as the ground truth.

Three evaluation metrics are adopted to evaluate our proposed algorithm: Dice coefficient (DC), accuracy (acc) and Hausdorff distance (HD).

$$DC = \frac{2 \times TP}{2 \times TP + FP + FN} \tag{6}$$

$$acc = \frac{TP + TN}{TP + FN + TN + FP} \tag{7}$$

$$HD(A, B) = max(h(A, B), h(B, A)) \tag{8}$$

$$h(A, B) = \max_{a \in A} \min_{b \in B} ||a - b|| \tag{9}$$

where TP, FP, TN and FN represent the number of true positives, false positives, true negatives and false negatives, respectively. And, A, B denote the prediction result and the Ground Truth, a, b represent the pixel belonging to the A and B, respectively.

4.2 Implementation details

The network was implemented by PyTorch[1], and Adam optimization algorithm [19] was used to train the network. The network was trained on a GPU of NVIDIA GeForce 3090 Super with 24 GBs graphic memory. Our multi-scale extractor employs pre-trained parameters based on ImageNet as initialization. During the training, we set the initial learning rate to 0.0001 and used Cosine Decay to adjust the learning rate. In our implementation, we set α and β to 0.25 and γ to 3. We set the mini-batch size to 8 for all training and performed 300 iterations on the network.

To improve the performance of the model, all images were cropped to 800×800 pixels centred on the OD. We used various transformations to augment the training set, including rotation by an angle of 90, 180, and 270 degrees.

[1] https://github.com/pytorch/pytorch.

4.3 Comparison of loss functions

Different loss functions are compared using the Drishti-GS dataset. Cross-Entropy loss, Lovasz_Softmax loss, and Dice loss were applied to train our network, respectively. The model was trained with an initial learning rate of 0.0001. As displayed in Fig. 6, when using multi-scale loss to train the network, the model converges at the loss of 0.008 around 300 epochs. When using Dice loss to train the network, the loss can converge to about 0.06. However, the convergence effect of Lovasz_Softmax loss and cross-entropy loss is not satisfactory, and it only converges to about 0.21 after 300 iterations. Therefore, the proposed multi-scale loss is proved to be more suitable for the training of the OD and OC segmentation network.

4.4 Segmentation results

Extensive experiments were conducted on two public databases. As shown in Table 1, our proposed method achieved scores of 97.59%, 99.21% and 0.099 in *Dice*, *acc* and *HD* for OD segmentation. Moreover, it achieves 89.87%, 98.77% and 0.882 for OC segmentation on the Drishti-GS database. On the RIM-ONE database, our proposed method achieved scores of 96.41%, 99.30% and 0.166 in terms of $Dice_{OD}$, acc_{OD} and $Avg.\ HD_{OD}$, respectively. For OC segmentation, it achieves 83.91%, 99.24% and 1.210 in *Dice*, *acc* and *Avg. HD*.

Based on the OD and OC segmentation results, the corresponding CDR values can be further calculated, which can be used to assist ophthalmologists in the diagnosis of glaucoma. We use the mean absolute error (MAE) to evaluate the accuracy of CDR estimation, which calculates the average error rate of all samples:

$$MAE = \sum_{i=1}^{N} |CDR_i^S - CDR_i^G|, \tag{10}$$

where N represents the number of test samples, CDR^G and CDR^S represent the ground truth of CDR provided by trained clinicians, and the CDR calculated by segmentation results of OD and OC, respectively. Our proposed method achieves MAE of 0.0499 and 0.0630 on the Drishti-GS and RIM-ONE datasets, respectively.

4.5 Accuracy analysis results

The performance comparison with the state-of-the-art approaches on two public databases is shown in Table 1. The results show that our method achieved higher segmentation performance than the state-of-the-art methods. On the Drishti-GS dataset, compared with the CCNet [17], our approach has an improvement of 0.48% and 0.13% in

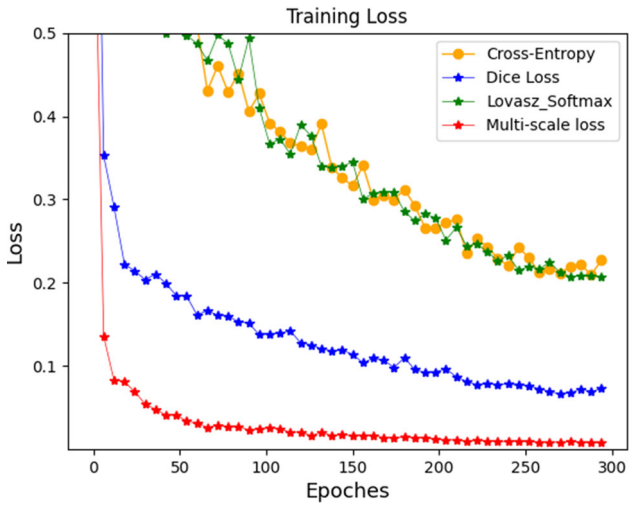

Fig. 6 Compare the variations of different loss in the process of training

Dice and *acc* for OD segmentation, respectively. Furthermore, it has an improvement of 1.73% and 0.18% in terms of *Dice* and *acc* for OC segmentation. On the RIM-ONE database, compared with CCNet, the *Dice* increases from 93.88% to 96.41% by 2.53% and the *acc* increases from 99.03% to 99.30% for OD segmentation. For OC segmentation, the *Dice* increases by 2.82%, and the *acc* increases by 0.18%, respectively. In terms of *HD* metric, a considerable improvement has also been achieved for OD and OC segmentation on Drishti-GS and RIM-ONE datasets. Compared with the state-of-the-art approaches, the proposed method showed superiority in three metrics, as shown in Table 1.

To compare the adaptability of the model on different databases, we provide a comprehensive cross-dataset performance analysis. Firstly, we used the Drishti-GS training dataset to train the model and directly evaluated it on the RIM-ONE testing datasets. Moreover, we also used the RIM-ONE training datasets to train the model and tested it on the Drishti-GS datasets. Since the first two methods in Table 2 do not compare the cross-dataset performance of the model and do not disclose the specific implementation, we cannot obtain its cross-dataset performance. From Table 2, the proposed method remarkably outperforms the U-Net, M-Net [10], AGNet [44], and CCNet models, indicating a solid generalization ability. On the RIM-ONE database, compared to AGNet, the proposed method achieved 7.27% and 1.95% improvements in *Dice* and *acc* for OD segmentation. And, it achieved 22.23% and 2.86% improvements in *Dice* and *acc* for OC segmentation. On the Drishti-GS database, compared with CCNet, the *Dice* increases by 6.82% and the *acc* increases by 3.04% for OD segmentation. For OC segmentation, the *Dice* increases by 0.99% and the *acc* increases by 2.83%. This improvement can also be witnessed for the *HD* metric, which

Table 1 Optic disc and cup segmentation performance on Drishti-GS and RIM-ONE datasets compared with other methods

Method		$Dice_{OD}(\%)\uparrow$	$Dice_{OC}(\%)\uparrow$	$acc_{OD}(\%)\uparrow$	$acc_{OC}(\%)\uparrow$	Avg. $HD_{OD}\downarrow$	Avg. $HD_{OC}\downarrow$
Drishti-GS	RACE-net [7]	97.00	87.00	–	–	–	–
	Son et al. [38]	96.74	–	–	–	–	–
	FCN-8s [25]	92.23	69.49	97.50	96.05	3.167	7.323
	U-Net	95.31	82.50	98.60	97.98	0.547	2.397
	M-Net	96.71	80.18	98.94	97.58	0.194	2.849
	AGNet	96.28	84.35	98.72	97.96	0.334	2.092
	CCNet	97.11	88.14	99.08	98.59	0.167	1.185
	ResFPN (resnet50)	97.34	89.80	99.14	98.70	0.117	0.900
	ResFPN (resnet101)	97.56	**89.87**	99.20	**98.77**	**0.099**	**0.882**
	ResFPN (resnet152)	**97.59**	89.61	**99.21**	98.73	0.102	1.001
RIM-ONE	RACE-net [7]	–	–	–	–	–	–
	Song et al. [38]	95.46	–	–	–	–	–
	FCN-8s [25]	86.01	60.58	97.29	98.17	7.092	11.151
	U-Net	93.99	79.22	98.83	98.94	1.727	2.605
	M-Net	91.17	70.10	98.34	98.77	1.362	4.897
	AGNet	94.35	80.84	98.89	99.02	1.749	2.627
	CCNet	93.88	81.09	99.03	99.06	0.548	1.845
	ResFPN (resnet50)	96.33	83.53	99.28	99.17	0.180	1.437
	ResFPN (resnet101)	**96.41**	**83.91**	**99.30**	**99.24**	0.183	**1.210**
	ResFPN (resnet152)	96.35	83.69	99.28	99.14	**0.166**	1.390

demonstrates the advantages of the proposed method over other approaches on adaptability.

The confusion matrix of segmentation results achieved by other competitive methods and our proposed method is shown in Fig. 7. Compared with other methods, our method can better distinguish the target region from the background and not divide the OC region into the background. Moreover, the number of misclassified pixels in OD and OC regions is lower than that of other methods.

4.6 Visual analysis results

We showed some typical results of the OD and OC segmentation in Fig. 8 to visually compare the proposed method with the competitive methods, including M-Net, AGNet and CCNet. From the comparison, it can be found that our method generates accurate segmentation results and exceeds other approaches. We constructed a multi-scale feature extractor to capture the edge information of the OD and OC. Compared with the previous methods (such as MNet, CCNet), our method is more accurate in depicting the edge information of the OD and OC. Meanwhile, our method used attention pyramid architecture to correlate the task of OD and OC segmentation, which can implicitly learn the relationship between them. It can be seen from Fig. 8, compared with other approaches, the proposed method is more accurate in locating the OD and OC.

We also conducted experiments on CDR calculation. The scatterplot of corresponding CDR values calculated based on OC and OD segmentation results derived by our proposed method and other competitive methods are visualized in Fig. 9. It can be observed that the CDR calculated by the proposed method has the highest correlation with the ground truth. On the Drishti-GS database, the M-Net achieved an MAE of 0.1003, and the AGNet achieved an MAE of 0.0816. In comparison, the proposed method achieved an MAE of 0.0499, which is a relative reduction of 0.0111 from 0.0610 by CCNet. While on the RIM-ONE dataset, the M-Net implemented an MAE of 0.0995, and the AGNet implemented an MAE of 0.0813. The proposed method implemented an MAE of 0.0630, which is a relative reduction of 0.0133 from 0.0763 by CCNet. Compared with other methods, the proposed method achieved the highest accuracy on CDR calculation.

4.7 Glaucoma screening

In this section, we evaluated the proposed method on glaucoma screening by using the calculated CDR value on Drishti-GS and RIM-ONE datasets. Moreover, we described the receiver operating characteristic (ROC) curve and area under the ROC curve (AUC) as the metric of the diagnostic accuracy shown in Fig. 10. From the ROC curves and AUC scores, it can be seen that the proposed method achieved the best performances on two public datasets. Comparing with the CCNet, the AUC scores increased from 0.8725 to 0.8947 on the Drishti-GS dataset. In the other database, comparing with the second-best method achieved by M-Net, the AUC scores increased by

Table 2 Cross-dataset performance on Drishti-GS and RIM-ONE datasets compared with other methods

Method		$Dice_{OD}(\%)\uparrow$	$Dice_{OC}(\%)\uparrow$	$acc_{OD}(\%)\uparrow$	$acc_{OC}(\%)\uparrow$	$Avg.\ HD_{OD}\downarrow$	$Avg.\ HD_{OC}\downarrow$
RIM-ONE	RACE-net [7]	–	–	–	–	–	–
	Song et al. [38]	–	–	–	–	–	–
	FCN-8s [25]	71.58	50.21	93.14	95.26	26.461	32.622
	U-Net	69.12	50.42	94.20	95.99	11.724	19.360
	M-Net	76.29	50.43	94.74	96.81	10.623	19.815
	AGNet	80.04	53.68	95.00	95.89	11.770	21.187
	CCNet	66.81	56.16	93.48	97.72	8.778	8.743
	ResFPN (resnet50)	**87.31**	**75.91**	**96.95**	**98.75**	**3.670**	**5.180**
Drishti-GS	RACE-net t [7]	–	–	–	–	–	–
	Song et al. [38]	–	–	–	–	–	–
	FCN-8s [25]	81.03	51.29	94.53	94.87	4.254	8.442
	U-Net	80.40	66.11	89.75	89.04	14.341	28.045
	M-Net	85.50	63.72	95.86	95.87	3.920	10.8
	AGNet	82.91	51.40	95.11	95.16	3.452	12.005
	CCNet	84.83	72.39	94.37	94.30	6.726	14.943
	ResFPN (resnet50)	**91.65**	**73.38**	**97.41**	**97.13**	**0.936**	**4.338**

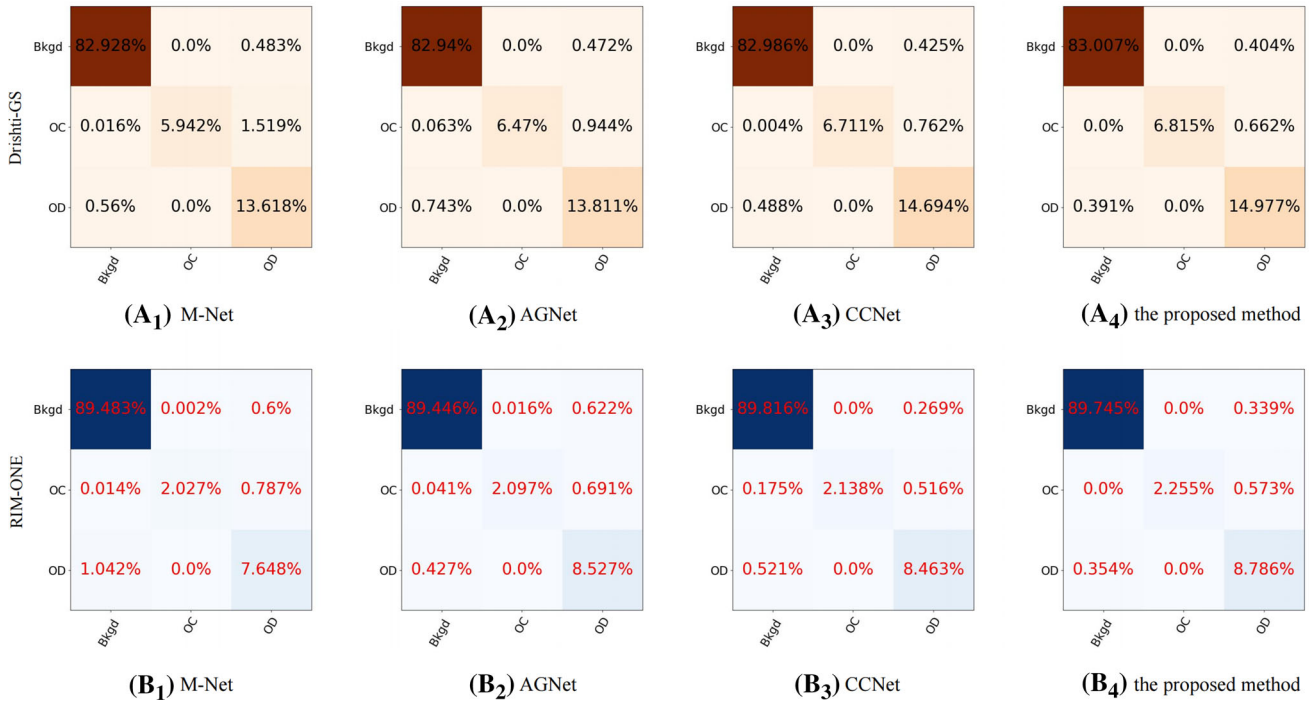

Fig. 7 Confusion matrix on the Drishti-GS database and RIM-ONE, respectively. ($\mathbf{A_1}$, $\mathbf{B_1}$) The M-Net results. ($\mathbf{A_2}$, $\mathbf{B_2}$) The AGNet results. ($\mathbf{A_3}$, $\mathbf{B_3}$) The CCNet results. ($\mathbf{A_4}$, $\mathbf{B_4}$) The proposed method result

1.7%. Compared with other methods, our method has higher accuracy in the diagnosis of glaucoma, which could be used to calculate clinical measurements and support ophthalmologists in clinical diagnosis.

4.8 Ablation experiments

Ablation experiments were conducted on the Drishti-GS dataset. For the sake of description, we used ME, MT, AP and MF to represent the multi-scale extractor, multi-scale segmentation transition, attention pyramid architecture and multi-loss function, respectively. The result achieved by different components of the model is shown in Table 3. We used the ResNet50+FPN network as the baseline model and adopted focal loss to train the model.

When ME, MT, AP and MF were gradually added into the segmentation model, all the evaluation indexes continuedly increased. Hence, the contribution of each improvement of the proposed model is verified. The ME module captures multi-scale features to preserve the boundary and other detailed information, which brings significant benefits to the OD and OC segmentation. Compared with baseline, the *Dice* increased by 1.30%, *acc* increased by 0.44% and the *Avg*. *HD* decreased by 0.105 for OD segmentation. For OC segmentation, the *Dice* increased by 4.96%, the *acc* increased by 0.70% and the *Avg*. *HD* decreased by 0.725. The MT module is integrated into the network to retain the multi-scale feature maps and reduces the burden of the decoder. From Table 3, it can be seen that the MT module has a great contribution to the improvement of segmentation accuracy. The AP module not only eliminates different levels of semantic gaps but also implicitly learns the internal relationship between the OD and OC. When the AP module replaces the corresponding module in the baseline model, the segmentation accuracy is also improved in varying degrees. Finally, we showed that MF supervision could improve the accuracy of the OD and OC. Experiments showed that combined learning these components and used the MF to trained, the network can achieve excellent segmentation results. Therefore, the MF is useful for our segmentation task.

5 Conclusion

In this work, we proposed a novel deep learning architecture that can achieve OD and OC segmentation simultaneously. The proposed ResFPN-Net is trained under multi-loss supervision and converges quickly in a limited time. We have evaluated our method on two public datasets, i.e. Drishti-GS and RIM-ONE. Comprehensive experiments demonstrated the superiority of each improvement and proved that our method could accurately segment OD/OC and outperformed other methods. The proposed multi-scale loss functions converge much quicker, and reached significant lower training loss than the compared loss

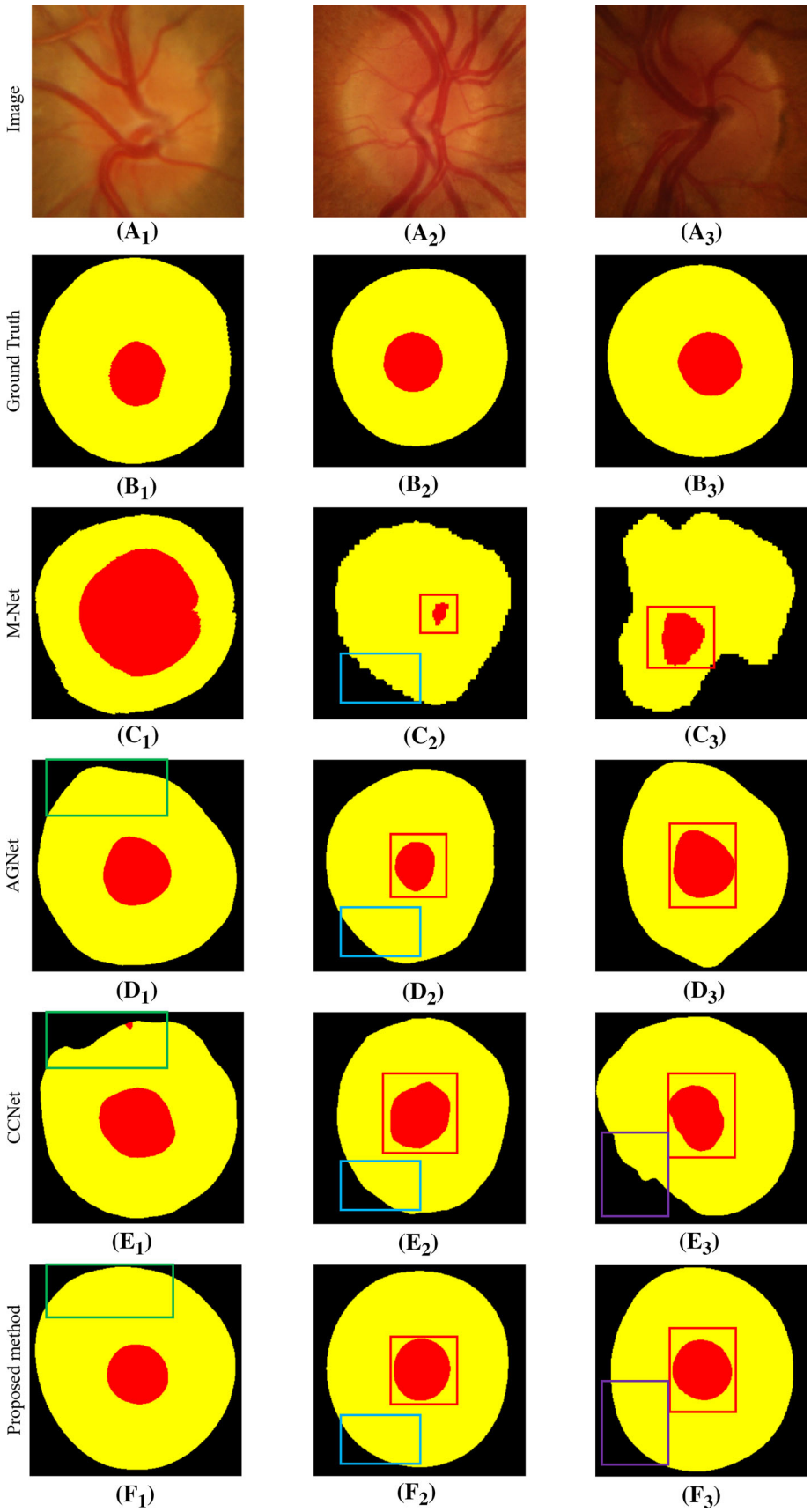

Fig. 8 Examples of visual segmentation results, where the yellow region denotes OD segmentation and the red region denotes OC segmentation result. ($\mathbf{A_1}$, $\mathbf{A_2}$, $\mathbf{A_3}$) Fundus images. ($\mathbf{B_1}$, $\mathbf{B_2}$, $\mathbf{B_3}$) Ground truth. ($\mathbf{C_1}$, $\mathbf{C_2}$, $\mathbf{C_3}$) The M-Net results. ($\mathbf{D_1}$, $\mathbf{D_2}$, $\mathbf{D_3}$) The AGNet results; ($\mathbf{E_1}$, $\mathbf{E_2}$, $\mathbf{E_3}$) The CCNet results. ($\mathbf{F_1}$, $\mathbf{F_2}$, $\mathbf{F_3}$) the proposed method results (The different coloured boxes represent the diverse region in fundus images)

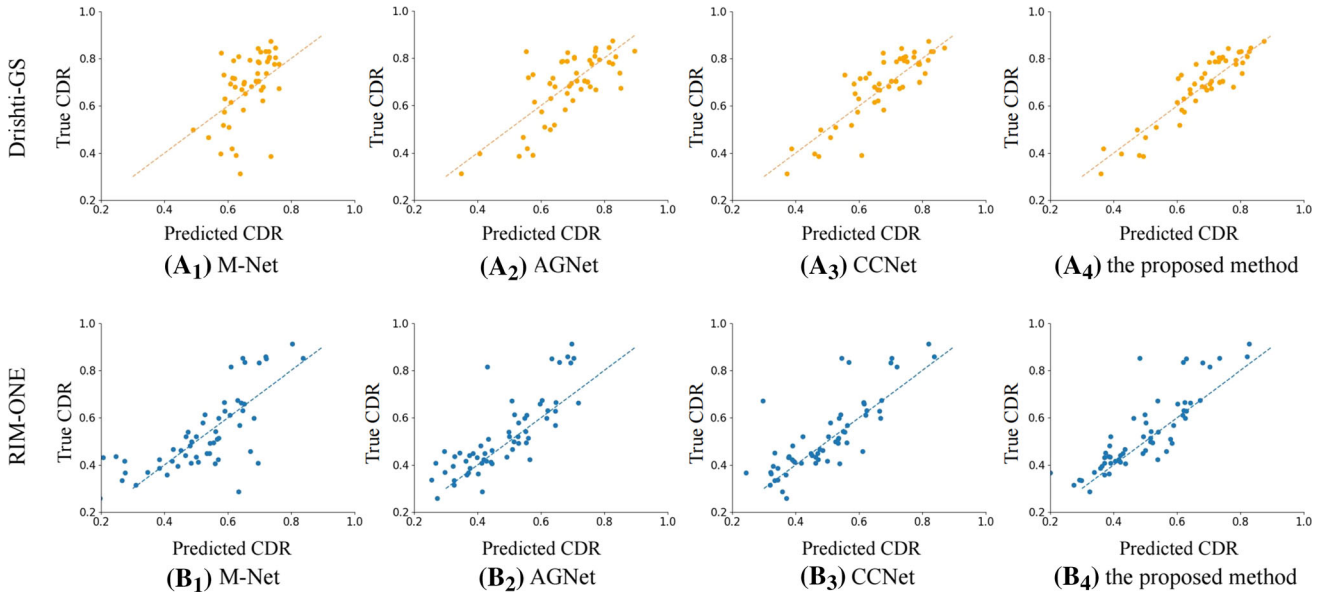

Fig. 9 Scatter plot of the CDR measurement on Drishti-GS and RIM-ONE datasets, respectively. ($\mathbf{A_1}$, $\mathbf{B_1}$) The M-Net results. ($\mathbf{A_2}$, $\mathbf{B_2}$) The AGNet results. ($\mathbf{A_3}$, $\mathbf{B_3}$) The CCNet results. ($\mathbf{A_4}$, $\mathbf{B_4}$) The proposed method result

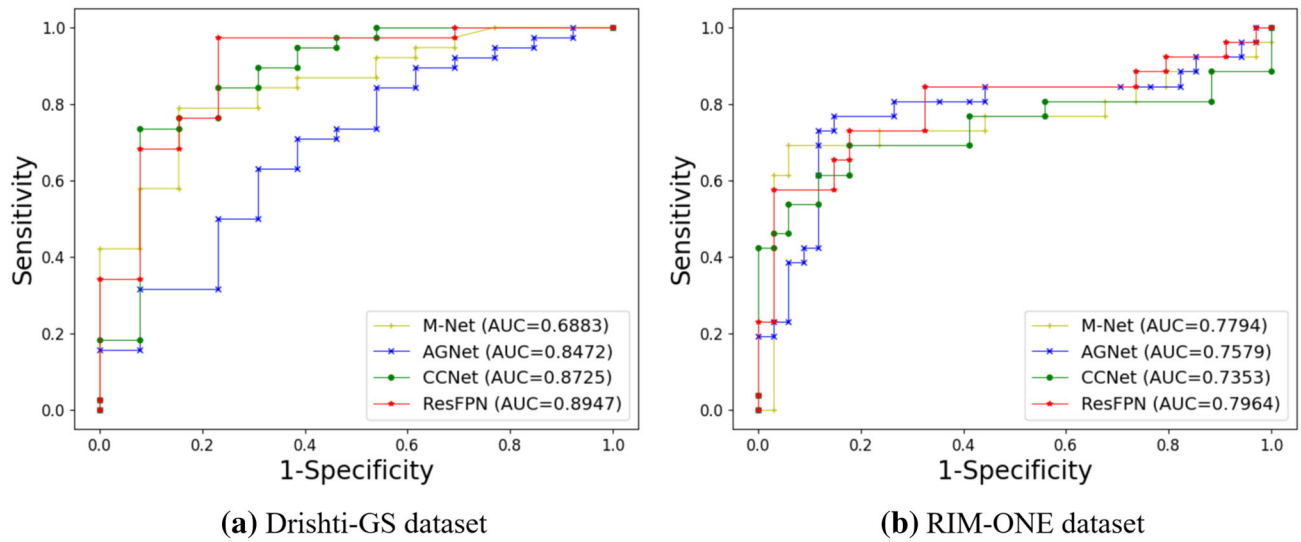

Fig. 10 ROC curves with AUC scores for glaucoma screening based on CDR on Drishti-GS and RIM-ONE datasets

Table 3 Effect of different components of our method on the Drishti-GS dataset

Model	ME	MT	AP	MF	$Dice_{OD}(\%)\uparrow$	$Dice_{OC}(\%)\uparrow$	$acc_{OD}(\%)\uparrow$	$acc_{OC}(\%)\uparrow$	$Avg.\ HD_{OD}\downarrow$	$Avg.\ HD_{OC}\downarrow$
Baseline	×	×	×	×	95.05	80.39	98.37	97.59	0.340	2.550
	✓	×	×	×	96.35	85.35	98.81	98.29	0.235	1.825
	✓	✓	×	×	96.96	87.40	99.01	98.52	0.247	1.517
	✓	✓	✓	×	97.20	88.63	99.10	98.64	0.135	1.106
	✓	✓	✓	✓	**97.34**	**89.80**	**99.14**	**98.70**	**0.117**	**0.900**

functions. By sharing the features from OD and OC for segmentation tasks, the proposed one-stage OD and OC segmentation network achieved both high accuracy and high efficiency. Cross-dataset experiments demonstrated the generalization performance of the network. Ablation experiments proved the contribution of each improvement of the proposed method. Based on the OD and OC segmentation results derived by the proposed ResFPN-Net, more accurate CDR can be calculated, which can provide key support for glaucoma diagnose. The proposed framework also has strong potential for other relative biomedical image segmentation tasks.

Acknowledgment This work was supported by Grants from National Key R&D Program of China (2017YFC0909600); Scientific Research Project of Beijing Educational Committee (KM202110005024) and the Natural Science Foundation of China (41706201).

Declaration

Conflict of interest The authors declare that they have no conflict of interest.

References

1. Abdel-Ghafar R, Morris T (2007) Progress towards automated detection and characterization of the optic disc in glaucoma and diabetic retinopathy. Med Inf Internet Med 32(1):19–25
2. Agarwal A, Gulia S, Chaudhary S, Dutta MK, Travieso CM, Alonso-Hernández JB (2015) A novel approach to detect glaucoma in retinal fundus images using cup-disk and rim-disk ratio. In: 2015 4th international work conference on bioinspired intelligence (IWOBI), IEEE, pp 139–144
3. Al-Bander B, Williams BM, Al-Nuaimy W, Al-Taee MA, Pratt H, Zheng Y (2018) Dense fully convolutional segmentation of the optic disc and cup in colour fundus for glaucoma diagnosis. Symmetry 10(4):87
4. Badrinarayanan V, Kendall A, Cipolla R (2017) Segnet: A deep convolutional encoder-decoder architecture for image segmentation. IEEE Trans Pattern Analy Machine Intell 39(12):2481–2495
5. Baid U, Baheti B, Dutande P, Talbar S (2019) Detection of pathological myopia and optic disc segmentation with deep convolutional neural networks. In: TENCON 2019-2019 IEEE Region 10 Conference (TENCON), pp. 1345–1350. IEEE
6. Bedke GC, Manza RR, Patil DD, Rajput YM (2015) Secondary glaucoma diagnosis technique using retinal nerve fiber layer arteries. In: 2015 International Conference on Pervasive Computing (ICPC), pp. 1–4. IEEE
7. Chakravarty A, Sivaswamy J (2018) Race-net: a recurrent neural network for biomedical image segmentation. IEEE J Biomed Health Inf 23(99), 1–1
8. Cheng J, Liu J, Xu Y, Yin F, Wong DWK, Tan NM, Tao D, Cheng CY, Aung T, Wong TY (2013) Superpixel classification based optic disc and optic cup segmentation for glaucoma screening. IEEE Trans Med Imaging 32(6):1019–1032
9. Fernandez-Granero M, Sarmiento A, Sanchez-Morillo D, Jiménez S, Alemany P, Fondón I (2017) Automatic cdr estimation for early glaucoma diagnosis. J Healthcare Eng , 2017
10. Fu H, Cheng J, Xu Y, Wong DWK, Liu J, Cao X (2018) Joint optic disc and cup segmentation based on multi-label deep network and polar transformation. IEEE IEEE Trans Med Imaging 37(7):1597–1605
11. Fumero F, Alayon S, Sanchez JL, Sigut J, Gonzalez-Hernandez M (2011) Rim-one: an open retinal image database for optic nerve evaluation. In: Computer-Based Medical Systems (CBMS), 2011 24th international symposium on computer-based medical systems (CBMS). IEEE, pp 1–6
12. Gu Z, Cheng J, Fu H, Zhou K, Hao H, Zhao Y, Zhang T, Gao S, Liu J (2019) Ce-net: context encoder network for 2d medical image segmentation. IEEE Trans Med Imaging 38(10):2281–2292
13. He K, Gkioxari G, Dollár P, Girshick R (2017) Mask r-cnn. In: Proceedings of the IEEE International Conference on Computer Vision, pp. 2961–2969
14. He K, Zhang X, Ren S, Sun J (2016) Deep residual learning for image recognition. In: Proceedings of the IEEE Conference on Computer Vision and Pattern recognition, pp. 770–778
15. He M, Foster PJ, Ge J, Huang W, Zheng Y, Friedman DS, Lee PS, Khaw PT (2006) Prevalence and clinical characteristics of glaucoma in adult chinese: a population-based study in liwan district, guangzhou. Investig Ophthalmol Visual Sci 47(7):2782–2788
16. Huang G, Liu Z, Van Der Maaten L, Weinberger KQ (2017) Densely connected convolutional networks. In: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, pp. 4700–4708
17. Huang Z, Wang X, Wei Y, Huang L, Huang TS (2020) Ccnet: criss-cross attention for semantic segmentation. IEEE Transactions on Pattern Analysis and Machine Intelligence (99), 1–1
18. Jiang Y, Duan L, Cheng J, Gu Z, Xia H, Fu H, Li C, Liu J (2019) Jointrcnn: a region-based convolutional neural network for optic disc and cup segmentation. IEEE Trans Biomed Eng 67(2):335–343
19. Kingma D, Ba J (2014) Adam: a method for stochastic optimization. Comput Sci
20. Lalonde M, Beaulieu M, Gagnon L (2001) Fast and robust optic disc detection using pyramidal decomposition and hausdorff-based template matching. IEEE Trans Med Imaging 20(11):1193–1200
21. Lim G, Cheng Y, Hsu W, Lee ML (2015) Integrated optic disc and cup segmentation with deep learning. In: 2015 IEEE 27th International Conference on Tools with Artificial Intelligence (ICTAI), pp. 162–169. IEEE
22. Lin TY, Dollár P, Girshick R, He K, Hariharan B, Belongie S (2017) Feature pyramid networks for object detection. In: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, pp. 2117–2125
23. Lin TY, Goyal P, Girshick R, He K, Dollár P (2017) Focal loss for dense object detection. In: Proceedings of the IEEE International Conference on Computer Vision, pp. 2980–2988
24. Long J, Shelhamer E, Darrell T (2015) Fully convolutional networks for semantic segmentation. In: Proceedings of the IEEE

Conference on Computer Vision and Pattern Recognition, pp. 3431–3440
25. Long J, Shelhamer E, Darrell T (2015) Fully convolutional networks for semantic segmentation. IEEE Trans Pattern Analy Machine Intell 39(4):640–651
26. Mary M, Rajsingh E, Naik G (2016) Retinal fundus image analysis for diagnosis of glaucoma: A Comprehensive Survey. IEEE Access pp. 4327–4354
27. Nayak J, Acharya R, Bhat PS, Shetty N, Lim TC (2009) Automated diagnosis of glaucoma using digital fundus images. J Med Syst 33(5):337–346
28. Omid S, Shanbehzadeh J, Ghassabi Z, Ostadzadeh SS (2015) Optic disc detection in high-resolution retinal fundus images by region growing. In: 2015 8th International Conference on Biomedical Engineering and Informatics (BMEI), pp. 101–105. IEEE
29. Osareh A, Mirmehdi M, Thomas B, Markham R (2002) Comparison of colour spaces for optic disc localisation in retinal images. In: Object recognition supported by user interaction for service robots, 1: 743–746. IEEE
30. Redmon J, Farhadi A (2018) Yolov3: An incremental improvement. arXiv preprint arXiv:1804.02767
31. Ronneberger O, Fischer P, Brox T (2015) U-net: Convolutional networks for biomedical image segmentation. In: International Conference on Medical image computing and computer-assisted intervention, pp. 234–241. Springer
32. Roy A.G., Navab N, Wachinger C (2018) Concurrent spatial and channel 'squeeze and excitation'in fully convolutional networks. In: International conference on medical image computing and computer-assisted intervention, pp. 421–429. Springer
33. Sarhan A, Al-Khaz'Aly A, Gorner A, Swift AJ, Rokne JG, Alhajj RS, Crichton A (2021) Utilizing transfer learning and a customized loss function for optic disc segmentation from retinal images
34. Sevastopolsky A (2017) Optic disc and cup segmentation methods for glaucoma detection with modification of u-net convolutional neural network. Pattern Recog Image Analy 27(3):618–624
35. Sevastopolsky A, Drapak S, Kiselev K, Snyder BM, Keenan JD, Georgievskaya A (2019) Stack-u-net: refinement network for improved optic disc and cup image segmentation. In: Medical Imaging 2019 Image Processing, International Society for Optics and Photonics, 10949: 1094928.
36. Singh VK, Rashwan HA., Akram F, Pandey N, Sarker MMK, Saleh A, Abdulwahab S, Maaroof N, Torrents-Barrena J, Romani S, et al (2018) Retinal optic disc segmentation using conditional generative adversarial network. In: CCIA, pp. 373–380
37. Sivaswamy J, Krishnadas SR, Joshi GD, Jain M, Tabish A (2014) Drishti-gs: Retinal image dataset for optic nerve head(onh) segmentation. In: 2014 IEEE 11th International Symposium on Biomedical Imaging (ISBI 2014)
38. Son J, Park SJ, Jung KH (2018) Towards accurate segmentation of retinal vessels and the optic disc in fundoscopic images with generative adversarial networks. J Digital Imaging ,32
39. Thakur N, Juneja M (2017) Clustering based approach for segmentation of optic cup and optic disc for detection of glaucoma. Current Med Imaging 13(1):99–105
40. Tham YC, Li X, Wong TY, Quigley HA, Aung T, Cheng CY (2014) Global prevalence of glaucoma and projections of glaucoma burden through 2040: a systematic review and meta-analysis. Ophthalmology 121(11):2081–2090
41. Xu Y, Lin S, Wong DWK, Liu J, Xu D (2013) Efficient reconstruction-based optic cup localization for glaucoma screening. In: International conference on medical image computing and computer-assisted intervention, pp. 445–452. Springer
42. Yin F, Liu J, Ong SH, Sun Y, Wong DW, Tan NM, Cheung C, Baskaran M, Aung T, Wong TY (2011) Model-based optic nerve head segmentation on retinal fundus images. In: 2011 annual international conference of the IEEE engineering in medicine and biology society. IEEE, pp 2626–2629
43. Yu S, Xiao D, Frost S, Kanagasingam Y (2019) Robust optic disc and cup segmentation with deep learning for glaucoma detection. Comput Med Imaging Graphics 74:61–71
44. Zhang S, Fu H, Yan Y, Zhang Y, Wu Q, Yang M, Tan M, Xu Y (2019) Attention guided network for retinal image segmentation
45. Zhao H, Shi J, Qi X, Wang X, Jia J (2017) Pyramid scene parsing network. In: Proceedings of the IEEE conference on computer vision and pattern recognition, pp 2881–2890