Physics in Medicine & Biology

PAPER

Segmentation of retinal detachment and retinoschisis in OCT images based on complementary multi-class segmentation networks

To cite this article: Fei Shi *et al* 2023 *Phys. Med. Biol.* **68** 115019

IOP Publishing Phys. Med. Biol. 68 (2023) 115019 https://doi.org/10.1088/1361-6560/acd223

Physics in Medicine & Biology

IPEM Institute of Physics and Engineering in Medicine

PAPER

Segmentation of retinal detachment and retinoschisis in OCT images based on complementary multi-class segmentation networks

Fei Shi[1], **Changqing Yang**[1], **Qingxin Jiang**[1], **Weifang Zhu**[1], **Xun Xu**[2], **Xinjian Chen**[1,3] **and Ying Fan**[2]

[1] MIPAV Lab, School of Electronic and Information Engineering, Soochow University, Suzhou 215006, People's Republic of China
[2] First People's Hospital Affiliated to Shanghai Jiao Tong University, Shanghai, 200080, People's Republic of China
[3] State Key Laboratory of Radiation Medicine and Protection, Soochow University, Suzhou 215123, People's Republic of China

E-mail: mdfanying@sjtu.edu.cn

Keywords: multi-class segmentation, OCT image, deep learning, retinal detachment, retinoschisis

RECEIVED
25 January 2023

REVISED
14 April 2023

ACCEPTED FOR PUBLICATION
3 May 2023

PUBLISHED
30 May 2023

Abstract

Retinal detachment (RD) and retinoschisis (RS) are the main complications leading to vision loss in high myopia. Accurate segmentation of RD and RS, including its subcategories (outer, middle, and inner retinoschisis) in optical coherence tomography images is of great clinical significance in the diagnosis and management of high myopia. For this multi-class segmentation task, we propose a novel framework named complementary multi-class segmentation networks. Based on domain knowledge, a three-class segmentation path (TSP) and a five-class segmentation path (FSP) are designed, and their outputs are integrated through additional decision fusion layers to achieve improved segmentation in a complementary manner. In TSP, a cross-fusion global feature module is adopted to achieve global receptive field. In FSP, a novel three-dimensional contextual information perception module is proposed to capture long-range contexts, and a classification branch is designed to provide useful features for segmentation. A new category loss is also proposed in FSP to help better identify the lesion categories. Experiment results show that the proposed method achieves superior performance for joint segmentation of RD and the three subcategories of RS, with an average Dice coefficient of 84.83%.

1. Introduction

High myopia can cause many pathological changes in the retina, among which retinal detachment (RD) and retinoschisis (RS) are the most common complications, which can seriously impair the visual function (Lai 2007). With optical coherence tomography (OCT), RD and RS can be observed clearly and non-invasively (Fujimoto *et al* 2010). Retinoschisis is characterized by the splitting of retinal neuroepithelium (RNE) layer. According to the retina layers where it occurs, retinoschisis can be divided into outer retinoschisis (ORS), middle retinoschisis (MRS), and inner retinoschisis (IRS). Retinal detachment refers to the separation of the RNE and retinal pigment epithelium (RPE). Figure 1 shows an example OCT B-scan with ORS, MRS, IRS and RD. In pathological myopia, RS generally occurs in the early stage. With the development of the disease, more number of RS will occur and the area will become larger. In a more advanced stage, RD will develop, and surgery is required (Takano 1999, Frisina *et al* 2020, Benhamou *et al* 2022). Quantization of RD, ORS, MRS, and IRS is important for the diagnosis, staging, management, and postoperative assessment of pathological myopia (Frisina *et al* 2020).

With the rise of the convolutional neural network (CNN), it has been used in more and more lesion segmentation tasks in OCT images, such as for segmentation of retinal edema (Feng *et al* 2020), retinal layer and fluid (Roy *et al* 2017), subretinal fluid and pigment epithelial detachment (Gao *et al* 2019, Hu *et al* 2019), macular hole and cystoid macular edema (Ye *et al* 2020). Its excellent capability of feature extraction results in superior segmentation performance. However, there are few works studying the automatic segmentation of RD, ORS, MRS, and IRS. The segmentation of these lesions faces some challenges: (1) uneven distribution of categories, because not all categories appear in a particular B-scan, (2) various sizes of target regions, some are small and

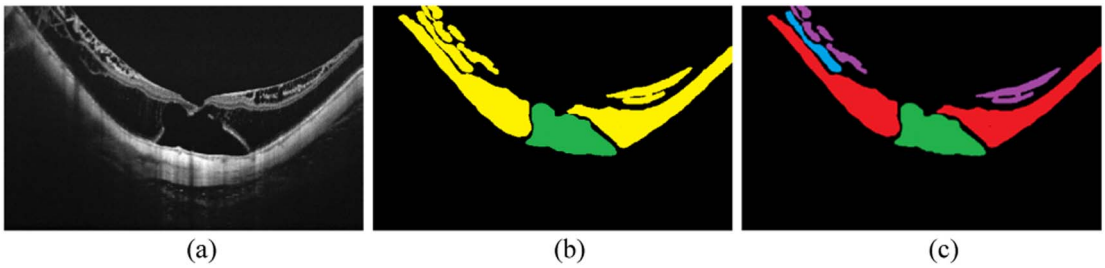

Figure 1. (a) Retinal detachment and retinoschisis in OCT images. (b) (c) the corresponding labels (RD is represented in green, RS is represented in yellow, ORS is represented in red, MRS is represented in blue, and IRS is represented in purple).

some have a large horizontal span, and (3) closeness in location and intensity. Especially, ORS, MRS, and IRS are more similar in shape and texture, and are thus more prone to segmentation errors.

In Yang *et al* (2021), we reported some preliminary work on segmentation of RD and RS, based on a U-shaped network embedded with a cross-fusion global feature module (CFGF). In this work, we substantially expand the previous work, and design the complementary multi-class segmentation networks (CMC-Net) to accomplish the task of RD, ORS, MRS, and IRS segmentation, and the results were obtained from a larger dataset.

It has been shown that ensemble of multiple segmentation models may improve the results (Causey *et al* 2022, Alsaih *et al* 2020, Golla *et al* 2021). Based on the prior knowledge that the difference between RD and RS is larger than that among subcategories of RS, we design a double-path framework. One path is a three-class segmentation network focusing on differentiation of RD and RS, which is a modified version of the network reported in (Yang *et al* 2021), while the other is a five-class segmentation network, emphasizing more on the subcategories of RS. Then, multiple convolutional layers are trained to fuse the results of the two paths on the decision level.

Capturing contextual information is essential for image segmentation tasks. In addition to the local context, long-distance contextual information is especially important when dealing with target regions with a large spatial span, and identifying different target categories in multi-class segmentation. Some networks used global or pyramid pooling operations to obtain useful contextual information (Zhao *et al* 2017, Zhang *et al* 2018, Gu *et al* 2019, Hu *et al* 2019). Atrous convolution, first proposed in Deeplabv3 (Chen *et al* 2017), was often adopted to increase the receptive field of the network to obtain contextual information (Chen *et al* 2018, Yang *et al* 2018, Mehta *et al* 2018a, Feng *et al* 2020). Some methods used non-local operations (Wang *et al* 2018), which allowed a single element in any location to perceive information from all other locations (Fu *et al* 2019, Zhu *et al* 2019, Mou *et al* 2021). However, these methods have their limitations. Pyramid pooling and atrous convolution can only obtain contextual information around a certain pixel, and improve the local perception of the network. Non-local models achieve a global receptive field, but at the expense of huge computational complexity and memory cost. To cope with the above problem, regarding the fact that RS may accounts for a large proportion of the entire image in the three-class segmentation, we use the CFGF module (Yang *et al* 2021) to fuse global information and obtain a global receptive field. For five-class segmentation, we propose a three-dimensional contextual information perception module (TCIP) to obtain the long-range contextual information in the channel, height, and width dimensions. Both modules expand the receptive field of the network while still have acceptable complexity.

Many studies have shown that multi-task learning can improve the performance of the model through the information sharing between different tasks (Mehta *et al* 2018b, Kawakami *et al* 2019, Xu *et al* 2020, Zhang *et al* 2021, Zhou *et al* 2021). For multi-class segmentation tasks, it is important for the segmentation network to identify the categories that appear in the image, which can alleviate the problem of wrong target labeling in the segmentation results. Therefore, for five-class segmentation, we propose to add a classification branch to the network, whose features are fused into the segmentation network and help to improve the segmentation performance. To further guide the model toward accurate identification of different target types, we also propose a new category loss to constrain the segmentation results.

In summary, the main contributions of this work are listed as follows:

- The CMC-Net, where the results from a three-class segmentation path (TSP) and a five-class segmentation path (FSP) are fused by decision fusion layers, are proposed for fully automatic segmentation of RD, ORS, MRS, and IRS in OCT images.

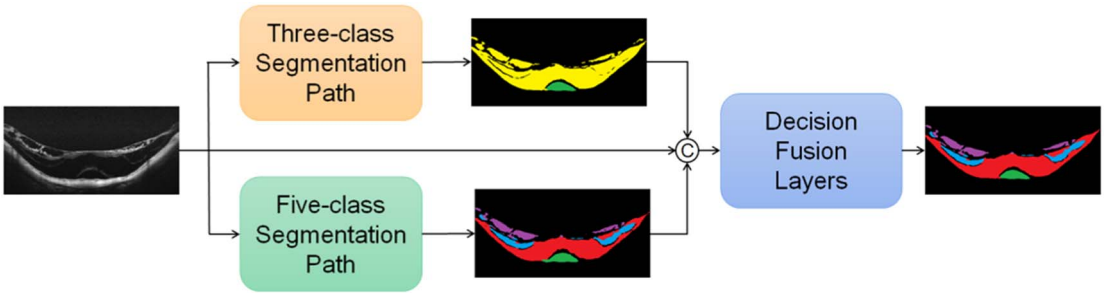

Figure 2. Overview of the proposed framework.

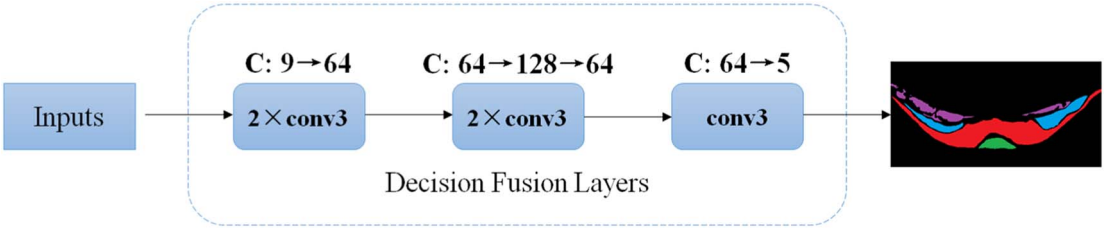

Figure 3. Structure of the decision fusion layers.

- A TCIP is proposed to obtain the contextual information fusion of channel, height, and width dimensions and increase the receptive field of the network.
- A classification-assisted segmentation idea is proposed, where a classification branch can provide auxiliary feature information to help the segmentation network.
- A category loss function is proposed to train the segmentation network to learn discriminative features between different types of targets, alleviating the problem of wrong label assignment in the segmentation results.

2. Methods

2.1. The complementary multi-class segmentation framework

As shown in figure 2, the proposed CMC-Net is composed of the TSP, the FSP, and the decision fusion layers. The TSP focuses on segmentation of RD and RS from the background, and the FSP is committed to segmentation of RD, ORS, MRS, and IRS from the background. The TSP and FSP are designed differently to suit their respective tasks. The difference is also necessary for the ensemble strategy to be effective.

As shown in figure 3, the decision fusion layers are composed of five 3×3 convolutional layers. They take a nine-channel input, which is the concatenation of the predicted probability maps output by the TSP and the FSP, and the original input image, and the output is five-channel corresponding to the probability maps for the five class, including RD, ORS, MRS, IRS and the background. The intermediate channel numbers are 64 or 128, as detailed in figure 3. Using convolution, for each location in the output, the result comes from a neighborhood of the TSP and FSP predictions, and also takes consideration of information from the original image. Therefore, the final segmentation fuses the two results by incorporating the spatial context.

2.2. The TSP framework

As shown in figure 4(a), the TSP network is a U-shaped structure (Ronneberger *et al* 2015), and adopts the general structure of our previous work (Yang *et al* 2021). The lower part of the encoder is replaced by the four stages of the middle part of Resnet18 (He *et al* 2016). The first two stages use a down-sampling operation with convolution strides of 2, and the latter two stages use dilated convolution instead of the downsampling to reduce the loss of detailed features. To make the network achieve global information, a CFGF module is added in the bottom of the TSP. In addition to Yang *et al* (2021), A deep supervision (DS) module is applied in each layer of the decoder, which will force the segmentation network to focus more on the target region and accelerate the

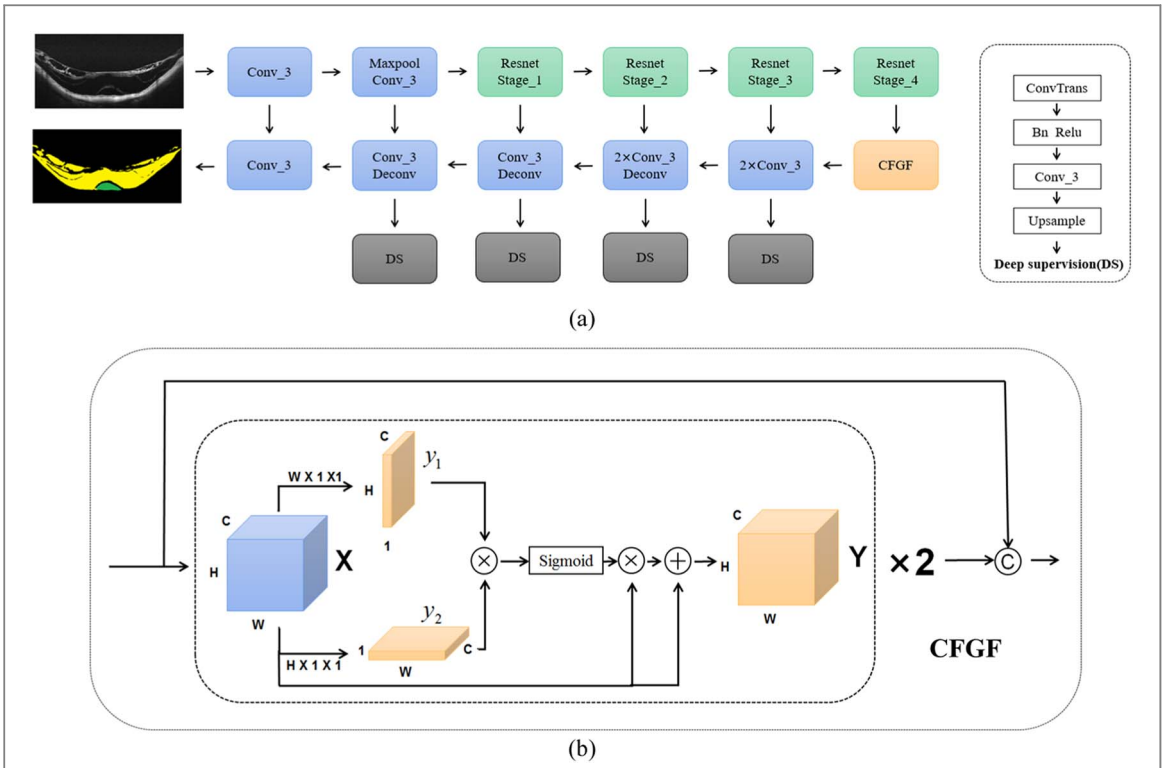

Figure 4. Structure of the three-class segmentation path (a) overall structure, (b) cross-fusion global feature module.

convergence in training. In the DS module, the feature map from each decoder layer is upsampled to the original image size after deconvolution and convolution operations. The result is compared with the ground truth and a DS loss is calculated.

2.3. Cross-fusion global feature module (CFGF)

In one OCT image, there can be multiple RS regions and they can account for a large proportion of the entire image. Besides, as the intensities of RS and RD are similar, global context such as the overall retinal structure is needed to distinguish them. Based on the above considerations, we adopted the CFGF module in the TSP.

As shown in figure 4(b), for a three-dimensional input tensor $X \in R^{C\times H\times W}$, where C is the channel number, H is the height, and W is the width, the cross-fusion operation for obtaining tensor $Y \in R^{C\times H\times W}$ can be written as:

$$y_{zij} = \text{Sigmoid}\left(\sum_{m=1}^{W} a_m x_{zim} \sum_{n=1}^{H} b_n x_{znj}\right) x_{zij} + x_{zij}, \tag{1}$$

where x_{zij} and y_{zij} are elements in X and Y with channel index z, row index i and column index j, and a_m and b_n are the learned fusion weights in the horizontal and vertical dimensions, respectively. Finally, after two such operations consecutively, each element in the output of CFGF will contain information from all locations in the corresponding input feature map. Therefore, embedded with the CFGF module, the TSP network obtains a global receptive field. Refer to Yang *et al* (2021) for more details of the module.

2.4. The FSP framework

As the task of FSP involves more categories, there come more uncertainties. Some categories may not occur in a certain image, and the difference among the three subcategories of RS is even smaller than that between RD and RS. These factors will cause wrong label assignments to segmented regions, resulting in error even when the region boundary is correctly delineated. To cope with the problem, we propose to use a classification-guided segmentation network in FSP, where features obtained by the classification task are fused with those of the segmentation network, and the classification loss and a new category loss are also employed in network optimization.

As shown in figure 5(a), the encoders of the classification network and the segmentation network both adopt four stages of pretrained Resnet18 (He *et al* 2016). A TCIP module is added to the bottom of the segmentation network to obtain the long-range contextual information in the channel, height, and width dimensions. The decoder of the segmentation network is composed of some feature merge decoder (FMD) modules (figure 5(b)),

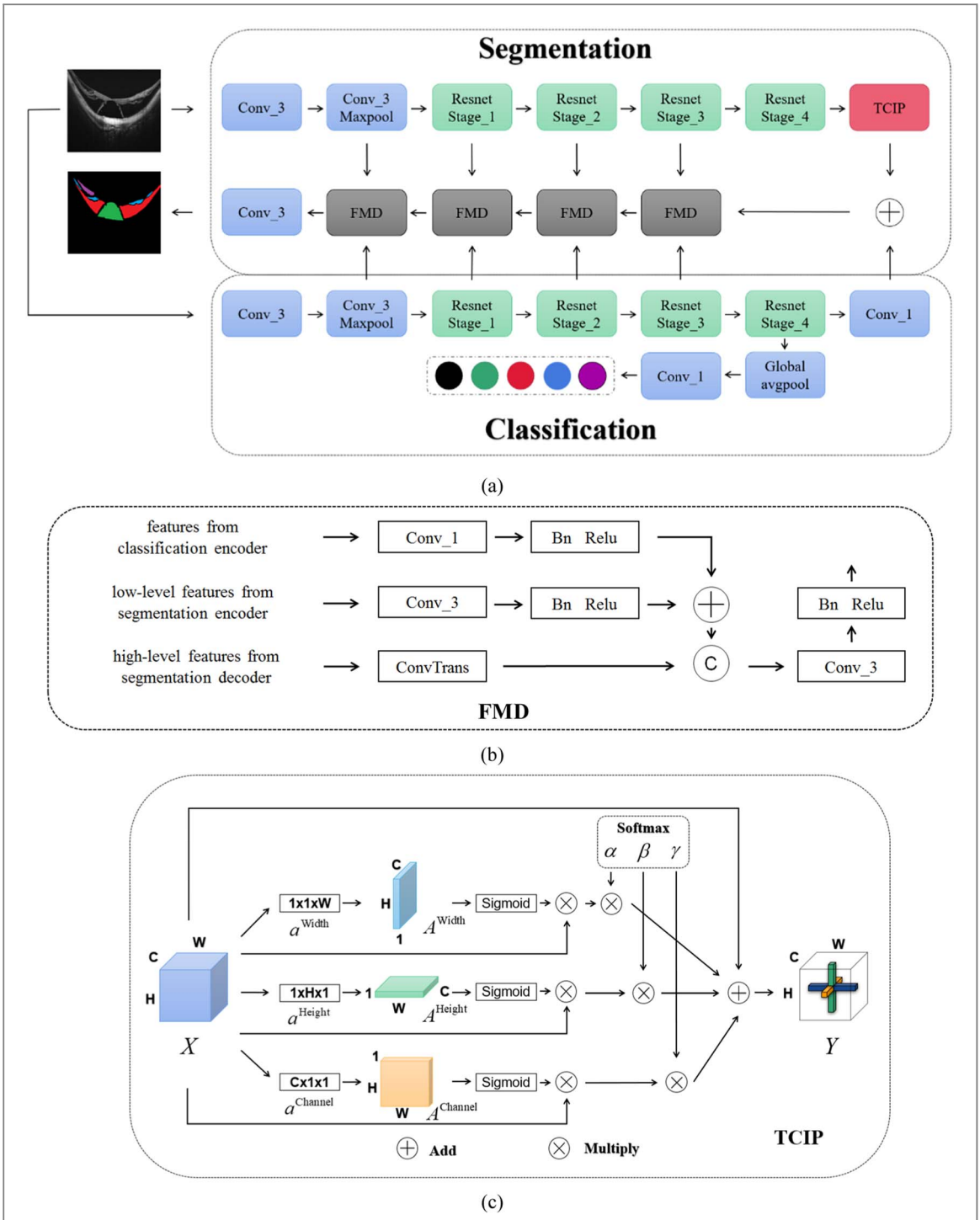

Figure 5. Structure of the five-class segmentation path (a) overall structure, (b) feature merge decoder module, (c) three-dimensional contextual information perception module.

which merge the high-level features with the low-level features extracted from both the classification and segmentation network encoders. The FMD module uses 1×1 convolution to process the semantic features of the classification network, hoping to extract information that is conducive to segmentation and add it to the segmentation features, while suppressing information that hinders the segmentation task.

2.5. Three-dimensional contextual information perception module (TCIP)

The RD, ORS, MRS, and IRS in the OCT image have large distribution span and big variation in size. In order to cope with such difficulties in segmentation, we propose the TCIP module to fuse the contextual information in channel, height, and wc receptive field of the network.

As shown in figure 5(c), the input tensor $X \in R^{C\times H\times W}$ is first fed into three parallel pathways, where it is squeezed in one dimension using a group of learnable weights $a^{\text{Width}} \in R^{1\times1\times W}$, $a^{\text{Height}} \in R^{1\times H\times1}$, or

$a^{\text{Channel}} \in R^{C\times 1\times 1}$. This gives $A^{\text{Width}} \in R^{C\times H\times 1}$, $A^{\text{Height}} \in R^{C\times 1\times W}$, and $A^{\text{Channel}} \in R^{1\times H\times W}$ respectively. Then, after the Sigmoid function, these squeezed features of different dimensions are multiplied with the original feature maps. Three learnable parameters constrained by the Softmax function are used to weight the feature maps from the three branches. Finally, the output feature map $Y \in R^{C\times H\times W}$ is obtained by adding all the feature maps from the three branches and the original feature map. Therefore, an element y_{zij} in the feature map Y is calculated from X as follows:

$$\begin{aligned} y_{zij} = {} & \alpha\ \text{Sigmoid}\left(\sum_{m=1}^{W} a_m^{\text{Width}} x_{zim}\right) x_{zij} \\ & + \beta\ \text{Sigmoid}\left(\sum_{n=1}^{H} a_n^{\text{Height}} x_{znj}\right) x_{zij} \\ & + \gamma\ \text{Sigmoid}\left(\sum_{t=1}^{C} a_t^{\text{Channel}} x_{tij}\right) x_{zij} + x_{zij}, \end{aligned} \tag{2}$$

where a_m^{Width}, a_n^{Height}, and a_t^{Channel} are elements from a^{Width}, a^{Height} and a^{Channel} respectively, and $\gamma = 1 - \alpha - \beta$.

Through the above process, each element in the output feature map contains information from elements in the same row, the same column, and at the same spatial location of all channels. Feature recalibration is achieved based on long-range context information from all dimensions.

2.6. Loss functions

To alleviate the problem of unbalanced pixel categories, we choose the Dice loss (Milletari *et al* 2016) as the main loss function of the segmentation network. Here we treat multi-class segmentation as multiple binary segmentation tasks, and define the Dice loss based on the average Dice score over all categories:

$$\text{loss}_{\text{Dice}} = 1 - \frac{1}{N}\sum_{n=1}^{N} \frac{2\sum_{i=1}^{S} p_{n,i} y_{n,i}}{\sum_{i=1}^{S} p_{n,i}^2 + \sum_{i=1}^{S} y_{n,i}^2}, \tag{3}$$

where N represents the number of categories, S represents the number of pixels, $p_{n,i} \in [0, 1]$ represents the predicted probability of the nth category for the ith pixel, and $y_{n,i} \in \{0, 1\}$ represents the ground truth label indicating whether the ith pixel belongs to the nth category.

For TSP, the total loss function is the weighted sum of the Dice loss of the final prediction and the four DS modules.

$$\text{loss}_{\text{TSP}} = \text{loss}_{\text{Dice}} + \lambda_{\text{l}} \sum_{k=1}^{4} \text{loss}_{\text{Dice}}^{k}, \tag{4}$$

where $N = 3$ when calculating Dice loss.

The classification network in the FSP uses the binary cross-entropy (BCE) loss function (Ridnik *et al* 2021). Again, multi-class classification is treated as multiple binary classification tasks, and the BCE loss is defined as the average over all categories:

$$\text{loss}_{\text{BCE}} = \frac{1}{N}\sum_{n=1}^{N} - t_n \log(P_n) - (1 - t_n)\log(1 - P_n), \tag{5}$$

where N represents the number of categories, $P_n \in [0, 1]$ represents the predicted classification probability for the nth category output by the classification branch, and $t_n \in \{0, 1\}$ represents the nth category label for each image. t_n is generated from the segmentation ground truth map, i.e. $t_n = \max_i \{y_{n,i}\}$ indicating whether any nth category pixel appears in the ground truth map.

Therefore, for FSP, the joint loss function of classification and segmentation is:

$$\text{loss}_{\text{Jo int}} = \text{loss}_{\text{Dice}} + \text{loss}_{\text{BCE}}, \tag{6}$$

where $N = 5$ when calculating the Dice and BCE loss.

The BCE loss mainly constrains the classification branch and assists the segmentation indirectly. To offer a more direct constraint of category information in the segmentation results, we further propose a novel category loss function, which helps to improve the ability of the segmentation network to correctly identify existing categories in images. As shown in (7), we find the largest probability value from each of the five output probability maps to represent the predicted probability of the category in the image, and then calculate the BCE loss.

$$\text{loss}_{\text{category}} = \frac{1}{N}\sum_{n=1}^{N} -t_n \log\left(\text{Sigmoid}\left(\max_i\left(p_{n,i}\right)\right)\right) - (1-t_n)\log\left(1-\text{Sigmoid}\left(\max_i\left(p_{n,i}\right)\right)\right). \quad (7)$$

Finally, the loss function of the FSP is calculated as:

$$\text{loss}_{\text{FSP}} = \text{loss}_{\text{Joint}} + \lambda_2 \text{loss}_{\text{category}}. \quad (8)$$

In training of the decision fusion layers which give the final segmentation results, the Dice loss is used.

3. Experiment settings

3.1. Datasets and evaluation metrics

The dataset used in this paper are two-dimensional OCT images acquired by Topcon Atlantis DRI-1 swept source OCT scanner (Topcon Corp., Tokyo, Japan) at the First People's Hospital Affiliated to Shanghai Jiao Tong University. The collection and analysis of image data were approved by the Institutional Review Board of the First People's Hospital Affiliated to Shanghai Jiao Tong University, and adhered to the tenets of the Declaration of Helsinki. An informed consent was obtained from each subject. The macula-centered 12-line radial scanning mode was used. The original image size was 1024 × 992 corresponding to 9 mm × 2.6 mm (width × height). The experimental dataset comprised of a total of 1596 OCT B-scans from 133 eyes with high myopia, with 12 OCT B-scans per eye. The ground truth is obtained by manual delineation under the supervision of a senior physician. A total of 972 images from 81 eyes were used as the training set, the validation set included 312 images from 26 eyes, and the test set included the rest 312 images from 26 eyes. The three sets are randomly divided on patient level.

To evaluate the segmentation results, four evaluation indicators are used: Dice coefficient, intersection over union (IoU), sensitivity (Sen), and specificity (Spe). These evaluation indicators are calculated for each type of lesion separately, as in a binary segmentation task, and the average over all types of lesions are also calculated in comparison with other existing methods.

3.2. Implementation details

The experiments were performed on the public platform PyTorch and on a GeForce RTX 3090 GPU with 24GB memory. The three parts of the proposed framework, TSP, FSP, and decision fusion layers were trained separately using the same training settings. The TSP and FSP were trained for 100 epochs, respectively, and the decision fusion layers were trained for 50 epochs. The batch size was 4. Considering the memory cost and training time cost, we resized the images to 256 × 512 before input. The stochastic gradient descent (SGD) algorithm was applied to optimize the network, and the 'poly' learning rate policy was used (Mishra and Sarawadekar 2019). The weights in the loss functions were determined according to performance on the validation set. The best values chosen were $\lambda_1 = 0.7$ and $\lambda_2 = 0.1$.

4. Results

4.1. Ablation experiments

Table 1 shows the results of ablation experiments for TSP, where 'Baseline' refers to the U-shaped structure shown in figure 4(a) without CFGF and DS. Compared to the Baseline, adding the CFGF module at the bottom of the segmentation network results in an improvement of 1.13%, 1.30%, and 1.18% in average Dice, IoU and Sen, respectively. Adding DS to the decoder further improves the average Dice, IoU and Sen by 1.24%, 1.35% and 0.86%, respectively. The Dice, IoU and Sen of each category are also improved in most cases, while Spe is kept high. These results validate the usefulness of the global context and the deep supervision strategy.

Table 2 shows the results of ablation experiments for FSP, where 'Baseline' refers to the U-shaped segmentation network without TCIP, whose encoder is the same as shown in figure 5(a), and the decoder is composed of a U-Net (Ronneberger *et al* 2015) decoder, but each layer has only one 3 × 3 convolution. The baseline was trained with Dice loss only. As shown in table 2, adding the classification network or adding the TCIP module can both improve the segmentation performance. Compared to the baseline, adding both leads to an improvement of 3.48%, 3.68%, and 2.31% in average Dice, IoU and Sen, respectively. By adding the proposed category loss to the total loss function, a further improvement of 1.36%,1.45% in Dice and IoU is achieved. The Dice, IoU and Sen of each category are also improved in most cases, while Spe is kept high.

Table 1. Results of ablation experiment of TSP.

Methods		Dice (%)	IoU (%)	Sen (%)	Spe (%)
Baseline	RS	88.84	80.85	88.23	99.61
	RD	87.90	85.99	97.55	99.60
	Average	88.37	83.42	92.89	99.61
Baseline + CFGF	RS	89.83	82.18	90.91	99.52
	RD	89.17	87.26	97.22	99.79
	Average	89.50	84.72	94.07	**99.66**
Baseline + CFGF + DS	RS	90.23	82.68	92.37	99.44
	RD	91.25	89.46	97.48	99.84
	Average	**90.74**	**86.07**	**94.93**	99.64

Table 2. Results of ablation experiment of FSP.

Methods		Dice (%)	IoU (%)	Sen (%)	Spe (%)
Baseline	ORS	89.36	81.58	92.03	99.59
	MRS	73.92	66.46	78.39	99.93
	IRS	71.40	66.24	76.76	99.94
	RD	80.25	77.88	96.54	99.70
	Average	78.73	73.04	85.93	99.79
Baseline + classification network	ORS	89.67	81.94	92.53	99.54
	MRS	76.24	69.11	77.22	99.95
	IRS	73.77	68.49	81.22	99.94
	RD	86.42	84.52	98.12	99.76
	Average	81.52	76.02	87.27	99.80
Baseline + TCIP	ORS	89.83	82.19	93.27	99.54
	MRS	74.96	67.79	76.36	99.94
	IRS	73.28	67.73	78.87	99.95
	RD	82.99	80.77	96.77	99.78
	Average	80.26	74.62	86.32	99.80
Baseline + TCIP + classification network	ORS	90.15	82.73	92.13	99.58
	MRS	76.73	69.47	82.53	99.92
	IRS	76.32	71.06	80.34	99.95
	RD	85.63	83.63	97.96	99.79
	Average	82.21	76.72	88.24	99.81
Baseline + TCIP + classification network + category loss	ORS	90.35	82.99	93.35	99.54
	MRS	77.49	70.54	81.19	99.93
	IRS	76.49	71.34	81.96	99.93
	RD	89.94	87.81	96.56	99.86
	Average	**83.57**	**78.17**	**88.27**	**99.82**

Table 3 shows the results of ablation experiments for the decision fusion layers, where '3 conv' means using a structure with the middle two convolution layers removed, and '7 conv' means duplicating the middle two convolution layers. Results without inputting the original image to the fusion are also compared. It can be seen that the proposed structure and input obtain the highest average Dice, IoU, Sen and Spe.

In table 4, the results of TSP, FSP, and the whole framework, CMC-Net, are compared. For FSP and CMC-Net, to show the segmentation of total RS, we merge the output regions of ORS, MRS, and IRS as one category. Student's paired t-test between the Dice scores of TSP and CMC-Net, as well as between FSP and CMC-Net is performed, and statistical significance with $p < 0.05$ are indicated. It can be seen that, regarding RS and RD segmentation, the performance of FSP is inferior to TSP, because the complexity of the task increases when the network is trained to distinguish more types of targets. According to the mean Dice values and results of statistical tests, by fusing the results of the TSP and FSP, the CMC-Net achieves comparable RS segmentation performance compared to TSP, and statistically better RD segmentation performance compared to both TSP and FSP. CMC-Net obtains statistically better performance over FSP in the total RS segmentation, and in segmentation of ORS and MRS. The performance on IRS segmentation is comparable. These results demonstrate that the CMC-Net combines results of the two paths in a complementary manner and thus gets an overall superior performance.

Table 3. Results of ablation experiment of decision fusion layers.

Methods		Dice(%)	IoU(%)	Sen(%)	Spe(%)
3 conv	ORS	90.51	83.37	93.57	99.51
	MRS	77.89	71.01	79.87	99.94
	IRS	76.05	70.77	78.73	99.97
	RD	92.47	90.69	97.34	99.85
	Average	84.23	78.96	87.38	**99.82**
7 conv	ORS	90.02	82.59	94.58	99.42
	MRS	78.44	71.52	79.36	99.95
	IRS	76.43	71.15	79.56	99.96
	RD	92.55	90.76	97.77	99.84
	Average	84.36	79.01	87.82	99.79
w/o original image	ORS	90.05	83.29	94.01	99.48
	MRS	78.42	71.55	80.39	99.94
	IRS	76.56	71.27	79.14	99.97
	RD	92.45	90.69	97.40	99.85
	Average	84.37	79.20	87.74	99.81
CMC-Net(5 convwith originalimage)	ORS	91.01	84.08	93.66	99.54
(5 conv with original image)	MRS	78.91	72.05	81.46	99.93
	IRS	76.86	71.67	80.91	99.94
	RD	92.54	90.76	97.54	99.84
	Average	**84.83**	**79.64**	**88.39**	**99.82**

Table 4. Results of ablation experiment of CMC-Net.

Methods		Dice (%)	IoU (%)	Sen (%)	Spe (%)
TSP	RS	90.23	82.68	92.37	99.44
	RD	91.25*	89.46	97.48	99.84
FSP	RS[a]	89.56*	81.63	91.35	99.47
	ORS	90.35*	82.99	93.35	99.54
	MRS	77.49*	70.54	81.19	99.93
	IRS	76.49	71.34	81.96	99.93
	RD	89.94*	87.81	96.56	99.86
CMC-Net	RS[a]	90.23	82.71	91.83	99.48
	ORS	91.01	84.08	93.66	99.54
	MRS	78.91	72.05	81.46	99.93
	IRS	76.86	71.67	80.91	99.94
	RD	92.54	90.76	97.54	99.84

* indicates statistically significant difference with $p < 0.05$, compared with CMC-Net.

[a] Calculated by treating the output ORS, MRS, and IRS regions as one category.

4.2. Comparisons with state-of-the-art

In this section, first we compare the segmentation results of our proposed TSP with some state-of-the-art networks on RD and RS segmentation. Then, we compare the segmentation results of the proposed FSP and CMC-Net with some state-of-the-art networks on RD, ORS, MRS, and IRS segmentation.

The proposed TSP are compared with methods including: PSPNet (Zhao *et al* 2017), DeeplabV3 (Chen *et al* 2017), R2U-Net (Alom *et al* 2018), Attention U-Net (Oktay *et al* 2018), UNet++ (Zhou *et al* 2020), CE-Net (Gu *et al* 2019), CPFNet (Feng *et al* 2020) and HRNet (Wang *et al* 2021). As shown in table 5, the proposed TSP obtains the best segmentation results, and the Dice coefficient of RS and RD reach 90.23% and 91.25%, respectively, and the average Dice coefficient reaches 90.74%.

For the FSP, the multi-task network Y-Net (Mehta *et al* 2018b) is also included for comparison. As shown in table 6, compared with these state-of-the-art networks, the proposed FSP achieves the best results, and the Dice coefficients of ORS, MRS, IRS, and RD reach 90.35%, 77.49%, 76.49%, and 89.94% respectively, and the average Dice coefficient reaches 83.57%. Finally, CMC-Net further improved the segmentation results of ORS, MRS, IRS, and RD with Dice coefficients of 91.01%, 78.91%, 76.86%, and 92.54% respectively, and the final average Dice coefficient reaches 84.83%.

The average test time for TSP, FSP and CMC-Net is also shown in tables 5 and 6 and compared with other existing methods. The average test time is 5.77 ms and 6.25 ms for TSP and FSP, respectively. With the added

Table 5. Comparisons of TSP with state-of-the-art networks.

Methods		Dice (%)	IoU (%)	Sen (%)	Spe (%)	Test time (ms)
PSPNet (Zhao *et al* 2017)	RS	83.24	72.31	84.75	99.25	5.29
	RD	83.59	80.99	96.08	99.69	
	Average	83.42	76.65	90.41	99.47	
DeeplabV3 (Chen *et al* 2017)	RS	74.00	60.10	73.75	99.06	**3.53**
	RD	78.64	75.01	93.90	99.52	
	Average	76.32	67.55	83.83	99.29	
R2U-Net (Alom *et al* 2018)	RS	64.03	51.38	67.80	99.11	9.94
	RD	66.81	62.88	91.44	98.74	
	Average	65.42	57.13	79.62	98.93	
Attention U-Net (Oktay *et al* 2018)	RS	89.69	82.00	90.76	99.55	5.93
	RD	86.23	84.37	98.27	99.61	
	Average	87.96	83.19	94.52	99.58	
UNet++ (Zhou *et al* 2020)	RS	89.50	81.73	90.70	99.51	5.13
	RD	86.40	84.25	97.03	99.64	
	Average	87.95	82.99	93.86	99.57	
CE-Net (Gu *et al* 2019)	RS	86.44	76.80	89.31	99.31	7.37
	RD	88.36	85.57	93.66	99.87	
	Average	87.40	81.18	91.48	99.59	
CPFNet (Feng *et al* 2020)	RS	86.93	77.61	86.54	99.50	3.85
	RD	82.85	80.59	95.56	99.73	
	Average	84.89	79.10	91.05	99.62	
HRNet (Wang *et al* 2021)	RS	85.68	75.68	87.09	99.34	4.81
	RD	84.42	81.86	95.97	99.68	
	Average	85.05	78.77	91.53	99.51	
TSP	RS	90.23	82.68	92.37	99.44	5.77
	RD	91.25	89.46	97.48	99.84	
	Average	**90.74**	**86.07**	**94.93**	**99.64**	

modules and the classification branch, TSP and FSP still have decent processing efficiency. The CMC-Net requires 16.35 ms when TSP, FSP and the decision layers are run sequentially. This time can be further reduced if TSP and FSP are run in parallel. Still, this test time can fulfill the real time requirement of clinical applications.

Figure 6 shows the segmentation results of ORS, MRS, IRS, and RD qualitatively. Compared with other state-of-the-art segmentation networks, the proposed FSP cannot only delineate the pathological regions more accurately, but also assign labels more correctly. For example, for the B-scan in the second column, the proposed FSP correctly determined that there are only two target categories presented. It can be seen from the first and third column that the proposed FSP has good segmentation results for both small and large targets. Figure 6(m) also gives the segmentation results of the TSP for RD and RS, and it can be seen that good segmentation results are achieved for both small and large RS. As shown in figure 6(n), in the fused results of CMC-Net, the results of TSP can make up for some detailed information lost in the results of FSP, making the segmentation areas more complete and accurate.

To further illustrate the results of multi-class segmentation, figure 7 shows some confusion matrices, which are computed on all pixels in the test set (different than the indices in table 6 which are averages on image-level), and are normalized row-wise. The row summary on the right of each matrix shows the total number of correctly and mistakenly labeled pixels. It can be seen that the total number of background pixels are the largest, which correpond to the surrounding retina tissues and non-retinal area, and the total number of RD and ORS pixels are larger than MRS and IRS. The confusion matrices from different methods share similar features. The segmentation of bigger lesions is better than smaller ones. For each lesion, major mistake occurs when they are confused as background. This is caused by the low contrast and blurred boundary of lesion regions, which is more profound for MRS and IRS. Mistakes also occur between lesions that are often adjacent in location, such as between RD and ORS, or between ORS and MRS. Compared with UNet++ and HRNet, the proposed FSP and CMC-Net can better distinguish different type of lesions, and the CMC-Net has the highest accuracy.

5. Discussion and conclusions

Accurate segmentation of RD, ORS, MRS, and IRS in OCT images has great clinical value for diagnosis and treatment of myopic maculopathy. The main challenges for segmentation of RD, ORS, MRS, and IRS in OCT images are unbalanced categories, large variations in size and span of the targets, and the similarity of shape and

Table 6. Comparisons of FSP and CMC-Net with state-of-the-art networks.

Methods		Dice (%)	IoU (%)	Sen (%)	Spe (%)	Tst time (ms)
PSPNet (Zhao *et al* 2017)	ORS	81.27	70.11	83.40	99.37	6.09
	MRS	67.41	59.30	67.02	99.95	
	IRS	63.07	58.22	68.32	99.90	
	RD	84.10	81.13	93.77	99.72	
	Average	73.96	67.19	78.13	99.74	
DeeplabV3 (Chen *et al* 2017)	ORS	73.95	60.67	81.76	98.96	4.83
	MRS	59.04	52.08	57.83	99.95	
	IRS	57.75	54.05	61.64	99.89	
	RD	77.14	73.41	93.37	99.49	
	Average	66.97	60.05	73.65	99.57	
R2U-Net (Alom *et al* 2018)	ORS	72.48	60.78	66.74	99.71	10.09
	MRS	63.80	57.87	64.33	99.96	
	IRS	65.67	61.19	67.15	99.98	
	RD	76.60	73.13	82.25	99.62	
	Average	69.64	63.24	70.12	99.82	
Attention U-Net (Oktay *et al* 2018)	ORS	90.17	83.01	91.33	99.62	6.41
	MRS	70.63	63.49	75.35	99.94	
	IRS	74.10	68.47	85.63	99.94	
	RD	83.19	81.33	98.21	99.54	
	Average	79.52	74.08	87.63	99.76	
UNet++ (Zhou *et al* 2020)	ORS	89.71	82.42	90.05	99.61	5.90
	MRS	73.47	66.66	75.82	99.95	
	IRS	74.74	69.16	80.74	99.95	
	RD	83.13	80.82	96.62	99.53	
	Average	80.26	74.76	85.81	99.76	
CE-Net (Gu *et al* 2019)	ORS	86.21	76.80	89.02	99.47	8.01
	MRS	64.96	56.97	73.45	99.95	
	IRS	59.43	52.65	82.34	99.88	
	RD	81.28	78.48	94.01	99.71	
	Average	72.97	66.23	84.70	99.75	
Y-Net (Mehta *et al* 2018b)	ORS	88.67	80.66	90.15	99.59	**4.79**
	MRS	72.77	65.37	73.90	99.96	
	IRS	72.46	67.02	76.18	99.94	
	RD	85.19	83.07	96.98	99.68	
	Average	79.77	74.03	84.30	99.79	
CPFNet (Feng *et al* 2020)	ORS	87.01	77.88	87.82	99.54	5.03
	MRS	64.61	56.99	62.68	99.98	
	IRS	59.88	53.97	73.62	99.95	
	RD	84.82	82.02	93.72	99.77	
	Average	74.08	67.72	79.46	99.81	
HRNet (Wang *et al* 2021)	ORS	89.83	82.19	93.27	99.54	5.99
	MRS	74.96	67.79	76.36	99.94	
	IRS	73.28	67.73	78.87	99.95	
	RD	82.99	80.77	96.77	99.78	
	Average	80.27	82.19	93.27	99.54	
FSP	ORS	90.35	82.99	93.35	99.54	6.25
	MRS	77.49	70.54	81.19	99.93	
	IRS	76.49	71.34	81.96	99.93	
	RD	89.94	87.81	96.56	99.86	
	Average	83.57	78.17	88.27	99.82	
CMC-Net	ORS	91.01	84.08	93.66	99.54	16.35
	MRS	78.91	72.05	81.46	99.93	
	IRS	76.86	71.67	80.91	99.94	
	RD	92.54	90.76	97.54	99.84	
	Average	**84.83**	**79.64**	**88.39**	**99.82**	

intensity among different types of targets, which will lead to category error in segmentation. In this paper, we propose the novel CMC-Net for automatic segmentation of RD, ORS, MRS, and IRS in OCT images. The CMC-Net consists of three independently trained sub-networks, namely TSP, FSP and decision fusion layers. Although our ultimate goal is to segment the four types of lesions, considering RD/RS segmentation is a simpler

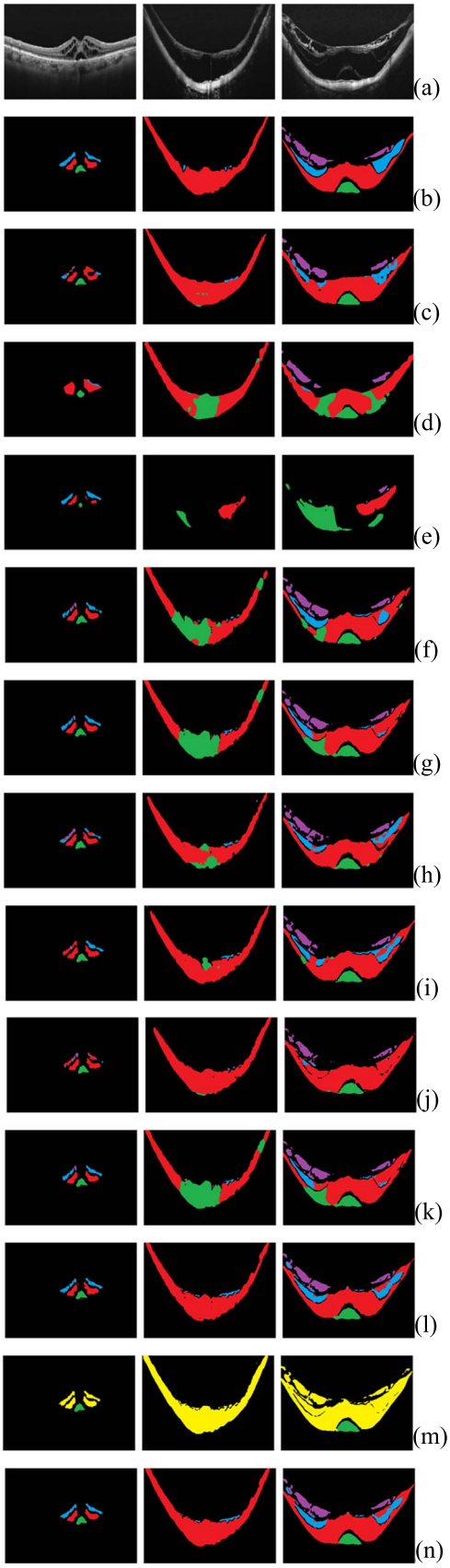

Figure 6. Visualization of segmentation results. (a) the original image (b) ground truth (c) PSPNet (d) DeeplabV3 (e) R2U-Net (f) Attention U-Net (g) UNet++ (h) CE-Net (i) Y-Net (j) CPFNet (k) HRNet (l) the proposed FSP (m) the proposed TSP (n) the proposed CMC-Net. (RD is represented in green, RS is represented in yellow, ORS is represented in red, MRS is represented in blue, and IRS is represented in purple).

Figure 7. Pixel-level confusion matrices calculated on the entire test set. The values in the matrices are normalized row-wise. The row summary on the right of each matrix shows the total number of correctly or mistakenly labeled pixels. (a) UNet++ (b) HRNet (c) FSP (d) CMC-Net.

task, we design both the three-class and FSPs so that by combination, the RD/RS segmentation results act to improve the final segmentation of RD, ORS, MRS, and IRS.

For the TSP, the network adopts a U-shaped structure, where ResNet blocks and dilated convolutions are integrated for better feature extraction. As sometimes RS occupies a large portion of the entire image, a CFGF module is placed at the bottom of the segmentation network to fuse the global features and make the network obtain global receptive field. In order to make the deep layers of the network pay more attention to the segmentation targets, deep supervision modules are further added to each layer of the decoder. Ablation experiments show that both the CFGF and DS modules contribute to the final segmentation performance. Comparative experiments show that the TSP outperforms some state-of-the-art networks in segmenting RD and RS.

For the FSP, the framework adopts a W-shaped structure, consisting of a classification encoder, a segmentation encoder and a decoder. Though each subcategory of RS is smaller in area than the total RS, some of them still have a large spatial span. Therefore, a novel TCIP module is added after the segmentation encoder, which can obtain the long-range contextual information and results in a long cross-shaped receptive field. Facing the challenges of discerning different types of targets, two strategies are proposed in FSP. First, following the idea of multi-task learning, the classification encoder, trained with the classification loss, produces supplementary features that are merged with those from the segmentation encoder. Secondly, the proposed category loss function can constrain the segmentation network to learn more distinguishing features. Ablation experiments show that the TCIP module, the classification branch, and the category loss all contribute to the final segmentation performance. Comparative experiments show that the FSP outperforms some state-of-the-art networks in segmenting RD, ORS, MRS, and IRS. Specifically, figure 6 shows that FSP can label the target regions more correctly while accurately segmenting them.

According to the design and the experimental results of TSP and FSP, the two models fulfill the requirements of the ensemble strategy, which have both 'accuracy' and 'diversity'. This ensures the final outcomes of the proposed CMC-Net to be a complementary combination and to achieve further improvement in performance. As the two paths give different number of labels, it is difficult to fuse the decisions by simple voting or weighting. Therefore some convolutional layers are designed for decision fusion. Note that the original image is also used in the fusion stage to provide more comprehensive information. As shown in table 3 and figure 6, the performance of RD/RS segmentation of TSP is better than that of FSP. The segmented region is more complete and the

boundaries are smoother. The main reasons are the simplicity of the task and the bigger receptive field. Then, after fusion, the final results of CMC-Net obtain improvement over both TSP and FSP, and the ultimate goal of segmenting RD and the three subcategories of RS is achieved.

Both TSP and FSP achieves big receptive field with acceptable model complexity. The number of parameters for TSP is 20.15 M, while the number of parameters for FSP is 15.76 M. The smaller size of FSP is associated with the smaller receptive field, but is more suitable for the five-class segmentation task, because the samples for each category become less. The total number of parameters for the proposed CMC-Net is 36.09 M, which is comparable to many state-of-the-art segmentation networks, such as PSPNet (48.79 M), R2U-Net (39.09 M), Attention U-Net (34.88 M), CE-Net (29.00 M), CPFNet (43.26 M), and HRNet (29.53 M).

The proposed work is an early attempt in automatic quantitative analysis of RD and RS. The CMC-Net achieves values over 90% for all performance indices in RD and ORS segmentation, and therefore is good for detection, localization, and tracking their changes which are needed in clinical diagnosis of pathological myopia. The Dice and IoU for MRS and IRS are lower. This may be due to their blurry boundaries, low contrast and small size. Still, the pixel-level sensitivity is over 80% and specificity is high. This indicates that the method can be used in automatic detection and localization of these early signs of pathological myopia and help clinical grading of the pathology.

In the future, to overcome the problem of incorrect segmentation of adjacent lesions caused by blurry boundaries and further improve the segmentation of MRS and IRS, prior knowledge such as constraints of the retinal layers can be integrated into model design. Other aspects for improvement include the following. The current framework requires separate training procedures for the three parts, which is inefficient. We will try feature fusion strategies which combine information from different segmentation tasks in an earlier stage, and make the segmentation end-to-end. In addition, we will extend the CMC-Net for other multi-class medical image segmentation tasks.

Data availability statement

The data cannot be made publicly available upon publication because they contain sensitive personal information. The data that support the findings of this study are available upon reasonable request from the authors.

Funding

This work was supported by the National Natural Science Foundation of China (62271337), and the National Key Research and Development Program of China (2018YFA0701700, 2019FYC1710204).

Disclosures

The authors declare no conflicts of interest.

References

Alom M Z *et al* 2018 Recurrent residual convolutional neural network based on U-Net (R2U-Net) for medical image segmentation arXiv:1802.06955

Alsaih K, Yusoff M Z, Faye I, Tang T B and Meriaudeau F 2020 Retinal fluid segmentation using ensembled 2-dimensionally and 2.5-dimensionally deep learning networks *IEEE Access* **8** 152452–64

Benhamou N, Massin P, Haouchine B, Erginay A and Gaudric A 2022 Macular retinoschisis in highly myopic eyes *Am. J. Ophthalmol.* **133** 794–800

Causey J, Stubblefield J, Qualls J, Fowler J, Cai L, Walker K, Guan Y and Huang X 2022 An ensemble of U-Net models for kidney tumor segmentation with CT images *IEEE/ACM Trans. Comput. Biol. Bioinf.* **19** 1387–92

Chen L C, Papandreou G, Schroff F and Adam H 2017 Rethinking atrous convolution for semantic image segmentation arXiv:1706.05587v3

Chen L C, Papandreou G, Schroff F and Adam H 2018 Encoder–decoder with atrous separable convolution for semantic image segmentation *Proc. of the European Conf. on Computer Vision (ECCV)* pp 801–18

Feng S, Zhao H, Shi F, Cheng X and Chen X 2020 CPFNet: context pyramid fusion network for medical image segmentation *IEEE Trans. Med. Imaging* **39** 3008–18

Frisina R, Gius I, Palmieri M, Finzi A, Tozzi L and Parolini B 2020 Myopic traction maculopathy: diagnostic and management strategies *Clin. Ophthalmol.* **14** 3699–708

Fu J, Liu J, Tian H, Li Y, Bao Y, Fang Z and Lu H 2019 Dual attention network for scene segmentation *Proc. of the IEEE/CVF Conf. on Computer Vision and Pattern Recognition* pp 3146–54

Fujimoto M, Hangai M, Suda K and Yoshimura N 2010 Features associated with foveal retinal detachment in myopic macular retinoschisis *Am. J. Ophthalmol.* **150** 863–70

Gao K, Niu S, Ji Z, Wu M, Chen Q, Xu R, Yuan S, Fan W and Chen Y 2019 Double-branched and area-constraint fully convolutional networks for automated serous retinal detachment segmentation in SD-OCT images *Comput. Methods Programs Biomed.* **176** 69–80
Golla A K, Bauer D F, Schmidt R, Russ T, Nörenberg D, Chung K, Tönnes C, Schad L R and Zöllner F G 2021 Convolutional neural network ensemble segmentation with ratio-based sampling for the arteries and veins in abdominal CT scans *IEEE Trans. Biomed. Eng.* **68** 1518–26
Gu Z, Cheng J, Fu H, Zhao K, Hao H, Zhao Y, Zhang T, Gao S and Liu J 2019 CE-Net: context encoder network for 2D medical image segmentation *IEEE Trans. Med. Imaging* **38** 2281–92
He K, Zhang X, Ren S and Sun J 2016 Deep residual learning for image recognition *Proc. of the IEEE Conf. on Computer Vision and Pattern Recognition* pp 770–8
Hu J, Chen Y and Yi Z 2019 Automated segmentation of macular edema in OCT using deep neural networks *Med. Image Anal.* **55** 216–27
Kawakami R, Yoshihashi R, Fukuda S, You S and Naemura T 2019 Cross-connected networks for multi-task learning of detection and segmentation *2019 IEEE Int. Conf. on Image Processing* pp 3636–40
Lai T 2007 Retinal complications of high myopia *Med. Bull.* **12** 18–20
Mehta S, Mercan E, Bartlett J, Weave D, Elmore J G and Shapiro L 2018b Y-Net: joint segmentation and classification for diagnosis of breast biopsy images *Int. Conf. on Medical Image Computing and Computer-assisted Intervention* pp 893–901
Mehta S, Rastegari M, Caspi A, Shapiro L and Hajishirzi H 2018a ESPnet: efficient spatial pyramid of dilated convolutions for semantic segmentation *Proc. of the European Conf. on Computer Vision (ECCV)* pp 552–68
Milletari F, Navab N and Ahmadi S A 2016 V-net: fully convolutional neural networks for volumetric medical image segmentation *2016 IEEE Fourth Int. Conf. on 3D Vision (3DV)* pp 565–71
Mishra P and Sarawadekar K 2019 Polynomial learning rate policy with warm restart for deep neural network *2019 IEEE Region 10 Conf. (TENCON)* pp 2087–92
Mou L *et al* 2021 CS2-Net: deep learning segmentation of curvilinear structures in medical imaging *Med. Image Anal.* **67** 101874
Oktay O, Schlemper J, Folgoc L L, Lee M, Heinrich M, Misawa K, Mori K, Mcdonagh S, Hammerla N Y and Kainz B 2018 Attention U-Net: learning where to look for the pancreas arXiv:1804.03999
Ridnik T, Ben-Baruch E, Zamir N, Noy A, Friedman I, Protter M and Zelnik-Manor L 2021 Asymmetric loss for multi-label classification *2021 IEEE/CVF Int. Conf. on Computer Vision (ICCV)* pp 82–91
Ronneberger O, Fischer P and Brox T 2015 U-net: convolutional networks for biomedical image segmentation *Int. Conf. on Medical Image Computing and Computer-assisted Intervention* pp 234–41
Roy A G, Conjeti S, Karri S P K, Sheet D, Katouzian A, Wachinger C and Navab N 2017 ReLayNet: retinal layer and fluid segmentation of macular optical coherence tomography using fully convolutional networks *Biomed. Opt. Express* **8** 3627–42
Takano M 1999 Foveal retinoschisis and retinal detachment in severely myopic eyes with posterior staphyloma *Am. J. Ophthalmol.* **128** 472–6
Wang J, Sun K, Cheng T, Jiang B and Xiao B 2021 Deep high-resolution representation learning for visual recognition *IEEE Trans. Pattern Anal. Mach. Intell.* **43** 3349–64
Wang X, Girshick R, Gupta A and He K 2018 Non-local neural networks *Proc. of the IEEE Conf. on Computer Vision and Pattern Recognition* pp 7794–803
Xu Q, Zeng Y, Tang W, Peng W, Xia T, Li Z, Teng F, Li W and Guo J 2020 Multi-task joint learning model for segmenting and classifying tongue images using a deep neural network *IEEE J. Biomed. Health Inform.* **24** 2481–9
Yang C, Chen X, Su J, Zhu W, Chen Q, Yu J, Fan Y and Shi F 2021 Segmentation of retinal detachment and retinoschisis in OCT images based on improved U-shaped network with cross-fusion global feature module *Proc. SPIE 11596, Medical Imaging : Image Processing*1159621
Yang M, Yu K, Chi Z, Li Z and Yang K 2018 DenseASPP for semantic segmentation in street scenes *Proc. of the IEEE Conf. on Computer Vision and Pattern Recognition* pp 3684–92
Ye L, Zhu W, Bao D, Feng S and Chen X 2020 Macular hole and cystoid macular edema joint segmentation by two-stage network and entropy minimization *Int. Conf. on Medical Image Computing and Computer-assisted Intervention* pp 735–44
Zhang H, Dana K, Shi J, Zhang Z, Wang X, Tyagi A and Grawal A A 2018 Context encoding for semantic segmentation *Proc. of the IEEE Conf. on Computer Vision and Pattern Recognition* pp 7151–60
Zhang Y, Li H, Du J, Qin J and Lei B 2021 3D Multi-attention guided multi-task learning network for automatic gastric tumor segmentation and lymph node classification *IEEE Trans. Med. Imaging* **40** 1618–31
Zhao H, Shi J, Qi X, Wang X and Jia J 2017 Pyramid scene parsing network *IEEE Conf. on Computer Vision and Pattern Recognition* pp 2881–90
Zhou Y, Chen H, Li Y, Liu Q and Yap P T 2021 Multi-task learning for segmentation and classification of tumors in 3D automated breast ultrasound images *Med. Image Anal.* **70** 101918
Zhou Z, Siddiquee M M R, Tajbakhsh N and Liang J 2020 UNet++: redesigning skip connections to exploit multiscale features in image segmentation *IEEE Trans. Med. Imaging* **39** 1856–67
Zhu Z, Xu M, Bai S, Huang T and Bai X 2019 Asymmetric non-local neural networks for semantic segmentation *Proc. of the IEEE/CVF Int. Conf. on Computer Vision* pp 593–602