Biomedical Signal Processing and Control 84 (2023) 104810

Contents lists available at ScienceDirect

Biomedical Signal Processing and Control

journal homepage: www.elsevier.com/locate/bspc

Classification and segmentation of OCT images for age-related macular degeneration based on dual guidance networks

Shengyong Diao [a,1], Jinzhu Su [a,1], Changqing Yang [a], Weifang Zhu [a], Dehui Xiang [a], Xinjian Chen [a,b], Qing Peng [c,*], Fei Shi [a,*]

[a] MIPAV Lab, School of Electronic and Information Engineering, Soochow University, Suzhou 215006, China
[b] The State Key Laboratory of Radiation Medicine and Protection, Soochow University, Suzhou 215123, China
[c] Department of Ophthalmology, Shanghai Tenth People's Hospital, Tongji University School of Medicine, Shanghai 200072, China

ARTICLE INFO

Keywords:
Age-related macular degeneration
Retinal OCT image
Convolutional neural network
Class activation map
Image classification
Image segmentation

ABSTRACT

Age-related macular degeneration (AMD) is one of the main causes of visual impairment in elderly people, with drusen and choroidal neovascularization (CNV) being two characterizing types of lesions. Based on optical coherence tomography (OCT), image classification can be used in AMD diagnosis, while image segmentation is necessary for quantitative assessment of the lesion area. In this paper, we propose a deep learning framework exploiting dual guidance between the two tasks. Firstly, a complementary mask guided convolutional neural network (CM-CNN) is proposed to perform classification of OCT B-scans with drusen or CNV from normal ones, where the guiding mask is generated by the auxiliary segmentation task. Secondly, a class activation map guided UNet (CAM-UNet) is proposed to achieve segmentation of drusen and CNV lesions, using CAM output from the CM-CNN. Tested on a subset of the public UCSD dataset, and compared with five classification networks, four segmentation networks, and three multi-task networks, the proposed dual guidance network has achieved higher accuracy both in classification and segmentation. The classification accuracy reaches 96.93% and the Dice coefficient for segmentation reaches 77.51%. Results on an extra dataset for detection of macular edema and segmentation of retinal fluids further show the generalizability of the proposed model.

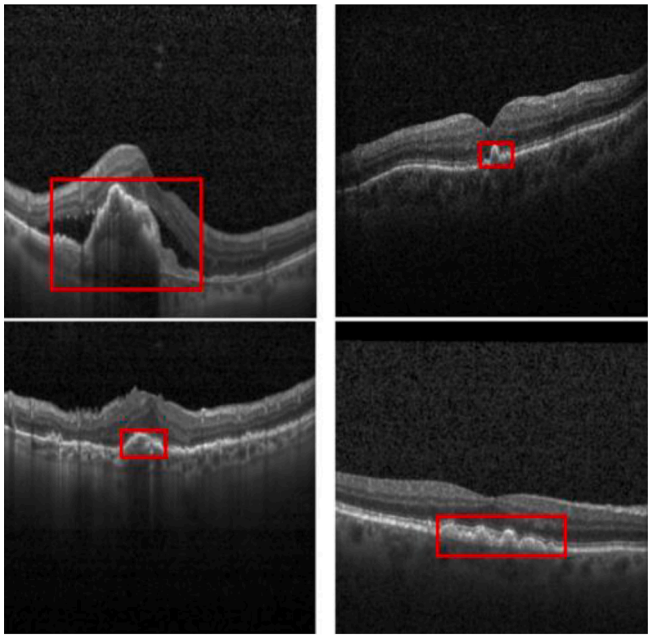
1. Introduction

Age-related macular degeneration (AMD) is a degenerative disease of the retina, and a leading cause of vision loss worldwide [1]. According to the pathogenesis and clinical manifestations of the disease, AMD can be divided into dry or wet AMD. Drusen is a typical manifestation of dry AMD, which is formed by the accumulation of metabolites between Bruch's membrane and the retinal pigment epithelium (RPE), causing focal elevations of the RPE [2]. Wet AMD, characterized by choroidal neovascularization (CNV), usually develops rapidly and causes more severe visual impairment than dry AMD [1]. CNV refers to the proliferating blood vessels growing from the choroid, passing through Bruch's membrane or even the RPE and resulting in subretinal or intraretinal fluid and hemorrhage [3]. In recent years, optical coherence tomography (OCT) has been widely used in diagnosis and treatment of AMD. OCT has the advantages of high resolution and non-invasiveness, and it can give sectional view of the retina and the lesions, which is important for observing and tracking of AMD [4]. Fig. 1 shows some OCT images with drusen or CNV. As the neovascularization complex is often obscured by hemorrhage and fluid exudates, or even hidden in the choroid [4], the CNV lesion hereafter refers to the hyperreflective or partly hyperreflective lesions underneath the uplifted RPE. In this work, we establish an automatic framework for both classification of OCT images with drusen, CNV or normal retina, and segmentation of the lesions. The classification helps for rapid screening and diagnosis of AMD, while the segmentation allows quantitative analysis necessary for disease tracking and treatment planning.

In recent years, automatic analysis of retinal pathologies from OCT images have attracted a lot of attention. Fig. 2 shows the number of papers year-wise from PubMed database [5] that reported works on classification or segmentation of retinal pathologies from OCT images, showing a general increasing trend, especially for deep learning based methods. Among these, various methods have been proposed for single tasks in AMD analysis. Regarding segmentation of drusen, Farsiu et al.

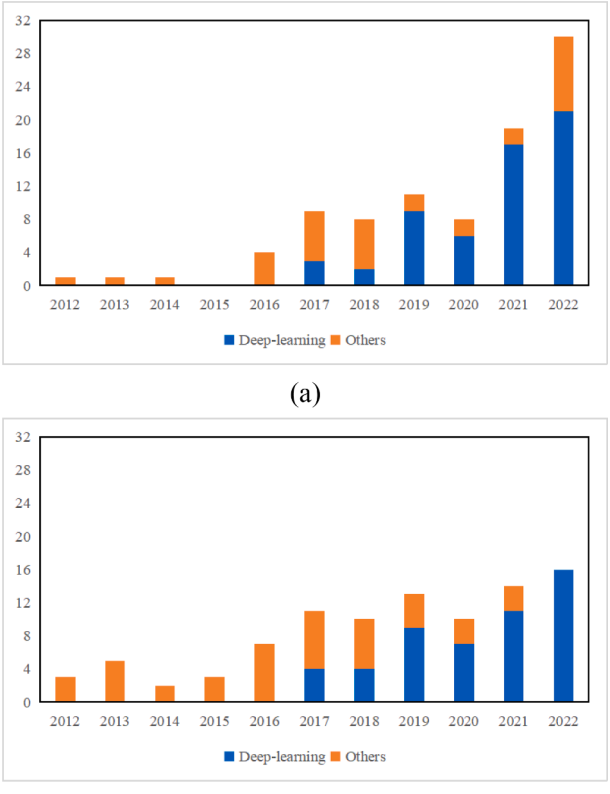
* Corresponding authors.
E-mail addresses: pengqing@tongji.edu.cn (Q. Peng), shifei@suda.edu.cn (F. Shi).
[1] Contributed equally.

https://doi.org/10.1016/j.bspc.2023.104810
Received 14 July 2022; Received in revised form 29 January 2023; Accepted 5 March 2023
Available online 11 March 2023



(a) (b)

Fig. 1. OCT images with (a) CNV or (b) drusen. (Red rectangles indicate the lesions.). (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)



(a)

(b)

Fig. 2. The number of papers from the PubMed database that reported works on (a) classification or (b) segmentation of retinal pathologies from OCT images.

[6], Yi et al. [7] and Chen et al. [8] solved it combining thresholding and quadratic curve fitting methods. These traditional methods involved many preprocessing steps and their generalization performance was poor. More recently, machine learning, especially deep learning methods have been applied. Oliveira et al. [9] proposed to use random forest with texture features. Shekoufeh et al. [10] proposed to use UNet for layer segmentation, followed by shortest path searching and polynomial fitting for drusen segmentation. Asgari et al. [11] used a multi-decoder network to achieve automatic segmentation of drusen. Wang et al. [12] proposed a novel multi-scale transformer global attention network combined with semi-supervised learning. With regard to segmentation of CNV, Li et al. [13] utilized the three-dimensional directional gradient histogram feature to train random forests. Xi et al. [14] proposed a U-shaped network with attention enhancement block and informative loss to segment CNV. Zhang et al. [15] used a multi-scale parallel branch convolutional neural network. Meng et al. [16] proposed a multi-scale information fusion network. Some other studies involved automatic classification of AMD. Wang et al. [17] made use of two identical feature extractors to extract features from fundus and OCT images, and fused them for classification of dry, wet AMD and normal eyes. Kermany et al. [18] used the InceptionV3 network pretrained on ImageNet to achieve classification of diabetic macular edema, drusen, CNV and normal OCT images. For the same task, Fang et al. [19] used a rough lesion segmentation map to guide the classification network, and Kamran et al. [20] proposed the Optic-Net which exploited improved residual learning element. More recently, Thomas et al. [21] and Sotoudeh-Paima et al. [22] proposed deep learning networks with multiscale structures for disease classification. Ma et al. [23] proposed a hybrid network utilizing both convolution and transformer structures.

Despite of the respective efforts in pathology classification and segmentation, in the clinic scenario, it is preferable and more efficient to perform both tasks in one unified framework. In recent years, multi-task learning algorithms have been developed to complete multiple closely-related but different tasks at the same time, which attempts to autonomously learn feature information that promotes one task from the other tasks. For example, in the Y-Net proposed by Mehta et al. [24], the classification and the segmentation tasks share the same encoder. Misra et al. [25] used the cross-stitch unit between two parallel networks to learn effective features in the other branch. Kawakami et al. [26] proposed to connect single-task networks through convolutional layers to transfer useful information for the counterpart. For multi-task learning, further research is needed in the design of the network structure and the trade-off of the loss function for different tasks.

In this work, we propose a new framework of dual guidance networks for the close-related tasks of AMD-associated OCT image classification and segmentation. Firstly, a complementary mask guided convolutional neural network (CM-CNN) is proposed to perform the classification of OCT images with drusen, CNV or normal retina. In CM-CNN, the auxiliary task of lesion segmentation is introduced. The resulting segmentation mask is used in a complementary form to guide the extraction of classification features, so as to improve the classification performance. Secondly, a class activation map guided UNet (CAM-UNet) is proposed to achieve automatic segmentation of AMD lesions, where the CAM from classification is fused into the features at each layer to guide the segmentation task.

2. Method

2.1. Overview of the dual guidance networks

In the proposed framework, two networks are trained sequentially for classification and segmentation, and the information in the other task is innovatively utilized to guide the learning of the current task. Specifically, dual guidance means introducing the segmentation mask to guide the classification network and then introducing the class activation map to guide the segmentation network.

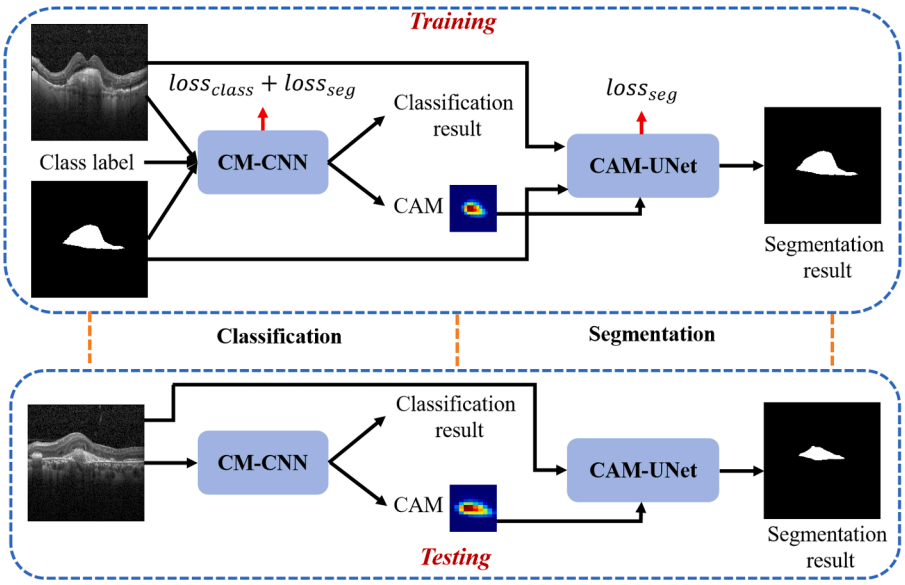


Fig. 3. The overall flowchart of dual guidance networks.

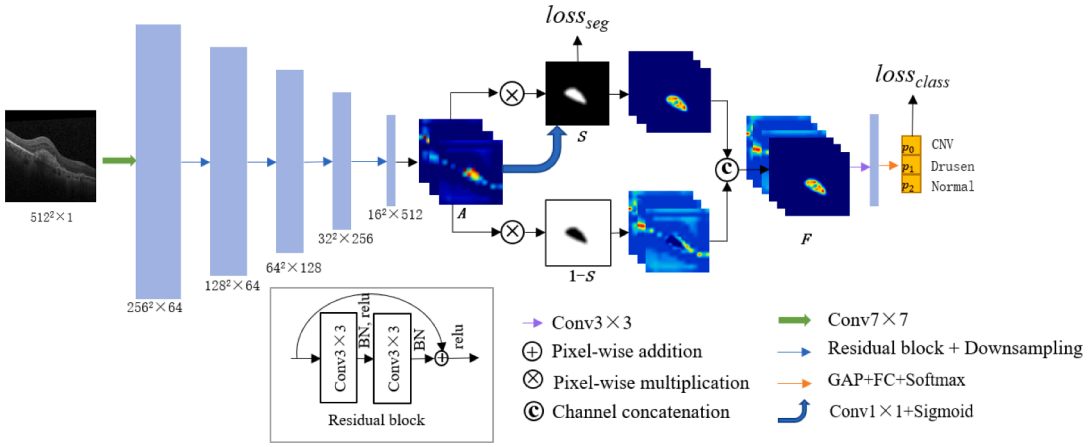


Fig. 4. The network structure of CM-CNN.

The overall flowchart of the dual guidance networks is shown in Fig. 3. In the classification stage, the CM-CNN guided by complementary segmentation masks achieves automatic classification of the input image. Then, the Grad-CAM algorithm is used to obtain the class activation map of the input image in CM-CNN. In the segmentation stage, the original input image and the class activation map are simultaneously fed into the CAM-UNet to obtain the segmentation results of AMD lesions.

2.2. Complementary mask-guided convolutional neural network

The network structure of the proposed CM-CNN is shown in Fig. 4. It adopts ResNet18 [27] as the main feature extractor, as the residual blocks can effectively extract features and prevent the gradient vanishing problem in the training process. The auxiliary task of segmentation mask is introduced at the bottom. Specifically, the input image is first passed through a convolutional layer with a kernel size of 7×7 to extract features with a large receptive field, and then the obtained features are sequentially passed through four basic layers consisting of

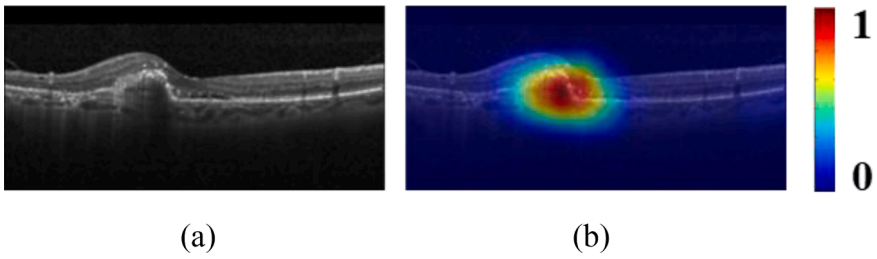


Fig. 5. An example of class activation map (a) raw image (b) class activation map overlaid on the raw image. Colors closer to red indicate higher importance in making the classification prediction. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

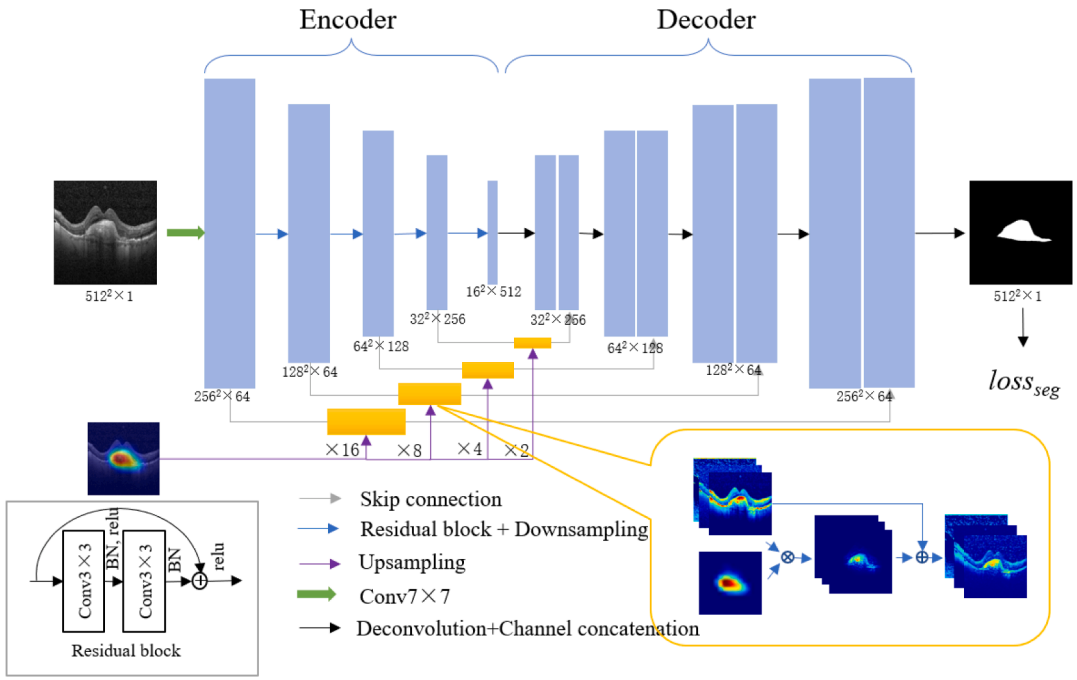


Fig. 6. The network structure of CAM-UNet.

residual blocks and downsampling operations. Each layer contains two residual blocks. Each residual block has two 3 × 3 convolution kernels, followed by Batch Normalization or Relu, as shown in Fig. 4. The output feature *A* passes through the 1 × 1 convolution and Sigmoid function to get the segmentation mask *S,* from which a segmentation loss is calculated. *A* is then multiplied by *S* and 1-*S* respectively to obtain the complementarily enhanced features. These two features are concatenated and then further processed by the convolution layer to obtain the complementary mask-guided feature *F*. Finally, *F* is sent to the fully connected layer to produce the predicted probability for each class, from which a classification loss is calculated.

Note that, multiplying *A* with the coarse lesion mask *S* emphasizes the characteristics of the lesion area and intuitively helps classification. However, when there is mis-segmentation or when there is no lesion area in the normal image, some important information in the original features can be lost. Therefore, we design another branch where $A\times(1\text{-}S)$ is retained. This not only preserves the context information outside the mask, but also increases the nonlinearity of the network. Putting the complementary features into separate channels allow the network to deal with them differently.

2.3. Class activation map of CM-CNN

Grad-CAM [28] is an algorithm that offers visual explanations of the classification decisions made by CNN, and the generated CAM indicates the importance of different regions in the input image for classification prediction. Fig. 5 shows an example of the original input image and the CAM that predicts the input image as choroidal neovascularization. As can be seen, the CAM can roughly give the location of the lesion area in the OCT image, so it is adopted to guide the segmentation of the lesion area in the proposed framework. More examples of CAM can be found in Section 4.

2.4. Class activation map-guided U-shaped segmentation network

The network structure of CAM-UNet is shown in Fig. 6. It adopts a symmetrical U-shaped structure. The encoder uses ResNet18 [27] with five layers, which is the same feature extractor as in CM-CNN, described in Section 2.2. The residual blocks are again used to improve feature extraction and facilitate training. The decoder restores the output to the size of the input image. It has four layers. Each layer upsamples features through a 2 × 2 transposed convolution. The encoder and decoder are connected by skip connections to fuse spatial and semantic information, avoiding information loss caused by downsampling in the encoder [29].

As shown in Fig. 6, the CAM corresponding to the input image is fused with the encoded features at different resolutions by multiplication and addition, in order to enhance the features of the region of interest while still keeping the original features. Since the size of the CAM obtained from CM-CNN is 16 × 16, different scales of upsampling are employed to match the size of the feature maps. By integrating the CAM into the skip connections at all levels, both the spatial information in the shallow layers and the semantic information in the deep layers can be guided by the CAM from the classification network.

2.5. Loss functions

The classification loss function is the cross entropy, expressed as.

$$loss_{class} = -\sum_{c=1}^{N} y_c log(p_c) \tag{1}$$

where *C* indicates the category, y_c represents the category label,p_c is the predicted probability, and *N* represents the number of categories.

The segmentation loss function is defined as the sum of binary cross entropy (BCE) and Dice loss. The Dice loss can make the network pay more attention to small objects. The loss functions are as follows:

$$loss_{BCE} = -\frac{1}{M}\sum_{i=1}^{M}[t_i lnq_i + (1-t_i)ln(1-q_i)] \tag{2}$$

$$loss_{Dice} = 1 - \frac{1}{M}\sum_{i=1}^{M}\frac{2q_i t_i}{q_i^2 + t_i^2} \tag{3}$$

$$loss_{seg} = loss_{BCE} + loss_{Dice} \tag{4}$$

where t_i denotes the value of the *i*-th pixel in the segmentation label map, q_i represents the value of the *i*-th pixel in the prediction result, and *M* indicates the total number of pixels in the image.

As shown in Fig. 3, As CM-CNN contains two sub-tasks of

Table 1
The number of different types of OCT images in the datasets.

	Dataset I			Dataset II	
	CNV	drusen	normal	macular edema	normal
Training set	1422	872	700	456	100
Validation set	100	100	100	124	100
Testing set	250	250	250	100	100

classification and auxiliary segmentation, its overall loss function adopts the sum of classification loss and segmentation loss. For CAM-UNet only the segmentation loss function is used.

3. Experiment settings

3.1. Datasets

Dataset I is a subset of the publicly available UCSD dataset [18]. All images were acquired by Heidelberg Spectralis OCT scanner (Heidelberg Engineering, Germany) from multiple centers. The original dataset included OCT B-scans of four categories, but in this study we only used images labeled as normal, drusen or CNV. The images in the original dataset only contained category labels, and the pixel-level ground truth of drusen and CNV lesions used in our study was labeled in-house under the guidance of an experienced clinician. As manual annotation was laborious, only a subset of the UCSD training set was labeled, which were divided into the training and validation set for our study. The entire UCSD testing set was labeled and used as our test set. The training set, validation set and test set were completely independent.

To further show the generalizability of the proposed model, we built Dataset II from two public datasets. It included OCT scans with macular edema (caused by AMD or retinal vein occlusion), and from healthy subjects. The pathological OCT scans were from the RETOUCH dataset [30], including B-scans randomly chosen from 24 eyes. Pixel-level annotations were provided. OCT scans from normal eyes were from the Srinivasan2014 dataset [31], including B-scans randomly chosen from 15 eyes. All OCT images were obtained from the same type of scanner (Heidelberg Spectralis), so that the classifiers wouldn't be distracted by the difference in image qualities. For this dataset, the classification task was to label each B-scan as normal or pathological, and the segmentation task was to segment all fluid regions. The images were split into training, validation and testing sets on eye level.

The detailed number of OCT images in the two datasets are shown in Table 1. All the OCT images used in this paper has undergone preprocessing steps before model training and testing, including filling the white edges of the image, image resizing to 512×512, and grey level normalization to [0,1].

3.2. Evaluation metrics

The evaluation indicators for the classification of drusen, CNV and normal OCT images includes area under the ROC curve (AUC), single-class accuracy (sAcc), Sensitivity (Sen) and Specificity (Spe), calculated for each category, and the overall accuracy (Acc) and Kappa coefficient [32] for multi-class classification. The evaluation indicators for the classification of macular edema and normal OCT images includes AUC, Sen and Spe, and Kappa coefficient for two-class classification.

The performance of the segmentation tasks is measured by five metrics [12,15]: Dice similarity coefficient (Dice), Intersection over Union (IoU), Sensitivity (Sen), Specificity (Spe), and pixel-wise accuracy (pAcc).

3.3. Implementation details

As shown in Fig. 3, for training of CM-CNN, the OCT image, the class label and the segmentation mask were needed. The sum of classification and segmentation losses were used to optimize the model. After the CM-CNN was trained, a CAM was calculated for each image. Then, for training of CAM-UNet, all pathological images with corresponding CAMs and masks were used. The segmentation loss was used to optimize the model. The testing followed similar procedures, except that no ground truth labels and masks were needed.

The implementation of the proposed framework was based on the public platform PyTorch and on a NVIDIA GeForce RTX 2080Ti graphics card with 11 GB of video memory. In the training process, the Adam optimizer with a learning rate of 0.001 was used to optimize the parameters in the network. The batchsize was 8, and the training was performed for 50 epochs for both CM-CNN and CAM-UNet. Random horizontal flipping was used for data augmentation. The model parameters with the highest classification or segmentation performance on the validation set were saved for testing. All hyperparameter settings in the comparative experiments were kept the same.

4. Experimental results

4.1. Comparison with other classification networks on Dataset I

We compare the performance of CM-CNN in the dual guidance

Table 2
Comparison with classification networks on Dataset I.

	Class	AUC	Sen	Spe	sAcc	Acc	Kappa	#Param (M)	Training time (h)	Test time (ms)
VGG16	CNV	0.9930	0.9960	0.9040	0.9347	0.9253	0.8880	14.93	0.78	18.40
[33]	Drusen	0.9790	0.7800	1	0.9267					
	Normal	0.9988	1	0.9840	0.9893					
MobileNet	CNV	0.9928	**1**	0.9340	0.9560	0.9373	0.9060	**12.26**	1.25	**18.13**
[34]	Drusen	0.9835	0.8160	1	0.9387					
	Normal	0.9979	0.9960	0.9720	0.9800					
SENet	CNV	0.9927	1	0.9500	0.9667	0.9467	0.9200	115.21	**0.47**	28.20
[35]	Drusen	0.9865	0.8480	0.9960	0.9467					
	Normal	0.9960	0.9930	0.9740	0.9800					
InceptionV3	CNV	0.9955	1	0.9540	0.9693	0.9520	0.9280	83.13	1.27	34.67
[18]	Drusen	0.9854	0.8600	0.9980	0.9520					
	Normal	0.9984	0.9960	0.9760	0.9827					
OpticNet	CNV	0.9951	0.9960	0.9420	0.9600	0.9400	0.9100	11.91	1.56	53.80
[20]	Drusen	0.9812	0.8240	**1**	0.9413					
	Normal	0.9995	1	0.9680	0.9787					
CM-CNN	CNV	**0.9988**	0.9960	**0.9680**	**0.9773**			60.63	0.53	18.33
	Drusen	**0.9874**	**0.9120**	0.9980	**0.9693**	**0.9693**	**0.9540**			
	Normal	**0.9999**	**1**	**0.9880**	**0.9920**					

Table 3
Comparison with segmentation networks on Dataset I.

	CAM	Dice	IoU	Sen	Spe	pAcc	#Param (M)	Training time (h)	Testing time (ms)
PSPNet [36]	×	0.6318	0.5128	0.6052	0.9957	0.9899	77.18	2.89	24.00
	√	0.7311	0.6032	0.7794	0.9932	0.9904			
Attention-UNet[37]	×	0.7543	0.6447	0.7373	0.9962	0.9913	33.29	3.87	25.93
	√	0.7644	0.6423	0.7578	0.9961	0.9918			
UNet [29]	×	0.7575	0.6439	0.7620	0.9953	0.9910	**29.62**	2.08	**12.33**
	√	0.7662	0.6382	**0.8758**	0.9928	0.9904			
CENet [38]	×	0.7661	0.6541	0.7618	0.9961	0.9919	110.07	6.12	20.27
	√	0.7680	0.6421	0.8713	0.9928	0.9906			
CAM-UNet	√	**0.7751**	**0.6638**	0.7610	**0.9963**	**0.9923**	60.79	**1.37**	14.60

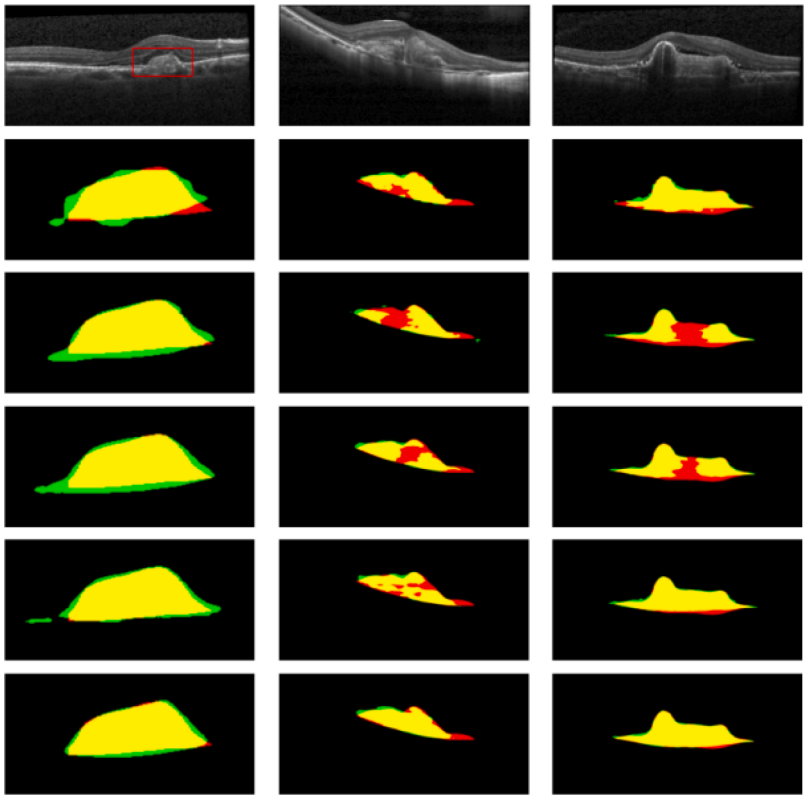


Fig. 7. Segmentation results of CNV using different algorithms (a) original image (b) PSPNet (c) Attention-UNet (d) UNet (e) CENet (f) CAM-UNet. In the first column, the segmentation maps correspond to the red rectangle in the original image. Yellow: true positive, Red: false negative, Green: false positive. Comparative results were obtained without CAM input to the networks. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

framework with some existing classification networks, including the traditional VGG16 [33], the lightweight MobileNet [34], the SENet [35] with the channel attention mechanism, the InceptionV3 used in the original paper of the UCSD dataset [18], and the OpticNet [20] with improved residual interpolation module, also proposed for OCT classification task. The experimental results are shown in Table 2. The overall classification accuracy reaches 96.93%, which is 1.73% higher than that of the second-ranked InceptionV3, and the Kappa value reaches 95.40%, which is 2.6% higher than that of InceptionV3. It also achieved the highest value in most of the single-class performance indicators.

The model complexity, training and testing time are also listed in Table 2. The number of parameters of the proposed model is moderate comparing to other networks. The training time is short, only slightly longer than that of SENet. The testing time is short too, only slightly longer than that of MobileNet.

4.2. Comparison with segmentation networks on Dataset I

We also compare the performance of CAM-UNet with some existing segmentation networks, include PSPNet [36], Attention-UNet [37], UNet [29], and CENet [38]. The latter three were also proposed for medical image segmentation tasks. Table 3 gives the experimental results. For the four established models, results were obtained with only the OCT image as input, or with the OCT image concatenated with the corresponding upsampled CAM from CM-CNN as input, indicated by the cross or tick in the second column. Based on Dice as the main index, it can be seen that adding the CAM to the input can improve the segmentation performance for all models. This proves that the CAM from the proposed CM-CNN indeed contains important information of pathological regions. Comparing the proposed CAM-UNet with all other models, it achieves the highest Dice coefficient of 77.51%. The IoU, Spe, and pAcc are also the highest among all networks compared. Though some models have high sensitivity, their lower specificity and IoU indicate over-segmentation with a lot of false positives.

As also shown in Table 3, regarding model complexity and efficiency, the number of parameters of the proposed model is moderate comparing to other networks. The training time is the shortest, and the testing time is shorter than other models except UNet.

Figs. 7 and 8 show the segmentation results of CNV and drusen lesion regions by different segmentation algorithms, respectively. From Fig. 7, it can be seen that the CNV region may have blurred boundaries or inhomogeneous intensity, and from Fig. 8, it can be seen that the drusen

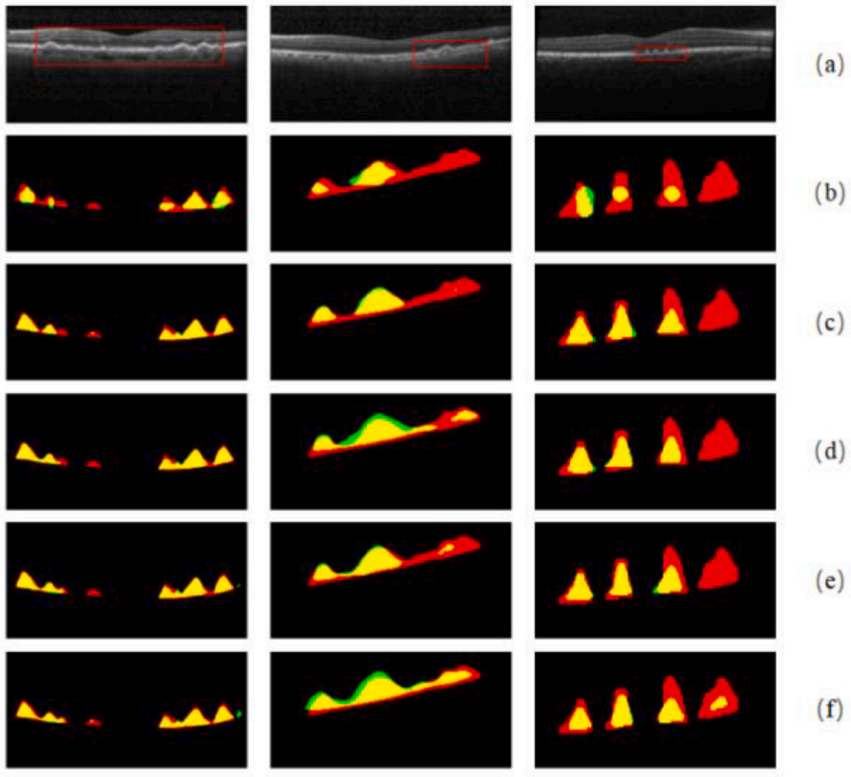


Fig. 8. Segmentation results of drusen using different algorithms (a) original image (b) PSPNet (c) Attention-UNet (d) UNet (e) CENet (f) CAM-UNet. The segmentation maps correspond to the red rectangles in the original image. Yellow: true positive, Red: false negative, Green: false positive. Comparative results were obtained without CAM input to the networks. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

Table 4
Comparison with multi-task networks on Dataset I.

	Acc	Kappa	Dice	IoU	#Param (M)	Trainingtime (h)	Testtime (ms)
Y-Net [24]	0.9213	0.8820	0.7362	0.6240	**54.61**	**1.02**	**13.33**
Cross-stitch [25]	0.9467	0.9200	0.7540	0.6400	97.21	1.63	16.60
Cross-connected [26]	0.9306	0.8960	0.7425	0.6349	103.40	1.72	15.33
CM-CMM + CAM-UNet	**0.9693**	**0.9540**	**0.7751**	**0.6638**	121.42	1.90	38.86

Table 5
The ablation experiments of CM-CNN.

	Class	AUC	Sen	Spe	sAcc	Acc	Kappa
ResNet18	CNV	0.9966	0.9920	0.9600	0.9770	0.9453	0.9180
	Drusen	0.9850	0.8520	0.9920	0.9453		
	Normal	0.9980	0.9920	0.9660	0.9747		
ResNet18(mask)	CNV	**0.9989**	0.9960	**0.9740**	**0.9813**	0.9560	0.9339
	Drusen	0.9851	0.8720	0.9980	0.9560		
	Normal	0.9986	1	0.9620	0.9747		
CM-CNN	CNV	0.9988	**0.9960**	0.9680	0.9773	**0.9693**	**0.9540**
	Drusen	**0.9874**	**0.9120**	**0.9980**	**0.9693**		
	Normal	**0.9999**	**1**	**0.9880**	**0.9920**		

regions are small, some separated and some adhered to each other. For both CNV and drusen, the proposed CAM-UNet can segment the lesion area more accurately than other methods.

4.3. Comparison with multi-task networks on Dataset I

We also compare our method with some existing multi-task deep learning networks including Y-Net[19], cross-stitch network [20] and cross-connected network [21] on both the classification and segmentation performance. As shown in Table 4, the proposed CM-CMM and CAM-UNet achieve the best classification and segmentation performance compared with these multi-task networks. However, the total number of parameters are larger, and both training and testing time are longer.

4.4. Ablation experiments on Dataset I

In this subsection, to show the effectiveness of the proposed

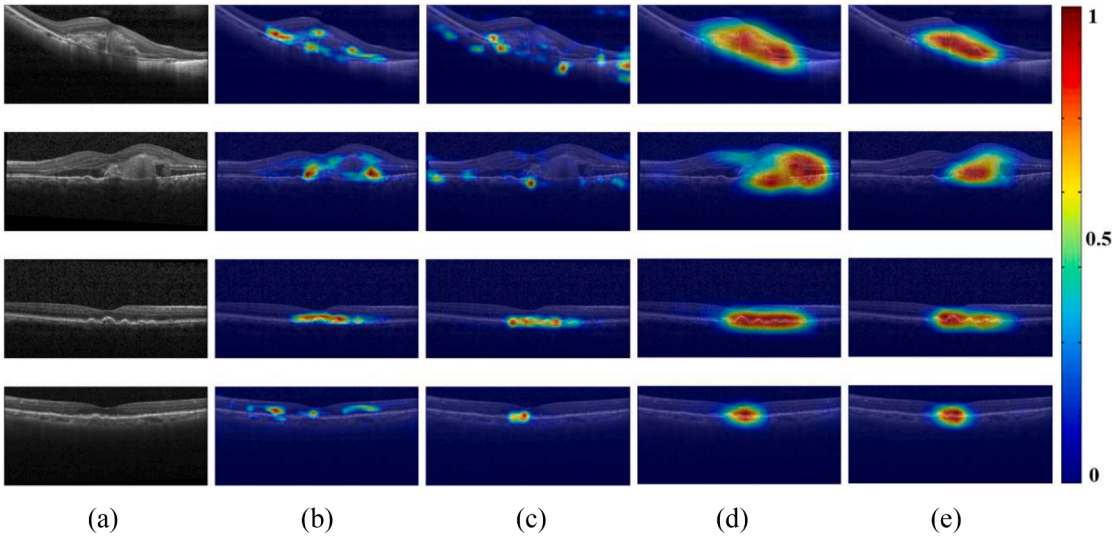


Fig. 9. Class activation maps of different classification networks (a) Original Image (b) VGG16 (c) MobileNet (d) ResNet18 (e) CM-CNN (1st and 2nd row: CNV, 3rd and 4th row: drusen) Colors closer to red indicate higher importance in making the classification prediction. The proposed CM-CNN obtains CAM that conform better to the pathological regions. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

Table 6
The ablation experiments of CAM-UNet.

	Dice	IoU	Sen	Spe	pAcc
Res18-UNet	0.7498	0.6378	0.7292	**0.9963**	0.9917
CAM-UNet*	0.7688	0.6557	**0.7616**	0.9959	0.9922
CAM-UNet	**0.7751**	**0.6638**	0.7610	**0.9963**	**0.9923**

components, we compare the proposed models with some of their variations.

1) Ablation tests on the usage of segmentation masks for classification.

Table 5 compares the classification performance of the baseline classification network ResNet18, the classification network where only the segmentation mask *S* is used, denoted as ResNet18 (mask), and CM-CNN using complementary masks *S* and *1-S*. Based on overall Acc and Kappa values, adding single mask information as guidance can improve the classification performance, while using complementary mask information can make better use of the features extracted by the network, resulting in better classification performance.

2) Ablation tests on usage of CAM for segmentation.

First, the CAM obtained by different classification networks using the Grad-CAM algorithm are compared. As shown in Fig. 9, the attention areas of different networks are quite different. VGG16 and MobileNet have multiple convolution and pooling operations, and the attention of the classification network is relatively scattered. However, thanks to the residual structure in ResNet18, the gradient can be transmitted to the deeper part of the network, avoiding the annihilation of.

effective features by multiple downsampling operations, and the resulting CAM has a larger highlighted area. In CM-CNN, with the guidance of proposed complementary segmentation masks, the CAM is more concentrated on the lesion area. From the comparison, it is clear that the CAMs obtained by VGG16 and MobileNet is not good to guide segmentation. While both the CAMs of ResNet18 and CM-CNN can cover the lesion area, CM-CNN locates the lesion area more accurately.

On this basis, we compare the segmentation performance of the baseline segmentation network without CAM guidance, denoted as Res18-UNet, the segmentation network guided by the CAM of ResNet18, denoted as and CAM-UNet*, with the proposed CAM-UNet, guided by the CAM of CM-CNN. The performance indices are shown in Table 6. The experimental results show that both the CAMs of the original ResNet18 and of the CM-CNN can effectively improve the segmentation performance. The proposed CM-CNN not only achieves higher classification accuracy, but offers CAM that more effectively promotes the segmentation network.

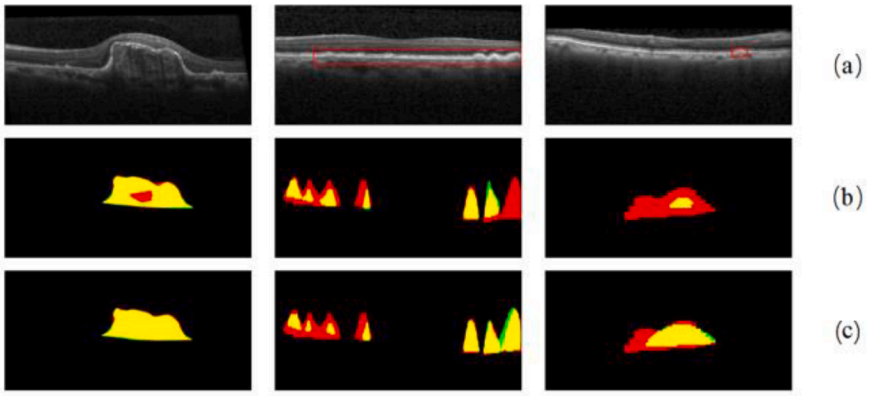


Fig. 10. Segmentation of CAM-UNet and Res18-UNet. (a) original image (b) Res18-UNet (c) CAM-UNet (The segmentation maps of the second and third column correspond to the red rectangle in the original image. Yellow: true positive, Red: false negative, Green: false positive). (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

Table 7
Comparison with classification networks on Dataset II.

	AUC	Sen	Spe	Acc	Kappa
VGG16[33]	0.9405	0.9600	0.7500	0.8550	0.7100
MobileNet[34]	0.9503	0.9100	0.9000	0.9050	0.8100
SENet[35]	0.9820	0.9100	**0.9800**	0.9450	0.8900
InceptionV3[18]	0.9826	**0.9800**	0.8000	0.8900	0.7800
OpticNet[20]	0.9565	0.9500	0.8300	0.8900	0.7800
CM-CNN	**0.9925**	**0.9800**	0.9600	**0.9700**	**0.9400**

Table 8
Comparison with segmentation networks on Dataset II.

	CAM	Dice	IoU	Sen	Spe	pAcc
PSPNet [36]	×	0.7025	0.5682	0.8134	0.9900	0.9879
	√	0.7039	0.5833	0.7624	0.9926	0.9891
Attention-UNet [37]	×	0.7626	0.6386	0.8366	0.9899	0.9880
	√	0.7655	0.6647	0.9063	0.9913	0.9908
UNet [29]	×	0.7665	0.6689	**0.9096**	0.9908	0.9903
	√	0.7685	0.6769	0.8379	0.9931	0.9919
CENet [38]	×	0.7428	0.6188	0.8755	0.9912	0.9902
	√	0.7561	0.6594	0.8415	0.9937	0.9923
CAM-UNet	√	**0.7800**	**0.7085**	0.8881	**0.9953**	**0.9946**

Table 9
Comparison with multi-task networks on Dataset II.

	Acc	Kappa	Dice	IoU
Y-Net [24]	0.8900	0.7800	0.7400	0.6178
Cross-stitch [25]	0.9250	0.8500	0.7550	0.6347
Cross-connected [26]	0.9050	0.8100	0.7413	0.6136
CM-CMM + CAM-UNet	**0.9700**	**0.9400**	**0.7800**	**0.7085**

Fig. 10 shows the segmentation results using Res18-UNet and CAM-UNet. After adding the guidance of the CAM, the network can segment the lesion areas more completely and accurately.

4.5. Results on Dataset II

The experimental results on Dataset II are shown in Table 7-9 and Fig. 11. Compared with the five classification models, four segmentation models and three multi-task models, the proposed method achieves the highest scores in most indices. It obtains a classification accuracy of 97.00%, and a Dice score of 78.00% in segmentation. Table 8 also shows concatenating the CAM with the original OCT image as the input can improve the performance of other models. The qualitative results in Fig. 11 show that, in segmenting the fluid regions, the proposed method have less false negatives and false positives than other models.

5. Discussion and conclusion

AMD is a serious threat to the vision health of middle-aged and elderly people. Automatic detection and quantization of AMD-related pathologies help speed up the diagnosis and analysis of the disease. The dual guidance networks proposed in this paper exploit the strong correlation between the tasks of classification and segmentation tasks. Segmentation mask-guided CM-CNN and class activation map-guided CAM-UNet are designed to achieve classification of OCT B-scans into normal, drusen and CNV, and segmentation of drusen or CNV lesions in these images, respectively. CM-CNN adopts complementary masks to enhance the extracted features, allowing the network to pay more attention to the features of the lesion area while still keeping the information from the non-lesion area. Ablation tests show that the adoption of segmentation guidance improved the classification accuracy. CAM-UNet fuses the CAM information obtained by CM-CNN to the U-shaped segmentation network, so that the network pays more attention to the regions considered important by the classification network. Ablation tests also prove the effectiveness of the guidance from classification. It is also shown that the proposed CM-CNN gives CAMs that better conform with the pathological regions and thus is more helpful for segmentation than other basic classification networks. The results of the proposed dual guidance networks are compared with those of some existing single-task or multi-task networks, and the proposed method achieves better performance. The classification accuracy reaches 96.93% and the Dice coefficient for segmentation reaches 77.51%. Results on an extra dataset for detection of macular edema and segmentation of retinal fluids further show the generalizability of the proposed model, with a classification accuracy of 97.00% and a Dice score of 78.00%. The proposed models have moderate complexity compared to other single-task networks, and require short training and testing time. Although the proposed method is more complex and less efficient than

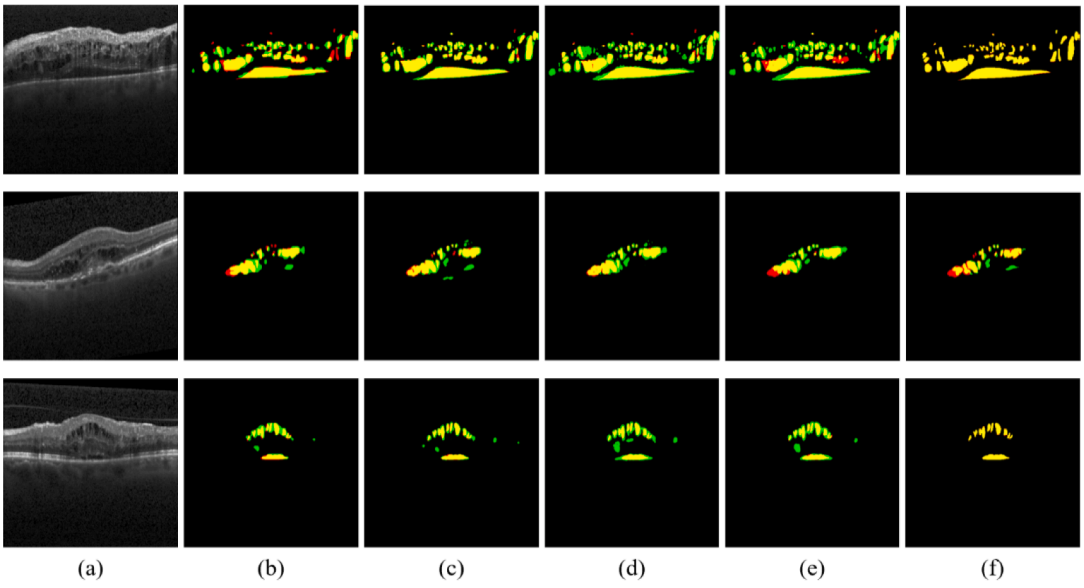


Fig. 11. Segmentation results of retinal fluids using different algorithms (a) original image (b) PSPNet (c) Attention-UNet (d) UNet (e) CENet (f) CAM-UNet. Yellow: true positive, Red: false negative, Green: false positive. Comparative results were obtained without CAM input to the networks. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

the multi-task networks, its testing time can readily fulfill the requirement of clinical applications.

Deep learning methods have been extensively applied to automatic OCT analysis, but most established methods are single-task ones aiming at classification [10–16] or segmentation [17–23]. As the two tasks are often both needed clinically, we design a unified framework that achieves both. Meanwhile, information from one task is used to guide the other, thus achieving improvement for both tasks. Compared to multi-task networks, in the proposed model, the information flow from one task to the other is explicitly defined and explainable. We believe this tailored guidance between tasks is more suitable for the specific tasks, making the proposed model achieve better performance than the general multi-task models. The limitations of the proposed model lie in two aspects. First, compared to single-task models, the training of the proposed networks requires both class and pixel-wise labels for each training image, which may be difficult to obtain. In the future, we'll further exploit semi-supervised or weakly supervised models when some annotations are missing, so as to make use of more clinical data. Secondly, compared to multi-task models, the model requires sequential training and testing, so that the model complexity is higher and the efficiency is lower. Next, we will try to integrate the two networks into an end-to-end framework.

Some other aspects for future improvement are as follows. First, the design of the dual guidance networks mainly focuses on effective utilization of information from different tasks. Although it has outperformed some networks with more complex structures, such as PSPNet [36] with pyramid pooling module, CENet [38] with dense atrous convolution module and residual multi-kernel pooling module, more complex network structure and more advanced modules can be applied to further improve the performance. Secondly, the current classification are single-label ones, and the segmentation treats all lesions as one type. We will further extend the model for multi-label classification and multi-class segmentation.

Funding

This work was supported by the National Natural Science Foundation of China (62271337, 61971298, U20A20170), and the National Key Research and Development Program of China (2018YFA0701700).

CRediT authorship contribution statement

Shengyong Diao: Methodology, Software, Writing – original draft. **Jinzhu Su:** Methodology, Software. **Changqing Yang:** Visualization, Validation. **Weifang Zhu:** Writing – review & editing. **Dehui Xiang:** Formal analysis. **Xinjian Chen:** Project administration, Funding acquisition. **Qing Peng:** Conceptualization, Data curation. **Fei Shi:** Supervision, Methodology, Writing – review & editing.

Declaration of Competing Interest

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

Data availability

Data will be made available on request.

References

[1] P. Mitchell, G. Liew, B. Gopinath, T.Y. Wong, Age-related macular degeneration. Lancet. 2018, 29;392(10153):1147-1159.
[2] S.A. Zweifel, Y. Imamura, T.C. Spaide, T. Fujiwara, R.F. Spaide, Prevalence and significance of subretinal drusenoid deposits (reticular pseudodrusen) in age-related macular degeneration, Ophthalmology 117 (9) (2010) 1775–1781.
[3] F. Zicarelli, C. Azzolini, E. Cornish, A. Agarwal, S. Khochtali, M. Airaldi, M. Khairallah, F. Viola, G. Staurenghi, P. McCluskey, A. Invernizzi, Optical coherence tomography features of choroidal neovascularization and their correlation with age, gender, and underlying disease, Retina 41 (5) (2021) 1076–1083.
[4] P.A. Keane, P.J. Patel, S. Liakopoulos, F.M. Heussen, A. Tufail, Evaluation of age-related macular degeneration with optical coherence tomography, Surv. Ophthalmol. 57 (5) (2012) 389–414.
[5] https://pubmed.ncbi.nlm.nih.gov/.
[6] S. Farsiu, S.J. Chiu, J.A. Izatt, C.A. Toth, Fast detection and segmentation of drusen in retinal optical coherence tomography images, Proc SPIE 4 (1) (2008) 68440D.
[7] K. Yi, M. Mujat, B. Park, W. Sun, J. Miller, J. Seddon, L. Young, J. Boer, T. Chen, Spectral domain optical coherence tomography for quantitative evaluation of drusen and associated structural changes in non-neovascular age-related macular degeneration, Br. J. Ophthalmol. 93 (2) (2009) 176–181.
[8] Q. Chen, T. Leng, L. Zheng, L. Kutzscher, J. Ma, L. de Sisternes, D.L. Rubin, Automated drusen segmentation and quantification in SD-OCT images, Med. Image Anal. 17 (8) (2013) 1058–1072.
[9] J. Oliveira, L. Goncalves, M. Ferreira, C.A. Silva, Drusen detection in OCT images with AMD using random forests, IEEE Portuguese Meeting on Bioengineering, 2017.
[10] G.Z. Shekoufeh, W.M.W. Maximilian, W. Vitails, T. Sarah, G.H. Frank, P.F. Robert, S. Thomas, CNNs enable accurate and fast segmentation of drusen in optical coherence tomography, International Workshop on Deep Learning in Medical Image Analysis International Workshop on Multimodal Learning for Clinical Decision Support, 2017,10553.
[11] R. Asgari, J.I. Orlando, S. Waldstein, F. Schlanitz, M. Baratsits, U. Schmidt-Erfurth, H. Bogunović, Multiclass segmentation as multitask learning for drusen segmentation in retinal optical coherence tomography, Medical Image Computing and Computer Assisted Intervention 11764 (2019) 192–200.
[12] M. Wang, W. Zhu, F. Shi, J. Su, H. Chen, K. Yu, Y. Zhou, Y. Peng, Z. Chen, X. Chen, MsTGANet: Automatic drusen segmentation from retinal OCT images, IEEE Trans. Medical Imaging 41 (2) (2022) 394–406.
[13] Y. Li, S. Niu, Z. Ji, W. Fan, S. Yuan, Q. Chen, Automated choroidal neovascularization detection for time series SD-OCT Images, International Conference on Medical Image Computing and Computer-Assisted Intervention (2018) 381–388.
[14] X. Xi, X. Meng, Z. Qin, X. Nie, Y. Yin, X. Chen, IA-Net: Informative attention convolutional neural network for choroidal neovascularization segmentation in OCT images, Biomed. Opt. Express 11 (11) (2020) 6122–6136.
[15] Y. Zhang, Z. Ji, Y. Wang, S. Niu, W. Fan, S. Yuan, Q. Chen, MPB-CNN: a multi-scale parallel branch CNN for choroidal neovascularization segmentation in SD-OCT images, OSA Continuum 2 (3) (2019) 1011–1027.
[16] Q. Meng, L. Wang, T. Wang, M. Wang, W. Zhu, F. Shi, Z. Chen, X. Chen, MF-net: multi-scale information fusion network for CNV segmentation in retinal OCT images, Front. Neurosci. 15 (2021), 743769.
[17] W. Wang, X. Li, Z. Xu, W. Yu, J. Zhao, D. Ding, Y. Chen, Two-Stream CNN with loose pair training for multi-modal AMD categorization. Medical Image Computing and Computer Assisted Intervention,2019.
[18] D.S. Kermany, M. Goldbaum, I.W. Ca, et al., Identifying medical diagnoses and treatable diseases by image-based deep learning, Cell 172 (5) (2018) 1122–1131. e9.
[19] L. Fang, C. Wang, S. Li, H. Rabbani, X. Chen, Z. Liu, Attention to Lesion: Lesion-Aware convolutional neural network for retinal optical coherence tomography image classification, IEEE Trans. Med. Imaging 38 (8) (2019) 1959–1970.
[20] S.A. Kamran, S. Saha, A.S. Sabbir, A. Tavakkoli, Optic-Net: A novel convolutional neural network for diagnosis of retinal diseases from optical tomography images. IEEE International Conference on Machine Learning And Applications, 2019,964-971.
[21] A. Thomas, P.M. Harikrishnan, A.K. Krishna, P. Palanisamy, V.P. Gopi, A novel multiscale convolutional neural network based age-related macular degeneration detection using OCT images, Biomed. Signal Process. Control 67 (2021), 102538.
[22] S. Sotoudeh-Paima, A. Jodeiri, F. Hajizadeh, H. Soltanian-Zadeh, Multi-scale convolutional neural network for automated AMD classification using retinal OCT images, Comput. Biol. Med. 144 (2022), 105368.
[23] Z. Ma, Q. Xie, P. Xie, F. Fan, X. Gao, J. Zhu, HCTNet: A hybrid convnet-transformer network for retinal optical coherence tomography image classification, Biosensors 12 (2022) 542.
[24] S. Mehta, E. Mercan, J. Bartlett, D. Weave, J.G. Elmore, L. Shapiro, Y-Net: Joint segmentation and classification for diagnosis of breast biopsy images. International Conference on Medical Image Computing and Computer-Assisted Intervention, 2018, 893-901.
[25] I. Misra, A. Shrivastava, A. Gupta, M. Hebert, Cross-stitch networks for multi-task learning, IEEE Conference on Computer Vision and Pattern Recognition (2016) 3994–4003.
[26] R. Kawakami, R. Yoshihashi, S. Fukuda, S. You, M. Iida, T. Naemura, Cross-connected networks for multi-task learning of detection and segmentation. IEEE International Conference on Image Processing (ICIP), 2019.
[27] K. He, X. Zhang, S. Ren, J. Sun, Deep residual learning for image recognition, IEEE Conference on Computer Vision and Pattern Recognition (2016) 770–778.
[28] R.R. Selvaraju, M. Cogswell, A. Das, R. Vedantam, D. Parikh, D. Batra, Grad-CAM: Visual explanations from deep networks via gradient-based localization, IEEE International Conference on Computer Vision 128 (2) (2019) 336–359.
[29] O. Ronneberger, P. Fischer, T. Brox, UNet: Convolutional networks for biomedical image segmentation. International Conference on Medical Image Computing and Computer-Assisted Intervention, 2015, 234-241.
[30] H. Bogunovic, F. Venhuizen, S. Klimscha, S. Apostolopoulos, A. Bab-Hadiashar, U. Bagci, M.F. Beg, L. Bekalo, Q. Chen, C. Ciller, K. Gopinath, A.K. Gostar, K. Jeon, Z. Ji, S.H. Kang, D.D. Koozekanani, D. Lu, D. Morley, K.K. Parhi, H.S. Park,

A. Rashno, M. Sarunic, S. Shaikh, J. Sivaswamy, R. Tennakoon, S. Yadav, S. De Zanet, S.M. Waldstein, B.S. Gerendas, C. Klaver, C.I. Sanchez, U. Schmidt-Erfurth, RETOUCH: The retinal OCT fluid detection and segmentation benchmark and challenge, IEEE Trans. Medical Imaging 38 (8) (2019) 1858–1874.

[31] P.P. Srinivasan, L.A. Kim, P.S. Mettu, S.W. Cousins, G.M. Comer, J.A. Izatt, S. Farsiu, Fully automated detection of diabetic macular edema and dry age-related macular degeneration from optical coherence tomography images, Biomed Opt Express 5 (10) (2014) 3568–3577.

[32] H.C. Sang, H. Kwon, H.W. Jin, H. Yoon, K.S. Park, Long Short-Term memory networks for unconstrained sleep stage classification using polyvinylidene fluoride film sensor, IEEE J. Biomed. Health Inform. 24 (12) (2020) 3606–3615.

[33] K. Simonyan, A. Zisserman, Very deep convolutional networks for large-scale image recognition, IEEE Conference on Computer Vision and Pattern Recognition, 2014.

[34] A.G. Howard, M. Zhu, B. Chen, et al. MobileNets: Efficient convolutional neural networks for mobile vision applications, IEEE Conference on Computer Vision and Pattern Recognition, 2017.

[35] H. Jie, S. Li, S. Gang, S. Albanie, Squeeze-and-excitation networks, IEEE Transactions on Pattern Analysis and Machine Intelligence, 2017,PP(99): 7132–7141.

[36] H. Zhao, J. Shi, X. Qi, X. Wang, Pyramid scene parsing network, Computer Vision Pattern Recognition 1 (2016) 6230–6239.

[37] O. Oktay, J. Schlemper, L.L. Folgoc, M. Lee, M. Heinrich, K. Mori, S. McDonagh, N. Y. Hammerla, B. Kainz, B. Glocker, Attention U-Net: Learning where to look for the pancreas, IEEE Conference on Computer Vision and Pattern Recognition, 2018.

[38] Z. Gu, J. Cheng, H. Fu, K. Zhou, H. Hao, Y. Zhao, T. Zhang, S. Gao, J. Liu, CE-Net: Context encoder network for 2D medical image segmentation, IEEE Trans. Med. Imaging 38 (10) (2019) 2281–2292.