

IA-net: Informative Attention Convolutional Neural Network for Choroidal Neovascularization Segmentation in OCT Images

XIAOMING XI,¹ XIANJING MENG,² ZHEYUN QIN,³ XIUSHAN NIE,¹ YILONG YIN,^{3,5} AND XINJIAN CHEN^{4,6}

¹ School of Computer Science and Technology, Shandong Jianzhu University, 250101, China

² School of Computer Science and Technology, Shandong University of Finance and Economics, 250014, China

³ School of Software, Shandong University, 250101, China

⁴ School of Electronic and Information Engineering, Soochow University, 215006, China

⁵* ylyin@sdu.edu.cn

⁶* xjchen@suda.edu.cn

Abstract: Choroidal neovascularization (CNV) is a characteristic feature of wet age-related macular degeneration (AMD). Quantification of CNV is useful to clinicians in the diagnosis and treatment of CNV disease. Before quantification, CNV lesion should be delineated by automatic CNV segmentation technology. Recently, deep learning methods have achieved significant success for medical image segmentation. However, some CNVs are small objects which are hard to discriminate, resulting in performance degradation. In addition, it's difficult to train an effective network for accurate segmentation due to the complicated characteristics of CNV in OCT images. In order to tackle these two challenges, this paper proposed a novel Informative Attention Convolutional Neural Network (IA-net) for automatic CNV segmentation in OCT images. Considering that attention mechanism has the ability to enhance the discriminative power of the interesting regions in the feature maps, the attention enhancement block is developed by introducing the additional attention constraint. It has the ability to force the model to pay high attention on CNV in the learned feature maps, improving the discriminative ability of the learned CNV features, which is useful to improve the segmentation performance on small CNV. For accurate pixel classification, the novel informative loss is proposed with incorporation of informative attention map. It can focus training on a set of informative samples which are difficult to be predicted. Therefore, the trained model has the ability to learn enough information to classify these informative samples, further improving the performance. The experimental results on our database demonstrate that the proposed method outperforms traditional CNV segmentation methods.

© 2020 Optical Society of America under the terms of the OSA Open Access Publishing Agreement

1. Introduction

Age-related macular degeneration (AMD) is one of the leading causes of blindness particularly in old people. Choroidal neovascularization (CNV) is a characteristic feature of AMD. It is characterized by the growth of abnormal blood vessels from the choroid layer [1,2]. These abnormal blood vessels expand from the choroid underneath the retina and leak. This leakage damages surrounding retinal tissue, thus causing deterioration in central vision [3, 4].

Optical Coherence Tomography (OCT) has been widely used for evaluation of CNV. High resolution OCT imaging technique enables the sensitive detection of multiple retinal cell layers and quantitative assessment of these macular lesions within retina [5,6]. Compared with other imaging modalities, such as Fluorescein angiography (FA), indocyanine green angiography

(ICGA), OCT has the following advantages [7,8]: (1) it's noninvasive. (2) It can obtain high resolution cross-sectional images of the neurosensory retina. (3) High-speed acquisition.

Accurate CNV segmentation could help doctors conducting auxiliary diagnosis and treatment. The image segmentation method can delineate CNV lesion automatically [9]. Based on the obtained CNV lesion, doctors acquire the properties of CNV lesion, including the area, volume, width, height, optical density value, etc. These properties play an important role in diagnosis and treatment of CNV. In order to obtain precise properties of CNV, it is necessary to develop effective segmentation method for accurate CNV segmentation.

In recent years, deep learning methods [10] have achieved remarkable success in image processing area, such as image denoising [11], image reconstruction [12,13], image segmentation[14,15,16,17]. However, satisfactory performance was hardly achieved by directly using the existing methods due to the complex characteristics of CNV in OCT images.

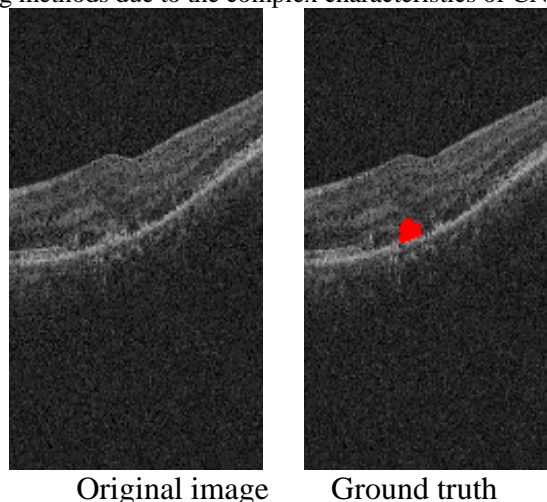


Fig. 1. Examples of CNV in OCT image.

There are two challenges for CNV segmentation in OCT images. (1) The existing methods fail to specially focus small CNV in the feature map learning process, result in unsatisfactory performance on small CNV. CNV has been occurred as a small object in the OCT image frequently. The mean proportion of CNV pixels in our dataset is 0.25%. Fig.1 also gives two examples of CNV in OCT image. We can infer that there is limited information of CNV in OCT image, as the spatial resolution of the feature maps is decreased and the large context information is integrated, the discrimination power of small object features may be easily weakened [18] in the low level feature maps. Meanwhile most of the small object features may be lost in the high-level feature maps. Therefore, it's difficult for existing methods to learn discriminative representations of small CNV, resulting in the performance degradation.

(2) CNV in OCT image has complicated image characteristics, result in inaccurate segmentation. As shown in Fig.1, the intensity distribution is complex. Fig. 2 gives the intensity distributions between CNV and background. As shown in this figure, we can see that there is a large intensity overlap between CNV and background (certain retinal structures), results in the large inter-class similarity and intra-class variation. Therefore, it's difficult to achieve the accurate classification for the CNV pixels in the intensity overlap interval.

In order to tackle these two challenges, a novel Informative Attention Convolutional Neural Network (IA-net) is proposed by introducing the attention mechanism. Considering that attention mechanism has the ability to enhance the discriminative power of the interesting regions [19, 20, 21, 22], the novel attention enhancement block is firstly developed by introducing the attention constraint. It can force the learned feature maps to be similar with the attention map with ideal discriminative information. In this way, it can improve the

discriminative ability of CNV features in the low-level feature maps and preserve some feature information of CNV in the high-level feature maps, improve the discriminative ability of the learned features of small CNV. For accurate pixel classification, the novel informative loss is developed by exploring the informative attention map. In this paper, the CNV samples whose class membership is hard to be decided are referred to as informative samples. According to the developed informative attention model, the informative samples will be assigned high attention in the training process. After training, the obtained model has ability to learn enough knowledge which is robust to classify these informative instances, further improving the performance. To demonstrate the effectiveness of the proposed network, we conduct the experiment on our dataset that contains 3034 image slices from 67 3D-OCT data with CNV. The experimental results demonstrate that the proposed method outperforms the traditional deep learning methods.

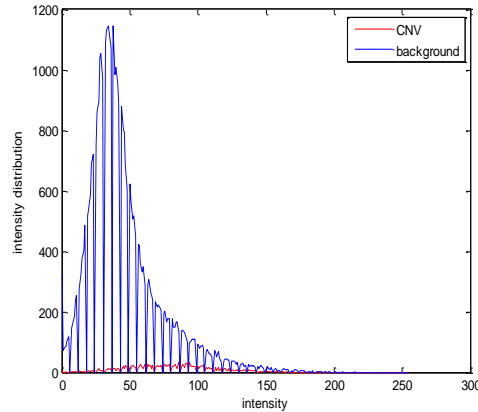


Fig. 2. Intensity distributions of CNV and background in our dataset

The main contributions of this paper are as follows:

(a) The attention enhancement block is developed by introducing the attention constraint. It has ability to force the model to pay high attention on CNV, improving the discriminative ability of the learned features of small CNV, which is useful to improve the segmentation accuracy on small CNV.

(b) In order to obtain accurate classification for CNV pixels which are difficult to be predicted, the informative loss is proposed with incorporation of informative attention map. It is helpful to learn enough information which is robust to classify these informative instances, further improving performance.

2. Related work

Recently, automatic CNV segmentation methods have been proposed. Xi et al. proposed a learned local similarity prior embedding active contour model [23]. The local similarity prior was firstly learned by using superpixels and local potential function. And then, the new energy function was constructed by combing the local similarity prior. Zhu et al. proposed a CNV growth prediction with treatment based on reaction-diffusion model in 3-D OCT images [24]. Before growth prediction, they performed CNV segmentation by using graph search. Xiang et al. proposed Neural Network and constrained graph search algorithm (NNCGS). In the proposed method, manual designed features and neural network classifier are firstly used for initial segmentation. Based on the initial result, a constrained graph search algorithm was proposed for finer segmentation of CNV [3]. Liet al. proposed a new 3D-histogram of oriented gradient (3D-HOG) feature and update the random forest models persistently [25]. However, it's difficult for handcrafted feature to extract enough discriminative information from smaller CNV. Considering the powerful learning ability of convolutional neural networks, Xi et al proposed multi-scale convolutional neural networks with structure prior [26]. The new

segmentation model was constructed by introducing the structure prior and multi-scale information into the convolutional neural networks. However, the proposed patchwise training method was time-consuming due to the large resolution of OCT image.

In the recent years, deep learning methods were proposed for medical images segmentation. Sparse autoencoders were proposed for efficient nuclei detection on high-resolution histopathological images [27,28]. Van Tulder et al., trained restricted Boltzmann machine with a generative learning objective for airway detection in CT images [29]. However, autoencoder and restricted Boltzmann machine belonged to the unsupervised deep framework without supervised information in the training processing. Generally speaking, as a classic supervised deep framework, convolutional neural networks (CNNs) architectures may achieve better performance. To obtain the multi-scale information about each voxel, multiple CNNs were trained based on 2D image patches with different sizes for segmentation of MR brain images [30]. In order to use multi-modality information of MR images, Zhang et al. proposed to use CNN for segmenting isointense stage brain tissues of multi-modality MR images [31]. In their work, 2D patches from T1, T2, and fractional anisotropy (FA) images were extracted as the training instances. However, the segmentation methods based on patches was time-consuming because the segmented medical image may generated large number of patches. In order to solve this problem, Fully Convolutional Networks (FCN) [32] was proposed. In order to input the image with arbitrary sizes, FCN replaced fully connected layers with the convolutional ones whose size was 1×1 , which can efficiently learn dense predictions for semantic segmentation. In addition, deconvolution layers are also introduced to obtain the segmentation result with the same size of input image. Compared with traditional segmentation methods, FCN achieved a significant improvement in segmentation accuracy and efficiency at inference. By using similar idea of FCN, Ronneberger et al. [33] proposed the U-net architecture for biomedical image segmentation. Chen et al. proposed a novel dual-force training scheme which was applied to U-net architecture [34]. Wei et al. proposed a multimodel, multi-size and multi-view deep neural network (M3Net) for brain MR image segmentation, which uses three identical modules to segment transaxial, coronal, and sagittal MR slices, respectively [35]. Each module consists of multi-size U-Nets and multi-size back propagation neural networks. Milletari et al. proposed V-net architecture which was 3D-variant of U-net [36]. U-net architecture based methods have ability to learn effective features of objects. However, these methods fail to specially focus small object in the feature map learning process, result in performance degradation. In addition, the existing methods ignored the accurate classification of hard pixels, result in the limitation of the performance improvement [37,38]. As an effective method to deal with hard examples, focal loss [39] was proposed to focus training on hard examples and achieve remarkable performance improvement. However, manually tuning hyperparameters in this loss is time-consuming.

3. Method

The basic architecture of the proposed network is shown in Fig. 3. The main modules of the proposed network contain attention enhancement block, max pooling, upsampling and prediction layer. The attention enhancement block is firstly proposed to guide to learn discriminative features of CNV. It can pay high attention on CNV features, which is helpful to improve the discriminative ability of CNV features. In order to reduce the complexity of features, max pooling is used to select the effective information from the learned feature maps. After that, upsampling is performed to obtain the segmentation result with the same size of input images by using deconvolution operation. At last, the prediction layer which introduced informative loss is used for pixel classification and obtains the final segmentation result.

To be noticed, Fig. 3 gives the basic architecture of the proposed network. In our experiment, U-net[33] is used as the backbone network. The basic network parameters such as kernel size, channel size, stride, are same as [33]. The proposed network can be regarded as a modified U-

network. Different from [33], we replace the cross-entropy loss by using informative loss. Moreover, the attention enhance block is introduced additionally.

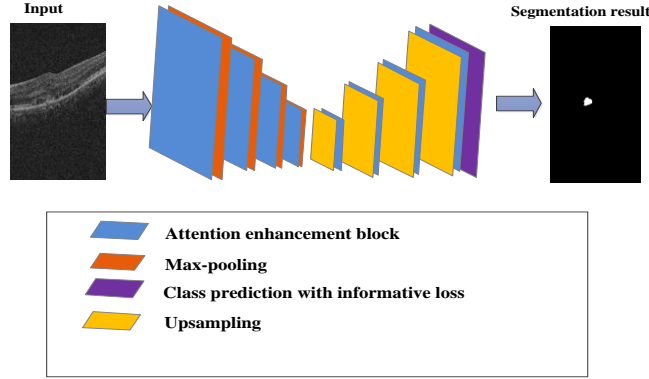


Fig. 3. Basic architecture of the proposed network

3.1 Attention Enhancement Block

For traditional convolutional neural network, convolutional layer performs convolution of the local patch of input maps with different filter banks. After that, corresponding convolutions are summed up, and then passed through a nonlinear activation function such as a ReLU to generate different feature maps which can capture local statistics of images.

The traditional convolution operation mainly consists of two steps: 1) sampling using a convolutional kernel with size S over the input feature map. 2) summation of sampled values by w . The output feature map $O(v_0)$ of the location v_0 can be calculated as:

$$p(v_0) = \sum_{v_n \in S} w(v_n) \cdot x(v_0 + v_n) \quad (1)$$

$$G(a) = \max(0, a) \quad (2)$$

$$O(v_0) = G(p(v_0)) \quad (3)$$

In above equations, $O(v_0)$ denotes the output feature map of the location v_0 . v_n represents the n -th element in the receptive field S . G denotes the active function. Considering that Rectified Linear Units (ReLU) can improve training speed, ReLU is used as the active function in this paper.

As shown in above equations, as the spatial resolution of the feature maps is decreased and the large context information is integrated, discriminative power of small object features may be weakened[18] in the low-level feature maps. In addition, most of the small object features may be lost in the high-level feature maps. Therefore, it is difficult to learn discriminative features for small CNV by using traditional networks, result in the performance degradation.

Attention is helpful to focus on the interesting object by finding important feature areas from feature maps. Inspired by this, we adopt attention mechanism to learn discriminative features of CNV by paying high attention on CNV regions in the feature learning process.

This paper proposed the attention enhancement block to introduce attention for discriminative feature learning. As shown in Fig. 4, the attention enhancement block mainly contains two parts: traditional convolutional layers and the attention constraint. We used traditional convolutional layers for feature learning and append a 1×1 convolution, followed by the last convolutional layer to introduce the attention constraint. The attention constraint is

constructed by embedding discriminative attention map into the feature learning process. The proposed constraint in the layer L is defined to be the difference between the learned feature and the introduced attention map, which is formulated as:

$$AL(O, \sigma) = \|\sigma O - af\|_2 \quad (4)$$

In above equation, af denotes the introduced attention map. Considering that CNV can be separated from background perfectly in the groundtruth image, we regarded groundtruth as the ideal attention map which is useful to contribute to the discriminative power of features [40,41]. O is the learned feature map and σ denotes the parameters of the 1×1 convolution in the layer L .

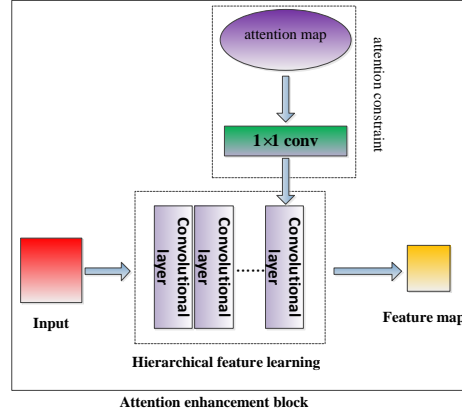


Fig. 4. Attention enhancement block

As shown in Eq.(4), the attention constraint has the ability to force the learned feature maps to be close to the corresponding attention map in the layer L . The attention map can be regarded as the discriminative features, in which CNV has high attention while background has low attention. Therefore, the proposed block can pay high attention on CNV features in each layer. It can improve the discriminative ability of CNV features in the low-level feature maps and preserve the feature information of CNV in the high-level feature maps, improving the discriminative ability of CNV features, which is useful to improve the segmentation accuracy on small CNV. Compared with existing attention mechanism, the proposed method has two advantages. (1)The network architecture of the proposed method is simpler. For example, in [19], in order to learn attention, Squeeze-and-Excitation-Unit (SE-Unit) block is introduced in the basic network. In the contrast, in our network, the attention is learned via attention constraint, without introduction of extra units or blocks. (2)The proposed method has advantages on small CNV segmentation. Existing methods pay attention on discriminative region [19]. However, it is difficult for them to pay attention on all CNV pixels. In contrast, the proposed method can pay attention on each CNV pixel via attention constraint in each feature map. It is useful to preserve the discriminative features of CNV pixel, avoid CNV features loss in the feature learning process. Therefore, the proposed method is more suitable for small CNV.

3.2 Informative Loss

It's helpful to improve the performance of the classification model trained based on informative samples [42,43,44]. Generally speaking, informativeness measures the ability of an instance in reducing the uncertainty of a learning model[43,44]. For example, in classification task, uncertainty has the ability to assess a training model's certainty in classifying an instance. If the uncertainty of the instances is high, it implies that the current model does not have enough knowledge to classify these instances, and presumably, focusing

the model on these uncertainty instances can help improve the robustness of underlying learning model [40].

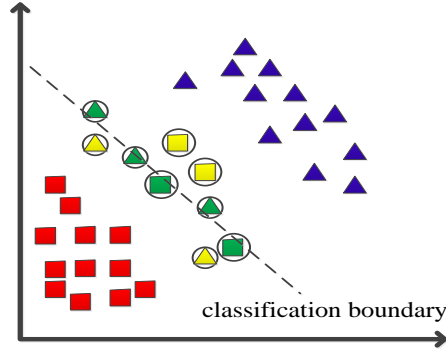


Fig. 5. An illustration example for informative samples

Considering that informative sample contains a high uncertainty [44], the CNV samples whose class membership is hard to be decided via the learning model are referred to as informative samples. Besides, the CNV instances that yield prediction error is also regarded as the informative samples. Fig. 5 gives a simple illustration example for informative samples. Triangle and the rectangle are two classes. The green elements near the classification boundary denote the samples whose class membership is hard to be decided. Meanwhile, the yellow elements denote the samples which are predicted incorrectly. Therefore, the yellow elements and the green elements are regarded as informative samples in this paper.

In order to mine these informative instances, informative attention map is generated firstly. And then, informative loss is developed by exploring informative attention map, further improving the performance.

In this paper, the informative attention map is calculated based on clinical prior. Here, the clinical prior is about features of CNV in OCT image[45]. Generally speaking, CNV in OCT image has the characteristics [23,45]: the global intensity of CNV is relative high and the local intensity variation occurred in CNV region.

Based on above idea, we represent the clinical prior by exploring global prior and the local prior. The prior probability of arbitrary pixel s is calculated as follows:

$$P(s) = GP(s)LP(s) \quad (5)$$

Where $GP(s)$ denotes the global prior probability while $LP(s)$ denotes the local prior probability for arbitrary pixel s .

The global prior is mainly used to capture the global intensity statistics of CNV. A histogram-based intensity method is used to obtain global prior probability, calculated as following:

$$GP(s) = \frac{Num(I(s))}{TNum} \quad (6)$$

Where $TNum$ denotes the total number of the training CNV pixels while $Num(I(s))$ denotes the number of training CNV pixels whose intensity value is $I(s)$.

The local prior is mainly used to capture the intensity variation in a neighbor region. A histogram-based local contrast method is used to obtain local prior probability, calculated as following:

$$LD(s) = \frac{\sum_{t \in R} |I(t) - I(s)|}{NNum} \quad (7)$$

$$LP(s) = \frac{Num(LD(s))}{TNum} \quad (8)$$

Where variable $LD(s)$ is calculated as the mean of the local contrast in its neighbor region R , capturing the local difference among the pixels and its neighbor pixels. $NNum$ denotes the number of the pixels in the neighbor region R . The local prior can be calculated according to (8), where $Num(LD(s))$ denotes the number of training CNV pixels whose value is $LD(s)$.

According to above equations, we can infer that the CNV pixels will be assigned to larger values.

After obtaining the prior, the informative attention model is developed to assign the attention for each pixel.

$$ia(s) = \begin{cases} e^{-(F(p(s)-pb))^2}, & \text{if } s \text{ belongs to CNV} \\ \epsilon, & \text{if } s \text{ belongs to background} \end{cases} \quad (9)$$

$$F(x) = \begin{cases} 0, & x \leq 0 \\ x, & x > 0 \end{cases} \quad (10)$$

Where ϵ is a small value and is set to 0.001 in our experiment. The segmentation task can be regarded as the binary classification problem. Variable pb denotes classification boundary and its value is set to 0.5 in our experiment. The class prior probability of CNV should be larger than 0.5 while the class prior probability of background should be less than 0.5. According to Eq.(9) and (10), if the class prior probability of CNV pixels is less than 0.5 or near 0.5, these CNV pixels are informative samples which should be assigned higher attention.

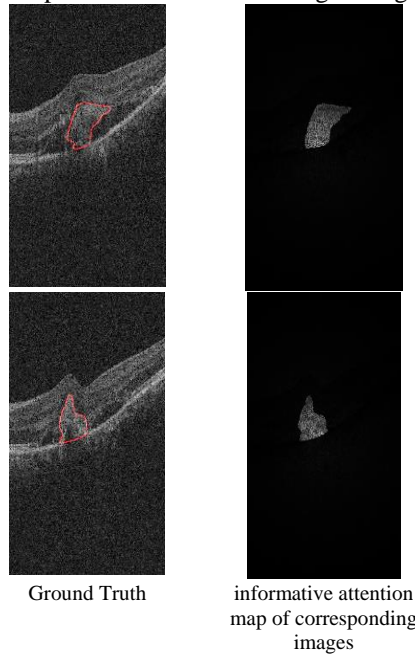


Fig. 6. Two examples of the informative attention map

Fig. 6 gives two examples of the informative attention map. The first row shows two slices with CNV and the second row shows the informative attention map of the corresponding images. As shown in this figure, some informative pixels such as pixels that are similar with background pixels have high attention values.

In order to improve the segmentation accuracy of informative pixels, informative loss is developed by introducing the informative attention map. The informative loss is formulated as:

$$L(\mathbf{w}, \theta) = -\frac{1}{N} \sum_{i=1}^N \sum_{t=1}^M ia(v_i) I\{y^{(i)} = t\} \log \frac{e^{\theta_t O_t(v_i)}}{\sum_{k=1}^M e^{\theta_k O_k(v_i)}} \quad (11)$$

In above equation, ia is introduced informative attention map, I is the indicator function. N is the total number of the elements in the feature map. M is the number of classes. In this paper, the value of M is 2, i.e. CNV and background. θ is the parameters of the network.

The proposed loss is designed to improve performance by assigning high attention to informative samples such that their contribution to the total loss is large. As shown in above function, the error prediction of informative samples will lead to a large value for the objective function due to the introduction of informative attention. In order to reduce the loss, the model should focus training on a set of informative samples. Therefore, the trained model has ability to learn enough knowledge to classify these informative instances, further improving performance.

4. Experiment

4.1 Experiment Setting

The experiments were performed on our 3D-OCT dataset. The data set was acquired using Topcon 3D-OCT-1000 (Topcon Corporation, Tokyo, Japan). Each SD-OCT volume contains $512 \times 1024 \times 128$ voxels. The dataset contains 67 eyes from 67 patients. Each OCT data contains 128 B-scan images. In our experiments, we only select the B-scans with CNV. There are total 3034 B-scan images. The range of proportions of CNV pixels is 0.000385 to 0.126. This study was approved by the Intuitional review board of Joint Shantou International Eye Center and adhered to the tenets of the Declaration of Helsinki. Because of its retrospective nature, informed consent was not required from subjects. The CNV was manually delineated by three retinal specialists, and the groundtruth of each slice was obtained by combining delineation results with major voting.

To evaluate the performance of the proposed method, Dice Similarity Coefficient (DSC), True Positive Volume Fraction (TPVF) and False Positive Volume Fraction (FPVF) are used as performance indices. DSC is used to measure the accuracy of the automatic CNV segmentation result as compared against reference standard delineation; TPVF indicates the fraction of the total amount of CNV in the true segmentation by the proposed method; FPVF denotes the amount of CNV pixels falsely identified by the proposed method. They are calculated as follows:

$$DSC = 2 \times \frac{|V_A \cap V_M|}{|V_A \cup V_M|} \quad (12)$$

$$TPVF = \frac{|V_A \cap V_M|}{|V_M|} \quad (13)$$

$$FPVF = \frac{|V_A| - |V_A \cap V_M|}{|V - V_M|} \quad (14)$$

Where $|\cdot|$ denotes volume, V_A denotes the CNV region segmented by the proposed method, V_M denotes the CNV region in the groundtruth, V denotes the total volume of the OCT data.

In our experiment, the first 57 OCT data are used for training and the remained 10 OCT data are used for testing. In the test set, there are 454 B-Scan images with CNV, and the range of CNV pixel proportions is 0.00077 to 0.068. The remain B-scans are used for training. In order to enlarge the size of training data, data augmentation is performed by horizontally flipping and rotating image by $30^\circ, 60^\circ$. In the training process, the learning rate, batch size, training epoch of U-net are set as 0.001, 5, 200, respectively.

4.2 Effectiveness of attention enhancement block evaluation

In this experiment, we compare IA-net and our network without attention enhancement block (Informative Neural Network, INN) to demonstrate the effectiveness of proposed attention enhancement block.

Table 1. Quantitative Results of IA-net and INN

	TPVF	DSC	FPVF
INN	0.9030±0.0841	0.8647±0.095	0.0044±0.0019
IA-net	0.9384±0.0617	0.8862±0.0785	0.0043±0.0014

Table 1 reports the quantitative results of INN and IA-net. As observed in the table, TPVF, DSC, FPVF of IA-net are better than INN. Fig.8 also gives the segmentation performance on small objects. IA-net achieves the best performance on small CNVs.

For IA-net, the attention enhancement block is introduced additionally. It can pay high attention on CNV in the feature map learning process, which is useful to improve the discriminative ability of CNV features in the low-level feature maps and preserve the feature information of CNV in the high-level feature maps. This implies that more discriminative features can be learned effectively by IA-net, resulting in accurate segmentation of more CNV pixels.

4.3 Effectiveness of informative loss evaluation

In this experiment, we compare with the cross-entropy loss, focal loss to demonstrate the effectiveness of proposed informative loss. Table 2 gives the comparison results. Here, U-net is used as the backbone network and the attention enhancement block is introduced in the U-net. The final loss of the network is used as weighted cross-entropy loss, focal loss and informative loss respectively.

As reported in this table, IA-net achieves the best performance on DSC, TPVF. Weighted cross-entropy loss assigned the weight according to the prediction accuracy of pixels. The pixels which are predicted incorrectly will be assigned a high weight. Different from it, in our loss, we only focus CNV pixels. The CNV pixels which are difficult to be predicted (the point near the classification boundary in Fig. 5) and the CNV pixels which are predicted incorrectly will be assigned high weight. Therefore, the proposed method achieved better TPVF and DSC than weight cross-entropy loss. Focal loss has the ability to deal with hard examples. However, its performance is independent on the manual tuning which is time-consuming. Different from focal loss, the proposed informative can mine the informative samples and assign the attention value to each sample automatically, eliminating manual parameters tuning. In addition, the proposed method only focuses on the hard samples which belong to CNV, result in better TPVF.

Table 2. Quantitative Results of different loss

	TPVF	DSC	FPVF
Weighted Cross-entropy loss	0.9087±0.0887	0.8639±0.093	0.00425±0.0013
Focal loss	0.9130±0.0724	0.8658±0.081	0.0042±0.0012
Informative loss	0.9384±0.0617	0.8862±0.0785	0.0043±0.0014

4.4 Comparison with other segmentation method

In this experiment, we also compare IA-net with existing CNV segmentation methods such as MS-CNN[26], ACM-LSP[23], NNCGS [3] and the backbone network U-net[33]. Table 3 shows TPVF, DSC, FPVF of different methods respectively. As shown in the table, IA-net obtaining better performance on TPVF, DSC and FPVF respectively. Compared with U-net, DSC of IA-net has been increased about 3 percentage points.

Table 3. Quantitative Results of different methods

	TPVF	DSC	FPVF
ACM -LSP	0.840	0.5821	0.015
MS-CNN	0.8323	0.8094	0.0075
NNCGS	0.8212	0.8454	0.005
U-net	0.8744	0.8556	0.0046
IA-net	0.9384	0.8862	0.0043

Fig.7 gives the segmentation comparison examples for small CNVs. The proposed IA-net achieves the best performance on small CNVs. U-net can not segment all the pixels of small CNVs because the feature of CNVs in high-level feature maps may be lost. Other methods can not achieve satisfactory performance on small CNV because it is hard to discriminate small CNV with limited information in the complex images.

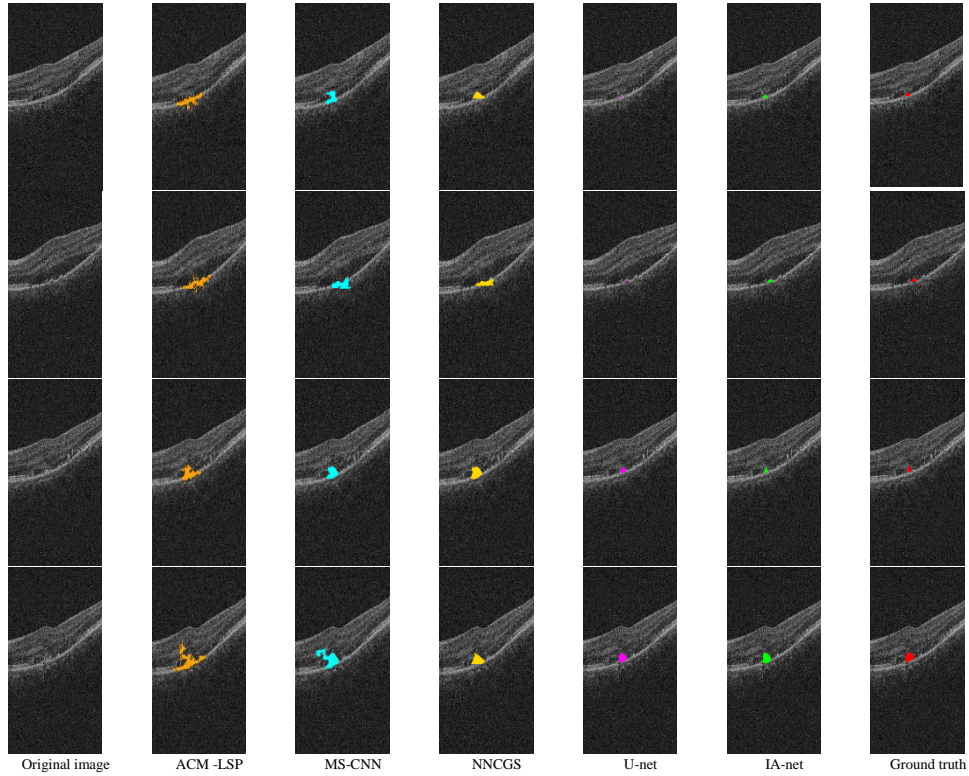


Fig. 7. Segmentation examples of small CNVs with different methods

Among these methods, ACM-LSP can be regarded as a shallow method which introduces the similarity prior into the active contour model. However, it is difficult to generate accurate segmentation result due to the complex characteristics. Generally speaking, deep learning

methods outperform shallow methods due to the powerful feature learning ability. Compared with traditional CNN, MS-CNN introduced the structure prior and multi-scale information in the CNN, improving the performance. However, MS-CNN is patch-based segmentation method, increasing the segmentation time due to the large resolution of OCT image. In addition, MS-CNN ignores to mine informative samples, result in performance degradation of informative samples. It's difficult for NNCGS to obtain accurate initial segmentation result by only using the handcraft feature and the shallow neural network. Therefore, the constrained graph search fails to achieve satisfactory segmentation result based on the inaccurate initial segmentation. The input of U-net is the whole image, which can reduce the segmentation time. However, the complicated intensity distribution of OCT images and small proportion of CNV may affect the effectiveness of the learned feature of U-net, resulting in the performance degradation. For proposed IA-net, attention enhancement block is developed to learn more discriminative information of CNV. In addition, informative loss is proposed to mine the informative samples and focus the trained model to learn enough knowledge of these samples which are difficult to predict by other models. Therefore, IA-net achieves best performance.

4.5 Quantitative evaluation

Table 1 gives the qualitative evaluation results of IA-net. From this table, we can see that IA-net can achieve precise segmentation results, TPVF, DSC and FPVF are 0.9384, 0.8662 and 0.0043 respectively. For each image, if the proportion of CNV pixels in total pixels in image is less than three percentages, the CNV is small CNV. Otherwise, the CNV is large CNV. For small cases and large cases, DSC, TP, FP of small CNVs are 0.8351 ± 0.0905 , 0.08617 ± 0.0923 , 0.0038 ± 0.0017 while large CNVs are 0.9181 ± 0.0373 , 0.9241 ± 0.0349 , 0.0049 ± 0.0009 . The segmentation performances on large CNVs is better than small CNVs because large CNVs have more enough information to be discriminated. Some typical CNV segmentation results of the proposed method are shown in Fig.8. As shown in this figure, most of CNV pixels can be segmented accurately. However, IA-net failed to achieve finer boundary segmentation. Compared with informative pixels in CVN region, the number of boundary pixels is too small, and some boundary pixels have some different characteristics with informative pixels. Therefore, it is difficult for the model to learn enough knowledge of these boundary pixels, result in the inaccurate segmentation of boundary.

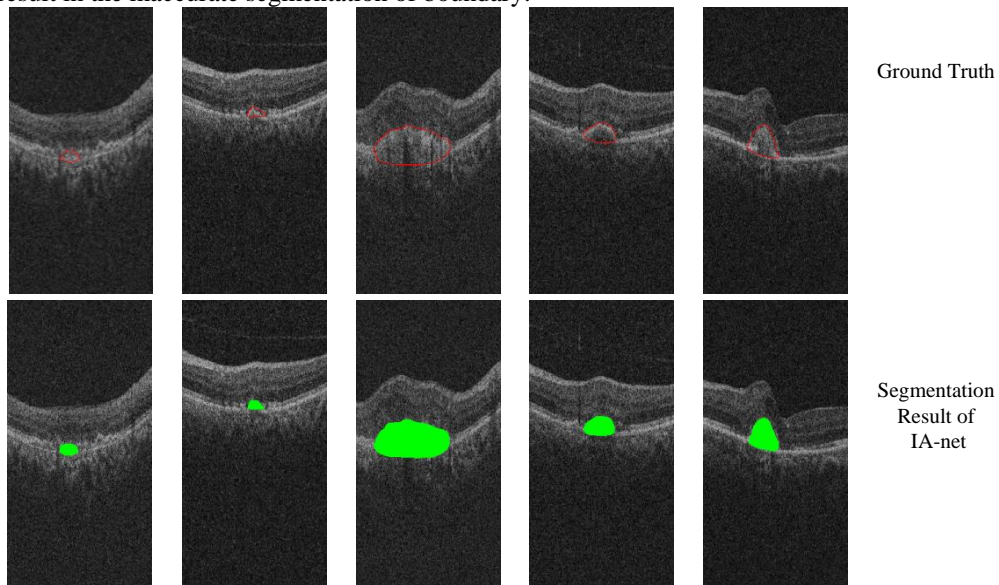


Fig. 8. Typical CNV segmentation results of the proposed method

5. Conclusion

This paper proposes IA-net for CNV segmentation in OCT images. The proposed method has ability to obtain the accurate CNV delineation result. The results will be provided to doctors. Based on the results, they can acquire the properties of CNV lesion, including the area, volume, width, height, optical density value, etc. These properties play an important role in diagnosis and treatment of CNV.

The novel attention enhancement block and informative loss are developed in the proposed network. We evaluate the performance of the proposed method on our database. The experimental results demonstrate that the proposed method significantly outperforms existing CNV segmentation methods such as ALM-LSP[23],MS-CNN[26], NNCGS[3] and U-net.

In the attention enhancement block, the discriminative attention map is embedded to guide the feature map learning process. Traditional methods fail to specially focus small CNV in the feature map learning process, result in the information loss of CNV features. The proposed attention enhance block is proposed to guarantee that the attention of CNV is higher while attention of background is lower. It is useful to improve the discriminative ability of CNV features in the low-level feature maps and preserve the feature information of CNV in the high-level feature maps. Therefore, compared with traditional convolutional block, the proposed attention enhancement block has the ability to focus the interesting region, improve the discriminative ability of the learned features of small CNV, which is useful to improve the segmentation accuracy on small CNV.

The proposed informative loss is proposed to deal with informative examples effectively. Focal loss has the ability to deal with hard examples. However, its performance is independent on the manually tuned hyperparameters. It's time-consuming to obtain the optimal hyperparameters. Different from focal loss [39], the proposed informative loss can assign the attention to each sample automatically, eliminating manual parameters tuning. In addition, the proposed method only focuses on the informative samples which belong to CNV. Therefore, the proposed loss can achieve better performance on the CNV pixels which are difficult to be predicted.

However, for 3D OCT data segmentation task, the proposed network is 2-D, ignoring relationship information between neighboring slices. In the future, we will extend our network by introducing the idea of three modules [35] to segment transaxial, coronal, and sagittal OCT slices.

Funding

Natural Science Foundation of China (61701280), National Key R&D Program of China (2018YFC0830100, 2018YFC0830102), Natural Science Foundation of China (61801263, 61876098).

Disclosures

The authors declare that there are no conflicts of interest related to this article.

References

1. Y. Jia, S. T. Bailey, D. J. Wilson, and O. Tan, "Quantitative Optical Coherence Tomography Angiography of Choroidal Neovascularization in Age-Related Macular Degeneration," *Ophthalmology* **121**(7), 1435–1444 (2014).
2. Abdelmoula, W., Shah, S., Fahmy, A.S., "Segmentation of choroidal neovascularization in fundus fluorescein angiograms," *IEEE Transactions on Biomedical Engineering* **60**(5), 1439-1445 (2013).
3. Xiang, D., Tian, H., Yang, X., Shi, F., Zhu, W., Chen, H., Chen, X. "Automatic Segmentation of Retinal Layer in OCT Images With Choroidal Neovascularization," *IEEE Transactions on Image Processing* **27**(12), 5880-5891 (2018).
4. A. Dewan, M. Liu, S. Hartman, S.S. Zhang, D.T. Liu, C. Zhao, P.O. Tam, W.M. Chan, D.S. Lam, M. Snyder. "HTRA1 promoter polymorphism in wet age-related macular degeneration," *Science* **314**(5801), 989-992 (2006).

5. Puliafito, C. A., Hee, M. R., Lin, C. P., Reichel, E., Schuman, J. S., Duker, J. S., et al., "Imaging of macular diseases with optical coherence tomography," *Ophthalmology* **102**(2), 217–229, (1995).
6. Hee, M. R., Bauman, C. R., Puliafito, C. A., Duker, J. S., Reichel, E., Wilkins, J. R., et al., "Optical coherence tomography of age-related macular degeneration and choroidal neovascularization," *Ophthalmology* **103**(8), 1260-1270, (1996).
7. Xinjian Chen, Li Zhang, Elliott H. Sohn, Kyungmoo Lee, Meindert Niemeijer, John Chen, Milan Sonka, Michael D. Abramoff, "Quantification of External Limiting Membrane Disruption Caused by Diabetic Macular Edema from SD-OCT", *Investigative Ophthalmology & Visual Science* **53**(13), 8042-8048 (2012).
8. Fei Shi, Xinjian Chen, Heming Zhao, Weifang Zhu, Dehui Xiang, Enting Gao, Milan Sonka Haoyu Chen, "Automated 3-D Retinal Layer Segmentation of Macular Optical Coherence Tomography Images with Serous Pigment Epithelial Detachments," *IEEE Transactions on Medical Imaging* **34**(2), 441-452 (2015).
9. C. L. Tsai, Y. L. Yang, S. J. Chen, K. S. Lin, C. H. Chan, W. Y. Lin, "Automatic characterization of classic choroidal neovascularization by using adaboost for supervised learning," *Invest Ophthalmol & visual science* **52**(5), 2767–2774 (2011).
10. Gu, J., Wang, Z., Kuen, J., Ma, L., Shahroudy, A., Shuai, B., Chen, T. "Recent advances in convolutional neural networks," *Pattern Recognition* **77**, 354-377 (2018).
11. Hong, I., Hwang, Y., Kim, D. "Efficient deep learning of image denoising using patch complexity local divide and deep conquer," *Pattern Recognition* **96**, 106945 (2019).
12. Mehta, J., A. Majumdar. "RODEO: Robust DE-aliasing autoencoder for Real-time Medical Image Reconstruction," *Pattern Recognition* **63** 499-510 (2017).
13. Li, F., Bai, H., Zhao, Y. "Detail-preserving image super-resolution via recursively dilated residual network," *Neurocomputing* **358**, 285-293(2019).
14. Stegmann, H., Werkmeister, R. M., Pfister, M., Garhöfer, G., Schmetterer, L., Dos Santos, V. A. "Deep learning segmentation for optical coherence tomography measurements of the lower tear meniscus," *Biomedical Optics Express* **11**(3), 1539-1554 (2020).
15. Wang, C., Gan, M., Zhang, M., Li, D. "Adversarial convolutional network for esophageal tissue segmentation on OCT images," *Biomedical Optics Express* **11**(6), 3095-3110 (2020).
16. Rong, Y., Xiang, D., Zhu, W., Shi, F., Gao, E., Fan, Z., Chen, X. "Deriving external forces via convolutional neural networks for biomedical image segmentation," *Biomedical optics express*, **10**(8), 3800-3814 (2019).
17. Wang, J., Hormel, T. T., Gao, L., Zang, P., Guo, Y., Wang, X., Jia, Y. "Automated diagnosis and segmentation of choroidal neovascularization in OCT angiography using deep learning," *Biomedical Optics Express* **11**(2), 927-944 (2020).
18. Li, Z., Peng, C., Yu, G., Zhang, X., Deng, Y., Sun, J "Detnet: Design backbone for object detection," *Proceedings of the European Conference on Computer Vision (ECCV)*, (2018).
19. Reza Rasti, Michael J. Allingham, Priyatham S. Mettu, Sam Kavusi, Kishan Govind, Scott W. Cousins, and Sina Farsi, "Deep learning-based single-shot prediction of differential effects of anti-VEGF treatment in patients with diabetic macular edema," *Biomedical Optics Express* **11**, 1139-1152 (2020)
20. Zheng, H., Fu, J., Mei, T., Luo, J. "Learning Multi-attention Convolutional Neural Network for Fine-Grained Image Recognition," *IEEE International Conference on Computer Vision (ICCV)*, (2017).
21. Yang, F., Yan, K., Lu, S., Jia, H., Xie, X., Gao, W. "Attention driven person re-identification," *Pattern Recognition* **86**, 143-155 (2019).
22. Zhou, S., Wang, J., Meng, D., Liang, Y., Gong, Y., Zheng, N. "Discriminative feature learning with foreground attention for person re-identification," *IEEE Transactions on Image Processing* **28**(9), 4671-4684 (2019).
23. Xi, X., Meng, X., Yang, L., Nie, X., Yu, Z., Zhang, C., Yin, Y., Chen, X. "Learned local similarity prior embedding active contour model for choroidal neovascularization segmentation in optical coherence tomography images," *Science China (Information Sciences)* **61**(9), 319-326 (2018).
24. Zhu, S., Shi, F., Xiang, D., Zhu, W., Chen, H., Chen, X. "Choroid Neovascularization Growth Prediction with Treatment Based on Reaction-Diffusion Model in 3-D OCT Images," *IEEE Journal of Biomedical and Health Informatics* **21**(6)1667-1674, (2017).
25. Li, Y., Niu, S., Ji, Z., Fan, W., Yuan, S., Chen, Q. "Automated Choroidal Neovascularization Detection for Time Series SD-OCT Images," *In International Conference on Medical Image Computing and Computer-Assisted Intervention (MICCAI)*, (2018).
26. Xi, X., Meng, X., Yang, L., Nie, X., Yang, G., Chen, H., Yin, Y., Chen, X.. "Automated segmentation of choroidal neovascularization in optical coherence tomography images using multi-scale convolutional neural networks with structure prior," *Multimedia Systems* **25**(2), 95-102 (2019).
27. J.Xu, L Xiang, Q Liu, H Gilmore, J Wu, Tang J, Madabhushi A. "Stacked sparse autoencoder (SSAE) for nuclei detection on breast cancer histopathology images," *IEEE Transactions on Medical Imaging* **35**(1), 119–130 (2016).
28. Hou, L., Nguyen, V., Kanevsky, A. B., Samaras, D., Kurc, T. M., Zhao, T., Saltz, J. H. . "Sparse Autoencoder for Unsupervised Nucleus Detection and Representation in Histopathology Images," *Pattern Recognition* **86**, 188-200 (2019).
29. van Tulder, G. and M. de Bruijne. "Combining Generative and Discriminative Representation Learning for Lung CT Analysis with Convolutional Restricted Boltzmann Machines," *IEEE Transactions on Medical Imaging* **35**(5), 1262-1272 (2016).

30. Viergever MA, Mendrik AM, de Vries LS, Benders MJ, Isgum I, "Automatic segmentation of MR brain images with a convolutional neural network," *IEEE Transactions on Medical Imaging* **35**(5), 1252–1261 (2016).
31. Zhang, W., Li, R., Deng, H., Wang, L., Lin, W., Ji, S., & Shen, D. "Deep convolutional neural networks for multi-modality isointense infant brain image segmentation," *NeuroImage* **108**, 214–224 (2015).
32. Long J, Shelhamer E, Darrell T. "Fully convolutional networks for semantic segmentation," *Proceedings of the IEEE conference on computer vision and pattern recognition (CVPR)*, (2015).
33. Ronneberger O, Fischer P, Brox T. "U-net: Convolutional networks for biomedical image segmentation," *International Conference on Medical image computing and computer-assisted intervention (MICCAI)*, (2015).
34. Chen, S., Ding, C., & Liu, M.. "Dual-force convolutional neural networks for accurate brain tumor segmentation," *Pattern Recognition* **88**, 90-100 (2019).
35. Wei, J., Xia, Y., Zhang, Y. "M3Net: A Multi-Model, Multi-Size, and Multi-View Deep Neural Network for Brain Magnetic Resonance Image Segmentation," *Pattern Recognition* **91**, 366-378. (2019).
36. Milletari F, Navab N, Ahmadi S A. "Fully convolutional neural networks for volumetric medical image segmentation," *International Conference on 3D Vision (3DV)*, (2016).
37. Yin, J., Xia, P., He, J. "Online Hard Region Mining for Semantic Segmentation," *Neural Processing Letters* **50**(3), 2665-2679, (2019).
38. Wang, Q., Zhao, W., Zhang, R., Li, Z., Cui, S. "Progressive Abdominal Segmentation with Adaptively Hard Region Prediction and Feature Enhancement," *International Symposium on Biomedical Imaging (ISBI)*, (2020).
39. Lin, T. Y., Goyal, P., Girshick, R., He, K., Dollár, P. "Focal loss for dense object detection," *In Proceedings of the IEEE international conference on computer vision (CVPR)*, (2017).
40. Roth, P. M., Hirzer, M., Köstinger, M., Belezni, C., Bischof, H.. *Mahalanobis Distance Learning for Person Re-identification*. Springer, London, 247-267 (2014).
41. Junejo, K. N., Karim, A., Hassan, M. T., Jeon, M. "Terms-based discriminative information space for robust text classification," *Information Sciences* **372**, 518-538 (2016).
42. Du, B., Wang, Z., Zhang, L., Zhang, L., Liu, W., Shen, J., Tao, D. "Exploring representativeness and informativeness for active learning," *IEEE transactions on cybernetics* **47**(1), 14-26 (2017).
43. Huang S J, Jin R, Zhou Z H. "Active Learning by Querying Informative and Representative Examples," *IEEE Transactions on Pattern Analysis and Machine Intelligence* **10**, 1936-1949, (2014).
44. Fu, Y., Zhu, X., Elmagarmid, A. K.. "Active Learning With Optimal Instance Subset Selection," *IEEE Transactions on Cybernetics* **43**(2), 464-475, (2013).
45. Abramoff M D, Garvin M K, Sonka M. "Retinal imaging and image analysis," *IEEE reviews in biomedical engineering* **3**, 169-208, (2010).