ARTICLE
Biomedical Engineering Communications 2026;5(2):9. https://doi.org/10.53388/BMEC2026009

C-SegNet: a practical approach for automated diabetic macular edema segmentation in optical coherence tomography images

Zhi-Yuan Guan[1#], Ge Deng[1#], Shi-Long Shi[1, 2], Zhen Tang[1], Xian-Kun Dong[1], Qiu-Yi Li[1], Shu-Jing Shen[3*], Yong-Ling He[4*], Xue-Jun Qiu[1, 2*]

[1]College of Health Science, Guangdong Pharmaceutical University, Guangzhou 510006, China.
[2]Guangdong Provincial Engineering and Technology Research Center of Light and Health, Guangzhou 510006, China.
[3]Department of Rehabilitation Medicine, People's Hospital of Yingde City Guangdong Province, Yingde 513000, China.
[4]College of Medical Information Engineering, Guangdong Pharmaceutical University, Guangzhou 510006, China.

[*]**Correspondence to:** Shu-Jing Shen, Department of Rehabilitation Medicine, People's Hospital of Yingde City Guangdong Province, No. 2 Jiaoyu East Road, Yingde 513000, China. E-mail: shujingsee@163.com. Yong-Ling He, College of Medical Information Engineering, Guangdong Pharmaceutical University, No. 280 Outer Ring East Road, Guangzhou 510006, China. E-mail: heyongling@gdpu.edu.cn. Xue-Jun Qiu, College of Health Science, Guangdong Pharmaceutical University, No. 280 Outer Ring East Road, Guangzhou 510006, China. E-mail: qiuxuejun@gdpu.edu.cn.

Author contributions
Guan ZY and Deng G are credited with substantial contributions to the study's design, drafting, and data analysis. Qiu XJ, Shen SJ and He YL were involved in interviewing and editing the manuscript. The document was revised critically for important intellectual content by Shi SL, Tang Z, Dong XK, Li QY. All authors participated in this study, read, and approved the final manuscript.

Competing interests
The authors declare no conflicts of interest.

Acknowledgments
This work was financially supported by the Guangdong Pharmaceutical University 2024 Higher Education Research Projects (GKP202403, GMP202402), and the Guangdong Pharmaceutical University College Students' Innovation and Entrepreneurship Training Programs (Grant No.202504302033, 202504302034, 202504302036, and 202504302244).

Peer review information
Biomedical Engineering Communications thanks Yan-Fei Li and all anonymous reviewers for their contribution to the peer review of this paper.

Abbreviations
OCT, optical coherence tomography; CNN, convolutional neural network; CAM, channel attention module; CBAM, convolutional block attention module; DME, diabetic macular edema; SAM, spatial attention module.

Citation
Guan ZY, Deng G, Shi SL, et al. C-SegNet: a practical approach for automated diabetic macular edema segmentation in optical coherence tomography images. *Biomed Eng Commun*. 2026;5(2):9. doi: 10.53388/BMEC2026009.

Executive editor: Xin-Yun Zhang.
Received: 13 June 2025; **Revised:** 14 July 2025; **Accepted:** 24 July 2025; **Available online:** 22 August 2025.

Abstract
Background: Diabetic macular edema is a prevalent retinal condition and a leading cause of visual impairment among diabetic patients' Early detection of affected areas is beneficial for effective diagnosis and treatment. Traditionally, diagnosis relies on optical coherence tomography imaging technology interpreted by ophthalmologists. However, this manual image interpretation is often slow and subjective. Therefore, developing automated segmentation for macular edema images is essential to enhance to improve the diagnosis efficiency and accuracy. **Methods:** In order to improve clinical diagnostic efficiency and accuracy, we proposed a SegNet network structure integrated with a convolutional block attention module (CBAM). This network introduces a multi-scale input module, the CBAM attention mechanism, and jump connection. The multi-scale input module enhances the network's perceptual capabilities, while the lightweight CBAM effectively fuses relevant features across channels and spatial dimensions, allowing for better learning of varying information levels. **Results:** Experimental results demonstrate that the proposed network achieves an IoU of 80.127% and an accuracy of 99.162%. Compared to the traditional segmentation network, this model has fewer parameters, faster training and testing speed, and superior performance on semantic segmentation tasks, indicating its highly practical applicability. **Conclusion:** The C-SegNet proposed in this study enables accurate segmentation of Diabetic macular edema lesion images, which facilitates quicker diagnosis for healthcare professionals.

Keywords: multi-scale input; diabetic macular edema; image segmentation; optical coherence tomography

Introduction

Optical coherence tomography (OCT) imaging is a cutting-edge, non-contact, and non-invasive diagnostic technique with significant potential for diagnosing and treating various organ diseases, particularly retinal conditions [1, 2]. Accurate diagnosis of retinal diseases usually relies on analyzing the ten layers of retinal structures depicted in OCT images, making precise identification of these layers crucial for effective disease management.

Historically, OCT image segmentation employed traditional image processing technologies such as threshold segmentation and graph search method [3, 4]. However, these methods are often sensitive to noise and lack reliability when addressing variations between images. With the advancement in technology, deep learning methods have emerged as the dominant approach for OCT image segmentation, supplanting traditional techniques. These modern methods encompass a variety of new techniques including support vector machines, convolutional neural network (CNN), and random forest classifiers [5–7]. In 2015, Ronneberger et al. introduced the U-Net model, which effectively addresses the automatic identification and segmentation of lesions in medical images [8]. Following this, the fully convolutional network model was proposed in 2017 by Shelhamer et al., offering flexibility in image input size and converting segmentation images to pixel-level representations [9]. SegNet, developed in 2016 at Cambridge for semantic segmentation tasks, modified the VGG-16 network but faced challenges in retaining neighboring information during low-resolution feature mapping [10]. The U-Net architecture, renowned for its impact on biomedical image analysis, is widely used for tasks such as neuronal structure and cell segmentation. However, its training phase requires more computational resources due to a higher number of learnable parameters compared to SegNet. Adiga et al. introduced the M-net architecture, which enhances fingerprint denoising but tends to lose local information when processing complete images [11]. All the three architectures of SegNet, U-Net, and M-net utilize encoder and decoder components, where the encoder progressively extracts features through convolutional and pooling layers, while the decoder part restores detailed features from various resolution levels. However, this symmetric structure often fails to recover lost information during up sampling.

To address these limitations, the attention mechanism was proposed by Bahdanau, Cho, and Bengio in 2014, enhancing the capability of Encoder-Decoder framework to capture key features by allowing more flexibility in processing the input information [12]. Subsequent developments, such as position attention module and channel attention module (CAM), further refined the capability to capture global information from spatial and channel dimensions [13]. The convolutional block attention module (CBAM) was later introduced by Sanghyun Woo et al. enhancing the perceptual capabilities of CNNs without increasing network complexity [14]. Zhou et al. introduced the U-net++ architecture, characterized by the interconnected encoder and decoder blocks that facilitate deeper network structures and enhanced information transfer [15].

The main contributions of this work are as follows:

(1) We propose an enhanced network architecture of SegNet, termed C-SegNet, which integrates jump connections and the CBAM attention mechanism to efficiently extract features from both channel and spatial dimensions.

(2) The model adopts multi-scale input for fusion features, enhancing the object perception across different scales. We evaluate the proposed C-SegNet on a public OCT dataset, demonstrating superior segmentation performance compared to several established deep networks.

Related works

In recent years, numerous studies have focused on the segmentation of retina and lesions in OCT scan images. To solve the issue of limited labelled data, Wang proposed a deep semi-supervised multi-instance learning framework to improve diabetic macular edema (DME) classification accuracy by using a small amount of roughly labelled data alongside a large amount of unlabeled data [16]. Zhang et al. developed a twin self-supervised semi-supervised learning method that utilizes both generative and discriminative self-supervised strategies to learn from a limited number of labelled images and a large number of unlabeled ones [17]. Significant advancements have been made in OCT lesion classification. Das et al. used a deep multiscale fusion convolutional neural network to effectively integrate features from various scales for retinal disease classification [18]. Rajagopalan et al. optimized hyper-parameters, including batch size and loss rate, using stochastic search methods to improve the classification of three common retinal diseases [19]. Suni ja et al. improved training efficiency by introducing down sampling and weight sharing in a deep convolutional neural network with six convolutional blocks, focusing on the classification of three distinct lesions from OCT images [20].

For single lesion region segmentation, researchers aimed to identify network models that offer high generalization ability, segmentation accuracy, and simple structure. proposed a method for detecting oedema and identifying different types of DME by detecting and processing the pixel values of the region of interest for cystic region segmentation. Roy et al. proposed ReLayNet, which enhances segmentation by transferring intermittent feature representations from encoder blocks to corresponding decoder blocks via skip connections, optimizing training with cross-entropy and Dice overlap loss functions [21].

Diao et al. developed a deep learning framework that leverages dual guidance between two tasks [22]. In this framework, a class of activation maps is used to guide the UNet (referred to as CAM-UNet) for segmenting glass wart and choroidal neovascularization lesions. Cao et al. proposed a U-shaped network that incorporates a self-attention mechanism to obtain a locally maximal receptive field by introducing the Transformer's self-attention mechanism [23]. Xu et al. designed a two-stage learning framework based on deep neural network for the automatic segmentation of pigment epithelial detachment in polypoidal choroidal vasculopathy patients, demonstrating effective segmentation results with their dual-framework deep neural network approach [24].

In other related research on medical image segmentation, Nagaraj et al. proposed a novel M-SegNet network featuring global attention [25]. This architecture employs convolutional kernels of varying sizes in both the encoding and decoding parts, incorporating jump connections with global attention to recover the spatial information lost in the down sampling process. Liu et al. proposed an enhanced nested U-Net architecture, which integrates multi-scale input, multi-scale side output, and a dual-attention mechanism to improve segmentation performance [26].

Methods

The general framework of the proposed C-SegNet is illustrated in Figure 1. This network adopts an encoder-decoder architecture that retains the core features of SegNet while incorporating multi-scale inputs and the CBAM attention mechanism. Additionally, it employs combinatorial connectivity to directly link the feature maps of the encoder to those of the decoder. A detailed description of the encoder and decoder structures follows.

The encoder consists of two consecutive 3 × 3 convolutions and a maximum pooling layer with a step size of 2. The convolutional layer uses a feed forward network to connect the output of layer (l-1) to layer l as the input of layer l. The output of layer (l-1) is used as the input of layer l to layer l. The formula is shown below:

$$a^{(l)} = f(w^{(l)} \otimes a^{(l-1)} + b^{(l)}) \quad (1)$$

Where $a^{(l)}$ is the output of layer l, w is the weight index, $b^{(l)}$ represents the bias parameter, and f (x) indicates that ReLU activation function processing is performed. Following the convolutional layers,

a down sampling operation is performed using a 2 × 2 maximum pooling layer with a stride of 2, which connects to the corresponding decoder layer via combinatorial connections. These jump connections facilitate the direct transfer of features from the encoder to the decoder, enabling the decoder to leverage both rich semantic and low-level detail information. This design also offers a more direct gradient pathway, mitigating the vanishing gradient problem and enhancing model convergence speed.

Decoder:

The decoder structure mirrors that of the encoder, with each layer consisting of two consecutive convolutional blocks of 3 × 3 kernels. Unlike the encoder, the decoder does not employ max pooling for up sampling, which reduces the number of parameters. Instead, the un pooling operation increases the receptive field size and enhances segmentation accuracy by restoring the spatial resolution and information of the feature map. The calculation for the un pooling operation is given by the calculation formula is as follows:

$$a^{(l)} = unpool(F(a^{(l-1)} * w^{(l)} + b^{(l)}), p^{(l)}\,\mathrm{ind}) \quad (2)$$

Where $a^{(l)}$ is the output of the un pooling operation, $F(x)$ represents the ReLU activation function of the convolution module, and $p^{(l)}\,\mathrm{ind}$ represents the pooling index passed from the encoder to the decoder. The un pooling operation aims to restore the spatial resolution of the feature map to a higher level, utilizing the positional features mapped from the encoder to aid in reconstructing the feature map during the decoding stage.

Multi-scale input

This network implements multi-scale inputs by utilizing image pyramids, enabling the transfer of information from images of varying resolutions into the encoding layer through jump connections [27]. It employs a maximum pooling operation to down sample images, enhancing the network's translational invariance without significantly increasing computational demands, thereby expanding the receptive field to some extent.

CBAM combined connection

The CBAM integrates both channel and spatial attention mechanisms. Unlike SegNet, which primarily focuses on channel attention and may overlook essential spatial features, CBAM effectively emphasizes critical features while suppressing irrelevant ones across both dimensions [10]. The CBAM architecture consists of a CAM and a spatial attention module (SAM) operating sequentially, with input feature layers undergoing processing by each attention mechanism. The structure is illustrated in Figure 2.

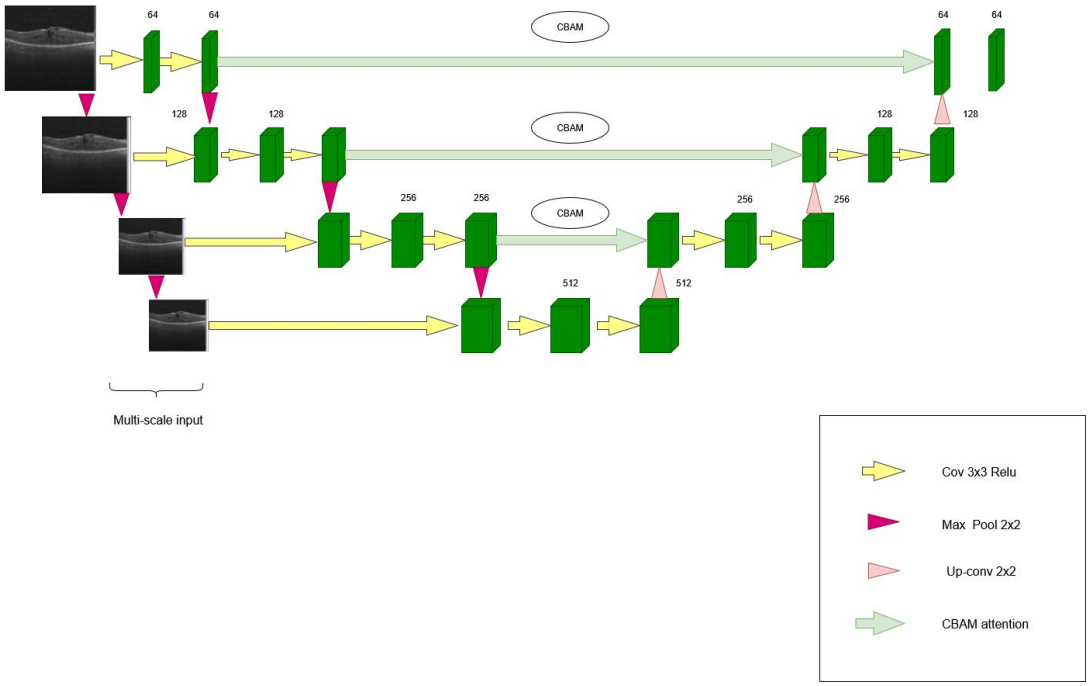

Figure 1 Architecture of the proposed C-SegNet model.
CBAM, convolutional block attention module.

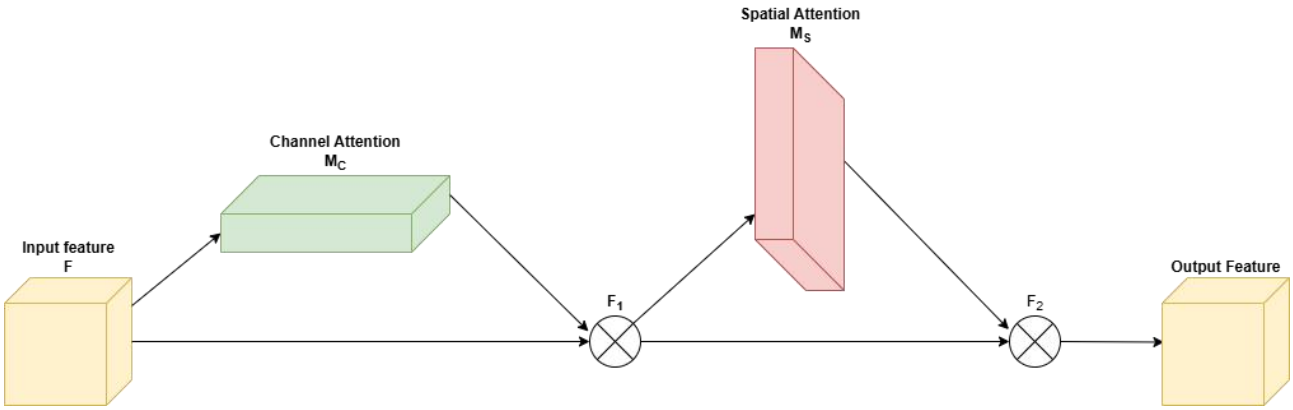

Figure 2 CBAM structure.
CBAM, convolutional block attention module.

CAM

The CAM is designed to assess the significance of feature maps along the channel dimension, enabling the model to concentrate on the most relevant feature channels. Its architecture is depicted in Figure 3.

Normalization uses the sigmoid function. In the channel attention module, for the input feature F, it is fed into a fully connected layer (MLP) after global average pooling and global maximum pooling respectively, then the global average pooling outputs and global maximum outputs are subjected to an addition operation, and finally normalized to form the weight matrix on the channel. The weight matrix calculation process is shown in the following equation:

$$M_c(F) = \sigma\big(MLP(Avgpool(F)) + MLP(MaxPool(F))\big) \quad (3)$$

The sigmoid function is defined as:

$$\sigma(x) = \frac{1}{1+e^{-x}} \quad (4)$$

Where x is the eigenvalue of the input. The output range of the Sigmoid function is (0,1), so it normalizes the input value to a probability value between 0 and 1. In this way, the model can focus on feature channels with higher weights, thereby extracting key information more effectively.

SAM

The SAM is used to capture the importance of the feature map in the spatial dimension so that the model can focus on important spatial locations in the feature map. Firstly, the channel attention processing. Its structure is shown in Figure 4.

The generated feature maps with f as input to the SAM are subjected to channel-based maximum pooling and average pooling operations, and the two generated feature maps are spliced together to form a composite feature map. Subsequently, a convolution operation is applied to this combined feature map with a Sigmoid activation function, resulting in a single-channel feature map representing the attentional weights with the same dimensions as the input feature map. Finally, the spatial attention mechanism is implemented by multiplying the attention weight feature map with the input feature map element by element. The calculation formula is shown in Equation 5:

$$M_s(F) = \sigma\big(f^{7\times7}([\text{AvgPool}(F); \text{MaxPool}(F)])\big) \quad (5)$$

Among them, f^{7*7} represents a 7*7 violation of the fusion feature. M_s represents the violation of the generated spatial features.

Finally, the attention-weighted feature map M_s is multiplied element-by-element with the input feature f from the initial module. The weighting operation on the input feature map is completed. The final CBAM attention calculation formula is shown in Equation 6 and 7:

$$F_1 = M_c(F) \otimes F \quad (6)$$

$$F_2 = M_s(F_1) \otimes F_1 \quad (7)$$

Experiments

The dataset utilized in this study was sourced from the Kaggle public DME dataset, specifically the Intraretinal Cystoid Fluid collection, which comprises 1000 OCT images. These images were meticulously selected and labeled by medical experts at Liaquat University of Medical and Health Sciences in Jamshoro [28]. The OCT images were acquired with the informed consent of patients at the Jamshoro Eye Institute. The selected images encompass a diverse range of cystoid macular edema (CME) regions, including various shapes and sizes, thereby ensuring comprehensive training for the identification of cystoid macular edema.

Experiments setting

The dataset comprising 1000 OCT images was randomly shuffled and divided into training, validation, and test sets with a ratio of 7:1.5:1.5. Stratified sampling was applied to ensure a balanced distribution of lesion types across the three subsets. To prevent data leakage, all image augmentation techniques such as random Gaussian noise addition and flipping were applied only to the training set. No augmented images were used in the validation or test sets. Furthermore, we conducted 5-fold cross-validation to evaluate the model's robustness. In each fold, 80% of the data was used for training and 20% for testing, and the reported metrics represent the average performance across all folds.

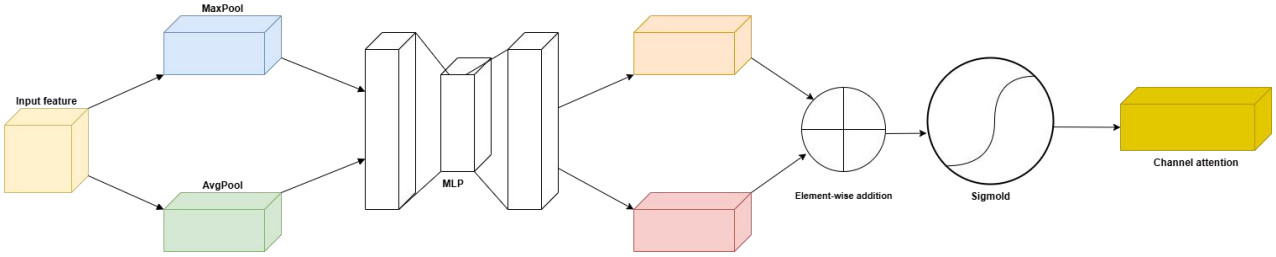

Figure 3 CAM structure.
CAM, channel attention module; MLP, multi-layer perceptron.

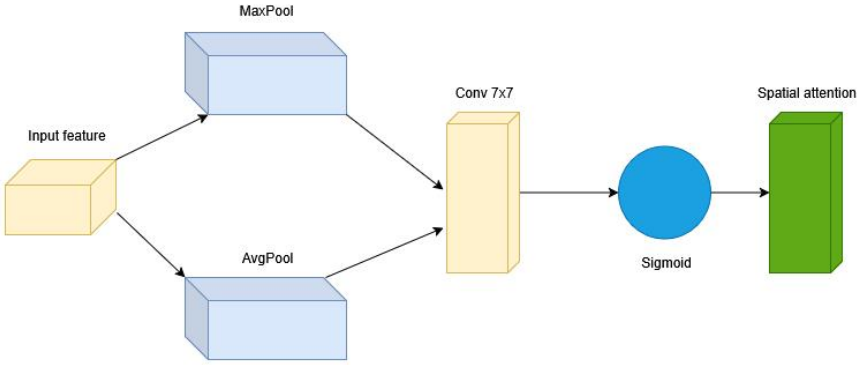

Figure 4 SAM structure.
SAM, spatial attention module.

Evaluation metrics

To test the effectiveness of different networks in the segmentation study of retinal layers and lesions in OCT scan images, this paper mainly uses IoU and accuracy to evaluate and compare the overall segmentation effect. The evaluation formula is as follows:

$$IoU = \frac{TP}{FN + FP + TP} \tag{8}$$

$$Accuracy = \frac{TP}{FN + TP} \tag{9}$$

$$SP = \frac{TP}{TN + FP} \tag{10}$$

$$SE = \frac{TP}{TP + FN} \tag{11}$$

Where TP, FP, TN, and FN denote true positives, false positives, true negatives, and false negatives, respectively. The TP portion is the intersection of the true and predicted values, while the portion of FN + FP + TP is the concatenation of the true and predicted values. The Intersection over the IoU is defined as the ratio of the intersecting set to the union of the true and predicted values. Accuracy, in this context, refers to the average pixel accuracy, which is calculated as the mean proportion of correctly segmented pixels.

Comparative experiments with other methods

To evaluate the performance of the proposed network model, we conducted comparative tests with UNet, SegNet, and UNet++ models on the same dataset [8]. The experiments utilized torch stat, a lightweight network analysis tool within the PyTorch framework, to compute the parameters of each model. All networks were trained under identical conditions. The training results were also compared with those of the latest networks. As shown in Table 1, the proposed C-SegNet demonstrates superior segmentation performance.

As indicated in the Table 1, C-SegNet achieves IoU and Accuracy scores of 80.127% and 99.161%, respectively, significantly outperforming the previous three networks. UNet and UNet++ exhibit lower IoU values due to the loss of spatial information during the up sampling process in the encoder and their failure to capture relevant contextual details. In contrast, C-SegNet enhances gradient propagation by integrating jump connections with the CBAM attention mechanism, facilitating the transfer of spatial information from the encoder to the decoder at all levels. Additionally, the multi-scale input module enriches the learned features. According to the data presented, UNet++ has the lowest parameter count at 9.16 M, while UNet has the highest at 31.01M. SegNet features fewer parameters than UNet, and the incorporation of the CBAM lightweight attention mechanism only increases C-SegNet's parameter count by 2.88 M compared to SegNet. While the number of parameters partially influences the computational burden, UNet++'s reduced parameter count does not mitigate its increased computational complexity resulting from numerous skip connections. The average test speed recorded reflects the time each model requires to complete testing on the same dataset, revealing that SegNet has the fastest test speed. C-SegNet maintains a relatively high speed while achieving improved accuracy.

To further validate the effectiveness of the proposed segmentation method, we conducted comparative tests with three other segmentation networks recently developed for image segmentation tasks, many of which have demonstrated superior performance. The segmentation results of these networks are illustrated in Figure 5, with regions exhibiting the largest discrepancies highlighted by white boxes. The labels A, B, and C in the figure correspond to three types of DME OCT images. It is evident from Figure 5 that SegNet's segmentation performance is inferior to that of the other three networks. This is attributed to SegNet's lack of jump connections, which leads to greater loss of feature information during up sampling reliant solely on pooled indices. Consequently, this method is more prone to omissions when segmenting complex medical images. While the differences in segmentation results between UNet, UNet++, and the ground truth are minimal across the three test images, Figure 5c reveals that based on qualitative analysis, some small lesion regions appear to be better segmented by C-SegNet compared to UNet and UNet++. However, further quantitative assessment is needed to isolate the effect of attention versus multi-scale input and skip connections.

Ablation experiment

To validate the proposed model's efficacy, we conducted ablation experiments as detailed in Table 2. We evaluated the initial SegNet, SegNet with combinatorial connectivity, and SegNet with only multi-scale inputs. The results demonstrate a significant improvement in segmentation accuracy following the introduction of CBAM combinatorial connections in SegNet, with IoU increasing by 0.05 compared to SegNet with multi-scale inputs. This indicates that combinatorial connectivity is crucial for enhancing model performance. The multi-scale input module aids the network in comprehending global and local relationships within images by fusing contextual information from various scales. The synergy of these two modules effectively enhances overall network performance.

While the experimental results highlight C-SegNet's strong performance on a specific dataset, the model's generalization capability across different types of OCT images or lesion types remains to be thoroughly investigated. Future research will aim to confirm the model's robustness and applicability using a diverse range of datasets. The dataset used in this study was retrieved from a Kaggle repository titled Intraretinal Cystoid Fluid Segmentation, available at: https://www.kaggle.com/datasets/zeeshanahmed13/intraretinal-cystoid-fluid/data. The OCT images in this dataset were annotated by medical professionals at Liaquat University of Medical and Health Sciences.

To further isolate the contribution of CBAM, we conducted an additional experiment using SegNet with CBAM alone, excluding the multi-scale input module. As shown in Table 2, introducing CBAM alone resulted in a notable increase in IoU from 70.321% to 75.902%. This result supports the assertion that CBAM significantly enhances feature extraction by enabling the network to focus on critical spatial and channel-wise features.

It is worth noting that while our ablation study assessed the combined impact of CBAM, further decomposition into its CAM and SAM could provide more granular insights into their respective contributions. Additionally, integrating CBAM into other backbone networks such as UNet or UNet++ could offer a more rigorous comparison. These aspects will be explored in future work to comprehensively evaluate the versatility of the CBAM mechanism. To further assess the stability and generalization performance of the proposed model in macular edema image segmentation, we conducted a 5-fold cross-validation on the C-SegNet framework. The detailed quantitative results, including accuracy, sensitivity, specificity, and IoU across different folds, are presented in Table 3. The consistently high IoU values indicate that the model effectively captures lesion boundaries, demonstrating robust segmentation capability and strong potential for clinical application.

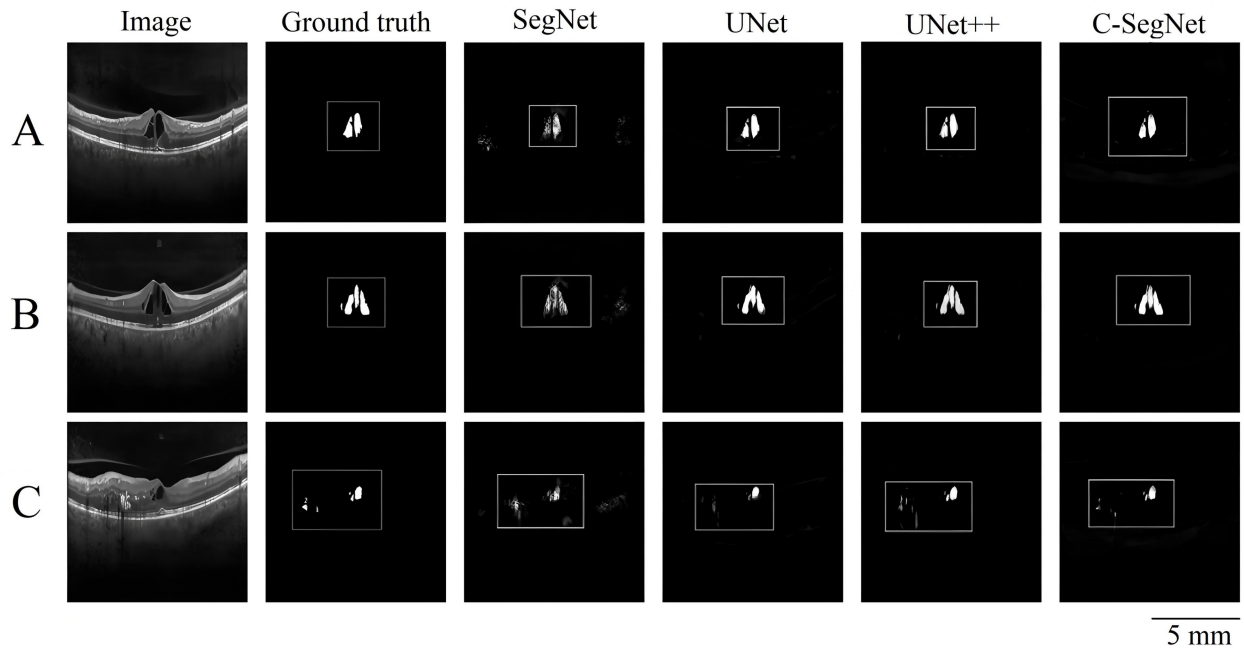

Figure 5 Prediction results of each segmentation network.
(A) DME lesion samples 1; (B) DME lesion samples 2; (C) DME lesion samples 3.

Table 1 Comparison of different network models

Model	IoU (%)	ACC (%)	SP (%)	SE (%)	Params (M)	Testing speed (s)
SegNet	70.321	98.923	99.938	88.094	22.44	17.13
UNet	76.392	99.110	99.972	92.524	31.01	22.30
UNet++	78.461	99.121	99.971	94.213	9.16	27.22
C-SegNet	80.127	99.162	99.963	95.126	25.32	20.92

ACC, accuracy; SP, spcificity; SE, sensitivity.

Table 2 Comparison of IoU, accuracy, training speed, testing speed of ablation experiments

Model	IoU (%)	ACC (%)	Training speed (s)	Testing speed (s)
SegNet	70.321	98.923	16.32	37.10
SegNet + CBAM	75.902	99.022	13.42	35.67
SegNet + CBAM (multi-scale)	77.735	99.110	12.77	35.12
SegNet + Multiscale input	73.261	98.996	15.38	33.13
C-SegNet	80.127	99.162	12.62	32.82

ACC, accuracy; CBAM, convolutional block attention module.

Table 3 Results of 5-fold cross-validation on the C-SegNet model

Fold	IoU (%)	ACC (%)	SP (%)	SE (%)
1	79.86	99.12	99.94	94.85
2	80.22	99.15	99.92	95.03
3	80.41	99.18	99.96	95.27
4	80.05	99.14	99.95	94.98
5	79.92	99.17	99.93	94.84
Mean	**80.09**	**99.15**	**99.94**	**94.99**
SD	**0.21**	**0.02**	**0.01**	**0.16**

SD, standard deviation; ACC, accuracy; SP, spcificity; SE, sensitivity.

Discussion and limitations

The proposed C-SegNet model demonstrates superior segmentation performance compared to traditional and attention-enhanced networks. The inclusion of CBAM and multi-scale input modules improved the model's ability to capture both contextual and fine-grained features, as reflected by the higher IoU and accuracy values.

However, several limitations should be acknowledged. First, the dataset used in this study comprises only 1000 OCT images from a single public source, which may not fully capture the diversity of clinical cases. Second, although the model benefits from data augmentation, external validation on independent datasets is needed to verify generalizability. Third, this work focuses exclusively on cystoid macular edema segmentation; the effectiveness of C-SegNet on other retinal lesions remains to be explored. Future research should address these aspects to improve robustness and clinical applicability.

Conclusion

In Summary, we present a novel segmentation algorithm for OCT retinopathy, designed to facilitate an end-to-end automatic segmentation model for OCT scan images. The proposed method enhances the original SegNet architecture introducing jump connections and CBAM attention, which sequentially applies channel and spatial attention mechanisms to strengthen critical features. Additionally, the integration of a multi-scale input module enables the model to obtain a broader range of receptive fields. The experimental results demonstrate that this method reaches an IoU of 80.127% and an accuracy of 99.161%. Furthermore, the model benefits from fewer parameters, resulting in faster training and testing. The C-SegNet proposed in this work demonstrates promising performance for the segmentation of DME lesions in OCT images. While the model shows potential to assist clinicians in improving diagnostic efficiency, further validation on diverse datasets is needed to confirm its generalizability across various OCT devices and lesion types.

References

1. Huang D, Swanson EA, Lin CP, et al. Optical Coherence Tomography. *Science.* 1991;254(5035):1178–1181. Available at: https://doi.org/10.1126/science.1957169
2. Qi L, Zheng K, Li X, Feng Q, Chen Z, Chen W. Automatic three-dimensional segmentation of endoscopic airway OCT images. *Biomed Opt Express.* 2019;10(2):642. Available at: https://doi.org/10.1364/BOE.10.000642
3. Chen Q, Leng T, Zheng L, et al. Automated drusen segmentation and quantification in SD-OCT images. *Med Image Anal.* 2013;17(8):1058–1072. Available at: https://doi.org/10.1016/j.media.2013.06.003
4. Wang C, Gan M, Yang N, et al. Fast esophageal layer segmentation in OCT images of guinea pigs based on sparse Bayesian classification and graph search. *Biomed Opt Express.* 2019;10(2):978. Available at: https://doi.org/10.1364/BOE.10.000978
5. Vapnik VN. *The Nature of Statistical Learning Theory*. 2nd ed. New York, NY: Springer;2000. Available at: https://doi.org/10.1007/978-1-4757-2440-0
6. Chiu SJ, Allingham MJ, Mettu PS, Cousins SW, Izatt JA, Farsiu S. Kernel regression based segmentation of optical coherence tomography images with diabetic macular edema. *Biomed Opt Express.* 2015;6(4):1172. Available at: https://doi.org/10.1364/BOE.6.001172
7. Karri SPK, Chakraborthi D, Chatterjee J. Learning layer-specific edges for segmenting retinal layers with large deformations. *Biomed Opt Express.* 2016;7(7):2888. Available at: https://doi.org/10.1364/BOE.7.002888
8. Ronneberger O, Fischer P, Brox T. U-Net: Convolutional Networks for Biomedical Image Segmentation. *Lecture Notes in Computer Science*. 2015:234–241. Available at: https://doi.org/10.1007/978-3-319-24574-4_28
9. Shelhamer E, Long J, Darrell T. Fully Convolutional Networks for Semantic Segmentation. *IEEE Trans Pattern Anal Mach Intell.* 2017;39(4):640–651. Available at: https://doi.org/10.1109/TPAMI.2016.2572683
10. Badrinarayanan V, Kendall A, Cipolla R. SegNet: A Deep Convolutional Encoder-Decoder Architecture for Image Segmentation. *IEEE Trans Pattern Anal Mach Intell.* 2017;39(12):2481–2495. Available at: https://doi.org/10.1109/TPAMI.2016.2644615
11. Adiga VS, Sivaswamy J. FPD-M-net: Fingerprint image denoising and inpainting using M-net based convolutional neural networks. *The Springer Series on Challenges in Machine Learning*. 2019:51–61. Available at: https://doi.org/10.1007/978-3-030-25614-2_4
12. Cho K, van Merriënboer B, Bahdanau D, Bengio Y. On the properties of neural machine translation: encoder – decoder approaches. *Proceedings of SSST-8, Eighth Workshop on Syntax, Semantics and Structure in Statistical Translation*. 2014:103–111. Available at: https://doi.org/10.3115/v1/W14-4012
13. Fu J, Liu J, Tian H, et al. Dual Attention Network for Scene Segmentation. *2019 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*. 2019:3141–3149. Available at: https://doi.org/10.1109/CVPR.2019.00326
14. Woo S, Park J, Lee JY, Kweon IS. CBAM: Convolutional block attention module. *Computer Vision – ECCV 2018*. 2018:3 – 19. Available at: https://doi.org/10.1007/978-3-030-01234-2_1
15. Zhou Z, Rahman Siddiquee MM, Tajbakhsh N, Liang J. UNet + +: A Nested U-Net Architecture for Medical Image Segmentation. *Deep Learning in Medical Image Analysis and Multimodal Learning for Clinical Decision Support*. 2018:3 – 11. Available at: https://doi.org/10.1007/978-3-030-00889-5_1
16. Wang X, Tang F, Chen H, Cheung CY, Heng P-A. Deep semi-supervised multiple instance learning with self-correction for DME classification from OCT images. *Med Image Anal.* 2023;83:102673. Available at: https://doi.org/10.1016/j.media.2022.102673
17. Zhang Y, Li M, Ji Z, et al. Twin self-supervision based semi-supervised learning (TS-SSL): Retinal anomaly classification in SD-OCT images. *Neurocomputing.* 2021;462:491–505. Available at: https://doi.org/10.1016/j.neucom.2021.08.051
18. Das V, Dandapat S, Bora PK. Automated Classification of Retinal OCT Images Using a Deep Multi-Scale Fusion CNN. *IEEE Sens J.* 2021;21(20):23256–23265. Available at: https://doi.org/10.1109/JSEN.2021.3108642
19. Rajagopalan N, N. V, Josephraj AN, E. S. Diagnosis of retinal disorders from Optical Coherence Tomography images using CNN. *PLoS One.* 2021;16(7):e0254180. Available at: https://doi.org/10.1371/journal.pone.0254180
20. A P S, Kar S, S G, Gopi VP, Palanisamy P. OctNET: A Lightweight CNN for Retinal Disease Classification from Optical Coherence Tomography Images. *Comput Methods Programs Biomed.* 2021;200:105877. Available at: https://doi.org/10.1016/j.cmpb.2020.105877
21. Roy AG, Conjeti S, Karri SPK, et al. ReLayNet: retinal layer and fluid segmentation of macular optical coherence tomography using fully convolutional networks. *Biomed Opt Express.* 2017;8(8):3627. Available at: https://doi.org/10.1364/BOE.8.003627
22. Diao S, Su J, Yang C, et al. Classification and segmentation of OCT images for age-related macular degeneration based on dual guidance networks. *Biomed Signal Process Control.* 2023;84:104810. Available at: https://doi.org/10.1016/j.bspc.2023.104810

23. Cao G, Wu Y, Peng Z, Zhou Z, Dai C. Self-attention CNN for retinal layer segmentation in OCT. *Biomed Opt Express.* 2024;15(3):1605. Available at: https://doi.org/10.1364/BOE.510464
24. Xu Y, Yan K, Kim J, et al. Dual-stage deep learning framework for pigment epithelium detachment segmentation in polypoidal choroidal vasculopathy. *Biomed Opt Express.* 2017;8(9):4061. Available at: https://doi.org/10.1364/BOE.8.004061
25. Yamanakkanavar N, Lee B. A novel M-SegNet with global attention CNN architecture for automatic segmentation of brain MRI. *Comput Biol Med.* 2021;136:104761. Available at: https://doi.org/10.1016/j.compbiomed.2021.104761
26. Liu W, Sun Y, Ji Q. MDAN-UNet: Multi-Scale and Dual Attention Enhanced Nested U-Net Architecture for Segmentation of Optical Coherence Tomography Images. *Algorithms.* 2020;13(3):60. Available at: https://doi.org/10.3390/a13030060
27. Fu H, Cheng J, Xu Y, Wong DWK, Liu J, Cao X. Joint Optic Disc and Cup Segmentation Based on Multi-Label Deep Network and Polar Transformation. *IEEE Trans Med Imaging.* 2018;37(7):1597–1605. Available at: https://doi.org/10.1109/TMI.2018.2791488
28. Ahmed Z, Panhwar SQ, Baqai A, Umrani FA, Ahmed M, Khan A. Deep learning based automated detection of intraretinal cystoid fluid. *Int J Imaging Syst Tech.* 2021;32(3):902–917. Available at: https://doi.org/10.1002/ima.22662