Macular Hole and Cystoid Macular Edema Joint Segmentation by Two-Stage Network and Entropy Minimization

Lei Ye[1], Weifang Zhu[1], Dengsen Bao[1], Shuanglang Feng[1],
and Xinjian Chen[1,2(✉)]

[1] School of Electronics and Information Engineering, Soochow University,
Suzhou, China
xjchen@suda.edu.cn
[2] State Key Laboratory of Radiation Medicine and Protection, Soochow University,
Suzhou, China

Abstract. The co-occurrence of macular hole (MH) and cystoid macular edema (CME) indicates the serious visual impairment in ophthalmology clinic. Joint segmentation and quantitative analysis of MH and CME can greatly assist the ophthalmologists in clinical diagnosis and treatment. Benefitting from the advancement of computer digital image processing technology, deep learning has shown remarkable performance in assisting doctors to diagnose diseases. In this paper, we propose a two-stage network for the segmentation of MH and CME, the MH auxiliary network and the joint segmentation network, in which the output of the Linknet based auxiliary network is used as the input of the joint segmentation network. The MH auxiliary network is designed to solve the problem that the top boundary of the MH is difficult to be discriminated by the joint segmentation network. In the joint segmentation network, we add a mixed downsampling module to retain more fine feature information during the downsampling. Furthermore, a new self-entropy loss function is proposed, which can pay more attention to the hard samples and reduce the uncertainty of the network prediction. Experimental results show that our method achieved an average Dice of 89.32% and an average IOU of 81.42% in segmentation of MH and CME, showing extremely competitive results.

1 Introduction

Cystoid macular edema(CME) is a common symptom that mostly occurs in retinal diseases such as age-related macular degeneration (AMD), diabetic retinopathy(DR), cataract, etc [10]. Macular hole(MH) is a full-thickness defect of retinal tissue in the macular area, thereby affecting central visual acuity [4]. Benefitting from non-invasive, radiation-free and high-resolution, optical coherence tomography(OCT) is widely used in ophthalmic clinics [12]. Ophthalmologists can judge

---

the degree of retinopathy according to the sizes of MH and CME through OCT images, but manually quantitative analysis of CME and MH from each image is time-consuming and labor-intensive. Therefore, the automatic CME and MH joint segmentation can greatly assist the ophthalmologists in clinical diagnosis and treatment.

In recent years, many methods have been proposed for the segmentation of MH or CME, including traditional algorithms and deep learning based methods. Zhu et al. [12] used graph cut and Adaboost classifier to improve the segmentation accuracy of CME with co-existence of MH. Wu et al. [9] combined Gaussian mixture model and active contour model to segment CME. Wilkins et al. [8] proposed a bilateral filter and thresholding based method to perform CME segmentation.

Some deep learning based methods have also been applied for the CME segmentation problem. Bai et al. [1] combined conditional random fields(CRF) and fully convolutional networks(FCN) [6] to refine the results of FCN in CME segmentation. On the basis of FCN, the U-Net [7], a more powerful and widely used network was proposed. Girish et al. [3] applied U-Net to the segmentation of intra-retinal cysts and achieved excellent results. To the best of our knowledge, there are no deep learning based methods for MH segmentation or CME and MH joint segmentation.

In this paper, we propose a two-stage encoder-decoder network for the joint segmentation of CME and MH. Our main contribution includes: (1) A MH auxiliary network is proposed to segment the top boundary of the MH, which can effectively improve the overall performance of the proposed network. (2) A mixed downsampling module is adopted to reduce the false positives of CME segmentation. (3) A self-entropy loss based loss function is proposed to improve the performances of CME and MH segmentation.

2   Method

2.1   Overall Architecture

The overall structure of the proposed two-stage network is shown in Fig. 1, which consists of MH auxiliary network and joint segmentation network. The original image is processed by the MH auxiliary network to generate an auxiliary line on the top of MH, which represents the top boundary of MH. Then the output of the MH auxiliary network is used as the input of the joint segmentation network. The characteristics of MH and CME lesions have high similarity in some aspects such as intensity value. If these two kinds of lesions are segmented separately, it is easy for the model to mistake CME as MH or MH as CME. Joint segmentation can use their relative position information to better distinguish these two types of lesions. Our baseline is composed of a shared encoder and two independent decoder branches. Similar as U-Net, each block is constructed by two convolution layers, and the number of feature channels is shown in Fig. 1(ii).
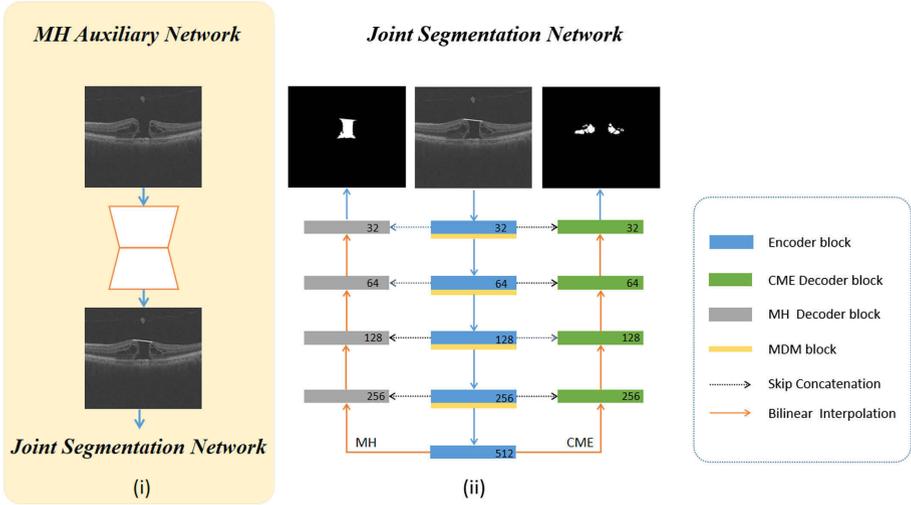
Fig. 1. Overview of the architecture of two-stage network.

2.2   MH Auxiliary Network

When the neural network model is used to segment the MH, it is prone to produce false positive or false negative in the top area of MH, shown as the second image in Fig. 3. It is very difficult for the model to learn the definition of the top boundary of MH directly from the OCT image because of the high similarity of intensity between the inner and outer region of MH, which is shown as the red line in Fig. 2(iii). According to the definition of top boundary of MH in medicine, we designed a MH auxiliary network as shown in Fig. 2(i). First, we feed the original image into a Linknet [2] based location network to find the highest point on both sides of the MH, which is shown as white dots in Fig. 2(iii). The location network outputs two Gaussian maps. The maximum likelihood estimation(MLE) method is used to estimate the center point of the Gaussian map as the predicted highest point. The auxiliary line of MH is generated by these two location points, which will be fused with the original image and finally sent it to the segmentation network. We will discuss different fusion options in the following ablation experiments.

2.3   Joint Segmentation Network

The structure of the joint segmentation network is shown in Fig. 1(ii). The network also adopts the classical encoder and decoder structure. Each block has two convolution layers and the number in block represents the amount of feature channels. Different from U-Net, the joint segmentation network proposed by us has two decoding paths and one encoding path. The outputs of the two decoder branches correspond to MH and CME respectively. We replace max
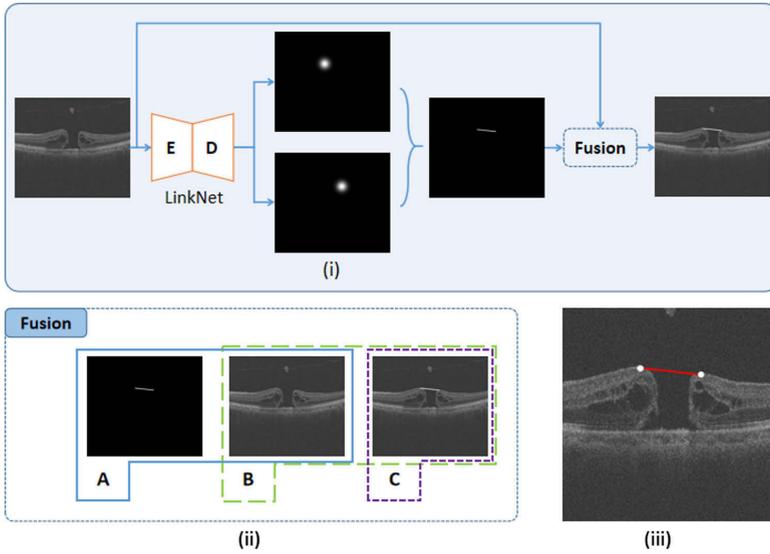
Fig. 2. Illustration of the MH auxiliary network. (i) Auxiliary line generator, we use Linknet as the backbone of our location network. (ii) A, B and C represent different fusion ways between the auxiliary line and original image. (iii) The enlarged view of lesions.

pooling with a mixed downsampling module(MDM), and propose a new loss function named as self- entropy(SE) loss.

MDM. Pooling is a very important operation in neural network models, which has the functions including feature dimension reduction, network parameter reduction, effective receptive field expansion and overfitting relief. Max pooling preserves texture features, while average pooling preserves overall features. Max pooling is the most widely used pooling method in neural network. Although max pooling can effectively reduce the computation, it inevitably loses some important features, which is disastrous for small targets such as CME, and cannot be recovered. In order to retain more features during the downsampling process, inspired by [5], we use a MDM that combines max pooling and average pooling. The MDM can be summarized as follows:

$$F_H = \alpha \cdot f_{avg}(F_L) + \beta \cdot f_{max}(F_L) \tag{1}$$

Where F_L means the low level features, while the F_H means the high level features. f_{avg} denotes the average pooling operation and f_{max} denotes the max pooling operation. α and β are the weights of two pooling methods respectively, which are both set as 1 in this paper.

SE. In order to enhance the generalization of the network, we propose a joint loss function, which is composed of binary cross entropy (BCE) loss, Dice loss and SE loss. The motivation of the SE loss function can be summarized as follows. First, when designing the loss function, we tend to narrow the distribution between the predicted image and the ground truth, but often ignore the constraint on the prediction image of the network. For a label, its entropy is 0, while for the prediction image, its entropy is not certain. Second, similar to focal loss for the description of hard samples, the SE loss can automatically pay more attention to those hard samples, because according to the definition of SE loss, the closer the predicted value is to 0.5, the higher is the loss. Dice loss can effectively alleviate the problem of data imbalance in image segmentation, which is beneficial to CME segmentation. But Dice loss makes the training of the model unstable, which will be relieved by the addition of BCE loss. The joint loss can be expressed as:

$$
\begin{aligned}
L_{\text{joint}} &= w_0 L_{Dice} + w_1 L_{BCE} + w_2 L_{SE} \\
&= w_0 \left(1 - \frac{2\sum_{i=1}^{C} g_i \times p_i}{\sum_{i=1}^{C} g_i + p_i} \right) - w_1 \left[\frac{1}{C} \sum_{i=1}^{C} g_i \log p_i + (1 - g_i) \log (1 - p_i) \right] \\
&\quad - w_2 \left[\frac{1}{C} \sum_{i=1}^{C} p_i \log p_i + (1 - p_i) \log (1 - p_i) \right]
\end{aligned}
\tag{2}
$$

Where w_0, w_1 and w_2 are weights of these three loss functions respectively, which are all set as 1 in our experiments. $g \in \{0,1\}$ is the ground truth, and $p \in [0,1]$ is the predicted probability. C is the sum of the pixels of the output results.

3   Experimental Results

3.1   Dataset, Evaluation Metrics and Implementation Details

The collection and analysis of image data were approved by the cooperative hospital and adhered to the tenets of the Declaration of Helsinki. Because of its retrospective nature, informed consent was not required from subjects. 240 images from 20 subjects (Topcon OCT-2000, radial scan model) were acquired as our experimental dataset. The corresponding ground truth were all annotated by professional ophthalmologists. It should be pointed out that the medical significance of this paper is to help doctors find the relationship between MH and CME, so there will be no case without MH. In order to make the experiment more representative, we split data into training and test set according to the subject and the strategy of 3-fold cross validation is adopted. All the experimental results are the average of the 3-fold cross validation results. The original image size is 1024×885, we resize to 512×448. For data augmentation, we apply a random horizontal flip, random vertical flip and random rotation between 0 and 45 degrees in our dataset. During training, batch size is set to 6 and the poly learning rate scheduling is adopted, decay coefficient is 0.9. The initial learning rate is 0.01 and 60 epochs are trained per fold. All the models are based on the pytorch framework and a NVIDIA GeForce RTX 2080Ti with 11 GB memory.

The evaluation metrics include Intersection-over-Union (IoU), Dice, Accuracy (Acc), Sensitivity (Sen) and Specificity(Spec) are adopted, among which, IoU is regarded as the most important evaluation metric. The epoch of all other metrics is the same as the epoch corresponding to the maximum IoU.
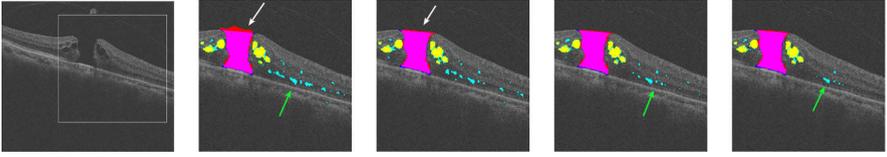
3.2   Ablation Experiments

MH Auxiliary Network. For the MH auxiliary network, we not only perform ablation experiments, but also discuss the fusion methods. We design three methods to fuse the original image and the auxiliary line (shown in Fig. 2(ii)). In scheme A, we concatenate the auxiliary line and the original image in the channel dimension, while in scheme C, these two images are directly merged into one image. In solution B, we concatenate the original image and the image generated by scheme C.

Table 1. The performance comparison of different fusion methods in MH segmentation

Methods		IoU(%)	Dice(%)	Acc(%)	Sen(%)	Spec(%)
Baseline	MH	86.04	92.36	99.70	93.26	99.83
A	MH	88.00	93.50	99.75	91.00	**99.92**
B	MH	88.72	93.92	99.77	**93.37**	99.90
C	MH	**89.36**	**94.30**	**99.79**	93.26	99.91

The experimental results show that no matter which kind of fusion method is used, it can greatly improve the performance of MH segmentation (shown in Table 1). Among them, scheme C is the best, which can improve about 3.3% in IoU compared with the baseline. The visualization of experimental results in Fig. 3 show that the auxiliary network can significantly improve the segmentation performance of the top boundary of MH(indicated by white arrows).

MDM and SE. The detailed ablation experiment results of MDM and SE loss function are presented in Table 2. As can be seen from Table 2, the addition of MDM and SE loss function significantly improves the segmentation performance of CME and improves the segmentation of MH relatively small, which is in line with our idea. CME lesions have a large number of hard samples with high degree of uncertainty due to the characteristics of big differences in size and spatial distribution. In the downsampling process of the network, MDM helps to retain more feature information about small CME lesions, while the SE loss can pay more attention to hard samples. For IoU index, the addition of MDM and SE loss improves by 1.26% and 1.18% in CME segmentation, respectively. As shown in Fig. 3(green arrow), MDM and SE loss can effectively reduce the false positives of CME.

Fig. 3. Examples results of ablation experiments. From left to right: original image, baseline, baseline+MH auxiliary network, baseline+MH auxiliary network+MDM, baseline+ MH auxiliary network +MDM+SE. Magenta, red and blue represent the true positive, false positive and false negative for MH, while yellow, green and cyan represent true positive, false positive and false negative for CME. (Color figure online)

3.3 Comparison Experiments

The details of the comparison experiments are shown in Table 2, where Aux represents the MH auxiliary network. As can be seen from Table 2, the proposed two-stage method achieves significant improvements in IoU, Dice, accuracy and

Table 2. The performance of different segmentation models on MH and CME.

Methods		IoU(%)	Dice(%)	Acc(%)	Sen(%)	Spec(%)
FCN8s [6]	CME	39.47	54.22	99.35	52.49	99.71
	MH	74.03	84.32	99.47	80.00	99.84
	AVG	56.75	69.27	99.41	66.24	99.78
U-Net [7]	CME	63.10	76.54	99.57	83.00	99.70
	MH	78.96	87.87	99.53	85.86	99.82
	AVG	71.03	82.21	99.55	84.43	99.76
U-Net++ [11]	CME	70.95	82.59	99.70	86.26	99.81
	MH	88.41	93.78	99.75	**94.50**	99.86
	AVG	79.68	88.19	99.72	90.38	99.83
Baseline	CME	70.20	82.05	99.68	88.15	99.78
	MH	86.04	92.36	99.70	93.26	99.83
	AVG	78.12	87.32	99.69	**90.67**	99.80
Baseline+Aux	CME	70.53	82.31	99.69	87.09	99.80
	MH	89.36	94.30	99.79	93.26	99.91
	AVG	79.95	88.31	99.74	90.17	99.86
Baseline+Aux+MDM	CME	71.79	83.22	99.71	**90.67**	99.81
	MH	89.49	94.36	99.78	93.33	**99.92**
	AVG	80.64	88.79	99.75	90.33	99.86
Baseline+Aux+MDM+SE	CME	**72.97**	**84.06**	**99.73**	86.30	**99.84**
	MH	**89.87**	**94.59**	**99.79**	93.96	99.91
	AVG	**81.42**	**89.32**	**99.76**	90.13	**99.87**
	CME_ONLY	69.66	81.60	99.67	86.55	99.78
	MH_ONLY	83.81	90.93	99.66	87.06	**99.92**

|(a) image|(b) FCN8s|(c) U-Net|(d) U-Net++|(e) Ours|
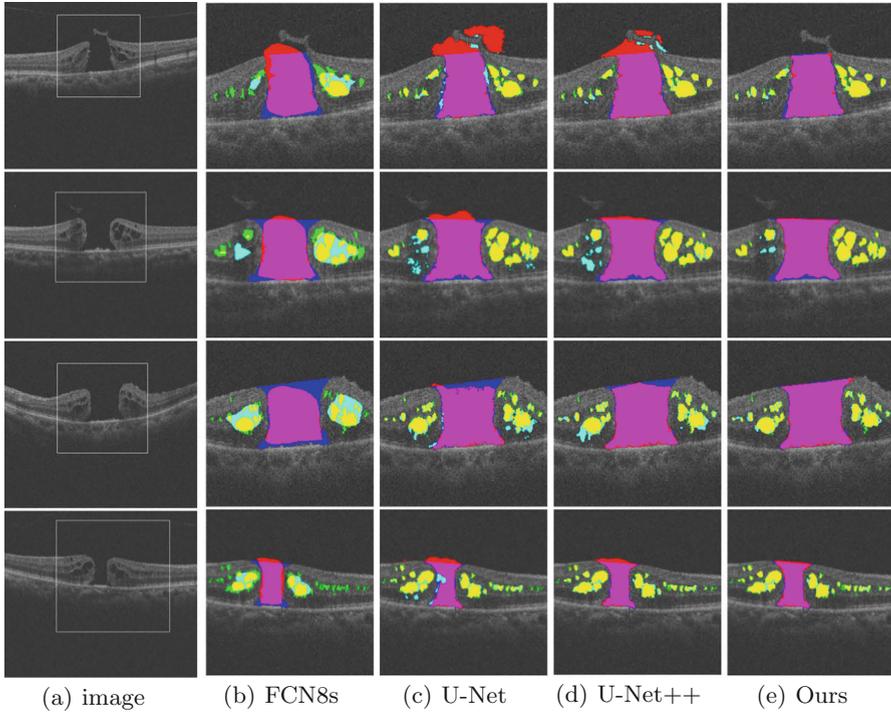
Fig. 4. MH and CME segmentation results with different models. Magenta, red and blue represent the true positive, false positive and false negative of MH respectively, while yellow, green and cyan represent the true positive, false positive and false negative of CME respectively. (Color figure online)

specificity compared with other models, such as FCN8s, U-Net and U-Net++ [11], which are popular in medical image segmentation.

Compared with U-Net++, our method improves average IoU by 1.74% and Dice by 1.13% respectively. We use the best model (Baseline+Aux+MDM+SE) to segment MH and CME separately. The corresponding experimental results prove the necessity of joint segmentation. For the sake of fairness, all models were trained and tested under the same set of hyperparameters.

Figure 4 shows the MH and CME segmentation results with different methods. It is obvious that our method achieves an impressive segmentation performance. For those models that have not been preprocessed by the MH auxiliary network, they are easy to produce serious false positives or false negatives in the top boundary of the MH, which indicates that the MH Auxiliary network plays an important role in MH segmentation. The addition of the MDM and SE loss effectively improves the segmentation performance on the hard CME samples.

4   Conclusion

In this paper, we propose a new two-stage encoder-decoder network for the joint segmentation of MH and CME in retinal OCT images. The overall network consists of two parts, the MH auxiliary network and the joint segmentation network. The MH auxiliary network is used as a pre-processing to help the segmentation network determine the top boundary of the MH. In the joint segmentation network, MDM can effectively retain more fine features, which is beneficial to the segmentation of CME. The SE loss function can automatically pay more attention to the hard samples and reduce the uncertainty of the network prediction. Experiment results show that the proposed network has compelling performance and great potential in the segmentation of MH and CME.

References

1. Bai, F., Marques, M.J., Gibson, S.J.: Cystoid macular edema segmentation of optical coherence tomography images using fully convolutional neural networks and fully connected crfs. arXiv preprint arXiv:1709.05324 (2017)
2. Chaurasia, A., Culurciello, E.: Linknet: exploiting encoder representations for efficient semantic segmentation. In: 2017 IEEE Visual Communications and Image Processing (VCIP), pp. 1–4 IEEE (2017)
3. Girish, G., Thakur, B., Chowdhury, S.R., Kothari, A.R., Rajan, J.: Segmentation of intra-retinal cysts from optical coherence tomography images using a fully convolutional neural network model. IEEE J. Biomed. Health Inform. **23**(1), 296–304 (2018)
4. Ho, A.C., Guyer, D.R., Fine, S.L.: Macular hole. Surv. Ophthalmol. **42**(5), 393–416 (1998)
5. Lee, C.Y., Gallagher, P., Tu, Z.: Generalizing pooling functions in cnns: mixed, gated, and tree. IEEE Trans. Pattern Anal. Mach. Intell. **40**(4), 863–875 (2017)
6. Long, J., Shelhamer, E., Darrell, T.: Fully convolutional networks for semantic segmentation. In: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, pp. 3431–3440 (2015)
7. Ronneberger, Olaf: Invited talk: U-net convolutional networks for biomedical image segmentation. Bildverarbeitung für die Medizin 2017. I, pp. 3–3. Springer, Heidelberg (2017). https://doi.org/10.1007/978-3-662-54345-0_3
8. Wilkins, G.R., Houghton, O.M., Oldenburg, A.L.: Automated segmentation of intraretinal cystoid fluid in optical coherence tomography. IEEE Trans. Biomed. Eng. **59**(4), 1109–1114 (2012)
9. Wu, J., Niu, S., Chen, Q., Fan, W., Yuan, S., Li, D.: Automated segmentation of intraretinal cystoid macular edema based on gaussian mixture model. J. Innov. Opt. Health Sci. **13**(1), 1950020 (2019)
10. Zhang, L., Zhu, W., Shi, F., Chen, H., Chen, X.: Automated segmentation of intraretinal cystoid macular edema for retinal 3d oct images with macular hole. In: 2015 IEEE 12th International Symposium on Biomedical Imaging (ISBI), pp. 1494–1497 IEEE (2015)

11. Zhou, Z., Rahman Siddiquee, M.M., Tajbakhsh, N., Liang, J.: Unet++: a nested u-net architecture for medical image segmentation. In: Stoyanov, D., et al. (eds.) DLMIA/ML-CDS -2018. LNCS, vol. 11045, pp. 3–11. Springer, Cham (2018). https://doi.org/10.1007/978-3-030-00889-5_1
12. Zhu, W., et al.: Automated framework for intraretinal cystoid macular edema segmentation in three-dimensional optical coherence tomography images with macular hole. J. Biomed. Opt. **22**(7), 076014 (2017)