CU-SEGNET : CORNEAL ULCER SEGMENTATION NETWORK

Tingting Wang[1], Weifang Zhu[1], Meng Wang[1], Zhongyue Chen[1], Xinjian Chen[1,2]

[1]School of Electronic and Information Engineering, Soochow University, Suzhou, China
[2]State Key Laboratory of Radiation Medicine and Protection, Soochow University, Suzhou, China

ABSTRACT

Corneal ulcer is a common-occurring illness in cornea. It is a challenge to segment corneal ulcer in slit-lamp image due to the different sizes and shapes of point-flaky mixed corneal ulcer and flaky corneal ulcer. These differences introduce inconsistency and effect the prediction accuracy. To address this problem, we propose a corneal ulcer segmentation network (CU-SegNet) to segment corneal ulcer in fluorescein staining image. In CU-SegNet, the encoder-decoder structure is adopted as main framework, and two novel modules including multi-scale global pyramid feature aggregation (MGPA) module and multi-scale adaptive-aware deformation (MAD) module are proposed and embedded into the skip connection and the top of encoder path, respectively. MGPA helps high-level features supplement local high-resolution semantic information, while MAD can guide the network to focus on multi-scale deformation features and adaptively aggregate contextual information. The proposed network is evaluated on the public SUSTech-SYSU dataset. The Dice coefficient of the proposed method is 89.14%.

***Index Terms*—**Slit-lamp image, Corneal ulcer, Image segmentation, Deep learning

1. INTRODUCTION

Cornea is a transparent membrane in the front of the eye. It is extensively exposed to the air and more likely to be infected with bacteria, which may lead to ophthalmic symptoms such as corneal ulcer. A corneal ulcer is an inflammatory or more seriously, infective condition of the cornea involving disruption of its epithelial layer with involvement of the corneal stroma [1]. According to the shape and distribution characteristics, corneal ulcer can be classified into three categories: point-like corneal ulcer, point-flaky mixed corneal ulcer and flaky corneal ulcer. Fluorescein staining and slit-lamp imaging are widely used for corneal disease diagnosis in clinic. Automatic and accurate corneal ulcer detection and segmentation in fluorescein staining slit-lamp images could help the ophthalmologists to diagnose corneal ulcer at its early stage and reduce the risk of blindness.

In the past decade, there are several corneal ulcer detection and segmentation methods [2][3][4]. L. Deng et al presented an automatic ulcer segmentation method by utilizing k-means clustering followed by morphological operations and region growing [2]. Z. Liu et al segmented the ulcer area within the cornea by employing a joint method of Otsu and Gaussian mixture modeling (GMM) with 150 images [3]. Q. Sun et al proposed a patch-based deep convolutional neural network for corneal ulcer area segmentation with 48 images [4].

U-Net [5] and its variants have been widely used in medical image segmentation due to its good performance. However, the original U-shape encoder-decoder structure still suffers from the insufficiency of the contextual information extraction capability, which is important for the lesion area segmentation with different sizes and shapes. To address this problem, we propose a novel CU-SegNet based on commonly used U-shape architecture for corneal ulcer area segmentation. Our main contributions are as follows:

1) To reduce the impact of large differences in morphology and size of lesions, two novel modules including multi-scale global pyramid feature aggregation (MGPA) module, which can improve the network's ability to learn to supplement low-level local high-resolution semantic information to high-level feature maps, and multi-scale adaptive-aware deformation (MAD) module, which guides the model to focus on multi-scale deformation of the targets and aggregates the contextual information, are proposed.

2) Comprehensive experiments are conducted based on SUSTech-SYSU dataset to demonstrate the effectiveness of our proposed method. The experimental results show that our proposed CU-SegNet achieves better segmentation performance compared with other excellent methods.

2. METHODOLOGY DEVELOPED

The overview of the proposed CU-SegNet is shown in Fig. 1. The commonly used U-shape encoder-decoder structure is adopted as our main framework, where the pretrained ResNet-34 [6] is employed as the encoder path to extract rich feature information from the input image. The novel proposed MAD module is embedded into the top of encoder path to guide the model to focus on the multi-scale deformation maps and aggregate the contextual information. The proposed MGPA module is embedded into the skip connection to supplement low-level local high-resolution semantic information to high-level feature maps.

2.1. Multi-scale Adaptive-aware Deformation Module

As can be seen form Figure 1, the proposed MAD module is embedded at the top of the encoder path. Figure 2 shows its structure, which contains 4 parts: parallel and deformable convolution module, multiple global spatial attention module, multiple global channel attention module and adaptive residual module.

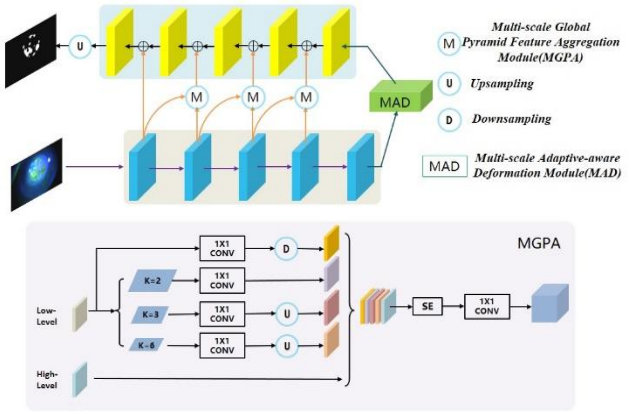

Fig. 1. Overview of the proposed network and the MGPA module.

2.1.1. Parallel and Deformable Convolution Module

In parallel and deformable convolution module, the multi-receptive field convolutional operation is stacked by parallel. It has four convolutional branches and one deformable convolution. First, four convolutional branches are used to squeeze the channels to reduce the cost of calculation by 1×1 convolution. Then 3×3 convolution and dilation convolution with rate 1, 3, 5, and 7 are followed to get the respective field size of 3, 11, 19 and 27, respectively. Next, these feature maps are concatenated and fed into a deformable convolution [7], which can augment the spatial sampling locations in the modules by additional offsets of kernel size in horizontal and vertical direction. Finally, the output feature maps are fed into the parallel-linked multiple global spatial attention module and the multiple global channel attention module.

2.1.2. Multiple Global Spatial Attention Module

Max-pooling can extract the most significant spatial response information in each channel of the feature maps. However, it may also introduce noise due to the different sizes and shapes of lesion. Meanwhile, average-pooling can represent the average of all channels in the corresponding position in the input feature maps. Although this approach can suppress some of the noise interference in the channels, it also suppresses the most significant spatial response information in all channels. Therefore, to get the most significant spatial response information in all channels and suppress noise interference, 2D average-pooling and max-pooling are performed simultaneously in multiple global spatial attention module. The feature maps are fed to the maximum map branch (h × w) and the mean map branch (h × w) in parallel. A convolutional operation to squeeze the channel of concatenated maps (2 × h × w) is followed. Finally, Sigmoid function is adopted to normalize the value of attention matrix to 0~1 (h × w). The contextual information in spatial dimension of original feature maps can be obtained. At the same time, the interference of noise can also be suppressed by this module.

2.1.3. Multiple Global Channel Attention Module

Multiple global channel attention module is similar to multiple global spatial attention module. The feature maps are firstly fed into two parallel branches to calculate the maximum and mean value of each feature map in all channels respectively. Then global channel maximum value maps (c×1×1) and global channel mean value maps (c×1×1) are concatenated and followed by a convolution layer to smooth and squeeze the feature maps. Finally, the results are reshaped (c×1) and fed into a fully connection layer followed by sigmoid function to obtain the weights of each feature map. This module can get the response of each feature map in all channels and suppress noise interference.

2.1.4. Adaptive Residual Module

The result of deformable convolution (X_1) multiplies feature maps from multiple global spatial attention module (X_2) and the multiple global channel attention module (X_3) by coefficients respectively. They are added together. Then the feature maps are smoothed by convolution. The learnable parameters γ and λ are initialized as a non-zero value (1.0 in this paper). The process can be summarized as:

$$Output = \text{Conv}\,(X_1 + \gamma X_1 X_3 + \lambda X_1 X_2). \quad (1)$$

Finally, we directly add the smoothed feature maps with the original feature maps to construct the residual mechanism.

2.2. Multi-scale Global Pyramid Feature Aggregation Module

The original skip-connection in the U-shape network will introduce irrelevant clutters and have semantic gap. To solve this problem, the MGPA module is designed and embedded into the skip connection to supplement low-level local high-resolution semantic information to high-level feature maps. In the MGPA module, the feature maps of current stage F_2 ($C_2 \times H_2 \times W_2$) and the low-level stage F_1 ($C_1 \times H_1 \times W_1$) are taken as input. First, F1 maps are fed into max pooling layers with size 2, 3 and 6 to capture targets with different sizes and shapes, which is followed by a 1×1 convolution layer. Then they are upsampled to the same size as the current stage feature maps via bilinear interpolation.

At the same time, F1 maps are fed into a 1×1 convolution layer and resized to the size of F2 to add more semantic information, individually. All of them are concatenated with the current stage feature maps F2. Finally, the concatenated feature map ($(4C_1+C_2) \times H_2 \times W_2$) is fed into squeeze-and-excitation (SE) module [8], followed by 1×1 convolution layer to match the channel number of decoder stage. By this module, the low-level local high-resolution semantic information is supplemented to high-level feature maps.

3. EXPERIMENTS AND RESULTS

3.1. Dataset and implementation details

We evaluate our network on a public slit-lamp fluorescein staining image dataset, the SUSTech-SYSU dataset, which is created to evaluate automatic corneal the ulcer segmentation algorithms and identify the general and specific ulcer patterns as well as the ulcer severity degree [9]. The dataset has 354 images with corneal ulcer, among which 263 are labeled as point-flaky mixed corneal ulcer and the other 91 are labeled as flaky corneal ulcer. The size of each RGB image is 2592 × 1728. The original images and labels are resized to 512×512 by bilinear interpolation in order to achieve the balance between computational efficiency and accuracy of prediction. 4-fold cross-validation strategy (90, 90, 90 and 84) is adopted. In order to avoid over-fitting and improve the robust ability of model, online data augmentation has performed, including rotations of -10 degrees to 10 degrees, horizontal flipping, vertical flipping, Gaussian noise addition and affine transformation.

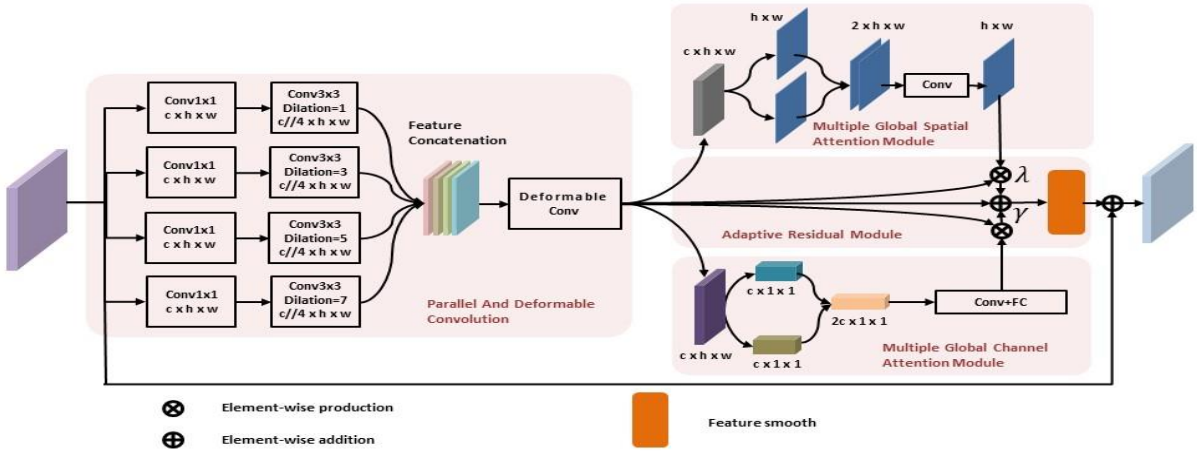

Fig.2. The MAD module with 4 parts: parallel and deformable convolution module, multiple global spatial attention module, multiple global channel attention module and adaptive residual module.

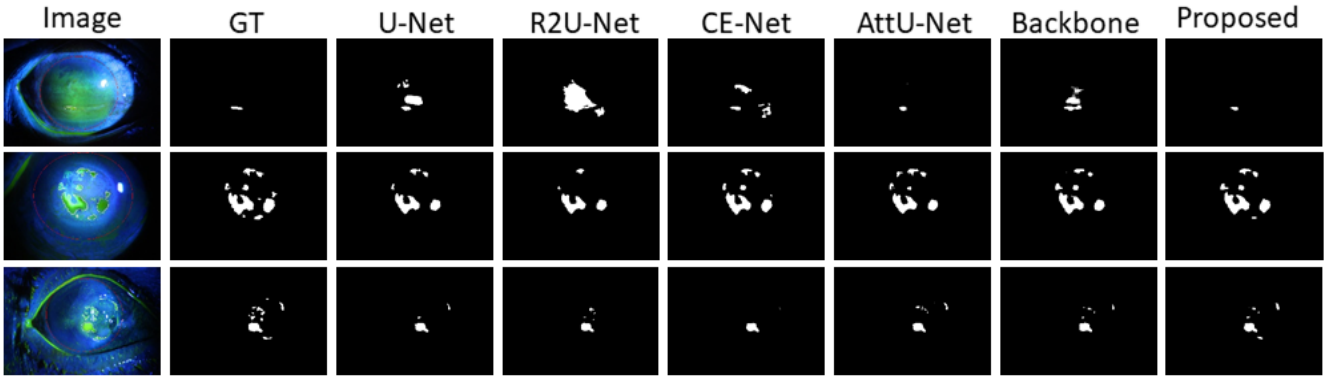

Fig.3. The examples of corneal ulcer segmentation. From left to right: original image, ground truth (GT), U-Net, R2U-Net, CE-Net, Attention U-Net, Backbone, and proposed method.

TABLE I. The result of contrast experiments and ablation studies (mean ± standard deviation)

Methods	Dice (%)	SEN (%)	SPE (%)	PCC (%)
U-Net	87.28±5.38	88.54±3.71	99.64±0.14	87.40±5.23
AttU-Net	86.40±6.17	88.05±3.28	99.62±0.15	86.59±6.03
R2U-Net	80.76±9.26	82.56±5.78	99.47±0.29	81.29±8.67
CE-Net	88.43±4.81	88.45±4.31	99.73±0.08	88.48±4.53
DeepLabv3+	88.29±5.41	89.19±4.90	99.69±0.10	88.33±5.27
PSPNet	89.09±4.64	90.20±3.34	99.70±0.10	89.08±4.56
backbone	87.71±5.54	88.53±4.03	99.70±0.11	88.00±5.29
backbone +MAD	88.82±4.51	88.99±3.17	99.73±0.10	88.82±4.44
backbone +MGPA	88.91±4.32	89.96±3.12	99.70±0.09	88.93±4.26
U-Net+MAD	88.70±4.39	89.21±2.87	99.65±0.17	88.13±4.39
U-Net+MGPA	87.93±4.27	88.35±3.54	99.69±0.12	87.98±4.14
Proposed	**89.14±4.59**	**89.65±4.06**	**99.70±0.13**	**89.16±4.51**

Cross-entropy loss and Dice loss are jointly used as the loss function to train the proposed network. The proposed network is performed on the public platform pytorch and a Tesla K40 GPU (12GB). Adam is used as the optimizer. Initial learning rate is set to 0.0005, and weight decay is set to 0.0001. The batch size is set as 4 and epoch is 100.

3.2. Results

4 metrics including Dice coefficient (Dice), sensitivity (SEN), specificity (SPE) and Pearson's correlation coefficient (PCC) are adopted to evaluate the corneal ulcer segmentation performance of different methods. The proposed method is compared with other excellent methods including U-Net, Attention U-Net [10], R2U-Net [11], CE-Net [12], PSPNet [13] and DeepLabv3+ [14]. All networks are trained with same parameters.

As shown in Table I and Fig. 3, the proposed method outperforms other networks. Compared with the result of backbone, the proposed method achieves an increase of 1.43% in Dice coefficient, which indicates the effectiveness of the proposed MAD and MGPA module. The ablation experiments both the backbone and U-Net with MAD and MGPA modules further show the effectiveness and generality of these two modules.

4. CONCLUSION

In this paper, we propose a novel CU-SegNet to segment corneal ulcer in fluorescein staining slit-lamp image automatically, which can improve the segmentation performance of the corneal ulcer area with different scales and shapes. The proposed CU-SegNet is evaluated on the recently published public SUSTech-SYSU dataset. The comparison and ablation experiments indicate the primary effectiveness of two proposed module. In the experiments, we also find that the point-flaky mixed corneal ulcer is not segmented accurately enough because of the great intra-eye and inter-eye differences of size and shape, which will be focused and settled in our near future work.

5. ACKOWLEDEGMENT

This work was supported in part by the National Key Research and Development Program of China under Grant 2018YFA0701700, in part by the National Nature Science Foundation of China under Grant 61622114, and in part by the National Basic Research Program of China under Grant 2014CB748600.

6. COMPLIANCE WITH ETHICAL STANDARDS

This research study was conducted retrospectively using human subject data made available in open access by (https://github.com/CRazorback/The-SUSTech-SYSU-dataset-for-automatically-segmenting-and-classifying-corneal-ulcers) and (https://doi.org/10.6084/m9.figshare.c.4526675). Ethical approval was not required as confirmed by the license attached with the open access data.

14. REFERENCES

[1] J. Chen and J. Yuan. "Strengthen the study of the ocular surface recon struction". [Zhonghua yan ke za zhi] Chinese journal of ophthalmology, vol. 46, no.1, pp. 3-5, 2010.

[2] L. Deng, H. Huang, J. Yuan and X. Tang. "Automatic segmentation of corneal ulcer area based on ocular staining images". Medical Imaging 2018: Biomedical Applications in Molecular, Structural, and Functional Imaging. International Society for Optics and Photonics, 2018, pp. 10578-105781D.

[3] Z. Liu, Y. Shi, P. Zhan, Y. Zhang, Y. Gong, and X. Tang. "Automatic Corneal Ulcer Segmentation Combining Gaussian Mixture Modeling and Otsu Method." 2019 41st Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC). IEEE, 2019, pp. 6298-6301.

[4] Q. Sun, L. Deng, J. Liu, H. Huang, J. Yuan, and X. Tang. "Patch-Based Deep Convolutional Neural Network for Corneal Ulcer Area Segmentation." International Workshop on Fetal & Infant Image Analysis International Workshop on Ophthalmic Medical Image Analysis. Springer, 2017, pp. 101-108.

[5] O. Ronneberger, P. Fischer, and T. Brox. "U-Net: Convolutional Networks for Biomedical Image Segmentation." Springer, 2015, pp. 234-241.

[6] K. He, X. Zhang, S. Ren, and J. Sun. "Deep residual learning for image recognition." Proceedings of the IEEE conference on computer vision and pattern recognition. IEEE, 2016, pp. 770-778.

[6] K. He, X. Zhang, S. Ren, and J. Sun. "Deep residual learning for image recognition." Proceedings of the IEEE conference on computer vision and pattern recognition. IEEE, 2016, pp. 770-778.

[7] J. Dai, H. Qi, Y. Xiong, Y. Li, G. Zhang, H. Hu, and Y. Wei. "Deformable Convolutional Networks." Proceedings of the IEEE international conference on computer vision. IEEE, 2017, pp. 764-773.

[8] J. Hu, L. Shen, and G. Sun. "Squeeze-and-Excitation Networks." IEEE Transactions on Pattern Analysis and Machine Intelligence, IEEE, 2018, pp. 7132-7141.

[9] L. Deng, J. Lyu, H. Huang, Y. Deng, J. Yuan, and X. Tang. "The SUSTech-SYSU dataset for automatically segmenting and classifying corneal ulcers." Scientific Data, vol. 7, no.1, pp. 1-7, 2020.

[10] O. Oktay, J. Schlemper, L. L. Folgoc, M. Lee, M. Heinrich, K. Misawa, K. Mori, S. McDonagh, N. Y. Hammerla, B. Kainz, B. Glocker, and D. Rueckert. "Attention U-Net: Learning Where to Look for the Pancreas." arXiv preprint arXiv:1804.03999, 2018.

[11] M. Z. Alom, M. Hasan, C. Yakopcic, T. M. Taha, and V. K. Asari. "Recurrent Residual Convolutional Neural Network based on U-Net (R2U-Net) for Medical Image Segmentation." arXiv preprint arXiv:1802.06955, 2018.

[12] Z. Gu, J. Cheng, H. Fu, K. Zhou, H. Hao, Y. Zhao, T. Zhang, S. Gao, and J. Liu. "CE-Net: Context Encoder Network for 2D Medical Image Segmentation." IEEE Transactions on Medical Imaging, vol. 38, no.10, pp. 2281-2292, 2019.

[13] H. Zhao, J. Shi, X. Qi, X. Wang, and J. Jia, "Pyramid scene parsing network," Proceedings of the IEEE conference on computer vision and pattern recognition, pp. 2881-2890,2017.

[14] L. C. Chen, Y. Zhu, G. Papandreou, F. Schroff, and H. Adam, "Encoder-Decoder with Atrous Separable Convolution for Semantic Image Segmentation", european conference on computer vision, pp. 833-851, 2018.