2019 IEEE 16th International Symposium on Biomedical Imaging (ISBI 2019)
Venice, Italy, April 8-11, 2019

A MULTI-SCALE DEEP CONVOLUTIONAL NEURAL NETWORK FOR JOINT SEGMENTATION AND PREDICTION OF GEOGRAPHIC ATROPHY IN SD-OCT IMAGES

Yuhan Zhang[1], *Zexuan Ji*[1], *Sijie Niu*[2], *Theodore Leng*[3], *Daniel L. Rubin*[4,5], *Qiang Chen*[1] *

[1]School of Computer Science and Engineering, Nanjing University of Science and Technology, China
[2]School of Information Science and Engineering, University of Jinan, China
[3]Byers Eye Institute at Stanford, Stanford University School of Medicine, Palo Alto, CA 94303, USA
[4]Department of Radiology, Stanford University, Stanford, CA 94305, USA
[5]Medicine (Biomedical Informatics Research), Stanford University, Stanford, CA 94305, USA

ABSTRACT

Geographic atrophy (GA) generally appears in the advanced stage of age-related macular degeneration (AMD). It is a principle cause of the severe central visual loss for elder adults with non-exudative AMD in developed countries. In this paper, a multi-scale deep convolutional neural network is proposed for the joint segmentation and prediction of GA. First, restricted summed-area projection (RSAP) technique was used to generate GA projection images from the SD-OCT volumetric data. Then, GA projection images were sent to the multi-scale branches to acquire multi-scale feature maps. The final GA segmentation results were obtained by refining the multi-scale feature maps with a voting decision strategy. In the end, those multi-scale feature maps were cascaded with low-level features computed from the original images to predict the growth of the GA lesion. The segmented and predicted GA lesion in the tested scenarios resulted in a satisfying accuracy, comparing with the observed ground truth.

Index Terms—geographic atrophy, segmentation, prediction, multi-scale, SD-OCT, neural network

1. INTRODUCTION

Age-related macular degeneration (AMD) is a leading cause of blindness among elderly individuals. Advanced AMD usually appears as a non-exudative form characterized by the presence of geographic atrophy (GA) [1-2]. The reduction in the worsening of atrophy is a well-known but crucial biomarker for estimating the effectiveness of the GA treatment [3]. Thus, automated GA segmentation and prediction, which could aid ophthalmologists in objectively measuring the regions of GA and forecasting the evolution of GA for further treatment decisions [4], is helpful for the clinical diagnosis. Manual segmentation is time consuming, and may not produce reliable results especially for the masses of data. In addition, the artificial prediction may produce a great deal of uncertainty. Therefore, an automated, accurate and reliable segmentation and prediction technology is urgently needed in the advanced care of GA.

The spectral-domain optical coherence tomography (SD-OCT) has proven successful in identifying the GA [5] and become a main imaging tool to capture the retinal structure these years. In recent years, several state-of-the-art algorithms were proposed for GA segmentation based on the SD-OCT images. [6-8]. *Chen et al.* [6] used geometric active contours to segment GA margin automatically. The performance of this model generally depend on the contour initializations. To further improve the segmentation accuracy and robustness, *Niu et al.* [7] proposed an automated GA segmentation method by using a Chan-Vese model via local similarity factor. *Ji et al.* [8] proposed a deep voting model for automated GA segmentation of SD-OCT images, which is capable of achieving high segmentation accuracy without layer segmentation. The segmentation methods mentioned above are time-consuming and of low efficiency. Until now, few literatures have focused on the automated growth prediction of GA in the field of computer science. *Niu et al.* [9] attempted to extract 19 comprehensive quantitative imaging features as the predictors for the future GA growth, and random forest classifier was selected for the prediction model, achieving preferable prediction accuracy. Similarly, it is fussy and time-consuming in elaborately designing the hand-crafted predictors.

In order to segment and predict the GA lesion more efficiently, we proposed a multi-scale deep convolutional neural network for the joint segmentation and prediction of GA in this paper, which is capable of achieving high segmentation accuracy and prediction accuracy simultaneously. Multi-scale branches are constructed to

* Corresponding author. E-mail: chen2qiang@njust.edu.cn.
This work was supported by the National Science Foundation of China (61671242) and Key R&D Program of Jiangsu Science and Technology Department (No.BE2018131).

capture multi-scale feature maps and then the segmentation results are obtained by refining multi-scale feature maps with a voting decision strategy. Besides, multi-scale feature maps were cascaded with low-level features computed from the original images to predict the growth of the GA lesion. Experiment results indicated that our method can provide reliable segmentation and prediction for GA from SD-OCT images.

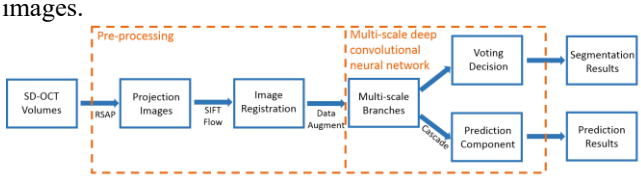

Fig. 1. The integrated flow chart of the joint segmentation and prediction of GA.

2. METHOD

The integrated flow chart of the joint segmentation and prediction of GA is illustrated in Fig. 1. The input data comprises a series of SD-OCT images and the output data includes the GA segmentation at the current time and GA prediction at the next time.

2.1. Pre-processing

The restricted summed-area projection (RSAP) [10], which restricts the axial projection of an SD-OCT volume to the regions beneath the Bruch's membrane (BM) and considers the choroidal vasculature's influence, generated more distinct GA projection images compared with other projection techniques [11]. Besides, due to the displacements among OCT volumes captured at different time points for the same patient, sift flow method [12] was selected to register the projection images into the same coordinate to guarantee the prediction accuracy. Finally, several frequently-used data augment methods, including flipping, rotation, cropping and scaling, were used to increase the sample size.

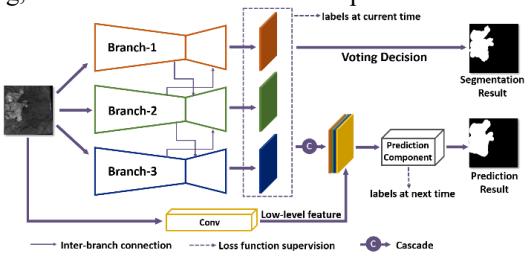

Fig. 2. The whole network architecture of the multi-scale deep convolutional nerual network.

2.2. Network Architecture

The whole network architecture of the proposed multi-scale deep convolutional neural network is illustrated in Fig. 2. In this network, branches used to capture the multi-scale features are arranged in parallel. The input GA projection images are sent to the multi-scale branches respectively. All branches share the same network structure, including the encoder part and the decoder part. For GA segmentation, a voting decision strategy was used to refine the multi-scale feature maps among branches. In addition, an extra path stacking four convolutional layers is constructed to extract the low-level features of the input image. Low-level features and multi-scale feature maps are cascaded together as the input of the prediction component of the network consisting of four convolutional layers to obtain the final GA prediction.

Multi-scale Branches: The fully convolutional network (FCN) [13] is selected as the base architecture for each branch. Based on the architecture of FCN (Fig. 3), which is a popular framework for image semantic segmentation, we replaced the standard convolution by an atrous convolution with a sparse convolutional kernel [14]. The atrous convolution is a powerful tool with the capability to capture the multi-scale image information through explicitly controlling the resolution of features computed by deep convolutional neural networks and can adjust the field-of-view of the convolutional kernel. Particularly, in the case of 2D images, for each location $\boldsymbol{i}$ on the output feature map $\boldsymbol{y}$ and the convolutional kernel $\boldsymbol{w}$, atrous convolution is applied over the input feature map $\boldsymbol{x}$ as follows:

$$\boldsymbol{y}[\boldsymbol{i}] = \sum_{\boldsymbol{k}} \boldsymbol{x}[\boldsymbol{i} + r \cdot \boldsymbol{k}]\boldsymbol{w}[\boldsymbol{k}] \qquad (1)$$

where the atrous rate r determines the stride with which we sample the input image, and $\boldsymbol{k}$ is the size of the convolutional kernel. Note that when $r = 1$, the atrous convolution can be regarded as a standard convolution. The field-of-view of the convolution kernel is adaptively modified by changing the rate r. To boost the training against the curse of gradient vanishing and augment the information propagation, we introduced the concept of dense connection into each branch, namely the intra-branch connection. The intra-branch connection also has the advantages of encouraging the reuse of features, reducing the number of parameters and benefiting the network optimization. Finally, several convolutional layers are stacked after each up-sampling process to promote the fitting ability of the branch. Modified multi-scale branches are represented in Fig. 3. We considered that the different scale information is helpful for each independent branch with the purpose of further improvement, thus an inter-branch connection was also added to interact information among branches, as shown in Fig. 2.

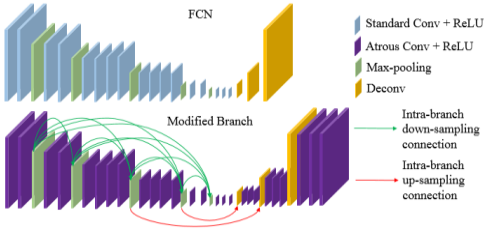

Fig. 3. The network architecture of each branch.

Loss Function: The pixel-wise cross-entropy loss was applied to all the training processes for the network. For the *i-th* branch with the pixel-wise cross-entropy loss E_i, the loss

was computed as follows:

$$E_i = \sum_{x \in \Omega} \omega(\boldsymbol{x}) \, log \, p_l(\boldsymbol{x}) \quad (2)$$

where $p_l(\boldsymbol{x})$ provides the estimated probability of pixel $\boldsymbol{x}$ belonging to class l, and $\omega(\boldsymbol{x})$ is the weight associated with pixel $\boldsymbol{x}$. To address the class imbalance, we introduced the concept of focal loss [15] to the cross-entropy loss. Then equation (2) can be reshaped as follows:

$$E_i = \sum_{x \in \Omega} \omega(\boldsymbol{x})(1 - p_l(\boldsymbol{x}))^{\gamma} \, log \, p_l(\boldsymbol{x}) \quad (3)$$

$(1 - p_l(\boldsymbol{x}))^{\gamma}$ is a modulating factor which is able to reduce the loss contribution from distinguishable examples and can extend the range in which an example can receive a low loss. We found the proposed architecture performs the best when $\gamma = 1$ in our experiments.

Implement Details: In our proposed network, each branch extracts feature maps in different scales. In this paper, the scale number was set to 3. The convolutional layers in each branch were composed of a 3×3 atrous convolution operation and a ReLU, followed by a dropout with the probability of 0.85. The rates of the atrous convolution of the multi-scale branch were set to 1, 2 and 4, respectively. The number of channels in each convolution layer was fixed to 128. We conducted the voting strategy on the segmentation results from the three branches and considered the labels with the probability greater than 2/3 as the final label of the current pixel.

In the training process, multi-scale branches were trained firstly for outputting the multi-scale feature maps. We set the GA segmentation label maps at the current time for supervision. GA label maps were annotated by experts manually. The loss function is represented as follows:

$$E = \sum E_i + exp(-\mu * G_i) \quad (4)$$

where E_i denotes the loss of the *i-th* branch, G_i denotes the sum of the boundary gradients for preserving the boundary of the segmentation, and μ controls the proportion of G_i which was set to 0.001. Then, we froze the parameters of multi-scale branches and trained the remaining parameters for GA prediction. We set the GA label maps at the next time for prediction supervision. The loss function for prediction can be seen in equation (3).

We used the gradient descent algorithm with the batch size of 20 to optimize the proposed network. The learning rate was set to 0.0001. The iterations in this network were set to 50000. Our network was implemented with Tensorflow based on Python3.5.

3. RESULTS

SD-OCT volumes from 38 eyes in 29 patients at the Byers Eye Institute of Stanford University comprising a total of 118 longitudinal SD-OCT examinations over a mean of 2.5 years were included in this study. The examinations were obtained from consecutive patients diagnosed with GA by an SD-OCT device (Cirrus OCT; Carl Zeiss Meditec, Inc., Dublin, CA). Each SD-OCT volume contains 1024×512×128 voxels with a corresponding trim size of 2mm×6mm×6mm.

In this paper, we used two criteria to quantitatively evaluate the performances of each comparison method: the overlap ratio (OR) and the dice index (DI).

3.1: Segmentation and Prediction of GA from Patient-Dependent Testing

Suppose each patient has n (n>2) volumes and the first n−2 volumes are used for training the network parameters to segment the n−1-*th* volume and predict the remaining n-*th* volume. In the training stage, the n − 1-*th* volume just provided the ground truth to validate the prediction of the n−2-*th* volume. In the testing stage, the n−1-*th* and the remaining n-*th* ground truth were used to validate the segmentation and prediction of the n − 1-*th* volume respectively. The evaluation comprised a total of 118 volumes from 38 eyes of 29 patients, where 74 out of 118 eyes were used for network training and only the eyes with more than 2 volumes were considered in the evaluation, which comprise 22 volumes from 22 eyes of 22 patients.

Table 1. The mean overlap ratio (OR) and the dice index (DI) of GA segmentation results over 38 eyes.

	OR	DI
Chen's method [6]	0.73	0.84
U-Net [16]	0.75	0.86
Niu's method [7]	0.81	0.89
Ji's method [8]	0.86	0.92
The proposed method	0.84	0.91

Table 2. The mean overlap ratio (OR) and the dice index (DI) of GA prediction results over 38 eyes.

	OR	DI
Niu's method [9]	0.68	0.81
The proposed method	0.69	0.82

GA segmentation results from our method were compared with *Chen*'s method [6], U-Net [16], *Niu*'s method [7] and *Ji*'s method [8] respectively, as listed in Table 1. We can find that the segmentation results from our proposed method are superior to U-Net, *Chen*'s and *Niu*'s method, but slightly lower than *Ji*'s method. That is because *Ji* utilized more complete information in the axial direction compared to the projection image. Besides, *Ji*'s method treated the semantic segmentation as the per-pixel classification and thus brought lower efficiency compared with our method. When testing a whole cube, the time cost of our method is 0.4s compared with the time cost in excess of 60s from *Ji*'s method.

GA prediction results from our method were compared with *Niu*'s method [9], as listed in Table 2. Our automated prediction results are better than those of *Niu*'s method. Besides, our method is simpler than the traditional machine learning methods based on hand picked features.

Fig. 4 shows the visualization of the joint segmentation and prediction results from our proposed method. First row shows the GA segmentation at the current time and the second row shows the GA prediction at the next time. As it

can be seen from Fig. 4, high agreements can be achieved between our automated results and the ground truth.

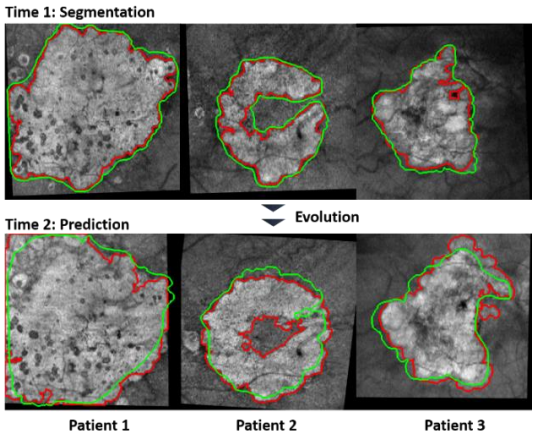

Fig. 4. Visualization of the joint segmentation and prediction results. The red indicates the ground truth and the green line indicates the automated results from our method.

3.2: Segmentation and Prediction of GA from Patient-Independent Testing

To further verify the performance of our proposed model on patient-independent testing, some experiments were designed via leave-one-out cross-validation. The evaluation comprised a total of 118 volumes from 38 eyes of 29 patients, where the training data from m−1 (m=38) eyes were used to build the model and the testing data at the first time of the remaining 1 eye were used for validation. Table 3 summarizes the average quantitative results with patient-independent testing. The performance of the segmentation and prediction algorithm from patient-independent testing declined approximately 4% and 3% respectively compared with the previous results. That is because the similarity between the training data and the testing data goes down. However, the quantitative results still keep relatively stable.

Table 3. The summarizations of GA segmentation and prediction results from patient-independent testing.

	OR	DI
Segmentation Accuracy	0.80	0.88
Prediction Accuracy	0.66	0.79

4. CONCLUSION

In this paper, we presented a multi-scale deep convolutional neural network for the joint segmentation and prediction of GA more efficiently in SD-OCT images. Our experiments showed that the proposed architecture can achieve an outstanding performance by the multi-scale feature extraction and the training in stages. The comparison with the existing methods also demonstrated the potential applications of neural network techniques for medical image segmentation and prediction. In the future work, the network architecture for the joint segmentation and prediction will be extended to 3D spaces based on the integrated SD-OCT volumetric data.

5. REFERENCES

[1] Resnikoff S, Pascolini D, Etya'als D, "Global data on visual impairment in the year 2002," Bull World Health Organ, vol. 82, pp. 844–851, 2004.
[2] Bressler NM, Bressler SB, Fine SL, "Age-related macular degeneration," Surv Ophthalmol, vol. 32, pp. 375–413, 1988.
[3] Tolentino MJ, Dennrick A, John E, Tolentino MS, "Drugs in phase ii clinical trials for the treatment of age-related macular degeneration," Expert Opin Invest Drugs, vol. 242, pp. 183–199, 2015.
[4] Chaikitmongkol V, Tadarati M, Bressler N M, "Recent approaches to evaluating and monitoring geographic atrophy," Curr Opin Ophthalmol, vol. 27, pp. 217–223, 2016.
[5] Yehoshua Z, Rosenfeld PJ, Grgori G, et al. Progression of geographic atrophy in age related macular degeneration imaged with spectral domain optical coherence tomography. Ophthalmology, Vol.118, pp. 679–686, 2011.
[6] Chen Q, de Sisternes L, Leng T, Zheng L, Kutzscher L, Rubin DL, "Semi-automatic geographic atrophy segmentation for SD-OCT images," Biomed Opt Express, vol. 4, pp. 2729–2750, 2013.
[7] Niu S, de Sisternes L, Chen Q, Leng T, Rubin DL, "Automated geographic atrophy segmentation for SD-OCT images using region-based CV model via local similarity factor," Biomed Opt Express, vol. 7, pp. 581–600, 2016.
[8] Z Ji, Q Chen, S Niu, T Leng, DL Rubin, "Beyond Retinal Layers: A Deep Voting Model for Automated Geographic Atrophy Segmentation in SD-OCT Images," Translational Vision Science & Technology, vol. 7, no. 1, 2018.
[9] Niu S, de Sisternes L, Chen Q, Rubin DL, Leng T, "Fully automated prediction of geographic atrophy growth using quantitative spectral-domain optical coherence tomography biomarkers," Ophthalmology, vol. 123, pp. 1737–1750, 2016.
[10] Q. Chen, S. Niu, H. Shen, T. Leng, L. de Sisternes, and D. L. Rubin, "Restricted summed-area projection for geographic atrophy visualization in SD-OCT images," Translational Vision Science & Technology, vol. 4, no. 5, 2015.
[11] Z. Yehoshua, C. A. A. Garcia Filho, F. M. Penha, G. Gregori, P. F. Stetson, W. J. Feuer, and P. J. Rosenfeld, "Comparison of geographic atrophy measurements from the OCT fundus image and the sub-RPE slab image," Ophthalmic Surg. Lasers Imaging Retina, vol. 44 no. 2, pp. 127–132, 2013.
[12] Liu C, Yuen J, Torralba A, "SIFT flow: dense correspondence across scenes and its applications," IEEE Trans Pattern Anal Mach Intell, vol. 33, pp. 1–17, 2011.
[13] E Shelhamer, J Long, and T Darrell. Fully Convolutional Networks for Semantic Segmentation. IEEE Conference on Computer Vision and Pattern Recognition (CVPR), 2015.
[14] L Chen, Y Zhu, G Papandreou, F Schroff, H Adam, "Encoder-Decoder with Atrous Separable Convolution for Semantic Image Segmentation," IEEE Conference on Computer Vision and Pattern Recognition (CVPR), 2018.
[15] TY Lin, P Goyal, R Girshick, K He, P Dollar, "Focal Loss for Dense Object Detection," IEEE Trans Pattern Anal Mach Intell, vol. 99, pp. 2999-3007, 2017.
[16] O Ronneberger, P Fischer, and T Brox, "U-Net: Convolutional Networks for Biomedical Image Segmentation," International Conference on Medical Image Computing and Computer Assisted Intervention (MICCAI), 2015.