Computers in Biology and Medicine 105 (2019) 102–111

Contents lists available at ScienceDirect

Computers in Biology and Medicine

journal homepage: www.elsevier.com/locate/compbiomed

Automated geographic atrophy segmentation for SD-OCT images based on two-stage learning model

Rongbin Xu[a], Sijie Niu[a,*], Qiang Chen[b], Zexuan Ji[b], Daniel Rubin[c,d], Yuehui Chen[a]

[a] Shandong Provincial Key Laboratory of Network based Intelligent Computing, School of Information Science and Engineering, University of Jinan, Jinan, China
[b] School of Computer Science and Engineering, Nanjing University of Science and Technology, Nanjing, 210094, China
[c] Department of Medicine (Biomedical Informatics Research), Stanford University School of Medicine Stanford, CA, 94305, USA
[d] Department of Radiology, Stanford University, Stanford, CA, 94305, USA

A B S T R A C T

Automatic and reliable segmentation for geographic atrophy in spectral-domain optical coherence tomography (SD-OCT) images is a challenging task. To develop an effective segmentation method, a two-stage deep learning framework based on an auto-encoder is proposed. Firstly, the axial data of cross-section images were used as samples instead of the projection images of SD-OCT images. Next, a two-stage learning model that includes offline-learning and self-learning was designed based on a stacked sparse auto-encoder to obtain deep discriminative representations. Finally, a fusion strategy was used to refine the segmentation results based on the two-stage learning results. The proposed method was evaluated on two datasets consisting of 55 and 56 cubes, respectively. For the first dataset, our method obtained a mean overlap ratio (OR) of 89.85 ± 6.35% and an absolute area difference (AAD) of 4.79 ± 7.16%. For the second dataset, the mean OR and AAD were 84.48 ± 11.98%, 11.09 ± 13.61%, respectively. Compared with the state-of-the-art algorithms, experiments indicate that the proposed algorithm can provide more accurate segmentation results on these two datasets without using retinal layer segmentation.

1. Introduction

Age-related macular degeneration (AMD) is the most common cause of irreversible blindness among the elderly individual and often presents with various phenotypic manifestations [1]. The advanced stage of non-exudative AMD is characterized by geographic atrophy (GA), which is mainly caused by atrophy of the retinal pigment epithelium (RPE) [2,3]. One of the major causes of visual acuity loss is the development of GA, which is generally associated with retinal thinning and loss of RPE and photoreceptors. To monitor the GA progression objectively or make treatment decisions, clinicians need to quantify and characterize morphologic alternations that appear within the atrophic area. However, manual labeling is time-consuming and subject to inter-observer variability, which can potentially result in qualitative differences. Therefore, automatic GA segmentation plays an important role in the diagnosis of advanced AMD and predicting future expansion of GA.

Most automatic or semi-automatic segmentation algorithms seem to be based on 2D color fundus photographs (CFP) or fundus autofluorescence (FAF), which can generally produce useful results [4–7]. However, these methods are just applied to quantify the atrophic area and failed to identify the retinal structure axially in fundus imaging modalities. Compared with fundus imaging, spectral-domain optical coherence tomography (SD-OCT) can non-invasively generate high-resolution three-dimensional (3D) representations of retinal structures, which allows the axial differentiation of retinal structures and the generation of volumetric image data. GA regions are visualized by considering an axial projection of the volumetric data (*en face* OCT fundus images), which can be used to accurately identify imaging characteristics of GA and provide detailed anatomic assessments.

Previous methods have been proposed to estimate the intra-retinal layers in images with GA using graph theory and dynamic programming, and the segmentation results were later used to measure the thickness [8] or volume [9] of RPE, which can be used for the evaluation of GA. However, it is hard for these methods to produce the outlines of the GA regions. Previous methods have used the projection images generated from SD-OCT volumetric data to identify the GA regions directly [10]. Tsechpenakis et al. [11] proposed a geometric deformable model driven by dynamically updated probability fields, which is used to segment the GA in dry AMD in human eyes. Chen et al. [12] proposed a semi-automated GA segmentation algorithm for SD-

* Corresponding author.
E-mail address: sjniu@hotmail.com (S. Niu).

https://doi.org/10.1016/j.compbiomed.2018.12.013
Received 20 September 2018; Received in revised form 27 December 2018; Accepted 27 December 2018

OCT images. A geometric active contour model was employed to detect and segment the extent of GA in the projections images automatically. A level set approach was also developed to segment GA regions in both SD-OCT and FAF images [13,14]. However, these methods need seed selections for initialization. To improve the segmentation accuracy, Niu et al. [15] proposed an automatic method that combines a region-based C−V model with a local similarity factor in projection images of a choroid sub-volume. However, all of the methods discussed thus far principally identify GA regions based on the *en face* OCT fundus images, which are sensitive to the segmentation of RPE.

With the development of deep neural networks, the capability of automatic feature extraction has seen tremendous improvement, and learned features are highly convolved to encode the intrinsic structures of an image for classification, recognition, and segmentation [16–20]. Deep learning has been successfully applied to medical image analysis, including liver segmentation [16], prostate segmentation [17], and brain tissue segmentation [18]. Ji et al. [19] proposed a deep voting model for automated GA segmentation of SD-OCT images without retinal layer segmentation, which achieved high segmentation accuracy. Accurately segmenting lesions in B-scan images is a challenging task because the same lesions have a wide variety of manifestations in different patients, so it is hard to learn the representative features with the same method. Therefore, a two-stage learning framework was designed to obtain deep representations and structures of SD-OCT images to segment lesions.

An effective two-stage learning model was proposed based on auto-encoder for automated GA segmentation. The model includes an offline-learning phase and a self-learning phase. The axial data of cross section images were used as input data to feed into the network, and then the offline-learning model and self-learning model were used to determine the categories of the axial data. Finally, the integration strategy was used to refine the segmentation results based on the two-stage learning results [21]. Experimental results show that the proposed method can provide more accurate segmentation results compared with state-of-the-art methods.

Compared with the existing methods, the main contributions of proposed method can be summarized as follows: (1) The self-learning model reduces the influence of image diversity on segmentation by learning the discriminative features of individual cube. (2) A fusion strategy is proposed to reduce the false positives effectively by combining two-stage learning results and to improve the accuracy of the segmentation results. (3) We conducted comprehensive experiments on two datasets, and experimental results indicate that the proposed method achieves superior performance over the state-of-art approaches on all these datasets. This paper is organized as follows. We introduce the auto-encoder model and the details of offline-learning and self-learning in section 2. Section 3 explains the results of the experiment. Important relevant issues are discussed in section 4.

2. Methodology

Fig. 1 shows a flowchart of the proposed method. An offline-learning model was first developed and used for learning the common features from training samples. A self-learning model is then used exploited for learning the discriminative features to effectively remove false positives generated by the offline-learning model. Finally, the outputs of these two models are fused to obtain the segmentation results.

2.1. Preprocessing

SD-OCT images contain speckle noise due to the wavelength of the imaging beam and the structural details of the imaged object. Therefore, a bilateral filter [5] is used to reduce the noise from the images, which is better for edge preservation compared with traditional denoising methods. As shown in Fig. 2, a variety of structural differences between the SD-OCT cube and B-scan images were obvious due to the differences of the retinal structure and the characteristics of SD-OCT imaging. Fig. 2 (a) shows a full projection image of one SD-OCT cube with GA. The ground truth and Chen's segmentation result are overlaid on projection image with red lines and green lines, respectively. The projection image clearly contains obvious intensity inhomogeneity, and the contrast between the background and lesion is low. These create challenges because they can easily lead to some GA lesions being omitted, as in Chen's segmentation result.

Based on Fig. 2 (a), three B-scans labeled with blue lines were selected, and the corresponding images are shown in Fig. 2 (b), where the GA lesions are covered by blue regions. However, B-scans cannot directly reflect the morphological characteristics of lesions, which can be identified as GA by the thickness of the RPE and the length of the reflected light band. Fig. 2 (c) shows three GA gray signals (A-scans) with red lines and three non-GA gray signals (A-scans) with blue lines, which were selected from three B-scans. There are obvious distinguishing differences between them. In this case, we can transform the image segmentation problem into a classification problem for the axial data of B-scans by using the axial data as samples.

Thus, 3D SD-OCT cubes were converted into 2D B-scan sets as in Fig. 2 (b). These B-scans were used to extract axial data to form training and testing samples. As shown in Fig. 2(a) and (b), for each OCT image, each pixel in the projection image is generated by averaging the D-dimensional vector x along the axial A-scan lines. The dataset is defined as $X = \{(x_i, y_i)|x_i \varepsilon R^D, y_i \varepsilon L, i = 1, ...M\}$, where M is the total numbers of A-scans in the dataset, and y_i is the class label of the vector x. For each OCT image, the dimension of each vector x is $D = 1024$. For the label set, our goal is to classify the axial data of OCT images as GA regions and non-GA regions, so the label set is $L = \{1, 0\}$. Therefore, the sample space can be mapped into the label space by the mapping function $f(\cdot): x \rightarrow y$.

2.2. Offline-learning model

As shown in Fig. 1, the proposed offline-learning model is stacked with three sparse auto-encoders (SAEs), and its structure is composed of five layers, including an input layer, three hidden layers, and an output layer. The axial data with 1024 dimensions were fed as input data into the model to obtain deep representations, and then a classifier was trained by the representations and its corresponding labels. Finally, a fine-tuning process was adopted to optimize the offline-learning model.

The auto-encoder is an unsupervised nonlinear neural network that attempts to make its reconstructed images approximating its input images. An auto-encoder is composed of' an encoder and a decoder. The encoder maps the input to a hidden representation through the function $h = f^{(1)}(W^{(1)}\ x+\ b^{(1)})$, where the superscript (1) represents the first layer, f is a transfer function of the encoder, W is the weight matrix, and b is the bias vector. The decoder then maps this representation h back to the original input x as $\hat{x} = f^{(2)}(W^{(2)}h + b^{(2)})$, where the superscript (2) represents the second layer, h is the representations of input data, and W, f, and b have the same meaning as in the first layer. Finally, the model was optimized based on a cost function that measures the error between the input data and the reconstruction. The cost function is defined as the mean squared error function:

$$J(W, b) = \left[\frac{1}{m}\sum_{i=1}^{m}\left(\frac{1}{2}||f(x^{(i)} - \hat{x}^{(i)})||^2\right)\right] + \lambda \times \Omega_{weights} + \beta \times \Omega_{sparsity} \quad (1)$$

$\Omega_{weights}$ is the L_2 regularization term constraint with coefficient λ, which is defined in equation (2), where L is the number of hidden layers, n is the number of samples, and k is the number of variables in the training data:

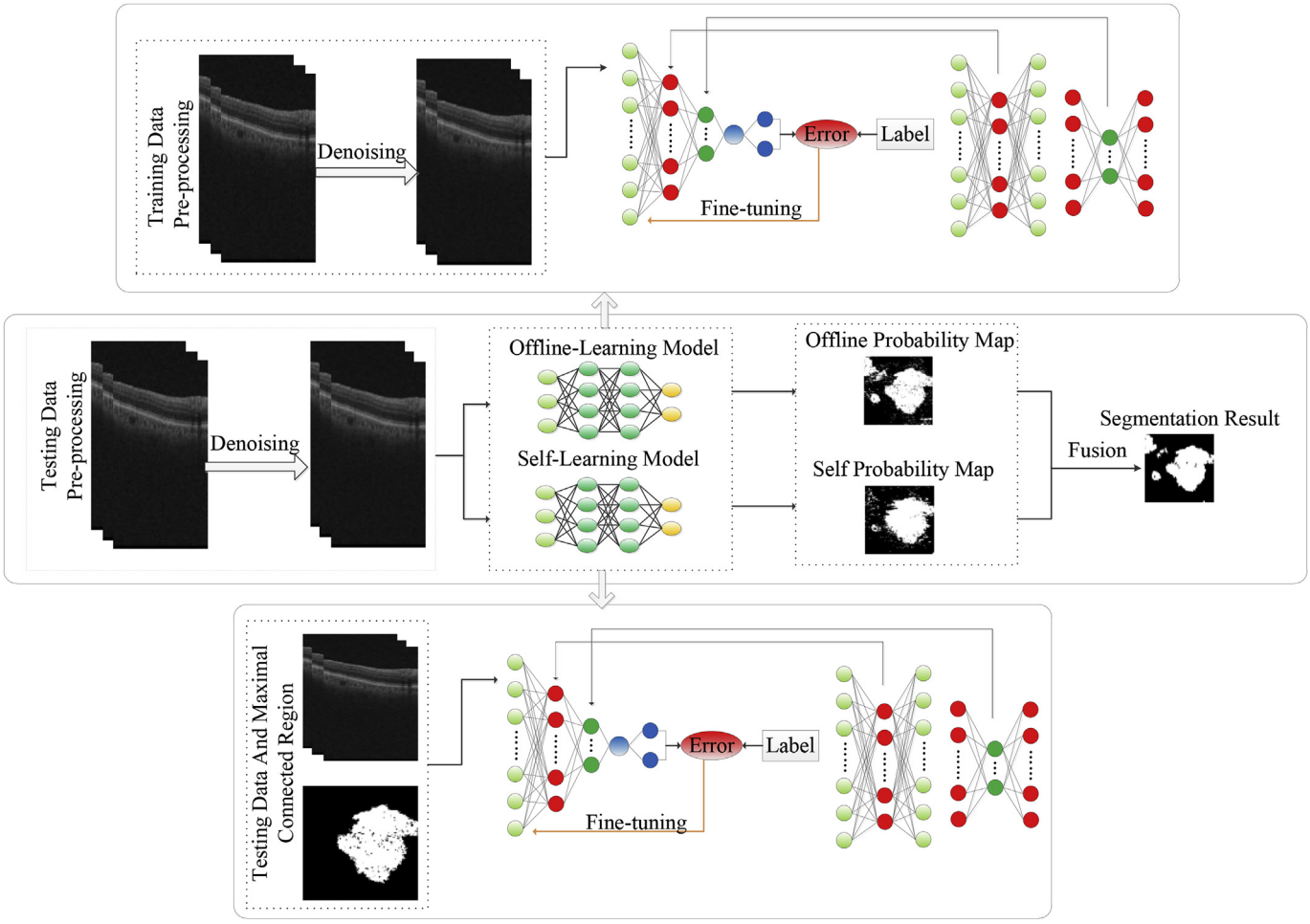

Fig. 1. The flowchart of the two-stage learning model.

$$\Omega_{weights} = \frac{1}{2}\sum_{l=1}^{L}\sum_{j=1}^{n}\sum_{i=1}^{k}(W_{ji}^{(l)})^2 \tag{2}$$

In order to ensure the sparsity of the output from the hidden layer, the sparsity regulation term constraint was adopted as follows:

$$\Omega_{sparsity} = \sum_{j=1}^{s2} KL(\rho||\hat{\rho}_j),\ where\ \hat{\rho}_j = \frac{1}{m}\sum_{i=1}^{m}[a_j^{(2)}(x^{(i)})] \tag{3}$$

For each auto-encoder, the main goal is to obtain the optimized parameters to obtain accurate deep representations by minimizing the loss calculated by equation (1). The gradient descent is used for training the auto-encoder.

Multiple SAEs are stacked together by feeding the output layer from the low-level SAE as the input layer of a high-level SAE to form the stacked sparse auto-encoder (SSAE), which is able to extract more useful and high level features. However, high-level features learned from the unsupervised SSAE are only data-adaptive and cannot discriminate the GA and background. To make the learned feature representation discriminative, supervised fine-tuning is often adopted by stacking another classification output layer on the top of the encoding part of the SSAE.

2.3. Self-learning model

It is difficult for the offline learning model to produce accurate segmentation results due to the variety of the images. In experimental observations, several normal regions were misclassified into lesion regions, which resulted in false positives (the most common segmentation errors). The reason for these misclassifications is that there are many similar features between the normal regions and lesion regions. In this case, if the network can learn to discriminate specific false positives from each cube, the performance can be efficiently improved. Thus, a self-learning model was proposed to learn the discriminative features to improve the segmentation performance further [22]. More specifically, we trained a specific model to learn the discriminative features of each individual cube to compensate for the deficiency of discrimination caused by different patients and limited training samples.

A key step for training the self-learning model is the selection of training samples, which should select representative samples to enhance the discrimination. In the self-learning stage, the coarse labels generated from the offline-learning probability map were used to train the classifier. Fig. 3 shows an example of selecting training samples for the self-learning model. Firstly, the probability map in Fig. 3 (a) resulting from the offline-learning model was converted into the binary map in Fig. 3 (b). The maximal connected region of the binary map was then selected to be the candidate region of positive labels. In order to obtain more accurate negative labels, a rectangle that can cover the candidate regions of positives was adopted, as shown in Fig. 3 (d). This can minimize the mistakes of negative sample selection.

When choosing negative labels, the low-probability regions of the offline-learning probability map were considered as candidates. Finally, the corresponding axial data were extracted based on the position information to form the training data. To ensure the balance of positive samples and negative samples, 2000 positive samples and 2000 negative samples were chosen. In addition, we used the same architecture as the offline-learning model to implement the self-learning model.

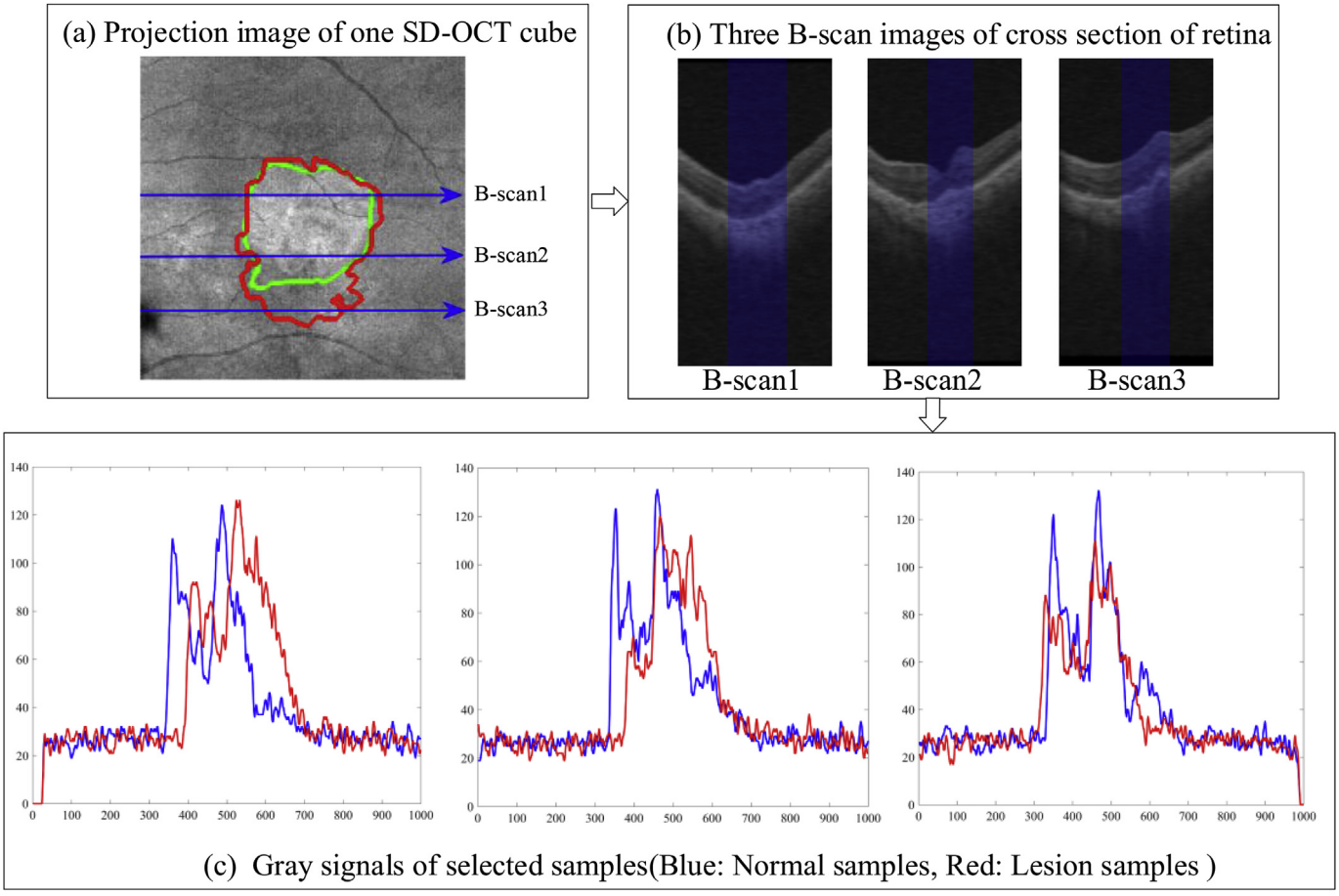

Fig. 2. The structural difference in SD-OCT data. (a) The full projection image of SD-OCT cube with GA where the ground truth and Chen's segmentation result are overlaid on projection image with red line and green line respectively. (b) Three B-scan images of the cross section selected from (a) where the blue regions are GA lesion. (c) The gray signals of selected A-scans, where the A-scans with GA and A-scans without GA are represented by red lines and blue lines respectively.

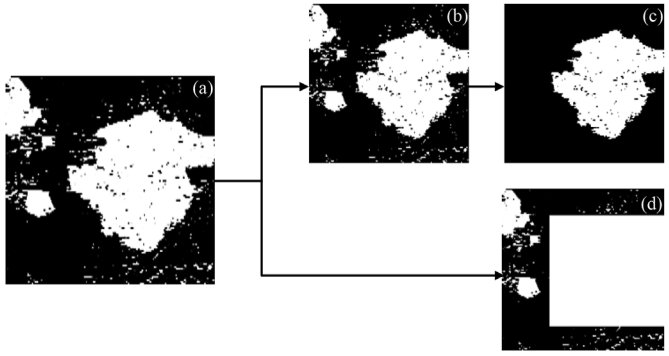

Fig. 3. An example of selecting training data for self-learning model. (a) The result produced by offline-learning model. (b) The binary image produced by (a). (c) The candidate area for positive samples. (d) The candidate area for negative samples.

2.4. Fusion strategy

As mentioned, the output of offline-learning model and self-learning model are fused to obtain the final segmentation result. The fusion strategy is defined as follows:

$$P = \begin{cases} (P_f + P_s)/2 & |P_f - P_s| \le T \\ P_f & |P_f - P_s| > T \end{cases} \quad (4)$$

where P is the fusion probability, and P_f and P_s are the predicted probability of offline-learning model and the self-learning model, respectively. There may be some mistakes in the self-learning segmentation because the provided labels for self-learning may be not correct. Therefore, a threshold was set for the output of the two models to reduce the influence of self-learning mistakes and to improve the robustness of the whole model. The average of the two outputs is taken as the final result when the absolute value of the difference between P_f and P_s is less than T. When this condition is not met, we use the result of the offline-learning model as the final result.

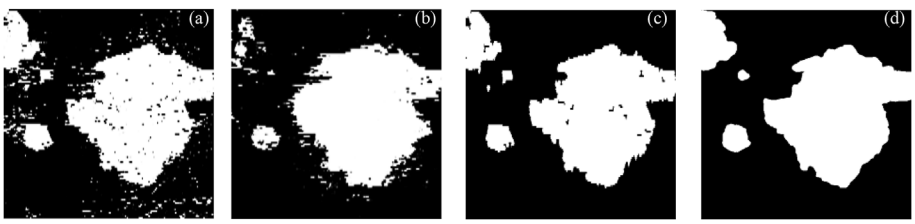

Fig. 4. An example of the fusion strategy. (a) The result of the offline-learning model. (b) The result of the self-learning model. (c) The fusion result. (d) Ground truth.

As shown in Fig. 4, there are some false positives in the result of offline-learning model because of the model cannot learn the detailed features of each individual cube. The self-learning model was used to overcome the problem, but some false negatives were produced because of the errors in the process of label selection. Therefore, we fused the results of offline-learning and self-learning, as shown in the final result in Fig. 4 (c). By comparing to ground truth, we can conclude that the proposed fusion strategy is effective for SD-OCT image segmentation.

3. Experiment and results

3.1. Datasets and experimental platform

Two different datasets [12,15,19] were used to evaluate the performance of the proposed method, and the training and testing datasets contain GA lesions. Each SD-OCT image set was acquired over a 6×6 mm^2 macular area with a 2-mm axial depth (corresponding to 1024 pixels) using a Cirrus device (Carl Zeiss Meditec, Inc., Dublin, CA). The first dataset (dataset 1) consists of 55 cubes (the data size of 51 of the cubes is $1024 \times 512 \times 128$, and the others are $1024 \times 200 \times 200$). Two different clinical ophthalmologists manually outlined each OCT cube twice in different sessions to form the "gold standard." A region was considered as a positive sample when two experts outlined the region as GA each time; otherwise, the region is considered as negative sample.

The second dataset (dataset 2) consists of 56 cubes (all of the data sizes are $1024 \times 200 \times 200$). A clinical ophthalmologist manually drew the outline of the GA area based on FAF images and registered the outline in the projection images to obtain the ground truth. In the training phase, we randomly selected 100,000 axial data with GA as positive samples and 100,000 normal axial data without GA as negative samples for each dataset. In the testing phase, 3D cubes were directly fed into the proposed model to obtain the results. Both the training and testing phases of the segmentation model were run on a platform with an Intel(R) Xeon(R) processor with 256 GB of RAM and no GPU.

3.2. Parameter analysis and valuation metrics

3.2.1. Network parameters

Table 1 summarizes the structure of the auto-encoder and the parameter settings for the auto-encoder training. The coefficients λ and β are adjusted manually by experiments. The SSAE network consists of five layers, including an input layer, three hidden layers, and an output layer. We investigated two models that have the exact same structure and parameter settings but different training data.

3.2.2. Evaluation metrics

Three criteria were used to evaluate the performance of the proposed method: the overlap ratio (OR), absolute area difference (AAD), and correlation coefficient (CC). OR is defined as the percentage of area in which both segmentation methods agree with respect to the presence of GA over the total area in which at least one of the methods detects GA (Jaccard index):

$$OR(X, Y) = Area(X \cap Y)/Area(X \cup Y) \quad (5)$$

where X and Y are the regions inside the segmented GA contour produced by two different methods (or graders), respectively. The operators $\cap$ and $\cup$ indicate union and intersection, respectively. The mean OR and standard deviation values are computed across scans in the datasets.

AAD measures the absolute difference between the GA areas segmented by two different methods:

$$AAD(X, Y) = |Area(X) - Area(Y)| \quad (6)$$

The mean AAD and standard deviation values are computed across scans in the datasets. CC was computed using linear correlation between the measured areas of GA computed by the segmentation of different methods or readers and measuring the linear dependence using each scan as an observation.

Table 1
The parameters of the network models.

Parameter	Value
Number of nodes in input layer	1024
Number of nodes in output layer	2
Number of nodes in layer 1	924
Number of nodes in layer 2	824
Number of nodes in layer 3	724
Unsupervised training epochs	1000
Supervised Training Epochs	2000
L2 weight regularization λ	0.004
Loss function	Msesparse
Sparsity regularization β	4
Training algorithm	trainscg

3.3. Fusion threshold setting

The final segmentation results are affected by the threshold T in the process of fusing the offline probability map and self-probability map. This threshold indicates the gap between two segmentation models. Fig. 5 shows the OR of the segmentation results of the two datasets by different threshold values in the range of [0.1, 1].

OR increases in the threshold range of [0.1, 0.9] and decreases in the range of [0.9, 1] in dataset 1. In dataset 2, the overlap ratio of the segmentation results remains stable in the threshold range of [0.1, 0.5] and decreases in the range of [0.5, 1]. For both datasets, we obtained the best performance at two different thresholds (0.9 and 0.5, respectively). To construct a unified segmentation model, we balanced the thresholds of the two segmentation models by setting it as 0.5.

3.4. Testing

3.4.1. Test I: segmentation results on the dataset with size of $1024 \times 512 \times 128$

In the first experiment, we validated the proposed model on dataset 1, which contains 55 SD-OCT cubes from eight patients. The testing time for each cube is 98 s, including the offline-learning testing and self-learning testing. As shown in Fig. 6, eight examples were selected to illustrate the performance of the model. The ground truth and the results of our segmentation results are overlaid on the projection image, where red lines represent the ground truth and green lines show our segmentation results. These examples show cases with different intensity inhomogeneity and complexity, which increase the difficulty of segmentation. However, the proposed method can avoid these problems and obtain more accurate results.

Fig. 7 shows a comparison of our segmentation results with three other methods. The red, white, green, yellow, and blue lines represent the ground truth and segmentation results from Chen [12], Niu [15], Ji [19], and the present study, respectively. For the second and fourth cases, all of these methods obtained better performance because the images have higher contrast. However, for the low-contrast images, the other methods fail to perform well. In the third case, Niu's method misclassifies the normal region as a lesion region, while Chen's method misclassifies the lesion region as a normal region. For the sixth case, Chen, Niu, and Ji's methods misclassify the lesion region as a normal region. For the first and fifth cases, both Chen and Niu's methods misclassify the lesion region as a normal region, and Ji's method misclassifies the normal region as a lesion region in the fifth case.

Tables 2 and 3 quantitatively compare between the segmentation results obtained with different methods and the ground truth on dataset 1 both with and without the training data. The proposed method fused the results of offline-learning and self-learning, while the SSAE

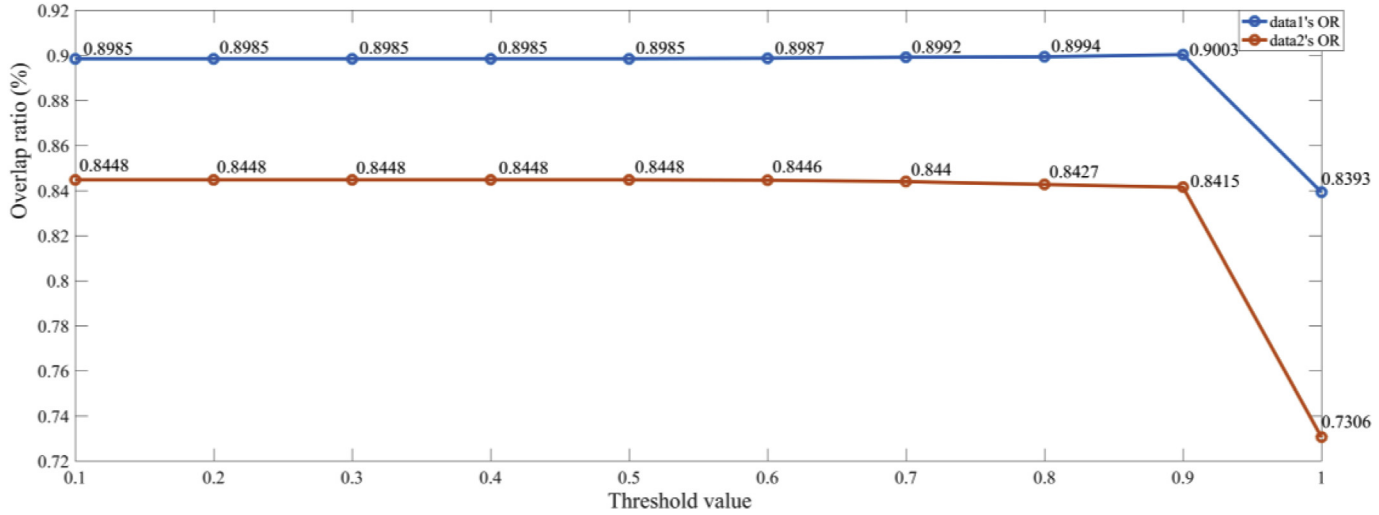

Fig. 5. The overlap ratio of segmentation result by different threshold value.

represents the results of offline-learning. By integrating the offline-learning and self-learning results, our segmentation result has a higher OR, lower AAD, and higher CC than Chen, Niu, and Ji's methods, indicating that our segmentation results are close to the manual outlines.

3.4.2. Test II: segmentation results on the dataset with a size of 1024 × 200 × 200

In the second experiment, we validated our models on dataset 2, which contains 56 SD-OCT cubes from 56 patients. The testing time for each cube is 50 s, including the offline-learning testing and self-learning testing. As shown in Fig. 8, eight examples were selected to illustrate the performance of the model. In each image, the ground truth and our segmentation results are overlaid on the projection. The red lines represent the boundary of the ground truth, and the green lines show the boundary of our segmentation results. The results of the proposed model are similar to the ground truth.

Fig. 9 qualitatively displays our segmentation results with those of the other three methods in six cases. In each figure, the red lines indicate the outline of the ground truth, and the white, green, yellow, and blue lines show Chen's [12], Niu's [15], Ji's [19], and our segmentation results, respectively. For the third case, all the method can perform well because of the obvious lesion features, while for other cases, Chen, Niu, and Ji's methods misclassify the normal region as a lesion region in the first and fifth cases. For the second case, Chen and Niu's methods misclassify the lesion region as a normal region. In the fourth and sixth cases, Chen, Niu, and Ji's methods misclassify the lesion region as a normal region. Thus, the segmentation results of our methods are better than those of the other three methods.

Tables 4 and 5 summarize the average quantitative results between the segmentation results and ground truth of the different methods on dataset 2. Table 4 shows the results from when the testing data included the training data, and Table 5 shows the results obtained without the training data. Compared with other methods, our segmentation result has a higher OR, lower AAD, and higher CC. The higher OR shows that our segmentation result is similar to the outline of the manual outline, and the low AAD indicates that the estimated areas of our model are similar to the manual productions. The higher CC indicated that our segmentation results are more similar to the ground truth.

For these two datasets, a majority of cases have a better segmentation performance, but there is a set of data with lower overlap. One of the main challenges for the automated segmentation of lesions from retinal images is that there are many interference factors, such as vascular and optic nerves. All of these factors affect the segmentation performance. The most direct solution is to remove these effects in the preprocessing step. However, there is a big difference in the structures of different people, and we have no common methods to address them.

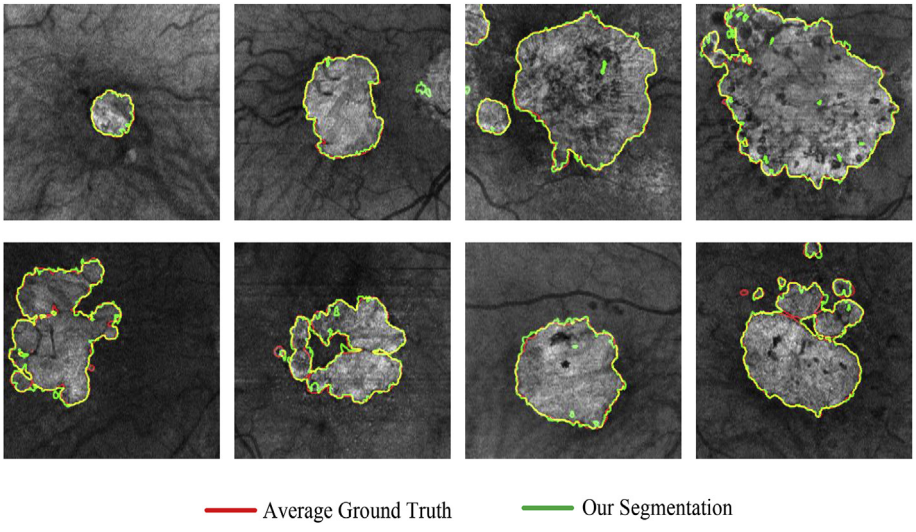

Fig. 6. Segmentation results and ground truth overlaid on full projection images for eight example cases in dataset 1.

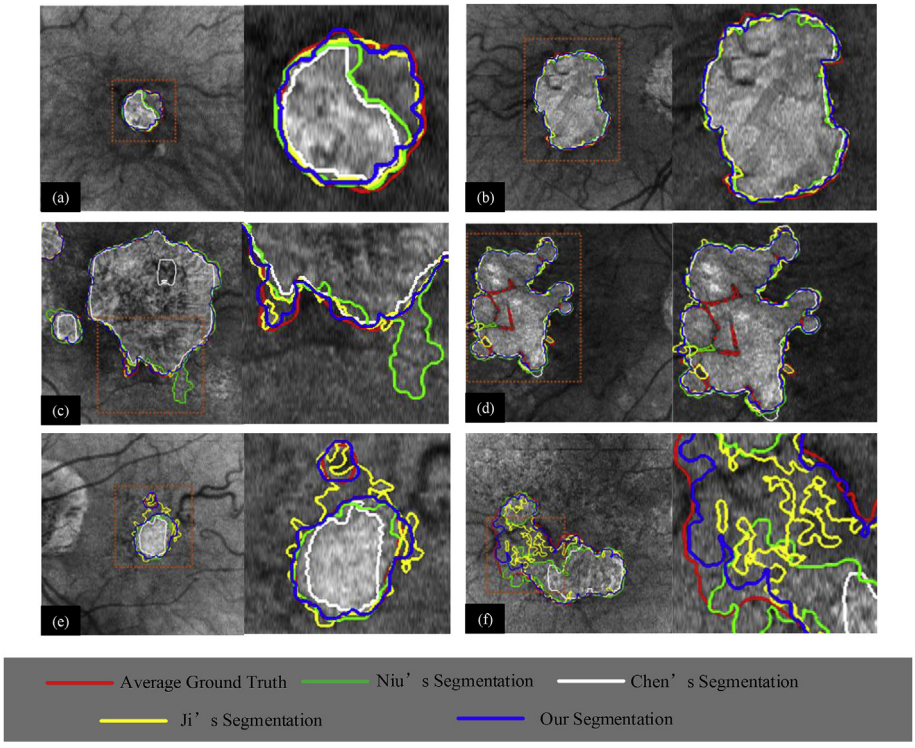

Fig. 7. Comparison of the segmentation result overlaid on projection images in dataset 1.

The self-learning strategy can train the models using the characteristics of images and effectively reduce the effect of the factors. Offline-learning can support self-learning to obtain the correct labels.

A case of failure is shown in Fig. 10. It is obvious that all the methods failed to segment the lesion, including the present method. This occurred because there are too many similar pixels that are

Table 2
Quantitative results (mean ± standard deviation) of segmentation methods and ground truth on dataset 1 including training data.

Methods	Criteria	Avg. expert	Expert A_1	Expert A_2	Expert B_1	Expert B_2
Chen's method	CC	0.970	0.967	0.964	0.968	0.977
	AAD[mm^2]	1.438 ± 1.26	1.308 ± 1.28	1.404 ± 1.31	1.597 ± 1.33	1.465 ± 1.14
	AAD[%]	27.17 ± 22.06	25.23 ± 22.71	26.14 ± 21.48	29.21 ± 22.17	27.62 ± 20.57
	OR[%]	72.60 ± 12.01	73.26 ± 15.61	73.12 ± 15.15	71.16 ± 15.42	72.09 ± 14.82
Niu's method	CC	0.979	0.975	0.976	0.976	0.975
	AAD[mm^2]	0.811 ± 0.94	0.758 ± 0.99	0.853 ± 1.04	0.984 ± 1.08	0.897 ± 1.05
	AAD[%]	12.95 ± 11.83	12.62 ± 12.86	13.32 ± 12.74	14.91 ± 12.65	14.07 ± 11.78
	OR[%]	81.86 ± 12.01	81.42 ± 12.12	81.61 ± 12.29	80.05 ± 13.05	80.65 ± 12.51
Ji's method	CC	0.986	0.986	0.985	0.985	0.991
	AAD[mm^2]	0.67 ± 0.73	0.55 ± 0.74	0.62 ± 0.80	0.82 ± 0.83	0.69 ± 0.66
	AAD[%]	11.49 ± 11.50	9.75 ± 11.35	10.32 ± 11.09	13.58 ± 12.41	11.73 ± 9.35
	OR[%]	86.94 ± 8.75	87.64 ± 8.75	87.71 ± 8.32	85.17 ± 9.40	86.37 ± 7.67
SSAE	CC	0.9829	0.9826	0.9814	0.9832	0.9832
	AAD[mm^2]	0.33 ± 0.26	0.27 ± 0.23	0.26 ± 0.25	0.22 ± 0.25	0.24 ± 0.23
	AAD[%]	1.67 ± 0.82	1.36 ± 0.74	1.28 ± 0.74	1.05 ± 0.75	1.18 ± 0.74
	OR[%]	71.54 ± 16.09	71.76 ± 15.64	72.05 ± 15.88	71.62 ± 15.73	72.06 ± 15.68
AlexNet	CC	0.9934	0.9901	0.9881	0.9877	0.9908
	AAD[mm^2]	0.36 ± 0.44	0.52 ± 0.61	0.64 ± 0.62	0.78 ± 0.72	0.68 ± 0.58
	AAD[%]	5.99 ± 6.57	7.88 ± 7.57	9.94 ± 8.00	12.30 ± 9.12	10.52 ± 6.19
	OR[%]	86.47 ± 8.24	84.03 ± 9.64	83.61 ± 9.08	81.70 ± 9.70	82.68 ± 8.88
VGG	CC	0.9945	0.9920	0.9897	0.9897	0.9915
	AAD[mm^2]	0.33 ± 0.46	0.40 ± 0.52	0.49 ± 0.58	0.65 ± 0.59	0.53 ± 0.54
	AAD[%]	5.91 ± 10.15	6.48 ± 9.12	7.91 ± 10.25	10.34 ± 9.98	8.53 ± 8.62
	OR[%]	86.61 ± 8.95	84.57 ± 9.92	84.21 ± 9.61	82.56 ± 10.23	83.46 ± 9.54
Our proposed model	CC	**0.9985**	**0.9945**	**0.9935**	**0.9930**	**0.9961**
	AAD[mm^2]	**0.18 ± 0.20**	**0.38 ± 0.53**	**0.45 ± 0.56**	**0.65 ± 0.64**	**0.52 ± 0.51**
	AAD[%]	**3.67 ± 6.02**	**6.02 ± 7.44**	**7.11 ± 7.83**	**9.86 ± 8.8**	**8.11 ± 6.16**
	OR[%]	**90.73 ± 5.75**	**88.29 ± 7.59**	**88.10 ± 7.24**	**85.98 ± 8.04**	**87.12 ± 6.74**

Table 3
Quantitative results (mean ± standard deviation) of segmentation methods and ground truth on dataset 1 without training data.

Methods	Criteria	Avg. expert	Expert A_1	Expert A_2	Expert B_1	Expert B_2
SSAE	CC	0.9754	0.9764	0.9754	0.9774	0.9767
	AAD[mm^2]	0.39 ± 0.28	0.34 ± 0.27	0.33 ± 0.28	0.30 ± 0.28	0.31 ± 0.27
	AAD[%]	2.44 ± 1.13	2.09 ± 1.08	2.0 ± 1.07	1.75 ± 1.08	1.90 ± 1.10
	OR[%]	68.41 ± 17.21	69.22 ± 16.72	69.62 ± 17.08	69.60 ± 16.89	69.86 ± 16.86
AlexNet	CC	0.9899	0.98.60	0.9832	0.9828	0.9866
	AAD[mm^2]	0.39 ± 0.45	0.56 ± 0.60	0.69 ± 0.62	0.82 ± 0.72	0.72 ± 0.60
	AAD[%]	7.76 ± 7.45	10.01 ± 8.39	12.39 ± 9.03	14.98 ± 10.24	12.97 ± 7.28
	OR[%]	83.45 ± 9.46	80.65 ± 10.83	80.04 ± 10.19	78.04 ± 10.79	79.00 ± 10.07
VGG	CC	0.9919	0.9888	0.9855	0.9858	0.9879
	AAD[mm^2]	0.33 ± 0.45	0.42 ± 0.52	0.52 ± 0.59	0.68 ± 0.60	0.56 ± 0.55
	AAD[%]	7.05 ± 10.82	7.87 ± 9.90	9.75 ± 11.27	12.45 ± 10.73	10.36 ± 9.43
	OR[%]	83.71 ± 10.02	81.18 ± 11.16	80.68 ± 10.73	78.94 ± 11.28	79.85 ± 10.68
Our proposed model	CC	**0.9979**	**0.9930**	**0.9918**	**0.9914**	**0.9951**
	AAD[mm^2]	**0.21 ± 0.20**	**0.33 ± 0.4**	**0.38 ± 0.53**	**0.57 ± 0.60**	**0.45 ± 0.48**
	AAD[%]	**4.79 ± 7.16**	**6.33 ± 8.25**	**7.23 ± 8.99**	**10.22 ± 9.69**	**8.26 ± 6.97**
	OR[%]	**89.85 ± 6.35**	**87.35 ± 8.20**	**87.23 ± 7.88**	**84.98 ± 8.66**	**86.18 ± 7.33**

seriously affected by the light. Furthermore, for our method, the lesion region is too small to obtain enough self-learning training data, so it produced an undesirable segmentation result.

In recent years, convolutional neural networks (CNNs) have been extensively applied in medical image analysis and have achieved remarkable success in many applications [23–32]. Therefore, in addition to comparing the proposed with traditional methods like Chen and Niu's, we also compared it to AlexNet [23] and VGG-19 [24]. The training data were the same as that of the proposed method, and we used the axial data as samples (there are 200,000 samples in total, and the size of each sample is 1024 × 1). From Tables 2 and 4, it is obvious that the proposed method achieved better results. AlexNet is seriously affected by the quality of images, and when the images have intensity inhomogeneity, the results of AlexNet produce many false positives, while the proposed method can distinguish them well by the fusion strategy. VGG-19 is composed of multiple convolution layers, pooling layers, and fully connected layers. Its accuracy has been effectively improved, but the manifold layers make the network use much more resources than other methods. Lastly, the inspirit function of AlexNet and VGG-19 easily causes non-repairable neuronal necrosis, which results in insufficient gradients to optimize the whole model. For the SSAE model, a very obvious gap is produced compared with the proposed method because of the individual differences. The comparison results make it obvious that the self-learning can have a great influence on the segmentation results.

4. Discussion

In this paper, we proposed the two-stage learning method based on stack sparse auto-encoder for automated geographic atrophy segmentation in SD-OCT images. The proposed method integrates the self-learning and offline-learning to determine the label of each axial data. Then, we use fusion strategy to refine the segmentation results based on the two-stage learning results. The quantitative experimental results demonstrated that the proposed method has higher segmentation accuracy than state-of-the-art methods [12,15,19]. This indicates that the model could potentially contribute to automatically identifying GA regions and provide an objective and reliable quantitative assessments for measuring and tracking non-exudative age-related macular degeneration.

One of the main challenges for automated segmentation is that many non-GA tissues influence the GA segmentation. Furthermore, it is hard to use simple methods to remove these non-GA regions because of the irregularity and differences of the tissues between different patients. Thus, we proposed the self-learning model for the GA segmentation, which learns the features of each individual cube to reduce the false

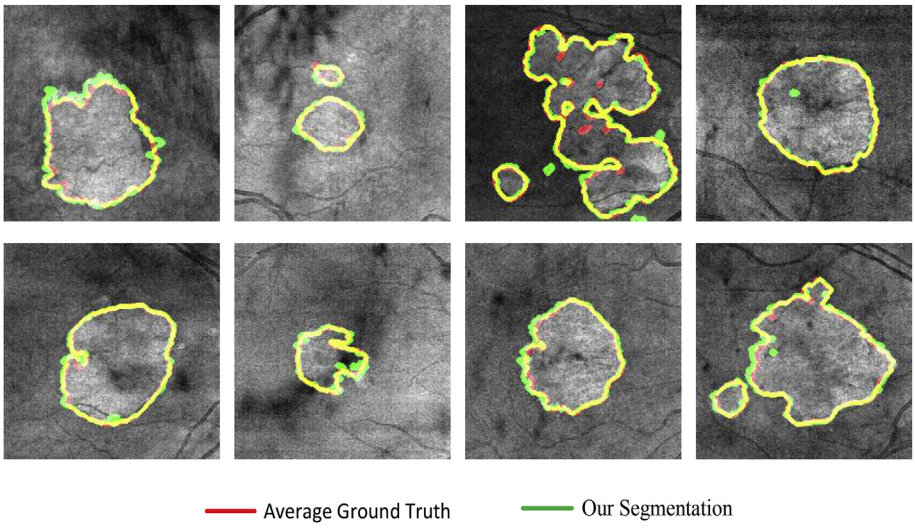

Fig. 8. Segmentation results and ground truth overlaid on full projection images for eight example cases in dataset 2.

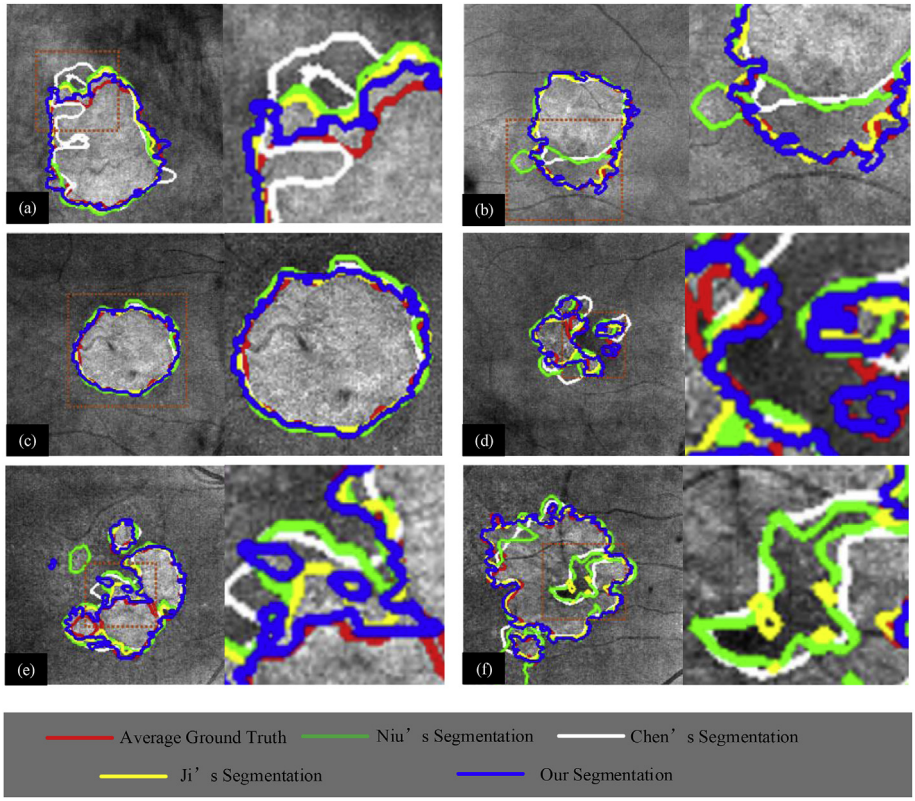

Fig. 9. Comparison of the segmentation result overlaid projection images in dataset 2.

Table 4
Quantitative results (mean ± standard deviation) of the segmentations and ground truth on dataset 2.

Method	OR (%)	AAD (mm^2)	AAD (%)	CC
Chen's method	65.88 ± 18.38	0.951 ± 1.28	19.68 ± 22.75	0.970
Niu's method	70.00 ± 15.63	1.215 ± 1.58	22.96 ± 21.74	0.979
Ji's method	81.66 ± 10.93	0.34 ± 0.27	8.30 ± 9.09	0.995
SSAE	68.97 ± 18.10	1.84 ± 1.09	36.71 ± 28.78	0.9757
AlexNet	75.79 ± 15.71	0.95 ± 0.84	21.67 ± 22.03	0.9839
VGG	74.72 ± 15.72	1.14 ± 0.79	24.66 ± 23.18	0.9867
Our proposed method	**84.55 ± 12.02**	**0.48 ± 0.46**	**11.09 ± 13.64**	**0.9953**

positives. Compared to other methods, the proposed method can improve the segmentation accuracy effectively by combining the results of offline-learning and self-learning.

On the other hand, the method has limitations as well. Some false positives and small holes are present in the segmentation results because we just used the axial data of A-scans as samples and failed to consider the spatial information of SD-OCT cubes. Another limitation is that we simply used the probability map from offline learning to obtain candidate labels, which could result in inaccurate self-learning models. As shown in Fig. 10, there is high similarity between the lesion region and the normal region, and the lesion region is too small to obtain enough features, so it is difficult to extract lesion features. Both the training and testing phases of the model were based on column data without considering the impact of patient independence. In the future, we plan to construct a 3D network to extract the spatial features and to consider the independence of different patients to improve the segmentation results further.

5. Conclusion

This paper presented an automatic algorithm for GA segmentation in SD-OCT images to quantify measurements of the extent and location of GA. The method uses axial data as samples to overcome the limitation of layer segmentation. Then, a two-stage deep learning model including offline-learning and self-learning was designed for the lesion segmentation. Quantitative experimental results showed that the algorithm has good consistency with different experts in manual segmentation. The algorithm could provide relatively reliable assistance for predicting the future expansion of GA.

Conflicts of interest

The authors declare that there are no conflicts of interest in this work.

Table 5
Quantitative results (mean ± standard deviation) of the segmentations and ground truth on dataset 2 without including training data.

Method	OR (%)	AAD (mm^2)	AAD (%)	CC
SSAE	68.88 ± 18.05	1.69 ± 1.03	30.76 ± 28.75	0.9736
AlexNet	75.7 ± 15.63	0.88 ± 0.79	21.73 ± 22.0	0.9826
VGG	74.62 ± 15.66	1.04 ± 0.74	24.7 ± 23.18	0.9856
Our proposed method	**84.48 ± 11.98**	**0.4418 ± 0.4334**	**11.09 ± 13.61**	**0.9948**

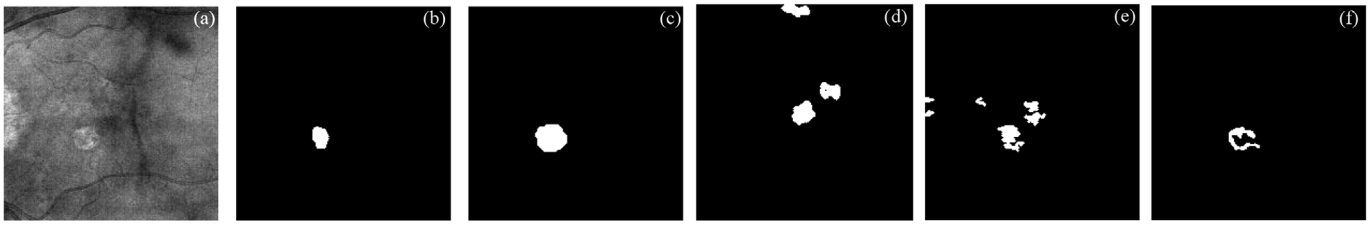

Fig. 10. Undesirable segmentation result.(a) The projection image of a SD-OCT cube. (b) Chen's segmentation result. (c) Niu's segmentation result. (d) Ji's segmentation result. (e) Our segmentation result. (f) Ground Truth.

Acknowledgments

This work is supported by the National Natural Science Foundation of China under Grant No. 61671242, 61701192, the Natural Science Foundation of Shandong Province, China, under Grant No. ZR2017QF004, the China Postdoctoral Science Foundation under Grants No. 2017M612178, the Shandong Provincial Key R&D Program (2016ZDJS01A12), the Shandong Provincial Key Research and Development Project (2017CXGC0810), and the National Key Research and Development Program of China (No. 2016YFC0106000).

References

[1] M. Fleckenstein, S. Schmitz-Valckenberg, C. Adrion, et al., Tracking progression with spectral-domain optical coherence tomography in geographic atrophy caused by age-related macular degeneration, Invest. Ophthalmol. Vis. Sci. 51 (8) (2010) 3846–3852.
[2] I. Bhutto, G. Lutty, Understanding age-related macular degeneration (AMD): relationships between the photoreceptor/retinal pigment epithelium/Bruch's membrane/choriocapillaris complex, Mol. Aspect. Med. 33 (4) (2012) 295–317.
[3] R.P. Nunes, G. Gregori, Z. Yehoshua, et al., Predicting the progression of geographic atrophy in age-related macular degeneration with SD-OCT en face imaging of the outer retina, Ophthalm. Surg., Lasers Imag. Retina 44 (4) (2013) 344–359.
[4] M. Niemeijer, B. Van Ginneken, J. Staal, et al., Automatic detection of red lesions in digital color fundus photographs, IEEE Trans. Med. Imag. 24 (5) (2005) 584–592.
[5] E. Trucco, A. Ruggeri, T. Karnowski, et al., Validating retinal fundus image analysis algorithms: issues and a proposal, Invest. Ophthalmol. Vis. Sci. 54 (5) (2013) 3546–3559.
[6] Z. Yehoshua, C.A.A. Garcia Filho, F.M. Penha, et al., Comparison of geographic atrophy measurements from the OCT fundus image and the sub-RPE slab image, Ophthalm. Surg., Lasers Imag. Retina 44 (2) (2013) 127–132.
[7] A.K. Feeny, M. Tadarati, D.E. Freund, et al., Automated segmentation of geographic atrophy of the retinal epithelium via random forests in AREDS color fundus images, Comput. Biol. Med. 65 (C) (2015) 124–136.
[8] S.J. Chiu, J.A. Izatt, R.V. O'Connell, et al., Validated automatic segmentation of AMD pathology including drusen and geographic atrophy in SD-OCT images, Invest. Ophthalmol. Vis. Sci. 53 (1) (2012) 53–61.
[9] F.A. Folgar, E.L. Yuan, M.B. Sevilla, et al., Drusen volume and retinal pigment epithelium abnormal thinning volume predict 2-year progression of age-related macular degeneration, Ophthalmology 123 (1) (2016) 39–50.
[10] Q. Chen, S. Niu, H. Shen, et al., Restricted summed-area projection for geographic atrophy visualization in SD-OCT images, Translat. Vision Sci. Technol. 4 (5) (2015) 2-2.
[11] G. Tsechpenakis, B. Lujan, O. Martinez, et al., Geometric deformable model driven by CoCRFs: application to optical coherence tomography, International Conference on Medical Image Computing and Computer-assisted Intervention, Springer, Berlin, Heidelberg, 2008, pp. 883–891.
[12] Q. Chen, L. de Sisternes, T. Leng, et al., Semi-automatic geographic atrophy segmentation for SD-OCT images, Biomed. Opt. Express 4 (12) (2013) 2729–2750.
[13] Z. Hu, G.G. Medioni, M. Hernandez, et al., Segmentation of the geographic atrophy in spectral-domain optical coherence tomography and fundus autofluorescence images, Invest. Ophthalmol. Vis. Sci. 54 (13) (2013) 8375–8383.
[14] L.A. Vese, T.F. Chan, A multiphase level set framework for image segmentation using the Mumford and Shah Model, Int. J. Comput. Vis. 50 (3) (2002) 271–293.
[15] S. Niu, L. de Sisternes, Q. Chen, et al., Automated geographic atrophy segmentation for SD-OCT images using region-based CV model via local similarity factor, Biomed. Opt. Express 7 (2) (2016) 581–600.
[16] Q. Dou, H. Chen, Y. Jin, et al., 3D deeply supervised network for automatic liver segmentation from CT volumes, International Conference on Medical Image Computing and Computer-assisted Intervention, Springer, Cham, 2016, pp. 149–157.
[17] Y. Guo, Y. Gao, D. Shen, Deformable MR prostate segmentation via deep feature learning and sparse patch matching, IEEE Trans. Med. Imag. 35 (4) (2017) 197–222.
[18] G. Zhao, X. Wang, Y. Niu, et al., Segmenting brain tissues from Chinese visible human dataset by deep-learned features with stacked autoencoder, BioMed Res. Int. 6 (2016) 1–12.
[19] Z. Ji, Q. Chen, S. Niu, et al., Beyond retinal layers: a deep voting model for automated geographic atrophy segmentation in SD-OCT images, Translational Vision Science & Technology 7 (1) (2018) 1-1.
[20] H. Shang, Z. Jiang, R. Xu, et al., The Dynamic Mechanism of a Novel Stochastic Neural Firing Pattern Observed in a Real Biological System, Cognitive Systems Research, 2018, https://doi: 10.1016/j.cogsys. 2018.04.09.
[21] L. Yu, H. Chen, Q. Dou, et al., Integrating online and offline three-dimensional deep learning for automated polyp detection in colonoscopy videos, IEEE J. Biomed. Health Informat. 21 (1) (2017) 65–75.
[22] J. Loo, L. Fang, D. Cunefare, et al., Deep longitudinal transfer learning-based automatic segmentation of photoreceptor ellipsoid zone defects on optical coherence tomography images of macular telangiectasia type 2, Biomed. Opt. Express 9 (6) (2018) 2681–2698.
[23] A. Krizhevsky, I. Sutskever, G.E. Hinton, Imagenet classification with deep convolutional neural networks, Adv. Neural Inf. Process. Syst. (2012) 1097–1105.
[24] K. Simonyan, A. Zisserman, Very Deep Convolutional Networks for Large-scale Image Recognition, (2014) arXiv preprint arXiv:1409.1556.
[25] A. Cameron, D. Lui, A. Boroomand, et al., Stochastic speckle noise compensation in optical coherence tomography using non-stationary spline-based speckle noise modelling, Biomed. Opt. Express 4 (9) (2013) 1769–1785.
[26] T. Schlegl, S.M. Waldstein, W.D. Vogl, et al., Predicting semantic descriptions from medical images with convolutional neural networks, International Conference on Information Processing in Medical Imaging, Springer, Cham, 2015, pp. 437–448.
[27] Q. Dou, L. Yu, H. Chen, et al., 3D deeply supervised network for automated segmentation of volumetric medical images, Med. Image Anal. 41 (2017) 40–54.
[28] K. Yan, L. Lu, R.M. Summers, Unsupervised Body Part Regression Using Convolutional Neural Network with Self-organization, (2017) arXiv preprint arXiv:1707.03891.
[29] Z. Ge, S. Demyanov, R. Chakravorty, et al., Skin disease recognition using deep saliency features and multimodal learning of dermoscopy and clinical images, International Conference on Medical Image Computing and Computer-assisted Intervention, Springer, Cham, 2017, pp. 250–258.
[30] S. Liao, Y. Gao, J. Lian, et al., Sparse patch-based label propagation for accurate prostate localization in CT images, IEEE Trans. Med. Imag. 32 (2) (2013) 419–434.
[31] B. Leng, S. Guo, X. Zhang, et al., 3D object retrieval with stacked local convolutional autoencoder, Signal Process. 112 (C) (2015) 119–128.
[32] S. Bu, L. Wang, P. Han, et al., 3D shape recognition and retrieval based on multi-modality deep learning, Neurocomputing 259 (2017) 183–193.