

Beyond Retinal Layers: A Deep Voting Model for Automated Geographic Atrophy Segmentation in SD-OCT Images

Zexuan Ji¹, Qiang Chen¹, Sijie Niu², Theodore Leng³, and Daniel L. Rubin^{4,5}

¹ School of Computer Science and Engineering, Nanjing University of Science and Technology, Nanjing, China

² School of Information Science and Engineering, University of Jinan, Jinan, China

³ Byers Eye Institute at Stanford, Stanford University School of Medicine, Palo Alto, CA, USA

⁴ Department of Radiology, Stanford University, Stanford, CA, USA

⁵ Medicine (Biomedical Informatics Research), Stanford University, Stanford, CA, USA

Correspondence: Qiang Chen, Professor, School of Computer Science and Engineering, Nanjing University of Science and Technology, Nanjing, China. e-mail: chen2qiang@njjust.edu.cn

Received: 30 June 2017

Accepted: 1 November 2017

Published: 2 January 2018

Keywords: spectral-domain optical coherence tomography; geographic atrophy; image segmentation; deep network; voting

Citation: Ji Z, Chen Q, Niu S, Leng T, Rubin DL. Beyond retinal layers: a deep voting model for automated geographic atrophy segmentation in SD-OCT images. *Trans Vis Sci Tech.* 2018;7(1):1, <https://doi.org/10.1167/tvst.7.1.1>

Copyright 2018 The Authors

Purpose: To automatically and accurately segment geographic atrophy (GA) in spectral-domain optical coherence tomography (SD-OCT) images by constructing a voting system with deep neural networks without the use of retinal layer segmentation.

Methods: An automatic GA segmentation method for SD-OCT images based on the deep network was constructed. The structure of the deep network was composed of five layers, including one input layer, three hidden layers, and one output layer. During the training phase, the labeled A-scans with 1024 features were directly fed into the network as the input layer to obtain the deep representations. Then a soft-max classifier was trained to determine the label of each individual pixel. Finally, a voting decision strategy was used to refine the segmentation results among 10 trained models.

Results: Two image data sets with GA were used to evaluate the model. For the first dataset, our algorithm obtained a mean overlap ratio (OR) $86.94\% \pm 8.75\%$, absolute area difference (AAD) $11.49\% \pm 11.50\%$, and correlation coefficients (CC) 0.9857; for the second dataset, the mean OR, AAD, and CC of the proposed method were $81.66\% \pm 10.93\%$, $8.30\% \pm 9.09\%$, and 0.9952, respectively. The proposed algorithm was capable of improving over 5% and 10% segmentation accuracy, respectively, when compared with several state-of-the-art algorithms on two data sets.

Conclusions: Without retinal layer segmentation, the proposed algorithm could produce higher segmentation accuracy and was more stable when compared with state-of-the-art methods that relied on retinal layer segmentation results. Our model may provide reliable GA segmentations from SD-OCT images and be useful in the clinical diagnosis of advanced nonexudative AMD.

Translational Relevance: Based on the deep neural networks, this study presents an accurate GA segmentation method for SD-OCT images without using any retinal layer segmentation results, and may contribute to improved understanding of advanced nonexudative AMD.

Introduction

As a chronic disease, age-related macular degeneration (AMD) is the leading cause of irreversible vision loss among elderly individuals, which is generally accompanied with various phenotypic manifestations.¹ The advanced stage of nonexudative AMD is generally characterized by geographic

atrophy (GA) that is mainly characterized by atrophy of the retinal pigment epithelium (RPE).² In the comparison of AMD treatments trial, the development of GA was one of the major causes for sustained visual acuity loss,³ which is generally associated with retinal thinning and loss of RPE and photoreceptors.⁴ A recent review article notes that the reduction in the worsening of atrophy is an important biomarker for

assessing the effectiveness of a given GA treatment.⁵ Thus, automatic detection and characterization of retinal regions affected by GA is a fundamental and important step for clinical diagnosis, which could aid ophthalmologists in objectively measuring the regions of GA and monitor the evolution of AMD to further make treatment decisions.^{6,7} GA characterization generally requires accurate segmentation. Manual segmentation is time consuming and subject to interrater variability, which may not produce reliable results especially for large data sets. Therefore, automatic, accurate, and reliable segmentation technologies are urgently needed to advanced care in AMD.

To the best of our knowledge, most semiautomated or automated image analysis methods to identify GA are applied to color fundus photographs, fundus autofluorescence (FAF), or optical coherence tomography (OCT) modalities.⁸ Semiautomatic and automatic segmentation GA segmentation methods applied to these modalities can generally produce useful results and have been found to agree with manually drawn gold standards.

Color fundus photographs have been widely used for measuring GA lesions, where GA is characterized by a strongly demarcated area.⁹ However, the performance of most methods mainly depends on the quality of the color fundus images. GA lesions can be easily identified in high-quality color images, while the boundaries may be more difficult to be identified in lower quality images.

As a noninvasive imaging technique for the ocular fundus, FAF can provide two-dimensional (2D) images with high contrast for the identification of GA. Both semiautomated and automated methods have been proposed for the segmentation of GA in FAF images. C. Panthier et al.¹⁰ proposed a semi-automated image processing approach for the identification and quantification of GA on FAF images, and constructed a commercial package (i.e., Region Finder software), which was widely used for the evaluation of GA in clinical setting. The interactive approaches including level sets,¹¹ watershed,¹² and region growing¹³ have also been used in GA segmentation of FAF images. Meanwhile, the supervised classification methods¹⁴ and clustering technologies¹⁵ are widely used to automatically segment GA lesions in FAF images.

Compared with fundus imaging, spectral-domain (SD) OCT imaging technology can obtain the axial differentiation of retinal structures and additional characterization of GA.¹⁶ Unlike the planar images

provided by fundus modalities, SD-OCT can generate three-dimensional (3D) cubes composed of a set of 2D images (i.e., B-scans), and provide more detailed imaging characteristics of disease phenotypes.^{17,18} Because GA is generally associated with retinal thinning and loss of the RPE and photoreceptors, earlier works mainly focused on the thickness measurement of RPE, which could be further used as the biomarkers of GA lesions.¹⁹ However, segmenting GA is not as straightforward as solely detecting RPE. To directly identify GA lesions by characterizing RPE, state-of-the-art algorithms principally segment the GA regions based on the projection image generated with the voxels between the RPE and the choroid layers.^{20–23} Chen et al.²⁰ used geometric active contours to produce a satisfactory performance when compared with manually defined GA regions. A level set approach was developed to segment GA regions in both SD-OCT and FAF images.²¹ However, the performance of these models were generally dependent on the initializations. To further improve the segmentation accuracy and robustness to initializations, Niu et al.²² proposed an automated GA segmentation method for SD-OCT images by using a Chan-Vese model via local similarity factor, and then used this segmentation algorithm to automatically predict the growth of GA.²³ However, as mentioned above, GA is generally associated with retinal thinning and loss of RPE and photoreceptors, and state-of-the-art algorithms mainly segment GA based on the projection image generated with the voxels between the RPE and the choroid layers, implying that these methods rely on the accuracy of retinal layer segmentation.

Recently, deep learning has gained significant success and obtained outstanding performance in many computer vision applications.²⁴ Much attention has been drawn to the field of computational medical imaging to investigate the potential of deep learning in medical imaging applications,²⁵ including medical image segmentation,²⁶ registration,²⁷ multimodal fusion,²⁸ diagnosis,²⁹ disease detection,³⁰ and so on. For ophthalmology applications, deep learning has also recently been applied to automated detection of diabetic retinopathy from fundus photos,³¹ visual field perimetry in glaucoma patients,³² grading of nuclear cataracts,³³ segmentation of foveal microvasculature,³⁴ AMD classification,³⁵ and identification of diabetic retinopathy.³⁶ Here, we use deep learning methods to automatically discover the representations and structures inside OCT data in order to segment GA. To our best knowledge, we are the first to

segment the GA lesions from OCT images with deep learning.

A deep voting model is proposed for automated GA segmentation of SD-OCT images, which is capable of achieving high segmentation accuracy without using any retinal layer segmentation results. A deep network is constructed to capture deep representations of the data, which contains five layers including one input layer, three hidden layers (sparse autoencoders; SA), and one output layer. During the training phase, the randomly selected labeled A-scans with 1024 features are directly fed into the network as the input layer to obtain the deep representations. Then a soft-max classifier is trained to determine the label of each individual pixel. Finally, a voting decision strategy is used to refine the segmentation results among ten trained models. Without retinal layer segmentation, the proposed algorithm can obtain higher segmentation accuracy and is more stable compared with the state-of-the-art methods that rely on the retinal layer segmentation results. Our method can provide reliable GA segmentations from SD-OCT images and be useful for evaluating advanced nonexudative AMD.

Methods

Experimental Data Characteristics

Two different data sets acquired with a Cirrus OCT device (Carl Zeiss Meditec, Inc., Dublin, CA) were used to evaluate the performance of the proposed algorithm, where all the training and testing cases contained advanced nonexudative AMD with GA. It should be noted that both data sets were described and used in previous work.^{20,22} The first data set contained 51 longitudinal SD-OCT cube scans from 12 eyes of 8 patients with a size of $512 \times 128 \times 1024$ corresponding to a $6 \times 6 \times 2\text{-mm}^3$ volume in the horizontal, vertical, and axial directions, respectively. Two independent experts manually drew the outlines of GA based on the B-scan images in two repeated separate sessions, which were used to generate the segmentation ground truths. **Figure 1a** shows one example study case with the manual segmentations by two different experts and at two different sessions and the average ground truth, which are all outlined on the full projection image. The red and green contours shows the manually segmentations by the first experts, and the blue and cyan contours shows the manually segmentations by the second experts. The second data set contained 54 SD-

OCT cube scans from 54 eyes of 54 patients with a size of $200 \times 200 \times 1024$ corresponding to the same volume in the horizontal, vertical, and axial directions, respectively. The manual outlines were drawn based on FAF images, and then were manually registered to the corresponding location in the projection images and considered as ground truth segmentations. **Figure 1b** shows the registration ground truth outlined on the full projection image. All the data processing and methods implementation were carried out with Matlab 2016a software (The MathWorks, Inc., Natick, MA). The research was approved by an institutional human subjects committee and followed the tenets of the Declaration of Helsinki. All federal, state, and local laws were abided by, and this study was conducted with respect to all privacy regulations.

Processing Pipeline

As shown in **Figure 2**, an automatic GA segmentation method for SD-OCT images based on the deep network is proposed, which is capable of capturing the deep representations of the data while achieving high segmentation accuracy. The structure of the SA deep network was composed of five layers, including one input layer, three hidden/SA layers, and one output layer. During the training phase, the labeled A-scans with 1024 features were directly fed into the network as the input layer to obtain the deep representations. Then a soft-max classifier was trained to determine the label of each individual pixel on the projection image. Finally, a voting decision strategy was used to refine the segmentation results among 10 trained models.

Data Preprocessing

As an interferometric method based on coherent optical beams, one of the fundamental challenges with OCT imaging is the presence of speckle noise in the tomograms.³⁷ To reduce the influence of the noise in OCT images, we used the BM4D software (The Matlab code can be found in <http://www.cs.tut.fi/~foi/GCF-BM3D/>) for volumetric data denoising,³⁸ which is one leading denoising method for OCT. The 3D, 2D, and 1D visualization results can be found in **Figure 2**.

Deep Network Training

For each OCT image, each pixel in the projection image is a D dimensional vector $x \in \mathcal{R}^D$ along the axial A-scan lines. The labeled dataset is represented

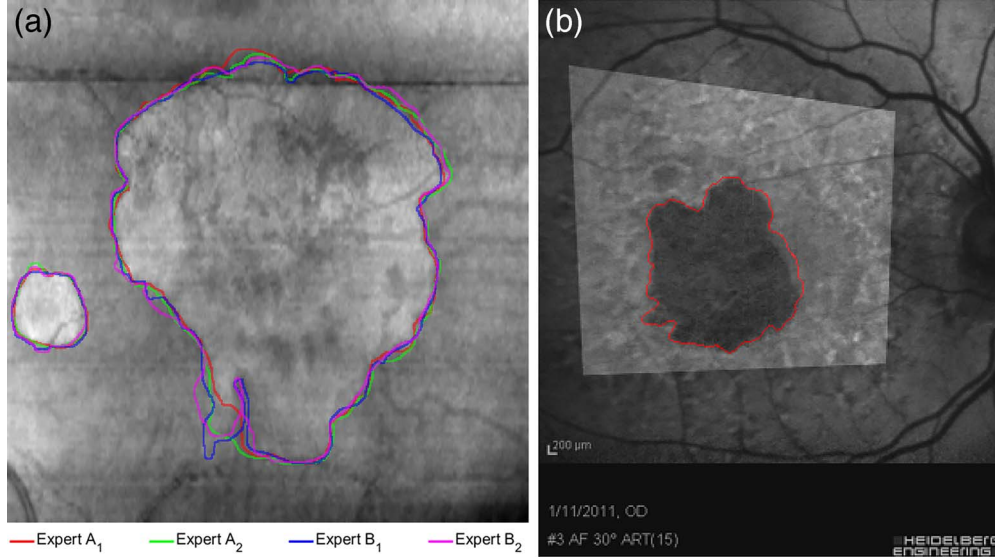


Figure 1. The example ground truths for two data sets. (a) One example study case with manual segmentations by two different experts during two different sessions, which are all outlined on the full projection image. (b) The registration ground truth outlined on the full projection image.

as $X = \{(x_i, y_i) | x_i \in \mathcal{R}^D, y_i \in L, i = 1, \dots, N\}$ where N is the number of samples in the dataset, which is the total number of A-scans in this paper, y_i is the class label of the corresponding vector x_i , and $L = \{l_i | i = 1, \dots, N, l_i = 1, \dots, K\}$ is the label set with size K . Generally, for an OCT image, the dimension of each vector x is $D = 1024$. Our target was to segment the GA tissues and non-GA tissues, so the label set was $K = 2$. Therefore, the target of training was to learn a mapping function $f(\cdot) : \mathcal{R}^D \rightarrow L$, which could map the input feature vector from the 3D space into the label space.

An autoencoder is a neural network, which attempts to replicate its input at its output. As mentioned above, we stacked three sparse autoencoders³⁹ as the hidden layers to construct our deep model. The training process was based on the optimization of a cost function, which measured the error between the input and its reconstruction at the output. An autoencoder is composed of an encoder and a decoder. For the input $x \in \mathcal{R}^D$ of one autoencoder, the encoder maps the vector x to another vector $z \in \mathcal{R}^{D^{(1)}}$ as $z^{(1)} = h^{(1)}(w^{(1)}x + b^{(1)})$, where the superscript (1) indicates the first layer. $h^{(1)} : \mathcal{R}^{D^{(1)}} \rightarrow \mathcal{R}^{D^{(1)}}$ is a transfer function for the encoder, $w^{(1)} \in \mathcal{R}^{D^{(1)} \times D}$ is a weight matrix, and $b^{(1)} \in \mathcal{R}^{D^{(1)}}$ is a bias vector. Then the decoder maps the encoded representation z back into an estimate of the original input vector x as $\hat{x} = h^{(2)}(w^{(2)}z^{(1)} + b^{(2)})$, where the superscript (2) represents the second layer. $h^{(2)} : \mathcal{R}^D \rightarrow \mathcal{R}^D$ is the

transfer function for the decoder, $w^{(2)} \in \mathcal{R}^{D \times D^{(1)}}$ is a weight matrix, and $b^{(2)} \in \mathcal{R}^D$ is a bias vector.

The cost function for training a sparse autoencoder is an adjusted mean squared error function as follows:

$$E = \frac{1}{N} \sum_{i=1}^N \sum_{k=1}^K (x_{ik} - \hat{x}_{ik})^2 + \lambda \times \Omega_{weights} + \beta \times \Omega_{sparsity} \quad (1)$$

$\Omega_{weights}$ is the L_2 regularization term with the coefficient λ , which can be defined as:

$$\Omega_{weights} = \frac{1}{2} \sum_{m=1}^2 \sum_{i=1}^N \sum_{k=1}^K (w_{ik}^{(m)})^2 \quad (2)$$

$\Omega_{sparsity}$ is the sparsity regularization term with the coefficient β , which can be defined as:

$$\Omega_{sparsity} = \sum_{i=1}^{D^{(1)}} KL(\rho || \hat{\rho}_i) \quad (3)$$

where $\hat{\rho}_i = \frac{1}{N} \sum_{j=1}^N h(w_i^{(1)T} x_j + b_j^{(1)})$

Sparsity regularization term attempts to enforce a constraint on the sparsity of the output from the hidden layer, which is constructed based on the Kullback-Leibler divergence.

For each hidden layer of the stacked autoencoders, the training target is to obtain the optimal parameter

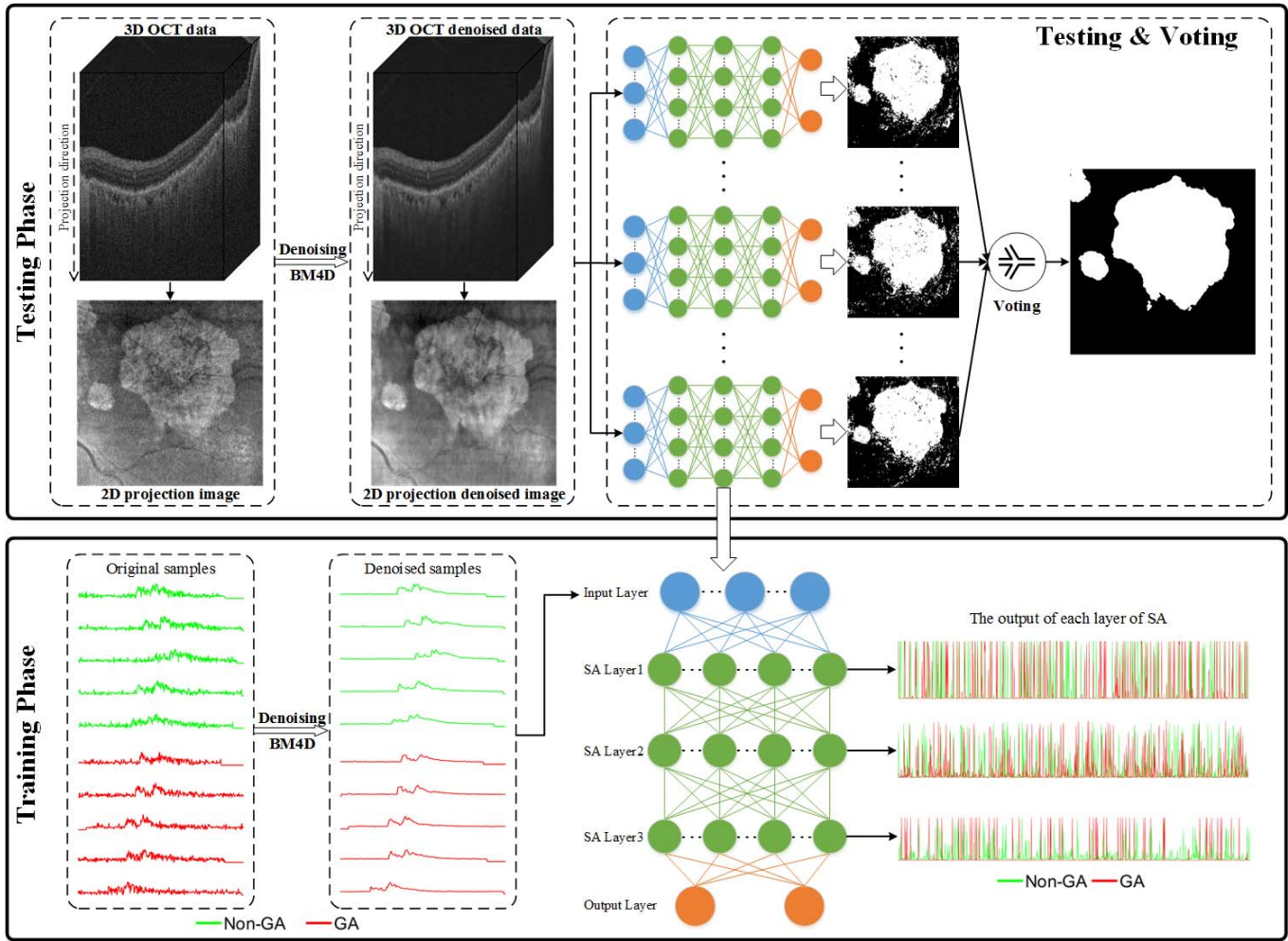


Figure 2. The pipeline of the proposed automatic GA segmentation method.

$\{W^*, b^*\}$ by minimizing the target cost function defined in Equation 1. The layers of stacked autoencoders are learned sequentially from top to bottom. As one of the most popular optimization methods, stochastic gradient descent is used for training the stacked autoencoders, and more details can be found in Ref. 40.

The learning of the stacked autoencoders is unsupervised learning. Lastly, behind the last autoencoder layer, we stacked another supervised classifier layer, which took the output of the last autoencoder layer as the input, and outputs classification results. By stacking this supervised layer, the deep network in this paper could be treated as a multilayer perceptron, where the parameters involved in autoencoders are learned by an unsupervised phase and further fine tuned by the backpropagation.⁴¹ Table 1 summarizes the parameter settings of the autoencoder structure and autoencoder training for all the experiments in

this paper. It should be noted that the coefficients λ and β for L_2 regularization term and sparsity regularization term are manually set based on the experimental results.

The representations learned by stacked autoencoders can decrease the redundant information of the input data, and preserve more useful information for the final classification. From the outputs of each layer of stacked autoencoders shown in Figure 2, a trend of sparsity can be clearly observed with the data propagation from the top layer to the bottom layer of the network.

Voting Strategy

As we mentioned before, during the training phase, the labeled A-scans with 1024 features were directly fed into the network as the input layer to obtain the deep representations, which meant that the spatial consistency among A-scans were not taken into

Table 1. The Parameter Settings in the Proposed Model

Parameter	Value
Number of hidden layers	3
Number of nodes in the input layer	1024
Number of nodes in the output layer	2
Number of nodes in SA-layer1	1024
Number of nodes in SA-layer2	1024
Number of nodes in SA-layer2	1024
Unsupervised training epochs	1000
Supervised training epochs	2000
L_2 weight regularization λ	0.001
Sparsity Regularization β	4

account. Moreover, due to the retinal structure and the characterizes OCT imaging, the corresponding OCT data (3D), B-scan images of the cross section (2D), and the A-scan samples (1D) contain various structural difference as shown in Figure 3. Figure 3a shows a full projection image of one study case with GA, where the ground truth is overlaid with the red line. Based on Figure 3a, three B-scan images of the cross section heightened with blue line are selected, and the corresponding images are shown in Figure 3b, where the GA lesions are overlaid with the blue regions. Then, for each B-scan image, two GA samples and two normal (non-GA) samples are selected highlighted with red and green lines, respec-

tively. The intensity profiles of the selected samples are shown in Figure 3c. From Figure 3a, we can find that the full projection image contains obvious intensity inhomogeneity. Moreover, the contrast between GA lesion and background is very low. Figure 3b shows various structural difference among the selected B-scan images of the cross section. The corresponding intensity profiles of the selected A-scans further demonstrate that the structure for GA and non-GA samples had high variability, which meant that it was very difficult for the corresponding deep learning model to capture the uniform or general structural information among these samples. Therefore, in our experiment, we found that it was very difficult to get an accurate classification result by only using one deep network.

To deal with the above observation, in this paper, we trained 10 deep network models, and a voting decision strategy was used to refine the segmentation results among 10 trained models. Specifically, we randomly selected ten thousand A-scans with GA as positive samples and 10,000 normal A-scans without GA as negative samples to train one model, and there was no intersection among the training data used in each model. Then we classified the 3D OCT data case with these 10 models and obtained ten classification results. Finally, the segmentation results were obtained with the voting decision strategy by setting the labels for each pixel as the voting probability greater than 70%. Finally, a 7×7 median filtering was

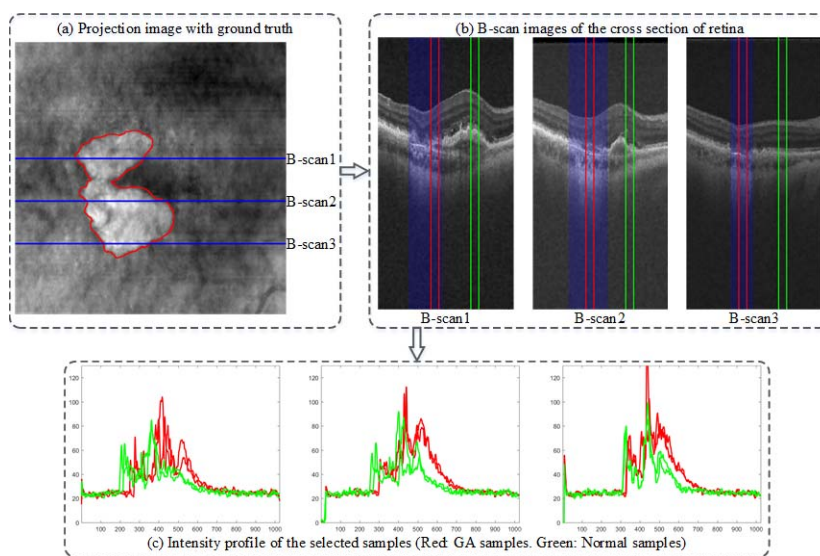


Figure 3. On example to show the various structural difference in OCT data. (a) A full projection image of one study case with GA, where the ground truth is overlaid with the red line. (b) Three B-scan images of the cross section selected from (a) heightened with blue line, where the GA lesions are overlaid with the blue regions. (c) The intensity profiles of the selected A-scans, where A-scans with GA and normal A-scans (A-scans without GA) are highlighted with red and green lines, respectively.

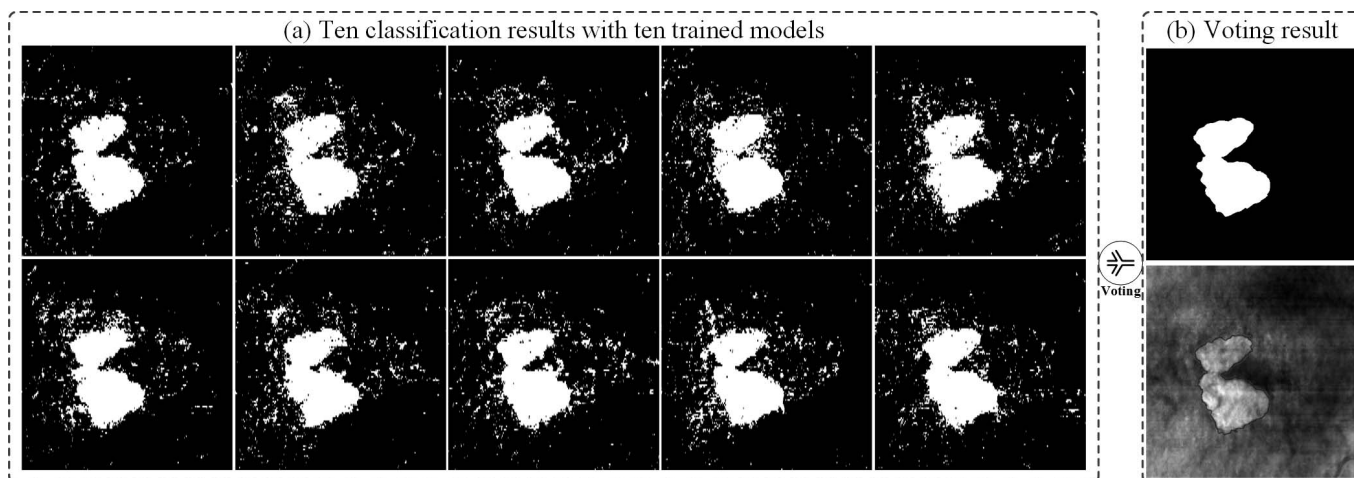


Figure 4. The voting decision strategy.

operated on the final voting results to ensure the smoothness of the final segmentations.

Figure 4 shows the voting decision strategy where the testing case is the same with that in Figure 3. From this figure, we can observe that each classification result obtained by these 10 models contains misclassifications due to the impact of the defects and various structural difference involved in OCT images. The voted classification result demonstrates that the proposed model can produce an accurate segmentation result, which is highly consistent with the ground truth.

Evaluation Criteria and Comparison Methods

In this paper, we used three criteria to quantitatively evaluate the performances of each comparison method: overlap ratio (OR), absolute area difference (AAD), and correlation coefficient (CC).

The overlap ratio 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):

$$OR(X, Y) = \frac{Area(X \cap Y)}{Area(X \cup Y)} \quad (4)$$

where X and Y indicate the regions inside the segmented GA contour produced by two different methods (or graders), respectively. The operators \cap and \cup indicate intersection and union, respectively.

The mean OR and standard deviation values are computed across scans in the data sets.

The absolute area difference measures the absolute difference between the GA areas as segmented by two different methods:

$$AAD(X, Y) = |Area(X) - Area(Y)| \quad (5)$$

Similar with OR, X , and Y indicate the regions inside the segmented GA contour produced by two different methods (or graders), respectively. The mean AAD and standard deviation values are computed across scans in the data sets.

The CC were computed using Pearson's linear correlation between the measured areas of GA computed by the segmentation of different methods or readers, measuring the linear dependence using each scan as an observation.

In the comparison experiments, we mainly compared with two related methods, called the Chen et al. method²⁰ and the Niu et al. method,²² respectively, the Chen et al. method is a semisupervised method based on the geometric active contours, while the Niu et al. method is an unsupervised method based on Chan-Vese model. It should be noted that both methods relied on the retinal layer segmentation results. They needed extract the RPE layers first, and then constructed the projection images based on the pixels below RPE layers. Finally, they performed their methods on 2D-projected images. Comparatively, our proposed algorithm directly processed 3D samples without using any retinal layer segmentation results.

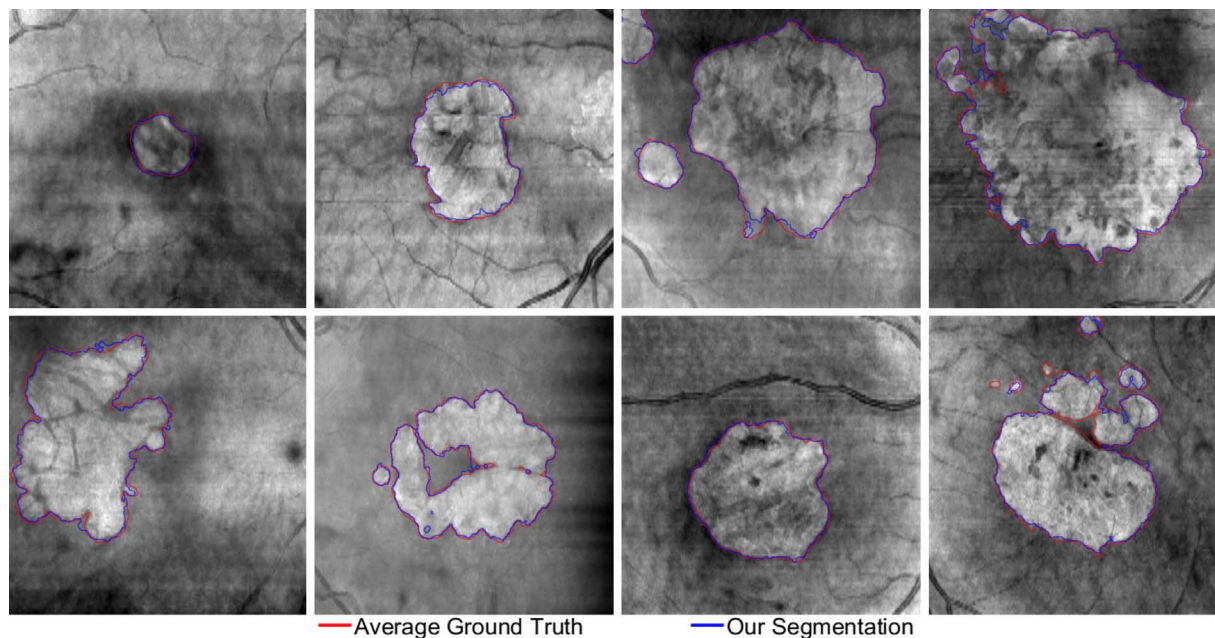


Figure 5. Segmentation results overlaid on full projection images for eight example cases selected from eight patients in Dataset 1, where the average ground truths are overlaid with a red line, and the segmentations obtained with the proposed model are overlaid with blue line.

Results

Testing I: Segmentation Results on the Dataset With a Size of $512 \times 128 \times 1024$

In the first experiment, we tested the proposed model on Dataset 1, which contained 51 longitudinal SD-OCT cube scans from 12 eyes of 8 patients with a size of $512 \times 128 \times 1024$. In the training phase, we randomly selected 10,000 A-scans with GA as positive samples and 10,000 normal A-scans without GA as negative samples to train one model, and there was no intersection among the training data used in each model. In the testing phase, we directly fed the testing 3D case into the proposed model to get the final segmentation result.

In Figure 5, eight example cases selected from eight patients were used to show the performance of the

Table 2. Intraobserver and Interobserver CC, AAD and OR Evaluations^{20,22}

	A_1 vs A_2	B_1 vs B_2	$A_{1\&2}$ vs $B_{1\&2}$
CC	0.998	0.996	0.995
AAD, mm^2	0.24 ± 0.21	0.24 ± 0.41	0.31 ± 0.47
AAD, %	3.70 ± 2.97	3.34 ± 5.37	4.68 ± 5.70
OR, %	93.29 ± 3.02	93.06 ± 5.79	91.28 ± 6.04

proposed model, where the red contours show the average ground truths and the blue contours are the segmentation results. In each figure, the ground truths and the segmentation results are overlaid on the full projection images, where the red line is the outline of average ground truth, and the blue line shows the outline of the segmentation results obtained by the proposed model. From this figure, we found that the full projection images contained obvious intensity inhomogeneity and low contrast between GA lesions and normal regions. Using the deep network and voting strategy, the proposed model can produce smooth and accurate segmentation results, which are highly consistent with the average ground truths.

The quantitative results in interobserver and intraobserver agreement evaluation for Dataset 1 are summarized in Table 2, where $A_i (i = 1, 2)$ represents the segmentations of the first grader in the i -th session, and $B_i (i = 1, 2)$ represents the segmentations of the second grader in the i -th session. Interobserver differences were computed by considering the union of both sessions for each grader: $A_{1\&2}$ and $B_{1\&2}$ represent the first and second grader, respectively. The intraobserver and interobserver comparison showed very high CC, indicating very high linear correlation and between different readers and for the same reader at different sessions. The overlap ratios (all $>90\%$) and the absolute GA area

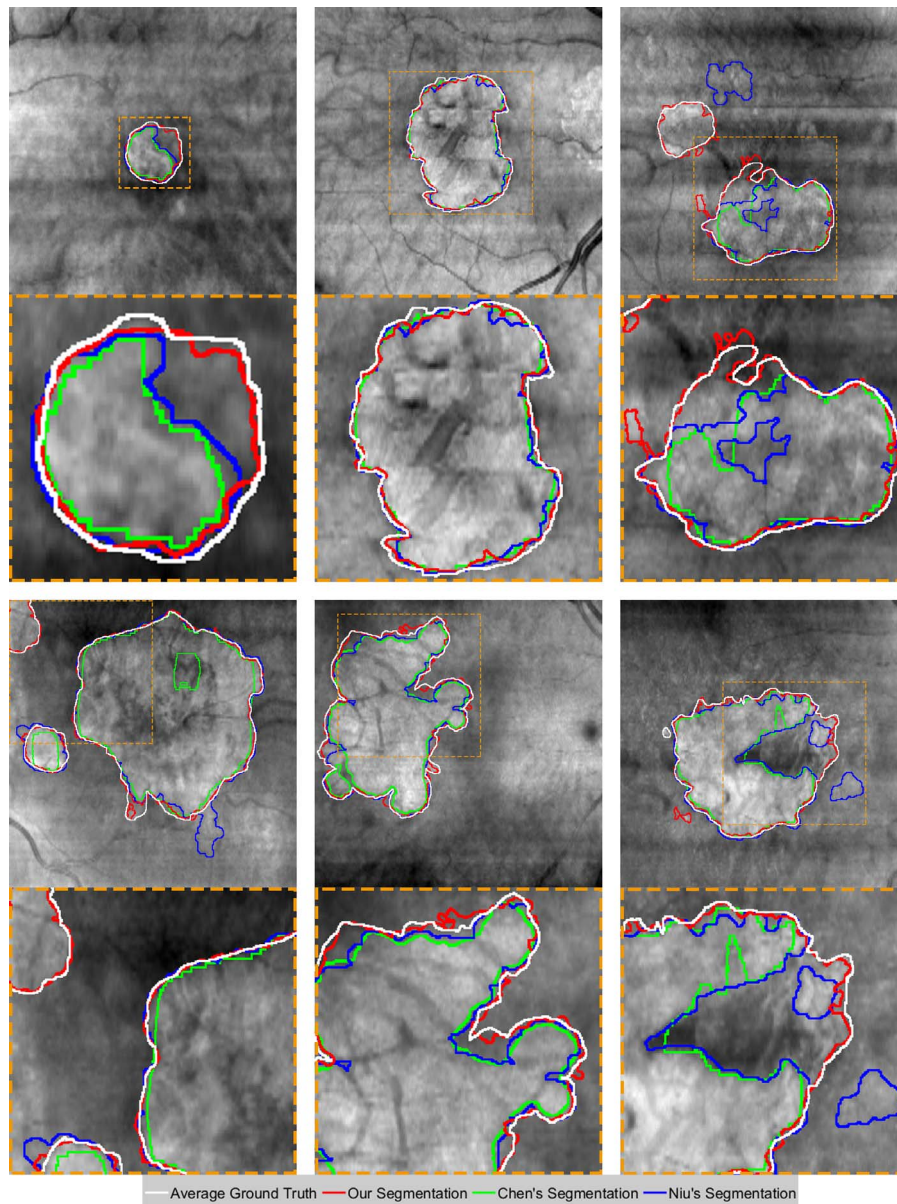


Figure 6. Comparison of segmentation results overlaid on full projection images for six example cases selected from six patients in Dataset 1, where the average ground truths are overlaid with a *white line*, and the segmentations obtained with the proposed model, Chen et al.'s²⁰ method and Niu et al.'s²² method are overlaid with *red, green, and blue lines*, respectively. In each subfigure, the *top image* shows the segmentation results overlaid on full projection images, and the *bottom image* shows the enlarged view of the rectangles marked by an *orange box*.

differences (all $<5\%$) indicate very high interobserver and intraobserver agreement, highlighting that the measurement and quantification of GA regions in the generated projection images seem effective and feasible.^{20,22}

Then we qualitatively compared the outlines of the segmentations obtained by the proposed model and two comparison methods on six examples in [Figure 6](#). In each figure, the white line shows the average

ground truth. The green, blue, and red lines show the Chen et al.,²⁰ Niu et al.,²² and our segmentations, respectively. For the second and the fifth cases, all the comparison methods could produce satisfactory results because the structure of GA is obvious and the corresponding contrast is higher. For the first and the sixth cases, due to the impact of the low contrast, both the Chen et al.²⁰ and Niu et al.²² methods failed to detect parts of the boundaries between GA lesions

and non-GA regions. The Niu et al.²² method misclassified normal regions as GA lesions for the third and sixth cases, while the Chen et al.²⁰ method misclassified GA lesions as normal regions for the fourth case. Moreover, in the fourth case, the upper left corner of the GA lesions were all missed by the Chen et al.²⁰ and Niu et al.²² methods. Comparatively, the proposed model qualitatively outperformed the other two methods even without using any retinal layer segmentation results, and obtained higher consistency with the average ground truths.

Figure 7 shows the quantitative comparison results between our segmentation results and manual gold standards (average expert segmentations) on all the cases in Dataset 1, where the top figure shows the OR comparison, the middle figure shows the AAD comparison measured with volume, and the bottom figure shows the AAD comparison results measured with percentage. In each subfigure, the green rhombus, blue squares, and red circles, respectively, indicate the segmentation accuracy of the Chen et al.,²⁰ Niu et al.,²² and our methods. From this figure, we can quantitatively observe that the proposed model can produce more accurate segmentation results in most cases. Table 3 summarizes the average quantitative results between the segmentation results and manual gold standards (individual reader segmentations and the average expert segmentations) on Dataset 1. Overall, our model is capable of producing a higher segmentation accuracy to the manual gold standard than both the Chen et al.²⁰ and Niu et al.²² methods by presenting higher CC (0.986 vs. 0.970 and 0.979), lower AAD (11.49% vs. 27.17% and 12.95%), and higher OR (86.94% vs. 72.6% and 81.86%). A higher CC indicates that our model resulted in more similar results to the ground truth. Lower AAD indicates the areas estimated by the proposed model are closer to those manual productions. Higher OR indicates the proposed model can obtain more similar results to the manual outlines. Moreover, the proposed model is more robust to all the cases in Dataset 1 due to the lower standard deviations. In conclusion, the proposed algorithm showed better segmentation performances than the other two comparison methods on Dataset 1.

Testing II: Segmentation Results on the Dataset With a Size of $200 \times 200 \times 1024$

In the second experiment, we tested the proposed model on Dataset 2, which contains 54 SD-OCT cube scans from 54 eyes of 54 patients with a size of $200 \times$

200×1024 . Similar to the first experiment, in the training phase, we randomly selected 10,000 A-scans with GA as positive samples and 10,000 normal A-scans without GA as negative samples to train one model, and there was no intersection among the training data used in each model. In the testing phase, we directly fed the testing 3D case into the proposed model to obtain the final segmentation results.

In Figure 8, eight example cases selected from eight patients were used to show the performance of the proposed model, where the red and blue contours show the average ground truths and segmentation results, respectively. In each figure, the ground truth and the segmentation results are overlaid on the full projection images with the red and blue lines, respectively. We obtained similar results in that the proposed model could produce accurate results highly consistent with the average ground truths.

We qualitatively compared the segmentations obtained by the proposed model and two comparison methods on six examples in Figure 9. In each figure, the average ground truths were overlaid with the white lines, and the segmentations obtained with the Chen et al.²⁰, Niu et al.,²² and the proposed methods were overlaid with the green, blue, and red lines, respectively. In the first and fifth cases, all the comparison methods could produce satisfactory results due to the higher contrast of GA lesions. In the second and the sixth examples, the Chen et al.²⁰ and Niu et al.²² methods obtained grossly misclassified regions. The Niu et al.²² method failed to segment the third case, while the Chen et al.²⁰ method failed to segment the fourth case. Moreover, in the last example, the region inside the GA lesions were all misclassified by both the Chen et al.²⁰ and Niu et al.²² methods. Comparatively, without using any retinal layer segmentation results, our proposed model qualitatively outperformed the other two methods, and obtained results more consistent with the average ground truths.

Figure 10 shows the quantitative comparison results between the segmentation results and average expert segmentations on all the cases in Dataset 2, where the figures from top to bottom show OR comparison, AAD comparison measured with volume, and AAD comparison measured with percentage. In each subfigure, the green rhombus, blue squares, and red circles, respectively, indicate the segmentation accuracy of the Chen et al.,²⁰ Niu et al.,²² and our methods. Table 4 summarizes the average quantitative results between the segmentation results and manual gold standards on Dataset 2. Overall, our model was capable of producing higher

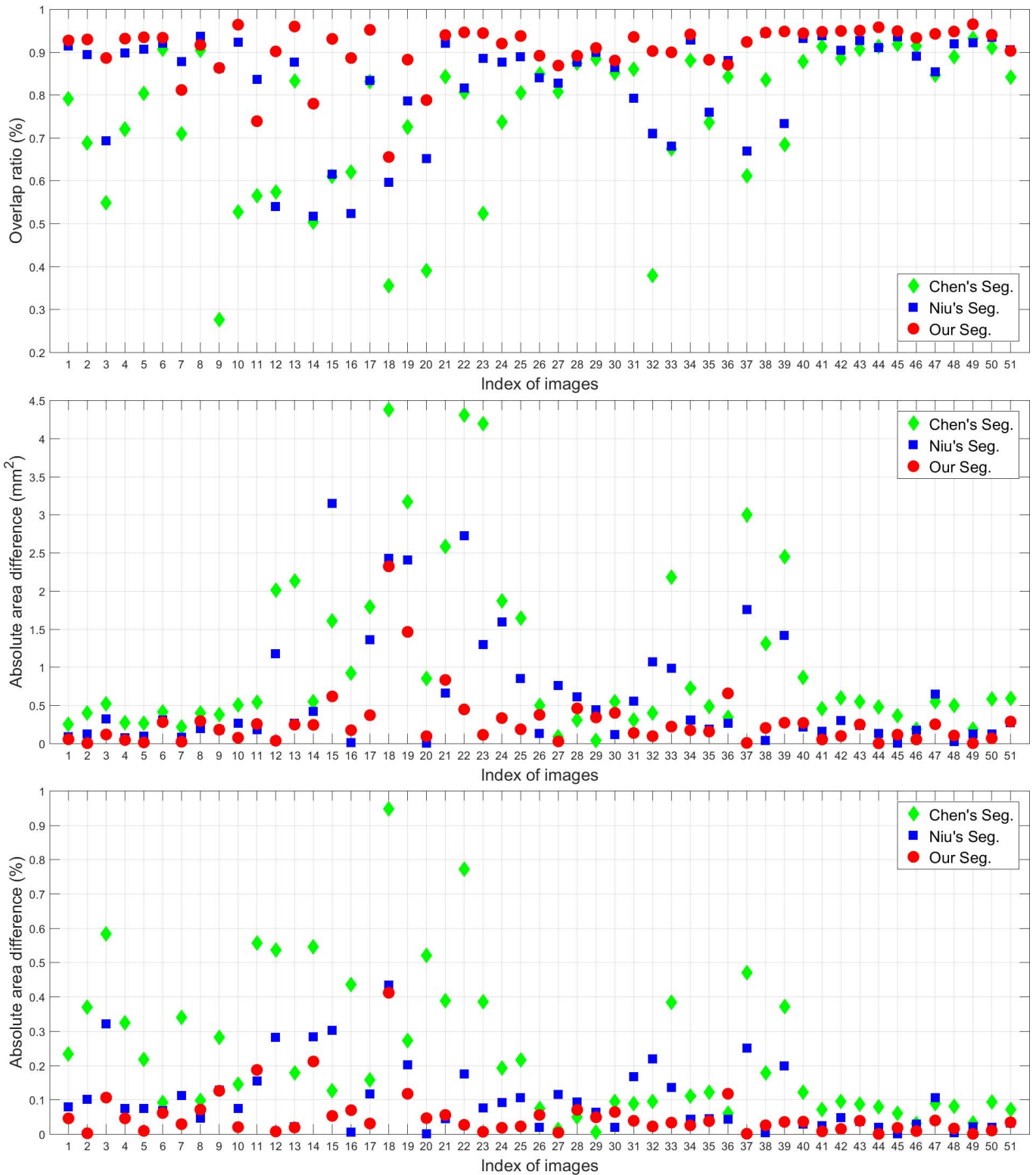


Figure 7. The quantitative comparisons between the segmentations and average expert segmentations on all the cases in Dataset 1.

Table 3. The Summarizations of the Quantitative Results (Mean \pm Standard Deviation) Between the Segmentations and Manual Gold Standards (Individual Reader Segmentations and the Average Expert Segmentations) on Dataset 1

Methods	Criteria	Average Expert	Expert A ₁	Expert A ₂	Expert B ₁	Expert B ₂
Chen et al. ²⁰ method	CC	0.970	0.967	0.964	0.968	0.977
	AAD, mm ²	1.44 \pm 1.26	1.31 \pm 1.28	1.40 \pm 1.31	1.60 \pm 1.33	1.47 \pm 1.14
	AAD, %	27.17 \pm 22.06	25.23 \pm 22.71	26.14 \pm 21.48	29.21 \pm 22.17	27.62 \pm 20.57
	OR, %	72.60 \pm 12.01	73.26 \pm 15.61	73.12 \pm 15.15	71.16 \pm 15.42	72.09 \pm 14.82
Niu et al. ²² method	CC	0.979	0.975	0.976	0.976	0.975
	AAD, mm ²	0.81 \pm 0.94	0.76 \pm 0.99	0.85 \pm 1.04	0.98 \pm 1.08	0.90 \pm 1.05
	AAD, %	12.95 \pm 11.83	12.62 \pm 12.86	13.32 \pm 12.74	14.91 \pm 12.65	14.07 \pm 11.78
	OR, %	81.86 \pm 12.01	81.42 \pm 12.12	81.61 \pm 12.29	80.05 \pm 13.05	80.65 \pm 12.51
The proposed model	CC	0.986	0.986	0.985	0.985	0.991
	AAD, mm ²	0.67 \pm 0.73	0.55 \pm 0.74	0.62 \pm 0.80	0.82 \pm 0.83	0.69 \pm 0.66
	AAD, %	11.49 \pm 11.50	9.75 \pm 11.35	10.32 \pm 11.09	13.58 \pm 12.41	11.73 \pm 9.35
	OR, %	86.94 \pm 8.75	87.64 \pm 8.75	87.71 \pm 8.32	85.17 \pm 9.40	86.37 \pm 7.67

Boldface values indicate the highest results.

segmentation accuracies to the manual gold standards than both the Chen et al.²⁰ and Niu et al.²² methods by presenting higher CC (0.995 vs. 0.937 and 0.955), lower AAD (8.30% vs. 19.68% and 22.96%), and higher OR (81.66% vs. 65.88% and 70.00%). Moreover, the proposed model was more robust to all the cases in Dataset 2 due to the lower standard deviations. In conclusion, the proposed algorithm

showed better segmentation performance than the other two comparison methods on Dataset 2.

Testing III: Segmentation Results With Patient-Independent Testing

In Testing I and Testing II, A-scans used for training came from the same patients that were later

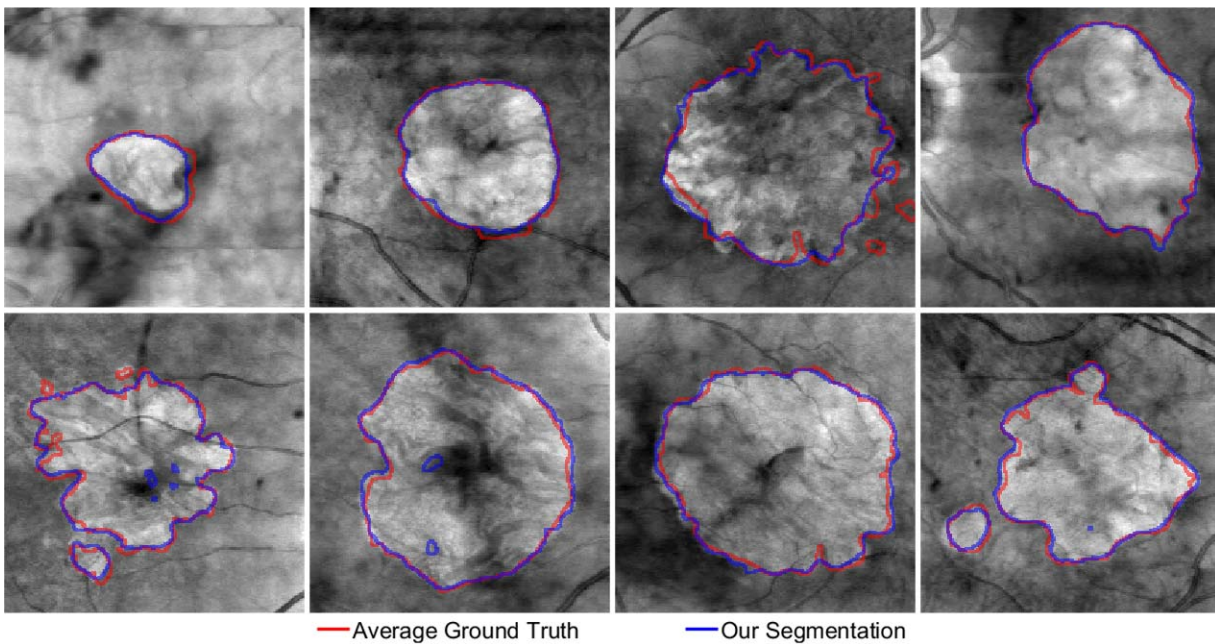


Figure 8. Segmentation results overlaid on full projection images for eight example cases selected from eight patients in Dataset 2, where the average ground truths are overlaid with red line, and the segmentations obtained with the proposed model are overlaid with blue line.

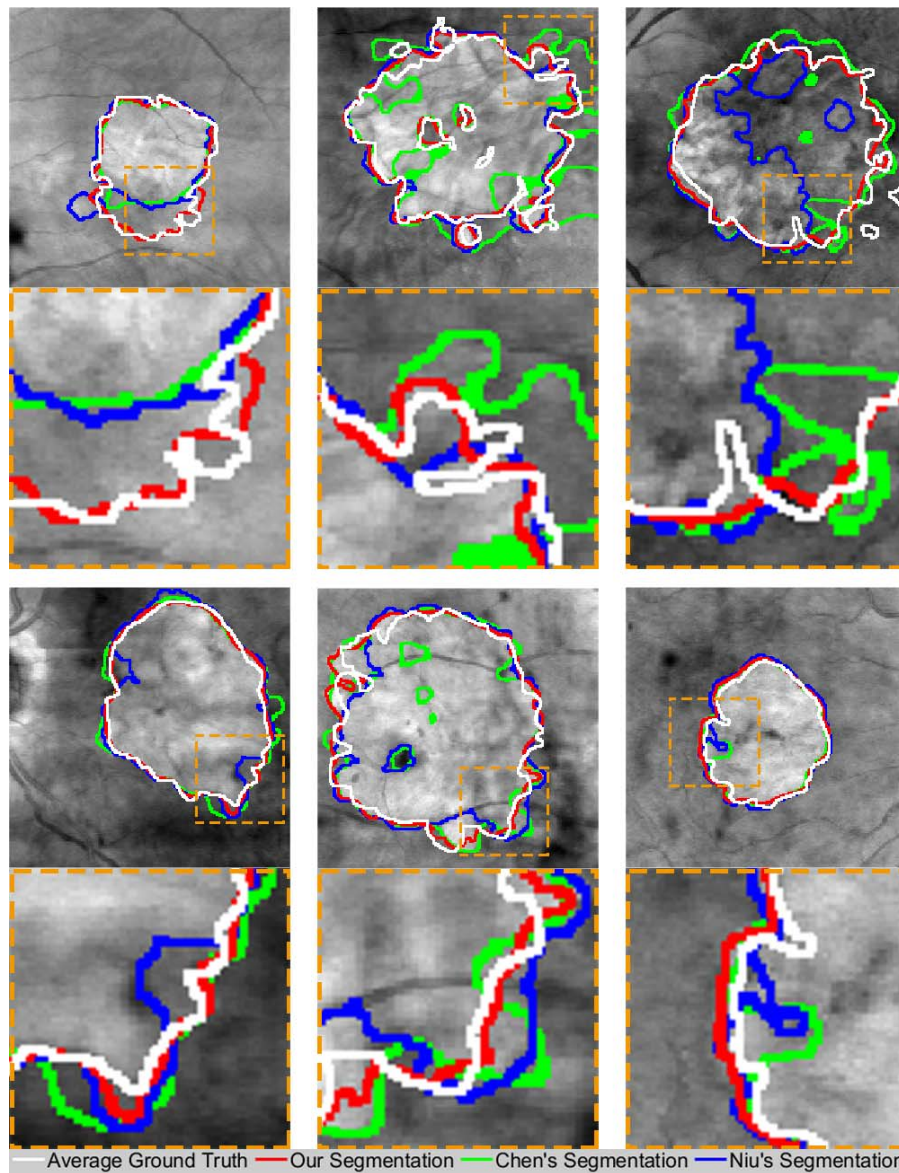


Figure 9. Comparison of segmentation results overlaid on full projection images for six example cases selected from six patients in Dataset 2, where the average ground truths are overlaid with a white line, and the segmentations obtained with the proposed model, the Chen et al.²⁰ and Niu et al.²² methods are overlaid with red, green, and blue lines, respectively. In each subfigure, the *top image* shows the segmentation results overlaid on full projection images, and the *bottom image* shows the enlarged view of the rectangles region marked by an orange box.

tested on, which means that these two experiments were not independent on the patient level. To further verify the performance of the propose model on patient-independent testing, in this experiment, we respectively divided Dataset 1 and Dataset 2 into two disjoint parts on the patient level. Specifically, for Dataset 1, 51 cases from eight patients were divided into two parts: the first part contains 25 images from four patients and the second part contains the other 26 images from the other four patients. For Dataset 2,

54 eyes from 54 patients were also divided into two disjoint parts, and each part contains 27 images from 27 patients without any overlap. In the training phase, we randomly selected 10,000 A-scans with GA as positive samples and 10,000 normal A-scans without GA as negative samples from one part to train the models. In the testing phase, we directly fed the testing 3D cases in the other part into the proposed model to get the final segmentation result. Therefore, the training and testing sets are totally independent

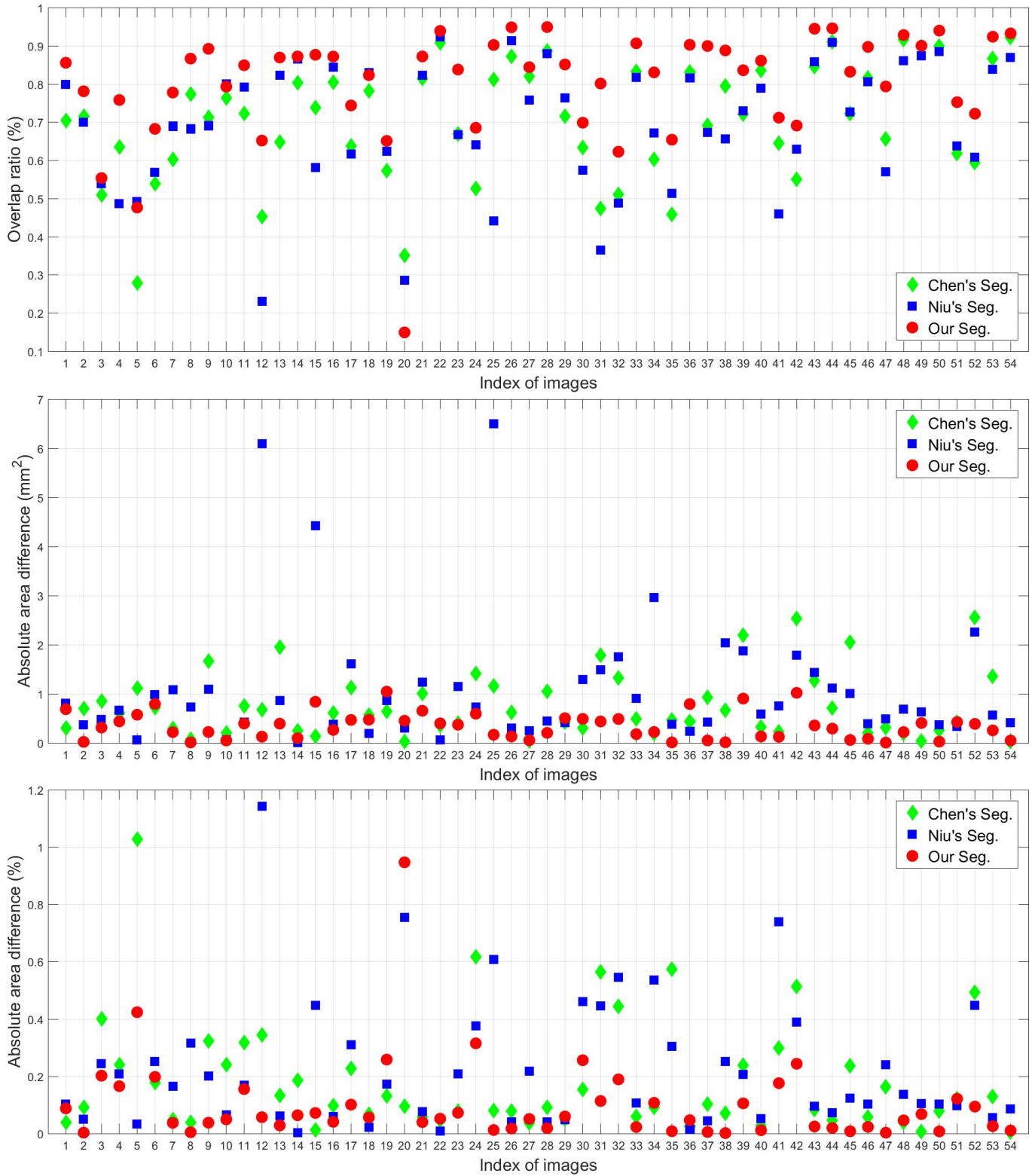


Figure 10. The quantitative comparisons between the segmentations and average expert segmentations on all the cases in Dataset 2.

Table 4. The Summarizations of the Quantitative Results (Mean \pm Standard Deviation) Between the Segmentations and Manual Gold Standards on Dataset 2

Methods	Chen et al. ²⁰ Segmentation	Niu et al. ²² Segmentation	Our Segmentation
CC	0.955	0.937	0.995
AAD, mm ²	0.95 \pm 1.28	1.21 \pm 1.58	0.34 \pm 0.27
AAD, %	19.68 \pm 22.75	22.96 \pm 21.74	8.30 \pm 9.09
OR,	65.88 \pm 18.38	70.00 \pm 15.63	81.66 \pm 10.93

Boldface values indicate the highest results.

with each other on the patient level. **Table 5** summarizes the average quantitative results between the segmentation results obtained with patient-independent testing procedure and manual gold standards on both data sets. For the first dataset, our algorithm on patient-independent testing procedure obtains a total mean OR 83.45% \pm 9.56%, AAD 14.49% \pm 14.30%, and CC 0.975. For the second dataset, the mean OR, AAD, and CC of the proposed method patient-independent testing procedure are 78.00% \pm 12.86%, 11.86% \pm 12.09%, and 0.992, respectively. The performances of patient-independent testing procedure declined approximately 4% compared with the quantitative results of the proposed model under patient-dependent procedure (Testing I and Testing II). However, performances under the patient-independent procedure still outperform the Chen et al.²⁰ and Niu et al.²² methods.

Testing IV: Segmentation Results With Cross Testing

In the last experiment, we executed a cross-testing procedure by using the trained models on one dataset to test the cases in the other dataset. Specificity, we tested all the cases in Dataset 1 with the models trained by Dataset 2, and we tested all the cases in Dataset 2 with the models trained by Dataset 1. It should be noted that we did not retrain the models, instead we directly used the models trained in the first and second experiments.

Figure 11 shows the cross testing results comparing with the original segmentations of the proposed

model and the ground truths. The figures in the first line shows the cross segmentations on four cases selected from Dataset 1, while the figures in the first line shows the cross segmentations on four cases selected from Dataset 2. In each figure, the average ground truths, the segmentations obtained with the proposed model, and the cross segmentation results were overlaid with red, green, and blue lines, respectively. From **Figure 11**, we can observe that the proposed model can still produce satisfactory results with cross testing procedure. However, compared with the original segmentations of the proposed model, the results obtained with cross testing contain misclassifications, especially for the regions near boundaries.

Table 6 summarizes the average quantitative results between the segmentation results obtained with cross-testing procedure and manual gold standards on both data sets. The performances of cross-testing procedure declined sharply (\sim 10%) compared with the original results of the proposed model. However, the cross-testing procedure still outperform the Chen et al.²⁰ method, and can produce similar accuracy comparing with the Niu et al.²² method.

Discussion

In this paper, based on the deep neural networks, we proposed an automatic and accurate GA segmentation method for SD-OCT images without using any retinal layer segmentation results. This is the first

Table 5. The Summarizations of the Quantitative Results (Mean \pm Standard Deviation) Between the Segmentations and Manual Gold Standards on Two Data Sets for Patient-Independent Testing

Strategy Criteria	Training on Part 1 and Testing on Part 2			Training on Part 2 and Testing on Part 1		
	AAD mm ²	AAD, %	OR, %	AAD, mm ²	AAD, %	OR, %
Dataset 1	1.07 \pm 1.33	19.54 \pm 17.46	80.27 \pm 11.78	0.59 \pm 0.43	9.64 \pm 8.15	86.50 \pm 5.45
Dataset 2	0.52 \pm 0.31	15.10 \pm 13.74	75.38 \pm 13.27	0.38 \pm 0.32	8.49 \pm 9.19	80.72 \pm 12.09

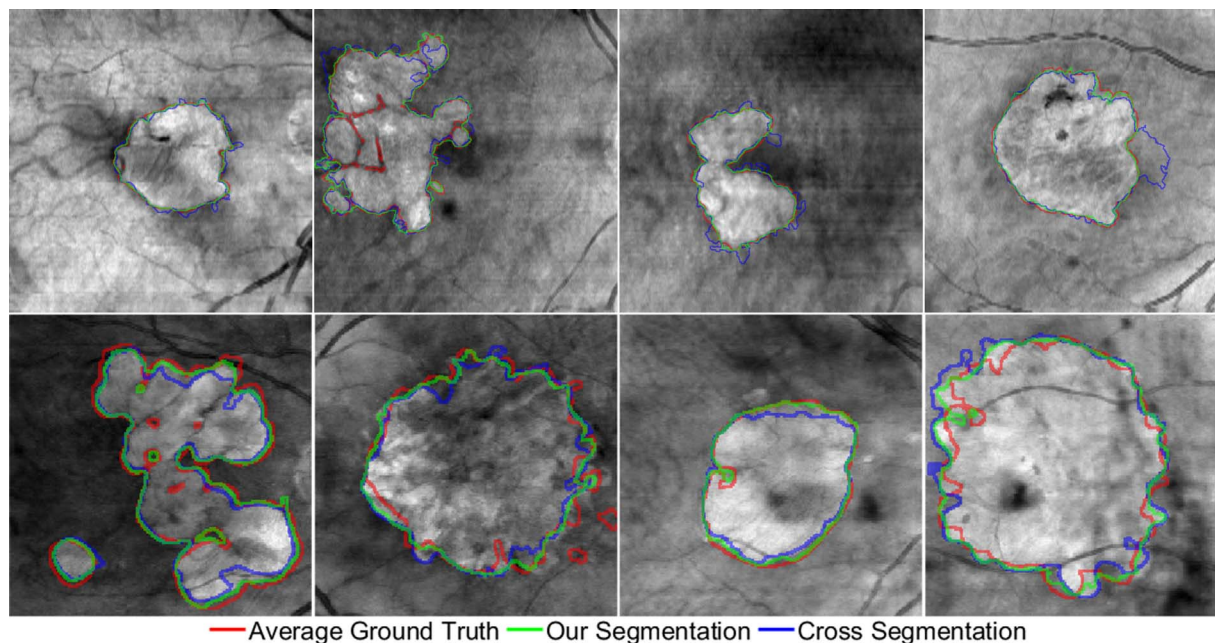


Figure 11. Segmentation results overlaid on full projection images for four cases selected Dataset 1 (*top row*) and four cases selected Dataset 2 (*bottom row*), where the average ground truths, the segmentations obtained with the proposed model, and the cross segmentation results are overlaid with *red, green, and blue lines*, respectively.

method that segments GA lesions from OCT images using deep learning technologies.

As listed in [Table 3](#), our model produced higher segmentation accuracies to the manual gold standards generated by two different readers and repeated at two separated sessions than both the Chen et al.²⁰ and Niu et al.²² methods. Our model had a higher CC (0.986), lower AAD (11.49%), and higher OR (86.94%). The proposed method also improved segmentation accuracy over 5% when compared with related algorithms on Dataset 1.

As summarized in [Table 4](#), the proposed model also obtained higher segmentation accuracies when compared with the registered ground truths manually drawn in FAF images (CC: 0.995, AAD: 8.30%, OR:

81.66%), and improve segmentation accuracy by over 10% when comparing with related algorithms on Dataset 2. The example segmentations shown in [Figures 5 and 8](#) corroborate the highly consistent with the average ground truths, and the comparison results shown in [Figures 6, 7, 9, and 10](#) further demonstrate the superior performances comparing with the related methods.

Compared with the results summarized in [Table 5](#) and the results of the proposed model listed in [Tables 3 and 4](#), the segmentation accuracy decreased approximately 10% on both data sets. The main reasons were: (1) the ground truths were inherently different. As shown in [Figure 1](#), the ground truths of two data sets were obtained through different

Table 6. The Summarizations of the Quantitative Results (Mean \pm Standard Deviation) Between the Segmentations and Manual Gold Standards on Two Data Sets

Dataset	Dataset 1					Dataset 2
	Average Expert	Expert A ₁	Expert A ₂	Expert B ₁	Expert B ₂	FAF Segmentation
CC	0.940	0.937	0.942	0.946	0.939	0.962
AAD, mm ²	1.03 \pm 1.14	1.02 \pm 1.21	1.03 \pm 1.12	1.01 \pm 1.18	1.07 \pm 1.15	0.92 \pm 0.76
AAD, %	14.36 \pm 9.30	12.96 \pm 9.15	14.05 \pm 9.36	13.65 \pm 9.18	14.81 \pm 9.39	16.49 \pm 13.85
OR, %	78.51 \pm 5.98	79.23 \pm 5.87	79.16 \pm 5.98	78.69 \pm 5.58	78.40 \pm 5.99	72.77 \pm 12.71

Table 7. The Summarizations of the Quantitative Results (Mean \pm Standard Deviation) Between the Segmentations and Manual Gold Standards (Individual Reader Segmentations and the Average Expert Segmentations) on Dataset 1 by Applying One Deep Model and Shallow Voting Model

Methods	Criteria	Average Expert	Expert A ₁	Expert A ₂	Expert B ₁	Expert B ₂
One deep model	CC	0.900	0.903	0.900	0.914	0.905
	AAD, mm ²	1.41 \pm 1.82	1.43 \pm 1.85	1.39 \pm 1.85	1.33 \pm 1.63	1.37 \pm 1.78
	AAD, %	20.49 \pm 24.69	20.95 \pm 24.68	20.81 \pm 25.28	19.37 \pm 23.52	19.86 \pm 24.05
	OR, %	72.86 \pm 15.94	72.72 \pm 16.31	73.14 \pm 16.26	72.43 \pm 15.58	72.67 \pm 15.51
Shallow voting model	CC	0.963	0.965	0.963	0.968	0.970
	AAD, mm ²	0.84 \pm 1.00	0.79 \pm 0.94	0.80 \pm 1.00	0.89 \pm 0.96	0.83 \pm 0.90
	AAD, %	15.35 \pm 16.89	14.29 \pm 16.06	14.26 \pm 17.19	15.91 \pm 17.03	15.07 \pm 15.29
	OR, %	79.57 \pm 12.61	79.95 \pm 12.85	80.18 \pm 12.67	78.34 \pm 12.72	79.16 \pm 11.93

procedures. The ground truths were drawn based on the OCT data itself for Dataset 1, while the ground truths were registered based on drawn outlines in FAF images for Dataset 2. Therefore, the performance of cross testing was better in Dataset 1 than those in Dataset 2. (2) The structure of the data varied. As shown in Figure 3, we found that even within one dataset, the intensity profiles of A-scans varied greatly, which meant that it was very difficult for the corresponding deep learning model to capture general structural information or general features among these A-scans. When we executed the cross testing, this difficulty was further magnified and the performance of the cross testing procedure declined sharply ($\sim 10\%$) compared with the original results of the proposed model. Ultimately though, the cross-testing procedure still outperformed the Chen et al.²⁰ method, and produced similar accuracy when compared with the Niu et al.²² method.

Consequently, without retinal layer segmentation, the proposed algorithm was able to obtain a higher segmentation accuracy when compared with the state-of-the-art methods relying on the retinal layer segmentations. Our methods may provide reliable GA segmentations for SD-OCT images and be useful for clinical diagnosis.

In this paper, a deep voting model is proposed to segment GA in SD-OCT images, which contains two keywords (i.e., deep and voting). To further test the efficiency of the proposed deep voting model, we implemented two other models: the first model was a shallow voting model in which the voting strategy was applied on a shallow neural network with a single hidden layer. Therefore, in shallow voting model, the structure of the SA neural network was composed of three layers, including one input layer, one hidden/SA layer, and one output layer. Finally, a voting decision

strategy was used to refine the segmentation results among 10 trained models. The second model was called one deep model by training a single deep model with 100% of the training set data without using the voting strategy. It should be noted that the structure of the SA deep network is the same with that in deep voting model.

Then, we tested the above two models (i.e., one deep model and shallow voting model, on Dataset 1 and Dataset 2). For the shallow voting model, we randomly selected 10,000 A-scans with GA as positive samples and 10,000 normal A-scans without GA as negative samples to train one model, and there was no intersection among the training data used in each model. For the one deep model, we used 10⁵ A-scans with GA as positive samples and 10⁵ normal A-scans without GA as negative samples to train the model. The testing phase is same for all the models. The quantitative results in interobserver and intraobserver agreement evaluation for Dataset 1 are summarized in Table 7. Table 8 summarizes the average quantitative results between the segmentation results and manual gold standards on Dataset 2. It should be noted that

Table 8. The Summarizations of the Quantitative Results (Mean \pm Standard Deviation) Between the Segmentations and Manual Gold Standards on Dataset 2 by Applying One Deep Model and Shallow Voting Model

Methods	One Deep Model	Shallow Voting Model
CC	0.860	0.964
AAD, mm ²	3.53 \pm 2.77	0.97 \pm 0.89
AAD, %	55.16 \pm 42.35	22.87 \pm 23.43
OR, %	52.44 \pm 23.00	56.91 \pm 19.30

Table 9. The Quantitative Results of the Proposed Model Based on the Patient Group and Eye Group

Dataset	Group	Index	Cubes <i>N</i>	Patient-Dependent Procedure		
				AAD, mm ²	AAD, %	OR, %
Dataset 1	Patient-based groups	P1	5	0.14 ± 0.09	12.00 ± 7.89	87.27 ± 5.43
		P2	6	0.27 ± 0.18	9.62 ± 5.09	85.22 ± 6.57
		P3	9	1.54 ± 1.50	24.75 ± 20.36	77.16 ± 14.62
		P4	4	1.02 ± 0.31	6.04 ± 2.32	92.16 ± 0.53
		P5	6	0.48 ± 0.41	6.87 ± 5.87	88.54 ± 3.86
		P6	10	0.73 ± 0.25	12.37 ± 6.35	86.92 ± 5.47
		P7	10	0.31 ± 0.14	4.79 ± 2.17	93.43 ± 1.29
		P8	1	0.72	8.41	88.52
Dataset 2	Eye-based groups	OD	31	0.63 ± 0.83	11.56 ± 13.98	87.11 ± 10.40
		OS	20	0.74 ± 0.74	11.37 ± 6.24	86.68 ± 5.53
		OD	31	0.38 ± 0.29	8.60 ± 9.47	80.57 ± 12.59
		OS	23	0.27 ± 0.24	8.58 ± 9.48	82.03 ± 9.79

the results of the proposed deep voting model were listed in Tables 3 and 4, respectively.

From Tables 7 and 8 we can observe that the shallow voting model outperforms the deep model without using the voting strategy, which means that the voting strategy is an efficient way to further improve the performance of the model. Comparing the results obtained with the shallow voting model and the proposed deep voting model, we can find that the representations learned by stacked autoencoders can decrease the redundant information of the input data, and preserve more useful information for the final classification.

In our experimental results, we quantitatively evaluated the proposed model over all the OCT cases without considering any patient group or eye group. To further demonstrate the robustness of proposed model on different patients and different eyes, Table 9 lists the grouping quantitative results based on the patient-dependent procedure (Testing I and Testing II) and patient-independent procedure (Testing III). For Dataset 1, 51 SD-OCT cubes from 12 eyes of 8 patients were firstly grouped based on the patients as 8 groups (from P1–P8), and then grouped based on eyes as two groups where right eye group is oculus dexter (OD) and left eye group is oculus sinister (OS). For Dataset 2, 54 SD-OCT cube scans from 54 eyes of 54 patients were grouped based on eyes as two groups (i.e., OD group and OS group). From Table 9 we can observe that the proposed model is robust to the patient group and eye group.

Our proposed model moves past the limitations that retinal layer segmentation present, thus making it more practical in real-life applications. Because GA is

generally associated with retinal thinning and loss of RPE and photoreceptors, state-of-the-art algorithms mainly segment the GA regions based on the projection image generated with the voxels between the RPE and the choroid layers, which means the corresponding methods rely on the accuracy of retinal layer segmentations. Comparatively, the data samples (A-scans) with 1024 features, without using any layer segmentation results, are directly fed into the network during the training and testing phases.

Our method also does not rely on large data sets. In the training phase for the Dataset 1 and Dataset 2, we only, respectively, needed approximately 5% and 9% of the total data to train our model. The proposed algorithm also showed data transfer capability, which was demonstrated with the third experiment. Even though the corresponding segmentation accuracy obtained by the cross testing procedure decreased approximately 10% on both data sets when compared with the original proposed model, the cross-testing procedure could still produce satisfactory results.

There are also some limitations about the proposed algorithm, which are summarized as follows:

(1) an interferometric method based on coherent optical beams, one of the fundamental challenges with OCT imaging is the presence of speckle noise in the tomograms. However, in the proposed model, the 1D data samples (A-scans) were directly fed into the network as the input layer, which meant that the spatial consistency among samples were not taken into account. Therefore, the proposed deep voting model was sensitive to the noise, and data preprocessing for image denoising was necessary. Our future work will focus on how to take the spatial consistency

Table 9. Extended

Patient-Independent Procedure		
AAD, mm ²	AAD, %	OR, %
0.21 ± 0.10	18.29 ± 9.06	82.72 ± 6.70
0.42 ± 0.41	14.95 ± 8.50	80.97 ± 6.83
1.88 ± 1.84	29.03 ± 25.05	73.90 ± 16.32
1.27 ± 0.86	7.76 ± 5.39	89.99 ± 1.73
0.55 ± 0.43	7.81 ± 6.06	85.21 ± 4.19
0.81 ± 0.38	14.29 ± 9.11	85.26 ± 7.49
0.38 ± 0.39	6.08 ± 6.37	88.53 ± 3.07
1.26	15.04	82.33
0.737 ± 1.03	13.85 ± 17.33	83.48 ± 11.08
0.964 ± 0.97	15.48 ± 7.90	83.41 ± 6.83
0.504 ± 0.36	11.13 ± 10.70	77.74 ± 13.38
0.377 ± 0.26	12.91 ± 14.05	78.36 ± 12.40

among samples into account in the deep learning models, for example, using convolutional neural networks or recurrent neural networks. (2) The voting strategy used in this paper is heuristic and intuitive, which treats each result obtained by ten models as equally important. In the future, we plan to automatically detect the importance for each model. (3) How deep should the deep network be? The deep network used in this paper is actually not very deep. In our experiments, we tried to add more hidden layers to further improve the performance. Unfortunately, in GA segmentation, we found that the accuracy improvement with more hidden layers was very limited and only served to increase the training cost. This is mainly due to the varying structural differences in OCT data. As shown in Figure 3, the intensity profiles of the selected samples demonstrate that the structure for A-scans with GA and normal A-scans without GA vary greatly, and it is very difficult for the corresponding deep learning model to capture the uniform or general structural information from these A-scans. In the future, we plan to detect the foveal center in the OCT data, which would further reduce the structural variance among different OCT scans and improve the performances of deep neural networks. (4) Instead of using the current widely used networks, like AlexNet and GoogleNet, in this paper, we used the sparse autoencoder to construct our deep model to discover and represent the sparsity of OCT data. From Figure 2, a trend of sparsity can be clearly observed with the data propagation from the top layer to the bottom layer of the network, which further indicates the efficiency of the proposed model. How to segment GA with pretrained deep network is

out of the scope of this paper and subjects to future research.

Acknowledgments

Supported by the National Natural Science Foundation of China under Grants No. 61401209, 61671242, 61701192, and 61672291, the Natural Science Foundation of Jiangsu Province, China (Youth Fund Project) under Grant No. BK20140790, the Natural Science Foundation of Shandong Province, China (Youth Fund Project) under Grant No. ZR2017QF004, the Fundamental Research Funds for the Central Universities under Grants No. 30916011324 and 30920140111004, China Postdoctoral Science Foundation under Grants No. 2014T70525, and 2013M531364 and 2017M612178, and Research to Prevent Blindness.

Disclosure: **Z. Ji**, None; **Q. Cheng**, None; **S. Niu**, None; **T. Leng**, None; **D.L. Rubin**, None

References

1. Klein R, Klein BE, Knudtson MD, Meuer SM, Swift M, Gangnon RE. Fifteen-year cumulative incidence of age-related macular degeneration: the Beaver Dam Eye Study. *Ophthalmology*. 2007; 114:253–262.
2. Sunness JS. The natural history of geographic atrophy, the advanced atrophic form of age-related macular degeneration. *Mol Vis*. 1999;5:25.
3. Ying GS, Kim BJ, Maguire MG, et al. Sustained visual acuity loss in the comparison of age-related macular degeneration treatments trials. *JAMA Ophthalmol*. 2014;32:915–921.
4. Nunes RP, Gregori G, Yehoshua Z, et al. Predicting the progression of geographic atrophy in age-related macular degeneration with SD-OCT en face imaging of the outer retina. *Ophthalmic Surg Lasers Imaging Retina*. 2013; 44:344–359.
5. 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*. 2015;242:183–199.
6. Chaikitmongkol V, Tadarati M, Bressler N M. Recent approaches to evaluating and monitoring geographic atrophy. *Curr Opin Ophthalmol*. 2016; 27:217–223.

7. Schmitz-Valckenberg S, Sadda S, Staurengi G, Chew EY, Fleckenstein M, Holz FG. Geographic atrophy: semantic considerations and literature review. *Retina*. 2016;36:2250–2264.
8. Abramoff M, Garvin M, Sonka M. Retinal imaging and image analysis. *IEEE Rev Biomed Eng*. 2010;3:169–208.
9. Feeny AK, Tadarati M, Freund DE, Bressler NM, Burlina P. Automated segmentation of geographic atrophy of the retinal epithelium via random forests in AREDS color fundus images. *Comput Biol Med*. 2015;65:124–136.
10. Panthier C, Querques G, Puche N, et al. Evaluation of semiautomated measurement of geographic atrophy in age-related macular degeneration by fundus autofluorescence in clinical setting. *Retina*. 2014;34:576–582.
11. Lee N, Laine AF, Barbazetto I, Busuoic M, Smith R. Level set segmentation of geographic atrophy in macular autofluorescence images. *Invest Ophthalmol Vis Sci*. 2006;47:2125.
12. Lee N, Smith RT, Laine AF. Interactive segmentation for geographic atrophy in retinal fundus images. *Conf Rec Asilomar Conf Signals Syst Comput*. 2008;42:655–658.
13. Deckert A, Schmitz-Valckenberg S, Jorzik J, Bindewald A, Holz FG, Mansmann U. Automated analysis of digital fundus autofluorescence images of geographic atrophy in advanced age-related macular degeneration using confocal scanning laser ophthalmoscopy (cSLO). *BMC Ophthalmol*. 2005;5:8.
14. Hu Z, Medioni GG, Hernandez M, Sadda SR. Automated segmentation of geographic atrophy in fundus autofluorescence images using supervised pixel classification. *J Med Imaging*. 2015;2:014501.
15. Ramsey DJ, Sunness JS, Malviya P, Applegate C, Hager GD, Handa JT. Automated image alignment and segmentation to follow progression of geographic atrophy in age-related macular degeneration. *Retina*. 2014;34:1296–1307.
16. Zohar Y, Rosenfeld PJ, Gregori G, Feuer WJ. Progression of geographic atrophy in age-related macular degeneration imaged with spectral-domain optical coherence tomography. *Ophthalmology*. 2011;118:679–686.
17. De Niro JE, McDonald HR, Johnson RN. Sensitivity of fluid detection in patients with neovascular AMD using spectral domain optical coherence tomography high-definition line scans. *Retina*. 2014;34:1163–1166.
18. Shuang L, Wang B, Yin B. Retinal nerve fiber layer reflectance for early glaucoma diagnosis. *J Glaucoma*. 2014;23:e45–e52.
19. Folgar FA, Yuan EL, Sevilla MB, et al.; for the Age Related Eye Disease Study 2 Ancillary Spectral-Domain Optical Coherence Tomography Study Group. Drusen volume and retinal pigment epithelium abnormal thinning volume predict 2-year progression of age-related macular degeneration. *Ophthalmology*. 2016;123:39–50.
20. 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*. 2013;4:2729–2750.
21. Hu Z, Medioni GG, Hernandez M, Hariri A, Wu X, Sadda SR. Segmentation of the geographic atrophy in spectral-domain optical coherence tomography and fundus autofluorescence images. *Invest Ophthalmol Vis Sci*. 2013;54:8375–8383.
22. 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*. 2016;7:581–600.
23. 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*. 2016;123:1737–1750.
24. LeCun Y, Bengio Y, Hinton G. Deep learning. *Nature*. 2015;521:436–444.
25. Shen DG, Wu GR, Suk HI. Deep learning in medical image analysis. *Annu Rev Biomed Engin*. 2017;19:221–248.
26. Kleesiek J, Urban G, Hubert A, et al. Deep MRI brain extraction: a 3D convolutional neural network for skull stripping. *NeuroImage*. 2016;129:460–469.
27. Wu G, Kim M, Wang Q, Munsell BC, Shen D. Scalable high-performance image registration framework by unsupervised deep feature representations learning. *IEEE Trans Biomed Engin*. 2016;63:1505–1516.
28. Suk HI, Lee SW, Shen DG. Hierarchical feature representation and multimodal fusion with deep learning for AD/MCI diagnosis. *NeuroImage*. 2014;101:569–582.
29. Suk HI, Shen D. Deep learning in diagnosis of brain disorders. In: Lee SW, Bulthoff HH, Muller KR, eds. *Recent Progress in Brain and Cognitive Engineering*. Dordrecht, the Netherlands: Springer; 2015:203–213.

30. Dou Q, Chen H, Yu L, et al. Automatic detection of cerebral microbleeds from MR images via 3D convolutional neural networks. *IEEE Trans Med Imaging*. 2016;35:1182–1195.
31. Abràmoff MD, Lou Y, Erginay A, Clarida W, Amelon R, Folk JC, Niemeijer M. Improved automated detection of diabetic retinopathy on a publicly available dataset through integration of deep learning. *Invest Ophthalmol Vis Sci*. 2016;57:5200–5206.
32. Asaoka R, Murata H, Iwase A, Araie M. Detecting preperimetric glaucoma with standard automated perimetry using a deep learning classifier. *Ophthalmology*. 2016;123:1974–1980.
33. Gao X, Lin S, Wong TY. Automatic feature learning to grade nuclear cataracts based on deep learning. *IEEE Trans Biomed Eng*. 2015;62:2693–2701.
34. Prentašić P, Heisler M, Mammo Z, et al. Segmentation of the foveal microvasculature using deep learning networks. *J Biomed Opt*. 2016;21:075008.
35. Lee CS, Baughman DM, Lee AY. Deep learning is effective for the classification of OCT images of normal versus age-related macular degeneration. *Ophthalmology*. 2017;8:1090–1095.
36. Gargeya R, Leng T. Automated identification of diabetic retinopathy using deep learning. *Ophthalmology*. 2017;124:962–969.
37. Cameron A, Lui D, Boroomand A, Glaister J, Wong A, Bizheva K. Stochastic speckle noise compensation in optical coherence tomography using non-stationary spline-based speckle noise modelling. *Biomed Opt Express*. 2013;4:1769–1785.
38. Maggioni M, Katkovnik V, Egiazarian K, Foi A. A nonlocal transform-domain filter for volumetric data denoising and reconstruction. *IEEE Trans Image Process*. 2013;22:119–133.
39. Ng A. Sparse autoencoder. CS294A Lecture notes. 2011;72:1–19.
40. Bottou L. Large-scale machine learning with stochastic gradient descent. In: Lechevallier Y, Saporta G, eds. *Proceedings of COMPSTAT'2010*. Heidelberg: Physica-Verlag HD; 2010:177–186.
41. Yegnanarayana B. *Artificial Neural Networks*. New Delhi: PHI Learning Pvt. Ltd.; 2009.