RETINAL LAYER SEGMENTATION USING 1D+2D U-NET FROM OCT IMAGES

Tsubasa Konno[⋆]*, Takahiro Ninomiya*[†]*, Kanta Miura*[⋆]*, Koichi Ito*[⋆]*,*
Noriko Himori[†]*, Parmanand Sharma*[†]*, Toru Nakazawa*[†]*, and Takafumi Aoki*[⋆]

[⋆] Graduate School of Information Sciences, Tohoku University, Japan.
[†] Department of Ophthalmology, Graduate School of Medicine, Tohoku University, Japan.

ABSTRACT

In ophthalmic diagnosis, it is crucial to observe the structure of the retinal layers, and the use of Optical Coherence Tomography (OCT) is growing for this purpose. Segmentation methods for OCT images have been proposed to measure the thickness of each retinal layer. Methods for detecting the boundaries between retinal layers using U-Net, which consists of 2D CNN, 3D CNN, or a combination of both, have exhibited high segmentation accuracy. On the other hand, these methods assume that the retinal shape of the OCT image is flattened to normalize the changes in the retinal shape due to individuality and diseases. Retinal diseases and poor-quality OCT images may prevent flattening, and therefore, methods without flattening are required. To address this problem, we propose a method for detecting the boundaries between retinal layers using 1D CNN, utilizing the fact that the pixels of each retinal layer exist in the vertical direction. The proposed method employs two U-Nets consisting of 1D CNN that detects boundaries pixel by pixel and 2D CNN that considers the horizontal continuity of the boundaries. Through experiments using public datasets, we demonstrate that the proposed method can segment retinal layers more accurately than conventional methods.

***Index Terms*—** OCT, segmentation, retinal layer, retina, U-Net

1. INTRODUCTION

Optical Coherence Tomography (OCT) is widely used in ophthalmology since it can noninvasively observe the retina in three dimensions. The thickness of retinal layers needs to be measured from OCT images for the diagnosis of diseases that affect the thickness of retinal layers, such as Multiple Sclerosis (MS), Age-related Macular Degeneration (AMD), glaucoma, etc. With the rapid development of deep learning techniques [1], segmentation methods of OCT images using deep learning [2] have achieved higher accuracy than conventional graph-based methods [3, 4, 5, 6].

The pioneering method using Convolutional Neural Networks (CNNs) for retinal layer segmentation is ReLayNet [7]. ReLayNet assigns retinal and other labels to each pixel using U-Net [8]. This method has a problem that the anatomical order of the retinal layers cannot be taken into account because of pixel-wise labeling. On the other hand, there are methods that detect the boundaries between retinal layers instead of segmenting the retinal layers. FCBR [9] detects the boundaries according to the anatomical order of the retinal layers. Advanced methods of FCBR have been proposed, such as SASR [10]. SASR [10] employs a 2D-3D hybrid network to take into account the displacement between OCT images. The above boundary detection methods apply flattening to the OCT image as preprocessing to approximate the boundaries as straight lines. Since the shape of the retinal layers varies depending on individuality, aging, and disease, a wide variety of shapes have to be taken into account for detection. Therefore, the above methods simplify the problem by normalizing the OCT image so that the Bruch's membrane (BM) of the retina is flat, approximating the boundary detection as a straight line detection [11]. If the quality of the OCT image is low or the shape of the retinal layers changes significantly due to disease, flattening may fail. Even if flattening can be applied to OCT images, it is not always possible to convert the boundaries of retinal layers into straight lines, depending on the shape of the retina. For stable OCT image segmentation, it is necessary to develop a boundary detection method independent of retinal shapes.

In this paper, we propose a boundary detection method for retinal layers utilizing the fact that each boundary between the retinal layers is represented by only one pixel in each longitudinal direction in B-scan images of OCT. The proposed method employs the combination of longitudinal 1D U-Net and 2D U-Net. Focusing on only one column in the longitudinal direction of OCT images, there are pixels that indicate the boundaries of all retinal layers, even if the retinal shapes are different. We consider detecting the retinal layer boundaries by extracting features from only the longitudinal direction of the OCT image using 1D CNN. Fig. 1 shows the application area of 2D and 1D convolution for flat shape and diagonal shape of the retinal layers. In 2D convolution, different features are extracted from flat and diagonal retinal layers since the application areas are different for each layer. In 1D convolution, the same features are extracted independent of the shape of the retinal layers since the application

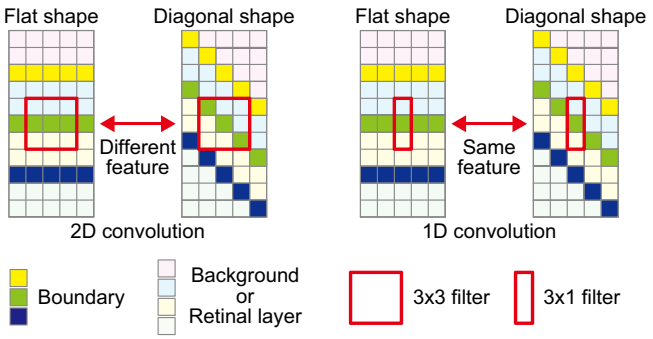

Fig. 1. Application area of 2D and 1D convolution for flat shape and diagonal shape of the retinal layers.

area is the same for the flat and diagonal retinal layers. We combine feature extraction by 2D CNN in addition to feature extraction by 1D CNN to achieve stable boundary detection since 1D CNN cannot consider the connections in the lateral direction. The proposed method does not require flattening of the OCT image and does not depend on the shape of the retinal layers unlike conventional boundary detection methods. This paper also proposes a new loss function to detect the boundaries smoothly. Through a set of experiments using three public datasets, we demonstrate the effectiveness of the proposed method compared to conventional methods.

2. METHOD

We describe the retinal layer segmentation method using longitudinal 1D U-Net and 2D U-Net proposed in this paper. Fig. 2 shows an overview of the proposed method. An input image for the proposed method is an OCT image acquired by B-Scan. First, an OCT image is input to the longitudinal 1D U-Net and 2D U-Net, and a feature map with 64 channels is obtained from each of them. The architecture of 1D U-Net and 2D U-Net is the same as shown in Fig. 2, except that max pooling, convolution, and batch normalization are longitudinal 1D and 2D operations, respectively. The feature maps extracted by 1D U-Net and 2D U-Net are concatenated in the channel direction to obtain the feature map with 128 channels. Next, the concatenated feature map is input to the layer branch, which performs pixel-wise labeling of retinal layers, and to the surface branch, which predicts the location of the boundaries between retinal layers. In the layer branch, after processing with res block and 1×1 convolution, a layer map representing the probability of a class for each pixel is output by channel-wise softmax. If C is the number of classes, i.e., the number of boundaries, the number of layers in the layer map is $C + 1$. The layer branch is only used to calculate the loss function in training as well as the conventional methods [9, 10]. If the retina contains an edema, the edema mask is created based on the edema label detected in the layer branch and superimposed on the boundaries detected in the

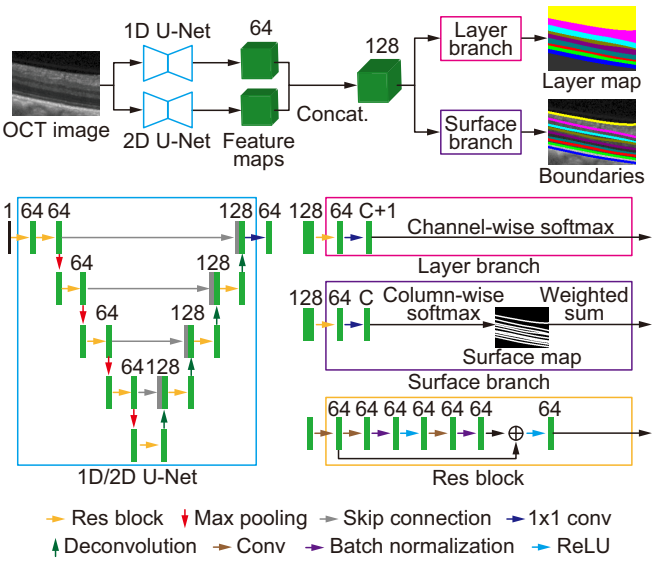

Fig. 2. Overview of the proposed method, where the numbers indicate the number of channels of features.

surface branch. In this case, a layer map with $C + 2$ classes is output since the class indicating the edema is added. In the surface branch, after processing with res block and 1×1 convolution, surface maps representing the existence probability of each boundary in each column are output by column-wise softmax. Similar to FCBR [9] and SASR [10], we obtain the sub-pixel level boundaries by weighting the sum of the existence probability of boundaries and the image coordinates in each column as

$$b_i^j = \sum_{h=1}^{H} p_i^j(h) \cdot h, \tag{1}$$

where i indicates the column index of the image, j indicates the class of boundaries, h indicates the row index of the image, and H indicates the height of the image. $p_i^j(h)$ indicates the existence probability of the boundary for class j at pixel (i, h). Finally, a segmentation map of the retinal layers can be obtained by mapping the retinal layers between the detected boundaries.

The network used in the proposed method is trained to minimize the loss functions for the layer branch and the surface branch in the same way as the conventional methods [9, 10]. For the layer branch, we use $\mathcal{L}_{CE+Dice}$, which is the combination of the Dice loss [12] and cross-entropy loss between the layer map and the segmentation mask of the ground truth. For the surface branch, we use $\mathcal{L}_{CE}$, which is the cross-entropy loss between the surface map and the ground truth, and $\mathcal{L}_{L1}$, which is the L1 loss between the location of the detected boundaries and the ground truth. In addition to the above loss functions, SASR [10] uses SmoothS loss to keep the continuity of the boundaries between columns to estimate the smooth boundaries. If the connection of the bound-

aries between columns is horizontal on the image, SmoothS loss becomes small, and therefore, SmoothS loss may result in estimating boundaries close to a straight line. Therefore, the smooth boundaries can be estimated by giving weights to SmoothS loss according to each retinal layer. Since these weights are hyperparameters, it is time-consuming to find the optimal weights experimentally as the number of target retinal layers increases. We therefore propose a new smooth loss, $\mathcal{L}_S$, that does not require any hyperparameters, which is defined by

$$\mathcal{L}_S = \frac{1}{C(W-1)} \sum_{j=1}^{C} \sum_{i=1}^{W-1} \left\{ (b_{i+1}^j - b_i^j) - (g_{i+1}^j - g_i^j) \right\}^2, \tag{2}$$

where i indicate the column index of the image, W indicates the width of the image, j indicates the class of boundaries, C indicates the number of classes, b_i^j indicates the estimated coordinate of the boundary j in column i, and g_i^j indicates the ground-truth coordinate of the boundary j in column i. By training the network so that the difference between the estimated boundary positions in adjacent columns is close to that of the ground truth, we can obtain smooth boundaries, which are independent of the type of retinal layers and images. The total loss function used in the proposed method is given by

$$\mathcal{L}_{all} = \mathcal{L}_{CE+Dice} + \mathcal{L}_{CE} + \mathcal{L}_{L1} + \alpha \mathcal{L}_S, \tag{3}$$

where α indicates a weight parameter. We use $\alpha = 10$ in this paper.

The order of retinal layers is anatomically determined, and therefore the boundaries must be detected according to this order. In the conventional methods [9, 10], the topology guarantee module is used to preserve the order of the boundaries. If the positions of the boundaries are switched, the order of the detected boundaries is made consistent with the anatomical order by replacing the coordinate of the lower boundary on the image with the upper coordinate. In FCBR [9], the L1 loss is calculated after correcting the order of the boundaries using the topology guarantee module. Correcting the order based on boundaries detected at incorrect positions may result in loss calculations with large errors. The proposed method also uses the topology guarantee module, while the L1 loss and the smooth loss are computed at the detected boundaries before correction.

3. EXPERIMENTS

This section describes experiments to evaluate the effectiveness of the proposed method for detecting the boundaries of retinal layers from OCT images.

3.1. Datasets

We use three public datasets in the experiments: OCT MS and Healthy Control (MSHC) dataset [13], Duke Cyst DME (Duke DME) dataset [3], and World's Largest Online Annotated (WLOA) SD-OCT dataset [14]. MSHC consists of OCT images acquired from 14 healthy subjects and 21 MS patients. Each OCT image consists of 49 B-scan images with $496 \times 1,024$ pixels. This dataset provides 9 boundaries as the ground truth. We use 6 healthy subjects and 9 MS patients from the end of the subject number for training and validation, and the remaining for test, as in FCBR [9]. Of the 15 subjects in the training and validation data, one healthy subject and two MS patients are used for validation in ascending order of subject number. Duke DME consists of OCT images acquired from 10 DME patients. Each OCT image consists of 11 B-scan images with 496×768 pixels. This dataset provides 8 boundaries and an edema mask as the ground truth. Note that we excluded pixels for which no boundary is defined from the calculation of the loss functions and the accuracy evaluation in the experiments. We use 6 patients for training, 2 patients for validation, and the remaining for test from the front of the subject number. WLOA consists of OCT images acquired from 115 healthy subjects and 269 AMD patients. Each OCT image consists of 100 B-scan images with $512 \times 1,000$ pixels. This dataset provides 3 boundaries as the ground truth. In this dataset, the ground truth of the boundaries is defined only around the central fovea. We use the regions of 512×400 pixels extracted from the center of 40 B-scan images around the central fovea, as in SASR [10]. Note that we cannot flatten 9 OCT images under the same conditions as SASR [10], and thus we do not use these images in the experiments[1]. We use 60% for training, 20% for validation, and the remaining for test from 115 healthy subjects and 260 AMD patients from the front of the subject number.

3.2. Experimental Condition

In training of the proposed method, Adam is used as the optimizer, the batch size is 4, the learning rate is 0.0001, and training continues until convergence. We employ data augmentation of horizontal flipping, vertical scaling, Gaussian noise, and contrast changes with probability 0.5 for each in training. In vertical scaling, the image is scaled vertically by a factor of 0.9–1.1 and is cropped to the same size before scaling. In random contrast, the pixel values are multiplied by 0.8–1.2. In addition, random erasing [15] is applied to the columns of the image, since some pixels in the vertical direction have low quality due to blood vessels.

We evaluate the accuracy of boundary detection with and without flattening of the input image. Flattening estimates BM using the intensity gradient method [11], and shifts each column up and down to make BM horizontal, as in the conventional methods [9, 10]. Note that upper and lower background areas are removed to reduce the memory usage as in [9, 10]. For each dataset, the size of the images after flat-

[1]Flattening is failed in subject numbers 1003, 1026, 1044, 1046, 1054, 1127, 1217, 1218, and 1250 of AMD patients.

Table 1. Experimental results of each method for MSHC, Duke DME, and WLOA, where $*$ indicates the result obtained in our reproduced experiments and the units for MAD and SD are μm.

Method	w/ Flattening			w/o Flattening		
	MSHC	Duke DME	WLOA	MSHC	Duke DME	WLOA
FCBR [9]	2.83±0.99	6.70	2.78±3.31	—	—	—
FCBR*	2.79±0.41	5.04±0.35	2.37±1.04	4.15±5.94	4.86±0.16	1.97±1.19
SASR [10]	—	—	2.71±2.25	—	—	—
SASR*	2.94±0.58	11.82±1.47	2.49±1.18	3.58±3.04	15.63±5.10	2.16±1.15
1D U-Net	3.14±0.60	5.30±0.39	4.03±2.37	3.36±0.97	5.31±0.46	4.38±2.78
Proposed	**2.77±0.48**	**4.47±0.10**	**2.19±1.23**	**3.33±1.62**	**4.78±0.14**	**1.90±1.19**

tening is $128 \times 1,024$ pixels for MSHC, 224×768 pixels for Duke DME, and 320×400 pixels for WLOA. The accuracy of boundary detection is evaluated by Mean Absolute Distance (MAD) and Standard Deviation (SD) between the detected boundaries and the ground truth. We compare the accuracy of the proposed method with FCBR [9] and SASR [10] to demonstrate the effectiveness of the proposed method. We use our implementation for FCBR, and use the public code[2] for SASR. For FCBR and SASR, the results reported in [9, 10] are compared for reference. Note that the experimental conditions of FCBR for Duke DME and FCBR and SASR for WLOA are different from those in this paper. We compare the detection accuracy between the method using 1D U-Net, which obtains the feature map with 128 channels using only 1D U-Net, and the proposed method to confirm the effectiveness of the combination of 1D U-Net and 2D U-Net. The feature map is input to the layer branch and the surface branch as in the proposed method, and the layer map and the boundaries are obtained.

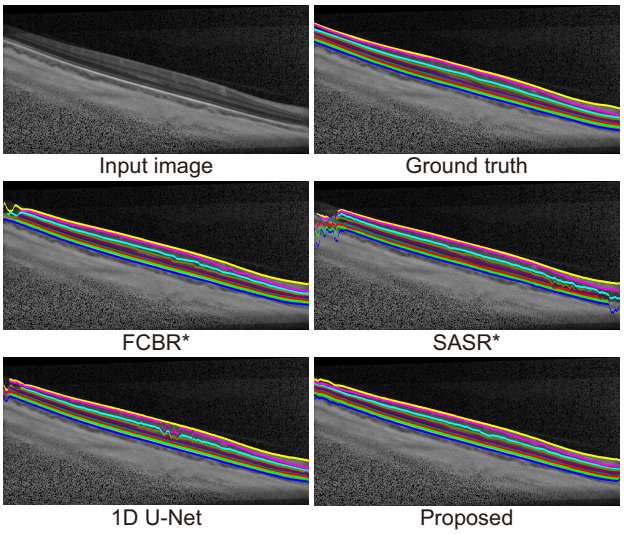

Fig. 3. An example of estimated boundaries for each method

3.3. Experimental Results and Discussion

Table 1 shows the experimental results for MSHC, Duke DME, and WLOA. In MSHC, when focusing on the case of flattening, MAD of FCBR [9] is 2.83, while the proposed method has the highest accuracy with MAD of 2.77. When focusing on the case without flattening, MAD of FCBR* is significantly high, and this method cannot deal with the variation of retinal shape. MAD of SASR* is 3.58, which may be highly accurate even without flattening, while MAD of the proposed method is 3.33, which is highly accurate in detecting the boundaries. Fig. 3 shows an example of boundary detection results for the image without flattening in MSHC. FCBR and SASR makes mistakes at the left side of the image. 1D U-Net cannot detect smooth boundaries in the center and left side of the image. The proposed method detects smooth boundaries similar to the ground truth. In Duke DME, MAD is relatively high since the images contain edemas. The proposed method has lower MAD than the conventional methods. The methods using 1D U-Net can detect boundaries independent of retinal shape since MAD of these methods almost does not change with or without flattening. The experimental results on WLOA are similar to those of other datasets. The difference is that MAD is lower for images without flattening in the proposed methods. The structure of retinal layers may be corrupted by flattening since the images cannot always be flattened exactly. Therefore, for WLOA containing images that are difficult to flatten, the accuracy of the proposed method is higher without flattening.

[2] https://github.com/ccarliu/Retinal-OCT-LayerSeg

4. CONCLUSION

We proposed a boundary detection method for retinal layers using the combination of longitudinal 1D U-Net and 2D U-Net for retinal layer segmentation. The proposed method does not require flattening of the OCT image and does not depend on the shape of the retina due to the use of 1D U-Net. Through a set of experiments using three public datasets, we demonstrated the effectiveness of the proposed method compared to conventional methods.

5. COMPLIANCE WITH ETHICAL STANDARDS

This research study was conducted retrospectively using human subject data made available in open access. Ethical approval was not required as confirmed by the license attached with the open access data.

6. ACKNOWLEDGMENTS

This work was supported in part by JSPS KAKENHI Grant Numbers 21H03457 and 23H00463, and the WISE Program for AI Electronics, Tohoku University.

7. REFERENCES

[1] I. Goodfellow, Y. Bengio, and A. Courville, *Deep Learning*, MIT Press, 2016.

[2] I. A. Viedma, D. Alonso-Caneiro, S. A. Read, and M. J. Collins, "Deep learning in retinal optical coherence tomography (OCT): A comprehensive survey," *Neurocomputing*, vol. 507, no. 1, pp. 247–264, Oct. 2022.

[3] S. J. Chiu, M. A. Allingham, P. S. Mettu, J. A. Cousins, S. W.and Izatt, and S. Farsiu, "Kernel regression based segmentation of optical coherence tomography images with diabetic macular edema," *Biomed. Opt. Express*, vol. 6, no. 4, pp. 1172–1194, Apr. 2015.

[4] S. J. Chiu, X. T. Li, P. Nicholas, C. A. Toth, J. A. Izatt, and S. Farsiu, "Automatic segmentation of seven retinal layers in SDOCT images congruent with expert manual segmentation," *Opt. Express*, vol. 18, no. 18, pp. 19413–19428, Aug. 2010.

[5] S. P. K. Karri, D. Chakraborthi, and J. Chatterjee, "Learning layer-specific edges for segmenting retinal layers with large deformations," *Biomed. Opt. Express*, vol. 7, no. 7, pp. 2888–2901, July 2016.

[6] F. Rathke, M. Desana, and C. Schnörr, "Locally adaptive probabilistic models for global segmentation of pathological OCT scans," *Proc. Int'l Conf. Medical Image Computing and Computer Assisted Intervention*, vol. 10433, pp. 177–184, Sept. 2017.

[7] A.G. Roy, S. Conjeti, S.P.K. Karri, D. Sheet, A. Katouzian, C. Wachinger, and N. Navab, "ReLayNet: Retinal layer and fluid segmentation of macular optical coherence tomography using fully convolutional networks," *Biomed. Opt. Express*, vol. 8, no. 8, pp. 3627–3642, Aug. 2017.

[8] O. Ronneberger, P. Fischer, and T. Brox, "U-Net: Convolutional networks for biomedical image segmentation," *Proc. Int'l Conf. Medical Image Computing and Computer Assisted Intervention*, pp. 234–241, Oct. 2015.

[9] Y. He, A. Carass, B. M. Jedynak, S. D. Solomon, S. Saidha, P. A. Calabresi, and J. L. Prince, "Fully convolutional boundary regression for retina OCT segmentation," *Proc. Int'l Conf. Medical Image Computing and Computer Assisted Intervention*, pp. 120–128, Oct. 2019.

[10] H. Liu, D. Wei, D. Lu, Y. Li, K. Ma, L. Wang, and Y. Zheng, "Simultaneous alignment and surface regression using hybrid 2D-3D networks for 3D coherent layer segmentation of retina OCT images," *Proc. Int'l Conf. Medical Image Computing and Computer Assisted Intervention*, pp. 108–118, Sept. 2021.

[11] A. Lang, A. Carass, M. Hauser, E. S. Sotirchos, P. A. Calabresi, H. S. Ying, and J. L. Prince, "Retinal layer segmentation of macular OCT images using boundary classification," *Biomed. Opt. Express*, vol. 4, no. 7, pp. 1133–1152, July 2013.

[12] F. Milletari, N. Navab, and S. Ahmadi, "V-Net: Fully convolutional neural networks for volumetric medical image segmentation," *Proc. Int'l Conf. 3D Vision*, pp. 565–571, Oct. 2016.

[13] Y. He, A. Carass, S. D. Solomon, S. Saidha, P. A. Calabresi, and J. L. Prince, "Retinal layer parcellation of optical coherence tomography images: data resource for multiple sclerosis and healthy controls," *Data Brief*, vol. 22, pp. 601–604, Feb. 2018.

[14] S. Farsiu, S. J. Chiu, R. V. O'Connell, F. A. Folgar, E. Yuan, J. A. Izatt, and C. A. Toth, "Quantitative classification of eyes with and without intermediate age-related macular degeneration using optical coherence tomography," *Ophthalmology*, vol. 121, no. 1, pp. 162–172, Jan. 2014.

[15] Z. Zhong, L. Zheng, G. Kang, S. Li, and Y. Yang, "Random erasing data augmentation," *Proc. AAAI Conf. Artificial Intelligence*, vol. 34, no. 7, pp. 13001–13008, Feb. 2020.