Boundary Aware U-Net for Retinal Layers Segmentation in Optical Coherence Tomography Images

Bo Wang, Wei Wei, Shuang Qiu, Shengpei Wang, Dan Li, and Huiguang He, *Senior Member, IEEE*

***Abstract*—Retinal layers segmentation in optical coherence tomography (OCT) images is a critical step in the diagnosis of numerous ocular diseases. Automatic layers segmentation requires separating each individual layer instance with accurate boundary detection, but remains a challenging task since it suffers from speckle noise, intensity inhomogeneity, and the low contrast around boundary. In this work, we proposed a boundary aware U-Net (BAU-Net) for retinal layers segmentation by detecting accurate boundary. Based on encoder-decoder architecture, we design a dual tasks framework with low-level outputs for boundary detection and high-level outputs for layers segmentation. Specifically, we first use the multi-scale input strategy to enrich the spatial information in the deep features of encoder. For low-level features from encoder, we design an edge aware (EA) module in skip connection to extract the pure edge features. Then, a U-structure feature enhanced (UFE) module is designed in all skip connections to enlarge the features receptive fields from the encoder. Besides, a canny edge fusion (CEF) module is introduced to aforementioned architecture, which can fuse the priory edge information from segmentation task to boundary detection branch for a better predication. Furthermore, we model each boundary as a vertical coordinates distribution for boundary detection. Based on this distribution, a topology guarantee loss with combined A-scan regression loss and structure loss is proposed to make an accurate and guaranteed topological boundary set. The method is evaluated on two public datasets and the results demonstrate that the BAU-Net achieves promising performance than other state-of-the-art methods.**

***Index Terms*—BAU-Net, boundary detection, OCT, retinal layers segmentation, topology guarantee loss.**

Manuscript received November 24, 2020; revised January 28, 2021; accepted March 9, 2021. Date of publication March 17, 2021; date of current version August 5, 2021. This work was supported in part by the National Key Research and Development Program of China under Grant 2017YFB1302704, in part by the National Natural Science Foundation of China under Grants 61976209, 62020106015, and 61906188, in part by the Chinese Academy of Sciences (CAS) International Collaboration Key Project under Grant 173211KYSB20190024; and in part by the Strategic Priority Research Program of CAS under Grant XDB32040000. *(Corresponding author: Huiguang He.)*

Bo Wang, Wei Wei, Shuang Qiu, Shengpei Wang, and Dan Li are with the School of Artificial Intelligence, University of Chinese Academy of Sciences, Beijing 100049, China, and also with the Research Center for Brain-Inspired Intelligence, National Laboratory of Pattern Recognition, Institute of Automation, Chinese Academy of Sciences, Beijing 100190, China (e-mail: wangbo2018@ia.ac.cn; weiwei2018@ia.ac.cn; shuang.qiu@ia.ac.cn; wangshengpei2014@ia.ac.cn; lidan2017@ia.ac.cn).

Huiguang He is with the School of Artificial Intelligence, University of Chinese Academy of Sciences, Beijing 100049, China with the Research Center for Brain-Inspired Intelligence, National Laboratory of Pattern Recognition, Institute of Automation, Chinese Academy of Sciences, Beijing 100190, China, and also with the Center for Excellence in Brain Science and Intelligence Technology, Chinese Academy of Sciences, Beijing 100190, China (e-mail: huiguang.he@ia.ac.cn).

Digital Object Identifier 10.1109/JBHI.2021.3066208

I. Introduction

OPTICAL coherence tomography (OCT) is a high resolution and non-invasive imaging modality, which utilizes the near-infrared light to capture 3D cross-sectional image and permits clear visualization of the various constituent layers and architecture of the retina [1]. The 3D OCT volumes are composed of a series of adjacent cross-sectional images named B-scans. Each column of B-scan image is named the A-scan which parallel to the direction of light propagation, as illustrated in Fig. 1. Since a rapidly growing imaging technology, OCT has been widely used in various biological and medical fields. Many retinal or optic nerve disease can change the morphology of retinal tissue, such as age-related macular degeneration (AMD) [2], multiple sclerosis (MS) [3], and glaucoma [4]. Therefore, we can analyze the morphological of each layer and compare it with normal layer through the layers information, which is contributed to understanding the evolution of a disease.

Since manual annotation of the layer boundaries is a subjective and time-consuming process, automatic retinal layers segmentation from OCT images gradually cause for concern. However, segment layers in OCT images is a daunting challenging for the following four reasons. First, the OCT images usually suffer from speckle noise and other artifacts, which can reduce the picture quality and influence accurate image analysis [5]. Second, since the refractive index difference is not remarkable in the smaller imaging depth scale, the contrast of adjacent tissue layers is very low, which can cloud the definition of boundaries [6]. Third, the presence of pathologies causes significant changes in retinal morphology and even disrupted retinal structures, which will make automatic retinal layer segmentation become difficult [7]. Last, there is a strict order from inner to outer for the retinal layer physiologically, and maintaining the correct topological structure can improve the interpretability of the segmentation algorithm [8].

In the past several years, deep learning (DL), especially convolutional neural network (CNN), has achieved great success

2168-2194 © 2021 IEEE. Personal use is permitted, but republication/redistribution requires IEEE permission.
See https://www.ieee.org/publications/rights/index.html for more information.

Authorized licensed use limited to: Lanzhou University. Downloaded on December 09,2021 at 13:09:56 UTC from IEEE Xplore. Restrictions apply.

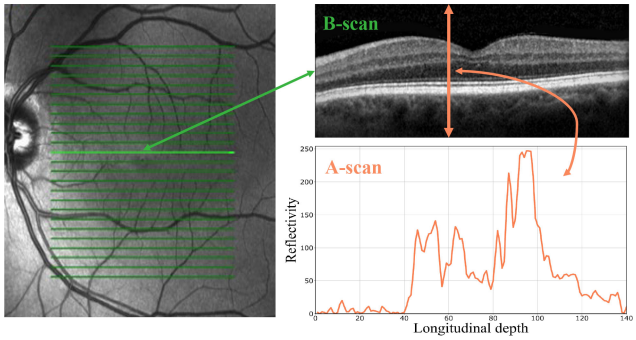

Fig. 1. Example of the relationship between B-scan and A-scan in OCT image. The left is a fundus image overlaid with lines and each line is the positions of a B-scans in 3D OCT volume. The thick line on the left represents the B-scan image (top right). The pixel value of a vertical A-scan in B-scan is shown (bottom right).

in biomedical image processing [9]. In retinal layers segmentation, state-of-the-art methods use CNN as feature extractor to generate coarse pixel-wise segmentation results and graph methods [10], [11] are adopt to layer boundaries estimation. However, these post-processing approaches require manual selection of parameters for the graph, with low generalization, and have poor performance on pathological images [7]. Therefore, many researches have focused on the end-to-end CNN framework based on U-Net [12] for retinal layers segmentation [13], [14]. Nevertheless, these methods have two major drawbacks. First, although U-Net with encoder and decoder structures achieves improved performance in layer segmentation by incorporating high-resolution low-level features, it still suffers from inaccurate object boundaries and unsatisfactory segmentation results [15]. Second, in order to obtain a layer structure that guarantees the topological order, the fully connected layer is often applied to model the boundary at the end of the network, which leads to excessive network parameters and overfitting.

To extract effective features of boundary detection, we propose a novel end-to-end framework boundary aware U-Net (BAU-Net) for retinal layers segmentation, which is based on the classical encoder-decoder structure. To obtain the retina layer with anatomical order by a fully convolutional operation, we introduce a new topology guarantee loss inspired by Surface Position Modeling [16], which using a guaranteed topology regression loss and a structure loss to compensate the independence between the adjacent boundaries and the adjacent pixels within a boundary, respectively.

In short, the contributions are summarized as follows:

1) We put forward a novel BAU-Net for retinal layers segmentation, which employs Multi-scale input, EA, UFE, CEF and different supervisions to produce more discriminative features for boundary detection and then generate more accurate retinal layers.
2) We design the topology guarantee loss with combined A-scan regression loss and structure loss for getting an accurate and guaranteed topological boundary set.
3) We conduct comprehensive experiments on two public datasets and achieve promising segmentation performance. The experimental results convince the effectiveness of the proposed framework.

II. Related Work

With the continuous development of image processing technology, various automatic retinal layer segmentation algorithms in OCT images have been proposed. According to extensive literature, we can roughly divide these existing methods into two categories: model driven approaches and data driven approaches.

A. Model Driven Methods

In the model driven approaches, algorithms that follow traditional image processing approach, like intensity gradient algorithms, deformable models, and graph-based methods.

1) Intensity Gradient Methods: The earliest model-driven algorithm is intensity gradient method. In each A-scan, it first analyzed the intensity profile by detecting the peak or the valley points, which may represent the position of the boundary on this column. Then, the detected points of adjacent A-scan are connected by curve fitting techniques to build a continuous boundary [17], [18]. The intensity gradient algorithms are easily for detecting boundaries, but it only can apply to extract few boundaries due to a small intensity difference at most layer.

2) Deformable Models: To overcome the shortcomings of intensity gradient methods, deformable models construct adaptive model based on the appearance and the shape priors of the retinal layers in OCT images, including active contours and level sets approaches. Mishra *et al.* [26] used an adaptive vector-valued kernel function to account for the presence of speckle noise based on Kass active contour model. Yazdanpanah et. al [19] adapted Chan-Vese model to segment multiple layers by applying a circular shape regularization. The edge-based level sets models were explored in [20], [21]. Novosel *et al.* [20] developed a segmentation method that used loosely coupled level sets of limiting boundary smoothness and layer thickness to segment eight retinal layers. Rossant *et al.* [21] proposed Parallel Double Snakes evolve simultaneously two contours by constraining the approximate parallelism between the adjacent boundaries. However, these methods require to define constraints for algorithm evolution based on shape prior and lead to a complex implementation and low robustness.

3) Graph-Based Methods: The graph-based methods constructed a graph for OCT image and employed energy minimization like graph cuts for OCT segmentation. Based on max-flow algorithm, Kaba *et al.* [22] adopted a multi-regional kernel graph cuts to segment the OCT layers. An automated globally optimal graph-theoretic algorithm was proposed by Antony *et al.* [23], which simultaneously segment the retinal layers and the lesion region. Dodo *et al.* [24] combined the fuzzy histogram hyperonization model and graph cut algorithm for layers segmentation in 3D OCT volumes. However, the graph-based methods are greatly affected by the initial pixel selection and the setting of search parameters, which also have expensive computations and a high running time.

Authorized licensed use limited to: Lanzhou University. Downloaded on December 09,2021 at 13:09:56 UTC from IEEE Xplore. Restrictions apply.

B. Data Driven Methods

Data driven methods are to use data with label for model training, and then the trained model is for retinal layers segmentation.

1) Classic Machine Learning Methods: Classic machine learning methods extract handcraft features of OCT images and then utilize some machine learning models (e.g, support vector machine (SVM) [25]) for pixels classification or combine the trained machine learning classifier and level set [26] or graph methods [27] for a better retinal boundary estimation. Vermeer *et al.* [25] extracted the intensities and gradients features of OCT images, and to trained a support vector machine classifier for five boundaries detection. Lang *et al.* [28] extracted features of each pixel and a random forest model is applied to classify nine layers categories, then the graph-search was used for post-processing. Xiang *et al.* [7] also used the random forest classifiers for pixels to find the initial boundaries, and a hyper live wire algorithm was proposed for refined the final boundaries. Since these methods require the hand-crafted features designed by special domain experts, generation of such features are difficult.

2) Deep Learning Methods: In recent years, as convolutional neural network has made great achievements in the medical imaging processing, many retinal layer segmentations based on convolutional neural network have been explored. Fang *et al.* [10] applied post-processing to the output of convolutional neural network based on the graph search algorithm, and achieved good performance for retinal layer segmentation. Hu *et al.* [29] designed a multiscale convolutional neural network to produce retinal layers probability maps, followed by graph search algorithm to detect the final boundaries. As fully convolutional neural network (FCN) [30] have achieved great success in segmentation task on natural images, many researches have focused on the end-to-end FCN framework [31], [32]. Lua Ngo *et al.* [33] proposed a feature learning regression network for layers segmentation without human bias. Anoop *et al.* [34] designed a DelNet based on a deep ensemble learning approach to selectively segment layers in OCT images. Hina *et al.* [35] used to classify the retinal layer pixels based on VGG-16 architecture. Since U-Net [5] achieves the excellent performance in many medical image processing tasks, many networks based on the encoder-decoder are proposed for medical image segmentation. An end-to-end deep network RelayNet [13] is proposed to segment retinal layers without any post-processing. He *et al.* [14] used two cascaded deep networks, called S-Net and T-Net. The S-Net was applied to classify each pixel and T-Net was used to guarantee correct layer ordering. In this paper, we focus on the deep encoder-decoder network for accurate retinal layers segmentation in the OCT image.

III. Materials

A. Dataset

The proposed BAU-Net is evaluated on two public datasets: HCMS [36] and DUKE DME [37].

1) Hcms: The HCMS dataset includes 35 SD-OCT volumes from the right eye of 35 subjects (including 14 healthy projects and 21 abnormal projects with MS, the mean age: 39.49), scanned on the Heidelberg Spectralis OCT device. Each volume consists of 49 B-scan images, and each B-scan's size is 496×1024. Manual delineated of nine boundaries is providing for every B-scan image. We can covert the surface positions into segmentation masks; thus, every pixel has one of ten labels (including Vitreous, RNFL, GCIP, INL, OPL, ONL, IS, OS, RPE, Choroid), which is shown in Fig. 2(a).

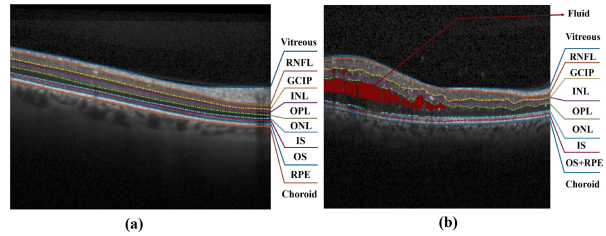

Fig. 2. An example of annotated OCT image in the dataset of HCMS (a) and DUKE DME (b). The HCMS includes nine layers and eight boundaries. RNFL is Retinal Nerve Fiber Layer; GCL is Ganglion Cell Layer; IPL is Inner Plexiform Layer; INL is Inner Nuclear Layer; OPL is Out Plexiform Layer; ONL is Outer Nuclear Layer; ISM is Inner Segment Myeloid; ISE is Inner Segment Ellipsoid; OS is Outer Segment; RPE is Retinal Pigment Epithelium. The DUKE DME includes nien retinal layers which regards the OS and RPE as the same class and a fluid class.

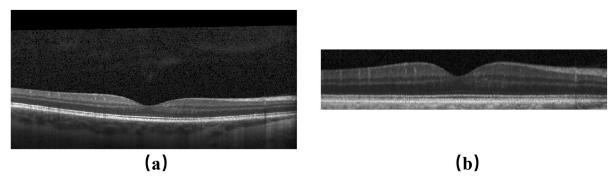

Fig. 3. An example of OCT image in the preprocessing stage. (a) the original OCT image. (b) flatten and crop.

2) Duke Dme: The DUKE DME dataset contains 110 annotated B-scan images from 10 subjects with Diabetic Macular Edema (DME, 11 B-scans per patient) of the size 512×740. These 110 B-scans are annotated for eight layers and fluid regions by expert clinician. In conjunction with the DME masks, every pixel has one of ten labels—either fluid or one of the following: Vitreous, RNFL, GCIP, INL, OPL, ONL, IS, OS+RPE, or choroid. An example of the annotation is shown in Fig. 2(b).

B. Data Preprocessing

A typical retinal B-scan requires large amounts of memory if directly processed by a deep network. Thus, we first use intensity gradient based method [14] to detect the boundary of RPE/Choroid and flatten a retinal B-scan image along this boundary. Then, all the flattened images must remove unnecessary portions of the choroid and vitreous, so we crop the size of each B-scan to 128×1024 in HCMS dataset and 224×500 in DUKE DME dataset, as shown in Fig. 3(b).

We use a sliding window mechanism to extract 128×128 pixels patches per B-scan in HCMS dataset and 128×224 pixels patches in DUKE DME dataset. Combined with our proposed method, two kinds labeled patches are used: the segmentation patches and the boundary patch. The example of the input patches and the corresponding label patches are shown in Fig. 4.

Authorized licensed use limited to: Lanzhou University. Downloaded on December 09,2021 at 13:09:56 UTC from IEEE Xplore. Restrictions apply.

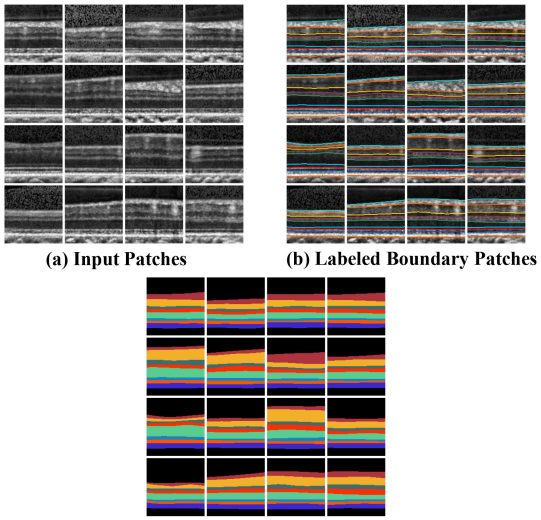

Fig. 4. Examples of extracted patches used in our methods. The input patches are shown in (a); the corresponding labeled boundary patches are shown in (b); the corresponding labeled segmentation patches are shown in (c).

IV. Methodology

Our task in this work is to assign each pixel x in a retinal OCT images, $x \in h \times w$, with a particular class of $\{0, 1, \ldots, C-1\}$ and $C = 10$. For the HCMS dataset, the 10 classes represent ten retinal layers, which are separated by nine boundary lines. For the DUKE DME, the 10 classes include nine retinal layers, separated by eight boundaries, and the Fluid. In this section, Fig. 5 introduce the construction of the BAU-Net. The EA and UEF modules of BAU-Net are presented in Fig. 6 and Fig. 7. Then, we describe a Canny Edge Fusion (CEF) module shown in Fig. 9. Finally, we discuss the training loss for our approach, especially the topology guarantee loss for boundaries detection and Fig. 8 shows the vertical coordinate distribution defined in this loss. It is worth noting that our final retinal layer result is generated by the boundary result, that is, pixel categories are assigned between adjacent boundaries in a topological order.

A. The BAU-Net

As an end-to-end deep neural network, we devised a modified U-Net model referred as Boundary Aware U-Net to detect retinal layer masks and boundaries. The BAU-Net consists of two parts: the basic BAU-Net and Canny Edge Fusion module, as illustrated in Fig. 5.

1) Backbone Encoder-Decoder With Multi-Scale Input: Since the U-Net structure has achieved promising performance in medical image segmentation, our BAU-Net also follows the encoder-decoder structure as the backbone of the network. Besides, the base block of our network is a modified resblock inspired by ResNet [38], which can mitigate the gradient vanishing problem. Furthermore, the channel of all resblock is setting to 64, empirically. Since the feature maps in deeper layers lack of spatial information, BAU-Net first employes pyramid analysis to the input image by using spatial pyramid pooling to obtain multi-scale input images with rich spatial information. In the encoder, the multi-scale inputs are turned into a 64-channel feature after a convolution operation and then summed to the backbone of the encoder.

2) Dual Tasks Framework: Motivated by a general knowledge that the features from deeper layers contain more context information, which are suitable to locate the layer regions correctly, and the low-level features contain more spatial details for recovering the object boundaries [39]. Therefore, a straightforward idea is to decompose the U-Net structure into two splits: shallow networks are used for boundary detection and deep networks are used for semantic segmentation by using different supervisions during training. As illustrated in Fig. 5, the shallow row networks are: X1-En1-De1-Feature1 and X2-En2-De2-Feature2; the deep row networks are: X3-En3-De3-Feature3 and X4-En4-De4-Feature4. We explicitly model the retinal layers mask for the features of deep network and the boundary line for the features of shallow network by using different supervisions during training. We call this structure is dual tasks framework. A notable character of the dual tasks framework is that we introduce deep supervision in BAU-Net to extract more discriminative features whilst with less redundancy. Besides, the segmentation features of the encoder are upsampling and integrate to the edge feature, which can add richer context information to improve the ability to identify different boundary instances.

3) Edge Aware Module: Generally, low-resolution feature maps often contain the most salient part and the edge information can be obtained by subtracting the down-sampled features from the original features. Inspired by this, we design an Edge Aware (EA) module to effectively extract the edge information on shallow network. In EA module, we first learn a flow field from the downsampling feature maps and warp it to obtain the explicit salient part feature representation. The edge feature is generate by subtracting the salient part from the original feature. As shown in Fig. 6, the origin feature map F and it's downsampling feature F_{low} are from the encoder in shallow row network. We first upsample F_{low} to the same size as F via bilinear interpolation, and concatenate them together followed by a 3×3 convolution layer to predict the flow map $\delta \in \mathbb{R}^{H \times W \times 2}$. Secondly, follow the the spatial transformer networks [40], each position (i, j) on standard spatial grid Ω is map to a new point $(i, j) + \delta(i, j)$ and then the differentiable bilinear sampling mechanism is applied to approximate each point $F^{salient}_{(i,j)}$. Finally, the edge feature is obtained by $F_{edge} = F - F^{salient}$.

4) U-Structure Feature Enhanced Module: Increasing the receptive field is significant in segmentation and detection tasks which bring more context information in feature extraction. The earlier methods often use atrous spatial pyramid pooling (ASPP) [41] to enlarge the receptive field but requires more computation. As shown in Fig. 5, the receptive field of the feature maps from multi-scale input is too small because it only goes through few filters. Besides, the receptive field of feature of shallow row networks is also very small. Motivated by this observation, we propose a U-structure feature enhanced (UFE)

Authorized licensed use limited to: Lanzhou University. Downloaded on December 09,2021 at 13:09:56 UTC from IEEE Xplore. Restrictions apply.

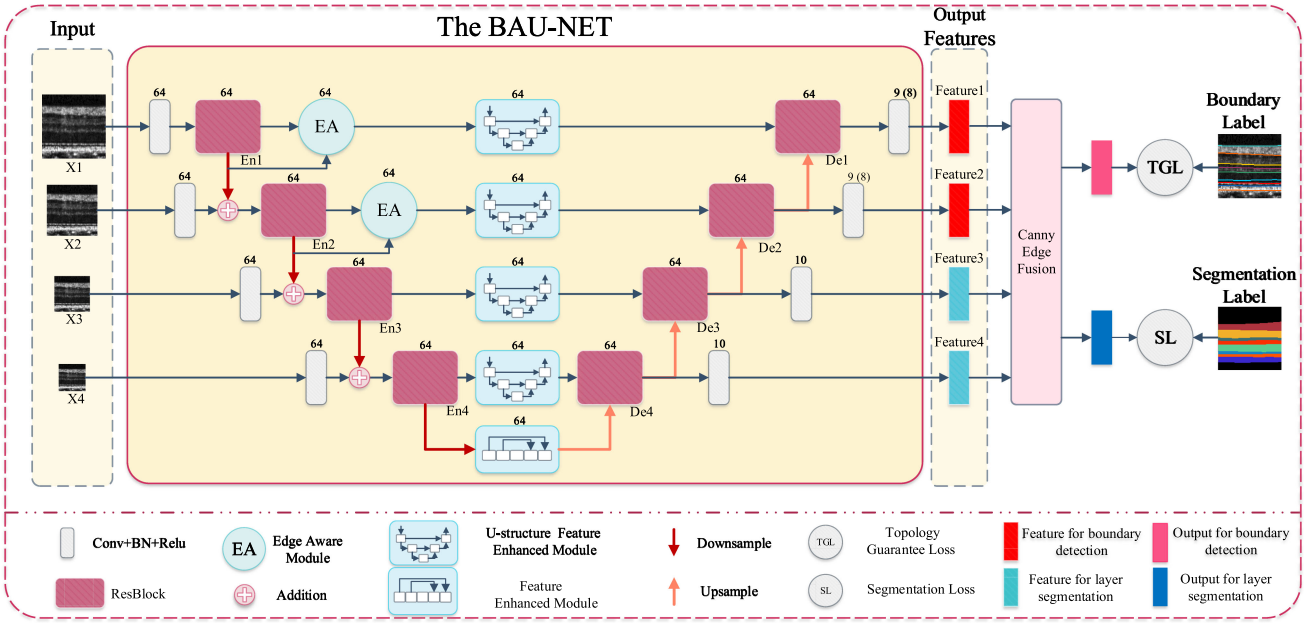

Fig. 5. The structure of the BAU-Net framework. Based on U-Net architecture, we first use the Multi-scale input to enrich the spatial information in the deep features in encoder. An edge aware (EA) module is designed in skip connection to extract pure edge features from low-level features. Then, a U-structure feature enhanced (UFE) module is designed in all skip connections to enlarge the features receptive fields from the encoder. Besides, a canny edge fusion (CEF) module is introduced on the output of the basic BAU-Net, which can fuse the priory edge information from segmentation task to boundary detection branch for a better predication. In training stage, segmentation loss and topology guarantee loss are applied to update the network weights.

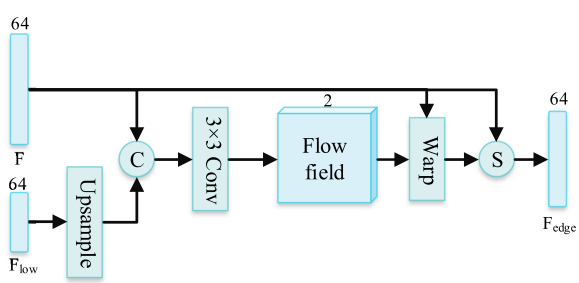

Fig. 6. Edge Aware Module. By learn a flow field from the downsamples feature maps and warp it to obtain the explicit salient part feature representation. The edge feature is generate by subtracting the salient part from the original feature.

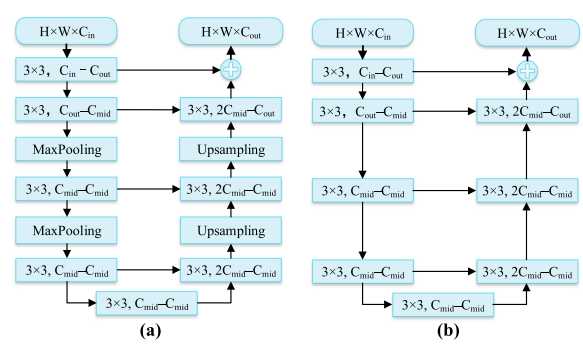

Fig. 7. U-structure Feature Enhanced module and Feature Enhanced module.

module in skip connections, which can extract and aggregate multi-scale features in the same stage. As presented in Fig. 7(a), the UFE is also an encoder-decoder structure. Given the feature maps $Z_i \in \mathbb{R}^{H\times W\times C_{in}}$ from the encoder, an intermediate map with size of $\mathbb{R}^{H\times W\times C_{out}}$ is computed by a 3×3 convolutional layer followed by batch normalization and ReLU. Then an encoder-decoder is implemented on the intermediate map to extract the multi-scale features with richer receptive fields. Finally, we add the intermediate feature and the multi-scale feature as the final output. More notably, the computation is small since the number of channel C_{middle} is much lower than C_{in} and C_{out}. Since $C_{in} = 64$ and $C_{out} = 64$, we set $C_{middle} = 16$ in this paper. Besides, we also design a Feature Enhanced module in the bottom of the encoder and decoder. As illustrated in Fig. 7(b), the structure of Feature Enhanced module is similar to UFE, but without pooling and upsampling operation, which also enlarge the receptive fields and provide more global features for a better prediction.

5) Canny Edge Fusion Module in BAU-Net: Naturally, the predicted maps of layer segmentation contain rich edge information, which may contribute to boundary detection. Besides, different scale features contain complementary context information. Based on this idea, we propose a canny edge fusion (CEF) module on the output of the basic BAU-Net, which can fuse the output features together and exploit the edge information of the segmentation task to make a better predication. We denote the output features of the basic BAU-Net as F_i, $i \in \{1, 2, 3, 4\}$. As illustrated in Fig. 8, the CEF module has three units: segmentation fusion (SF) boundary fusion (BF) and Canny. Firstly, in the SF, we concatenate the segmentation outputs of $\{F_3, F_4\}$, and apply the 1×1 convolution followed

Authorized licensed use limited to: Lanzhou University. Downloaded on December 09,2021 at 13:09:56 UTC from IEEE Xplore. Restrictions apply.

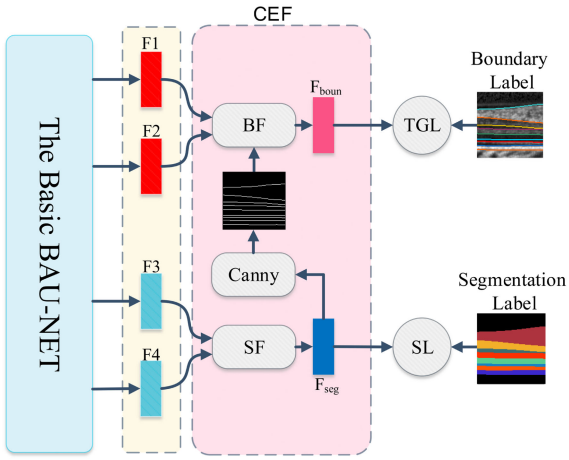

Fig. 8. Canny Edge Fusion Module. We can fuse different scales output features together and a Canny edge detector is applied to find the prominent boundaries information which then is fused to the boundary detection branch.

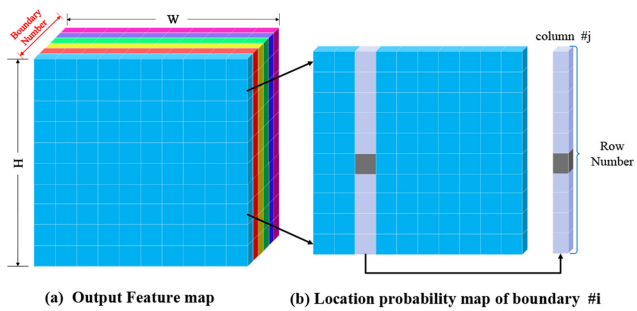

Fig. 9. The vertical coordinate distribution for the output of boundary detection. Each dimension of (a) represents a boundary and each column of (b) the distributions of the vertical coordinatefor the corresponding boundary pixel.

by softmax operation to generate the finally segmentation output F_{seg}. Secondly, in canny unit, we apply the argmax operation on the F_{seg}, which's channels are decreased to 1, and scaled it to [0,255]. Next, the Canny edge detector [42] is applied on the scaled feature followed by 1×1 convolution. Since the Canny algorithm uses a double thresholding strategy, we set the upper threshold (T2) as 100 and the lower threshold (T1) as 10, empirically. Subsequently, the canny features are mapped to a range of 0 to 1. Finally, in the BF unit, we concatenate the canny features, F_1 and F_2 followed by 3×3 convolution to get the final boundary prediction F_{boun}. The most notable thing is that the canny operation is not differentiable, so there is no gradient backpropagation in canny unit during the training process, which can reduce the excessive influence between different supervision. With this module, BAU-Net is able to fuse different level features and find the prominent boundaries information for the better boundary detection.

B. Loss Function of BAU-Net

For a patch of B-scan image X, $X \in \mathbb{R}^{H \times W}$, where W is the number of A-scan and H is the number of pixels per A-scan, there two kinds corresponding groundtruth are providing: the segmentation masks G^S and the boundarys G^B. The size of G^B is $K \times W$, wherein each row vector defines the vertical coordinates of the K boundaries in order from top to bottom and each column corresponds to an A-scan. The G^S is a $H \times W$ matrix with values in $\{1, \ldots, C\}$, where C is the number of predefined layer categories. Since the proposed BAU-Net is a multi-task network, given a image X, we extract two kinds output features: $F^S \in \mathbb{R}^{C \times H \times W}$ and $F^B \in \mathbb{R}^{K \times H \times W}$. In particular, we append two supervised losses for the two kinds outputs, resceptively.

1) Segmentation Loss: We define a hybrid segmentation loss consisting of pixel-wise cross-entropy loss and multi-class Dice loss for the F^S. Mathematically, the hybrid loss is defined as:

$$L_{Seg} = -\frac{1}{H \times W} \sum_{i=1}^{H} \sum_{j=1}^{W} \sum_{c=1}^{C} \times \left(\hat{G^S}_{i,j,c} \log F^S_{i,j,c} + \frac{2\hat{G^S}_{i,j,c} F^S_{i,j,c}}{\hat{G^S}^2_{i,j,c} + F^{S2}_{i,j,c}} \right) \quad (1)$$

where $\hat{G^S}$ is one-hot format of G_S.

2) Topology Guarantee Loss: For an input image X, we first model a location distribution of each boundary as $f_k = \prod_{j=1}^{W} a_j(x_j|I;\Theta)$, $k \in (1, \ldots, K)$, where Θ is the parameters of the BAU-Net and the network generates W independent A-scan distributions $a_j(x_j|I;\Theta)$ at each column for image I. As illustrated in Fig. 9 (a), the output feature F^B has size of $H \times W \times K$. Each channel in the feature map represented the location probability of one boundary. Each column in the map represents the distributions of the vertical coordinate for the corresponding boundary pixel of that A-scan segment in the axial direction. The network parameters Θ are trained by minimizing the K-L divergence between the distribution of F^B and G^B. But we must make the row vector in F^B become a probability map by transform each column of the row vector to become a distribution of H-dimensional one-hot vector where only the vertical position with the corresponding column has the value of 1, represented by $\hat{G^B} \in \mathbb{R}^{H \times W \times L}$. The optimization of our formulation corresponds to:

$$\begin{aligned} L_{KL} &= \sum_{k=1}^{K} \sum_{j=1}^{W} KL(\hat{G^B}_{k,j} || F^B_{k,j}) \\ &= \sum_{k=1}^{K} \sum_{j=1}^{W} \sum_{i=1}^{H} \hat{G^B}_{k,i,j} \log \frac{\hat{G^B}_{k,i,j}}{F^B_{k,i,j}} \end{aligned} \quad (2)$$

There two drawbacks of the proposed boundary location distribution: one is that the different boundaries from the same images are independent without any topological constraints; the other is that the adjacent coordinate distributions are also independent of each other in the A-scan dimension in the same boundary, which betrays our common sense: there is a horizontal dependency on adjacent pixels at specific distances. Therefore, we define more loss functions to constrain the final output.

Firstly, we can calculate the predicted location matrix of the boundary $P \in \mathbb{R}^{K \times W}$ from the column of F^B. The intuitive

Authorized licensed use limited to: Lanzhou University. Downloaded on December 09,2021 at 13:09:56 UTC from IEEE Xplore. Restrictions apply.

idea is finding the maximum response peak at each column. For any boundary index k and column index j, the predicted vertical coordinate of $P_{k,j}$, can be represented as:

$$P_{k,j} = \arg\max_{i} F_{k,i,j}, st.i \in (1, \ldots, W) \tag{3}$$

However, the argmax function is not differentiable and can not be used with further constraints. To solve this problem, we propose to use the soft-argmax operation as an approximation of location.

$$Prob_{k,:,j} = Softmax(F_{k,1:W,j}) \tag{4}$$

$$P_{k,j} = \sum_{i=1}^{H} i \times Prob_{k,i,j} \tag{5}$$

The benefits of this localization method are twofold. The first one is that the expectation function is differentiable. The other is that this operation recovers the continuous location with the discrete random varia.

For the matirx $P \in \mathbb{R}^{K \times W}$, each column corresponds to an A-scan and the values of P in its rows from $1 \rightarrow K$ define the vertical coordinate of the K boundaries in order from top to bottom. There exist a strict anatomical order in the retinal layer boundaries. For K boundaries $P_1, P_2, \ldots, P_K$ from inner to outer retina, the anatomy requires $P_1(j) \leq P_2(j) \leq \cdots \leq P_L(j)$, $j = 1, 2, \ldots L$ where $P_i(j)$ is the position of the jth boundaries at the ith column. In order to satisfy the topological constraint, we update the each $P_i(j)$ from the matirx P by $P_i(j)^{new} = P_{i-1}(j) + ReLU(P_i(j) - P_{i-1}(j))$ iteratively. Then, a smooth L1 loss is applied between P^{new} and the label of boundary G^B:

$$L_{Re} = \sum_{k=1}^{K} \sum_{j=1}^{W} \left\| P_{k,j}^{new} - G_{k,j}^{B} \right\|_1 \tag{6}$$

In order to model the location relations between adjacent A-scan within a boundary, we further propose a structural loss function to encourage more structural information on the predicted boundary. The structure of a boundary can be represented by the derivative of several adjacent pixels, which is used to describe the change relationship between adjacent pixels. In this way, the structural loss function can be:

$$L_{Str} = \sum_{k=1}^{K} \left(\sum_{j=1}^{W-2} \left\| P_{k,j+2}^{new} - P_{k,j}^{new} \right\|_1 + \sum_{j=1}^{W-4} \left\| P_{k,j+4}^{new} - P_{k,j}^{new} \right\|_1 \right) \tag{7}$$

Finally, the overall topology guarantee loss can be:

$$L_{TG} = L_{KL} + L_{Re} + L_{Str} \tag{8}$$

3) The Overall Loss: During training the BAU-Net, all above loss functions together give rise to defined as follows:

$$L_{All} = \lambda_1 L_{TG} + \lambda_2 L_{Seg} \tag{9}$$

where λ_1 and λ_2 represent the weight of different terms, respectively.

V. Experimentation

A. Experimental Setting

For HCMS dataset in our experiments, following the work of [14], [16], we train on the last six of the HC volumes and the last nine of the MS volumes (735 images come form 15 volumes), and test on the remaining 20 volumes (980 images). For the DUKE DME dataset, we still divide the image into training and testing, and 8 subjects with 88 images for training and the left 2 subjects of 22 images for testing. It's remarkable that both the training set and the testing set contain images of normal and lesions. We sample 20 patches from each image with fixed stride on both datasets. Besides, for the training set, we randomly choose 20% of the patches for validating, which can determine the best model in the training stage.

The hyperparameters of the overall loss function are set as $\lambda_1 = 1$ and $\lambda_2 = 2$ empirically. For our experiments, we use the adam optimizer [43] with default parameters to update the model weights. The batch size is 10. Besides, we set the learning rate of 0.001, 0.0005, 0.0001 on epochs 0, 15, and 25, respectively. The maximum number of training epochs is 50 on both datasets. The BAU-Net is implemented based on Pytorch framework and equipped with NVIDIA TITAN X GPUs.

B. Compared Methods

We compare our BAU-Net with start-of-the-art retinal layers segmentation algorithms to evaluate the performance on two datasets. For the HCMS dataset, the comparative algorithms can be divided into two categories: AURA [28] is a graph-based random forest method, U-Net [12], RelayNet [13], SR-Net [14], and BR-Net [16] are based on deep convolutional neural network. For DUKE DME, we also compare our results with some graph-based method: graph-based dynamic programming (GDP) [27] and layer-specic structured learning (LSE) [44], and some deep learning mathod: U-Net [12], RelayNet [13], BR-Net [16].

C. Evaluation Metrics

To comprehensively evaluate the performance of the predicted boundaries and retinal layer maps against the ground truth, two matrices are defined:

1) Mad: the mean absolute difference (MAD) is the mean of the absolute difference between the predicted boundaries and the annotations along each column. Besides, because the physical distance of each pixel is 3.90 μm in HCMS image and 3.87 μm in DUKE DME image, the metric is calculated with multiplying the MAD of the pixel by the corresponding physical distance.

2) Dice Score: the dice is overlap between the predicated retinal layers and the ground truth, which is defined as:

$$Dice = \frac{2|prediction \cap groundtruth|}{|prediction| + |groundtruth|} \tag{10}$$

where $|*|$ and $\cap$ denote the size and the intersection operation in the set, respectively.

Authorized licensed use limited to: Lanzhou University. Downloaded on December 09,2021 at 13:09:56 UTC from IEEE Xplore. Restrictions apply.

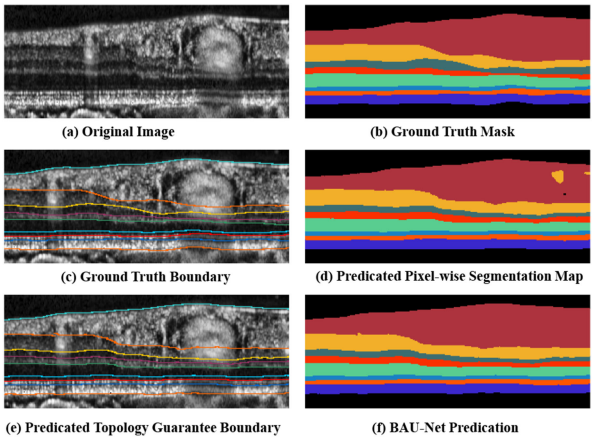

Fig. 10. The outputs images of BAU-Net from the HCMS are shown. (a) is the original image from the HCMS dataset; (b) and (c) are the annotation of segmentation mask and boundary; (d) is the output of the pixel-wise segmentation output; (e) is the output of the topology guarantee output; and (f) is the prediction of BAU-Net generated by the boundary (e).

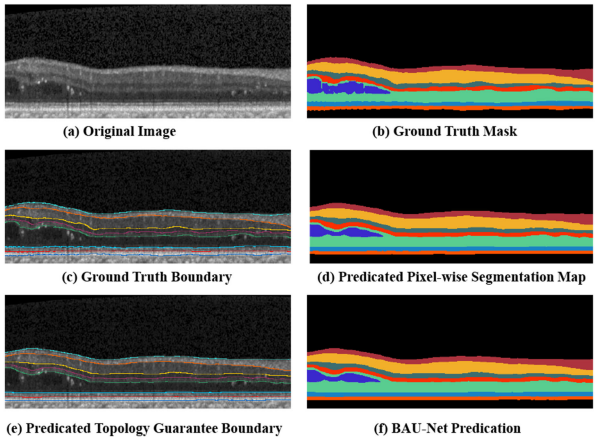

Fig. 11. The outputs images of BAU-Net from the DUKE DME are shown. (a) is the original image from the DUKE DME dataset; (b) and (c) are the annotation of segmentation mask and boundary; (d) is the output of the pixel-wise segmentation output; (e) is the output of the topology guarantee output; and (f) is the prediction of BAU-Net generated by the boundary (e) and the fuild result from (d).

VI. Results and Discussion

In this section, we present the results of experiments for analyzing and validating the performance of the proposed BAU-Net. First, we present the overall performance of the model and show the performance on abnormal images. Then, we compare BAU-Net with some other methods as illustrated in Section V. Finally, we conduct the ablation experiments on the BAU-Net to prove the effectiveness of module designs.

A. Overall Performance

Our proposed BAU-Net has two outputs: pixel-wise segmentation map and topology guarantee boundary. It is worth noting that our final retinal layer result is generated by the boundary result, that is, pixel categories are assigned between adjacent boundaries in a topological order. In order to clearly explain the outputs extracted from the proposed method, we visualize all the results of one B-scan image of two datasets. For HCMS dataset, examples of nine boundaries automatically segmented using the proposed scheme are shown in Fig. 10. We can see the result of pixel-wise seg-mentation map is topologically incorrect results but the output of boundary satisfies the ordering constraint. Thus, we only consider the topology guarantee boundaries to build the final retinal layers result that is guaranteed to have the correct topology, presented in Fig. 10(f). Fig. 11 presents an example results of our BAU-Net for DUKE DME dataset. In addition to using boundaries to generate topology guarantee layers, we also embed the result of fluid from pixel-wise segmentation maps to the topology guarantee layers for the final predication. Specifically, we first assign layer categories between adjacent boundaries in topological order to generate a retinal layer map. Then, we obtain the position of the lesion pixel in the predicated pixel-wise segmentation map, and assign the lesion category to the same position in the retinal layer map. From this two examples, we can see that the BAU-Net can output the retinal layers that is accurate and meets the topological structure.

TABLE I
Dice Results (%) of Normal and Pathological Images on HCMS Dataset

Image	Eight layers of OCT images in the HCMS dataset							
	RNFL	GCIP	INL	OPL	ONL	IS	OS	RPE
Normal	94.5	96.0	88.8	89.8	95.0	89.1	89.4	92.7
Pathologcial	93.8	95.0	88.3	89.1	94.8	87.6	88.7	92.0
All Images	94.1	95.4	88.5	89.4	94.9	88.2	89.0	92.3

TABLE II
Dice Results (%) of Normal and Pathological Images on DUKE DME Dataset

Image	Seven layers of OCT images in the DUKE DME dataset						
	RNFL	GCIP	INL	OPL	ONL	IS	OS-RPE
Normal	91.4	89.6	80.0	86.7	91.8	88.9	88.6
Pathological	85.4	95.6	82.6	80.8	90.1	89.6	91.8
All Images	87.3	93.7	81.8	82.6	90.6	89.4	90.8

Since the presence of fluid region or other lesion causes significant changes in retinal morphology, layers segmentation on pathological images is a difficult task. To analysis of the performance of BAU-Net in pathological cases, we divide the testing set into two groups: normal and pathological according to their expert diagnosis. Table I and Table II illustrate the comparison performance of two groups on HCMS and DUKE DME datasets. Although pathological cases tend to have distorted shapes, obscure boundaries, the results show that the dice score of HCMS for pathological images only a tiny percentage gap lower than the normal. Besides, for DUKE DME dataset, the dice score of some layers in pathological images are even slightly higher than those of the normal case. The comparison performance illustrates that our BAU-Net performs high robustness on the pathological images in retinal layers segmentation.

In order to measure the consistency between our method and the doctor's annotations, we conducted the Consistency of Diagnostic Test. Since retinal layers segmentation is a classification task of assigning each pixel with a particular class label,

Authorized licensed use limited to: Lanzhou University. Downloaded on December 09,2021 at 13:09:56 UTC from IEEE Xplore. Restrictions apply.

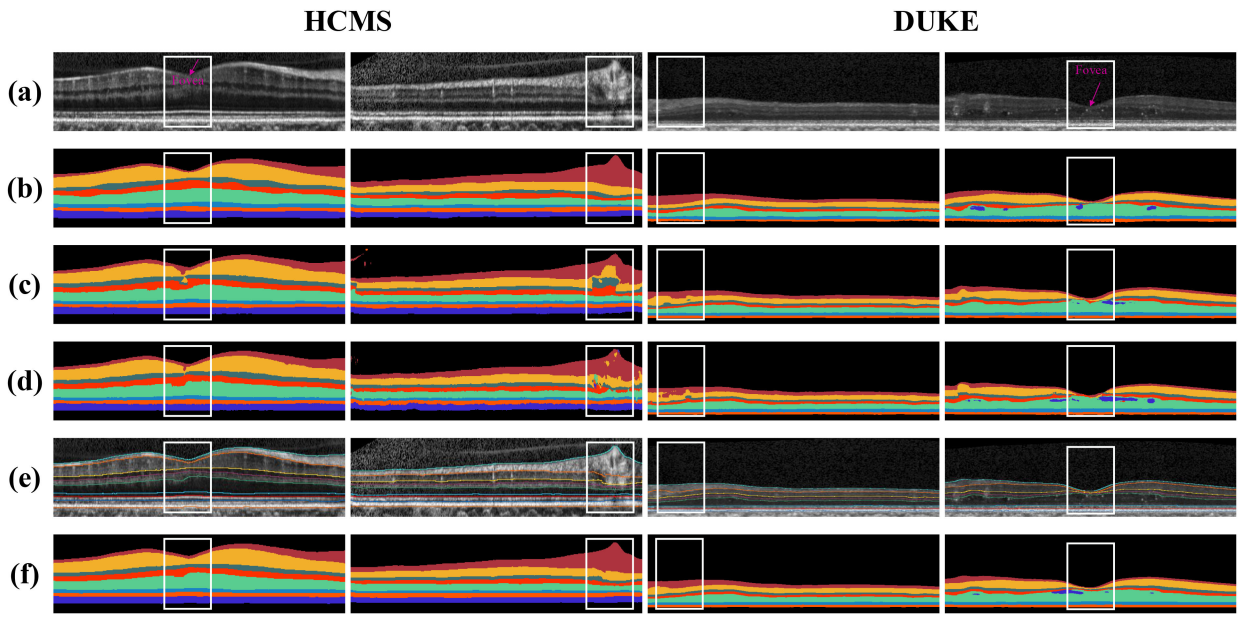

Fig. 12. The results of different methods on the test image on two datasets. (a) The original OCT image. (b) The groundtruth. (c) The U-Net method. (d) The RelayNet method. (e) The boundary result of the proposed method. (f) The retinal layer generate by the boundary result of BAU-Net.

the Cohen's kappa coefficient [45] is adapted to evaluate this consistency. The kappa value is between 0 and 1. The larger the value, the higher the consistency of the two methods. In this experiment, the kappa value we obtained on the two data sets are: HCMS (0.9447) and DUKE DME (0.9334), which proves that our BAU-Net has high consistency with the ground truth.

B. Comparison With the Existing Methods

The results of the proposed BAU-Net and other comparative methods of all nine layers in HCMS and eight layers combined with fluid in DUKE DME are presented in Fig. 12. The first two columns show the segmentation results of foveal image and non-foveal image on HCMS. In the foveal area, the RNFL layer will become narrower, which can easily lead to poor results. Both U-Net and RelayNet have a distortion in this area, and the topology of the boundary has also changed. However, the boundaries of IML and RNFL-GCL near RNFL are still very discriminative in the boundary output obtained by our method, so the final generated retinal layer segmentation map maintains the correct topological structure. In the non-foveal image, the original picture has huge noise and shadows at the marker region, and it has a poor impact in U-Net and RelayNet, which leads to very poor segmentation results. But the boundary output of BAU-net still has a good topology, so the generated retinal layer segmentation map also maintains perfect results. The last two columns of Fig. 12 show the results of the DUKE DME dataset, and also include the foveal and non-foveal image. In non-foveal result, the marker position in the original image has blurred boundary. Therefore, there are topological distortions in the results of U-Net and RelayNet. In foveal result, the first few retinal layers are also very thin, and our method still obtains topologically guaranteed results. All in all, we can see that our method can make better layer segmentation results than other methods.

To quantitate the performance of our BAU-Net compared with the comparative methods, we conducted experiments and evaluated it on the two metrics defined in Section V.C. Table III shows the MAD between the predicted boundary and manual delineation on HCMS dataset which measures the similarity of them along the column. As shown in Table III, the BAU-Net achieves seven best performance in the retinal tissue boundaries of Vitreous-RNFL, RNFL-GCL, IPL-INL, ELM, IS-OS, OS-RPE and RPE-Choroid. The overall average error of all tissue boundaries for our BAU-Net is 2.63 which reduces 0.20 compared with the previous best result 2.83. Table IV presents the results of the mean absolute difference on the DUKE DME dataset. The BAU-Net achieves the seven best results which are Vitreous-RNFL, RNFL-GCL, IPL-INL, INL-OPL,OPL-ONL, IS-OS and RPE-Choroid and the overall average error of all tissue boundaries of the BAU-Net is 4.12, which reduces 1.55 compared to the best 5.67 of the previous methods.

Table V shows the results of dice score of all layers on HCMS dataset. The BAU-Net makes the best performance in six layers, and only the results of OPL and ONL are worse than best result. Table VI presents the dice results on the DUKE DME dataset, where our model also achieves the best performance in most of the tissue layers comparing with other comparative methods. Besides, for the Fluid class, our method achieves the best performance of 52.7, with improvements of 7.5 and 3.2 over U-Net and RelayNet, respectively. In summary, under the condition of ensuring the topology of the segmentation result, the proposed model can still achieve higher performance with better robustness on two datasets than current methods.

Authorized licensed use limited to: Lanzhou University. Downloaded on December 09,2021 at 13:09:56 UTC from IEEE Xplore. Restrictions apply.

TABLE III
MEAD ± STANDARD (%, μM) OF MAD FOR EACH TESTED METHOD COMPARED TO MANUAL DELINEATION ON THE HCMS DATASET

Methods	Nine boundaries of OCT images in the HCMS dataset									
	Vitreous-RNFL	RNFL-GCL	IPL-INL	INL-OPL	OPL-ONL	ELM	IS-OS	OS-RPE	RPE-Choroid	Overall
AURA [28]	2.37±0.36	3.09±0.64	3.43±0.53	3.25±0.48	2.96±0.55	2.96±0.44	2.07±0.81	3.77±0.94	2.89±2.18	2.95±1.04
U-Net [12]	2.76±1.65	3.45±2.03	3.26±1.69	3.33±1.59	3.18±1.76	2.65±1.41	2.07±1.38	3.57±1.66	3.14±2.28	3.04±1.79
ReLayNet [13]	2.86±1.42	3.70±2.33	3.95±3.07	4.19±3.30	3.20±1.84	2.84±1.84	2.40±1.88	3.77±1.91	3.05±2.07	3.33±2.33
BR-Net [16]	2.41±0.40	2.96±0.71	2.87±0.46	3.19±0.53	**2.72±0.61**	2.65±0.73	2.01±0.57	3.55±1.02	3.10±2.02	2.83±0.99
SR-Net [14]	2.38±0.36	3.10±0.55	2.89±0.42	**3.15±0.56**	2.76±0.59	2.65±0.66	2.10±0.75	3.81±1.17	2.71±2.27	2.95±1.10
BAU-Net	**2.21±0.32**	**2.68±0.48**	**2.82±0.40**	3.26±0.55	3.10±0.65	**2.42±0.63**	**1.80±0.51**	**3.18±0.87**	**2.21±1.76**	**2.63±0.82**

TABLE IV
MEAD ± STANDARD (%, μM) OF MAD FOR EACH TESTED METHOD COMPARED TO MANUAL DELINEATION ON THE DUKE DME DATASET

Methods	Eight boundaries of OCT images in the DUKE DME dataset								
	Vitreous-RNFL	RNFL-GCL	IPL-INL	INL-OPL	OPL-ONL	ELM	IS-OS	RPE-Choroid	Overall
GDP [27]	4.25±(-)	15.33±(-)	22.97±(-)	21.66±(-)	20.55±(-)	4.35±(-)	4.03±(-)	5.22±(-)	12.19±(-)
LSE-GDP [44]	3.75±(-)	6.29±(-)	6.58±(-)	6.59±(-)	8.31±(-)	3.34±(-)	4.20±(-)	6.28±(-)	5.67±(-)
U-Net [12]	6.23±11.87	7.71±8.27	10.77±12.98	14.11±16.24	19.97±18.32	**3.13±0.75**	4.20±1.57	3,23±0.81	8.67±12.17
ReLayNet [13]	5.95±6.37	12.08±14.06	17.11±19.38	19.16±19.25	21.34±16.36	3.45±0.72	4.80±1.91	3.21±0.75	10.89±14.11
BR-Net [16]	4.51±1.18	6.71±2.79	8.29±6.63	10.71±11.38	9.88±7.07	4.41±1.87	4.52±1.60	4.61±1.82	6.70±2.62
BAU-Net	**3.47±0.79**	**4.77±1.53**	**3.92±0.88**	**4.95±1.26**	**5.94±2.17**	3.15±0.93	**3.57±1.07**	**3.16±0.95**	**4.12±1.55**

TABLE V
DICE SCORES (%) FOR EACH TESTED METHOD COMPARED TO MANUAL DELINEATION ON THE HCMS DATASET

Methods	RNFL	GCIP	INL	OPL	ONL	IS	OS	RPE
AURA [28]	93.8	94.5	87.4	89.9	94.7	87.4	87.5	91.6
U-Net [12]	92.9	94.4	87.8	89.8	94.7	87.7	87.8	91.6
ReLayNet [13]	92.5	94.3	87.5	89.7	94.6	86.9	87.6	91.5
BR-Net [16]	93.9	95.0	87.9	**90.3**	94.8	86.8	87.3	91.3
SR-Net [14]	93.7	95.1	88.3	90.2	95.0	87.2	87.2	90.7
BAU-Net	**94.1**	**95.4**	**88.5**	89.4	94.9	**88.2**	**89.0**	**92.3**

TABLE VI
DICE SCORES (%) FOR EACH TESTED METHOD COMPARED TO MANUAL DELINEATION ON THE DUKE DME DATASET

Methods	RNFL	GCIP	INL	OPL	ONL	IS	OS-RPE	Fulid
GDP [27]	77.8	77.2	65.2	67.0	86.8	87.8	82.3	-
LSE-GDP [44]	87.0	90.8	80.5	77.2	94.2	88.0	86.2	-
U-Net [12]	85.5	91.4	79.8	80.9	89.0	86.5	90.7	45.2
ReLayNet [13]	86.3	91.5	78.7	76.2	90.5	82.3	90.4	49.5
BRUNet [16]	84.3	86.0	72.7	71.5	81.1	82.3	79.2	-
BAU-Net	**87.3**	**93.7**	**81.8**	**82.6**	90.6	**89.4**	**90.8**	**52.7**

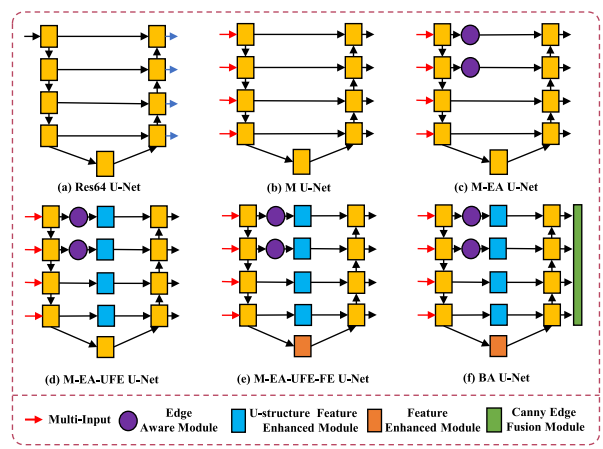

Fig. 13. Simplified illustration of different models tested in our experiments.

TABLE VII
DICE SCORES (%) FOR EACH ABLATION MODEL COMPARED TO MANUAL DELINEATION ON THE HCMS DATASET

Methods	RNFL	GCIP	INL	OPL	ONL	IS	OS	RPE	Overall
Res64 U-Net	92.0	93.5	86.3	89.1	94.3	84.4	86.8	90.6	89.6
M U-Ne	91.9	94.0	86.5	89.3	94.1	85.2	86.9	91.9	89.9
M-EA U-Net	92.2	94.3	87.3	89.9	94.8	87.0	87.4	91.2	90.6
M-EA-UFE Net	93.8	94.5	87.9	**90.5**	94.1	87.3	**89.2**	91.4	91.1
M-EA-UFE-FE Net	93.7	94.7	87.7	**90.5**	94.5	87.5	89.1	91.7	91.2
BAU-Net	**94.1**	**95.4**	**88.5**	89.4	**94.9**	**88.2**	89.0	**92.3**	**91.5**

C. Ablation Analysis

Comparing with the basic U-Net, BAU-Net mainly has the following improvements: (1) using the multi-scale input to enrich the spatial information; (2) adding edge-aware (EA) module to extract pure edge features; (3) designing U-structure feature enhanced (UFE) and feature enhanced (FE) module to enlarge the receptive fields; (4) adding canny edge fusion (CEF) module to transfer rich edge information. To analyze the effects of these module designs, we conduct experiments on six different architectures, namely Res64 U-Net, M U-Net, M-EA U-Net, M-EA-UFE U-Net, M-EA-UFE-FE U-Net and BAU-Net, which are presented in Fig. 13. Then, we evaluate all the models on HCMS test data and the dice result is shown in Table VII and Fig. 14.

The results in Table VII show that (1) BAU-Net performs better than other improved U-Nets in dice score on HCMS dataset, which indicates the effectiveness of module designed; (2) the performance with the U-Net having CEF is better than non-CEF models, which indicates that CEF structure can exploit the edge information and fuse it to the boundary detection branch for better predication; (3) skip connection with EA can extract more edge features and UFE-FE module increase the receptive field, which both make performance improvement.

Fig. 14 visualizes the segmentation results of different ablation models. We can see that our method achieves the smoother result, and the visualization result of our model is better than any of the ablation models. Specifically, the baseline model can also obtain topologically ordered segmentation results due to

Authorized licensed use limited to: Lanzhou University. Downloaded on December 09,2021 at 13:09:56 UTC from IEEE Xplore. Restrictions apply.

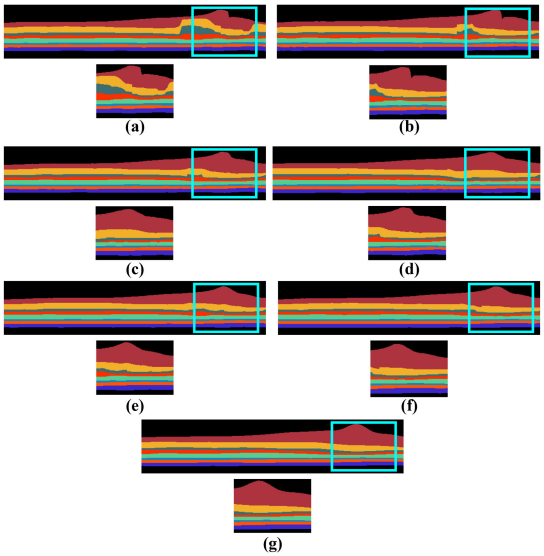

Fig. 14. Test results with different ablations of our model on HCMS dataset. (a) the baseline of Res64 U-Net; (b) M U-Net; (c) M-EA U-Net; (d) M-EA-UFE U-Net; (e) M-EA-UFE-FE U-Net; (f) BAU-Net. (g) groundtruth.

TABLE VIII
DICE SCORES (%) FOR TWO MODELS COMPARED TO MANUAL DELINEATION ON THE HCMS DATASET

Methods	RNFL	GCIP	INL	OPL	ONL	IS	OS	RPE	Overall
Dual tasks framework	**93.7**	**94.7**	**87.7**	90.5	94.5	**87.5**	**89.1**	**91.7**	**91.2**
One task framework	93.0	94.5	87.3	**90.7**	**95.1**	86.3	86.8	90.6	90.6

the use of topological guarantee loss, but the segmentation map has mutation in the local retina. From the comparison between module ablations and ours, with the help of multi-scale input, our model obtains richer spatial information in encoder, and the EA module provides more boundary information to low-level network. Meanwhile, the UFE and FE modules enlarge the receptive fields of the features, which enables the BAU-Net to capture more global context information. In addition, the CEF module provides the edge prior of the segmentation task to boundary detection branch. This further demonstrates that all improvements effectively enhance the final segmentation performance.

In this study, we are also interested in investigating the performance of the proposed dual tasks framework of using high level output for segmentation and low level output for boundary detection. For this purpose, we replace the two high level segmentation outputs of the basic BAU-Net without CEF for boundary detection and performed a comparing experiment on HCMS dataset to evaluate the performance based on dice metric. The results are shown in Table VIII. It can be seen that the basic BAU-Net with dual tasks framework has a better segmentation accuracy compared with the model of all outputs for boundary detection (named one task framework). It may be because dual tasks framework can obtain more discriminating features in high level network. Besides, the dual tasks framework can not only perform boundary detection, but also provide segmentation mask of the lesion.

D. Limitation

Although the BAU-Net has low trainable parameters with size of 1.27 M and achieves best retinal layers segmentation results, it has high computational complexity. In our experiments, the average inference time for a testing image of the HCMS and DUKE DME datasets are 6.40 s, and 6.43 s, respectively. The main reason is that the steps of image reconstruction from patch results need to consume a lot of memory and operations, where we must first use the patch boundary result to reconstruct the boundary result of entire image, and then assign pixel categories between each adjacent boundary to generate the final retinal layers segmentation map.

VII. CONCLUSION AND OUTLOOK

In this paper, we proposed a novel end-to-end BAU-Net for retinal layers segmentation. Motivated by a general knowledge that the deep features contain more context information and low-level features possess rich context information, we used the shallow networks for boundary detection and deep networks for layers segmentation. Specifically, we first use the Multi-scale input to enrich the spatial information in the deep features of encoder. For low-level features from encoder, an edge aware (EA) module is designed in skip connection to extract pure edge features. Then, a U-structure feature enhanced (UFE) module is designed in all skip connections to enlarge the receptive fields of the multi-scale features from the encoder. Besides, a canny edge fusion (CEF) module is introduced to aforementioned architecture, which can fuse the priory edge information from segmentation task to boundary detection branch for a better prediction. Furthermore, we introduce a new topology guarantee loss, which can guarantee the topological order for all boundaries by a fully convolutional operation. Experimental evaluations on two OCT datasets demonstrate the effectiveness of the proposed approach. For future work, we plan to extend our BAU-Net to support 3D OCT segmentation tasks.

REFERENCES

[1] W. Drexler and J. G. Fujimoto, "State-of-the-art retinal optical coherence tomography," *Prog. Retinal Eye Res.*, vol. 27, no. 1, pp. 45–88, 2008.
[2] P. A. Keane *et al.*, "Evaluation of optical coherence tomography retinal thickness parameters for use in clinical trials for neovascular age-related macular degeneration," *Invest. Ophthalmol. Vis. Sci.*, vol. 50, no. 7, pp. 3378–3385, 2009.
[3] S. Saidha *et al.*, "Visual dysfunction in multiple sclerosis correlates better with optical coherence tomography derived estimates of macular ganglion cell layer thickness than peripapillary retinal nerve fiber layer thickness," *Mult. Scler. J.*, vol. 17, no. 12, pp. 1449–1463, 2011.
[4] S. C. Lin *et al.*, "Optic nerve head and retinal nerve fiber layer analysis: A report by the American academy of ophthalmology," *Ophthalmology*, vol. 114, no. 10, pp. 1937–1949, 2007.
[5] E. Götzinger *et al.*, "Speckle noise reduction in high speed polarization sensitive spectral domain optical coherence tomography," *Opt. Exp.*, vol. 19, no. 15, pp. 14 568–14 584, 2011.
[6] A. F. Fercher, C. K. Hitzenberger, G. Kamp, and S. Y. El-Zaiat, "Measurement of intraocular distances by backscattering spectral interferometry," *Opt. Commun.*, vol. 117, no. 1-2, pp. 43–48, 1995.
[7] D. Xiang *et al.*, "Automatic retinal layer segmentation of OCT images with central serous retinopathy," *IEEE J. Biomed. Health Inform.*, vol. 23, no. 1, pp. 283–295, Jan. 2019.

Authorized licensed use limited to: Lanzhou University. Downloaded on December 09,2021 at 13:09:56 UTC from IEEE Xplore. Restrictions apply.

[8] S. Lee *et al.*, "Atlas-based shape analysis and classification of retinal optical coherence tomography images using the functional shape (FSHAPE) framework," *Med. Image Anal.*, vol. 35, pp. 570–581, 2017.

[9] J. Zhang, Y. Xia, Y. Xie, M. Fulham, and D. D. Feng, "Classification of medical images in the biomedical literature by jointly using deep and handcrafted visual features," *IEEE J. Biomed. Health Inform.*, vol. 22, no. 5, pp. 1521–1530, Sep. 2018.

[10] L. Fang, D. Cunefare, C. Wang, R. H. Guymer, S. Li, and S. Farsiu, "Automatic segmentation of nine retinal layer boundaries in OCT images of non-exudative AMD patients using deep learning and graph search," *Biomed. Opt. Exp.*, vol. 8, no. 5, pp. 2732–2744, 2017.

[11] J. Kugelman, D. Alonso-Caneiro, S. A. Read, S. J. Vincent, and M. J. Collins, "Automatic segmentation of oct retinal boundaries using recurrent neural networks and graph search," *Biomed. Opt. Exp.*, vol. 9, no. 11, pp. 5759–5777, 2018.

[12] O. Ronneberger, P. Fischer, and T. Brox, "U-NET: Convolutional networks for biomedical image segmentation," in *Proc. Int. Conf. Med. Image Comput. Comput.-Assist. Interv.*, 2015, pp. 234–241.

[13] A. G. Roy *et al.*, "Relaynet: Retinal layer and fluid segmentation of macular optical coherence tomography using fully convolutional networks," *Biomed. Opt. Exp.*, vol. 8, no. 8, pp. 3627–3642, 2017.

[14] Y. He *et al.*, "Deep learning based topology guaranteed surface and MME segmentation of multiple sclerosis subjects from retinal OCT," *Biomed. Opt. Exp.*, vol. 10, no. 10, pp. 5042–5058, 2019.

[15] R. Wang, S. Chen, C. Ji, J. Fan, and Y. Li, "Boundary-aware context neural network for medical image segmentation," 2020, *arXiv:2005.00966*.

[16] Y. He *et al.*, "Fully convolutional boundary regression for retina oct segmentation," in *Proc. Int. Conf. Med. Image Comput. Comput.-Assist. Interv.*, 2019, pp. 120–128.

[17] H. Ishikawa, D. M. Stein, G. Wollstein, S. Beaton, J. G. Fujimoto, and J. S. Schuman, "Macular segmentation with optical coherence tomography," *Invest. Ophthalmol. Vis. Sci.*, vol. 46, no. 6, pp. 2012–2017, 2005.

[18] M. Shahidi, Z. Wang, and R. Zelkha, "Quantitative thickness measurement of retinal layers imaged by optical coherence tomography," *Amer. J. Ophthalmol.*, vol. 139, no. 6, pp. 1056–1061, 2005.

[19] A. Yazdanpanah, G. Hamarneh, B. Smith, and M. Sarunic, "Intra-retinal layer segmentation in optical coherence tomography using an active contour approach," in *Proc. Int. Conf. Med. Image Comput. Comput.-Assist. Interv.*, 2009, pp. 649–656.

[20] J. Novosel, G. Thepass, H. G. Lemij, J. F. de Boer, K. A. Vermeer, and L. J. van Vliet, "Loosely coupled level sets for simultaneous 3D retinal layer segmentation in optical coherence tomography," *Med. Image Anal.*, vol. 26, no. 1, pp. 146–158, 2015.

[21] F. Rossant, I. Bloch, I. Ghorbel, and M. Paques, "Parallel double snakes. application to the segmentation of retinal layers in 2D-OCT for pathological subjects," *Pattern Recognit.*, vol. 48, no. 12, pp. 3857–3870, 2015.

[22] D. Kaba *et al.*, "Retina layer segmentation using kernel graph cuts and continuous max-flow," *Opt. Exp.*, vol. 23, no. 6, pp. 7366–7384, 2015.

[23] B. J. Antony *et al.*, "Simultaneous segmentation of retinal surfaces and microcystic macular edema in SDOCT volumes," in *Proc. Med. Imag. Image Process. Int. Soc. Opt. Photon.*, vol. 9784, 2016, Art. no. 97841C.

[24] B. I. Dodo, Y. Li, K. Eltayef, and X. Liu, "Graph-cut segmentation of retinal layers from OCT images," *Proc. 11th Int. Joint Conf. Biomed. Eng. Syst. Technol.*, 2018, pp. 35–42.

[25] K. Vermeer, J. Van der Schoot, H. Lemij, and J. De Boer, "Automated segmentation by pixel classification of retinal layers in ophthalmic oct images," *Biomed. Opt. Exp.*, vol. 2, no. 6, pp. 1743–1756, 2011.

[26] A. Carass, A. Lang, M. Hauser, P. A. Calabresi, H. S. Ying, and J. L. Prince, "Multiple-object geometric deformable model for segmentation of macular OCT," *Biomed. Opt. Exp.*, vol. 5, no. 4, pp. 1062–1074, 2014.

[27] 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. Exp.*, vol. 18, no. 18, pp. 19 413–19 428, 2010.

[28] A. Lang *et al.*, "Retinal layer segmentation of macular OCT images using boundary classification," *Biomed. Opt. Exp.*, vol. 4, no. 7, pp. 1133–1152, 2013.

[29] K. Hu, B. Shen, Y. Zhang, C. Cao, F. Xiao, and X. Gao, "Automatic segmentation of retinal layer boundaries in OCT images using multi-scale convolutional neural network and graph search," *Neurocomputing*, vol. 365, pp. 302–313, 2019.

[30] J. Long, E. Shelhamer, and T. Darrell, "Fully convolutional networks for semantic segmentation," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, 2015, pp. 3431–3440.

[31] T. G. P. Reddy *et al.*, "Retinal-layer segmentation using dilated convolutions," in *Proc. 3rd Int. Conf. Comput. Vis. Image Process.*, 2020, pp. 279–292.

[32] Q. Li *et al.*, "Deepretina: Layer segmentation of retina in OCT images using deep learning," *Transl. Vis. Sci. Technol.*, vol. 9, no. 2, pp. 61–61, 2020.

[33] L. Ngo, J. Cha, and J.-H. Han, "Deep neural network regression for automated retinal layer segmentation in optical coherence tomography images," *IEEE Trans. Image Process.*, vol. 29, pp. 303–312, 2020.

[34] B. Anoop, R. Pavan, G. Girish, A. R. Kothari, and J. Rajan, "Stack generalized deep ensemble learning for retinal layer segmentation in optical coherence tomography images," *Biocybern. Biomed. Eng.*, vol. 40, no. 4, pp. 1343–1358, 2020.

[35] H. Raja *et al.*, "Extraction of retinal layers through convolution neural network (CNN) in an OCT image for glaucoma diagnosis," *J. Digit. Imag.*, vol. 33, no. 6, pp. 1428–1442, 2020.

[36] 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, 2019.

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

[38] K. He, X. Zhang, S. Ren, and J. Sun, "Deep residual learning for image recognition," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, 2016, pp. 770–778.

[39] T. Zhao and X. Wu, "Pyramid feature attention network for saliency detection," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, 2019, pp. 3085–3094.

[40] M. Jaderberg *et al.* "Spatial transformer networks," in *Proc. Adv. Neural Inf. Process. Syst.*, 2015, pp. 2017–2025.

[41] L.-C. Chen, Y. Zhu, G. Papandreou, F. Schroff, and H. Adam, "Encoder-decoder with atrous separable convolution for semantic image segmentation," in *Proc. Eur. Conf. Comput. Vis.*, 2018, pp. 801–818.

[42] L. Ding and A. Goshtasby, "On the canny edge detector," *Pattern Recognit.*, vol. 34, no. 3, pp. 721–725, 2001.

[43] D. P. Kingma and J. Ba, "Adam: A method for stochastic optimization," 2014, *arXiv:1412.6980*.

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

[45] S. M. Vieira, U. Kaymak, and J. M. Sousa, "Cohen's kappa coefficient as a performance measure for feature selection," in *Proc. Int. Conf. Fuzzy Syst.*, 2010, pp. 1–8.

Authorized licensed use limited to: Lanzhou University. Downloaded on December 09,2021 at 13:09:56 UTC from IEEE Xplore. Restrictions apply.