IEEE Access
Multidisciplinary ⋮ Rapid Review ⋮ Open Access Journal

Received 27 September 2023, accepted 30 October 2023, date of publication 6 November 2023,
date of current version 13 November 2023.

Digital Object Identifier 10.1109/ACCESS.2023.3330493

RESEARCH ARTICLE

Multi-Task Dual Boundary Aware Network for Retinal Layer Segmentation

CE YANG[1], WENYU WANG[1], CHENGYU WU[1], KAI JIN[2], YAN YAN[2], JUAN YE[2], AND SHUAI WANG[1,3,4]

[1]Department of Mechanical, Electrical and Information Engineering, Shandong University, Weihai 264209, China
[2]Department of Ophthalmology, The Second Affiliated Hospital, Zhejiang University School of Medicine, Hangzhou 310018, China
[3]School of Cyberspace, Hangzhou Dianzi University, Hangzhou 310018, China
[4]Suzhou Research Institute, Shandong University, Suzhou 215123, China

Corresponding author: Shuai Wang (shuaiwang.tai@gmail.com)

This work was supported in part by the Zhejiang Provincial Natural Science Foundation of China under Grant LDT23F01015F01, in part by the National Natural Science Foundation of China under Grant 62201323, and in part by the Natural Science Foundation of Jiangsu Province under Grant BK20220266.

ABSTRACT Layer segmentation of Optical Coherence Tomography (OCT) images is an important step in diagnosing retinal diseases. However, the presence of some artifacts and noise in OCT images often leads to unsatisfactory layer segmentation results. Especially when the number of layers to be segmented is particularly large, the boundaries between layers are indistinguishable, which poses a great challenge to automatic and accurate segmentation. To solve these problems, we propose a novel multi-task dual boundary-aware network to improve the retinal layer segmentation performance in OCT images. Specifically, based on the hierarchical relationship between retinal layers, we design a dual boundary representation method to encode the bidirectional boundary information between layers. Then we design a multi-task architecture and a novel consistency loss to utilize the boundary representation to make the segmentation more accurate. For evaluation, we have built a large-scale OCT layer segmentation dataset with 1,200 images. The comprehensive experimental results show that our method achieves superior performance over other state-of-the-art algorithms.

INDEX TERMS Boundary representation, consistency constraint, multi-task network, optical coherence tomography, retinal layer segmentation.

I. INTRODUCTION

Retinal diseases are the main causes of visual impairment and blindness. The survey by the World Health Organization shows that 36 million people in the world are perpetually blind, and 253 million people have disturbances of visual acuity [1]. In fact, 80% of cases of visual impairment can be prevented or cured at an early stage with an appropriate retinal screening and treatment program [2], [3]. For example, diabetic retinopathy and age-related macular degeneration can be diagnosed by examining changes in retinal layer thickness and structure [4], [5]. Because optical coherence tomography (OCT) can obtain a high-resolution cross-sectional view of the human retina without invasion, it has been widely used for retinal disease diagnosis [6]. While retinal layer segmentation can provide a very intuitive analysis of the shape and thickness of the retinal layer, it is a very critical step in the diagnosis of retinal diseases [7], [8], [9]. However, manual segmentation of retinal layers is time-consuming and subjective, which greatly reduces the efficiency of clinical diagnosis. Therefore, automated retinal segmentation techniques have been developed to efficiently and accurately carry out the task of OCT layer segmentation.

However, automated retinal layer segmentation faces the following challenges: First, there exists interference information from the background and other surrounding tissues in OCT images. For example, as shown in Fig. 1(a), blood vessels in the retinal region disrupt the continuous shape of the retinal layers. Second, the retinal layers of OCT images have low contrast and narrow width, which makes

The associate editor coordinating the review of this manuscript and approving it for publication was Kumaradevan Punithakumar.

© 2023 The Authors. This work is licensed under a Creative Commons Attribution-NonCommercial-NoDerivatives 4.0 License.
For more information, see https://creativecommons.org/licenses/by-nc-nd/4.0/

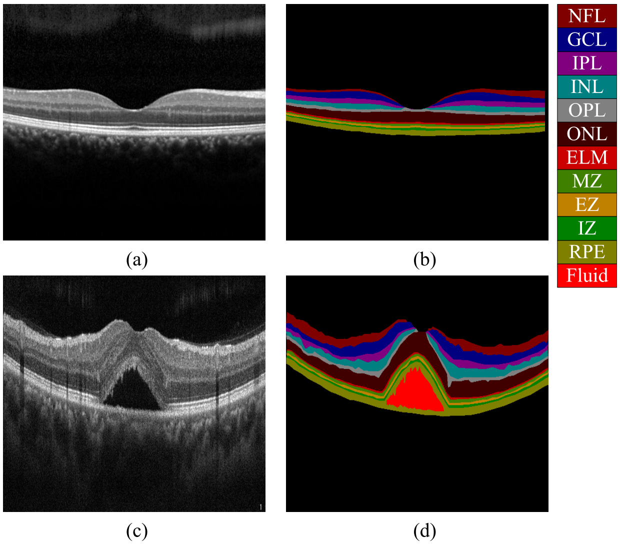


FIGURE 1. Examples of our OCT retinal layer segmentation dataset. (a) and (c) are original OCT images without and with lesion, respectively. (b) and (d) are the corresponding retinal layering masks.

it difficult to distinguish the boundaries between the retinal layers. As shown in Fig. 1(b), the layer boundary between the Ganglion Cell Layer (GCL) and Inner Plexiform Layer (IPL) is blurry and difficult to distinguish. In addition, the narrow widths of the lower retinal layers makes it a challenge for models to accurately separate distinct layers like the Myoid Zone (MZ), Ellipsoid Zone (EZ), and Interdigitation Zone (IZ). Third, the presence of the lesion disrupts the intrinsic anatomy of the retina [10]. As shown in Fig. 1(c), the presence of effusion disrupts the shape of the retinal layers, making it difficult for automatic segmentation algorithms to learn the structural features of the retina.

Currently, some deep learning based methods have been proposed to address these challenges in retinal layer segmentation, which can be divided into the following categories: The first category of methods uses convolutional neural networks (CNNs) to classify the central pixels of sliding patches [10], [11], but they are very inefficient. The second category of methods adopts fully convolutional networks (FCNs) to segment the whole image using dense prediction [12], [13], but their performance on boundaries is usually unsatisfactory. The third category of methods first uses an FCN to obtain preliminary retinal layer segmentation results and then uses a graph search algorithm to optimize the boundaries of each retinal layer [14], [15]. However, graph search is a post-processing operation that requires manual parameter setting and takes a long time to run. The fourth category of methods uses an end-to-end multi-task FCN to generate layer segmentation results and boundary regression results simultaneously, which prompts the network to focus on the boundaries between retinal layers [16], [17]. However, the boundary representations of these methods are too simplistic and do not take full advantage of the relationship between layers. The fifth category of methods is the Transformer-based networks like Swin-Unet [18] and TransUnet [19]; global features can be acquired through Transformer structure. However, Transformer-based networks always fall into heavy computation costs [20], which may cause unsatisfactory performance in small-shape object segmentation on small datasets due to overfitting.

To address these challenges, we propose a multi-task dual boundary aware network (DBA-Net) that not only segments retinal layers but also obtains rich boundary information. To better highlight the unclear retinal layer boundary, we propose a novel dual boundary representation that encodes the boundary based on the hierarchical relationship between adjacent layers. To better strengthen mutual assistance between different tasks, we designed a consistency loss to make the segmentation prediction and boundary regression prediction mutually constrained.

Our main contributions can be summarized as follows:

- A dual boundary representation method is proposed to address the indistinguishable boundaries between retinal layers. Different from conventional methods, the proposed method is based on the hierarchical spatial relationship between adjacent layers from two perspectives.
- A consistency constraint between segmentation tasks and boundary regression tasks is designed to enhance the task consistency in multi-task learning, which can effectively reduce the semantic gap in multi-task optimization.
- Comprehensive experiments have been performed on 1,200 OCT images, with results demonstrating the superiority of our method over state-of-the-art (SOTA) methods.

II. RELATED WORKS

A. RETINAL LAYER SEGMENTATION

With the continuous development and improvement of deep learning techniques, more and more methods based on them have been applied to retinal layer segmentation. Fang et al. used a CNN to classify the central pixel of sliding patches in an image as background or boundary, thus layering the retina by locating all boundary pixels [11]. Xiang et al. used a custom feature extractor and neural networks to classify each pixel point as one of seven retinal layers, background, or neovascularization [10].

However, the efficiency of using sliding windows and CNN classifiers is too low, which means that each pixel has to undergo a separate classification process. Therefore, some semantic segmentation algorithms based on FCNs are widely used for retinal layer segmentation. For example, Roy et al. proposed a variant of Unet named ReLayNet to segment the retina into 7 layers, edema, and background, which adopted the unpooling operation in the upsampling process to recover the fine-grained location information lost in the pooling operation and used a joint loss consisting of both cross-entropy loss and Dice loss to constrain the network optimization [12]. Wang et al. used the higher-level

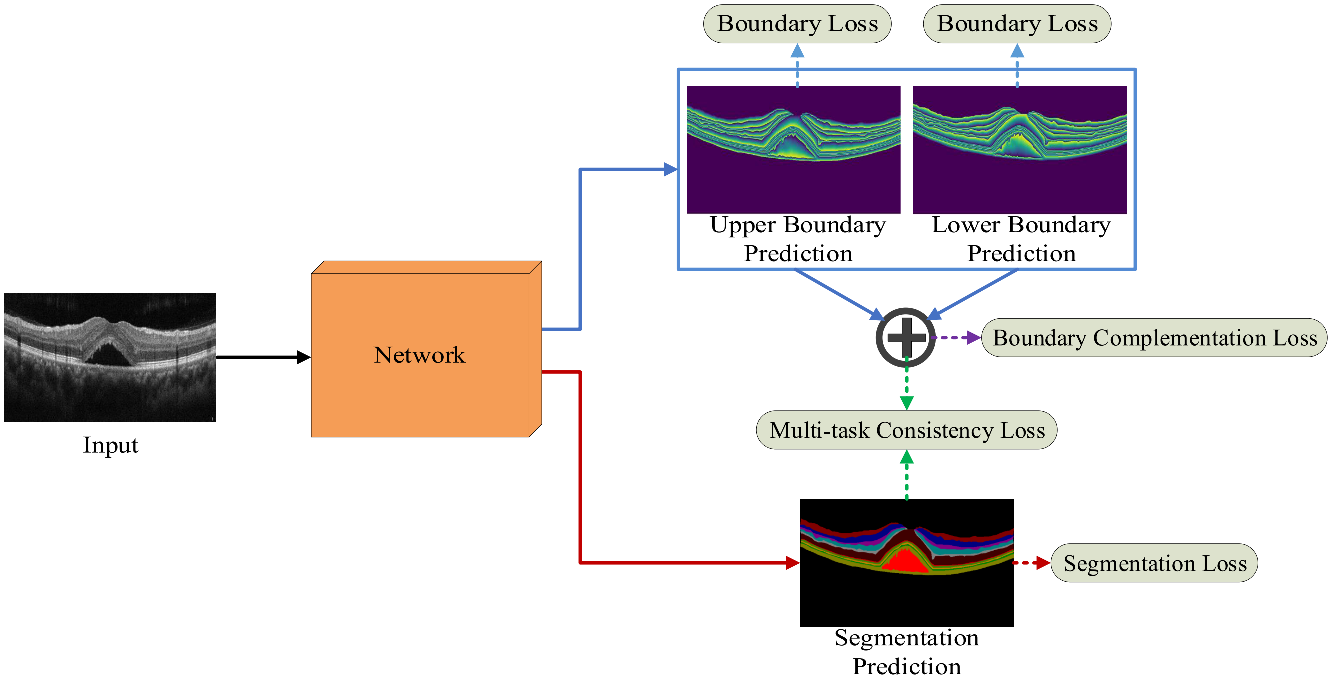

Boundary Loss
Boundary Loss
Upper Boundary Prediction
Lower Boundary Prediction
Boundary Complementation Loss
Network
Input
Multi-task Consistency Loss
Segmentation Loss
Segmentation Prediction

FIGURE 2. The overall multi-task architecture of our proposed method includes three tasks: one main segmentation task and two boundary regression tasks. In addition to the loss for each task, a complementation loss and a consistency loss are used to enhance the model capability based on the constraints between different boundary representations and between boundary representations and segmentation, respectively.

features of the encoder to generate the region segmentation results and the lower-level features of the encoder to generate the boundary segmentation results, and then the two results were combined to obtain the final segmentation results [16]. While the traditional pooling operation will lead to resolution loss, dilated convolution, and spatial pyramid pooling are adopted by some works to increase the receptive field. For example, Apostolopoulos et al. used multi-scale input and dilated convolution to compensate for the loss of resolution caused by downsampling [21]. Li et al. proposed an FCN that adopted dilated convolution layers and a modified spatial pyramid pooling layer to obtain multi-scale information to accomplish fine retinal layer segmentation [13].

One of the main disadvantages of the convolution layer is that it can only extract local features, so it cannot encode the relationship between global pixels. To address this problem, some methods based on Recurrent Neural Networks (RNN) have been proposed to extract global pixel dependencies for retinal layer segmentation [15], [22]. For example, Gopinath et al. used CNN for the layer of interest extraction and edge detection and then used Long Short Term Memory to trace continuous boundaries following edge detection [22]. Hu et al. constructed an RNN-based image feature extraction module and embedded this module in ResNet [23], which extracted global information from images in four directions to improve the segmentation performance [15]. Another approach is a Transformer-based network. This method uses multi-head self-attention in the Transformer module to build a global dependency of the feature map, which solves the problem of the local receptive field of CNN. Xue et al. propose a method for retinal layer segmentation called CTS-Net [24], which is based on the CSWin Transformer [25] architecture. The CTS-Net combines the advantages of the Transformer's global modeling capabilities with convolutional operations to achieve accurate retinal layer segmentation and smooth boundary extraction.

B. BOUNDARY AWARE SEGMENTATION

The accuracy of boundaries is very important in the image segmentation task, so some boundary-based segmentation methods have been investigated to produce accurate boundaries, which can be divided into polygon-based methods and multi-task learning-based methods.

Polygon-based methods regard the segmentation task as the coordinate regression of boundary points and then connect these points into polygons as the segmentation result [26], [27], [28]. For example, Tam et al. first used a regression network to get the coordinates of 50 boundary points and then adopted a manifold regularization to constrain the spatial correlation between boundary points [29]. Meng et al. proposed CABNet to represent the boundaries of objects with vertices and then explicitly predicted the coordinates of these vertices, which achieved good performance in optic disc (OD) and optic cup (OC) segmentation tasks [28]. Similarly, Xie et al. proposed a PolarMask that first represented segmentation polygons in the polar coordinate system and then used a CNN to predict the length of a ray at each angle [30].

Multi-task learning-based methods usually pay attention to the dependency of segmented regions and boundaries, implicitly or explicitly [31], [32]. For example, Zhang et al. proposed a network for OD and OC segmentation that used an edge guide mechanism to emphasize and highlight the segmentation boundaries [33]. Zhang et al. and Fan et al. proposed a similar idea of boundary attention, where object boundaries were implicitly extracted from region predictions through a foreground elimination mechanism [31], [32]. Meng et al. proposed cross-domain graph reasoning with regional nodes and boundary nodes to improve the interactive aggregation ability of regional features and boundary features [34]. Typically, these methods regard segmentation as a multi-task learning problem by extracting features of regions and boundaries using a shared backbone network.

Boundary-based segmentation methods have also been used in retinal layer segmentation tasks. While retinal layer segmentation predictions often have discontinuous boundaries, two kinds of methods have been proposed to address this problem. The first one usually adopted post-processing strategies such as graph search and level sets to optimize the segmentation results to generate continuous boundaries [35], [36]. For example, Kugelman et al. first trained an RNN-based network and then optimized the boundaries using a graph search algorithm [14]. The second one adopted end-to-end networks to get continuous retinal layer boundaries. For example, Ngo et al. predicted the retinal boundary pixels by feeding the image patches augmented with boundary and location information into a regression network [2]. He et al. proposed two cascaded U-structured networks, S-Net and T-Net, in which S-Net classifies each pixel and T-Net generates consecutive layers with the correct layer order based on detected retinal layer boundaries [37]. He et al. designed a multi-task architecture to do region segmentation and boundary delineation together [17].

Although post-processing algorithms such as graph search can optimize the boundaries of the segmentation results to obtain continuous boundaries, manual parameters need to be set, and the algorithm takes a long time. Multi-task networks can assist segmentation by adding boundary tasks, but their boundary modeling fails to adequately represent boundary information, and the relationships between multi-tasks are not fully exploited.

III. METHOD

Fig. 2 shows the overall multi-task architecture of our proposed method, which includes three tasks: one main segmentation task, and two boundary regression tasks. Different from the widely used boundary representation methods, we design a pair of boundary representation methods based on the spatial context between adjacent layers, and there is a strong correlation between them, namely, the sum of two values under two representation methods is always 1 for each pixel. Based on this, we design a complementation loss between the predictions of two boundary representations to enhance the boundary learning ability. Moreover, to improve mutual supervision between the segmentation task and boundary tasks, we design a consistency loss. In our implementation, the image pixels will be classified into 13 classes, including 11 retinal layers, fluid, and background. Identifying an abnormality in a specific layer of the retina can greatly assist clinicians in refining their differential diagnosis when interpreting OCT scans, which emphasizes the importance of being able to identify and distinguish the 11 retinal layers. For instance, exudates and drusen may appear similar with an ophthalmoscope or in fundus photos. However, they can be easily differentiated based on their location within the retinal layers. Exudates are typically found in or adjacent to the outer plexiform layer, as they are lipid residues originating from damaged capillaries in the inner retina. On the other hand, drusen are deposits located between the retinal pigment epithelium (RPE) and Bruch's membrane due to the malfunctioning of RPE. Next, we will give details about our network.

A. MULTI-TASK ARCHITECTURE

Fig. 3 shows the detailed architecture of our proposed method. Since any FCN can be the backbone of our method, we adopt a U-shaped architecture in our implementation due to its excellent performance in the field of medical image segmentation. Specifically, we deepened the original Unet by adding one encoder block and one decoder block to increase its receptive field. Each encoder block includes two convolution layers and a max pooling layer, while each decoder block includes one up-sampling layer and two convolution layers.

The initial shared blocks consist of five encoder blocks and two decoder blocks; the independent blocks for respective tasks consist of three decoder blocks. This shared structure has the advantage of reducing the parameters and computational effort of the network and improving the feature extraction capability of the encoder blocks.

The predictions of the conventional segmentation task are generated by the last learned feature maps through a 1×1 convolution layer and a Softmax function. The predictions of each boundary regression task are generated by the last feature map through a 1×1 convolution layer and a Sigmoid activation function. Based on our innovative definition, two boundary regression predictions can generate segmentation results by adding operations.

B. DUAL BOUNDARY REPRESENTATION

In our retinal segmentation task, we segment a retinal OCT image into 11 layers, fluid, and background. Uncertain lesion regions pose significant challenges for achieving accurate segmentation.

Many methods attempt to improve the segmentation performance with the help of boundary representation, which enhances the feature learning on the boundary by encoding the boundary. But most of them are based on simple pixel distance context to implement boundary coding [38],

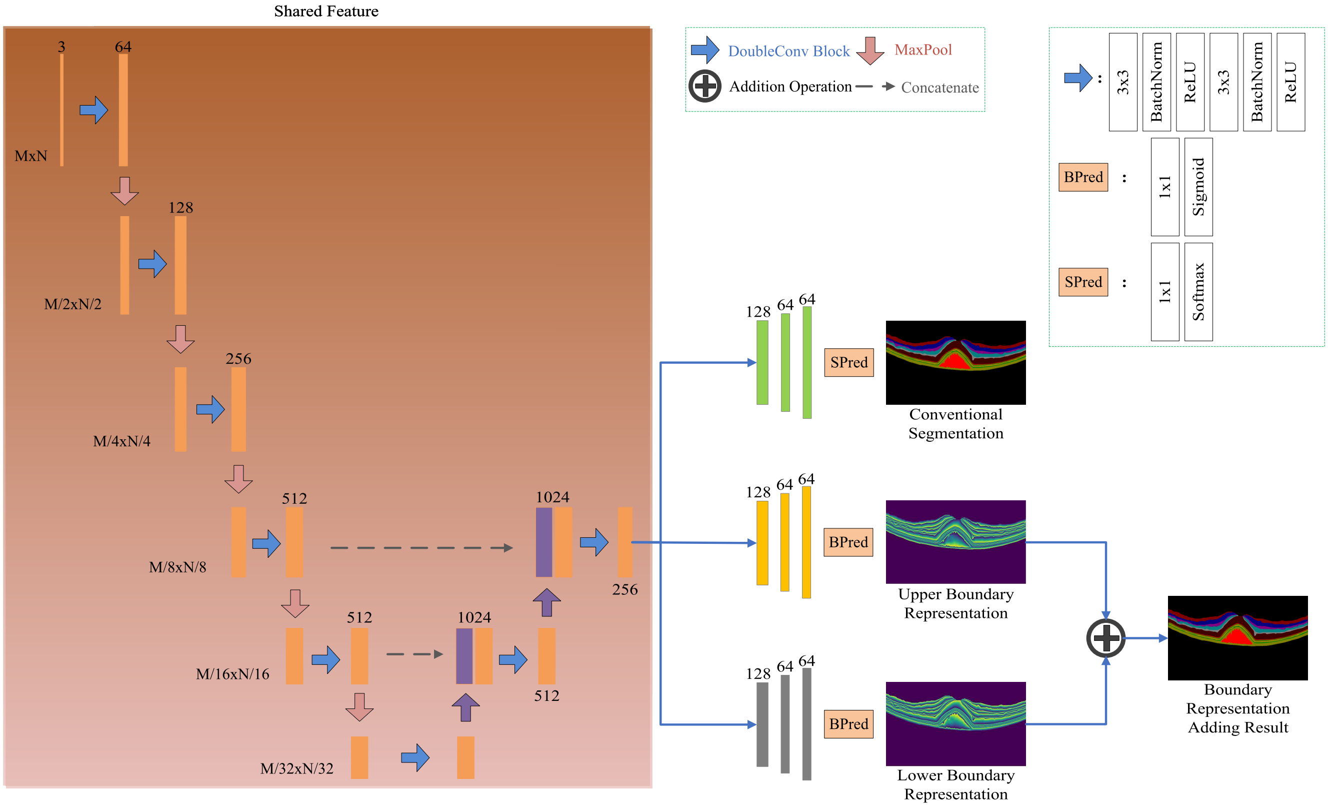


FIGURE 3. The detailed illustration of our proposed method.

[39], [40], which makes it hard to capture the hierarchical relationship between adjacent layers of retinal in OCT images. For example, the background is always above the first retinal layer. Based on this finding, we designed a dual boundary representation to highlight the indistinguishable boundary, which is defined based on the nearest distance between the pixel and its adjacent layer boundary from both the upper and lower directions. A visualization example of our dual boundary representation is shown in Fig. 4.

Our dual boundary representation consists of an upper boundary representation and a lower boundary representation. Before giving their definitions, we first define BR^{dis}_{upper} and BR^{dis}_{lower} as:

$$BR^{c}_{upper}(x) = \begin{cases} \inf\limits_{y \in B^{c}_{upper}} \|x - y\|, & GT^{c}(x) = 1 \\ 0, & otherwise \end{cases} \tag{1}$$

$$BR^{dis}_{upper} = cat(BR^{1}_{upper}, BR^{2}_{upper}, \ldots, BR^{C-1}_{upper}) \tag{2}$$

$$BR^{c}_{lower}(x) = \begin{cases} \inf\limits_{y \in B^{c}_{lower}} \|x - y\|, & GT^{c}(x) = 1 \\ 0. & otherwise \end{cases} \tag{3}$$

$$BR^{dis}_{lower} = cat(BR^{1}_{lower}, BR^{2}_{lower}, \ldots, BR^{C-1}_{lower}). \tag{4}$$

where GT uses One-hot encoding and $GT^{c}(x)$ refer to the x-th pixel in the c-th channel of segmentation mask. B^{c}_{upper} and B^{c}_{lower} refer to the upper boundary and lower boundary of the c-th retinal layer, respectively. $\|x - y\|$ refers to the Euclidean distance between x and y. The function inf refers to the minimum value taken from the set, which serves to obtain the shortest distance from the pixel to the upper or lower boundary of the layer to which it belongs. The function cat refers to combining many single-channel data into multi-channel data. Based on BR^{dis}_{upper} and BR^{dis}_{lower}, our dual boundary representation is defined as:

$$BR_{upper} = \frac{BR^{dis}_{upper}}{BR^{dis}_{upper} + BR^{dis}_{lower}}. \tag{5}$$

$$BR_{lower} = \frac{BR^{dis}_{lower}}{BR^{dis}_{upper} + BR^{dis}_{lower}}. \tag{6}$$

As can be seen from Eq. (5) and (6), the values of BR_{upper} and BR_{lower} all fall in the range of 0 to 1, which can improve the robustness of the representation. The upper boundary representation and the lower boundary representation can be derived from each other, which further enhances the representation capability. Moreover, the added results of these two boundary representations can be converted to segmentation results, which facilitates mutual assistance with the segmentation predictions. Please note that our dual boundary representation applies to 11 retinal layers and fluid, excluding the background, so BR_{upper} and BR_{lower} have $C - 1$ channels.

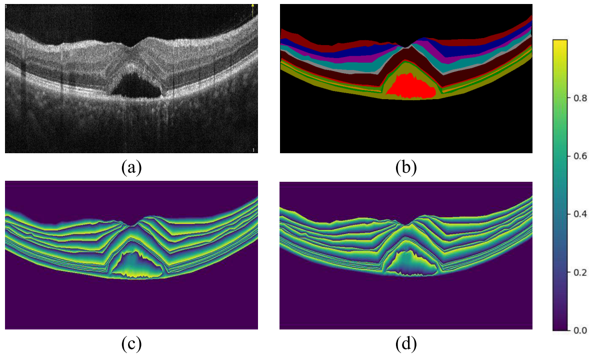


FIGURE 4. Example of our dual boundary representation. (a) is an OCT image with lesion, and (b) is its corresponding segmentation mask. (c) and (d) are its corresponding upper and lower boundary representations, respectively.

C. LOSS FUNCTION

There are four types of loss functions in our proposed network, namely, segmentation loss for the segmentation task, boundary loss for the boundary regression task, boundary complementation loss for mutual constraint between two boundary representations, and multi-task consistency loss for mutual constraint between the boundary regression task and the segmentation task.

1) SEGMENTATION LOSS

Instead of using conventional cross-entropy loss, we adopt Dice loss to supervise the segmentation prediction while it can better deal with the class imbalance problem and has better segmentation ability for smaller targets [41]. The detailed definition of Dice loss is as follows:

$$L_{segmentation} = 1 - \frac{1}{C}\sum_{c=0}^{C-1}\frac{2P^c GT^c + \tau}{P^c + GT^c + \tau}. \tag{7}$$

where GT^c refers to the c-th channel of the ground-truth label mask, and P^c refers to the c-th channel of the segmentation predicted mask. $C = 13$ refers to the number of channels, which corresponds to the number of classes. τ is a constant close to zero that is used to prevent calculation exceptions.

2) BOUNDARY LOSS

There are two items in our boundary loss while we have two boundary regression tasks, one for the upper boundary representation prediction and the other for the lower boundary representation prediction. Here, we adopt mean square error as the optimization objective, and the boundary loss is defined as:

$$L_{upper} = \frac{1}{C-1}\sum_{c=1}^{C-1}||PB^c_{upper} - BR^c_{upper}||^2, \tag{8}$$

$$L_{lower} = \frac{1}{C-1}\sum_{c=1}^{C-1}||PB^c_{lower} - BR^c_{lower}||^2, \tag{9}$$

$$L_{boundary} = L_{upper} + L_{lower}. \tag{10}$$

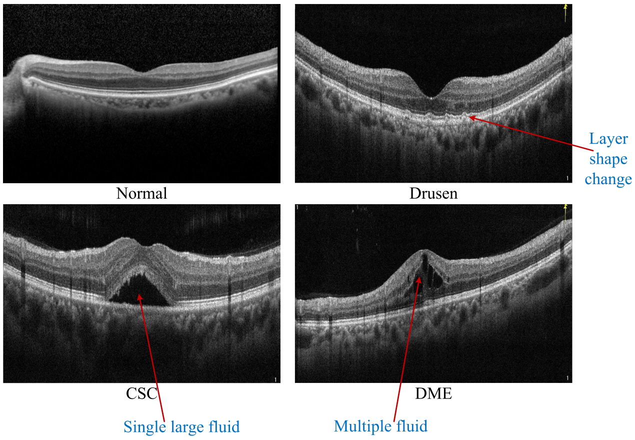


FIGURE 5. Examples of the effects of lesions on retinal structure.

where PB^c_{upper} and PB^c_{lower} are the c-th channel of upper and lower boundary representation prediction results and BR^c_{upper} and BR^c_{lower} are the c-th channel of upper and lower boundary representation labels, respectively. Here c starts from 1 instead of 0, which means that the calculation of the background is not performed.

3) BOUNDARY COMPLEMENTATION LOSS

In terms of Eq. (5) and (6), we can find that for each pixel, the adding value of the upper boundary representation and the lower boundary representation is always equal to 1, which is the innovation characteristic of our definition. To better capture the boundary, we propose a boundary complementation loss to enhance the learning capability for a boundary as follows:

$$L_{complementation} = \frac{1}{|PB_{add}|}\sum_{c=1}^{C-1}\sum_{x \in PB^c_{add}} d(PB^c_{add}(x)), \tag{11}$$

in which

$$PB_{add} = PB_{upper} + PB_{lower}, \tag{12}$$

$$d(PB^c_{add}(x)) = \|PB^c_{add}(x) - 1\|^2, x \in \{x \mid GT^c(x) = 1\}. \tag{13}$$

where PB_{upper} and PB_{lower} are the upper and lower boundary representation predictions, respectively, and PB_{add} is the result of adding dual boundary representation. $PB^c_{add}(x)$ represents the x-th pixel in the c-th channel of PB_{add}, $d(PB^c_{add}(x))$ is the square of the L2 norm of error between $PB^c_{add}(x)$ and 1, and $|PB_{add}|$ refers to the number of pixels in PB_{add}.

4) MULTI-TASK CONSISTENCY LOSS

There is a strong consistency between our boundary representation and the segmentation, while the added results of two boundary predictions can be converted to the segmentation results. So we designed a multi-task consistency loss to improve the mutual supervision between the boundary regression and the segmentation task. The definition is:

$$L_{consistency} = \frac{1}{C-1}\sum_{c=1}^{C-1} ||PB_{add}^{c} - P^{c}||^{2}. \qquad (14)$$

$L_{consistency}$ calculates the consistency difference between the predicted dual boundary representation and the predicted conventional segmentation.

5) TOTAL LOSS

When training our network, our total loss is a combination of the four losses mentioned above:

$$\begin{aligned} P_{total} =& w_1 L_{segmentation} + w_2 L_{boundary} \\ &+ w_3 L_{complementation} + w_4 L_{consistency}. \end{aligned} \qquad (15)$$

where w_1, w_2, w_3, and w_4 are loss weights to balance the importance of different losses. We set $w_1 = 10$ and $w_2 = w_3 = w_4 = 1$ in our implementation based on the empirical results.

IV. EXPERIMENTS

A. DATASET

We verified our method on an OCT image dataset collected from the Second Affiliated Hospital of Zhejiang University, and the retinal layer annotation was done by clinicians with the help of our developed annotation tool. The dataset contains healthy OCT images and also OCT images with ocular diseases, including Central Serous Chorioretinopathy (CSC), Diabetic Macular Edema (DME), and Drusen. As shown in Fig. 5, CSC causes one large fluid in the retina, DME causes multiple small fluids in the retina, and Drusen causes a change in the shape of retinal layers. Specifically, 950 OCT images without any diseases, 100 OCT images with CSC, 50 OCT images with DME, and 100 OCT images with Drusen were collected. And images with different diseases make the dataset more challenging.

These images have different resolutions, mainly 425 × 927, 496 × 769, and 496 × 528. The retina is divided into 11 layers, which are the Nerve Fiber Layer (NFL), Ganglion Cell Layer (GCL), Inner Plexiform Layer (IPL), Inner Nuclear Layer (INL), Outer Plexiform Layer (OPL), Outer Nuclear Layer (ONL), External Limiting Membrane (ELM), Myoid Zone (MZ), Ellipsoid Zone (EZ), Interdigitation Zone (IZ), and Retinal Pigment Epithelium (RPE). In some OCT images with diseases such as CSC and DME, there will be fluid in the retina, and we have also marked the fluid to assist in the diagnosis of the disease.

B. EXPERIMENTAL SETTING AND EVALUATION METRICS

To remove interference such as the optic nerve, we center cropped these images and unified them into 411 × 451 according to the minimum size of these images, so the input size of the network is 411 × 451 × 3. The whole dataset was divided into the training set, the validation set, and the test set in the ratio of 7:1:2 in terms of each type of image. We used the RMSProp optimization algorithm with $1e-8$ weight decay of 0.9 momentum to update the network weights and a learning rate scheduler with a decay of 0.5 every 10 steps. Batch size was set to 8, and the maximum number of training epochs was 50. We normalized the pixel values of the input image between 0 and 1 by dividing them by 255.

To better evaluate the layer segmentation performance, we adopt the Dice coefficient and Boundary Intersection-over-Union (BIoU) [42] for evaluation. The dice coefficient is used to evaluate the performance of region segmentation, and BIoU is used to measure the boundary accuracy. The definitions are as follows:

$$Dice = \frac{2|P\bigcap GT|}{|P| + |GT|}, \qquad (16)$$

$$BIoU = \frac{|(P_b\bigcap B)\bigcap(P\bigcap GT)|}{|(P_b\bigcap B)\bigcup(P\bigcap GT)|}. \qquad (17)$$

where GT refers to the ground-truth label mask, and P refers to the segmentation predicted result. B and P_b are the boundaries of GT and P obtained through pixel-by-pixel boundary judgment operations, respectively. $\bigcap$ represents the intersection of two sets, $\bigcup$ represents the union of two sets, and $+$ represents the sum of the number of pixels in the two sets.

C. COMPARISON WITH SOTA METHODS

To better evaluate the performance of our model in the retinal layering task, we compared our method with some SOTA image segmentation methods, including Unet [43], Relaynet [12], Deeplabv3+ [44], HarDNet-MSEG [45], BASNet [46], CTS-Net [24], Swin-Unet [18], DuAT [47], MultiResUnet [48], and TransSegNet [49]. Unet and Relaynet are traditional U-structured networks that use skip connections to merge detailed features and high-dimensional features. Deeplabv3+ uses dilated convolution and spatial pyramid pooling to increase the perceptual domain, while HarDNet-MSEG achieved SOTA in both accuracy and inference speed on five medical datasets. BASNet uses two cascaded U-structure segmentation networks, where the first network uses a deep supervision strategy to generate preliminary segmentation results, and the second network optimizes the segmentation results to obtain refined results. CTS-Net and Swin-Unet applied CSwin Transformer and Swin Transformer [50], respectively, as the basic architecture of the encoder and decoder in the model. DuAT applied the Pyramid Transformer [51] module combined with global-to-local spatial aggregation and a selective boundary aggregation block to generate fine-grained segmentation results. TransSegNet combined Unet and Vision Transformer to take advantage of the complementary benefits of CNN-based network and Transformer-based network. MultiResUnet improves on the architecture of Unet, enabling the network to learn image features of different scales and solving the semantic gap between the corresponding levels of encoder and decoder. For a fair comparison, we use the same training strategy for all methods.

TABLE 1. Comparison results in terms of Dice coefficient. The red fonts indicate the best results among all methods.

	Unet DuAT	Relaynet MultiResUnet	Deeplabv3+ CTS-Net	HarDNet-MSEG TransSegNet	BASNet Swin-Unet	DBA-Net(Ours)
NFL	91.160% 89.214%	82.049% 89.013%	89.218% 85.760%	90.559% 90.132%	86.138% 88.456%	91.194% (1st)
GCL	89.894% 91.843%	78.792% 91.178%	88.934% 87.248%	89.657% 91.895%	77.655% 87.877%	90.217% (4th)
IPL	87.807% 89.406%	76.277% 88.810%	86.425% 84.003%	87.332% 88.256%	64.940% 84.243%	88.250% (4th)
INL	90.998% 91.809%	77.803% 90.878%	89.992% 86.723%	90.748% 89.586%	76.658% 87.454%	91.477% (2nd)
OPL	87.062% 85.275%	74.531% 84.928%	85.768% 78.833%	86.159% 78.139%	68.796% 79.384%	87.354% (1st)
ONL	94.340% 93.157%	81.082% 93.688%	92.529% 90.732%	94.122% 92.290%	88.176% 91.213%	94.674% (1st)
ELM	84.388% 2.139%	72.457% 82.400%	81.365% 74.830%	82.765% 66.357%	65.593% 78.320%	84.592% (1st)
MZ	85.923% 68.345%	73.409% 83.602%	82.836% 78.023%	84.538% 69.541%	72.192% 80.961%	86.184% (1st)
EZ	87.094% 82.579%	75.320% 85.163%	84.621% 82.740%	85.792% 80.657%	78.050% 82.862%	87.307% (1st)
IZ	84.632% 82.083%	72.747% 83.874%	82.525% 81.738%	82.540% 79.494%	71.732% 78.984%	84.599% (2nd)
RPE	91.322% 89.807%	80.086% 89.281%	88.660% 86.827%	89.677% 89.597%	82.952% 88.074%	91.423% (1st)
Fluid	87.736% 97.859%	57.087% 15.745%	94.044% 68.999%	91.171% 87.846%	87.607% 70.054%	92.292% (3rd)
mean	88.530% 80.857%	75.137% 82.780%	87.243% 83.173%	87.922% 84.877%	76.707% 84.424%	89.130% (1st)

TABLE 2. Comparison results in terms of BIoU. The red fonts indicate the best results among all methods.

	Unet DuAT	Relaynet MultiResUnet	Deeplabv3+ CTS-Net	HarDNet-MSEG TransSegNet	BASNet Swin-Unet	DBA-Net(Ours)
NFL	66.008% 53.814%	59.608% 54.467%	63.009% 48.504%	62.330% 58.706%	48.259% 57.189%	66.221% (1st)
GCL	53.645% 49.367%	48.871% 50.001%	52.503% 34.986%	52.329% 51.066%	20.615% 39.610%	53.959% (1st)
IPL	53.702% 50.184%	48.735% 51.106%	52.462% 40.997%	52.675% 48.050%	15.359% 40.083%	53.857% (1st)
INL	60.904% 54.715%	54.620% 55.643%	58.532% 44.541%	58.823% 47.712%	18.552% 45.871%	61.125% (1st)
OPL	60.277% 47.956%	53.407% 51.897%	58.373% 42.431%	58.724% 42.164%	25.694% 43.428%	60.728% (1st)
ONL	61.327% 43.487%	55.137% 57.810%	58.640% 44.294%	59.526% 44.676%	29.095% 45.259%	61.705% (1st)
ELM	74.148% 13.577%	65.052% 72.187%	70.482% 61.180%	72.267% 51.120%	49.785% 65.844%	74.427% (1st)
MZ	75.286% 41.566%	65.926% 72.317%	71.062% 65.211%	73.435% 54.080%	56.274% 68.725%	75.668% (1st)
EZ	76.442% 68.082%	67.391% 74.208%	72.599% 70.796%	74.595% 67.623%	62.913% 70.958%	76.754% (1st)
IZ	70.929% 64.401%	62.701% 67.805%	67.589% 60.607%	66.997% 64.162%	50.120% 62.609%	70.950% (1st)
RPE	61.096% 50.147%	54.666% 49.189%	53.231% 47.040%	56.514% 50.217%	31.132% 47.862%	61.308% (1st)
Fluid	81.132% 89.575%	55.103% 9.889%	88.335% 69.911%	83.777% 87.655%	87.607% 64.264%	85.630% (2nd)
mean	66.241% 50.847%	57.601% 54.670%	63.901% 52.091%	64.333% 55.171%	41.284% 54.135%	66.861% (1st)

We calculated the Dice coefficient and BIoU for each class as well as the average results of all classes. Quantitative results of our method and other comparison methods are reported in Table 1 and Table 2. Among the 11 layers and fluid of the retina, our method achieved the top results in 7 layers (NFL, OPL, ONL, ELM, MZ, EZ, RPE) and the second highest results in 2 layers (INL, IZ) in terms of Dice coefficient; the average value was 5.965% higher than the best results of all compared methods. Our method achieved the top results in all layers in terms of BIoU, whose average value was 10.833% higher than the best results of all compared methods. BIoU represents the accuracy of

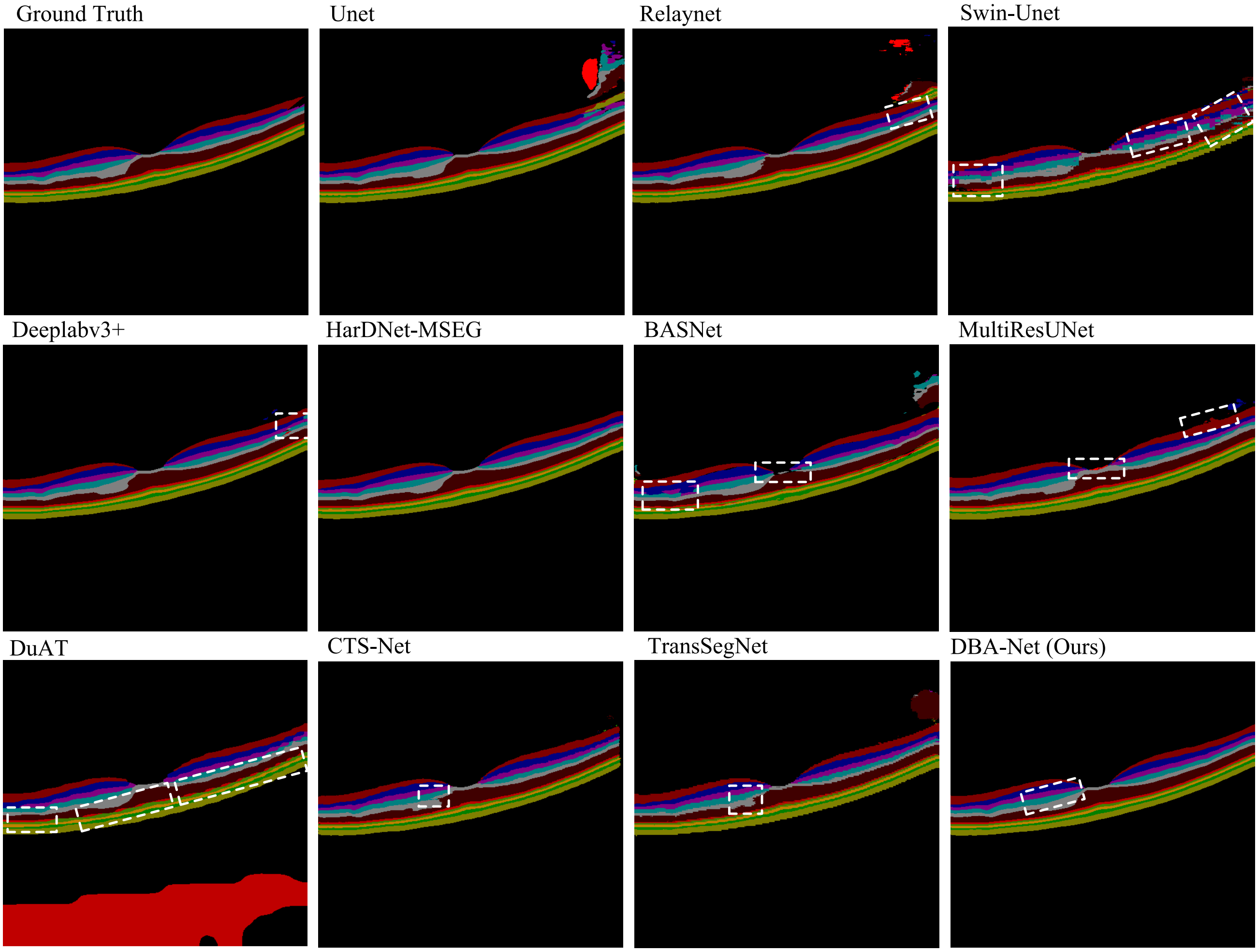


FIGURE 6. Segmentation results of all methods on one image without lesion. The area circled by the white box means segmentation errors occur compared to our model.

the boundaries in the segmentation results, and the best average BIoU value of our method indicates that our method can obtain more accurate boundaries. Also, we performed a qualitative analysis to visually compare the segmentation results of one OCT image without a lesion and one with a lesion, as shown in Fig. 6 and Fig. 7. Other methods suffered segmentation errors in the white boxes compared to ours, which can better deal with the case when noise or interference occurs than other methods. From the experimental results, we can conclude that our model achieves satisfactory results in the segmentation of narrower layers and interlayer boundaries in the retina compared to other current methods, including the Transformer-based networks, demonstrating the effectiveness of our proposed dual boundary constraint loss functions.

D. ABLATION STUDY

In our method, there are several improvements over the backbone, including upper boundary representation, lower boundary representation, boundary complementary loss, and multi-task consistency loss. To demonstrate their effectiveness, we have implemented several networks to perform ablation experiments by gradually adding the corresponding modules from the backbone, namely, the original backbone, the backbone with the upper boundary representation branch (backbone+upper), the backbone with the lower boundary

TABLE 3. Ablation studies in terms of Dice coefficient.

	backbone	+upper	+lower	+dual	all
NFL	90.783%	91.388%	91.597%	91.475%	91.194%
GCL	90.069%	90.271%	90.271%	90.363%	90.217%
IPL	87.861%	87.887%	88.381%	88.424%	88.250%
INL	91.100%	91.089%	91.520%	91.177%	91.477%
OPL	87.093%	87.159%	87.433%	87.291%	87.354%
ONL	94.395%	94.634%	94.697%	94.540%	94.674%
ELM	84.538%	84.642%	84.832%	84.681%	84.592%
MZ	86.113%	86.109%	86.336%	86.124%	86.184%
EZ	87.323%	87.000%	87.389%	87.252%	87.307%
IZ	84.655%	84.704%	84.745%	84.627%	84.599%
RPE	91.026%	91.318%	91.321%	91.324%	91.423%
Fluid	83.587%	89.123%	87.172%	88.997%	92.292%
mean	88.212%	88.777%	88.808%	88.856%	89.130%

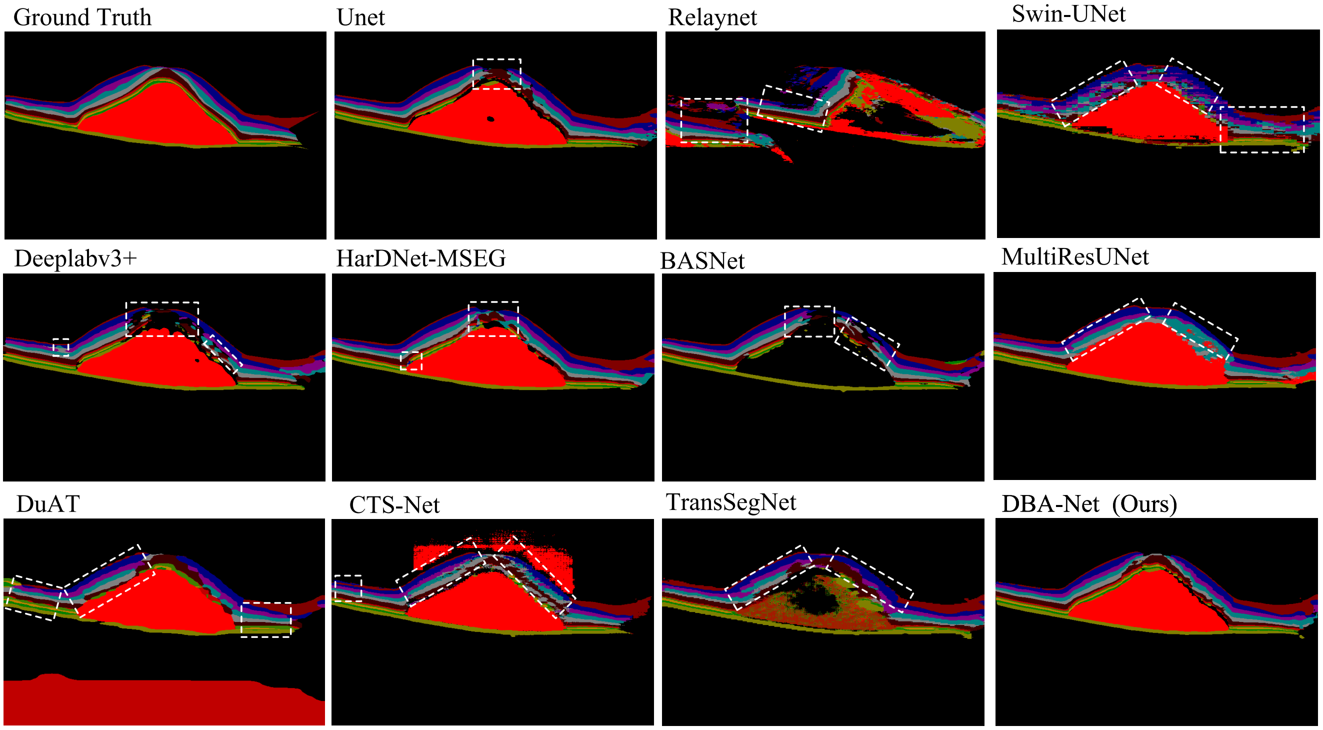


FIGURE 7. Segmentation results of all methods on one image with a lesion. The area circled by the white box means segmentation errors occur compared to our model.

TABLE 4. Ablation studies in terms of BIoU.

	backbone	+upper	+lower	+dual	all
NFL	66.221%	66.602%	66.852%	66.564%	66.221%
GCL	53.938%	54.360%	54.441%	54.225%	53.959%
IPL	53.511%	54.217%	54.155%	54.394%	53.857%
INL	60.650%	61.235%	61.474%	61.078%	61.125%
OPL	60.322%	60.919%	61.072%	60.512%	60.728%
ONL	61.587%	62.159%	62.245%	62.140%	61.705%
ELM	74.367%	74.497%	74.761%	74.650%	74.427%
MZ	75.588%	75.539%	75.907%	75.626%	75.668%
EZ	76.751%	76.423%	76.907%	76.720%	76.754%
IZ	70.665%	70.977%	71.068%	70.897%	70.950%
RPE	60.103%	61.092%	61.227%	61.073%	61.308%
Fluid	77.772%	82.494%	80.496%	82.879%	85.630%
mean	65.956%	66.710%	66.717%	66.730%	66.861%

representation branch (backbone+lower), the backbone with both the upper and lower boundary representation branches and also the boundary complementary loss (backbone+dual), and the backbone with both the upper and lower boundary representation branches, the boundary complementary loss, and the multi-task consistency loss (ours).

The quantitative results of the ablation experiments are reported in Table 3 and Table 4. It can be seen that the Dice coefficient and BIoU of the backbone were improved when the upper boundary representation branch or the lower boundary representation branch was added, which indicated that the learned boundary-related features under the supervision of the boundary representation could enhance the segmentation performance. When the backbone added both the upper and lower boundary representation branches, its Dice coefficient and BIoU were higher than both backbone+upper and backbone+lower, which demonstrated the effectiveness of our dual boundary representation. Finally, our proposed method achieved a higher Dice coefficient and BIoU than backbone+dual, which demonstrated the effectiveness of the multi-task consistency loss.

E. SHARE STRUCTURE STUDY

Network weight sharing strategy was used in our method, which is inspired by [52]. Different sharing structures determined the number of network parameters and the degree of sharing between multiple tasks. In order to explore a more suitable network-sharing structure, we selected five network-sharing structures, as shown in Fig. 8, for comparison.

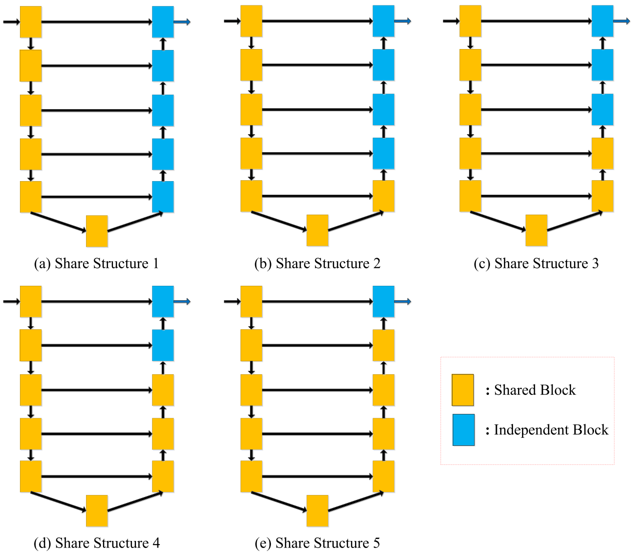
The segmentation metrics and network parameters of different network sharing structures are shown in Table 5. It can be seen that shared structure 3 has achieved the best segmentation result. When the degree of sharing is large, it is difficult for the shared features to satisfy multiple tasks at the same time. When the degree of sharing is small, the multi-tasks cannot improve the feature extraction ability. We finally chose shared structure 3 as our network shared structure.

F. LOSS WEIGHT STUDY

In our experiment, the selection of four loss function weight hyperparameters is very important and plays a decisive role in the final experimental result. In this regard,

TABLE 5. Quantitative segmentation results and parameter quantities for different network sharing structures.

	Share Structure 1	Share Structure 2	Share Structure 3	Share Structure 4	Share Structure 5
Mean Dice	88.800%	88.730%	**89.130%**	88.464%	88.980%
Mean BIoU	66.586%	66.539%	**66.861%**	66.228%	66.834%
Params	58.947M	44.785M	**32.984M**	30.033M	29.294M



(a) Share Structure 1 (b) Share Structure 2 (c) Share Structure 3

(d) Share Structure 4 (e) Share Structure 5

FIGURE 8. Network share structures. From (a) to (e), the degree of network sharing increases sequentially.

TABLE 6. Comparison results of different loss weights in terms of Dice coefficient.

	All Equal	w4 Max	w3 Max	w2 Max	w1 Max
NFL	53.150%	74.579%	69.388%	80.572%	**91.194%**
GCL	66.326%	49.429%	74.093%	81.289%	**90.217%**
IPL	67.940%	45.945%	67.861%	20.417%	**88.250%**
INL	81.201%	16.699 %	14.459%	78.467%	**91.477%**
OPL	65.652%	55.593%	59.139%	69.527%	**87.354%**
ONL	82.553%	84.132%	80.2645%	87.468%	**94.674%**
ELM	61.766%	66.020%	64.546%	71.280%	**84.592%**
MZ	56.276%	66.173%	58.086%	61.151%	**86.184%**
EZ	76.417%	80.710%	2.125%	79.815%	**87.307%**
IZ	0.251%	77.103%	70.901%	51.093%	**84.599%**
RPE	70.799%	77.399%	73.262%	80.288%	**91.423%**
Fluid	0.305%	0.228%	87.606%	0.662%	**92.292%**
Mean	59.421%	59.940%	62.185%	65.323%	**89.130%**

TABLE 7. Comparison results of different loss weights in terms of BIoU coefficient.

	All Equal	w4 Max	w3 Max	w2 Max	w1 Max
NFL	15.817%	32.741%	28.750%	41.418%	**66.221%**
GCL	22.866%	14.099%	29.700%	37.157%	**53.959%**
IPL	29.214%	14.950%	27.783%	11.067%	**53.857%**
INL	39.135%	6.563%	9.103%	38.249%	**61.125%**
OPL	31.306%	26.639%	25.514%	35.584%	**60.728%**
ONL	33.910%	30.304%	31.111%	38.173%	**61.705%**
ELM	45.877%	50.454%	48.626%	56.424%	**74.427%**
MZ	39.878%	50.812%	40.716%	44.516%	**75.668%**
EZ	56.500%	67.696%	1.055%	63.180%	**76.754%**
IZ	1.158%	60.049%	46.225%	32.728%	**70.950%**
RPE	19.163%	29.141%	25.018%	25.840%	**61.308%**
Fluid	1.204%	0.0395%	87.606%	0.145%	**85.630%**
Mean	25.705%	29.552%	30.929%	32.733%	**66.861%**

we designed a comparison experiment to verify the difference in segmentation results caused by different weight choices of loss functions. In practice, we assign the weights of each loss function to integers from 1 to 10 and then perform the experiment. The comparison experiment results can be seen in Table. 6 and Table. 7. Here, we analyze the model's performance by separately setting the weights of each loss function ($w_i = 1, \ i = \{1, 2, 3, 4\}$) to 10, while assigning the weights of the remaining three loss functions as 1. It can be seen that the model performs best when the segmentation loss $w1$ is set to the maximum value of 10 for both the segmentation effect of a single retinal layer and multiple retinal layers. From the perspective of experience, the weight allocation of the loss function should be set higher for the segmentation loss so that the model can pay more attention to the segmentation task, while the boundary constraint loss function is to increase the performance of the model based on the segmentation task and should not be placed in the first place. Thus, our loss function weight allocation is reasonable for tasks primarily based on segmentation.

V. CONCLUSION

In this paper, we propose an advanced boundary-aware multi-task image segmentation network, DBA-Net, to perform layer segmentation in retinal OCT images. Different from conventional methods, our boundary representation is based on the hierarchical relationship between adjacent layers and enhances the segmentation task learning through a consistency constraint between the segmentation task and the boundary task. Extensive experiments demonstrate that our method achieves superior performance than other SOTA methods.

ACKNOWLEDGMENT

(Ce Yang, Wenyu Wang, and Chengyu Wu contributed equally to this work.)

REFERENCES

[1] R. R. Bourne, S. R. Flaxman, T. Braithwaite, M. V. Cicinelli, A. Das, J. B. Jonas, J. Keeffe, J. H. Kempen, J. Leasher, and H. Limburg, "Magnitude, temporal trends, and projections of the global prevalence of blindness and distance and near vision impairment: A systematic review and meta-analysis," *Lancet Global Health*, vol. 5, no. 9, pp. e888–e897, 2017.

[2] 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.

[3] J. Yang, Y. Tao, Q. Xu, Y. Zhang, X. Ma, S. Yuan, and Q. Chen, "Self-supervised sequence recovery for semi-supervised retinal layer segmentation," *IEEE J. Biomed. Health Informat.*, vol. 26, no. 8, pp. 3872–3883, Aug. 2022.

[4] B. J. Antony, M. Chen, A. Carass, B. M. Jedynak, O. Al-Louzi, S. D. Solomon, S. Saidha, P. A. Calabresi, and J. L. Prince, ''Voxel based morphometry in optical coherence tomography: Validation and core findings,'' *Proc. SPIE*, vol. 9788, pp. 180–187, Mar. 2016.

[5] S. Lee, N. Charon, B. Charlier, K. Popuri, E. Lebed, M. V. Sarunic, A. Trouvé, and M. F. Beg, ''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, Jan. 2017.

[6] W. Drexler and J. Fujimoto, ''State-of-the-art retinal optical coherence tomography,'' *Prog. Retinal Eye Res.*, vol. 27, no. 1, pp. 45–88, Jan. 2008.

[7] S. Niu, L. de Sisternes, Q. Chen, T. Leng, and D. L. Rubin, ''Automated geographic atrophy segmentation for SD-OCT images using region-based CV model via local similarity factor,'' *Biomed. Opt. Exp.*, vol. 7, no. 2, pp. 581–600, 2016.

[8] Z. Ji, Q. Chen, S. Niu, T. Leng, and D. L. Rubin, ''Beyond retinal layers: A deep voting model for automated geographic atrophy segmentation in SD-OCT images,'' *Translational Vis. Sci. Technol.*, vol. 7, no. 1, p. 1, Jan. 2018.

[9] J. Arslan, G. Samarasinghe, A. Sowmya, K. K. Benke, L. A. B. Hodgson, R. H. Guymer, and P. N. Baird, ''Deep learning applied to automated segmentation of geographic atrophy in fundus autofluorescence images,'' *Transl. Vis. Sci. Technol.*, vol. 10, no. 8, p. 2, Jul. 2021.

[10] D. Xiang, H. Tian, X. Yang, F. Shi, W. Zhu, H. Chen, and X. Chen, ''Automatic segmentation of retinal layer in OCT images with choroidal neovascularization,'' *IEEE Trans. Image Process.*, vol. 27, no. 12, pp. 5880–5891, Dec. 2018.

[11] 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.

[12] 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. Exp.*, vol. 8, no. 8, pp. 3627–3642, 2017.

[13] Q. Li, S. Li, Z. He, H. Guan, R. Chen, Y. Xu, T. Wang, S. Qi, J. Mei, and W. Wang, ''DeepRetina: Layer segmentation of retina in OCT images using deep learning,'' *Transl. Vis. Sci. Technol.*, vol. 9, no. 2, p. 61, Dec. 2020.

[14] 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.

[15] K. Hu, D. Liu, Z. Chen, X. Li, Y. Zhang, and X. Gao, ''Embedded residual recurrent network and graph search for the segmentation of retinal layer boundaries in optical coherence tomography,'' *IEEE Trans. Instrum. Meas.*, vol. 70, pp. 1–17, 2021.

[16] B. Wang, W. Wei, S. Qiu, S. Wang, D. Li, and H. He, ''Boundary aware U-Net for retinal layers segmentation in optical coherence tomography images,'' *IEEE J. Biomed. Health Informat.*, vol. 25, no. 8, pp. 3029–3040, Aug. 2021.

[17] Y. He, A. Carass, Y. Liu, B. M. Jedynak, S. D. Solomon, S. Saidha, P. A. Calabresi, and J. L. Prince, ''Structured layer surface segmentation for retina OCT using fully convolutional regression networks,'' *Med. Image Anal.*, vol. 68, Jan. 2021, Art. no. 101856.

[18] H. Cao, Y. Wang, J. Chen, D. Jiang, X. Zhang, Q. Tian, and M. Wang, ''Swin-Unet: Unet-like pure transformer for medical image segmentation,'' in *Proc. Eur. Conf. Comput. Vis.* Cham, Switzerland: Springer, 2022, pp. 205–218.

[19] J. Chen, Y. Lu, Q. Yu, X. Luo, E. Adeli, Y. Wang, L. Lu, A. L. Yuille, and Y. Zhou, ''TransUNet: Transformers make strong encoders for medical image segmentation,'' 2021, *arXiv:2102.04306*.

[20] H. Xiao, L. Li, Q. Liu, X. Zhu, and Q. Zhang, ''Transformers in medical image segmentation: A review,'' *Biomed. Signal Process. Control*, vol. 84, Mar. 2023, Art. no. 104791.

[21] S. Apostolopoulos, S. De Zanet, C. Ciller, S. Wolf, and R. Sznitman, ''Pathological OCT retinal layer segmentation using branch residual U-shape networks,'' 2017, *arXiv:1707.04931*.

[22] K. Gopinath, S. B. Rangrej, and J. Sivaswamy, ''A deep learning framework for segmentation of retinal layers from OCT images,'' in *Proc. 4th IAPR Asian Conf. Pattern Recognit. (ACPR)*, Nov. 2017, pp. 888–893.

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

[24] S. Xue, H. Wang, X. Guo, M. Sun, K. Song, Y. Shao, H. Zhang, and T. Zhang, ''CTS-Net: A segmentation network for glaucoma optical coherence tomography retinal layer images,'' *Bioengineering*, vol. 10, no. 2, p. 230, 2023.

[25] X. Dong, J. Bao, D. Chen, W. Zhang, N. Yu, L. Yuan, D. Chen, and B. Guo, ''CSWin transformer: A general vision transformer backbone with cross-shaped windows,'' in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jun. 2022, pp. 12114–12124.

[26] H. Kervadec, J. Bouchtiba, C. Desrosiers, E. Granger, J. Dolz, and I. B. Ayed, ''Boundary loss for highly unbalanced segmentation,'' in *Proc. Int. Conf. Med. Imag. Deep Learn.*, 2019, pp. 285–296.

[27] D. Cheng, R. Liao, S. Fidler, and R. Urtasun, ''DARNet: Deep active ray network for building segmentation,'' in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jun. 2019, pp. 7423–7431.

[28] Y. Meng, M. Wei, D. Gao, Y. Zhao, X. Yang, X. Huang, and Y. Zheng, ''CNN-GCN aggregation enabled boundary regression for biomedical image segmentation,'' in *Proc. Int. Conf. Med. Image Comput. Comput.-Assist. Intervent.* Cham, Switzerland: Springer, 2020, pp. 352–362.

[29] C. M. Tam, D. Zhang, B. Chen, T. Peters, and S. Li, ''Holistic multitask regression network for multiapplication shape regression segmentation,'' *Med. Image Anal.*, vol. 65, May 2020, Art. no. 101783.

[30] E. Xie, P. Sun, X. Song, W. Wang, X. Liu, D. Liang, C. Shen, and P. Luo, ''PolarMask: Single shot instance segmentation with polar representation,'' in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jun. 2020, pp. 12190–12199.

[31] D.-P. Fan, G.-P. Ji, T. Zhou, G. Chen, H. Fu, J. Shen, and L. Shao, ''PraNet: Parallel reverse attention network for polyp segmentation,'' in *Proc. Int. Conf. Med. Image Comput. Comput.-Assist. Intervent.* Cham, Switzerland: Springer, 2020, pp. 263–273.

[32] R. Zhang, G. Li, Z. Li, S. Cui, D. Qian, and Y. Yu, ''Adaptive context selection for polyp segmentation,'' in *Proc. Int. Conf. Med. Image Comput. Comput.-Assist. Intervent.* Cham, Switzerland: Springer, 2020, pp. 253–262.

[33] Z. Zhang, H. Fu, H. Dai, J. Shen, Y. Pang, and L. Shao, ''ET-Net: A generic edge-attention guidance network for medical image segmentation,'' in *Proc. Int. Conf. Med. Image Comput. Comput.-Assist. Intervent.* Cham, Switzerland: Springer, 2019, pp. 442–450.

[34] Y. Meng, H. Zhang, Y. Zhao, X. Yang, Y. Qiao, I. J. C. MacCormick, X. Huang, and Y. Zheng, ''Graph-based region and boundary aggregation for biomedical image segmentation,'' *IEEE Trans. Med. Imag.*, vol. 41, no. 3, pp. 690–701, Mar. 2022.

[35] T. Wang, J. Yang, Z. Ji, and Q. Sun, ''Probabilistic diffusion for interactive image segmentation,'' *IEEE Trans. Image Process.*, vol. 28, no. 1, pp. 330–342, Jan. 2019.

[36] A. Shah, L. Zhou, M. D. Abramoff, and X. Wu, ''Multiple surface segmentation using convolution neural nets: Application to retinal layer segmentation in OCT images,'' *Biomed. Opt. Exp.*, vol. 9, no. 9, pp. 4509–4526, 2018.

[37] Y. He, A. Carass, Y. Liu, B. M. Jedynak, S. D. Solomon, S. Saidha, P. A. Calabresi, and J. L. Prince, ''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.

[38] C. Tan, L. Zhao, Z. Yan, K. Li, D. Metaxas, and Y. Zhan, ''Deep multitask and task-specific feature learning network for robust shape preserved organ segmentation,'' in *Proc. IEEE 15th Int. Symp. Biomed. Imag. (ISBI)*, Apr. 2018, pp. 1221–1224.

[39] J. Cao, X. Liu, Y. Zhang, and M. Wang, ''A multi-task framework for topology-guaranteed retinal layer segmentation in OCT images,'' in *Proc. IEEE Int. Conf. Syst., Man, Cybern. (SMC)*, Oct. 2020, pp. 3142–3147.

[40] X. Liu, J. Cao, S. Wang, Y. Zhang, and M. Wang, ''Confidence-guided topology-preserving layer segmentation for optical coherence tomography images with focus-column module,'' *IEEE Trans. Instrum. Meas.*, vol. 70, pp. 1–12, 2021.

[41] F. Milletari, N. Navab, and S.-A. Ahmadi, ''V-Net: Fully convolutional neural networks for volumetric medical image segmentation,'' in *Proc. 4th Int. Conf. 3D Vis. (3DV)*, Oct. 2016, pp. 565–571.

[42] B. Cheng, R. Girshick, P. Dollár, A. C. Berg, and A. Kirillov, ''Boundary IoU: Improving object-centric image segmentation evaluation,'' in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jun. 2021, pp. 15329–15337.

[43] O. Ronneberger, P. Fischer, and T. Brox, "U-Net: Convolutional networks for biomedical image segmentation," in *Proc. Int. Conf. Med. Image Comput. Comput.-Assist. Intervent.* Cham, Switzerland: Springer, 2015, pp. 234–241.

[44] 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. (ECCV)*, 2018, pp. 801–818.

[45] C.-H. Huang, H.-Y. Wu, and Y.-L. Lin, "HarDNet-MSEG: A simple encoder–decoder polyp segmentation neural network that achieves over 0.9 mean dice and 86 FPS," 2021, *arXiv:2101.07172*.

[46] X. Qin, D.-P. Fan, C. Huang, C. Diagne, Z. Zhang, A. C. Sant'Anna, A. Suarez, M. Jagersand, and L. Shao, "Boundary-aware segmentation network for mobile and web applications," 2021, *arXiv:2101.04704*.

[47] F. Tang, Q. Huang, J. Wang, X. Hou, J. Su, and J. Liu, "DuAT: Dual-aggregation transformer network for medical image segmentation," 2022, *arXiv:2212.11677*.

[48] N. Ibtehaz and M. S. Rahman, "MultiResUNet: Rethinking the U-Net architecture for multimodal biomedical image segmentation," *Neural Netw.*, vol. 121, pp. 74–87, Jan. 2020.

[49] Y. Zhang, Z. Li, N. Nan, and X. Wang, "TranSegNet: Hybrid CNN-vision transformers encoder for retina segmentation of optical coherence tomography," *Life*, vol. 13, no. 4, p. 976, Apr. 2023.

[50] Z. Liu, Y. Lin, Y. Cao, H. Hu, Y. Wei, Z. Zhang, S. Lin, and B. Guo, "Swin transformer: Hierarchical vision transformer using shifted windows," in *Proc. IEEE/CVF Int. Conf. Comput. Vis. (ICCV)*, Oct. 2021, pp. 9992–10002.

[51] W. Wang, E. Xie, X. Li, D.-P. Fan, K. Song, D. Liang, T. Lu, P. Luo, and L. Shao, "Pyramid vision transformer: A versatile backbone for dense prediction without convolutions," in *Proc. IEEE/CVF Int. Conf. Comput. Vis. (ICCV)*, Oct. 2021, pp. 548–558.

[52] J. Zhuang, "LadderNet: Multi-path networks based on U-Net for medical image segmentation," 2018, *arXiv:1810.07810*.

CE YANG received the B.S. degree in software engineering from Qufu Normal University, China, in 2020. He is currently pursuing the M.S. degree in computer science with Shandong University, China.

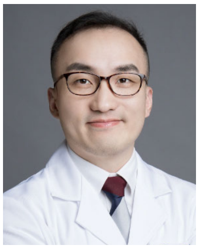
His current research interests include medical image analysis and compute vision.

WENYU WANG received the B.S. degree in computer science from the Dalian University of Technology, China, in 2000, and the Ph.D. degree from Shandong University, Weihai, China, in 2015.

She was a Postdoctoral Fellow with Shandong University, from 2015 to 2018, where she is currently an Associate Professor. Her current research interests include data mining and machine learning.

CHENGYU WU received the B.S. degree in computer science and technology from Jiangsu Normal University, China, in 2022. He is currently pursuing the M.S. degree in computer science with Shandong University, China.

His current research interests include medical image analysis and compute vision.



KAI JIN received the M.D. and Ph.D. degrees in ophthalmology from Zhejiang University, in 2018.

He is currently a Physician and a Research Scientist with the Department of Ophthalmology, The Second Affiliated Hospital, Zhejiang University School of Medicine. To date, he has published more than 40 peer-reviewed papers. His current research interests include artificial intelligence-based decision making in ophthalmology and medical image analysis.

Dr. Jin's awards and honors include the Prize of Zhejiang Province Science and Technology Progress Award and Outstanding Graduates of Zhejiang Province.

YAN YAN received the master's degree in medicine (ophthalmology specialty) from Zhejiang University. He is currently pursuing the Ph.D. degree with the Department of Ophthalmology, The Second Affiliated Hospital, Zhejiang University School of Medicine, China.

His current research interest includes the computer-aided diagnosis of fundus disease.

JUAN YE received the M.D. degree from Zhejiang University, China, and the Ph.D. degree from Yonsei University, South Korea.

She is currently a Professor with the Department of Ophthalmology, The Second Affiliated Hospital, Zhejiang University School of Medicine, and the Deputy Director of the Chinese Society of Oculoplastic Surgery and Orbital Disease. Her current research interests include artificial intelligence-based decision making in ophthalmology, biological/medical treatment techniques, and epidemiology of age-related eye disease.

Dr. Ye's awards and honors include the Prize of National Science and Technology Progress Award and the Prize of Zhejiang Province Science and Technology Progress Award.

SHUAI WANG received the B.S. degree from the China University of Mining and Technology, Xuzhou, China, in 2010, and the Ph.D. degree from the State Key Laboratory of Robotics, Chinese Academy of Sciences, Shenyang, China, in 2017.

He was a Postdoctoral Fellow with the School of Medicine, The University of North Carolina at Chapel Hill, Chapel Hill, NC, USA, from 2017 to 2019, and a Visiting Fellow with the Clinical Center, National Institutes of Health, Bethesda, MD, USA, from 2020 to 2021. He is currently a Full Professor with Hangzhou Dianzi University, Hangzhou, China. His current research interests include medical image analysis, compute vision, and machine learning.