Dual Difficulty-aware Adaptive Pseudo Labeling for Semi-supervised CNV Segmentation

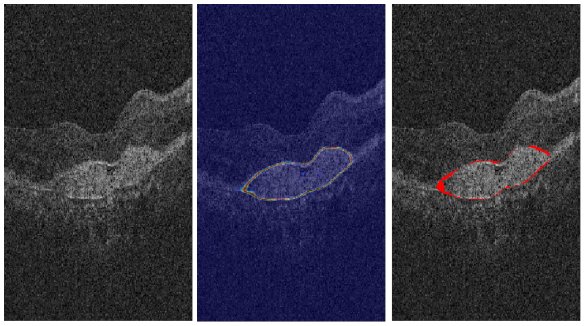
Jie Guo*, Liangyun Sun*, Lishan Qiao, Xiushan Nie, Jixin Yang, Weicui Li, Ying Guo, Xiaoming Xi†, Xinjian Chen†, Yilong Yin

***Abstract*—In clinical practice, obtaining a large amount of labeled CNV data is very difficult. Semi-supervised learning can effectively utilize a large amount of unlabeled CNV data. Since CNV has complex features such as blurred and unevenly distributed pixels on the edges, there are differences in the segmentation difficulty between pixels in the same image. Existing semi-supervised segmentation methods do not consider the segmentation difficulty of pixels, which will reduce the segmentation accuracy. To address this problem, we propose a dual difficulty-aware adaptive pseudo-label learning (D2APL) method for semi-supervised CNV segmentation. The proposed dual difficulty awareness includes segmentation difficulty perception of pixels in labeled and unlabeled data. For labeled data, we propose a classification confidence-guided difficulty perception method. For unlabeled data, we propose a model stability-guided difficulty perception method. Finally, we propose a difficulty-aware self-training method to dynamically adjust the threshold of pseudo-labels according to the difficulty, thereby improving the utilization of difficult-to-segment pixels in unlabeled data. Experimental results show that our method outperforms the state-of-the-art method in CNV segmentation.**

***Index Terms*—CNV segmentation, Semi-supervised learning, Dual difficulty aware, Pseudo labeling**

I. Introduction

LATE-STAGE age-related macular degeneration (AMD) typically involves the formation of choroidal neovascularization (CNV), which is a major cause of vision loss and irreversible blindness [1]–[3]. Optical Coherence Tomography (OCT) has been widely used in the diagnosis of CNV in recent years. Compared to other imaging methods, OCT offers the advantages of non-invasiveness, rapid image acquisition, and high safety [4].



(a) (b) (c)

Fig. 1. An example showcasing the challenges of semi-supervised CNV segmentation. (a) Original image (b) Uncertainty map (c) Visualization of the difficult-to-segment pixel regions.

Accurate segmentation of CNV can assist physicians in diagnosing and treating diseases. Through observation of these regions, physicians can gain insights into CNV lesion areas, volume, and other information, facilitating disease diagnosis and treatment. In clinical practice, obtaining a large amount of labeled CNV data is very difficult, which limits the accuracy of current automatic segmentation methods. Therefore, it is particularly important to use semi-supervised learning to realize the utilization of a large amount of unlabeled CNV data.

Existing semi-supervised segmentation methods have achieved good performance in medical image segmentation. However, challenges such as blurred and unevenly distributed pixels at the boundaries of CNV images pose significant difficulties for semi-supervised CNV segmentation. These characteristics often lead to large intra-class variations. Therefore, the complex characteristics of CNV result in difficult-to-segment regions in OCT images, leading to significant uncertainty in CNV segmentation results. Most existing semi-supervised segmentation methods employ difficulty awareness to address the challenge of difficult-to-segment pixels. These methods typically characterize the importance of pixels through uncertainty, thereby guiding the network to focus on difficult-to-segment pixels. However, these methods have some limitations in semi-supervised CNV segmentation: 1) They have insufficient knowledge learning of difficult-to-segment pixels. In the uncertainty calculation, they do not fully leverage the classification confidence information from labeled data and the reliability of unlabeled data. 2) They did not consider the

Manuscript received April 19, 2021; revised August 16, 2021.

Jie Guo, Liangyun Sun, Lishan Qiao, Xiushan Nie, Xiaoming Xi are with the School of Computer Science and Technology, Shandong Jianzhu University, Jinan, 250101, China (e-mail: guojiesdu@163.com; sly315630@126.com; qlishan@163.com; niexiushan@163.com; fyzq10@126.com)

Jixin Yang is with the Kailin Environmental Protection Equipment, Heze, China (e-mail: linxiangxls@126.com)

Weicui Li is with Shandong Institute of Scientific and Technical Information, Jinan, 250101, China (e-mail: 15964028157@163.com)

Ying Guo is with the Key Laboratory of Computing Power Network and Information Security, Ministry of Education, Qilu University of Technology(Shandong Academy of Sciences), Jinan, 250014, China (e-mail: guoying@sdas.org)

Xinjian Chen is with the School of Electronics and Information Engineering, Soochow University, Jiangsu, 215006, China (e-mail: xjchen@suda.edu.cn)

Yilong Yin is with the School of Sofware, Shandong University, Jinan, 250101, China (e-mail: ylyin@sdu.edu.cn)

* These authors contributed equally to this work and should be considered co-first authors.

†Corresponding authors: Xiaoming Xi(e-mail:fyzq10@126.com); Xinjian Chen(e-mail: xjchen@suda.edu.cn)

difficulty information of pixels when generating pseudo-labels. This issue leads to the knowledge of difficult-to-segment pixels with confidence below the threshold failed to be learned. Thereby reducing the quality of pseudo-label generation and the accuracy of CNV segmentation. In Fig. 1(b) and Fig. 1(c), we respectively show the estimated uncertainty and the difficult-to-segment pixel regions. As shown in Fig. 1, the high-uncertainty regions in Fig. 1(b) cover only a subset of the difficult-to-segment pixels shown in Fig. 1(c).

To address the above problems, we propose a dual difficulty-aware adaptive pseudo-label learning method for semi-supervised CNV segmentation. In supervised learning, we propose a classification confidence-guided difficulty-aware module (CCDM). This module uses the Monte Carlo dropout method to estimate the uncertainty of segmentation results and combine it with classification confidence to enhance the importance of difficult-to-segment pixels during training. In unsupervised learning, we propose a model stability-guided difficulty-aware module (MSDM). This module uses stability scores to evaluate prediction reliability, and then integrates prediction uncertainty and reliability information to facilitate learning of difficult-to-segment pixels in unlabeled samples. Additionally, we introduce a difficulty-aware pseudo-labeling module (DPM). This module uses the uncertainty and reliability information of sample predictions as prior knowledge to perceive pixel-wise differences in segmentation difficulty. It dynamically adjusts thresholds based on the segmentation difficulty of pixels, thereby improving the utilization of difficult-to-segment pixels in unlabeled data.

The main contributions of the proposed method are summarized as follows:

1) We introduce pixel segmentation difficulty in semi-supervised CNV segmentation and propose a dual-difficulty-aware adaptive pseudo-labeling method for semi-supervised CNV segmentation.

2) For labeled and unlabeled data, we respectively propose classification confidence-guided and model stability-guided pixel segmentation difficulty-aware methods, enabling the network to focus on learning difficult-to-segment pixels during semi-supervised training.

3) We propose a difficulty-aware self-training method that considers the difficulty of pixel segmentation in the process of selecting pseudo-labels, enhancing the utilization of difficult-to-segment pixels in unlabeled data.

The remainder of the paper is organized as follows. Section II presents a review of related work. Sections III provide a detailed report on the proposed method. We report the dataset, experimental settings, and results in Section IV, and conclusion in Section V.

II. Related Work

A. CNV Segmentation

With the advancement of machine learning technology [56]–[59], a series of automatic CNV segmentation methods have been successively proposed. Traditional shallow machine learning and graph-based models have gradually been employed for CNV segmentation.

Lang et al. [6] applied a random forest classifier to the segmentation of macular cube images. The segmentation of retinal layers is achieved by classifying boundary pixels of the retinal layers to determine the segmentation boundaries. Xu et al. [7] utilized a layer-correlated hierarchical sampling strategy in a novel voxel classification-based approach to achieve automatic segmentation of three-dimensional retinal fluid. Li et al. [8] proposed a CNV segmentation method based on 3D HOG features and updated random forests, achieving automatic CNV segmentation in 3D spectral OCT images. Zhu et al. [9] employed a reaction-diffusion model to fit the growth of CNV over spatial and temporal dimensions, presenting a 3D OCT-based model for predicting CNV growth. Xiang et al. [10] proposed a fully supervised method for automatic segmentation of retinal layers in OCT images containing CNV.

Given the significant breakthroughs of deep learning in medical image segmentation, there is increasing attention from researchers on deep learning-based CNV segmentation methods. Researchers have undertaken many studies to address the challenges posed by CNV-induced retinal layer morphological changes, variability in size and shape of CNV regions, quality issues in OCT images and challenge of blurry CNV boundaries.

For CNV-induced retinal layer morphological changes, Meng et al. [14] proposed a multi-scale information fusion network based on a U-shaped structure. Shen et al. [15] introduced graph attention encoder and graph de-correlation modules into the U-Net framework, proposing a graph attention-based U-shaped segmentation network for retinal layer surface detection and CNV segmentation. Wang et al. [5] proposed a dynamic multi-level weighted segmentation network to achieve simultaneous segmentation of the retinal layer and CNV. Ke et al. [18] proposed a data-driven dual-path network that includes an expansive path and a guidance path, which guide the network's focus on learning CNV boundary information.

For variability in size and shape of CNV regions, Xi et al. [12] introduced an Information Attention Convolutional Neural Network for CNV segmentation. They incorporated an attention enhancement module into the U-Net framework to boost the network's ability to learn discriminative information. Wei et al. [16] proposed a Spatial Pyramid Pooling module that enables the model to extract contextual information at different scales. Zhang et al. [19] introduced a Cross Convolution module based on anisotropic convolutions into the 3D U-Net framework to learn features of CNV with different shapes.

For quality issues in OCT images and the challenge of blurry CNV boundaries, Wang et al. [11] introduced a two-stage CNN structure based on OCTA images, which achieves fully automated diagnosis and segmentation of CNV. Su et al. [13] enhanced the U-Net network by introducing a Discrete Dilated Module capable of extracting contrast information between background and foreground pixels. Chen et al. [17] proposed a reliable boundary-guided network for simultaneous segmentation of CNV regions and vessels.

B. Semi-supervised Learning

The success of semi-supervised learning methods in classification [20]–[23] has promoted the development of semi-supervised semantic segmentation methods [24]–[27]. Currently, the two mainstream semi-supervised learning methods are consistency-based semi-supervised learning [28]–[32] and self-training-based semi-supervised learning [33], [34].

Consistency-based semi-supervised learning methods enforce models to maintain consistent predictions for the same unlabeled data under different perturbations. French et al. [35] utilized CutOut and CutMix as strong augmentation techniques to perturb input images. CPS [36] employed two models with identical structures but different initializations during the semi-supervised learning process. Liu et al. [37] introduced feature perturbations to explore consistency at the feature level, enhancing the predictive performance of student networks by introducing a novel auxiliary teacher network. In semantic segmentation tasks, consistency-based semi-supervised learning methods have been demonstrated to outperform pseudo-labeling based semi-supervised learning methods [38].

The semi-supervised learning method based on self-training was originally used for image classification by utilizing a large amount of unlabeled data. Subsequently, it has been applied to image segmentation tasks [39]–[44]. Wang et al. [45] proposed a semi-supervised semantic segmentation method using unreliable pseudo-labels. Yuan et al. [42] pointed out that excessive perturbations in unlabeled images are disastrous for cleaning data distribution. Furthermore, to address the issue of class imbalance in pseudo-labels, He et al. [41] adjusted the class distribution between real and pseudo-labels. To further consider the reliability of unlabeled data, Yang et al. [46] proposed a more advanced self-training framework. This framework uses stability scores to select highly trustworthy unlabeled samples for subsequent iterative training.

Existing CNV segmentation methods do not take into account the differences in pixel segmentation difficulty. Although semi-supervised image segmentation methods consider pixel segmentation difficulty, they primarily rely on uncertainty to measure sample difficulty, leading to insufficient utilization of valuable information in hard samples. This leads to poor pseudo-label generation quality and reduces CNV segmentation accuracy. Therefore, we incorporate pixel difficulty into both the training of the segmentation network and the pseudo-label generation. Through classification confidence-guided and model stability-guided pixel segmentation difficulty-aware methods, the network can focus on learning difficult-to-segment pixels during semi-supervised training process. Through a difficulty-aware self-training strategy, the threshold for pseudo-labels is dynamically adjusted based on difficulty. This enhances the utilization of difficult-to-segment pixels in unlabeled data and improves the performance of CNV segmentation.

III. METHOD

A. Problem formulation

Medical image segmentation requires pixel-level labels. However, manually labeling images at the pixel level is time-consuming and costly. In reality, a large amount of medical image data is unlabeled. Therefore, we adopt a semi-supervised learning approach, leveraging both labeled and unlabeled data to learn richer semantic information. We formalize the problem of CNV semantic segmentation as follows.

Given labeled dataset $D^l = \{(x_1, y_1), \dots (x_n, y_n)\}$,which contains n labeled samples, each sample comprised of an input image x_i with dimensions $H \times W$ and its corresponding pixel-level label $y_i \in R^{H \times W \times C}$, where C is the number of classes and $H \times W$ is spatial dimension. Given unlabeled dataset $D^u = \{u_1, u_2, \dots, u_m\}$ including m unlabeled samples, with $m \gg n$. The purpose of CNV segmentation based on semi-supervised learning is to learn a model f based on the samples D, with $D = D^l \cup D^u$. We employ the standard U-Net as the Segmentation Model in our D2APL.

B. Proposed approach

In this paper, a novel semi-supervised segmentation framework is proposed to address the aforementioned challenges in the CNV segmentation. As shown in Fig. 2, the proposed framework contains three modules: Classification Confidence-guided Difficulty-aware Module (CCDM), Model Stability-guided Difficulty-aware Module (MSDM), and Difficulty-aware Pseudo-labeling Module (DPM). For labeled data, we utilize CCDM to perceive the segmentation difficulty of pixels. For unlabeled data, we generate pseudo-labels using DPM and then employ MSDM to perceive the segmentation difficulty of pixels. The difficulty-aware modules are used to guide the segmentation process to focus on challenging pixels.

1) Classification Confidence-guided Difficulty-aware Module: Inspired by uncertainty estimation methods based on Bayesian networks, the CCDM first utilizes the Monte Carlo Dropout method [50], [51] to estimate uncertainty. Specifically, during training, the model performs T forward passes under random dropout. Based on this, a set of predictions is obtained, and the information entropy of these predictions is calculated to approximate uncertainty. A higher entropy value indicates greater prediction uncertainty and thus greater difficulty in predicting the sample by the network. The calculating prediction entropy of the j-th pixel in the i-th labeled image can be summarized as:

$$\mu_i^c = \frac{1}{T}\sum_t P_t^c,\ r_i^j = \sum_c \mu_{ij}^c log \mu_{ij}^c \tag{1}$$

Where P_t^c is the probability of the c-th class in the t-th time prediction for the image x_i, c is the category of a pixel. The uncertainty of the labeled pixel is estimated in CCDM, and the uncertainty map of the labeled image is r_i.

Perceiving the segmentation difficulty of pixels only from the perspective of uncertainty will cause the network to ignore difficult pixels which have low uncertainty but are predicted incorrectly. To address this, we add the classification confidence in the process of computing sample importance weights. By combining classification confidence with uncertainty, it perceives the difficulty of samples and guides the network to focus on learning difficult-to-segment pixels. Specifically, the module utilizes cross-entropy loss to measure the distance

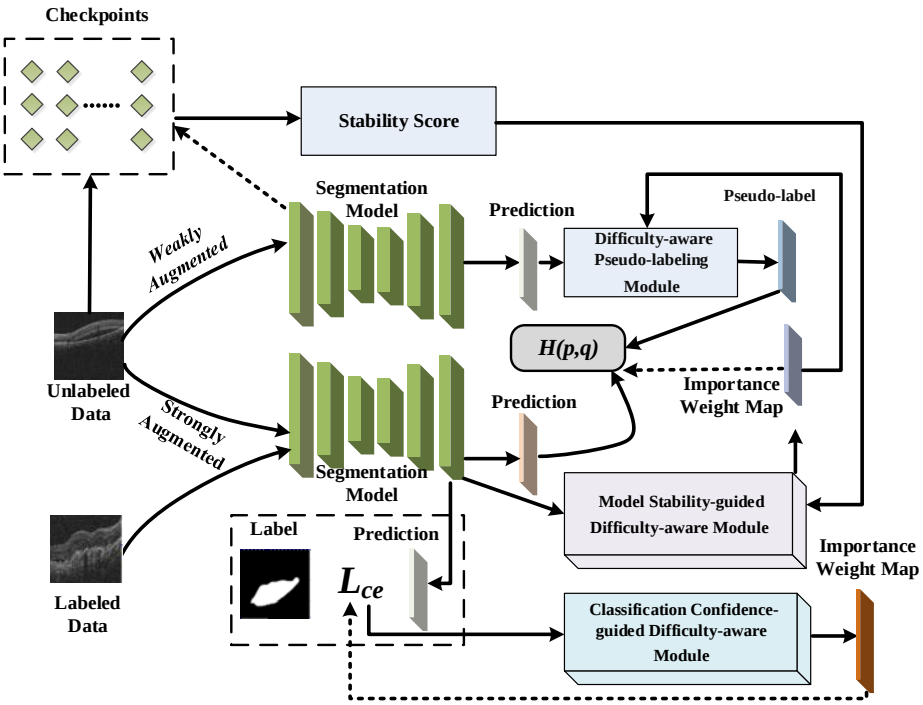


Fig. 2. The flowchart of the proposed D2APL framework. For labeled data, we utilize Classification Confidence-guided Difficulty-aware Module (CCDM) to perceive the segmentation difficulty of pixels. For unlabeled data, we generate pseudo-labels using Difficulty-aware Pseudo-labeling Module (DPM) and then employ Model Stability-guided Difficulty-aware Module (MSDM) to perceive the segmentation difficulty of pixels. The difficulty perception modules are used to guide the segmentation process to focus on challenging pixels.

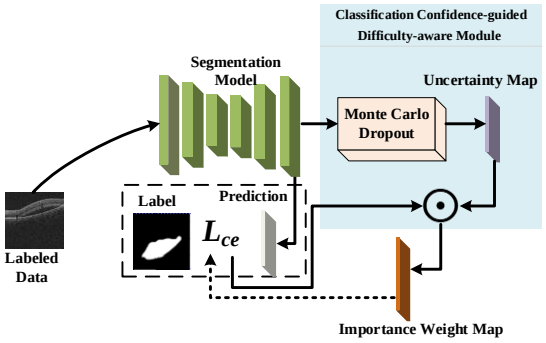


Fig. 3. The flowchart of CCDM. It utilizes the Monte Carlo Dropout method to estimate uncertainty. By combining label information with uncertainty, it perceives the difficulty of samples and guides the network to focus on learning difficult-to-segment pixels.

between each pixel's prediction and its corresponding label and multiplies this distance by the uncertainty estimate to obtain the importance weight of the pixel. The formula for calculating the importance weight of the pixel in labeled image x_i is:

$$w_i^j = e^{lr_i^j} \tag{2}$$

Where $l = e^L$, and L is the cross-entropy loss. $w_i^j \in R^{H\times W}$ is the importance weight of the pixel in labeled image x_i. In the supervised learning stage, the importance weight map w_i is used for the supervised loss L_s. The L_s can be summarized as:

$$L_s = \frac{1}{B_l}\sum_{i=1}^{B_l}\left(\sum_{j=1}^{H\times W} w_i^j D\left(y_i^j, f\left(\alpha\left(x_i^j\right);\theta\right)\right)/(H\times W)\right) \tag{3}$$

Where B_l is the size of a mini-batch of labeled images, D$(\cdot)$ is the cross-entropy loss, and $\alpha(\cdot)$ is the weakly augmented operation. In all of our experiments, weak augmentation is a standard flip-and-shift augmentation strategy. x_i^j is the j-th pixel in labeled image x_i, y_i^j is the label of x_i^j. The framework of CCDM is shown in Fig.3. Guided by the importance weight map w_i and classification confidence, the network pays more attention to learning difficult-to-segment pixels in the labeled data, thereby improving the segmentation performance of difficult-to-segment pixels in OCT images of labeled CNV.

2) Model Stability-guided Difficulty-aware Module: In the model stability-guided difficulty-aware module, the uncertainty of unlabeled data is initially estimated using the Monte Carlo Dropout method, as shown in the following equation:

$$\rho_i^c = \frac{1}{T}\sum_t P_t^c,\ z_i^j = \sum_c \rho_{ij}^c log\rho_{ij}^c \tag{4}$$

Where P_t^c is the probability of the c-th class in the t-th time prediction for the image u_i, c is the category of pixel. The uncertainty of an unlabeled pixel is estimated in MSDM, and the uncertainty map of an unlabeled image is z_i.

To mitigate the impact of noise and pseudo-labels on the network, traditional semi-supervised segmentation methods only calculate the importance weight map of unlabeled data based on the uncertainty map, thereby guiding the network to

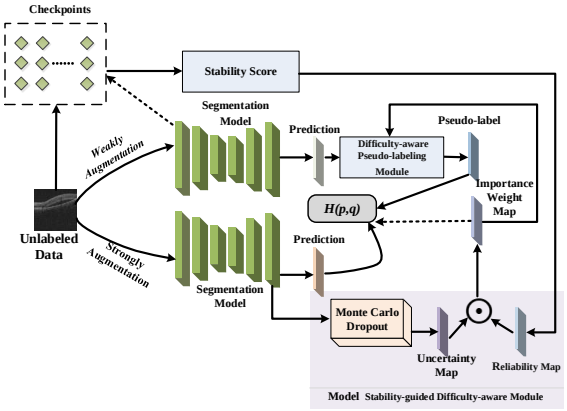


Fig. 4. The flowchart of MSDM. It utilizes the Monte Carlo Dropout method to estimate uncertainty and adopts a distributed hierarchical clustering prototype learning method and checkpoints to compute stability scores. By combining stability scores with uncertainty, it perceives the difficulty of samples and guides the network to focus on learning difficult-to-segment pixels with low uncertainty but incorrectly predicted.

focus more on learning pixels with low uncertainty. However, this approach overlooks learning difficult-to-segment pixels with high prediction uncertainty in unlabeled data, which is not conducive to learning knowledge from difficult-to-segment pixels. To enhance the segmentation performance of difficult-to-segment pixels in unlabeled data, we add reliability estimation in MSDM. By computing stability scores of pixels to assess the reliability of pixel prediction results, and fusing the obtained reliability map with the uncertainty map to calculate the importance weight of unlabeled data, the network is guided to focus more on learning difficult-to-segment pixels in unlabeled data. The framework of MSDM is shown in Fig.4.

In MSDM, the pixel-level stability score is calculated to measure the reliability of the prediction results for each pixel. Since the calculation of stability scores requires historical prediction results as a basis, we add MSDM and DPM (in subsection 3.2.3) into the proposed D2APL model after a period of training. Specifically, we use a manually set time point H to control the timing of introducing MSDM and DPM into the model. When $EP > H$, MSDM and DPM are added to the model, where EP represents the current training epoch. After each training epoch, the model saves a checkpoint, resulting in a sequence of checkpoints $\{C_{ep}\}_{ep=1}^{EP}$. We use the saved checkpoints to predict the unlabeled image u_i, the results of the prediction is $\{\hat{y}_i^{ep}\}_{ep=1}^{EP}$.

To compute more accurate stability scores, we adopt a distributed hierarchical clustering prototype learning method to obtain $K-1$ representative prediction results from $\{\hat{y}_i^{ep}\}_{ep=1}^{EP}$. Specifically, all prediction results except $\hat{y}_i^{EP}$ are divided into $K-1$ groups, and hierarchical clustering is performed on each group to obtain $K-1$ representative prediction results. These $K-1$ prediction results are considered as prototypes. As the model tends to converge and achieve optimal performance in the later stages of training, we calculate the Kullback-Leibler (KL) divergence between prototypes and templates (the prediction results of unlabeled images from the last checkpoint), to measure the similarity between prototypes and templates. The collection consisting of $K-1$ prototypes and the template can be represented as $\{V_j\}_{j=1}^{K}$. The value of KL divergence serves as a measure of stability for quantifying the similarity between the learned prototypes and the predefined templates. The formula for calculating stability scores is as follows:

$$s_i = \sum_{j=1}^{K-1} V_K log\left(\frac{V_K}{V_j}\right) \tag{5}$$

MSDM estimates the reliability of prediction results for unlabeled data at the pixel level, resulting in a reliability map s_i for unlabeled image u_i. The reliability score s_i^j of the j-th pixel in the unlabeled image u_i and the uncertainty score z_i^j are further utilized for computing the importance weights, as shown in the following equation:

$$w_i^j = z_i^j e^{s_i^j} \tag{6}$$

The importance weight map w_i of unlabeled images distinguishes the segmentation difficulty of pixels from the perspectives of prediction uncertainty and reliability. It guides the network to focus on learning reliable but difficult-to-segment pixels in unlabeled images, thereby enhancing the segmentation performance of difficult-to-segment pixels in unlabeled images. Furthermore, we designed a difficulty-aware unsupervised loss, as shown in the following equation:

$$L_u = \frac{1}{B_u}\sum_{i=1}^{B_u}\left(\frac{\sum_{j=1}^{H\times W} w_i^j 1\left(\max\left(q_i^j\right)\geq\tau\right)D\left(\hat{q}_i^j, p_i^j\right)}{H\times W}\right) \tag{7}$$

Where, B_u is the size of a mini-batch of unlabeled images, $q_i^j = f\left(\alpha\left(u_i^j\right);\theta\right)$ is the predicted probability of pixels in weakly augmented images. $\hat{q_i^j} = argmax\left(q_i^j\right)$ is the corresponding pseudo-label. u_i^j is the j-th pixel in the unlabeled image u_i, $p_i^j = f\left(A\left(u_i^j\right);\theta\right)$ is the predicted probability of pixels in strongly augmented images, A(·) is the strongly augmented operation. The strongly augmented operation is RandAugment and Cutout. τ represents the confidence threshold, where 1(·) denotes the conditional function. Pseudo-labels are generated only when the predicted confidence of a pixel is greater than the corresponding threshold.

3) Difficulty-aware Pseudo-labeling Module: The threshold τ is uniformly set for all pixels in the image and does not consider the variability of pixel segmentation difficulty. Difficult-to-segment pixels with confidence below the threshold cannot be learned by the network, reducing the network's utilization of difficult-to-segment pixels in unlabeled data. To address this issue, we propose a Difficulty-Aware Pseudo Labeling Module (DPM) to obtain pseudo-labels. Since the importance weight w_i of unlabeled data perceives the segmentation difficulty of pixels from the perspectives of prediction stability and uncertainty, DPM further adjusts the confidence threshold τ using w_i, the formula is as follows:

$$\tau_{ep}^{ij} = \begin{cases} \frac{\tau}{w_i^j}, & \text{if } ep < T_{max} \\ \tau, & \text{otherwise} \end{cases} \tag{8}$$

Where, τ is the initial threshold, and T_{max} is the maximum number of training epochs for dynamic threshold adjustment. During unsupervised training, difficult-to-segment pixels with reliable predictions are assigned smaller thresholds, thereby increasing the network's utilization of reliable difficult-to-segment pixels in unlabeled data. The pseudo-labels obtained through DPM will be merged into the labeled dataset for subsequent iterative training. After adding the DPM, the unsupervised loss is shown as follows:

$$L_u = \frac{1}{B_u}\sum_{i=1}^{B_u}\left(\frac{\sum_{j=1}^{H\times W} w_i^j 1\left(\max\left(q_i^j\right) \geq \tau_{ep}^{ij}\right) D\left(\hat{q}_i^j, p_i^j\right)}{H\times W}\right) \tag{9}$$

The unsupervised loss L_u and the supervised loss L_s jointly constitute the objective loss L. The objective loss is shown as follows:

$$L \ = L_s \ + \lambda L_u \tag{10}$$

Where λ is a balancing factor.

IV. Experiments

We evaluate the effectiveness of the proposed method from multiple aspects. Through ablation studies, we demonstrate the effectiveness of the CCDM, MSDM, and DPM modules. We also conduct ablation studies on the time point H and the number of prototypes K. Furthermore, we compare our proposed method with other semi-supervised segmentation methods and CNV segmentation methods, providing evidence of its effectiveness.

The dataset includes 3510 images, with 2174 images for the training set and 1336 images for the test set. The ratio of unlabeled to labeled data in the training set is approximately 2:1. In our experiments, we set the time step H to 20, and the value of K is calculated as $\lceil ep/5 \rceil$, where $\lceil \cdot \rceil$ denotes the ceiling operation. The learning rate, batch size, and number of epochs are set to 0.001, 16, and 150, respectively. All experiments are conducted using the Adam optimizer in PyTorch under Python 3.6.9.

A. Ablation Study

This section is to validate the effectiveness of the CCDM, MSDM, and DPM modules. From Table I and Fig.5, we can observe that the absence of the CCDM and MSDM modules leads to a significant decrease in segmentation accuracy. Notably, the segmentation accuracy is the lowest when the CCDM module is missing. This is because guiding uncertainty estimation through category confidence can direct the network to focus on learning difficult-to-segment pixels that have low uncertainty but are predicted incorrectly. The absence of this module results in some pixels being predicted inaccurately. Additionally, uncertainty estimation and model stability can guide the network to focus on learning difficult-to-segment pixels while ensuring model stability. The lack of this module leads to model instability, which affects segmentation performance. Furthermore, the segmentation accuracy also declines when the DPM module is missing. This is because, during the generation of pseudo-labels, thresholds are not set based on the segmentation difficulty of the pixels. This results in a decrease in the quality of pseudo-labels, thereby impacting segmentation performance.

TABLE I
THE EFFECTIVENESS VALIDATION OF THE MODULES IN D2APL

Method			TPR	Dice	IoU
CCDM	MSDM	DPM			
✓	✓	-	0.8264	0.8153	0.7221
✓	-	✓	0.8121	0.8044	0.7090
-	✓	✓	0.7929	0.7997	0.7047
✓	✓	✓	**0.8615**	**0.8337**	**0.7489**

TABLE II
ABLATION STUDY ON THE TIME POINT

The value of H	20	40	60	80
IoU	**0.7489**	0.7352	0.7207	0.7064

TABLE III
ABLATION STUDY ON THE TOTAL NUMBER OF THE PROTOTYPES

The value of K	$\lceil ep/15 \rceil$	$\lceil ep/10 \rceil$	$\lceil ep/6 \rceil$	$\lceil ep/5 \rceil$
IoU	0.7302	0.7243	0.7409	**0.7489**

TABLE IV
COMPARISON OF UNCERTAINTY ESTIMATION METHODS IN D2APL

method	TPR	Dice	IoU
D2APL(TEU)	0.8106	0.7996	0.7040
D2APL(Correctness)	0.8403	0.8229	0.7307
D2APL(MC Dropout)	**0.8615**	**0.8337**	**0.7489**

We conduct ablation studies on the time point H, which is used to control the timing of introducing MSDM and DPM. As shown in Table II, setting H to 20 proves to be effective. Notably, as H increases, the IoU values exhibit a decreasing trend, indicating that an earlier introduction of MSDM and DPM leads to better results. In summary, an earlier introduction of MSDM and DPM ensures that the model progressively adapts to pixel segmentation difficulty from the initial training stages. This results in better pseudo-label selection and higher segmentation performance. We also conduct ablation studies on K, which represents the total number of the prototypes in MSDM for each sample. As shown in Table III, setting K to $\lceil ep/5 \rceil$proves to be effective, where ep is the number of epochs.

In order to validate the necessity of using Monte Carlo Dropout to estimate uncertainty, we have conducted a comparative analysis in our experiments. Specifically, we compared the Monte Carlo Dropout (MC Dropout) method with the Traditional Entropy-based Uncertainty (TEU) estimation method and Correctness [60] within our framework. From Table IV, we can observe that the Monte Carlo Dropout method achieves the best performance across all evaluation metrics. The main reason lies in the fact that TEU and Correctness rely on a single prediction to estimate uncertainty. In contrast, MC Dropout performs multiple stochastic forward passes with dropout, which effectively captures variations in the model's confidence, particularly in challenging areas. This makes uncertainty estimation more robust.

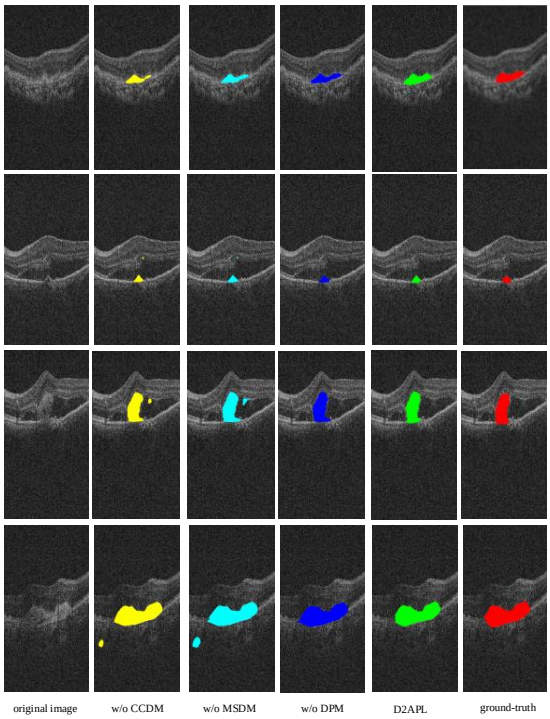


Fig. 5. The segmentation results of the ablation study.

B. Comparison with semi-supervised methods

D2APL is compared with existing semi-supervised segmentation methods to evaluate the effectiveness of the proposed approach. Methods such as PseudoSeg [43] , PC2Seg [44] , CPS [36] , ST++ [46] , and Unimatch [49] leverage a large amount of unlabeled data through pseudo-supervision. However, they did not consider the difficulty of pixel segmentation during model training. This results in valuable information from a significant number of difficult-to-segment pixels being overlooked, thereby limiting the improvement of model performance in segmenting difficult pixels.

D2APL proposes dual difficulty-aware mechanisms, CCDM and MSDM, for labeled and unlabeled data respectively. These mechanisms perceive the segmentation difficulty of pixels based on the classification confidence and reliability of prediction results, guiding the network to focus more on learning difficult-to-segment pixels, thus improving the segmentation performance. Furthermore, compared to traditional semi-supervised segmentation methods, D2APL utilizes the difficulty-aware results as prior information to dynamically adjust threshold. This enables effective utilization of most unlabeled samples with low confidence but high reliability, ultimately demonstrating good performance on difficult-to-segment pixels. Table V shows the performance of D2APL compared to other semi-supervised segmentation methods in terms of TPR, Dice, and IoU. From Table V, it can be observed that D2APL achieves approximately 7% improvement in IoU compared to Unimatch [49]. Furthermore, D2APL demonstrates the best performance among other semi-supervised segmentation methods in terms of TPR, Dice, and IoU. Fig.6 illustrates the segmentation results of different semi-supervised segmentation methods. Among them, UCMT and MiDSS are semi-supervised segmentation methods for medical images.

TABLE V
THE COMPARISON BETWEEN D2APL AND OTHER SEMI-SUPERVISED METHODS

Method	TPR	Dice	IoU
$PseudoSeg^{[43]}$	0.7112	0.7061	0.6160
$PC2Seg^{[44]}$	0.7207	0.7141	0.6198
$CPS^{[36]}$	0.7573	0.7534	0.6571
$ST++^{[46]}$	0.7796	0.7632	0.6643
$UCMT^{[52]}$	0.8063	0.7650	0.6702
$Unimatch^{[49]}$	0.8192	0.7690	0.6785
$MiDSS^{[53]}$	0.8309	0.8097	0.7109
D2APL	**0.8615**	**0.8337**	**0.7489**

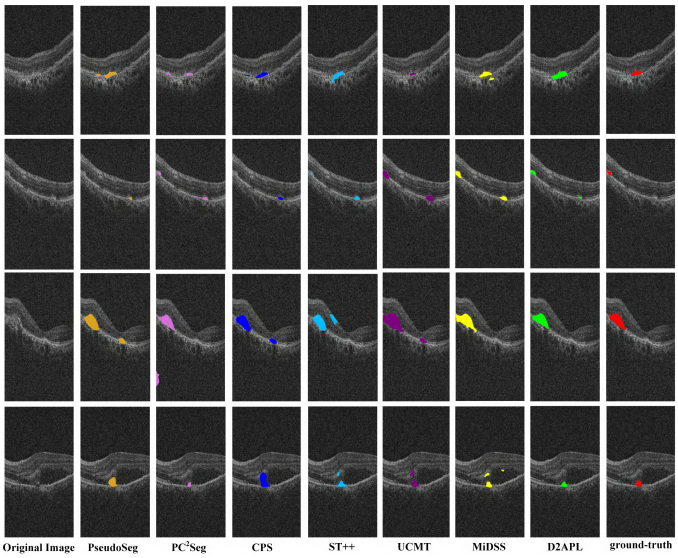


Fig. 6. The segmentation results of different semi-supervised segmentation methods.

C. Comparison with CNV segmentation methods

Compared to other CNV segmentation methods such as ACM-LSP [6], MS-CNN [48], NNCGS [10], and IA-net [12], as well as the baseline segmentation network U-Net [47], D2APL focuses more on learning difficult-to-segment pixels. Additionally, by introducing DPM, D2APL effectively utilizes a large amount of unlabeled data. Table VI presents the segmentation performance comparison between D2APL and other CNV segmentation methods as well as the baseline network. Overall, D2APL exhibits the best performance across all three evaluation metrics: TPR, Dice, and IoU. Fig.7 illustrates the segmentation results of different CNV segmentation methods.

D. Effectiveness and Intermediate Process Visualization of the CCDM Module

For labeled data, the difficulty-aware model guided by classification confidence combines misclassified pixels with pixel uncertainty to produce importance weights. These weights help the model focus on learning difficult-to-segment pixels. Fig.8 illustrates the effects of the intermediate processes involved in the CCDM module. As shown in Fig.8, the importance

TABLE VI
THE COMPARISON BETWEEN D2APL AND OTHER CNV SEGMENTATION METHODS

Method	Label Proportion	TPR	Dice	IoU
$ACM-LSP^{[6]}$	33%	0.5531	0.3921	0.4963
$MS-CNN^{[48]}$	33%	0.5962	0.6214	0.5107
$NNCGS^{[10]}$	33%	0.6253	0.6534	0.5430
$U-net^{[47]}$	33%	0.6542	0.6764	0.5793
$IA-net^{[12]}$	33%	0.7025	0.6954	0.5962
D2APL	33%	**0.8615**	**0.8337**	**0.7489**

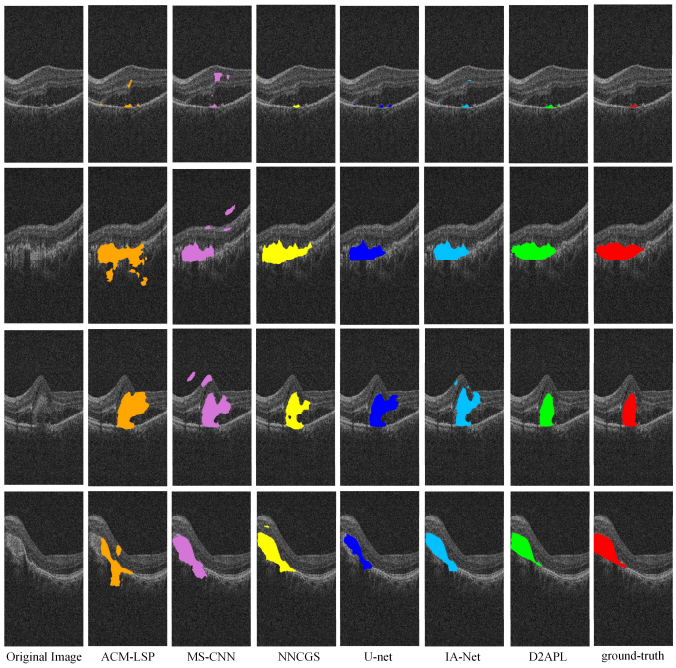


Fig. 7. The segmentation results of different CNV segmentation methods.

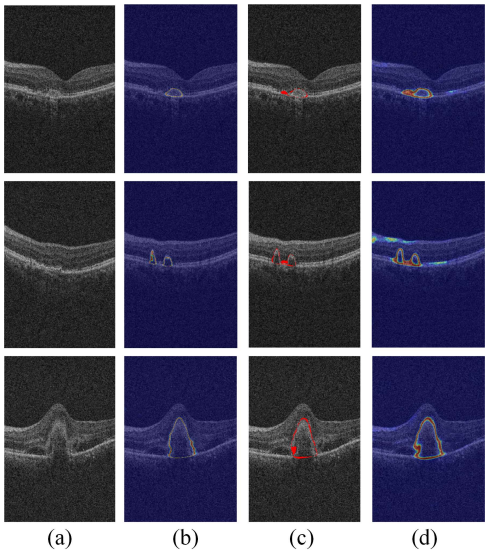


Fig. 8. Intermediate Process Visualization of the CCDM Module. (a) Original image (b) Uncertainty map (c) Misclassified regions (d) Importance weight map.

weight map (Fig.8 (d)) further enhances the weights of pixels in the misclassified regions with low prediction uncertainty (Fig.8 (c)) on the uncertainty map (Fig.8 (b)). This effectively guides the network to focus more on learning from difficult samples.

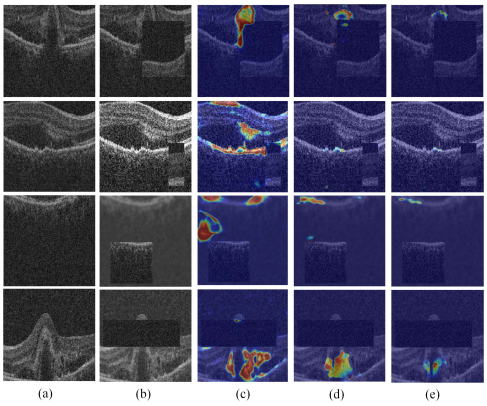


Fig. 9. Intermediate Process Visualization of the MSDM Module. (a) Weakly augmentation image (b) Strongly augmentation image (c) Uncertainty map (d) Reliability map (e) Importance weight map.

E. Effectiveness and Intermediate Process Visualization of the MSDM Module

For unlabeled data, the stability-guided difficulty-aware model combines pixel classification stability with pixel uncertainty to produce importance weights. These weights help guide the model to focus on learning from difficult-to-segment pixels in the unlabeled data. Fig.9 illustrates the effects of the intermediate processes involved in the MSDM module. Fig.9(a) shows the image after weak augmentation, while Fig.9(b) shows the image after strong augmentation. By comparing the uncertainty map (Fig.9(c)), reliability map (Fig.9(d)), and importance weight map (Fig.9(e)), it is evident that samples in the unlabeled data with high prediction uncertainty and high prediction reliability are assigned higher weights. This effectively guides the network to focus on learning from difficult samples in the unlabeled data.

V. CONCLUSION

Since CNV has complex features such as blurred and unevenly distributed pixels on the edges, there are difficult-to-segment pixels in CNV segmentation. To guide the network to focus on difficult-to-segment pixels in CNV segmentation, we proposed a dual difficulty-aware adaptive pseudo-label learning method for semi-supervised CNV segmentation. For labeled data, we proposed a difficulty-aware method guided by classification confidence. For unlabeled data, we proposed a difficulty-aware method guided by model stability. Finally, we proposed a difficulty-aware self-training strategy to dynamically adjust the threshold of pseudo-labels according to the difficulty, thereby improving the utilization of difficult-to-segment pixels in unlabeled data and the performance of CNV segmentation.Considering the rapid development of three-dimensional (3D) optical coherence tomography (OCT) imaging, we plan to extend our method to 3D OCT data.

Unlike 2D fundus images, 3D OCT provides richer spatial information, which could help in more accurate CNV segmentation. However, the increased data complexity and computational cost present new challenges. Future research will explore efficient 3D feature extraction techniques and adaptive pseudo-labeling strategies suitable for volumetric data.

ACKNOWLEDGEMENT

This work was supported in part by the Natural Science Foundation of China under Grant U23A20389, in part by the Science and Technology Innovation Program for Distinguished Young Scholars of Shandong Province Higher Education Institutions under Grant 2021KJ036, in part by the Major Basic Research Project of the Natural Science Foundation of Shandong Province under Grant ZR2021ZD15, in part by the Natural Science Foundation of Shandong Province under Grant ZR2022MF285, ZR2021QF119, in part by the Shandong Provincial Natural Science Foundation for Distinguished Young Scholars under Grant ZR2021JQ26, in part by the Taishan Scholar Project of Shandong Province under Grant tsqn202103088, in part by the special funds for distinguished professors of Shandong Jianzhu University and the Natural Science Foundation of Shandong Province under Grant ZR2020QF029, National Science Foundation for Young Scientists of China under Grant 62402296, and in part by the National Natural Science Foundation of China under Grant 62102235, 82001987, and 62176141,in part by the Open project of Key Laboratory of Computer Power Internet and Information Security of Ministry of Education under Grant 2023ZD030.

REFERENCES

[1] N. Congdon, B. O'Colmain, C.C. Klaver, R. Klein, B. Muñoz, D.S. Friedman, J. Kempen, H.R. Taylor, and P. Mitchell, "Causes and prevalence of visual impairment among adults in the United States," *Arch. Ophthal.*, vol. 122, no. 4, pp. 477–485, 2004.

[2] D.S. Friedman, B.J. O'Colmain *et al.*, "Prevalence of age-related macular degeneration in the United States," *Arch. Ophthal.*, vol. 122, no. 4, pp. 564–572, 2004.

[3] L.S. Lim, P. Mitchell, J.M. Seddon, F.G. Holz, and T.Y. Wong, "Age-related macular degeneration," *Lancet*, vol. 379, no. 9827, pp. 1728–1738, 2012.

[4] G. Trichonas and P.K. Kaiser, "Optical coherence tomography imaging of macular oedema," *Brit. J. Ophthalmol.*, vol. 98, no. Suppl 2, pp. ii24–ii29, 2014.

[5] L. Wang, M. Wang, T. Wang, Q. Meng, Y. Zhou, Y .Peng, W. Zhu, Z. Chen and X. Chen, "DW-Net: Dynamic multi-hierarchical weighting segmentation network for joint segmentation of retina layers with choroid neovascularization," *Front. Neurosci-switz*, vol. 15, p. 797166, 2021.

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

[7] X. Xu, K. Lee, L. Zhang, M. Sonka and M.D. Abramoff, "Stratified sampling voxel classification for segmentation of intraretinal and subretinal fluid in longitudinal clinical OCT data," *IEEE T. Med. Imaging*, vol. 34, no. 7, pp. 1616–1623, 2015.

[8] Y. Li, S. Niu, Z. Ji, W. Fan, S. Yuan and Q. Chen, "Automated choroidal neovascularization detection for time series SD-OCT images," in *Proc. Int. Conf. Med. Image Comput. Comput.-Assist. Intervention*,Granada, Spain , 2018, pp. 381–388.

[9] S. Zhu, F. Shi, D. Xiang,W. Zhu, H. Chen, X. Chen, "Choroid neovascularization growth prediction with treatment based on reaction-diffusion model in 3-D OCT images," *IEEE J. Biomed. Health* , vol. 21, no. 6, pp. 1667–1674, 2017.

[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 T. Image Process.*, vol. 27, no. 12, pp. 5880–5891, 2018.

[11] J. Wang, T.T. Hormel, L. Gao, P. Zang, Y. Guo, X. Wang, S.T. Bailey and Y. Jia, "Automated diagnosis and segmentation of choroidal neovascularization in OCT angiography using deep learning," *Biomed. Opt. Express*, vol. 11, no. 2, pp. 927–944, 2020.

[12] X. Xi, X. Meng, Z. Qin, X. Nie, Y. Yin and X. Chen, "IA-Net: Informative attention convolutional neural network for choroidal neovascularization segmentation in OCT images," *Biomed. Opt. Express*, vol. 11, no. 11, pp. 6122–6136, 2020.

[13] J. Su, X. Chen, Y. Ma, W. Zhu and F. Shi, "Segmentation of choroid neovascularization in OCT images based on convolutional neural network with differential amplification blocks," *Medical Imaging 2020: Image Processing*, vol. 11313, pp. 491–497, 2020.

[14] Q. Meng, L. Wang, T. Wang, M. Wang, W. Zhu, F. Shi, Z. Chen and X. Chen, "MF-Net: Multi-scale information fusion network for CNV segmentation in retinal OCT images," *Front. Neurosci-switz*, vol. 15, p. 743769, 2021.

[15] Y. Shen, J. Li, W. Zhu, K. Yu, M. Wang, Y. Peng, Y. Zhou, L. Guan and X Chen, "Graph attention U-Net for retinal layer surface detection and choroid neovascularization segmentation in OCT images," *IEEE T. Med. Imaging*, vol. 42, no. 11, pp. 3140–3154, 2023.

[16] W. Feng, M. Duan, B. Wang, Y. Du, Y. Zhao, B. Wang, L. Zhao, Z. Ge and Y. Hu, "Automated segmentation of choroidal neovascularization on optical coherence tomography angiography images of neovascular age-related macular degeneration patients based on deep learning," *Journal of Big Data*, vol. 10, no. 1, p. 111, 2023.

[17] T. Chen, Y. Zhao, L. Mou, D. Zhang, X. Xu, M. Liu, H. Fu and J. Zhang, "RBGNet: Reliable Boundary-Guided Segmentation of Choroidal Neovascularization," in *Proc. Int. Conf. Med. Image Comput. Comput.-Assist. Intervention*, Vancouver, Canada, 2023, pp. 163–172.

[18] J. Ke, Z. Ji, Q. Chen, W. Fan and S. Yuan, "Data-Dependence Dual Path Network for Choroidal Neovascularization Segmentation in SD-OCT Images," in *Proc. Int. Conf. Image Graphics*, Haikou, Zhejiang, China, 2021, pp. 535–546.

[19] X. Zhang, M. Li, Y. Zhang, S. Yuan and Q. Chen, "Choroidal Neovascularization Segmentation Based on 3D CNN with Cross Convolution Module," in *Proc. BenchCouncil Int. Federated Intell. Comput. Block Chain Conf.*, Qingdao, Shandong, China. 2020, pp. 14–21.

[20] X. Cui, C. Wang, D. Ren, Y. Chen and P. Zhu, "Semi-supervised image deraining using knowledge distillation," *IEEE T. Circ. Syst. Vid.*, vol. 32, no. 12, pp. 8327–8341, 2022.

[21] X. Wang, L. Fu, Y. Zhang, Y. Wang and Z. Li, "MMatch: Semi-supervised discriminative representation learning for multi-view classification," *IEEE T. Circ. Syst. Vid.*, vol. 32, no. 9, pp. 6425–6436, 2022.

[22] R. Xu, L. Xing, S. Shao, L. Zhao, B. Liu, W. Liu and Y. Zhou, "GCT: Graph co-training for semi-supervised few-shot learning," *IEEE T. Circ. Syst. Vid.*, vol. 32, no. 12, pp. 8674–8687, 2022.

[23] D. Li, Y. Liu and L. Song, "Adaptive weighted losses with distribution approximation for efficient consistency-based semi-supervised learning," *IEEE T. Circ. Syst. Vid.*, vol. 32, no. 11, pp. 7832–7842, 2022.

[24] I. Alonso, A. Sabater, D. Ferstl, L. Montesano and A.C. Murillo, "Semi-supervised semantic segmentation with pixel-level contrastive learning from a class-wise memory bank," in *Proc. IEEE/CVF Int. Conf. Comput. Vision*, online, 2021, pp. 8219–8228.

[25] Z. Zhao, S. Long, J. Pi, J. Wang and L. Zhou, "Instance-specific and model-adaptive supervision for semi-supervised semantic segmentation," in *Proc. IEEE/CVF Conf. Comput. Vision Pattern Recognit.*, online, 2023, pp. 23705–23714.

[26] Z. Zhao, L. Yang, S. Long, J. Pi, L. Zhou and J. Wang, "Augmentation matters: A simple-yet-effective approach to semi-supervised semantic segmentation," in *Proc. IEEE/CVF Conf. Comput. Vision Pattern Recognit.*, Paris, France, 2023, pp. 11350–11359.

[27] Y. Zhong, B. Yuan, H. Wu, Z. Yuan, J. Peng and Y.X. Wang, "Pixel contrastive-consistent semi-supervised semantic segmentation," in *Proc. IEEE/CVF Int. Conf. Comput. Vision*, online, 2021, pp. 7273–7282.

[28] J. Fan, B. Gao, H. Jin and L. Jiang, "UCC: Uncertainty guided cross-head co-training for semi-supervised semantic segmentation," in *Proc. IEEE/CVF Conf. Comput. Vision Pattern Recognit.*, New Orleans, Louisiana, USA, 2022, pp. 9947–9956.

[29] X. Wang, D. Kihara, J. Luo and G.J. Qi, "ENAET: A self-trained framework for semi-supervised and supervised learning with ensemble transformations," *IEEE T. Image Process.*, vol. 30, pp. 1639–1647, 2020.

[30] X. Lai, Z. Tian, L. Jiang, S. Liu, H. Zhao, L. Wang and J. Jia, "Semi-supervised semantic segmentation with directional context-aware

consistency," in *Proc. IEEE/CVF Conf. Comput. Vision Pattern Recognit.*, Nashville, TN, USA, 2021, pp. 1205–1214.
[31] V. Olsson, W. Tranheden, J. Pinto and L. Svensson, "ClassMix: Segmentation-based data augmentation for semi-supervised learning," in *Proc. IEEE/CVF Winter Conf. Appl. Comput. Vision*, online, 2021, pp. 1369–1378.
[32] Y. Zhou, H. Xu, W. Zhang, B. Gao and P.A. Heng, "C3-SemiSeg: Contrastive semi-supervised segmentation via cross-set learning and dynamic class-balancing," in *Proc. IEEE/CVF Int. Conf. Comput. Vision*, online, 2021, pp. 7036–7045.
[33] D. Kwon and S. Kwak, "Semi-supervised semantic segmentation with error localization network," in *Proc. IEEE/CVF Conf. Comput. Vision Pattern Recognit.*, New Orleans, Louisiana, USA, 2022, pp. 9957–9967.
[34] Y. Momoki, A. Ichinose, Y. Shigeto, U. Honda, K. Nakamura and Y. Matsumoto, "Characterization of pulmonary nodules in computed tomography images based on pseudo-labeling using radiology reports," *IEEE T. Circ. Syst. Vid.*, vol. 32, no. 5, pp. 2582–2591, 2021.
[35] G. French, S. Laine, T. Aila, M. Mackiewicz and G. Finlayson, "Semi-supervised semantic segmentation needs strong, varied perturbations," *arXiv preprint arXiv:1906.01916*, 2019.
[36] X. Chen, Y. Yuan, G. Zeng and J. Wang, "Semi-supervised semantic segmentation with cross pseudo supervision," in *Proc. IEEE/CVF Conf. Comput. Vision Pattern Recognit.*, online, 2021, pp. 2613–2622.
[37] Y. Liu, Y. Tian, Y. Chen, F. Liu, V. Belagiannis and G. Carneiro, "Perturbed and strict mean teachers for semi-supervised semantic segmentation," in *Proc. IEEE/CVF Conf. Comput. Vision Pattern Recognit.*, New Orleans, Louisiana, USA, 2022, pp. 4258–4267.
[38] L. Hoyer, D. Dai and L. Van Gool, "DAFormer: Improving network architectures and training strategies for domain-adaptive semantic segmentation," in *Proc. IEEE/CVF Conf. Comput. Vision Pattern Recognit.*, New Orleans, Louisiana, USA, 2022, pp. 9924–9935.
[39] L. Hoyer, D. Dai and L. Van Gool, "HRDA: Context-aware high-resolution domain-adaptive semantic segmentation," in *Proc. Eur. Conf. Comput. Vision*, Amsterdam, Netherlands, 2022, pp. 372–391.
[40] D. Guan, J. Huang, A. Xiao and S. Lu, "Unbiased subclass regularization for semi-supervised semantic segmentation," in *Proc. IEEE/CVF Conf. Comput. Vision Pattern Recognit.*, New Orleans, Louisiana, USA, 2022, pp. 9968–9978.
[41] R. He, J. Yang and X. Qi, "Re-distributing biased pseudo labels for semi-supervised semantic segmentation: A baseline investigation," in *Proc. IEEE/CVF Int. Conf. Comput. Vision*, online, 2021, pp. 6930–6940.
[42] J. Yuan, Y. Liu, C. Shen, Z. Wang and H. Li, "A simple baseline for semi-supervised semantic segmentation with strong data augmentation," in *Proc. IEEE/CVF Int. Conf. Comput. Vision*, online, 2021, pp. 8229–8238.
[43] Y. Zou, Z. Zhang, H. Zhang, C.L. Li, X. Bian, J.B. Huang and T. Pfister, "PseudoSeg: Designing pseudo labels for semantic segmentation," *arXiv preprint arXiv:2010.09713*, 2020.
[44] Y. Zhong, B. Yuan, H. Wu, Z. Yuan, J. Peng and Y.X. Wang, "Pixel contrastive-consistent semi-supervised semantic segmentation," in *Proc. IEEE/CVF Int. Conf. Comput. Vision*, online, 2021, pp. 7273–7282.
[45] Y. Wang, H. Wang, Y. Shen, J. Fei, W. Li, G. Jin, L. Wu, R. Zhao and X. Le, "Semi-supervised semantic segmentation using unreliable pseudo-labels," in *Proc. IEEE/CVF Conf. Comput. Vision Pattern Recognit.*, New Orleans, Louisiana, USA, 2022, pp. 4248–4257.
[46] L. Yang, W. Zhuo, L. Qi, Y. Shi and Y. Gao, "ST++: Make self-training work better for semi-supervised semantic segmentation," in *Proc. IEEE/CVF Conf. Comput. Vision Pattern Recognit.*, New Orleans, Louisiana, USA, 2022, pp. 4268–4277.
[47] O. Ronneberger, P. Fischer and T. Brox, "U-Net: Convolutional networks for biomedical image segmentation," in *proc. Int. Conf. Medical Image Computing and Computer-Assisted Intervention*, Munich, Germany, 2015, pp. 234–241.
[48] X. Xi, X. Meng, L. Yang, X. Nie, G. Yang, H.Chen, X. Fan, Y. Yin and X. Chen, "Automated segmentation of choroidal neovascularization in optical coherence tomography images using multi-scale convolutional neural networks with structure prior," *Multimedia Syst.*, vol. 25, pp. 95–102, 2019.
[49] L. Yang, L. Qi, L. Feng and W. Zhang, "Revisiting weak-to-strong consistency in semi-supervised semantic segmentation," in *Proc. IEEE/CVF Conf. Comput. Vision Pattern Recognit.*, Vancouver, Canada, 2023, pp. 7236–7246.
[50] L. Yu, S. Wang, X. Li, C.W. Fu and P.A. Heng, "Uncertainty-aware self-ensembling model for semi-supervised 3d left atrium segmentation," in *Proc. Medical Image Computing and Computer Assisted Intervention*, Shenzhen, China, 2019, pp. 605–613.
[51] A. Kendall, Y. Gal, "What uncertainties do we need in bayesian deep learning for computer vision?," in *Proc. Conference on Neural Information Processing Systems*, California, USA, 2017, pp. 5580–5590.
[52] Z. Shen, P. Cao, H. Yang, X. Liu, J. Yang, and O.R. Zaiane, "Co-training with high-confidence pseudo labels for semi-supervised medical image segmentation," *arxiv:2301.04465*,2023.
[53] Q. Ma, J. Zhang, L. Qi, Q. Yu, Y. Shi and Y. Gao, "Constructing and Exploring Intermediate Domains in Mixed Domain Semi-supervised Medical Image Segmentation," in *Proc. IEEE/CVF Conf. Comput. Vision Pattern Recognit.*, Seattle, USA, 2024, pp. 11642–11651.
[54] P. Zhu, M. Qi, X. Li, W. Li and H. Ma, "Unsupervised Self-Driving Attention Prediction via Uncertainty Mining and Knowledge Embedding", in *Proc. IEEE/CVF Int. Conf. Comput. Vision*, Paris, France, 2023, pp. 8558-8568.
[55] M. Qi, W. Li, Z. Yang, Y. Wang and J. Luo, "Attentive Relational Networks for Mapping Images to Scene Graphs", in *Proc. IEEE/CVF Conf. Comput. Vision Pattern Recognit.*, Long Beach, CA, USA, 2019, pp. 3957–3966
[56] M. Qi, J. Qin, Y. Yang, Y. Wang and J. Luo, "Semantics-Aware Spatial-Temporal Binaries for Cross-Modal Video Retrieval", *IEEE T. Image Process.*, vol. 30, pp. 2989–3004, 2021.
[57] C. Lv, S. Zhang, Y. Tian, M. Qi and H. Ma, "Disentangled Counterfactual Learning for Physical Audiovisual Commonsense Reasoning", in *Proc. Adv. Neural Inform. Process. Syst.*, Vancouver, British Columbia, Canada, 2024.
[58] M. Qi, Y. Wang, A. Li and J. Luo, "STC-GAN: Spatio-Temporally Coupled Generative Adversarial Networks for Predictive Scene Parsing", *IEEE T. Image Process.*, vol. 29, pp. 5420–5430, 2020.
[59] M. Qi, Y. Wang, J. Qin, A. Li, J. Luo and L. Van Gool, "stagNet: An Attentive Semantic RNN for Group Activity and Individual Action Recognition", *IEEE T. Circ. Syst. Vid.*, vol. 30, no. 2, pp. 549–565, 2019.
[60] J. Moon, J. Kim, Y. Shin, S. Hwang, "Confidence-aware learning for deep neural networks", in *Proc. int.conf. machine learning*, Online, pp.1-20,2020.

Jie Guo received the Ph.D. degree at Shandong University. She is currently a lecturer with School of Computer Science and Technology, Shandong Jianzhu University, Jinan, China. Her research interests include multimodal machine learning, medical image processing and multimedia analysis.

Liangyun Sun received his Master's degree from the School of Computer Science and Technology of Shandong Jianzhu University. He is currently pursuing a doctorate degree at Nanjing University of Science and Technology. His research interests include computer vision, medical image processing and multimodal large language model.

Lishan Qiao is a professor with School of Computer Science and Technology, Shandong Jianzhu University, doctoral advisor at the University of Camerino, Italy, and the principal investigator of Shandong Province's Qingchuang Team Introduction and Cultivation Plan. He has a Ph.D. in Computer Application Technology from Nanjing University of Aeronautics and Astronautics and has conducted postdoctoral research and academic visits at institutions such as Nanjing University and the University of North Carolina at Chapel Hill. His research interests span artificial intelligence, machine learning, pattern recognition, and brain function data analysis.

Xiushan Nie received the PhD degree from Shandong University, China, in 2011. He is currently a full professor with the School of Computer Science and Technology, Shandong Jianzhu University, China. He was a research fellow under the supervision of Prof. Wenjun (Kevin) Zeng with the University of Missouri-Columbia, Columbia, Missouri, (2013- 2014). His research interests include multimedia retrieval and indexing, multimedia security, and computer vision.

Jixin Yang is mechanical and electrical engineer of Shandong Kailin Environmental Protection Equipment Co., Ltd. His research interests include image segmentation, and scene understanding.

Weicui Li received the master's degree from Shandong University of Finance and Economics in 2011. She is currently an associate researcher with Shandong Institute of Science and Technology Information, Jinan, China. Her research interests include computer vision and medical image processing.

Ying Guo received a Master degree in Dongbei University of Finance and Economics in 2004. She is currently a research fellow in the School of Computer Science and Technology, Qilu University of Technology (Shandong Academy of Science) in Jinan, China. Her research interests include cloud computing, big data and software engineering. She has published about 20 academic papers. Meanwhile, she has many years of experience in software development and project management. Her industrial applications in the fields of cloud computing and big data have been widely used and highly evaluated by users.

Xiaoming Xi is professor and master's supervisor at Shandong Jianzhu University, leads the "Intelligent Visual Computing" team, a young and innovative scientific group in Shandong Province's higher education institutions. His research is centered on machine learning, pattern recognition, computer vision, and medical image processing.

Xinjian Chen received the Ph.D. degree from the Chinese Academy of Sciences, in 2006. He is currently a Distinguished Professor with Soochow University, Suzhou, China. He conducted postdoctoral research at the University of Pennsylvania, the National Institute of Health, and the University of Iowa, USA, from 2008 to 2012. He has published over 150 international journals and conference papers. He holds 14 patents.

Yilong Yin is the leader of the Artificial Intelligence discipline, a distinguished professor, and doctoral supervisor at Shandong University. He is an IET Fellow, an outstanding individual in Shandong Province, and an outstanding member of the CCF. He has published over 80 papers in international journals such as TPAMI, TIP, TKDE, TIFS, TMM, and at conferences including ICML, AAAI, IJCAI, CVPR, and MM. He leads key projects funded by the National Natural Science Foundation of China, the National Key R&D Program, and the major basic research plan of Shandong Province in the field of artificial intelligence theory and methods.