Received: 14 September 2023 | Revised: 23 January 2024 | Accepted: 20 February 2024

DOI: 10.1002/mp.17012

RESEARCH ARTICLE

MEDICAL PHYSICS

Two-stage adversarial learning based unsupervised domain adaptation for retinal OCT segmentation

Shengyong Diao[1] | Ziting Yin[1] | Xinjian Chen[1,2] | Menghan Li[3] | Weifang Zhu[1] | Muhammad Mateen[1] | Xun Xu[3] | Fei Shi[1] | Ying Fan[3]

[1] MIPAV Lab, the School of Electronics and Information Engineering, Soochow University, Suzhou, China

[2] The State Key Laboratory of Radiation Medicine and Protection, Soochow University, Suzhou, China

[3] Shanghai General Hospital, Shanghai Jiao Tong University School of Medicine, Shanghai, China

Correspondence
Fei Shi, MIPAV Lab, the School of Electronics and Information Engineering, Soochow University, Suzhou 215006, China.
Email: shifei@suda.edu.cn

Ying Fan, Shanghai General Hospital, Shanghai Jiao Tong University School of Medicine, Shanghai 200080, China.
Email: mdfanying@sjtu.edu.cn

Funding information
National Natural Science Foundation of China, Grant/Award Numbers: 62271337, 62371326; National Key Research and Development Program of China, Grant/Award Numbers: 2018YFA0701700, 2019FYC1710204; Natural Science Foundation of Jiangsu Province, Grant/Award Number: BK20231310

Abstract

Background: Deep learning based optical coherence tomography (OCT) segmentation methods have achieved excellent results, allowing quantitative analysis of large-scale data. However, OCT images are often acquired by different devices or under different imaging protocols, which leads to serious domain shift problem. This in turn results in performance degradation of segmentation models.

Purpose: Aiming at the domain shift problem, we propose a two-stage adversarial learning based network (TSANet) that accomplishes unsupervised cross-domain OCT segmentation.

Methods: In the first stage, a Fourier transform based approach is adopted to reduce image style differences from the image level. Then, adversarial learning networks, including a segmenter and a discriminator, are designed to achieve inter-domain consistency in the segmentation output. In the second stage, pseudo labels of selected unlabeled target domain training data are used to fine-tune the segmenter, which further improves its generalization capability. The proposed method was tested on cross-domain datasets for choroid or retinoschisis segmentation tasks. For choroid segmentation, the model was trained on 400 images and validated on 100 images from the source domain, and then trained on 1320 unlabeled images and tested on 330 images from target domain I, and also trained on 400 unlabeled images and tested on 200 images from target domain II. For retinoschisis segmentation, the model was trained on 1284 images and validated on 312 images from the source domain, and then trained on 1024 unlabeled images and tested on 200 images from the target domain.

Results: The proposed method achieved significantly improved results over that without domain adaptation, with improvement of 8.34%, 55.82% and 3.53% in intersection over union (IoU) respectively for the three test sets. The performance is better than some state-of-the-art domain adaptation methods.

Conclusions: The proposed TSANet, with image level adaptation, feature level adaptation and pseudo-label based fine-tuning, achieved excellent cross-domain generalization. This alleviates the burden of obtaining additional manual labels when adapting the deep learning model to new OCT data.

KEYWORDS
adversarial learning, deep learning, retinal OCT image segmentation, unsupervised domain adaptation

1 INTRODUCTION

Optical coherence tomography (OCT), a high resolution, non-invasive, and highspeed imaging technique, capable of showing cross sections of the retina, is widely used in the clinical practice for diagnosis and management of retinal diseases. Accurate quantization of retinal tissues and lesions is important for clinical decisions and thus comes the need for automatic segmentation of OCT images. With the development of deep convolutional neural networks (CNN),[1–3] they have become powerful tools for automatic segmentation of retinal OCT images. This has largely alleviated the problem of difficult lesion identification, greatly reduced the workload, and improved the diagnostic efficiency of ophthalmologists, and is of great importance for treatment guidance of diseases.

The ideal application scenario for deep learning is to have the test data with the same distribution as the training data. However, this assumption often does not hold in real-world applications. It is well known that medical images obtained from devices of different manufacturers, or under different imaging protocols, vary greatly in image quality. For retinal OCT images, for example, images obtained by spectral domain OCT (SD-OCT) and swept-source OCT (SS-OCT) devices have different contrast between retinal layers, due to the usage of light sources of different wavelengths.[4] Different post-processing procedures embedded in the scanners, such as different numbers of averaging, give images with different noise levels.[5] Moreover, in many real-world applications, collecting enough labeled training data is often time-consuming and expensive. Therefore, it is often the case that the CNN model is trained on one dataset (source domain) where training labels are available and is expected to work on other datasets (target domain) where no training labels are provided. The domain shift between data often limits the generalization and knowledge reuse ability of traditional deep learning models.[6]

Unsupervised domain adaptation (UDA) aims to address the problem of model performance degradation caused by domain shift when only source domain annotations are available. Currently, unsupervised domain adaptation is widely used in CNN-based image classification tasks.[7–12] As for semantic segmentation, many works on unsupervised domain adaptation have been carried out based on adversarial learning. Many of them adopted the CycleGAN[13] scheme, some with additional constraints, to reduce the domain shift in pixel level. Both Zhang et al.[14] and Zhang et al.[15] proposed CycleGAN structures with add-on segmentation supervision for medical image segmentation. The methods of cycle-consistent adversarial domain adaptation (CyCADA)[16] and domain adaptation via deeply synergistic image and feature alignment (SIFA)[17] were proposed which achieved both pixel-level and feature alignment based on CycleGAN.

However, the double generators and discriminators of the CycleGAN can make the network quite complex and difficult to train. Instead, a variety of works applied adversarial learning in the feature space or output space to tackle domain shift efficiently. Wang et al.[18] proposed to use a patch-based discriminator to enforce similarity in the output. Dou et al.[19] proposed a plug-and-play adversarial domain adaptation network, consisting of a source segmentation network, a domain adaptation module, a feature discriminator, and an output mask discriminator. Tsai et al. proposed AdaptSeg[20] which adopted two adaptation modules for output space adversarial learning at different levels. Yan et al.[21] also used two-level output adaption modules, where the Canny edge detector was introduced to enhance attention to edges during adversarial learning. Tsai et al. further proposed AdaptPatch[22] which aligned output patch distributions. Saito et al. proposed maximum classifier discrepancy for domain adaptation (MCDDA)[23] which aligned distributions of source and target by utilizing the task-specific decision boundaries. Vu et al. proposed the adversarial entropy minimization (ADVENT) [24] method which minimized the prediction entropy of the target sample both directly and through adversarial learning. Chen et al. further proposed the Maxsquare[25] method using the maximum squares loss for the entropy minimization setting. Many of these previous works have shown that alignment at multiple levels is beneficial for domain adaptation performance.

Not many works have focused on domain adaptation of OCT image segmentation. Bian et al.[26] proposed uncertainty-aware domain alignment in the feature level for retina and choroid segmentation. Chai et al.[27] proposed a network for choroid segmentation with perceptual loss and adversarial loss in the output space. Chen et al.[28] proposed a CycleGAN-based domain adaptation method for intraretinal layer segmentation. He et al.[29] proposed a cross-domain fluid segmentation network using layer structures guidance, which also performed domain alignment in the output space. These works only applied single-level domain adaptation, and their models were tested on one specific OCT segmentation task respectively. Some were only tested on a specific pair of source and target domains.

In this paper, we propose a two-stage method for OCT image segmentation based on unsupervised domain adaptation. The domain shift problem is approached in multiple levels and using multiple strategies. The first stage takes the adversarial learning approach. At the image level, we avoid the complex CycleGAN method and adopted Fourier coefficient replacement[30] to reduce image style differences. Then at the feature level, adversarial learning is applied in the output space to narrow the gap of inter-domain feature distribution. The second stage adopts the pseudo-labeling technique. Samples in the target domain with low-entropy prediction are collected, and the model is fine-tuned with pseudo-labels to further improve the generalization performance. The proposed framework was tested on two tasks of cross-domain choroid segmentation and retinoschisis segmentation, respectively, with images from various scanners and protocols. The results demonstrate its superiority over existing UDA methods and the generalizability for different source and target domains. Preliminary version of this work was presented in Ref. [31].

2 | MATERIALS AND METHODS

2.1 | Datasets

In this paper, we apply the proposed model for two segmentation tasks respectively, which are choroid and retinoschisis segmentation. The choroid thickness and volume are important indices of the progress of various eye diseases,[32] while retinoschisis is one of the most common lesions in high myopia retina and its location and size are important for diagnosis and treatment of pathological myopia.[33] The datasets were made up of clinical data acquired at Shanghai General Hospital. The collection and analysis of image data were approved by the Institutional Review Board of Shanghai General Hospital, and adhered to the tenets of the Declaration of Helsinki.

For the choroid segmentation task, OCT B-scans came from both normal and high myopia subjects. The choroid thickness varies greatly among subjects. The lower boundary of the choroid is often not well defined, especially when the noise level is high. The source domain data were obtained by a Topcon DRI-1 SS-OCT scanner (Topcon Corp., Japan) in 12-line radial scan mode. Two datasets were collected as target domain data respectively. Target domain I data were obtained by the same Topcon DRI-1 scanner in 256-line volumetric scan mode, and target domain II data were obtained by a Zeiss Cirrus HD-OCT 4000 (Carl Zeiss Meditec. Inc, USA) SD-OCT scanner in 128-line volumetric scan mode.

For the retinoschisis segmentation task, OCT B-scans came from high myopia subjects. The lesion can appear in various retinal layers, and the size also varies greatly. The source domain data were generated by the Topcon DRI-1 SS-OCT scanner in 12-line radial scan mode, and the target domain data were obtained by the SVision SS-OCT scanner (SVision Imaging, Ltd., China) in 96-line volumetric scan mode.

All B-scans are centered at the macula but may cover different physical lengths. The image quality was considered acceptable for clinical diagnosis by inspection of ophthalmologists, and no other quality control was performed. The source domain data were divided into training and validation sets, while the target domain data were divided into training and testing sets. All the divisions are on patient-level. Manual delineation of the boundaries was performed by an ophthalmologist using the biomedical image visualization and analysis software ITK-SNAP.[34] Smooth curves were drawn where the boundaries were undefined or broken. The boundaries were then converted to binary regional masks as ground truth. Manual annotation was done on the source domain data for training and validation, and the target domain test data for performance evaluation only. The details of each dataset are shown in Table 1.

Figures 1 and 2 show some OCT B-scans in our datasets, acquired by different imaging protocols and/or different devices. It can be observed that these images differ in noise level, contrast, resolution, etc. The 12-line radial scans are the results of averaging dozens of repeated B-scans by the built-in algorithms, and they have high image quality. With low-level noise and high contrast, manual annotation is easier and more reliable, and thus they are used as the labeled source domain data. However, these B-scans can only cover 12 radial lines centered at the macula, and cannot give a complete view of the whole retina. In contrast, B-scans obtained by the volumetric scanning protocol are raw data or averaging results of only several B-scans, and the noise level is usually high. Manual labeling of these images is more difficult and time-consuming, and automatic segmentation becomes more challenging as well. However, segmentation in these B-scans is of more interest, as 3D reconstruction of the layers or the lesions can be further obtained to allow more complete analysis. Therefore, in our study, they are treated as the unlabeled target domain.

2.2 | Overview of the two-stage method

For UDA, denote a pair of labeled source domain data as (x^s, y^s) where y^s is the ground truth corresponding to the image x^s, and an unlabeled target image as x^t. The framework of the proposed method is shown in Figure 3. The whole framework consists of two stages: the first-stage domain adaptation, with a segmenter S and a discriminator D, and the second-stage fine-tuning of the segmenter. Among them, the first-stage domain adaptation is further divided into two levels: image-level style transfer and feature-level adversarial domain adaptation.

2.3 | Image-level style transfer

Generative adversarial networks (GAN) have achieved excellent results in image style transfer,[13–17] but the model structure is complex and the network training is difficult and time-consuming. Therefore, we adopt

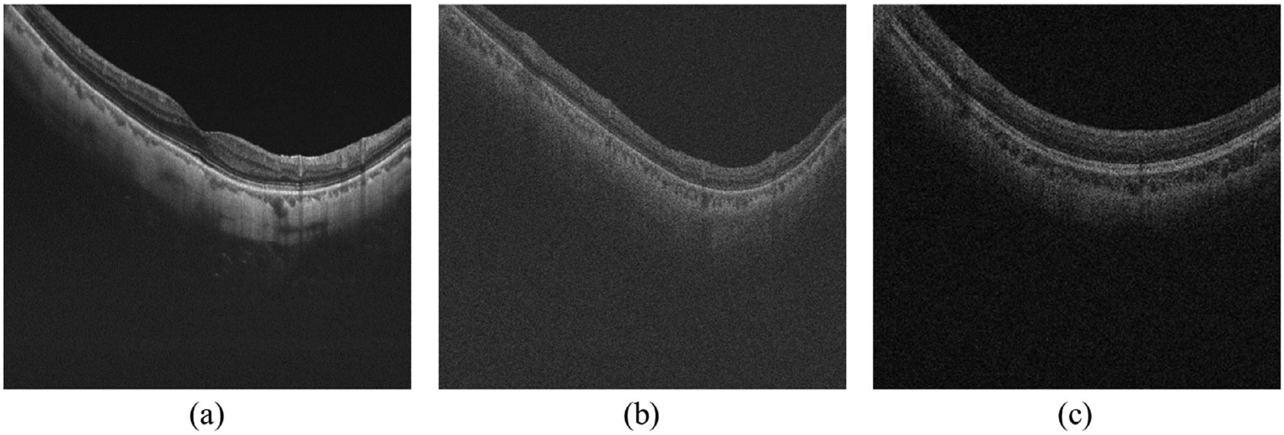

FIGURE 1 Choroid segmentation dataset with domain shifts. (a) Acquired by Topcon DRI-1 SS-OCT scanner using 12-line radial scan mode. (b) Acquired by Topcon DRI-1 scanner using 256-line volumetric scan mode. (c) Acquired by Zeiss 4000 SD-OCT scanner using 128-line volumetric scan mode.

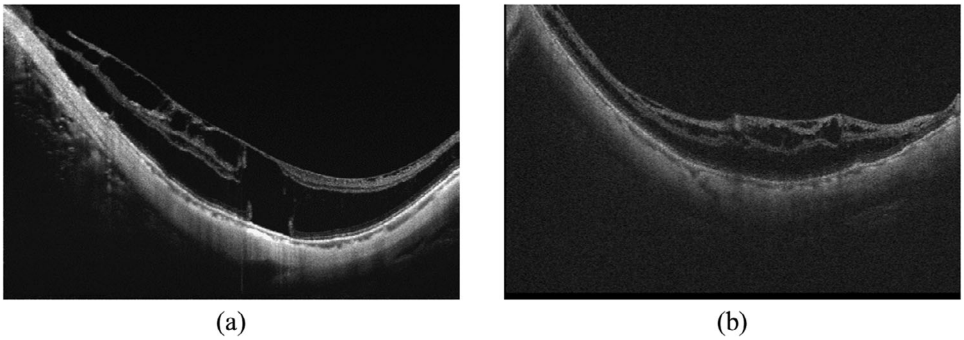

FIGURE 2 Retinoschisis segmentation dataset with domain shifts. (a) Acquired by Topcon DRI-1 SS-OCT scanner using 12-line radial scan mode (b) Acquired by SVision SS-OCT scanner using 96-line volumetric scan mode.

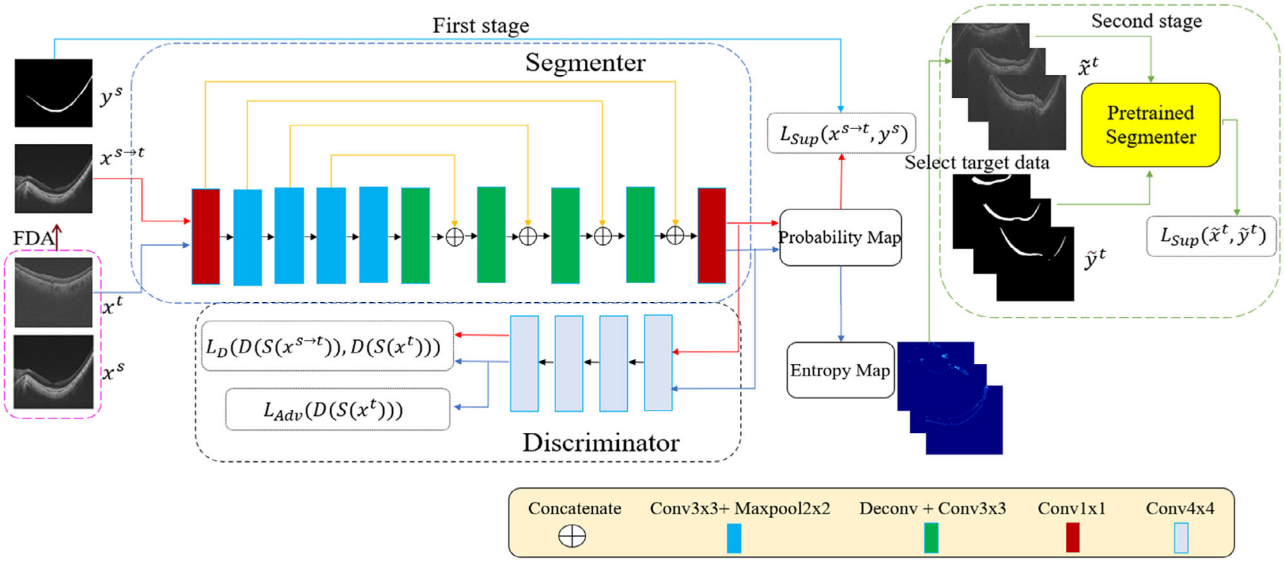

FIGURE 3 An overview of the proposed method.

TABLE 1 Optical coherence tomography (OCT) datasets for unsupervised domain adaptation (UDA).

Task	Domain	Training	Validation	Testing	Imaging protocol	Scanner	Imaging wavelength (nm)	Image size	B-scan length (mm)
Choroid	Source	400	100	–	12-line radial scan	Topcon DRI-1	1310	1024 × 992	9
	Target I	1320	–	330	256-line volumetric scan	Topcon DRI-1	1310	512 × 992	9
	Target II	400	–	200	128-line volumetric scan	Zeiss 4000	840	512 × 1024	6
Retino- schisis	Source	1284	312	–	12-line radial scan	Topcon DRI-1	1310	1024 × 992	9
	Target	1024	–	200	96-line volumetric scan	SVision	1310	2048 × 1382	12

the Fourier transform-based domain adaptation (FDA)[30] where no training is required. In FDA, the low-frequency part of the source domain image is replaced with that of a target domain image, to achieve a simple style transfer between images and reduce the problem of domain shift from the image level.

For a source domain image x^s, calculate its 2-D Fourier transform, and define F^A as the magnitude and F^P as the phase, for which the low-frequency part is in the center. In addition, use M_β to denote a mask whose value is 1 for the central region and 0 otherwise, expressed as:

$$M_\beta(h, w) = 1_{(h,w)\in[-\beta H:\beta H, -\beta W:\beta W]} \tag{1}$$

where H and W represent the height and width of the image respectively, and $\beta \in [0, 1]$ controls the size of nonzero regions. Then the transferred image after replacement can be represented as:

$$x^{s\to t} = F^{-1}\left(\left[M_\beta \cdot F^A\left(x^t\right)\right] + \left(1 - M_\beta\right) \cdot F^A\left(x^s\right), F^P\left(x^s\right)\right) \tag{2}$$

where F^{-1} represents the inverse Fourier transform.

Figure 4 shows the transferred source domain images obtained by different values of β on different datasets. By replacing the low-frequency part, the reconstructed source domain image appears closer to the target image in intensity. Larger β brings higher similarity. However, too large β values introduce ringing and blurring artifacts. By visual inspection and ablation tests, the value of β is set as 0.01 in all experiments in this paper.

2.4 | Feature-level domain adaptation

Image style transfer is task independent and thus has insufficient discriminative power. Adversarial learning enables models to learn how to extract domain-invariant features by using adversarial loss to impose strong constraints. In the pixel-level classification task of medical image segmentation, the output space contains rich information, and this information should be shared in different domains. Alignment in the output space is more task-oriented and effective. Therefore, in this paper, we adopt the strategy of adversarial learning in the output space to achieve the purpose of feature-level domain adaptation.

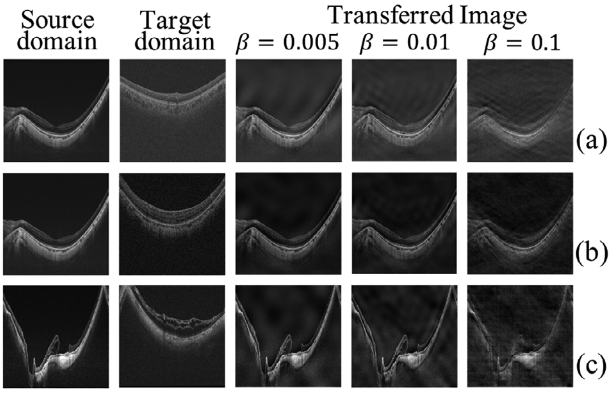

FIGURE 4 Fourier transform based domain adaptation (FDA) image style transfer. (a) Target domain I for choroid segmentation (b) Target domain II for choroid segmentation (c) Target domain for retinoschisis segmentation.

As shown in Figure 3, both the transferred source domain images $x^{s\to t}$ and target domain images x^t pass through the segmenter S, and their predicted probability maps are input into the discriminator D. The goal of S is to segment the source domain images correctly while making the predictions of two domains undiscernible by D, and therefore it is optimized by minimizing the supervised loss L_{seg} using ground truth for source domain images, and the unsupervised adversarial loss L_{adv}, which is the binary cross entropy (BCE) loss that characterizes the difference between the discriminator output for target domain image and the source domain labels. The goal of D is to differentiate the predictions from the two domains to its best, and it is optimized by minimizing the discriminator loss L_D, which contains the BCE losses that characterize the difference between the discriminator output for images from both domain and their true domain labels, respectively. Joint minimization of these loss functions enables the networks to obtain satisfactory segmentation results while extracting domain-invariant features.

2.5 | Second-stage fine-tuning

After image and feature-level domain adaptation, an initial segmenter is trained. To further fit the segmenter to target domain data, the pseudo-labeling technique, which is originally used in semi-supervised learning,[35] is adopted to fine-tune the segmenter. Specifically, the initial segmenter is applied on the unlabeled target domain training images, to get their pseudo-labels. Then the segmenter is fine-tuned using paired target domain images and pseudo-labels.

The pseudo-labeling technique solves the problem of model dependence on labeled data to a certain extent. However, the pseudo-labels are often noisy. Directly retraining the model with these noisy outputs as pseudo-labels may degrade the segmentation performance of the original model.[35] To solve this problem, pseudo-labels with high confidence should be selected. In this paper, we use information entropy, a measure of uncertainty of random variables, as the confidence measure of the prediction result. Specifically, the segmenter trained in the first stage is used to get the output for the target domain training images. The entropy for each pixel in the prediction map is calculated and summed up to get the confidence score of the prediction result of this image, as follows:

$$E_{x^t} = -\sum_{H,W}\left[P_{x^t}^{(h,w)}\log\left(P_{x^t}^{(h,w)}\right) + \left(1-P_{x^t}^{(h,w)}\right)\log\left(1-P_{x^t}^{(h,w)}\right)\right] \quad (3)$$

where $P_{x^t}^{(h,w)}$ is the output probability at pixel location (h, w).

Smaller entropy values mean lower uncertainty and more reliable predictions. Figure 5 visualizes the entropy maps for some target domain images for choroid segmentation, where (a) and (b) represent low entropy predictions, and (c) and (d) represent high entropy predictions. It can be seen that the highly uncertain regions are often consistent with segmentation errors. For images with lower total entropy values, the segmentation is more accurate.

Then, k% of target domain training data with lowest entropy values, with the first-stage prediction as their pseudo-labels, denoted as $(\tilde{x}^t, \tilde{y}^t)$, are selected to fine-tune the segmenter.

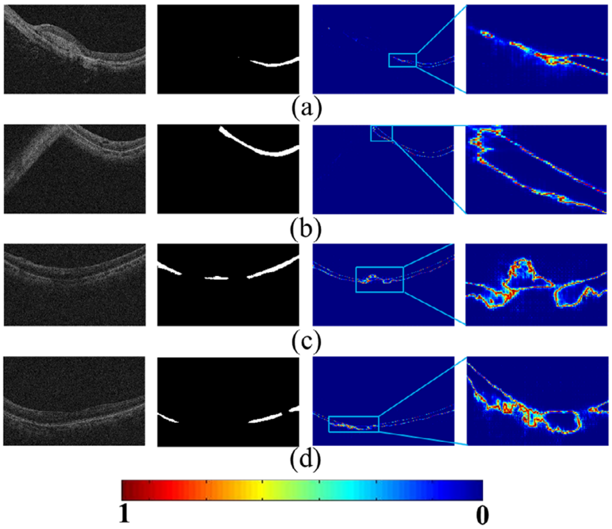

FIGURE 5 Visualization of entropy maps of unlabeled target domain data. (a) and (b) for low entropy predictions, and (c) and (d) for high entropy predictions. First column: the original image, second column: coarse segmentation result, third column: entropy maps, fourth column: entropy maps zoomed in.

2.6 | Network architecture

In this paper, we use the traditional U-shaped network structure[36] as the segmenter S. The feature encoder contains four layers, each consisting of a convolution kernel of size 3×3 with a step size of 1, batch normalization and ReLU activation function. The image is downsampled after each layer of the encoder using maximum pooling to reduce the resolution of the image while increasing the receptive field. Each decoder layer consists of deconvolution and 3×3 convolutions. Skip connections help avoid the loss of low-level semantic information while reconstruction of the high-resolution feature maps.

The discriminator D is patch-based, consists of four convolutional layers with a kernel size of 4×4, each of which is followed by a Leaky ReLU layer (Leak = 0.2), and its output is a 32×32 matrix. This can make the discriminator pay attention to each local patch, and therefore effectively align the inter-domain features.

2.7 | Loss functions

Let $S(x^t)$ and $D(S(x^t))$ denote the output of the segmenter S and the discriminator D for target domain data, and let $S(x^{s\to t})$ and $D(S(x^{s\to t}))$ denote the output of S and D for style transferred source domain data, respectively. In the first stage, for segmenter, the supervised loss is a combination of the dice loss and the pixel-wise BCE loss functions, and the adversarial loss function is a BCE loss that measure the discrepancy between $D(S(x^t))$ and the source domain label. Therefore, the total loss function L_S^1 for the segmenter is expressed as:

$$L_S^1 = L_{Dice} + L_{BCE} + \alpha L_{Adv} \tag{4}$$

Specifically, L_{Dice}, L_{BCE} and L_{Adv} are calculated as follows:

$$L_{Dice} = 1 - \frac{1}{N} \sum_{i=1}^{N} \frac{2S_i\left(x^{s\to t}\right) y_i^s}{\left(S_i\left(x^{s\to t}\right)\right)^2 + \left(y_i^s\right)^2} \tag{5}$$

$$L_{BCE} = -\frac{1}{N} \sum_{i=1}^{N} \times \left[y_i^s \log S_i\left(x^{s\to t}\right) + \left(1 - y_i^s\right) \log\left(1 - S_i\left(x^{s\to t}\right)\right)\right] \tag{6}$$

$$\begin{aligned} L_{Adv} &= \sum_{x^t \in X^t} L_{BCE}\left(D\left(S\left(x^t\right)\right), 0\right) \\ &= -\frac{1}{M} \sum_{x^t \in X^t} \sum_{m=1}^{M} \log\left(1 - D_m\left(S\left(x^t\right)\right)\right) \end{aligned} \tag{7}$$

where $S_i(x^{s\to t})$ represents the value of the ith pixel in the predicted segmentation result of source domain image, y_i^s represents the ground truth label of the ith pixel, N represents the total number of pixels in the image, $D_m(S(x^t))$ represents the mth element value in the output matrix of the discriminator, M represents the number of elements in the matrix, 0 represents the source domain label, α represents the weight of the adversarial loss. In this paper, α is set to 0.001 empirically.

For the discriminator, the following BCE loss function is chosen to optimize it.

$$\begin{aligned} L_D &= L_{BCE}\left(D(S(x^{s\to t})), 0\right) + L_{BCE}(D(S(x^t)), 1) \\ &= -\frac{1}{M} \sum_{m=1}^{M} \log(1 - D_m(S(x^{s\to t}))) \\ &\quad - \frac{1}{M} \sum_{m=1}^{M} \log(D_m(S(x^t))) \end{aligned} \tag{8}$$

where 0 and 1 represent the source and target domain labels, respectively.

For the second stage, loss function L_S^2 for the segmenter only contains the supervised loss functions:

$$L_S^2 = L_{Dice} + L_{BCE} \tag{9}$$

3 | EXPERIMENT SETTINGS

3.1 | Evaluation metrics

In this paper, we use four evaluation metrics commonly used in medical image segmentation tasks, namely dice similarity coefficient (DSC), intersection over union (IoU), sensitivity (Sen), and specificity (Spe).

3.2 | Implementation details

The proposed method was implemented in Python using the Pytorch framework on a GeForce GTX 2080Ti graphics card with 11 GB GPU memory. The parameters of the segmenter were updated by the stochastic gradient descent (SGD) algorithm (momentum = 0.9, weight decay = 0.0001) and its initial learning rate is 0.01. Differently, the discriminator adopts Adam optimizer with an initial learning rate of 0.001. The segmenter and discriminator were optimized alternatively. For both stages, the networks were trained for 60 epochs. The batch size was set as 2.

All images were resized to 512 × 512 and the intensities were normalized to [0, 255] before input to the networks. Data augmentation including random horizontal flipping and Gaussian noise addition were applied for training in both stages. The source domain validation set was used to choose the best segmenter in the first stage. For the second stage, in the selected target domain training data with pseudo-labels, 80% were used for training and 20% for validation, to give the best model for testing. For both pseudo-label generation and segmentation output, a threshold of 0.5 was applied to the predicted probability maps to get the binary mask.

4 | EXPERIMENTAL RESULTS

4.1 | Ablation tests

In the second stage, the target domain training data with high-confidence pseudo-labels are selected to fine-tune the segmenter. Table 2 shows the test results when different percentage value k is set in this selection. It can be seen that $k = 70$ gives the best overall performance, because it can reach a balance between retaining enough training data and removing enough data with noisy labels. Therefore, 70% of target domain training data is used for all our experiments.

Table 3 shows results of the ablation tests for components of the proposed method. Here the baseline means directly applying the segmenter trained on source domain to target domain data. In the choroid segmentation task, target domain I was obtained by the same scanner but different protocol with source domain. Compared with baseline, FDA based image-level style transfer improves the IoU for 2.24%, adversarial learning (ADL) based feature-level alignment improves the IoU for 5.81%, and combining them, the IoU is improved by 6.93%. The proposed TSANet, with the second stage added, has a total improvement of 8.34% for IoU. Target domain II was obtained by the SD-OCT scanner instead of the SS-OCT scanner for source domain, resulting in bigger difference in image quality, and the

TABLE 2 Ablation tests on percentage of data for pseudo-label training mean (std).

	Choroid segmentation											
	Target domain I				Target domain II				Retinoschisis segmentation			
k(%)	IoU(%)	DSC(%)	Sen(%)	Spe(%)	IoU(%)	DSC(%)	Sen(%)	Spe(%)	IoU(%)	DSC(%)	Sen(%)	Spe(%)
50	**80.72** (1.61)	88.34 (1.33)	89.83 (1.63)	**89.08 (1.52)**	70.05 (2.63)	80.47 (2.30)	90.55 (0.98)	75.52 (2.89)	85.64 (2.60)	91.19 (2.37)	96.23 (5.39)	88.57 (5.01)
60	80.37 (1.53)	88.19 (1.31)	91.47 (1.46)	87.64 (1.49)	69.82 (2.39)	80.65 (2.08)	91.47 (0.89)	74.55 (2.61)	86.02 (2.11)	91.50 (1.70)	96.70 (5.20)	88.19 (4.98)
70	80.51 (1.28)	**88.63** (0.95)	**92.56 (1.17)**	86.15 (1.72)	**73.40** (2.67)	**82.47** (2.34)	92.66 (0.72)	**78.21** (2.98)	**86.84** (2.06)	**91.92** (1.74)	95.89 (5.25)	**90.78** (5.06)
80	80.34 (1.56)	88.11 (1.59)	90.9 (1.59)	87.8 (1.48)	71.51 (2.35)	81.93 (1.97)	91.19 (0.85)	76.78 (2.58)	84.79 (2.34)	90.52 (2.07)	97.24 (5.27)	87.12 (4.96)
90	79.79 (1.56)	87.83 (1.27)	90.71 (1.51)	87.23 (1.55)	72.03 (2.72)	81.67 (2.43)	92.65 (0.60)	77.08 (3.02)	84.41 (2.40)	90.18 (2.12)	**97.32** (5.28)	87.22 (4.95)
100	79.38 (1.57)	87.55 (1.30)	91.43 (1.52)	86.14 (1.58)	71.36 (2.66)	81.34 (2.33)	**94.02** (0.49)	75.31 (2.92)	85.71 (2.15)	91.23 (1.82)	96.93 (5.26)	88.39 (4.98)

Abbreviations: DSC, dice similarity coefficient; IoU, intersection over union; Sen, sensitivity; Spe, specificity.

TABLE 3 Ablation test of components of the proposed TSANet mean(std).

	Choroid segmentation											
	Target domain I				Target domain II				Retinoschisis segmentation			
Method	IoU (%)	DSC (%)	Sen (%)	Spe (%)	IoU (%)	DSC (%)	Sen (%)	Spe (%)	IoU (%)	DSC (%)	Sen (%)	Spe (%)
Baseline	72.17 (1.48)	82.87 (1.33)	91.87 (1.34)	77.70 (1.63)	17.58 (2.64)	25.68 (3.49)	75.28 (5.05)	35.41 (2.90)	83.31 (2.73)	88.86 (2.54)	96.08 (5.26)	89.02 (5.07)
+FDA	74.41 (1.69)	84.17 (1.40)	84.78 (1.83)	86.09 (1.44)	28.93 (3.15)	39.89 (3.88)	73.83 (4.17)	40.87 (3.35)	86.05 (2.31)	91.09 (2.11)	**96.57** (5.28)	89.85 (5.11)
+ADL	77.98 (1.78)	86.81 (1.61)	91.96 (1.30)	84.38 (1.93)	67.41 (3.12)	77.50 (3.02)	88.95 (1.00)	74.49 (3.51)	84.96 (2.51)	90.28 (2.28)	95.98 (5.31)	89.50 (5.15)
+FDA + ADL	79.10 (1.44)	87.50 (1.22)	91.95 (1.19)	85.51 (1.80)	70.94 (2.89)	80.55 (2.69)	89.81 (1.08)	77.17 (3.23)	86.62 (2.11)	91.79 (1.81)	96.00 (5.25)	90.65 (5.08)
TSANet	**80.51** (1.28)	**88.63** (0.95)	**92.56** (1.17)	**86.15** (1.72)	**73.40** (2.67)	**82.47** (2.34)	**92.66** (0.72)	**78.21** (2.98)	**86.84** (2.06)	**91.92** (1.74)	95.89 (5.25)	**90.78** (5.06)

Abbreviations: ADL, adversarial learning; DSC, dice similarity coefficient; FDA, fourier transform based domain adaptation; IoU, intersection over union; Sen, sensitivity; Spe, specificity; TSANet, two-stage adversarial learning based network.

effect of domain adaptation is more profound. Without domain adaptation, the model fails in the target domain with IoU of only 17.58%. Applying FDA improves it to 28.93%, and applying ADL effectively improves the IoU to 67.41%. Combining the two improves the IoU to 70.94%, and adding the second stage further improves it to 73.40%. In the retinoschisis segmentation task, the source and target domain data were from two different SS-OCT scanners. FDA improves the IoU by 2.74%, ADL improves the IoU by 1.65%, combining them improves the IoU by 3.31%. The proposed TSANet has a total improvement of 3.53% for IoU. Similar improvements can be observed in most other performance indices. Although for retinoschisis segmentation, the sensitivity of TSANet is slightly lower than the other variations, its specificity is higher, and the two indices are more balanced, thus giving a better overall segmentation performance. These ablation tests show that all the proposed components contribute to the performance of the proposed UDA scheme.

4.2 | Comparison with other UDA methods

We compare the proposed method with state-of-the-art unsupervised domain adaptation semantic segmentation methods, including AdaptSeg,[20] AdaptPatch,[22] ADVENT,[24] CycleGAN,[13] CYCADA,[16] MCDDA,[23] Maxsquare,[25] SIFA[17], and VarDA.[37] Table 4 shows the comparison results.

Comparing the proposed TSANet with all other methods, it achieved the highest IoU of 80.51% and 73.40% in the two target domains for choroid segmentation, respectively, and achieved the highest IoU of 86.84% in the target domain for retinoschisis segmentation. Wilcoxon rank sum test was performed on IoU values between the results of TSANet and other methods. Except for CycleGAN on the Retinoschisis dataset, the improvement achieved by the proposed TSANet has statistical significance with $p < 0.05$. The dice is also the highest among all methods compared. The sensitivity and specificity are relatively balanced among all test sets. The inference time of the proposed method is 42 ms for each B-scan, which is decent and can meet the requirements of clinical applications.

In addition, visual comparisons of the results with existing methods are shown in Figures 6–8. It can be seen from the three figures that compared with the existing methods, the proposed method can segment the target area more correctly, and the cases of false positives and false negatives are reduced.

5 | DISCUSSION AND CONCLUSIONS

There is domain shift between OCT images obtained by different scanners or different imaging protocols. In

TABLE 4 Comparison with unsupervised domain adaptation methods mean(std).

	Choroid segmentation												
	Target domain I				Target domain II				Retinoschisis segmentation				Testing
Method	IoU (%)	DSC (%)	Sen (%)	Spe (%)	IoU (%)	DSC (%)	Sen (%)	Spe (%)	IoU (%)	DSC (%)	Sen (%)	Spe (%)	time (ms)
AdaptSeg[20]	71.89 (1.34)	82.91 (1.11)	93.33 (1.14)	76.27 (1.63)	56.75 (3.12)	68.94 (3.24)	86.86 (1.65)	61.70 (3.42)	83.26 (2.34)	89.47 (2.02)	93.28 (5.09)	90.17 (5.08)	33.4
AdaptPatch[22]	74.88 (1.59)	84.54 (1.43)	90.51 (1.38)	82.04 (1.58)	47.58 (2.89)	61.14 (3.17)	86.78 (1.92)	50.69 (3.09)	83.17 (2.62)	89.07 (2.37)	**96.55** (5.28)	86.48 (5.05)	40.5
ADVENT[24]	72.71 (1.63)	83.05 (1.45)	93.36 (1.38)	77.28 (1.78)	54.59 (3.27)	66.67 (3.47)	78.42 (2.68)	62.22 (3.74)	84.11 (2.53)	89.17 (2.23)	96.32 (5.16)	88.88 (4.79)	35.6
CycleGAN[13]	73.32 (1.30)	83.90 (1.13)	93.03 (1.30)	77.92 (1.43)	68.09 (2.66)	78.95 (2.43)	87.89 (1.32)	74.06 (2.89)	84.86 (2.37)	90.20 (2.16)	94.49 (5.12)	91.37 (5.00)	27.7
CYCADA[16]	76.39 (1.32)	85.93 (1.10)	91.99 (1.31)	82.21 (1.42)	68.72 (2.16)	80.18 (1.86)	86.69 (1.23)	76.36 (2.40)	85.19 (2.38)	90.54 (2.09)	94.22 (5.22)	91.17 (5.10)	36.9
MCDDA[23]	74.89 (1.85)	84.13 (1.69)	85.04 (1.71)	87.02 (1.77)	65.85 (3.17)	76.29 (3.04)	75.42 (2.88)	**80.28** (3.40)	84.31 (2.61)	89.73 (2.34)	95.69 (5.14)	88.60 (5.07)	36.3
Maxsquare[25]	73.88 (1.51)	84.06 (1.25)	88.86 (1.54)	81.88 (1.51)	55.15 (3.40)	66.87 (3.59)	87.12 (1.80)	60.12 (3.81)	85.75 (2.18)	91.08 (1.92)	94.88 (5.22)	**91.49** (5.06)	30.3
SIFA[17]	74.61 (1.58)	84.38 (1.41)	**93.41** (1.22)	79.50 (1.78)	57.37 (1.44)	72.27 (1.30)	78.68 (1.04)	68.25 (1.72)	83.52 (2.65)	88.98 (2.35)	96.10 (5.02)	89.22 (5.09)	53.6
VarDA[37]	79.10 (1.48)	87.46 (1.25)	89.33 (1.33)	**87.91** (1.65)	67.51 (2.90)	78.10 (2.69)	88.44 (1.13)	74.03 (3.28)	85.02 (2.73)	90.32 (2.09)	96.05 (5.67)	89.15 (4.96)	50.3
TSANet	**80.51** (1.28)	**88.63** (0.95)	92.56 (1.17)	86.15 (1.72)	**73.40** (2.67)	**82.47** (2.34)	**92.66** (0.72)	78.21 (2.98)	**86.84** (2.06)	**91.92** (1.74)	95.89 (5.25)	90.78 (5.06)	42.0

Abbreviations: ADVENT, adversarial entropy minimization; DSC, dice similarity coefficient; IoU, intersection over union; MCDDA, maximum classifier discrepancy for domain adaptation; Sen, sensitivity; SIFA, synergistic image and feature alignment; Spe, specificity; TSANet, two-stage adversarial learning based network.

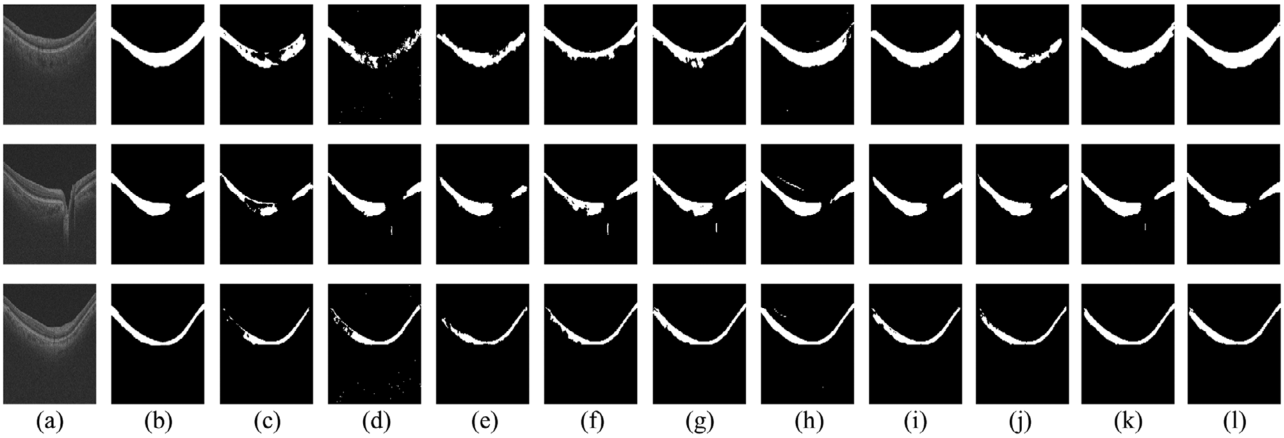

FIGURE 6 Choroid segmentation results of target I dataset using different algorithms (a) original image, (b) ground truth, (c) AdaptSeg, (d) AdaptPatch, (e) Advent, (f) CycleGAN, (g) CYCADA, (h) MCDDA, (i) Maxsquare, (j) SIFA, (k) VarDA, and (l) TSANet.

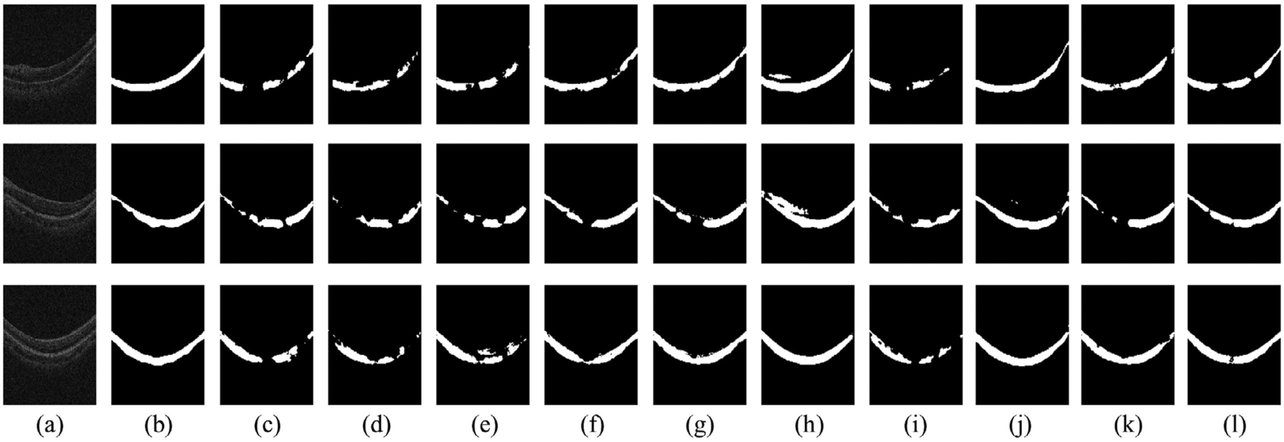

FIGURE 7 Choroid segmentation results of target II dataset using different algorithms (a) original image, (b) ground truth, (c) AdaptSeg, (d) AdaptPatch, (e) Advent, (f) CycleGAN, (g) CYCADA, (h) MCDDA, (i) Maxsquare, (j) SIFA, (k) VarDA, and (l) TSANet.

this paper, we aim at the task of unsupervised domain adaptation for OCT segmentation and propose a two-stage adversarial learning-based network. We address the problem of domain shifts between images from three perspectives: image-level adaptation through a simple style transfer technique, feature-level adaptation through discrimination in the output space, and model fine-tuning based on highly confident pseudo-labels from target domain. We conduct extensive experiments to validate the performance of the proposed method. The experiments involve two segmentation tasks, corresponding to anatomical structure and lesion area, respectively, and three different scanners with various imaging protocols. Ablation tests show that all steps in the proposed method contributed to the final performance. Compared with single image-level adaptation using CycleGAN,[12] single output space adaptation methods such as AdaptSeg[27] and AdaptPatch,[28] and multi-level adaptation method CYCADA,[11] and other state-of-the-art UDA methods,[17,29–31,37] the proposed method achieved superior performance. The experiments demonstrate the applicability and generalizability of the proposed model for OCT segmentation.

There are some limitations of the proposed model. First, there are some parameters, including the ratio β in FDA that controls the range of replaced frequency components and the percentage k that controls the proportion of pseudo-labeled data for model refinement, that may need to be tuned for different datasets to get optimal performance. Second, in the joint loss function, fixed weights are used. In the future, some data-driven adaptation strategies can be explored to optimize the loss function. Third, the proposed method requires two separate training stages. We will investigate end-to-end domain adaptation methods in our future work. Furthermore, we will also explore the design of network structures and make the segmenter and discriminator more effective, to further improve the segmentation performance.

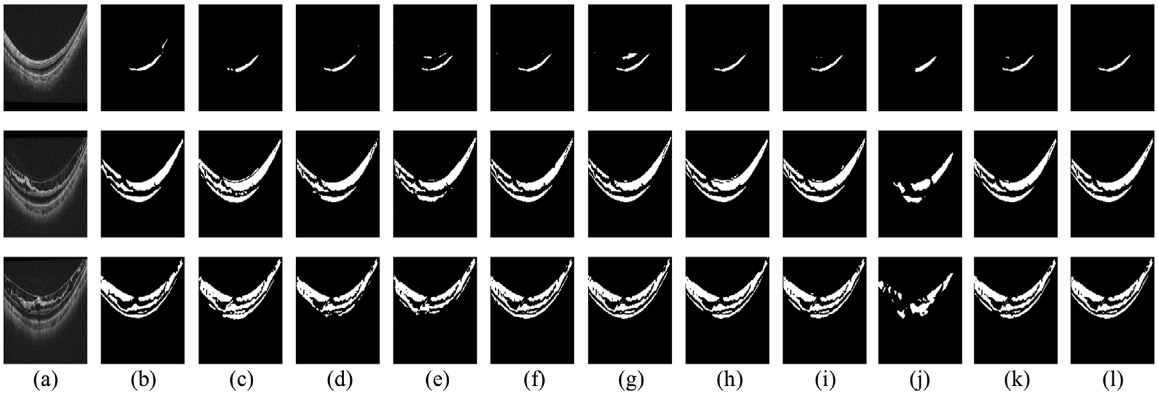

FIGURE 8 Retinoschisis segmentation results of target dataset using different algorithms (a) original image, (b) ground truth, (c) AdaptSeg, (d) AdaptPatch, (e) Advent, (f) CycleGAN, (g) CYCADA, (h) MCDDA, (i) Maxsquare, (j) SIFA, (k) VarDA, and (l) TSANet.

Although the proposed method is not end-to-end in training, in testing only the trained segmenter is needed. This ensures its efficiency when used in real clinical applications. In addition, the proposed method is potentially applicable for segmentation of other retinal layers or lesions. In clinical practice, it is often the case that the same analysis needs to be performed on images from different domains. With the proposed UDA method, for a specific segmentation task, only one set of labeled OCT data in a particular domain is needed and can be repeatedly used. When it comes to any new OCT scanner or new imaging protocol, only unlabeled images are needed to retrain the model, thus saving the great deal of efforts in manual labeling each time for a new domain. Furthermore, if a small amount of annotated data is available in the new domain, it can be readily integrated into the second stage training and can potentially improve the model performance.

In conclusion, we propose a UDA method that can effectively improve the segmentation performance when adapting the deep learning model to new OCT data, without the need of additional manual annotations.

ACKNOWLEDGMENTS

This work was supported by the National Natural Science Foundation of China (62271337, 62371326), the National Key Research and Development Program of China (2018YFA0701700, 2019FYC1710204), and the Natural Science Foundation of Jiangsu Province under Grant BK20231310.

CONFLICT OF INTEREST STATEMENT

The authors have no relevant conflicts of interest to disclose.

DATA AVAILABILITY STATEMENT

The dataset underlying the results presented in this paper is not publicly available at this time but may be obtained from the authors upon reasonable request.

REFERENCES

1. Xing G, Chen L, Wang H, et al. Multi-scale pathological fluid segmentation in OCT with a novel curvature loss in convolutional neural network. *IEEE Trans Med Imaging*. 2022;41(6):1547-1559.
2. Wang M, Shi F, Zhou Y, et al. MsTGANet: automatic drusen segmentation from retinal OCT images. *IEEE Trans Med Imaging*. 2022;41(2):394-406.
3. Guha RA, Sailesh C, Debdoot S, et al. ReLayNet: retinal layer and fluid segmentation of macular optical coherence tomography using fully convolutional network. *Biomed Opt Express*. 2017;8(8):3627-3642.
4. Lee ECW, de Boer JF, Mujat M, Lim H, Yun SH. In vivo optical frequency domain imaging of human retina and choroid. *Opt Express*. 2006;14:4403-4411.
5. Desjardins AE, Vakoc BJ, Tearney GJ, Bouma BE. Speckle reduction in OCT using massively-parallel detection and frequency-domain ranging. *Opt Express*. 2006;14:4736-4745.
6. Guan H, Liu M. Domain adaptation for medical image analysis: a survey. *IEEE Trans Biomed Eng*. 2022;69(3):1173-1185.
7. Sohn K, Liu S, Zhong G, et al. Unsupervised domain adaptation for face recognition in unlabeled videos. *Proceedings of the International Conference on Computer Vision*. 2017: 3210-3218.
8. Tzeng E, Hoffman J, Darrell T, Saenko, K. Simultaneous deep transfer across domains and tasks. *Proceedings of the International Conference on Computer Vision*. 2015:4068-4076.
9. Ganin Y, Lempitsky V. Unsupervised domain adaptation by backpropagation. *International Conference on Machine Learning*. 2015:1180-1189.
10. Ganin Y, Ustinova E, Ajakan H, et al. Domain-adversarial training of neural networks. *J Mach Learn Res*. 2016;17(1):2096-2030.
11. Rozantsev A, Salzmann M, Fua P. Beyond sharing weights for deep domain adaptation. *IEEE Trans Pattern Anal Mach Intell*. 2018;41(4):801-814.
12. Long M, Cao Y, Wang J, Jordan MI. Learning transferable features with deep adaptation networks. *International Conference on Machine Learning*. 2015:97-105.
13. Zhu J, Park T, Isola P, Efros AA. Unpaired image-to-image translation using cycle-consistent adversarial networks. *Proceedings of the International Conference on Computer Vision*. 2017:2223-2232.
14. Zhang Y, Miao S, Mansi T, Liao R. Task driven generative modeling for unsupervised domain adaptation: application to X-ray image segmentation. *Med Image Comput Comput Assist Interv*. 2018:599-607.
15. Zhang Z, Yang L, Zheng Y. Translating and segmenting multimodal medical volumes with cycle- and shape-consistency generative adversarial network. *Proceedings of the IEEE*

Conference on Computer Vision and Pattern Recognition. 2018:9242-9251.

16. Hoffman J, Tzeng E, Park T, et al. CyCADA: cycle-consistent adversarial domain adaptation. *International Conference on Machine Learning*. 2018:1989-1998.
17. Chen C, Dou Q, Chen H, Qin J, Heng PA. Unsupervised bidirectional cross-modality adaptation via deeply synergistic image and feature alignment for medical image segmentation. *IEEE Trans Med Imaging*. 2020;39(7):2494-2505.
18. Wang S, Yu L, Yang X, FU CW, Heng P-A. Patch-based output space adversarial learning for joint optic disc and cup segmentation. *IEEE Trans Med Imaging*. 2019;38(11):2485-2495.
19. Dou Q, Ouyang C, Chen C, et al. PnP-Adanet: plug-and-play adversarial domain adaptation network at unpaired cross-modality cardiac segmentation. *IEEE Access*. 2019;7:99065-99076.
20. Tsai Y, Hung W, Schulter S, Sohn K, Yang M-H, Chandraker M. Learning to adapt structured output space for semantic segmentation. *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*. 2018:7472-7481.
21. Yan W, Wang Y, Xia M, Tao Q. Edge-guided output adaptor: highly efficient adaptation module for cross-vendor medical image segmentation. *IEEE Signal Process Lett*. 2019;26(11):1593-1597.
22. Tsai Y, Sohn K, Schulter S, Chandraker M. Domain adaptation for structured output via discriminative patch representations. *Proceedings of the IEEE/CVF International Conference on Computer Vision*. 2019:1456-1465.
23. Saito K, Watanabe K, Ushiku Y, Harada T. Maximum classifier discrepancy for unsupervised domain adaptation. *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*. 2018:3723-3732.
24. Vu T, Jain H, Bucher M, Cord M, Pérez, P. ADVENT: adversarial entropy minimization for domain adaptation in semantic segmentation. *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*. 2019:2517-2526.
25. Chen M, Xue H, Cai D, Domain adaptation for semantic segmentation with maximum squares loss. *Proceedings of the International Conference on Computer Vision*. 2019:2090-2099.
26. Bian C, Yuan C, Wang J, et al. Uncertainty-aware domain alignment for anatomical structure segmentation. *Med Image Anal*. 2020;64:101732.
27. Chai Z, Zhou K, Yang J, et al. Perceptual-assisted adversarial adaptation for choroid segmentation in optical coherence tomography. *2020 IEEE 17th International Symposium on Biomedical Imaging (ISBI)*. 2020:1966-1970.
28. Chen S, Ma D, Lee S, et al. Segmentation-guided domain adaptation and data harmonization of multi-device retinal optical coherence tomography using cycle-consistent generative adversarial networks. *Comput Biol Med*. 2023;159:106595.
29. He X, Zhong Z, Fang L, He M, Sebe N. Structure-guided cross-attention network for cross-domain OCT fluid segmentation. *IEEE Trans Image Process*. 2023;32:309-320.
30. Yang Y, Soatto S. FDA: Fourier domain adaptation for semantic segmentation. *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*. 2020:4085-4095.
31. Diao S, Chen X, Xiang D, Zhu W, Fan Y, Shi F, A two-stage unsupervised domain adaptation method for OCT image segmentation. *Proceedings of SPIE, Medical Imaging 2023: Image Processing*. 2023:124643M.
32. Shi F, Cheng X, Feng S, et al. Group-wise context selection network for choroid segmentation in optical coherence tomography. *Physics in Med Biol*. 2021;66:245010.
33. Shimada N, Tanaka Y, Tokoro T, Ohno-Matsui, K. Natural course of myopic traction maculopathy and factors associated with progression or resolution. *Am J Ophthalmol*. 2013;156(5):948-957.
34. Yushkevich PA, Piven J, Hazlett HC, et al. User-guided 3D active contour segmentation of anatomical structures: significantly improved efficiency and reliability. *NeuroImage*. 2006;31(3):1116-1128.
35. Lee D. Pseudo-label: the simple and efficient semi-supervised learning method for deep neural networks. Workshop on challenges in representation learning. *ICML*. 2013;3(2):896.
36. Ronneberger O, Fischer P, Brox T. U-net: convolutional networks for biomedical image segmentation. *Medical Image Computing and Computer-Assisted Intervention*. 2015:234-241.
37. Wu F, Zhuang X. Unsupervised domain adaptation with variational approximation for cardiac segmentation. *IEEE Trans Med Imaging*. 2021;40(12):3555-3567.

How to cite this article: Diao S, Yin Z, Chen X, et al. Two-stage adversarial learning based unsupervised domain adaptation for retinal OCT segmentation. *Med Phys*. 2024;1-12. https://doi.org/10.1002/mp.17012