This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/JBHI.2021.3085770, IEEE Journal of Biomedical and Health Informatics

Unsupervised Domain Adaptation Based Image Synthesis and Feature Alignment for Joint Optic Disc and Cup Segmentation

Haijun Lei, Weixin Liu, Hai Xie, Benjian Zhao, Guanghui Yue, and Baiying Lei*, *Senior Member, IEEE*

***Abstract*—Due to the discrepancy of different devices for fundus image collection, a well-trained neural network is usually unsuitable for another new dataset. To solve this problem, the unsupervised domain adaptation strategy attracts a lot of attentions. In this paper, we propose an unsupervised domain adaptation method based image synthesis and feature alignment (ISFA) method to segment optic disc and cup on fundus images. The GAN-based image synthesis (IS) mechanism along with the boundary information of optic disc and cup is utilized to generate target-like query images, which serves as the intermediate latent space between source domain and target domain images to alleviate the domain shift problem. Specifically, we use content and style feature alignment (CSFA) to ensure the feature consistency among source domain images, target-like query images and target domain images. The adversarial learning is used to extract domain-invariant features for output-level feature alignment (OLFA). To enhance the representation ability of domain-invariant boundary structure information, we introduce the edge attention module (EAM) for low-level feature maps. Eventually, we train our proposed method on the training set of the REFUGE challenge dataset and test it on Drishti-GS and RIM-ONE_r3 datasets. On the Drishti-GS dataset, our method achieves about 3% improvement of Dice on optic cup segmentation over the next best method. We comprehensively discuss the robustness of our method for small dataset domain adaptation. The experimental results also demonstrate the effectiveness of our method. Our code is available at https://github.com/thinkobj/ISFA.**

***Index Terms*—Optic disc and cup segmentation, Unsupervised domain adaptation, Image synthesis, Adversarial learning**

This work was supported partly by National Natural Science Foundation of China (Nos. 61871274, 61801305 and 81571758), National Natural Science Foundation of Guangdong Province (Nos.2020A1515010649, 2019B1515 120029 and 2019A1515111205), Guangdong Province Key Laboratory of Popular High Performance Computers (No. 2017B030314073), Guangdong Laboratory of Artificial-Intelligence and Cyber-Economics (SZ), Shenzhen Peacock Plan (Nos. KQTD2016053112051497 and KQTD2015033016 104926), Shenzhen Key Basic Research Project (Nos. JCYJ201908 08165209410, 20190808145011259,JCYJ20180507184647636, GJHZ20190 822095414576,GCZX2017040715180580, GJHZ20180418190529516, and JSGG20180507183215520),Special Project in Key Areas of Ordinary Universities of Guangdong Province (No. 2019KZDZX1015), Guangdong Basic and Applied Basic Research Foundation(No. 2020A1515110605). Asterisk indicates corresponding author (e-mail: leiby@szu.edu.cn).

H. Lei, W. Liu and B. Zhao are with the Key Laboratory of Service Computing and Applications, Guangdong Province Key Laboratory of Popular High Performance Computers, College of Computer Science and Software Engineering, Shenzhen University, Shenzhen, China, 518060.

H. Xie, G. Yue and B. Lei are with the National- Regional Key Technology Engineering Laboratory for Medical Ultrasound, Guangdong Key Laboratory for Biomedical Measurements and Ultrasound Imaging, School of Biomedical Engineering, Health Science Center, Shenzhen University, Shenzhen, China, 518060.

I. Introduction

GLAUCOMA is one of the most important diseases leading to irreversible blindness in the world. It is estimated that the large number of people with glaucoma will reach 111.8 million by 2040 [1]. Detection and screening in the early stage are essential to preserve vision and improve life quality. Optic nerve head (ONH) [2] assessment is a practical detection and screening method for early glaucoma. However, the manual assessment method is subjective and requires clinicians with sufficient experience, which leads to the unavailable application for large-scale screening. Therefore, the objective and quantitative clinical assessment indicators are necessary, such as the morphology change degree of the disc diameter and cup diameter [3], rim to disc area ratio and the cup-to-disc ratio (CDR) [4]. Among them, CDR is commonly used to indicate the risk degree of glaucoma by clinicians. A higher risk of glaucoma is with a larger CDR value. The accurate segmentation results of optic disc (OD) and optic cup (OC) can be used to calculate vertical cup diameter (VCD) and vertical disc diameter (VDD). Since the CDR is calculated by the ratio of VCD to VDD, the accurate segmentation of OC and OD can provide assistance to doctors in terms of glaucoma screening and detection [5].

Many early studies implemented OC or OD segmentation using hand-crafted features [6-8]. These methods usually use morphological edge detection or circular transformation for detecting the OD boundary firstly and a pixel-level classification is adopted to segment OD and OC. However, these methods rely heavily on the representation ability of the hand-crafted features. Recently, the convolution neural network (CNN) achieves great success in the segmentation of OD and OC [5, 9-12]. These methods often assume that the appearance distribution of training and test images is the same. Otherwise, the deep segmentation network performs well on the training dataset (i.e., source domain) while the performance will be decreased when the trained model is applied to a new unseen test dataset (i.e., target domain). As the medical images have different characteristics due to inter-scanner and cross-modality, various fundus image datasets usually exist domain shift. As shown in Fig. 1, the fundus image datasets acquired from different devices have distinct style appearances due to the difference of parameter settings, light source intensities and image resolution ratios. Such domain shift would cause severe performance degradation of deep convolution neural networks. For instance, the M-Net [5] reached a relatively good result on the ORIGA dataset but gained an unsatisfactory performance on other fundus image datasets [13].

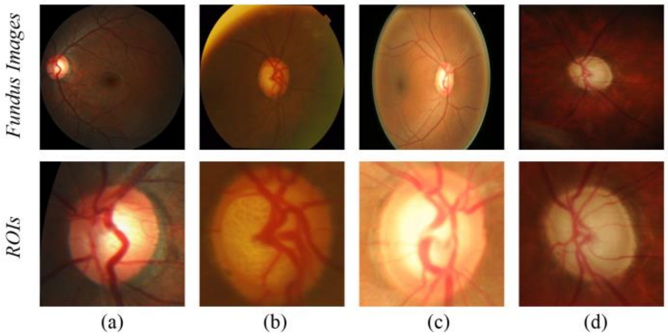

Fig. 1 Illustration of severe inter-scanner domain shift on the fundus images (top) and ROIs (bottom). From left to right: the training set of REFUGE dataset, the Drishti-GS dataset, the ORIGA dataset and the RIM-ONE_r3 dataset.

To deal with the performance degradation caused by domain shift, domain adaptation methods [14, 15] are widely studied. A straightforward approach is to train a deep model with abundant annotated target domain samples. Since it is time-consuming for professional clinicians to gain extra annotations for medical images, which leads to the insufficient application in realistic life. Therefore, unsupervised domain adaptation methods are explored to effectively transfer the knowledge learned from the labeled source domain to the unlabeled target domain. Recent advances on unsupervised domain adaptation can be summarized as image alignment [16-21] using image-to-image transformation and feature alignment [13, 14, 22-27] using adversarial learning. Image alignment mainly uses Generative Adversarial Network (GAN) [28] to map source domain samples to an intermediate latent space, which can shrink the statistical distribution gap between two different domains. Feature alignment applies an adversarial learning strategy to prediction probability maps obtained by the segmentation network in a low-dimensional space, where the probability distribution of the target domain is aligned to source domain. Considering that two kinds of alignment modes can address the domain shift complementarily, we use GAN-based image synthesis for image alignment and adversarial learning for output-level feature alignment. To obtain better target-like query images, two-stage training strategy is exploited to separate the image synthesis and feature alignment.

In this paper, we propose an unsupervised domain adaptation based image synthesis and feature alignment (ISFA) method for OD and OC segmentation. Considering that the quality of the generated image cannot be guaranteed (e.g., CycleGAN [16]), we first integrate the boundary structure information of OD and OC into the generator to obtain high-quality target-like query images, which includes the style appearance of the target domain and the content information of the source domain by a brief manual selection. Since the feature space is difficult to be directly aligned due to high-dimension [29], we use content encoder and style encoder to obtain content features and style features in two compact lower-dimensional spaces. To ensure the feature consistency among the synthesized target-like query images, source domain images and target domain images, we conduct a content and style feature alignment in these compact lower-dimensional spaces. Since the morphologic boundary of OD and OC are alike among different domain images, we integrate the edge prediction map in low-level features with high-level features for guiding the network to obtain an accurate segmentation prediction map. Furthermore, an edge attention module (EAM) [30] is leveraged to enhance the representation ability of domain-invariant edge structure information. Inspired by [26], we apply the discriminators along with adversarial learning to the edge prediction map and segmentation prediction probability map to encourage the prediction result of target-like query images and target domain images to be similar to source domain images. The main contributions of this work can be summarized as:

(1) The GAN-based image synthesis (IS) along with the boundary structure information of OD and OC is utilized to generate high-quality target-like query images, which serves as the intermediate latent space between target domain and source domain images to alleviate the domain shift.

(2) We use content and style feature alignment (CSFA) to ensure the feature consistency among the source domain images, synthesized target-like query images and target domain images. To encourage the predicted results of target-like query images and target domain images to be similar to source domain images, we employ adversarial learning to extract domain-invariant features for output-level feature alignment (OLFA).

Experimentally, we evaluate the proposed method using several fundus image datasets to prove the effectiveness. For the Drishti-GS dataset, our method achieves about 3% improvement of Dice on OC segmentation. Considering the annotation scarcity of fundus images, we comprehensively discuss the robustness of our method for small dataset domain adaptation. The experimental results also demonstrate that our method obtains superior generalization.

II. Related Work

Domain adaptation is a hot research topic since medical images have a wide range of domain shift problems due to different image acquisition equipment. Recent studies of addressing domain shift are mainly divided into two types. The first type is semi-supervised learning method, which aims to reduce the requirement of annotated target data. The other type is unsupervised domain adaptation method, which is not desired for extra labeled target data. Considering that the labeled fundus images are limited, we adopt unsupervised domain adaptation method for OD and OC segmentation. In this section, we will discuss the image synthesis and feature alignment methods that are related to our studies.

A. Image Synthesis

The image synthesis methods [31, 32] mainly map source domain samples to an intermediate latent space, which can shrink the statistical distribution gap on two different domains. The GAN is employed to synthesize realistic-like samples by style transfer techniques. In terms of visual appearance, the synthesized samples retain massive domain-invariant feature information. As the representative framework of the unpaired image-to-image translation, CycleGAN could perform cross-domain image translation in terms of both source-to-target and target-to-source direction. Zhang *et al.* applied a modified CycleGAN method for pixel-to-pixel translation [18]. The bidirectional learning method based on adaptation strategy adopted alternately training mode to obtain positive feedback from image-to-image translation [15]. Chen *et al.* [33] designed semantic-aware GAN to transform the target domain images into the appearance of source domain images. Moreover, the segmentation networks were trained with the source domain images and the trained models were directly applied to the transformed test images. In contrast, Huo *et al.* [34] use CycleGAN as the backbone to transform modality images of source domain into the appearance of target domain modality images. R-sGAN [20] synthesized target domain samples, which takes the vessel structure of source domain samples and the textural appearance of the target domain samples as inputs. In this way, the network can consider unlabeled target domain images as annotated query images. However, the vessel annotation datasets are usually very small, which is inappropriate for generating abundant fundus images. For example, DRIVE dataset only contains 20 training images and 20 test images. Furthermore, Chen *et al.* [27] separated every domain image into style features and content features. They integrated style features of target domain images and the content features of source domain images to generate target-like images. However, the quality of the generated image cannot be guaranteed, and the poor quality of the image will affect the results of the segmentation network. To address these two limitations, we leverage two-stage training strategy and the GAN-based IS along with the boundary structure information of OD and OC to obtain high-quality target-like images by a brief manual selection.

B. Feature alignment

Feature alignment is utilized to extract domain-invariant features by adversarial learning, which is also an effective method to train robust deep networks. The recent advances focus on the domain shift in terms of semantic prediction space, which aims to minimize the distribution gap between the result of the target domain and source domain. For example, a Geometrically Guided Input-Output Adaptation Approach [35] provided the domain discriminator to exploit the correlation between semantics segmentation maps and depth prediction maps. High prediction confidence regions are extracted from the input images using a Category-level Adversarial co-training strategy [36]. Tzeng *et al.* [14] combined untied weight sharing, discriminative modeling and a GAN loss to implement domain adaptation. Zhang *et al.* [37] proposed Collaborative and Adversarial Network (CAN) through domain adversarial training and domain-collaborative of the networks. Javanmardi *et al.* [38] used an adversarial training approach for vasculature segmentation for the output-level adaptation. Due to the similarity of predicted results between source and target domains, Tsai *et al.* [24] utilized an adversarial learning strategy for output space domain adaptation.

Recently, various works focus on unsupervised domain adaptation for OD and OC segmentation using adversarial learning [13, 26, 27, 39]. For example, Liu *et al.* [39] combined self-ensembling weights and adversarial learning to realize exquisite collaborative adaptation. Meanwhile, they maintained an exponential moving average of the predictions to achieve a better prediction for the target domain images. Wang *et al.* [13] designed a morphology-aware segmentation loss function to guide the segmentation network to obtain accurate prediction maps. A discriminator is used to enforce the prediction results of the target domain that are similar to the source domain. However, the segmentation prediction map obtained from the output space adaptation has large entropy, which obtain not smooth segmentation contours. Therefore, Wang *et al.* [26] devised the networks to obtain low-entropy segmentation prediction maps and boundary prediction maps using adversarial learning. To accelerate computation and reduce the number of parameters, a handy and lightweight network was adopted. Chen *et al.* [27] used image translation and adversarial learning for input and output space alignment. Inspired by the above methods, considering that the morphologic boundary of OD and OC are alike among different domain images, we firstly obtain the edge prediction map from a low-level feature map. Furthermore, we integrate it with low-level and high-level feature maps to obtain segmentation prediction map inspired by [26]. Two discriminators are utilized to the edge prediction map and segmentation prediction map for output-level feature alignment. The adversarial learning can encourage the result of target-like query images and target domain images to be similar to source domain images.

III. Methodology

In this paper, the segmentation task of the OD and OC on fundus image consists of three parts: 1) IS for input-level domain adaptation. In detail, we devise two generators named target generator G_t and reconstructed generator G_r along with the boundary information of OD and OC to synthesize target-like query images. A discriminator D_t is utilized to judge whether the synthesized images come from the target images or the target-like query images. 2) CSFA ensures feature consistency of different domains to extract domain-invariant features. The shared encoder E is utilized to extract low-level feature maps and high-level feature maps. For low-level feature maps, we use a shared content encoder E_C to extract content features of target-like query images and source domain images. Similarly, the shared style encoder E_s is used to extract style features of target-like query images and target domain images. Therefore, the low-level feature maps extracted from encoder E of target-like query images can be aligned to the source domain and target domain images, which can alleviate the domain shift on two different domains. 3) OLFA. We apply the edge decoder U_e and mask decoder U_m to obtain the edge prediction map and segmentation probability prediction map, respectively. An EAM is utilized to enhance the representation ability of domain-invariant edge structure information. The edge discriminator D_e and mask discriminator D_m are used to identify the results from source images, target-like query images or target images. The adversarial learning enforces the prediction results on the target-like query samples and target domain samples to be close to the source domain samples. The flowchart of the proposed method is shown in Fig. 2

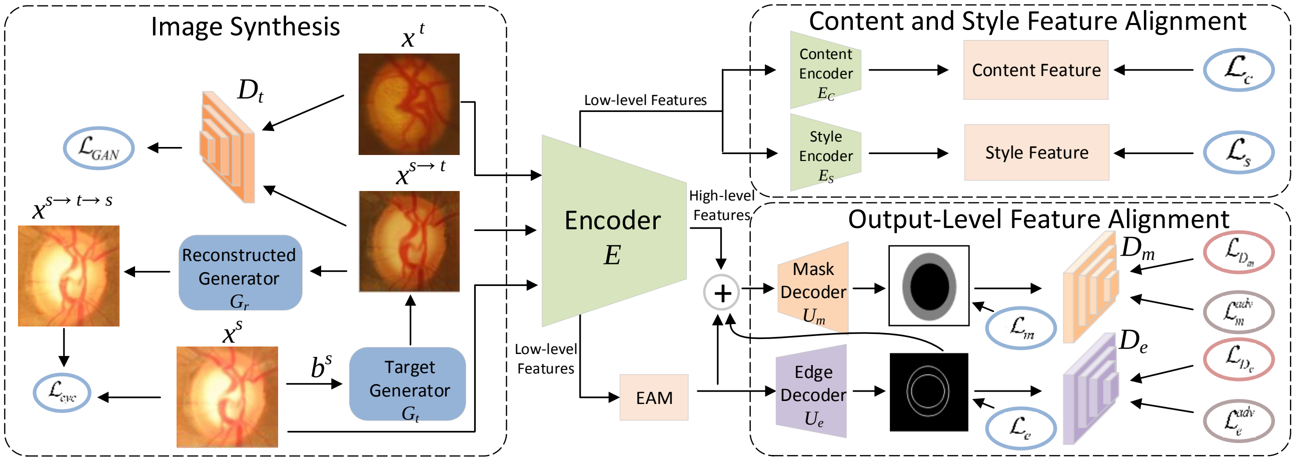

Fig. 2 Flowchart of our proposed method. For IS, the target generator G_t along with the boundary structure information b_s is utilized to generate target-like query images, while the reconstructed generator G_r is used to regain the source-like domain images. For CSFA, we use the shared content encoder E_C and style encoder $E_S$to extract content features and style features of different domains. For OLFA, we adopt adversarial learning by applying two discriminators to align the distribution of the edge prediction maps and segmentation prediction maps, respectively. An EAM is utilized to extract domain-invariant boundary structure.

A. *Image Synthesis in Input-level Adaptation*

The different visual appearances of different fundus image dataset usually triggers domain shift. The GAN-based IS method is utilized to obtain target-like query images, which serves as the intermediate latent space between target domain images and source domain images to alleviate the domain shift. To enable the content structure information of target-like query images to be more similar to the source domain images, we integrate the boundary structure information of OD and OC to the generator. We formulate a source domain image set $\mathcal{I}_S \in \mathbb{R}^{H\times W\times 3}$, a boundary set $\mathcal{B}_S \in \mathbb{R}^{H\times W}$ obtained from the ground truth $\mathcal{Y}_S \in \mathbb{R}^{H\times W}$, and a target domain image set $\mathcal{I}_T \in \mathbb{R}^{H\times W\times 3}$. For a source domain sample $x^s \in \mathcal{I}_S$ and its corresponding boundary map $b^s \in \mathcal{B}_S$, the target generator along with the boundary information of OD and OC aims to integrate the style appearance information of target domain images and content structure information of source domain images into target-like query images $x^{s\to t} = G_t(x^s, b^s)$. The discriminator D_t competes with G_t to correctly distinguish between target-like query samples $x^{s\to t}$ and target domain samples x^t. We define the objective function as:

$$\mathcal{L}_{\mathrm{GAN}} = \mathbb{E}_{x^t\sim\mathcal{I}_T}[\log D_t(x^t)] + \mathbb{E}_{x^s\sim\mathcal{I}_S}[\log(1 - D_t(G_t(x^s, b^s))] \quad (1)$$

Furthermore, we use a reconstruction generator G_r to ensure target-like query images retaining the content structure information of source domain images. The generator G_r is optimized via cycle consistency loss:

$$\mathcal{L}_{\mathrm{cyc}} = \mathbb{E}_{x^s\sim\mathcal{I}_S}[||x^{s\to t\to s} - x^s||_1] \quad (2)$$

Therefore, the total loss function of IS is:

$$\mathcal{L}_1 = \mathcal{L}_{\mathrm{GAN}} + \lambda_1\mathcal{L}_{\mathrm{cyc}} \quad (3)$$

where λ_1 is a balance coefficient ($\lambda_1 = 10$ in our experiments).

B. *Content and Style Feature Alignment*

We obtain target-like query images through the above image synthesis, which includes the style features of the target domain images and the content features of the source domain images. The target-like query images can serve as the intermediate latent space between target and source domain images. Since the feature space is difficult to be directly aligned due to high-dimension. We use content encoder E_C and style encoder E_S to obtain content features and style features in two compact lower-dimensional spaces. Therefore, CSFA is employed in low-level feature maps, which can ensure the style feature of the target-like query images to be similar to target domain images and the content feature of the target-like query images to be similar to source domain images. Furthermore, source domain images contain private style features and shared content features with target-like query images. For target domain images, we aim to learn the shared style features and private content features with target-like query images. Therefore, we use a content encoder E_C to extract content features of target-like images and source domain images. A style encoder E_S is utilized to obtain style features of target-like and target domain images. The $\mathcal{L}_1$ loss function is adopted directly in feature space to enforce the content and style feature alignment. The corresponding feature map is sparse, which can reduce the complexity and improve the generalization ability of the model. The content loss $\mathcal{L}_{\mathrm{C}}$ and style loss $\mathcal{L}_{\mathrm{S}}$ are defined as:

$$\mathcal{L}_{\mathrm{C}} = \mathbb{E}_{x^s\sim\mathcal{I}_S, x^{s\to t}\sim\mathcal{I}_{S\to T}}[||E_C(E(x^{s\to t})) - E_C(E(x^s))||_1] \quad (4)$$

$$\mathcal{L}_{\mathrm{S}} = \mathbb{E}_{x^t\sim\mathcal{I}_T, x^{s\to t}\sim\mathcal{I}_{S\to T}}[||E_S(E(x^{s\to t})) - E_S(E(x^t))||_1] \quad (5)$$

C. *Output-level Feature Alignment*

As mentioned in the last section, we adopt the shared encoder E to extract low-level and high-level feature maps, respectively. Since the morphologic boundary of OD and OC are alike among different domains, the edge decoder U_e is introduced to obtain the edge prediction map (i.e., p_e^s, $p_e^{s\to t}$ and p_e^t) of OD and OC in low-level feature maps. The edge prediction map can guide the network to obtain an accurate segmentation probability prediction map. To improve the representation ability of the network for extracting edge information from low-level feature maps, the EAM is used to suppress the noise in the foreground and highlight the edge structure information. As shown in Fig. 3, we introduce a conventional EAM inspired by [30]. We use 1×1 convolution layer to obtain low-dimension and most relevant compressed

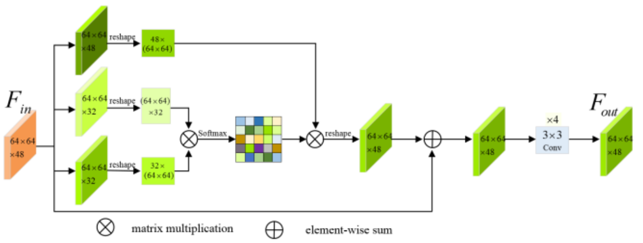

Fig. 3 Illustration of the edge attention module.

feature map (64×64×32) from original low-level feature map (64×64×48). To reduce computation and facilitate matrix multiplication, we reshape the three-dimension tensor into two-dimension tensor. Furthermore, a matrix multiplication operation and a Softmax function are utilized to get an attention map that can represent important boundary information. We integrate it into original feature maps by another matrix multiplication and an element-wise sum operation. The obtained feature map is encoded to get the final edge attention features based on four convolution layers with 3×3 kernel sizes. The edge decoder U_e is introduced to obtain the edge prediction map of OD and OC. For the high-level feature map, we concatenate it with the low-level feature maps and edge prediction map to obtain the shared feature map, which contains abundant domain-invariant information obtained from the above methods. Our mask decoder U_m is devised to produce the segmentation prediction maps (i.e., p_m^s, $p_m^{s\to t}$ and p_m^t). Specifically, the weight of the segmentation network is optimized via supervised learning and unsupervised learning.

1) *Predicted Result by Supervised Learning*

For source domain images, we use the ground truth information of them to optimize the segmentation network by supervised learning. Since the morphologic boundary information of OD and OC are alike among different domain images, the edge prediction map (i.e., p_e^s) of source domain images can serve as a predicted benchmark for other domain images. Therefore, we optimize the result of the prediction edge in low-level feature maps by the binary boundary map (i.e., b^s) obtained from the ground truth:

$$\mathcal{L}_e = \frac{1}{N}\sum(b^s - p_e^s)^2 \tag{6}$$

For the high-level feature map, we concatenate it with the low-level feature maps and edge prediction map to obtain the shared feature map. The segmentation prediction map (i.e., p_m^s) is produced by the mask decoder U_m. Similarly, we use the binary ground truth (i.e., y_m^s) to optimize the segmentation result:

$$\mathcal{L}_m = 1 - \frac{2\sum p_m^s y_m^s}{\sum(p_m^s)^2 + (y_m^s)^2} \tag{7}$$

2) *Feature Alignment by Unsupervised Adversarial Learning*

With the above image synthesis, we obtain several target-like query images. Compared with the source domain samples, the target-like query samples are with the content structure information of the source domain samples and the style representation information of the target domain samples, which can help the network learn the domain-invariant features. Considering the rigor of segmentation network training, we still regard them as unlabeled samples and apply unsupervised adversarial learning to implement feature alignment along with target domain samples. Specifically, we construct an edge discriminator D_e to classify the edge prediction map corresponding to x^s, $x^{s\to t}$ or x^t. If the features obtained from $x^{s\to t}$ and x^t are domain-invariant with the features extracted from x^s, the discriminator D_e is not able to classify their corresponding edge prediction map, for the reason that the morphologic boundary of OD and OC are consistent. Otherwise, the network weights of the shared encoder E and edge decoder U_e will be updated by the adversarial gradients, which aims to minimize the discrepancy between the feature distributions from $x^{s\to t}$ and x^t to x^s:

$$\mathcal{L}_{D_e} = \frac{1}{N}\sum_{x^s\in\mathcal{I}_S}\mathcal{L}_D(p_e^s,1) + \frac{1}{M}\sum_{x^t\in\mathcal{I}_T}\mathcal{L}_D(p_e^t,0) + \frac{1}{K}\sum_{x^{s\to t}\in\mathcal{I}_{S\to T}}\mathcal{L}_D(p_e^{s\to t},0) \tag{8}$$

where M and K are the total sample size of target domain images and target-like query images, respectively. $\mathcal{L}_D$ is the binary cross-entropy loss. Furthermore, although the fundus image datasets acquired from different devices have distinct style appearances, the morphologic boundary of OD and OC are alike among different domains. To make the $p_e^{s\to t}$ and p_e^t similar to the p_e^s, we confuse the edge discriminator D_e to judge the $p_e^{s\to t}$ and p_e^t as true. Therefore, the edge adversarial loss is employed to align the edge structure distribution on target domain samples and target-like query samples with the source domain samples:

$$\mathcal{L}_e^{adv} = \frac{1}{M}\sum_{x^t\in\mathcal{I}_T}\mathcal{L}_D(p_e^t,1) + \frac{1}{K}\sum_{x^{s\to t}\in\mathcal{I}_{S\to T}}\mathcal{L}_D(p_e^{s\to t},1) \tag{9}$$

Similarly, we construct a mask discriminator D_m to classify the segmentation prediction map corresponding to x^s, $x^{s\to t}$ or x^t. If the features obtained from $x^{s\to t}$ and x^t are domain-invariant with features extracted from x^s, the discriminator D_m is not able to classify their corresponding segmentation prediction map. Otherwise, the segmentation network would be optimized by:

$$\mathcal{L}_{D_m} = \frac{1}{N}\sum_{x^s\in\mathcal{I}_S}\mathcal{L}_D(p_m^s,1) + \frac{1}{M}\sum_{x^t\in\mathcal{I}_T}\mathcal{L}_D(p_m^t,0) + \frac{1}{K}\sum_{x^{s\to t}\in\mathcal{I}_{S\to T}}\mathcal{L}_D(p_m^{s\to t},0) \tag{10}$$

Additionally, for target-like query images and target domain images, segmentation prediction maps (i.e., $p_m^{s\to t}$ and p_m^t) are optimized by adversarial learning. To make the $p_m^{s\to t}$ and p_m^t similar to the p_m^s, we confuse the mask discriminator D_m to judge the $p_m^{s\to t}$ and p_m^t as true. The mask adversarial loss can encourage the segmentation prediction map on target domain samples and target-like query samples with the source domain samples:

$$\mathcal{L}_m^{adv} = \frac{1}{M}\sum_{x^t\in\mathcal{I}_T}\mathcal{L}_D(p_m^t,1) + \frac{1}{K}\sum_{x^{s\to t}\in\mathcal{I}_{S\to T}}\mathcal{L}_D(p_m^{s\to t},1) \tag{11}$$

The encoder-decoder structures and discriminators are optimized alternately. The objective function of the encoder-decoder networks can be summarized as:

$$\mathcal{L}_2 = \mathcal{L}_e + \mathcal{L}_m + \mathcal{L}_C + \mathcal{L}_S + \lambda_2\left(\mathcal{L}_e^{adv} + \mathcal{L}_m^{adv}\right) \tag{12}$$

where λ_2 is a balance coefficient ($\lambda_2 = 0.01$ in our experiments).

IV. Experimental and Results

A. *Datasets and Evaluation Metrics*

To better illustrate the effectiveness of our proposed method, we use several public retinal fundus image datasets: the training set of the REFUGE dataset [40], Drishti-GS dataset [41], RIM-ONE_r3 dataset [42] and ORIGA dataset [43]. The 400 images of training set of the REFUGE dataset are acquired by a Zeiss Bisucam 500 fundus camera with a resolution of 2124×2056 pixels. The RIM-ONE_r3 dataset are acquired by a Cannon EOS 5D fundus camera with a resolution of 1072×1424, which is split into 99 training images and 60 test images. Although the Drishti-GS dataset and the ORIGA dataset do not give the information of the fundus cameras, the fundus image on these datasets have distinct style appearances, as shown in Fig.1. The Drishti-GS dataset contains 50 training images and 51 test images. For ORIGA dataset, it is not divided into training data and test data at the stage of collection. In this paper, we split it into 500 training images and 150 test images empirically. The detailed statistical distribution of the four datasets is shown in Table I. In the comparative experiments, the training set of the REFUGE dataset is used as the source domain. The Drishti-GS dataset and RIM-ONE_r3 dataset are used as the target domain. In the discussion of small dataset adaptation, since the images number of RIM-ONE_r3 dataset and Drishti-GS dataset are small, we use one of the dataset as the source domain and the other dataset combined with ORIGA dataset as the target domain.

We use the Dice similarity coefficient (Dice), Mean Intersection over Union (MIoU), mean absolute CDR error (MAE) and Pixel Accuracy (PA) to evaluate the segmentation performance of network in pixel-level and region-level:

$$Dice = \frac{2\times TP}{(TP+FN)+(TP+FP)} \tag{13}$$

$$MIoU = \frac{1}{2}\left(\frac{TP}{TP+FP+FN} + \frac{TN}{TN+FP+FN}\right) \tag{14}$$

$$MAE = \sum_{c=0}^{C} \left| \frac{VD^{p}_{cup}}{VD^{p}_{disc}} - \frac{VD^{g}_{cup}}{VD^{g}_{disc}} \right| \tag{15}$$

$$PA = \frac{TP+TN}{TP+TN+FP+FN} \tag{16}$$

where TP, FP, FN, TN indicates true positive, false positive, false negative and true negative, respectively. C is the image number of the test set. VD^{g}_{cup} and $VD^{p}_{cup}$are the VCD of the ground truth and segmentation prediction map, respectively. Similarly, VD^{g}_{disc} and $VD^{p}_{disc}$are the VDD of the ground truth and segmentation prediction map, respectively. Noted that the closer to 0, the better for MAE and the closer to1, the better for Dice, MIoU and PA.

B. Network Architecture and Training procedure

Our network structure mainly consists of three parts: IS, CSFA and OLFA. Considering that the quality of generated images acquired from IS will affect feature alignment and adversarial learning. We train the network of IS solely and select high-quality images as target-like query images by a brief manual selection. The target generator G_t and reconstructed generator G_r have the same architectures, which contain of three convolutional layers, nine residual blocks and two deconvolutional layers. All the residual blocks and the convolutional layers are equipped with Instance Normalization. The settings of the discriminator D_t are similar to PatchGAN [44], which output prediction map with one channel. It contains three convolutional layers along with a LeakyReLU activation function.

We use MobilenetV2 [45] as the encoder E to extract the shared low-level feature maps and high-level feature maps. The content encoder E_C is composed of 3 convolutional layers and 1 residual blocks together with an Instance Normalization and a ReLU activation function. The style encoder consists of 5 convolutional layers, 1 global average pooling layer. The edge decoder U_e and mask decoder U_m consists of 2 convolutional layers with 3×3 kernel size. Each convolutional layer is followed by a batch normalization and a ReLU activation function, respectively. Finally, we use a 1×1 convolutional layer to obtain the prediction map. The edge discriminator D_e and mask discriminator D_m have the same architectures, which contain five convolutional layers with 4×4 kernel sizes and a LeakyReLU activation function.

C. Implementation Details

Our framework is implemented with PyTorch library. The generators, encoder-decoder networks and a discriminator (i.e., G_t, G_r, D_t, E, E_C, E_S, U_e, U_m) are optimized with Adam algorithm, while the SGD algorithm is utilized to optimize the discriminator D_e and D_m. We define the initial learning rate of G_t, G_r and D_t as 0.0002. The learning rate of the E, E_C, E_S, U_e, U_m is 0.001 and it will be multiplied by 0.1 every 100 epochs for 400 epochs in total. The learning rate of the discriminator D_e and D_m are fixed as $2.5e^{-5}$. As shown in Table I, the image size of different datasets is different, which can affect the segmentation network to obtain accurate results. Therefore, we crop the original fundus image to obtain 512×512 ROIs around OD as shown in Fig. 1. We further resize it to 256×256 for saving memory and computation time. We use common data augmentations such as adding Gaussian noise, elastic transformation, contrast adjustment, random erasing, random flip, random rotation and random scale in the stage of training [13].

D. Segmentation Results

Considering the quality of generated images obtained from IS will affect the subsequent feature alignment and adversarial learning, we train the IS solely and select high-quality images as target-like query images by a brief manual selection. We compare the generated results of several IS method as shown in Fig. 4. We can see that IOSUDA retains the content information of the source domain images, but it has a poor ability to transfer the style appearance information of the target domain images to generated images. As shown in the 4th column, the generated image using CycleGAN has different content information from the source domain image, which

TABLE I
THE STATISTICAL DISTRIBUTION OF THE USED DATASETS.

Datasets	REFUGE(training)	Drishti-GS	RIM-ONE_r3	ORIGA
Image size	2124×2056	2047×1760	1072×1424	3072×2048
Camera	Zeiss Viscucam 50	unknown	Canon EOS 5D	unknown
Number of training images	400	50	99	500
Number of test images	0	51	60	150

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/JBHI.2021.3085770, IEEE Journal of Biomedical and Health Informatics

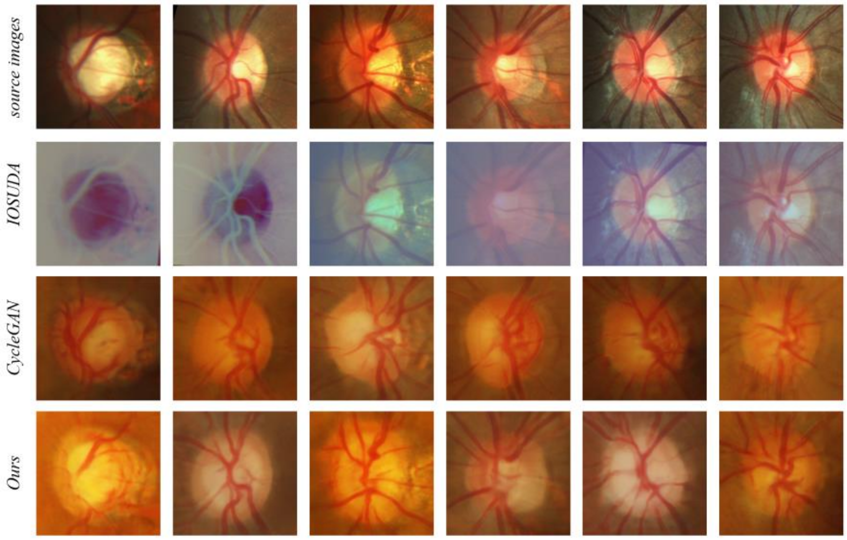

Fig. 4 The comparative results of the generated fundus images of IOSUDA, CycleGAN and our proposed method.

would mislead the content feature alignment. Compared to the above methods, the generated images using our method can effectively retain the content structure information of the source domain images and integrate the style representation information of the target domain images.

For OD and OC segmentation, we evaluate our proposed method with other six unsupervised domain adaptation methods. TD-GAN [18] presented the cross-modality domain adaptation method, which contains a modified CycleGAN method for pixel-to-pixel translation. Hoffman *et al.* [46] proposed a latent feature alignment method to solve the domain

TABLE II
THE SEGMENTATION PERFORMANCE OF DIFFERENT METHODS ON THE DRISHTI-GS DATASET AND RIM-ONE_r3 DATASET.

Method	**RIM-ONE_r3**		**Drishti-GS**	
	Dice$_{cup}$	**Dice$_{disc}$**	**Dice$_{cup}$**	**Dice$_{disc}$**
TD-GAN	0.728	0.853	0.747	0.924
Hoffman *et al.*	0.755	0.852	0.851	0.959
Javanmardi *et al.*	0.779	0.853	0.849	0.961
OSAL-pixel	0.778	0.854	0.851	0.962
pOSAL	0.787	0.865	0.858	0.965
BEAL	0.810	0.898	0.862	0.961
ISFA(Ours)	**0.822**	**0.908**	**0.892**	**0.966**

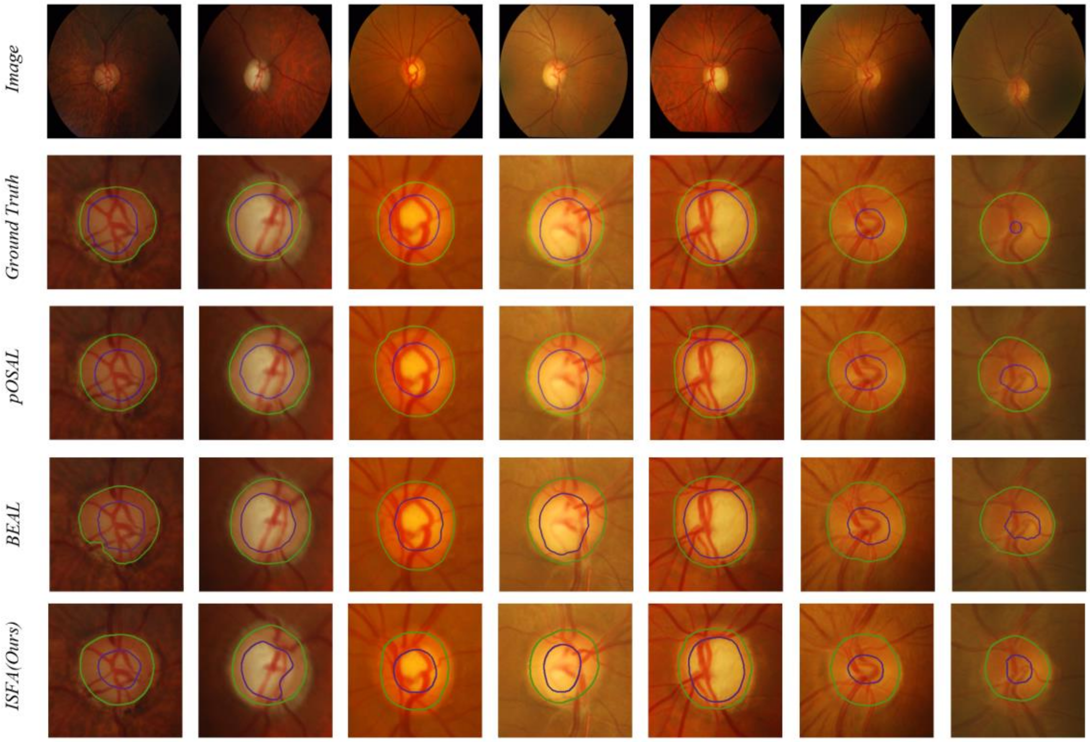

Fig. 5 Visualization of segmentation results on the Drishti-GS dataset. From top to bottom: original image (1st row), ROIs with ground truth (2nd row), the segmentation results of *pOSAL* (3rd row), BEAL (4th row) and our ISFA (5th row). The region enclosed by the blue and green curve denotes the boundary of OC and OD, respectively.

shift problem. Javanmardi *et al.* [38] used adversarial learning to implement the domain adaptation problem of vessel segmentation for the fundus images. OSAL-pixel [13], *p*OSAL [13] and BEAL [26] are three state-of-the-arts methods, which are focused on the OD and OC segmentation using the unsupervised domain adaptation. The results of the mentioned methods are inherited from the previous work [26]. For the OD and OC segmentation, our proposed method outperforms the state-of-the-art methods, as reported in Table II. For the RIM-ONE_r3 dataset, our proposed method achieves about 1% Dice improvement of the OD and OC segmentation. Specifically, our method achieves about 3% improvement of Dice on OC segmentation on the Drishti-GS dataset. The visualization of segmentation results of the OD and OC on the Drishti-GS dataset are shown in Fig. 5. Specifically, for the image in the last column, it is quite hard to distinguish OC for other methods due to the sheltered vascular, while our method is still able to segment OC with accurate boundaries, which is closest to the ground truth. Accordingly, we can observe that the segmentation boundaries of our method are closer to the ground truth.

E. Ablation Study

We conduct an ablation study to estimate the effectiveness of four key modules in our ISFA. We similarly use the training set of the REFUGE dataset as the source domain. The Drishti-GS dataset and RIM-ONE_r3 dataset are utilized as the target domain. Specifically, the following six results are utilized to complete the OD and OC segmentation: baseline, baseline + IS, baseline + IS + CSFA, baseline + OLFA, baseline + OLFA + EAM, ISFA. MoblienetV2 is also used as the baseline. As shown in Table III, IS, OLFA and EAM can improve both OD and OC segmentation results. The OC segmentation result is

TABLE III
ABLATION STUDY ON EVERY MODULE OF OUR PROPOSED METHOD.

Target Domain	baseline	IS	CSFA	EAM	OLFA	Dice_{disc}	Dice_{cup}	MIoU_{disc}	MIoU_{cup}	PA_{disc}	PA_{cup}
Drishti-GS	✓					0.9513	0.8642	0.9202	0.8537	0.9593	0.9522
	✓	✓				0.9560	0.8671	0.9269	0.8551	0.9631	0.9528
	✓	✓	✓			0.9538	0.8877	0.9237	0.8750	0.9614	0.9596
	✓				✓	0.9621	0.8771	0.9374	0.8651	0.9687	0.9569
	✓			✓	✓	0.9620	0.8866	0.9372	0.8742	0.9687	0.9598
	✓	✓	✓	✓	✓	**0.9658**	**0.8921**	**0.9433**	**0.8798**	**0.9718**	**0.9615**
RIM-ONE_r3	✓					0.8720	0.8069	0.8398	0.8246	0.9289	0.9723
	✓	✓				0.9026	0.8109	0.8788	0.8276	0.9501	0.9736
	✓	✓	✓			0.9015	0.8171	0.8776	0.8327	0.9496	0.9743
	✓				✓	0.8722	0.8140	0.8403	0.8299	0.9295	0.9734
	✓			✓	✓	0.8921	0.8198	0.8648	0.8345	0.9424	0.9745
	✓	✓	✓	✓	✓	**0.9076**	**0.8218**	**0.8841**	**0.8373**	**0.9521**	**0.9754**

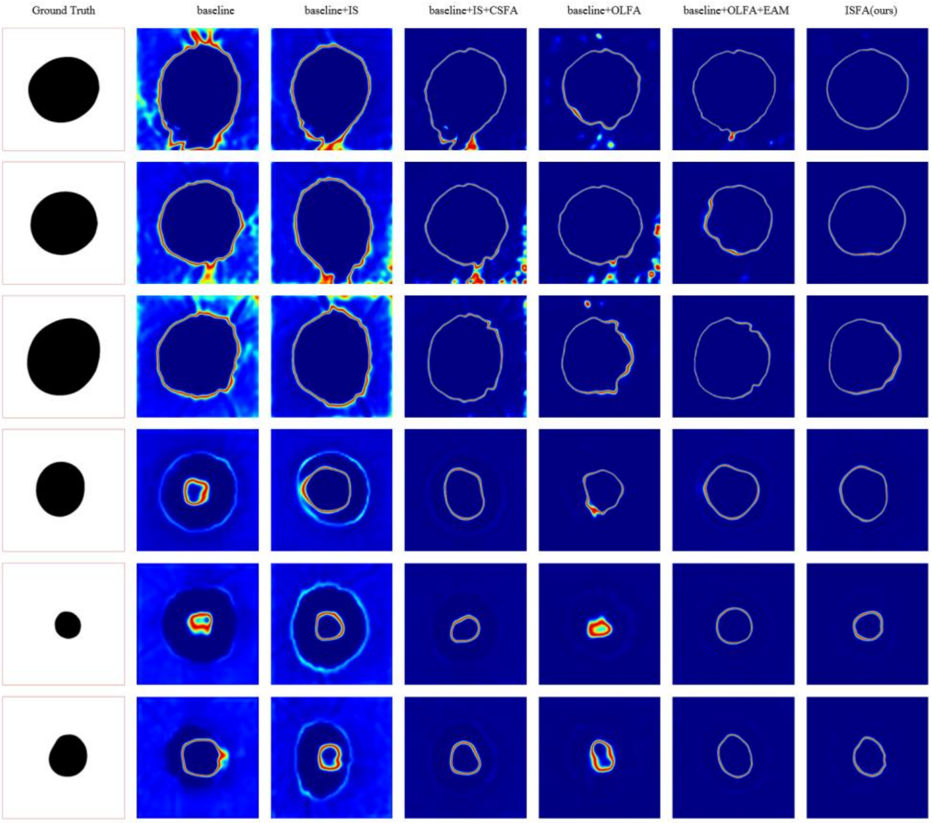

Fig. 6 The entropy map of the ablation study on the Drishti-GS test set.

improved using CSFA. Fig. 6 shows a representative entropy map of the ablation study on the Drishti-GS test set. The first three rows are the results of OD, and the last three rows represent the results of OC. For the results of OD, the entropy values of the OD boundary prediction map is reduced using IS. For the results of OC, CSFA can reduce the influence of OD boundary on OC boundary prediction. Intuitively, EAM can convert the high-entropy prediction map to the low-entropy prediction map, which suppresses the noise in the foreground and highlight the edge structure information effectively. The results integrated with four modules are with the lowest entropy value, which indicates the advance of our four key modules.

V. Discussions

A. Small datasets domain adaptation

The number of retinal fundus images with annotation is quite limited, which makes the network model too difficult to train, where the overfitting problem occurs easily. To demonstrate the good performance of our proposed method for the domain adaptation with respect to small dataset, we also conduct sufficient experiments. Since the image number of RIM-ONE_r3 dataset and Drishti-GS dataset is small, we use one of the datasets as the source domain and the other dataset along with ORIGA dataset as the target domain.

As shown in Tables IV and V, for the domain adaptation between Drishti-GS dataset and RIM-ONE_r3 dataset, our proposed method outperforms other two unsupervised domain adaptation methods. Specifically, the Dice of OC segmentation reaches 0.8385 and 0.8220. For the ORIGA dataset, the result demonstrates that our ISFA also obtains the best result on OD segmentation, while the performance on OC segmentation has a slight decrease than *p*OSAL and BEAL. Fig. 7 shows the OD and OC segmentation results of the Drishti-GS dataset and ORIGA dataset using RIM_ONE_r3 dataset as the source domain. The four left columns show the results from the Drishti-GS dataset, and the results on the four right columns are

Table IV
THE OD and OC SEGMENTATION PERFORMANCE ON THE DRISHTI-GS DATASET AND ORIGA DATASET.

Source Domain	Target Domain	Model	Dice_{disc}	Dice_{cup}	MIoU_{disc}	MIoU_{cup}	PA_{disc}	PA_{cup}
RIM-ONE_r3	Drishti-GS	*p*OSAL	0.9262	0.8331	0.8875	0.8340	0.9440	0.9509
		BEAL	0.9273	0.7992	0.8891	0.8053	0.9457	0.9421
		ISFA(ours)	**0.9324**	**0.8385**	**0.8961**	**0.8413**	**0.9492**	**0.9549**
	ORIGA	*p*OSAL	0.9458	0.8450	0.9151	0.8426	0.9578	0.9542
		BEAL	0.9074	**0.8491**	0.8658	**0.8476**	0.9333	**0.9566**
		ISFA(ours)	**0.9481**	0.8449	**0.9192**	0.8439	**0.9604**	0.9558

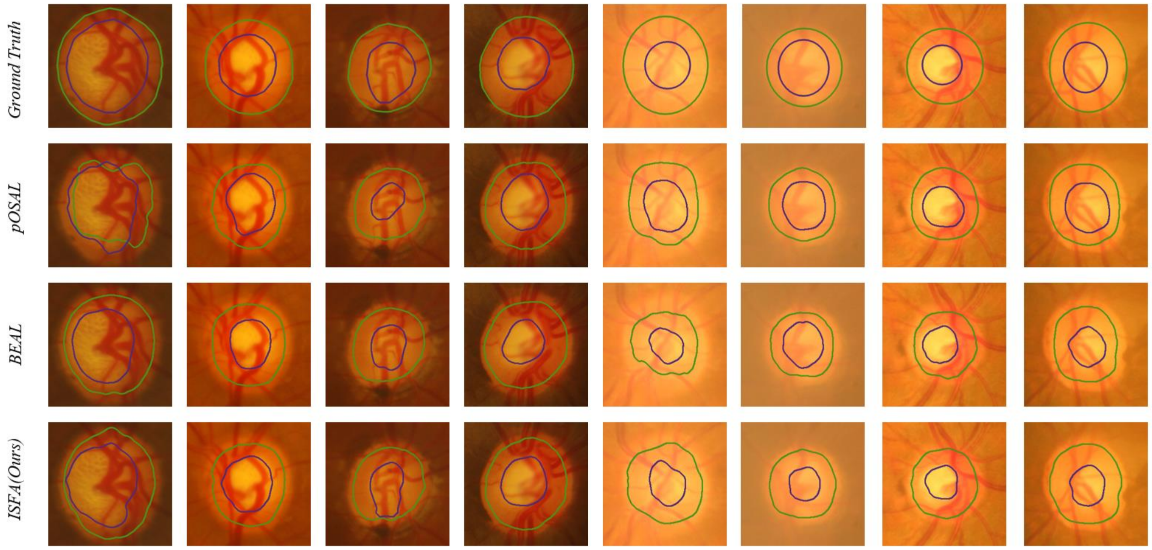

Fig. 7 Visualization of OD and OC segmentation results on the Drishti-GS dataset and ORIGA dataset. The four left columns show the results from the Drishti-GS dataset, and the results on the four right columns are from the ORIGA dataset. From top to bottom: ROIs with ground truth (1st row), the segmentation results of *pOSAL* (2nd row), BEAL (3rd row) and our ISFA (4th row). The region enclosed by the blue and green curve denotes the boundary of OC and OD, respectively.

Table V
THE OD and OC SEGMENTATION PERFORMANCE ON THE RIM-ONE_r3 DATASET AND ORIGA DATASET

Source Domain	Target Domain	Model	Dice_{disc}	Dice_{cup}	IOU_{disc}	IOU_{cup}	PA_{disc}	PA_{cup}
Drishti-GS	RIM-ONE_r3	*p*OSAL	0.9191	0.8131	0.8994	0.8412	0.9594	0.9787
		BEAL	0.9236	0.7980	0.9037	0.8322	0.9609	0.9782
		ISFA(ours)	**0.9252**	**0.8220**	**0.9059**	**0.8471**	**0.9616**	**0.9804**
	ORIGA	*p*OSAL	0.9491	0.8548	0.9195	0.8515	0.9601	0.9569
		BEAL	0.9536	0.8684	0.9255	0.8641	0.9630	0.9600
		ISFA(ours)	**0.9587**	**0.8728**	**0.9339**	**0.8676**	**0.9674**	**0.9616**

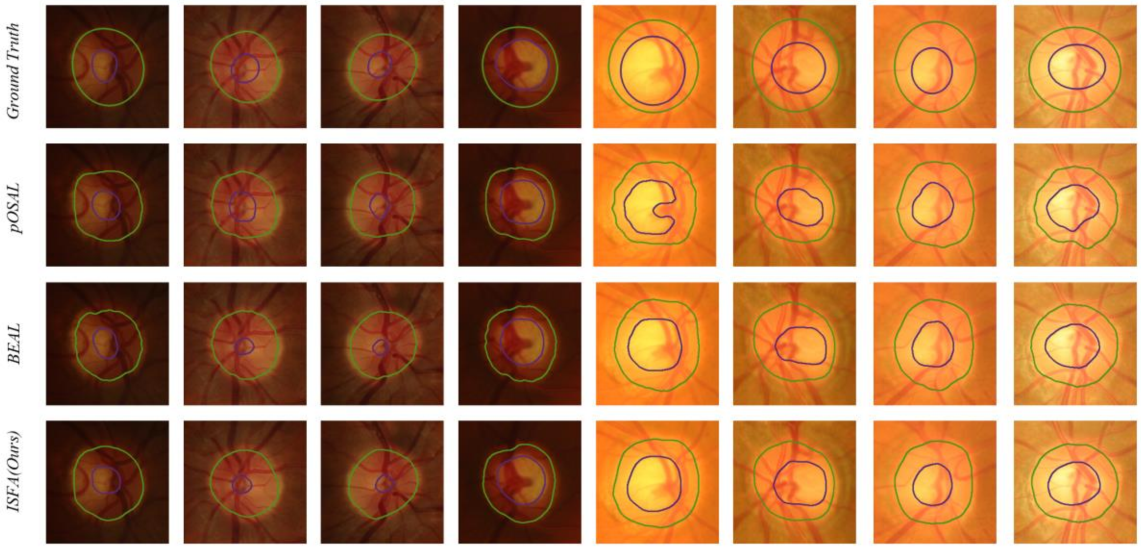

Fig. 8 Visualization of OD and OC segmentation results on the RIM-ONE_r3 dataset and ORIGA dataset. The four left columns show the results from the RIM-ONE_r3 dataset, and the results on the four right columns are from the ORIGA dataset. From top to bottom: ROIs with ground truth (1st row), the segmentation results of *p*OSAL (2nd row), BEAL (3rd row) and our ISFA (4th row). The region enclosed by the blue and green curve denotes the boundary of OC and OD, respectively.

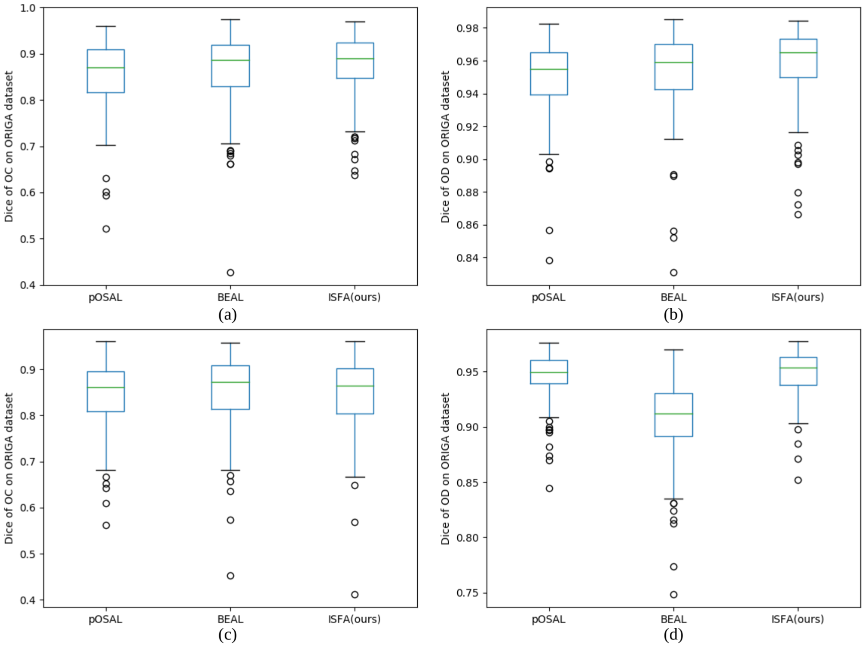

Fig. 9 Boxplot of Dice values of *p*OSAL, BEAL and our ISFA on the test part of the ORIGA datasets. (a) and (b) use the Drishti-GS dataset as the source domain. (c) and (d) use the RIM-ONE_r3 dataset as the source domain.

from the ORIGA dataset. Fig. 8 shows the OD and OC segmentation results of the RIM_ONE_r3 dataset and ORIGA dataset when Drishti-GS is used as the source domain. The four left columns show the results from the RIM_ONE_r3 dataset, and the results on the four right columns are from the ORIGA dataset. From top to bottom: ROIs with ground truth (1st row), the segmentation results of *p*OSAL (2nd row), BEAL (3rd row) and our proposed method (4th row). The regions enclosed by the blue and green curves denote the boundary of OC and OD, respectively. The visualization of OD and OC segmentation results also demonstrates that the performance of our proposed method outperforms other methods.

To demonstrate the effectiveness of our proposed method, we make statistical analyses in terms of Dice values of OD and OC, respectively. Fig. 9 shows the Dice values on the test part of the ORIGA using boxplot. In Fig. 9 (a) and 9(b), our ISFA has a smaller quartile and a larger median, which depicts that the overall OD and OC segmentation results are significantly better on the Drishti-GS source domain dataset. The result is similar to that in TABLE V. In Fig. 9 (c), the performance of other methods is with a smaller quartile, while our method consists of fewer outliers. Similarly, the Dice values of OD also have fewer outliers in Fig 9 (d). According to the above analysis, we can conclude that the image synthesis and feature

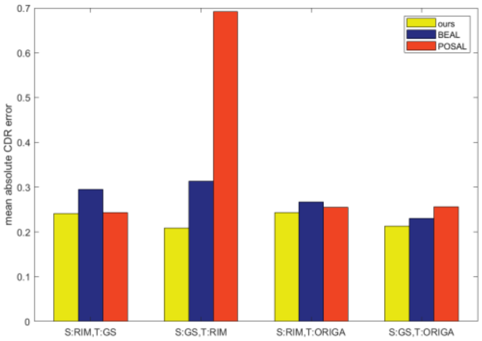

Fig. 10 The mean absolute CDR error results on small dataset domain adaptation. The capital letters S and T represent the source domain and target domain, respectively.

alignment method are beneficial to small dataset domain adaptation.

The accurate segmentation results of OD and OC can be used to calculate the CDR to provide assistance to doctors in terms of glaucoma screening and detection. Therefore, we adopt the MAE to evaluate the effectiveness of our method on the domain adaptation for small dataset. As shown in Fig. 10, we use the histogram with red, blue and yellow to show the MAE of *p*OSAL, BEAL and our proposed method, respectively. When using the RIM-ONE_r3 dataset as the source domain and Drishti-GS as the target domain, the performance of *p*OSAL is close to that of our method. However, when the source and target domains are interchanged, *p*OSAL method has an unsatisfactory performance. Similarly, when using the RIM-ONE_r3 dataset as the source domain or the Drishti-GS dataset as the source domain, we can see that the mean absolute error of our method is smaller than other methods on the ORIGA dataset. Therefore, the segmentation results obtained by our method are more helpful to provide assistance to doctors in terms of glaucoma screening and detection.

B. Limitation and Future Works

Although the proposed method has a good segmentation performance of OD and OC on different domain datasets, there exist some limitations in this paper. These limitations can be summarized into two aspects. Firstly, since the quality of image synthesis can affect feature alignment and adversarial learning, we use a two-stage training strategy by training the image synthesis solely. Although this training strategy can help select better synthetic images, it also increases the time and computational cost of the training network. Chen *et al.* [27] separated every domain image into style features and content features for input space alignment. They use input space and output space alignment to address domain shift in an end-to-end framework, which sets an appropriate number to stop the training of image synthesis. Inspired by them, we aim to devise a loss function to align the content structure information and style appearance information among source domain images, generated target-like query images and target domain images in image synthesis. Secondly, it is limited to apply the domain adaptation method from a single source domain to abundant unseen target datasets, for the reason that the fundus images are collected by different devices, which usually have the problem of domain shift. To improve the generalization of network, Wang *et al.* [47] proposed a novel domain-oriented feature embedding framework, which use multi-source domains to train the networks. However, it is hard to acquire from professional ophthalmologists to annotate multiple datasets. Inspired by them, we aim to use GAN mechanism to generate multi-target-like domain images, which is beneficial to extract domain-invariant features more effectively.

Ultra-widefield fundus images can cover a wide 200º field-of-view of the retinal, while the original fundus images only provide 30º-60º area of the retinal [48]. Ultra-widefield fundus images allow more clinically relevant lesions to be detected. In the future, we would like to explore how to extend our method to address the cross-modality adaptation between ultra-widefield images and original fundus images.

VI. Conclusions

We propose a novel unsupervised domain adaptation based image synthesis and feature alignment for OD and OC segmentation using fundus images. Specifically, we leverage the GAN-based IS along with the boundary information of OD and OC to generate high-quality target-like query samples, which serves as the intermediate latent space between target domain images and source domain images to alleviate the domain shift. CSFA is used to ensure feature consistency to align consistent content features and style features of source domain images, synthesized target-like query images and target domain images. For OLFA, an EAM is utilized in low-level feature maps to enhance the representation ability of boundary structure information. In addition, the adversarial learning is employed to extract domain-invariant features. The comparative experimental results and ablation studies demonstrate that our method outperforms other unsupervised domain adaptation methods in terms of the segmentation task of OD and OC. For the Drishti-GS dataset, our method achieves about 3% improvement of Dice value on OC segmentation. Moreover, our proposed method is more effective in small dataset adaptation.

References

[1] Y.-C. Tham, X. Li, T. Y. Wong, H. A. Quigley, T. Aung, and C.-Y. Cheng, "Global Prevalence of Glaucoma and Projections of Glaucoma Burden through 2040: A Systematic Review and Meta-Analysis," *Ophthalmology*, vol. 121, pp. 2081-2090, 2014.

[2] D. Garway-Heath and R. Hitchings, "Quantitative evaluation of the optic nerve head in early glaucoma," *Br. J. Ophthalmol.*, vol. 82, pp. 352-361, 1998.

[3] Michael D. and Hancox O.D., "Optic disc size, an important consideration in the glaucoma evaluation," *Clin. Eye Vis. Care,* vol. 11, pp. 59-62, 1999.

[4] J. B. Jonas, A. Bergua, P. Schmitz–Valckenberg, K. I. Papastathopoulos, W. M. Budde, "Ranking of optic disc variables for detection of glaucomatous optic nerve damage," *Invest. Ophthalmol.,* vol. 41, pp. 1764-1773, 2000.

[5] H. Fu, J. Cheng, Y. Xu, D. W. K. Wong, J. Liu, and X. Cao, "Joint optic disc and cup segmentation based on multi-label deep network and polar transformation," *IEEE Trans. Med. Imaging,* vol. 37, pp. 1597-1605, 2018.

[6] A. Aquino, M. E. Gegúndez-Arias, and D. Marín, "Detecting the optic disc boundary in digital fundus images using morphological, edge detection, and feature extraction techniques," *IEEE Trans. Med. Imaging,* vol. 29, pp. 1860-1869, 2010.

[7] S. Lu, "Accurate and efficient optic disc detection and segmentation by a circular transformation," *IEEE Trans. Med. Imaging,* vol. 30, pp. 2126-2133, 2011.

[8] J. Cheng, J. Liu, Y. Xu, F. Yin, D. W. K. Wong, N.-M. Tan, *et al.*, "Superpixel classification based optic disc and optic cup segmentation for glaucoma screening," *IEEE Trans. Med. Imaging,* vol. 32, pp. 1019-1032, 2013.

[9] Z. Gu, J. Cheng, H. Fu, K. Zhou, H. Hao, Y. Zhao, *et al.*, "Ce-net: Context encoder network for 2d medical image segmentation," *IEEE Trans. Med. Imaging,* vol. 38, pp. 2281-2292, 2019.

[10] S. M. Shankaranarayana, K. Ram, K. Mitra, and M. Sivaprakasam, "Joint optic disc and cup segmentation using fully convolutional and adversarial networks," in *Proc. Int. Workshop on Fetal, Infant and Ophthalmic Med. Image Anal.*, 2017, pp. 168-176.

[11] J. G. Zilly, J. M. Buhmann, and D. Mahapatra, "Boosting convolutional filters with entropy sampling for optic cup and disc image segmentation from fundus images," in *Int. Workshop on Mach. Learn. in Med. Imaging*, 2015, pp. 136-143.

[12] S. Zhang, H. Fu, Y. Yan, Y. Zhang, Q. Wu, M. Yang, *et al.*, "Attention guided network for retinal image segmentation," in *Proc. Int. Conf. Med. Image Comput. Comput.-Assisted Intervention*, 2019, pp. 797-805.

[13] S. Wang, L. Yu, X. Yang, C.-W. Fu, and P.-A. Heng, "Patch-based output space adversarial learning for joint optic disc and cup segmentation," *IEEE Trans. Med. Imaging*, vol. 38, pp. 2485-2495, 2019.

[14] E. Tzeng, J. Hoffman, K. Saenko, and T. Darrell, "Adversarial discriminative domain adaptation," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, 2017, pp. 7167-7176.

[15] Y. Li, L. Yuan, and N. Vasconcelos, "Bidirectional learning for domain adaptation of semantic segmentation," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, 2019, pp. 6936-6945.

[16] J.-Y. Zhu, T. Park, P. Isola, and A. A. Efros, "Unpaired image-to-image translation using cycle-consistent adversarial networks," in *Proc. IEEE Int. Conf. Comput. Vis.*, 2017, pp. 2223-2232.

[17] P. Russo, F. M. Carlucci, T. Tommasi, and B. Caputo, "From source to target and back: symmetric bi-directional adaptive gan," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, 2018, pp. 8099-8108.

[18] Y. Zhang, S. Miao, T. Mansi, and R. Liao, "Task driven generative modeling for unsupervised domain adaptation: Application to x-ray image segmentation," in *Proc. Int. Conf. Med. Image Comput. Comput.-Assisted Intervention*, 2018, pp. 599-607.

[19] K. Bousmalis, N. Silberman, D. Dohan, D. Erhan, and D. Krishnan, "Unsupervised pixel-level domain adaptation with generative adversarial networks," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, 2017, pp. 3722-3731.

[20] H. Zhao, H. Li, S. Maurer-Stroh, Y. Guo, Q. Deng, and L. Cheng, "Supervised segmentation of un-annotated retinal fundus images by synthesis," *IEEE Trans. Med. Imaging,* vol. 38, pp. 46-56, 2018.

[21] H. Hao, Y. Zhao, Q. Yan, R. Higashita, J. Zhang, Y. Zhao, *et al.*, "Angle-closure assessment in anterior segment OCT images via deep learning," *Med. Image Anal.,* vol. 69, p. 101956, 2021.

[22] Q. Dou, C. Ouyang, C. Chen, H. Chen, and P.-A. Heng, "Unsupervised cross-modality domain adaptation of convnets for biomedical image segmentations with adversarial loss," *arXiv preprint arXiv:1804.10916,* 2018.

[23] Y. Ganin, E. Ustinova, H. Ajakan, P. Germain, H. Larochelle, F. Laviolette, *et al.*, "Domain-adversarial training of neural networks," *The J. of Mach. Learn. Res.,* vol. 17, pp. 2096-2030, 2016.

[24] Y.-H. Tsai, W.-C. Hung, S. Schulter, K. Sohn, M.-H. Yang, and M. Chandraker, "Learning to adapt structured output space for semantic segmentation," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, 2018, pp. 7472-7481.

[25] S. Sankaranarayanan, Y. Balaji, A. Jain, S. N. Lim, and R. Chellappa, "Learning from synthetic data: Addressing domain shift for semantic segmentation," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, 2018, pp. 3752-3761.

[26] S. Wang, L. Yu, K. Li, X. Yang, C.-W. Fu, and P.-A. Heng, "Boundary and Entropy-driven Adversarial Learning for Fundus Image Segmentation," in *Proc. Int. Conf. Med. Image Comput. Comput.-Assisted Intervention*, 2019, pp. 102-110.

[27] C. Chen and G. Wang, "IOSUDA: an unsupervised domain adaptation with input and output space alignment for joint optic disc and cup segmentation," *Appl Intell*, pp. 1-19, 2020.

[28] I. J. Goodfellow, J. Pouget-Abadie, M. Mirza, B. Xu, D. Warde-Farley, S. Ozair, *et al.*, "Generative adversarial networks," *arXiv preprint arXiv:1406.2661,* 2014.

[29] C. Chen, Q. Dou, H. Chen, J. Qin, and P. A. Heng, "Unsupervised bidirectional cross-modality adaptation via deeply synergistic image and feature alignment for medical image segmentation," *IEEE Trans. Med. Imaging,* vol. 39, pp. 2494-2505, 2020.

[30] X. Chen, Y. Lian, L. Jiao, H. Wang, Y. Gao, and S. Lingling, "Supervised Edge Attention Network for Accurate Image Instance Segmentation," in *Proc. Eur. Conf. Comput. Vis.*, 2020, pp. 617-631.

[31] T. Zhou, H. Fu, G. Chen, J. Shen, and L. Shao, "Hi-net: hybrid-fusion network for multi-modal MR image synthesis," *IEEE Trans. Med. Imaging,* vol. 39, pp. 2772-2781, 2020.

[32] T. Zhang, H. Fu, Y. Zhao, J. Cheng, M. Guo, Z. Gu, *et al.*, "SkrGAN: Sketching-rendering unconditional generative adversarial networks for medical image synthesis," in *Proc. Int. Conf. Med. Image Comput. Comput.-Assisted Intervention,* 2019, pp. 777-785.

[33] C. Chen, Q. Dou, H. Chen, and P.-A. Heng, "Semantic-aware generative adversarial nets for unsupervised domain adaptation in chest x-ray segmentation," in *Int. Workshop on Mach. Learn. in Med. Imaging*, 2018, pp. 143-151.

[34] Y. Huo, Z. Xu, H. Moon, S. Bao, A. Assad, T. K. Moyo, *et al.*, "Synseg-net: Synthetic segmentation without target modality ground truth," *IEEE Trans. Med. Imaging,* vol. 38, pp. 1016-1025, 2018.

[35] Y. Chen, W. Li, X. Chen, and L. V. Gool, "Learning semantic segmentation from synthetic data: A geometrically guided input-output adaptation approach," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, 2019, pp. 1841-1850.

[36] Y. Luo, L. Zheng, T. Guan, J. Yu, and Y. Yang, "Taking a closer look at domain shift: Category-level adversaries for semantics consistent domain adaptation," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, 2019, pp. 2507-2516.

[37] W. Zhang, W. Ouyang, W. Li, and D. Xu, "Collaborative and adversarial network for unsupervised domain adaptation," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, 2018, pp. 3801-3809.

[38] M. Javanmardi and T. Tasdizen, "Domain adaptation for biomedical image segmentation using adversarial training," in *IEEE Int. Symp. Biomed. Imaging*, 2018, pp. 554-558.

[39] P. Liu, B. Kong, Z. Li, S. Zhang, and R. Fang, "CFEA: collaborative feature ensembling adaptation for domain adaptation in unsupervised optic disc and cup segmentation," in *Proc. Int. Conf. Med. Image Comput. Comput.-Assisted Intervention*, 2019, pp. 521-529.

[40] J. I. Orlando, H. Fu, J. B. Breda, K. van Keer, D. R. Bathula, A. Diaz-Pinto, *et al.*, "Refuge challenge: A unified framework for evaluating automated methods for glaucoma assessment from fundus photographs," *Med. Image Anal.,* vol. 59, p. 101570, 2020.

[41] J. Sivaswamy, S. Krishnadas, G. D. Joshi, M. Jain, and A. U. S. Tabish, "Drishti-gs: Retinal image dataset for optic nerve head (ONH) segmentation," *in IEEE Int. Symp. Biomed. Imaging,* 2014, pp. 53-56.

[42] F. Fumero, S. Alayón, J. L. Sanchez, J. Sigut, and M. Gonzalez-Hernandez, "RIM-ONE: An open retinal image database for optic nerve evaluation," in *Proc. IEEE Symp. Comput.-Based Med. Syst.*, 2011, pp. 1-6.

[43] Z. Zhang, F. S. Yin, J. Liu, W. K. Wong, N. M. Tan, B. H. Lee, *et al.*, "Origa-light: An online retinal fundus image database for glaucoma analysis and research," in *Annu. Int. Conf. IEEE Eng. Med. Biol. Soc.*, 2010, pp. 3065-3068.

[44] P. Isola, J.-Y. Zhu, T. Zhou, and A. A. Efros, "Image-to-image translation with conditional adversarial networks," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, 2017, pp. 1125-1134.

[45] M. Sandler, A. Howard, M. Zhu, A. Zhmoginov, and L.-C. Chen, "Mobilenetv2: Inverted residuals and linear bottlenecks," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, 2018, pp. 4510-4520.

[46] J. Hoffman, D. Wang, F. Yu, and T. Darrell, "Fcns in the wild: Pixel-level adversarial and constraint-based adaptation," *arXiv preprint arXiv:1612.02649*, 2016.

[47] S. Wang, L. Yu, K. Li, X. Yang, C.-W. Fu, and P.-A. Heng, "DoFE: Domain-oriented Feature Embedding for Generalizable Fundus Image Segmentation on Unseen Datasets," *IEEE Trans. Med. Imaging,* vol. 39, pp. 4237-4248, 2020.

[48] L. Ju, X. Wang, X. Zhao, P. Bonnington, T. Drummond, and Z. Ge, "Leveraging Regular Fundus Images for Training UWF Fundus Diagnosis Models via Adversarial Learning and Pseudo-Labeling," *IEEE Trans. Med. Imaging,* DOI: 10.1109/TMI.2021.3056395, 2021.