Unsupervised Domain Adaptation Segmentation Network based on Cross-Domain Consistency Self-supervised Learning for Fundus Images

Xiji Huang, Hui Jiang, Mingzhi Chen, Minsheng Tan, Shiyu Yan, Lingna Chen

***Abstract*— Segmentation of the optic cup and optic disk in fundus images is an important medical image-processing task that plays a key role in evaluating and diagnosing glaucoma. However, traditional supervised learning methods have poor generalization performance over different data domains due to the differences in data distribution of fundus images collected from different medical institutions. In this work, we proposed a new unsupervised domain adaptation segmentation network based on adversarial and self-supervised learning for fundus images. To address the domain shift problem, unsupervised domain adaptation techniques receive extensive attention in fundus image segmentation. We add consistency and category regularization to the unsupervised domain adaptation framework. We train a consistency-adversarial loss network with Fourier transform to minimize the difference in feature distribution between the source and target domains. We use a self-supervised learning strategy to filter out the pseudo-labels located inside the categories by introducing category boundary constraints to improve the credibility of the pseudo-labels. The experiment results show our method is better than other methods, with an improvement of 1.19% Dice and 0.94 ASD on the Drishti-GS dataset and 3.49% Dice and 3.17 ASD on the RIM-ONE-r3 dataset. Experimental results on three publicly available retinal fundus image datasets show that the proposed unsupervised domain adaptation method for fundus image segmentation method achieves excellent performance compared to other state-of-the-art methods.**

***Index Terms*— Unsupervised domain adaptation, fundus image segmentation, cross-domain consistency, self-supervised learning**

This study was funded by the Natural Science Foundation of Hunan Province of China (grant numbers 2020JJ4514, 2022JJ50153 and 2020JJ4519), Natural Science Foundation of Xinjiang Uygur Autonomous Region of China (grant number 2022D01C461), Clinical Research 4310 Program of the First Affiliated Hospital of the University of South China (grant number 4310-2021-K06). (Xiji Huang and Hui Jiang contributed equally to this work.) (Corresponding authors: Shiyu Yan; Lingna Chen.)

Xiji Huang is with the Computer School, University of South China, Hengyang 421001, China (e-mail: lfeng5676@gmail.com).

Hui Jiang and Minsheng Tan are with the Computer School, University of South China, Hengyang 421001, China (e-mail: 3218515@qq.com; tanminsheng65@163.com).

Mingzhi Chen is with the School of Mechanical Engineering, Shanghai Jiao Tong University, Shanghai 200240, China (e-mail: 1297008453@qq.com).

Shiyu Yan and Lingna Chen are with the Computer School, University of South China, Hengyang 421001, China (e-mail: 1154528294@qq.com; linda_cjx@163.com).

I. Introduction

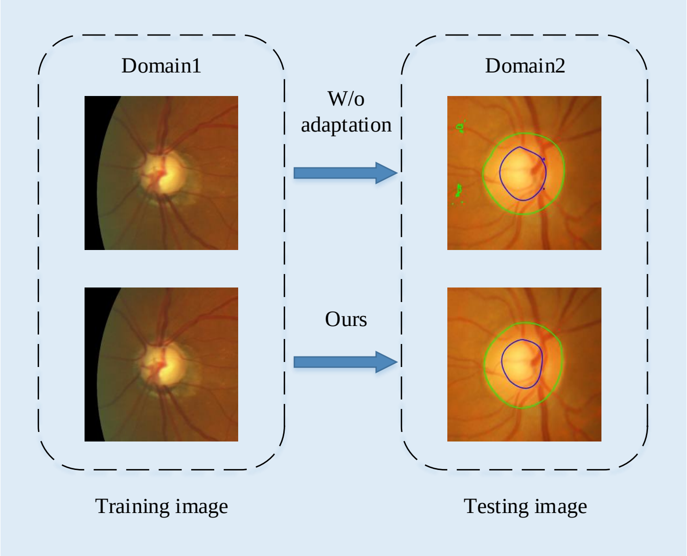

Fig. 1. Segmentation degradation due to domain shift.

GLAUCOMA is a group of eye conditions characterized by progressive and irreversible vision loss [1]. The glaucoma optic-nerve detection focused mainly on the cup-disc ratio [2]. Fundus images play a crucial role in diagnosing glaucomas by capturing the retinal structure of the eye [3], [4]. Previous studies have used deep learning models to improve the glaucoma diagnosis from fundus images [5], [6], [7]. However, most existing diagnosis models mainly use a convolutional neural network (CNN) and labeled data to detect glaucoma, which wastes rich knowledge in the feature space. Therefore, the knowledge use of unlabeled data is further studied for the glaucoma diagnosis task.

Accurate optic disc and optic cup segmentation is essential for early diagnosis of glaucoma [8]. Traditional methods already have limited performance on new datasets [9], [10], [11], [12], [13]. In recent years, deep learning methods have been widely applied in fundus image segmentation [14], [15], [16], [8], [17], [18], [19]. The deep learning models for image segmentation achieve effective local feature learning and remote contextual information interaction. However, traditional deep learning methods require that datasets have sufficient pixel-level annotations and the same feature distribution between the training and test data. Some of the existing deep learning models cause segmentation degradation due to domain shifts from different datasets, as shown in Fig.1. Due to the variations in imaging conditions, acquisition devices, and patient populations, a significant domain shift exists between different datasets, which decreases the performance of segmentation models across different domains. Therefore, it is worth studying how to understand and overcome the domain shift.

To reduce the performance degradation because of domain shift, domain adaptation methods have been developed to bridge this gap by adapting a segmentation model trained on a source domain to perform well on a target domain [20], [21], [22], [23], [24], [25], [26]. Domain adaptation methods perform well in domain shift if there are enough labels from the target domain to fine-tune the segmentation network. For example, the DANN [25] method introduces a domain classifier and an adversarial loss, minimizing the differences in the classifier to make the model feature representation insensitive to domain information. However, obtaining labeled datasets is inherently challenging, especially in medical imaging, as labeling requires professional knowledge and significant time and effort. Therefore, unsupervised domain adaptation (UDA) techniques are more competitive in leveraging the availability of labeled data from the source domain and unlabeled data from the target domain. Unsupervised domain adaptation is used for fundus image segmentation tasks to improve model accuracy and generalization ability in clinical settings, assisting healthcare professionals in diagnosing and treating ocular diseases [27], [28], [29], [30]. pOSAL [29] designs the network patch-based for fine-grained recognition of local segmentation details and improves the model performance in fundus image segmentation but remains unstable in the boundary segmentation, and the network fails on new data. The domain shift mainly manifests in distribution and label differences for unsupervised domain adaptation tasks.

Regarding distribution differences, the existing methods of unsupervised domain adaptation are image alignment and feature alignment [31]. Feature alignment is often preferred over image alignment in situations where the visual appearance or pixel-level information may not be sufficient or reliable for aligning domains. However, both methods focus only on global distribution and ignore distribution constraints at the category. Moreover, in terms of label differences, models typically need to learn knowledge from the source domain to generate pseudo labels in an unsupervised manner and transfer them to the target domain. Thus, performing feature alignment and category-level constraints in a unified framework focuses on global and local distributions to improve domain adaptation performance. Meanwhile, we recognize that reducing the label differences is important for increasing the confidence of pseudo labels.

In this work, we propose a new unsupervised domain adaptation framework by combining adversarial and self-supervised learning. In the first stage, we use a generative adversarial network with Fourier transform for cross-domain consistency loss to generate images that consider the styles of both the source and target domains. The Fourier transform is added to counteract domain shifts. We use a discriminator to connect the generated new style image with the target image, which has the advantage of better feature alignment and helps with cross-domain consistency. In the second stage, we use the pseudo-labels learned through self-supervision to label the unlabeled target domain images. A category boundary constraint module regularizes the pseudo-label to improve confidence. Our framework is unified because the encoders and weights of the segmentation network are shared. Our method fuses adversarial and self-supervised learning, and we fully validate its domain adaptation effect on the retinal fundus image dataset.

Our main contributions are summarized as follows:

1) We propose a new unsupervised domain adaptation framework, which combines adversarial learning and self-supervised learning to address the domain shift problem.
2) We use two stages to generalize the model to obtain generative adversarial networks with Fourier transforms for cross-domain consistency and self-supervised learning with the addition of category regularization.
3) To demonstrate the framework's effectiveness, we have performed extensive experiments on three publicly available datasets of retinal fundus images to segment optic cups and discs.

II. Related Work

Professional fundus cameras or scanners usually acquire fundus images and have high-resolution and complex structures [3]. Segmentation of the optic cup and disc in retinal fundus images is a key task in medical image processing in ophthalmology. Cup and disc are two important segmentation structures in fundus images for disease diagnosis and evaluation of fundus lesions [32]. Deep learning techniques, especially Convolutional Neural Networks (CNNs), have achieved remarkable results in fundus image segmentation tasks in recent years [33], [34], [35], [36]. Effective network architecture design helps to improve the performance of methods for deep learning. UT-net [35] proposes an attention-gated bilinear fusion scheme that utilizes the advantages of U-Net and transformers in its coding layer to capture the low-level features. However, CNN-based methods are often degraded when the training and testing datasets are from different domains. Domain shift is typical on different datasets of the same class. For a domain shift task, domain adaptation methods are more advantageous than traditional deep learning methods.

Recently, various approaches have been proposed to tackle the domain shift problem, which is mainly caused by inconsistent edge probability distributions of the output labels [37], [20], [38], [39], [40]. Following the previous work, DAE [37] uses an independently trained denoising auto-encoder to model the implicit prior on rational anatomical segmentation labels. FDA [20] reduces the discrepancy between the source and target distributions by exchanging the low-frequency spectrum of one with the other. BEAL [38] utilizes adversarial learning to encourage boundary predictions and uncertainty maps in the target domain to be similar to the source domain, generating more accurate boundaries and suppressing high uncertainty predictions of OD and OC segregation.

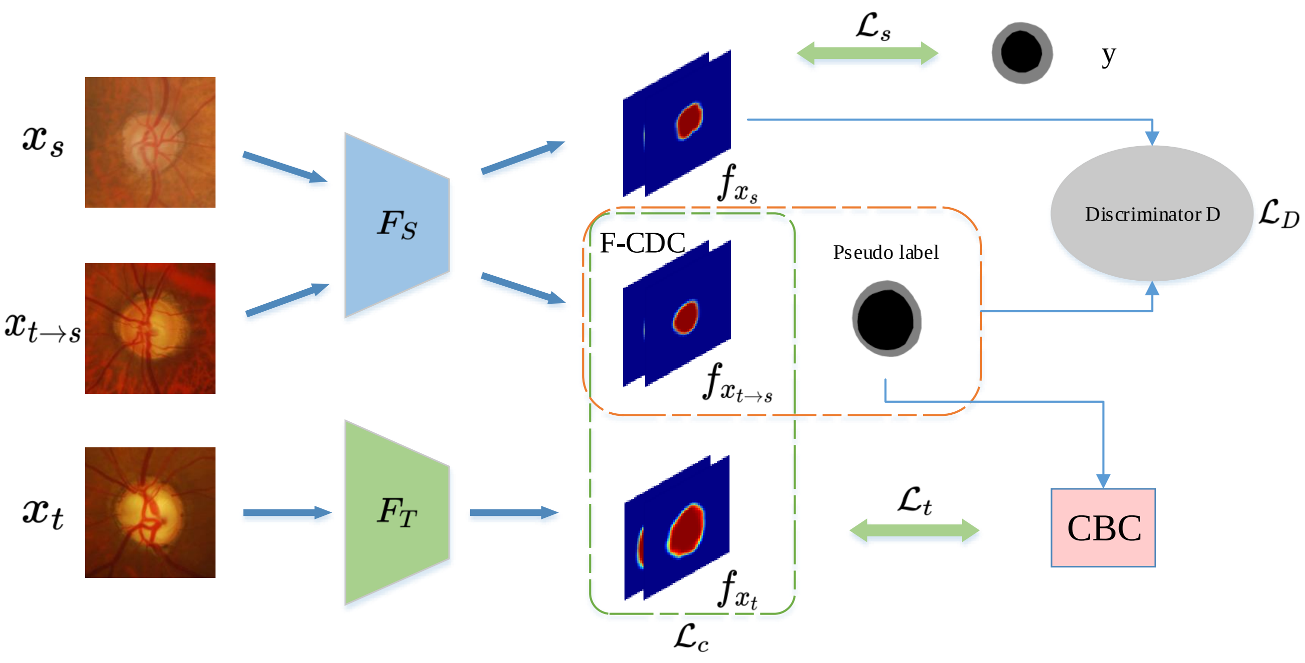

Fig. 2. Overview of the pOSAL framework.

Adversarial domain adaptation methods aim to minimize this discrepancy and enable the model to generalize well to the target domain [41], [42]. Cycle-GAN [41], proposed by Jun-Yan Zhu et al., is particularly effective when paired training data is unavailable, making it suitable for tasks such as style migration, image synthesis, and domain adaptation. This approach encourages the model to learn domain-invariant representations useful for the main task. The cyclic consistency loss during training ensures that the translated image can be converted to the original domain without losing information. Domain adaptation methods are also based on self-training and leverage self-training or pseudo-labeling to address the domain shift problem. Self-training for Domain Adaptation (STA) [43] combines self-training with domain adaptation. Therefore, our approach combines self-training and domain adversarial training to align the feature distributions across domains. Initially, a model is trained on the labeled source domain data. Then, pseudo-labels are generated for the unlabeled target domain data. In addition to the self-training process, an adversarial domain classifier is introduced to encourage domain-invariant representations. The model is iteratively refined by updating the pseudo-labels, optimizing the model parameters, and training the domain classifier.

In clinical practice, datasets are subject to domain shifts and may also be missing labeled data. Retinal fundus images are even less prone to labeling data due to their complex structure. Unsupervised domain adaptation methods are advantageous in such cases. In the domain adaptation process, to perform well on the target domain, the model's output should be as consistent as possible with the small perturbations of the inputs in the target domain. Pixmatch [40] proposes to use pixel consistency as a simple and easy-to-implement method to generalize the model. Due to the lack of labels, labels on the target domain are usually replaced by pseudo-labels generated by the source model. Domain transfer inevitably leads to pseudo-labels with noise. DPL [39] introduces two complementary pixel-level and class-level denoising schemes with uncertainty estimation and prototype estimation to reduce the noisy pseudo-labels and select reliable pseudo-labels to enhance the pseudo-labeling efficacy. This iterative process of co-training in generating pseudo-labels allows the models to learn from each other and utilize their agreement to improve their performance in the target domain. These above methods narrow the difference in distribution between the source and target domains and improve the model's performance in the target domain without requiring extensive labeled data in the target domain. However, these methods ignore the more important point that different styles of images in different domains are closely related to the segmentation results. Even for self-supervised strategies, different styles of images affect the generation of pseudo-labels. Therefore, we consider adding consistency and category regularization to the method to conduct the domain adaptation for the optic cup and disc segmentation in retinal fundus images.

III. Methodology

A. Method Overview

As shown in Fig.2, we present a method for unsupervised domain adaptation in fundus image segmentation, which consists of two main modules: a cross-domain consistency module based on the Fourier transform and a class boundary constraint (CBC) module combined with regularization. We first input the preprocessed source and target images into the segmentation network to generate corresponding predictions. Then use a discriminator to induce the network to produce similar outputs for the source and target images. Since fewer labels exist in the source domain, we use the self-supervised learning strategy to create a high-confidence pseudo label for the unlabeled target image by adding category boundary constraints.

We embark on the challenging endeavor of unsupervised domain adaptation for fundus image segmentation tasks. In this context, we are confronted with a scenario where we possess a source image set denoted as X_S, a corresponding source label set denoted as Y_S, and an unlabeled target image set represented by X_T. We aim to acquire a proficient task network, F_T, capable of seamlessly and precisely segmenting the optic disc and cup for every fundus image residing in the target domain.

The proposed network takes an image x_s from the source domain X_S and another image x_t from the target domain X_T as inputs. We first perform Fourier fusion of image x_s and image x_t to obtain a new input $x_{t\to s}$. We then pass x_s and $x_{t\to s}$ to F_S to obtain their task predictions. The function of discriminator D is not only to make the output similar but also to calculate their consistency loss, which can be used to initialize the target network. At the same time, input the generated labels from the source domain into the category boundary constraint module to assist in the segmentation task of the target network.

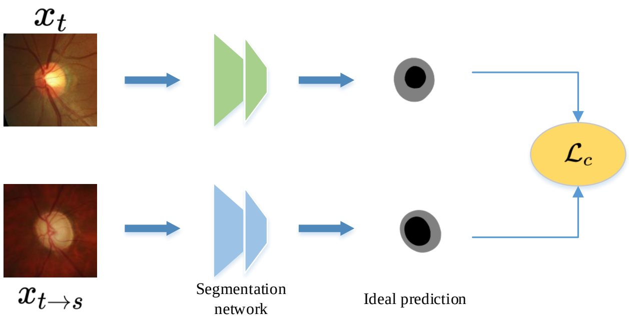

Fig. 3. Diagram of F-CDC block.

B. Fourier-based Cross-domain Consistency

Fourier-based cross-domain consistency leverages the power of consistency training and pseudo-labeling to enforce robust consistency on the target domain. The elegantly designed loss function comprises two fundamental components: two cross-entropy loss terms and a pivotal cross-domain consistency loss. Initially, the model is guided by the standard supervised cross-entropy loss, applied proficiently on the source domain:

$$\mathcal{L}_s = -\frac{1}{N_{X_S}} \sum_{x_s \in X_S} \sum_{i=1}^{h\cdot w} H\left(y_s^{(i)}, p^{(i)}\left(y \mid x_s\right)\right) \tag{1}$$

where $p^{(i)}$ is the output probability distribution at pixel i for source input x_s, y is the ground truth semantic map, and N_{X_S} is the number of images in the source dataset X_S.

The second of these losses is a cross-entropy loss on the source domain image with the style of the target domain. To compute this loss, we first generate a cross-domain-styled source domain image $x_{t\to s}$ through the model. We then use the true label of the corresponding target domain image to compare with the predicted pseudo-label $\hat{y_t}$ of $x_{t\to s}$. Our cross-domain style image loss function is as followed:

$$\mathcal{L}_t = -\frac{1}{N_{X_T}} \sum_{x_t \in X_T} \sum_{i=1}^{h\cdot w} H\left(\hat{y_t}^{(i)}, p^{(i)}\left(y \mid x_t\right)\right) \tag{2}$$

We connected the outputs of two domain-specific model segmentation networks (i.e., F_S and F_T) using a cross-domain consistency loss $\mathcal{L}_c$, as shown in Fig.3. In the semantic segmentation task, we introduce the bidirectional KL divergence loss, a novel and powerful component of our methodology. These components contribute to the robustness and efficacy of our approach in achieving accurate and reliable semantic segmentation results. We define the cross-domain consistency loss as

$$\begin{aligned} \mathcal{L}_c = &-\mathbb{E}_{x_t} \sum_{h,w,c} f_{x_{t\to s}}(h,w,c) \log\left(f_{x_t}(h,w,c)\right) \\ &-\mathbb{E}_{x_t} \sum_{h,w,c} f_{x_t}(h,w,c) \log\left(f_{x_{t\to s}}(h,w,c)\right) \end{aligned} \tag{3}$$

where f_{x_t} and $f_{x_{t\to s}}$ are the task predictions for x_s and $x_{t\to s}$, respectively.

To facilitate the comparison of the experiments, we give here the formula for Mean Squared Error Loss.

$$\mathcal{L}_{mse} = \frac{1}{N} \sum_{i=1}^{N} \left(y_i - \hat{y_i}\right)^2 \tag{4}$$

C. Class Boundary Constraints with Regularization

In an unsupervised domain adaptation (UDA) task setting, no explicit labeling information is usually used to guide the learning process. Therefore, we introduce category boundary constraints in Fig.4 to guide the algorithm in learning distinguishable and discriminative feature representations. We apply regularization to control model complexity and prevent overfitting while imposing specific constraints on decision boundaries between classes. Specifically, We introduce a regularization term that penalizes deviations from a straight line, such as the sum of squared distances of the boundary points to the line. By including this term in the loss function, the model will be encouraged to find a decision boundary that is as close to a straight line as possible while minimizing the overall loss.L2 regularization (Ridge) can be employed in conjunction with the class boundary constraints to strike a balance between fitting the training data and adhering to the desired boundary properties.

D. Cross-Domain Self-supervised Learning

We designed this training process as shown in Fig.2. This adversarial and self-supervised learning process ensures that we constrain the segmentation mask.

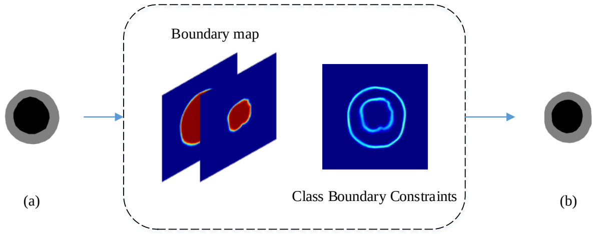

Fig. 4. Illustration of CBC block.

1) Discriminator: The discriminator uses a GAN network for adversarial learning. To complement the Fourier-based cross-domain consistency, we impose constraints on the source domain image prediction and the transformed image prediction, intending to direct the segmentation network to focus on the similarity of the image feature distributions.

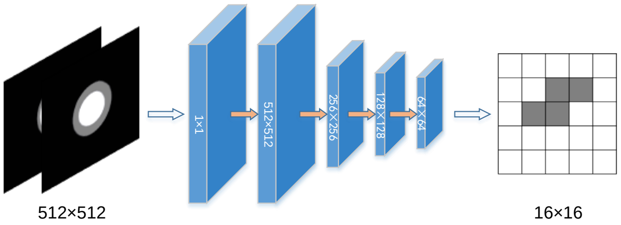

Fig. 5. Network architecture of the discriminator D.

A fully convolutional network is used to implement the discriminator, as shown in Fig.5. The network contains five convolutional layers with a kernel size of 4 × 4 and a stride of 2 × 2. The channel numbers of the five convolutional layers are 64, 128, 256, 512, and 1. The activation function following each convolutional layer is LeakyReLU, with an alpha value of 0.2. We use this self-supervised learning strategy to enforce that the feature maps generated in the target domain prediction are similar to those in the source domain.

2) Objective function: Self-supervised learning leverages the feature structure learned from the source domain data to generate pseudo-labels or supervisory signals for training. The weights of the source model, the target model, and the discriminator D get updated in this framework. The discriminator assesses the loss of consistency between the target and source domains. We formulate the training objective for the discriminator as

$$\begin{aligned}\mathcal{L}_D = & -\sum_{i=1}^{h\cdot w} z\log\left(D\left(f_{x_s}\left(h,w,c\right)\right)\right) \\ & -\sum_{i=1}^{h\cdot w}\left(1-z\right)\log\left(1-D\left(f_{x_{t\rightarrow s}}\left(h,w,c\right)\right)\right)\end{aligned} \tag{5}$$

where z = 1 if the patch prediction is from the source domain, and z = 0 if the patch prediction is from the target domain.

The overall training objective $\mathcal{L}$ for training the proposed network consists of three loss terms. First, The loss on the source domain image helps bridge the gap between the source and target domains by encouraging the model to learn domain-invariant features. Second, losing the target domain image encourages the model to learn discriminative representations that generalize well to the target domain. Third, consistency loss helps segmentation models better align the feature representations of the source and target domains.

$$\mathcal{L} = \mathcal{L}_s + \mathcal{L}_{t\rightarrow s} + \lambda c \cdot \mathcal{L}_c \tag{6}$$

wa here λc is the hyperparameter used to control the relative importance of the loss terms.

3) Training strategy: The ground truth masks of the target domain are not available in the UDA setup. As compensation, in self-supervised learning, the model learns to predict certain aspects of the data itself, creating its pseudo-labels from the available data without relying on external annotations. The first step is to preprocess the data to create meaningful pretext tasks. Pretext tasks are designed to create surrogate supervisory signals that allow the model to learn useful data representations. This preprocessing process and pretext tasks are described in Section 3.1. We create high-confidence pseudo-labels by using the training model's prediction probability $p^{(i)}$ on the i-th pixel of the target image x_t. The pseudo-label can be defined as $\hat{y}^{(i)}$= arg max($p^{(i)}$(y | x_t))[$p^{(i)} \geq \gamma$], where $p^{(i)}$ is the output probability distribution at pixel i for target input x_t and $\gamma \in$ (0, 1) is the probability threshold to determine the binary mask. During training, the model processes the data and generates pseudo-labels based on the pretext task. After the initial pretraining using self-supervised learning, we proceeded to fine-tune the model on labeled data from the target task. This crucial step enables the model to adapt its learned feature representation, acquired through self-supervised learning, to the specific requirements and intricacies of the target task. Continuously iterating through the abovementioned process, we diligently apply the same procedure for each training iteration.

E. Network Configurations and Implementation Details

The framework is implemented with Pytorch 1.7.1 using one NVIDIA 1080Ti GPU. We first train our segmentation network with images and labels on the source domain and then train the entire framework with self-supervised learning. We employ a MobileNetV2 adapted from DeepLabv3+ to backbone the network and then use the MobileNetV2 weights already trained on the ImageNet dataset to initialize the backbone network weights. Data augmentation was adopted to expand the training dataset by random scale, rotation, flip, elastic transformation, contrast adjustment, adding noise, and random erasing. The segmentation network was trained using Adam optimizer with a momentum of 0.99 and a learning rate of 1e-3. We trained a total of 200 epochs with a batch size of 8. When training the whole framework, we initialized the target model with weights obtained from training in the source domain, which can continuously benefit from the discriminator D. The stochastic gradient descent (SGD) algorithm was adapted to optimize the discriminator with the initial learning rate of 2.5e-5 and the weight decay of 0.0005.

IV. Experiments

A. Dataset

We performed extensive experiments on three publicly available retinal fundus image datasets, the Drishti-GS dataset, the RIM-ONE-r3 dataset, and the REFUGE challenge dataset. The statistics of these three datasets are listed in TableI. Due to different acquisition devices, the REFUGE challenge dataset is divided into a REFUGE Train dataset and a REFUGE Validation/Test dataset, which have 400 images each. We divided these 400 images into 320 training sets and 80 test sets in the specific use process, respectively. During the training process of the unsupervised domain adaptation framework, we used the REFUGE Train dataset as the source domain, the Drishti-GS dataset, and the RIM-ONE-r3 dataset as the target domain. As shown in Fig.6, the source and target domain data are acquired by different devices and have different texture structures, so they can be used to verify the model's generalization. Among them, the labels (or a small number of labels) can be obtained from the source domain data, but not the labels of the target domain data can be accessed. With this setup, we compared our approach with other state-of-the-art UDA segmentation methods on the target domain dataset.

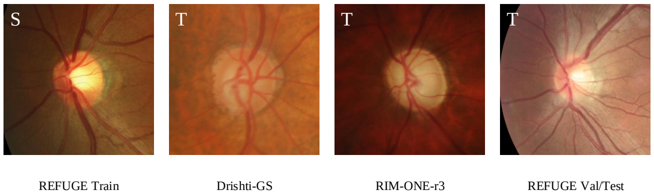

Fig. 6. Comparison of images from different datasets.

B. Evaluation Metrics

We adopt two commonly-used evaluation metrics, including the Dice coefficient (Dice) for pixel-wise accuracy measure and the Average Surface Distance (ASD) for boundary agreement assessment to assess our approach. The criteria are defined as:

$$Dice = \frac{2 \times TP}{2 \times TP + FP + FN}, \tag{7}$$

$$ASD\left(X,Y\right) = \frac{\sum_{x \in X} min_{y \in Y} d\left(x,y\right)}{\left|X\right|} \tag{8}$$

where TP, FP, and FN represent the number of true positive, false positive, and false negative pixels, respectively. d(x,y) is a 3-D matrix consisting of the Euclidean distances between the two image volumes X and Y. The higher Dice and lower ASD indicate good performance.

C. Comparison with State-of-the-art Methods

We compare DAE generates and optimizes labels for the target domain with a denoising autoencoder trained from the source domain. FDA reduces discrepancy by learning low-frequency spectral features between two domain distributions with Fourier transforms. PixMatch introduces a new module to explore the pixel-wise consistency of the source and target domain outputs. BEAL proposes a more accurate boundary prediction over the target domain, one of the best UDA methods for cross-domain fundus images. DPL is a source-free unsupervised domain adaptation method that employs pseudo-label denoising, which improves the confidence of pseudo labels by reducing the noisy pseudo labels. We followed the same experimental setup for each method for a fair comparison. The target domain image used in the training phase differs from the target domain image used in the testing phase.

The quantitative comparison results of the different methods are shown in TableII, and the visual comparison results of the two target domain datasets are shown in Fig.7. It can be seen that our method outperforms the other comparative methods with an average improvement of 1.19% Dice and 0.94 ASD on the Drishti-GS dataset and 3.49% Dice and 3.17 ASD on the RIM-ONE-r3 dataset. DPL proposes pixel-level and class-level denoising schemes, achieving better segmentation performance than the other methods, which shows that handling pseudo-labeling is critical in unsupervised domain adaptation. Bad pseudo-labeling will directly affect the segmentation effect of the model. The better performance achieved by our method compared to DPL proves the effectiveness of adding category boundary constraints with regularization in our framework. It is worth noting that there is a significant difference in the segmentation results of the FDA on two different datasets due to the domain bias affecting our perturbation function, which indirectly proves the effectiveness of our proposed cross-domain consistency with the Fourier transform.

The proposed framework contains a MobileNetV2 backbone and two modules. Using the REFUGE dataset as the source domain and the Drishti-GS dataset as the target domain, we analyze the model complexity of the different approaches in terms of runtime memory, floating-point operations, number of parameters, and inference time. FLOPs and parameters are related to the method's time and space complexity, respectively. TableIII offers a comprehensive comparative overview of our method, BEAL, and DPL. Remarkably, while our method marginally boasts a higher parameter count than its counterparts, it is noteworthy that its GFLOPs and inference time exhibit remarkable proximity to those of BEAL and DPL. Our method is slightly more complex than BEAL and DPL, but its GFLOPs are still acceptable, and its inference speed is faster than other methods.

Additionally, we further validate the effectiveness of FFT-based adversarial learning with class boundary constraints on the REFUGE validation dataset. Following the division in Table 1, we randomly divide the 400 validation images into two parts and use them as unlabeled target domain training data to train the network and target domain test data to evaluate the network, respectively. We report in TableIV the performance of our framework and the same network without domain adaptation. We can see that the proposed framework also improves the Dice of the optic cups and disks on the REFUGE validation dataset.

TABLE I
STATISTICS OF THE FUNDUS DATASETS USED IN OUR METHOD.

Domain	Dataset	Cameras	Samples(train+test)
Source	REFUGE Train	Zeiss Visucam 500	400(320+80)
Target	Drishti-GS	Unknown(Aravind eye hospital)	50+51
Target	RIM-ONE-r3	Canon EOS 5D	99+60
Target	REFUGE Validation/Test	Canon CR-2	400(320+80)

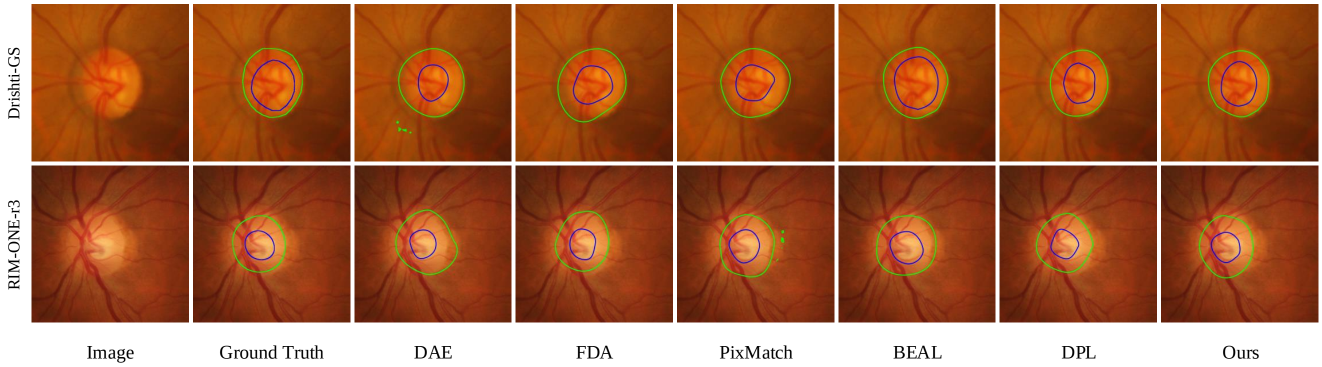

Fig. 7. Qualitative results on the Drishti-GS testing dataset.

TABLE II
COMPARISON WITH OTHER METHODS ON TARGET DOMAIN DATASETS.

Methods	Dice Metric			ASD Metric		
	Cup	Disc	Avg	Cup	Disc	Avg
Drishti-GS						
DAE	73.3	93.63	83.47	15.22	9.5	12.36
FDA	69.38	78.07	73.72	21.15	28.86	25.01
PixMatch	64.91	75.88	70.39	18.6	30.5	24.55
BEAL	76.39	95.69	86.06	16.15	5.14	10.64
DPL	77.95	95.79	86.87	15.05	4.79	9.92
Ours	79.74	96.38	88.06	13.93	4.03	8.98
RIM-ONE-r3						
DAE	73.58	88.3	80.94	13.93	11.99	12.96
FDA	75.35	92.13	83.74	12.4	7.68	10.04
PixMatch	76.45	91.97	80.21	17.06	10.92	13.99
BEAL	75.57	89.02	82.29	11.02	11.74	11.38
DPL	72.97	90.1	81.53	12.63	9.79	11.21
Ours	75.8	94.24	85.02	10.95	5.13	8.04

TABLE III
THE COMPARISON OF MODEL COMPLEXITY ON DRISHTI-GS DATASET.

Models	Memory(GB)	FLOPs(G)	Params(M)	Inference Time(s)
BEAL	6.28	6.64	5.81	0.021
DPL	6.99	6.49	5.69	0.021
Ours	7.26	6.49	5.69	0.022

TABLE IV
RESULTS OF SEGMENTATION ON THE REFUGE VALIDATION DATASET.

Method	Dice$_{cup}$	Dice$_{disc}$	ASD$_{cup}$	ASD$_{disc}$
Ours W/o adaptation	76.36	92.84	15.05	9.76
Ours	79.86	96.27	14.05	4.59

D. Ablation Analysis of Key Components

To verify the contributions of key modules of our method, we further conducted a series of experiments on the test dataset. The quantitative results are presented in TableV. To observe the effect of domain shifting on segmentation performance, we first test the target image by directly applying the model learned from the source domain without using any domain adaptation method to obtain a lower bound named "w/o adaptation." We use the model without any module, i.e., the backbone network, as the baseline, and we first test the generative adversarial network with Fourier transforms for cross-domain consistency loss. We then ablate the category boundary constraint module. Both Dice and ASD metrics show that the proposed method significantly improves the adaptation capability of the segmentation network in this experiment setting.

To investigate the effects of hyper-parameter λc in an Equation6 that controls the contributions of different losses, we evaluated the model's performance with different λc on the Test dataset. From TableVI, we observed that when λc=0.25, the Dice is higher, and the ASD is lower. Therefore, we set λc=0.25 in the experiments.

We compared the effect of different loss functions on the segmentation network. The results are shown in Fig.8. We can find that the Dice Loss achieved a better Dice for OC and a lower Dice for OD than the BCE Loss. Combining the BCE and Dice loss, the segmentation network achieved an unsatisfactory Dice of OD and OC predictions. And the segmentation network achieves high-quality predictions when combined with the Mean Squared Error Loss.

TABLE V
THE ABLATION STUDY ON DIFFERENT COMPONENTS.

Target Domain	Baseline	F-CDC	CBC	Dice$_{cup}$	Dice$_{disc}$	ASD$_{cup}$	ASD$_{disc}$
W/o adaptation Oracle				76.36	92.84	15.05	9.76
Drishti-GS	✓			77.73	95.32	15.11	9.68
	✓	✓		78.95	95.37	14.55	7.04
	✓		✓	79.49	95.86	13.94	5.25
	✓	✓	✓	79.74	96.38	13.93	4.03

TABLE VI
HYPER-PARAMETER ANALYSIS FOR λc.

λc	Dice$_{cup}$	Dice$_{disc}$	ASD$_{cup}$	ASD$_{disc}$
0.1	76.09	94.52	15.24	10.07
0.2	78.01	95.19	14.87	7.32
0.25	79.74	96.38	13.93	4.03
0.3	78.81	95.27	14.81	6.85
0.4	76.33	94.83	15.03	9.5

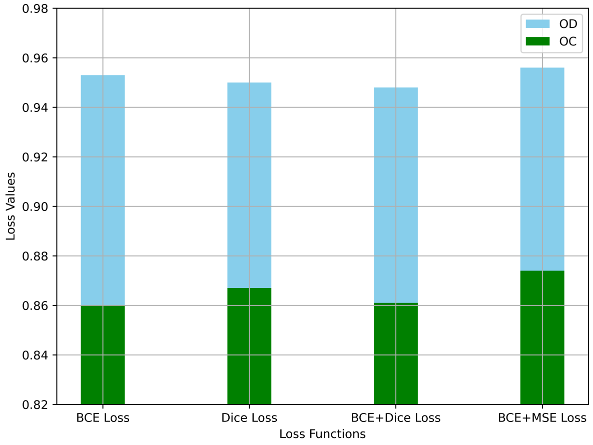

Fig. 8. Comparison of different loss functions.

V. Discussion

Optic cup and disk segmentation is an important part of the fundus image segmentation task. It is vital to detect diseases early, evaluate treatments' effectiveness, and predict disease progression to accurately segment the optic cup and optic disk in ocular structures. Deep neural networks remain susceptible to domain shifts between training (source) and test (target) data obtained under different conditions. Preferred modalities and scanning protocols used in different hospitals can vary significantly. Different CT equipment and protocols can lead to different scan spacing, slice thicknesses, and differences in the intensity and texture of organs. Due to the domain shift, models usually cannot perform better when validated on new datasets. Unsupervised domain adaptation methods improve the model's generalization ability over the target domain by learning the difference in data distribution between the source and target domains and minimizing the inter-domain bias. Recently, some work has explored the feasibility of unsupervised domain adaptation for medical image segmentation with promising results. Therefore, we proposed the unsupervised domain adaptation for optic cup and optic disk segmentation. Our goal is consistency of images on both source and target domains. Experiments on three publicly available datasets show the proposed method's effectiveness in improving the model's generalization ability.

Domain shift causes models trained on the target domain to perform poorly in real applications. The target domain tends to have relatively little labeled data in domain adaptation. Different from [39], Our approach incorporates consistency and category regularization in the unsupervised domain adaptation framework to solve the problem of domain shift. By adding consistency constraints on the target domain, the network makes the unlabeled data of the target domain better utilized in training and brings the distance between the source and target domains closer in the feature space, which reduces the distributional differences between the domains and improves the generalization ability of the model. In addition, consistency can help mitigate label shifts, allowing the model to make more accurate predictions on the target domain. We improve pseudo-labels' credibility by adding category boundary constraints when generating pseudo-labels. The pseudo-labels increase the diversity of data in the target domain, thus allowing the model to learn the features of the target domain more comprehensively.

The validity of our method was demonstrated, but the proposed method has some limitations. In clinical practice, the source domains lack labels on some medical datasets. Therefore, how to introduce diffusion models [44] to generate pseudo-labels for unsupervised domain adaptation on unlabeled source domain datasets is worth developing. This paper's dataset is unimodal, and the model segmentation performance performs poorly when a multimodal dataset appears. In the future, we will continue to explore the application of the Fusion of multimodal [12] and unsupervised domain adaptation.

VI. Conclusion

We propose a novel unsupervised domain adaptation segmentation network based on adversarial and self-supervised learning for fundus images to segment optic cups and discs. We first incorporate Fourier knowledge in cross-domain consistency to guide the model to converge the outputs in the source and target domains to enhance further the generalization ability of the unsupervised domain adaptation segmentation network. In particular, we also incorporate a category boundary constraint module into the model, which is essential for local information segmentation. Experiments conducted on three publicly available retinal fundus image datasets fully demonstrate the effectiveness of our proposed consistency framework.

References

[1] M. C. V. Stella Mary, E. B. Rajsingh, and G. R. Naik, "Retinal fundus image analysis for diagnosis of glaucoma: A comprehensive survey," *IEEE Access*, vol. 4, p. 4327–4354, Jan 2016.

[2] R. Zhao, X. Chen, X. Liu, Z. Chen, F. Guo, and S. Li, "Direct cup-to-disc ratio estimation for glaucoma screening via semi-supervised learning," *IEEE Journal of Biomedical and Health Informatics*, vol. 24, no. 4, pp. 1104–1113, 2020.

[3] Z. Shen, H. Fu, J. Shen, and L. Shao, "Modeling and enhancing low-quality retinal fundus images," *IEEE Transactions on Medical Imaging*, vol. 40, no. 3, pp. 996–1006, 2021.

[4] 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 Transactions on Medical Imaging*, vol. 38, no. 1, pp. 46–56, 2019.

[5] Y. Jiang, L. Duan, J. Cheng, Z. Gu, H. Xia, H. Fu, C. Li, and J. Liu, "Jointrcnn: A region-based convolutional neural network for optic disc and cup segmentation," *IEEE Transactions on Biomedical Engineering*, vol. 67, no. 2, pp. 335–343, 2020.

[6] S. M. Shankaranarayana, K. Ram, K. Mitra, and M. Sivaprakasam, "Fully convolutional networks for monocular retinal depth estimation and optic disc-cup segmentation," *IEEE Journal of Biomedical and Health Informatics*, vol. 23, no. 4, pp. 1417–1426, 2019.

[7] V. Agrawal, A. Kori, V. Alex, and G. Krishnamurthi, "Enhanced optic disk and cup segmentation with glaucoma screening from fundus images using position encoded cnns," 2018.

[8] 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 Transactions on Medical Imaging*, vol. 37, no. 7, pp. 1597–1605, 2018.

[9] J. Cheng, J. Liu, Y. Xu, F. Yin, D. W. K. Wong, N.-M. Tan, D. Tao, C.-Y. Cheng, T. Aung, and T. Y. Wong, "Superpixel classification based optic disc and optic cup segmentation for glaucoma screening," *IEEE Transactions on Medical Imaging*, vol. 32, no. 6, pp. 1019–1032, 2013.

[10] S. Sedai, P. K. Roy, D. Mahapatra, and R. Garnavi, "Segmentation of optic disc and optic cup in retinal fundus images using shape regression," in *2016 38th Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC)*, pp. 3260–3264, 2016.

[11] G. D. Joshi, J. Sivaswamy, and S. R. Krishnadas, "Optic disk and cup segmentation from monocular color retinal images for glaucoma assessment," *IEEE Transactions on Medical Imaging*, vol. 30, no. 6, pp. 1192–1205, 2011.

[12] M. S. Miri, M. D. Abràmoff, K. Lee, M. Niemeijer, J.-K. Wang, Y. H. Kwon, and M. K. Garvin, "Multimodal segmentation of optic disc and cup from sd-oct and color fundus photographs using a machine-learning graph-based approach," *IEEE Transactions on Medical Imaging*, vol. 34, no. 9, pp. 1854–1866, 2015.

[13] A. M. Jose and A. A. Balakrishnan, "A novel method for glaucoma detection using optic disc and cup segmentation in digital retinal fundus images," in *2015 International Conference on Circuits, Power and Computing Technologies [ICCPCT-2015]*, pp. 1–5, 2015.

[14] J. Cheng, J. Liu, Y. Xu, F. Yin, D. W. K. Wong, N.-M. Tan, D. Tao, C.-Y. Cheng, T. Aung, and T. Y. Wong, "Superpixel classification based optic disc and optic cup segmentation for glaucoma screening," *IEEE Transactions on Medical Imaging*, vol. 32, no. 6, pp. 1019–1032, 2013.

[15] Q. Wu and A. Cheddad, "Segmentation-based deep learning fundus image analysis," in *2019 Ninth International Conference on Image Processing Theory, Tools and Applications (IPTA)*, pp. 1–5, 2019.

[16] V. G. Edupuganti, A. Chawla, and A. Kale, "Automatic optic disk and cup segmentation of fundus images using deep learning," in *2018 25th IEEE International Conference on Image Processing (ICIP)*, pp. 2227–2231, 2018.

[17] E. Moris, N. Dazeo, M. P. A. de Rueda, F. Filizzola, N. Iannuzzo, D. Nejamkin, K. Wignall, M. Leguía, I. Larrabide, and J. I. Orlando, "Assessing coarse-to-fine deep learning models for optic disc and cup segmentation in fundus images," 2023.

[18] Y. Jiang, N. Tan, and T. Peng, "Optic disc and cup segmentation based on deep convolutional generative adversarial networks," *IEEE Access*, vol. 7, pp. 64483–64493, 2019.

[19] 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 Transactions on Medical Imaging*, vol. 37, pp. 1597–1605, jul 2018.

[20] Y. Yang and S. Soatto, "Fda: Fourier domain adaptation for semantic segmentation," 2020.

[21] 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 Transactions on Medical Imaging*, vol. 39, no. 12, pp. 4237–4248, 2020.

[22] J. Lyu, Y. Zhang, Y. Huang, L. Lin, P. Cheng, and X. Tang, "Aadg: Automatic augmentation for domain generalization on retinal image segmentation," *IEEE Transactions on Medical Imaging*, vol. 41, no. 12, pp. 3699–3711, 2022.

[23] J. Chen, Z. Zhang, X. Xie, Y. Li, T. Xu, K. Ma, and Y. Zheng, "Beyond mutual information: Generative adversarial network for domain adaptation using information bottleneck constraint," *IEEE Transactions on Medical Imaging*, vol. 41, no. 3, pp. 595–607, 2022.

[24] Y.-C. Chen, Y.-Y. Lin, M.-H. Yang, and J.-B. Huang, "Crdoco: Pixel-level domain transfer with cross-domain consistency," 2020.

[25] D. Quang, Y. Chen, and X. Xie, "Dann: a deep learning approach for annotating the pathogenicity of genetic variants.," *Bioinformatics*, p. 761–763, Mar 2015.

[26] H. Yang, C. Chen, M. Jiang, Q. Liu, J. Cao, P. A. Heng, and Q. Dou, "Dltta: Dynamic learning rate for test-time adaptation on cross-domain medical images," 2022.

[27] H. Lei, W. Liu, H. Xie, B. Zhao, G. Yue, and B. Lei, "Unsupervised domain adaptation based image synthesis and feature alignment for joint optic disc and cup segmentation," *IEEE Journal of Biomedical and Health Informatics*, vol. 26, no. 1, pp. 90–102, 2022.

[28] Z. Chen, Y. Pan, and Y. Xia, "Reconstruction-driven dynamic refinement based unsupervised domain adaptation for joint optic disc and cup segmentation," *IEEE Journal of Biomedical and Health Informatics*, vol. 27, no. 7, pp. 3537–3548, 2023.

[29] 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 Transactions on Medical Imaging*, vol. 38, no. 11, pp. 2485–2495, 2019.

[30] W. Feng, L. Wang, L. Ju, X. Zhao, X. Wang, X. Shi, and Z. Ge, "Unsupervised domain adaptive fundus image segmentation with category-level regularization," 2022.

[31] Y. Meng, H. Zhang, Y. Zhao, D. Gao, B. Hamill, G. Patri, T. Peto, S. Madhusudhan, and Y. Zheng, "Dual consistency enabled weakly and semi-supervised optic disc and cup segmentation with dual adaptive graph convolutional networks," *IEEE Transactions on Medical Imaging*, vol. 42, no. 2, pp. 416–429, 2023.

[32] J. Cheng, F. Yin, D. W. K. Wong, D. Tao, and J. Liu, "Sparse dissimilarity-constrained coding for glaucoma screening," *IEEE Transactions on Biomedical Engineering*, vol. 62, no. 5, pp. 1395–1403, 2015.

[33] Y. Wang, G. Yang, D. Ding, and J. Zao, "Segmentation-based information extraction and amalgamation in fundus images for glaucoma detection," 2022.

[34] H. Fu, J. Cheng, Y. Xu, C. Zhang, D. W. K. Wong, J. Liu, and X. Cao, "Disc-aware ensemble network for glaucoma screening from fundus image," *IEEE Transactions on Medical Imaging*, vol. 37, pp. 2493–2501, nov 2018.

[35] R. Hussain and H. Basak, "Ut-net: Combining u-net and transformer for joint optic disc and cup segmentation and glaucoma detection," 2023.

[36] Z. Deng, Y. Cai, L. Chen, Z. Gong, Q. Bao, X. Yao, D. Fang, W. Yang, S. Zhang, and L. Ma, "Rformer: Transformer-based generative adversarial network for real fundus image restoration on a new clinical benchmark," *IEEE Journal of Biomedical and Health Informatics*, vol. 26, no. 9, pp. 4645–4655, 2022.

[37] N. Karani, E. Erdil, K. Chaitanya, and E. Konukoglu, "Test-time adaptable neural networks for robust medical image segmentation," *Medical Image Analysis*, p. 101907, Feb 2021.

[38] 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," 2019.

[39] C. Chen, Q. Liu, Y. Jin, Q. Dou, and P.-A. Heng, "Source-free domain adaptive fundus image segmentation with denoised pseudo-labeling," 2021.

[40] L. Melas-Kyriazi and A. K. Manrai, "Pixmatch: Unsupervised domain adaptation via pixelwise consistency training," 2021.

[41] J.-Y. Zhu, T. Park, P. Isola, A. Efros, S. Winter, V. Gogh, C. Monet, and U.-E. Photos, "Unpaired image-to-image translation using cycle-consistent adversarial networks,"

[42] 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 Transactions on Medical Imaging*, vol. 40, no. 10, pp. 2911–2925, 2021.

[43] A. Kumar, T. Ma, and P. Liang, "Understanding self-training for gradual domain adaptation," *International Conference on Machine Learning,International Conference on Machine Learning*, Jul 2020.

[44] J. Wu, R. Fu, H. Fang, Y. Zhang, Y. Yang, H. Xiong, H. Liu, and Y. Xu, "Medsegdiff: Medical image segmentation with diffusion probabilistic model," 2023.