Local-Global Pseudo-label Correction for Source-free Domain Adaptive Medical Image Segmentation

Yanyu Ye^a, Zhengxi Zhang ^b, Chunna Tian^{b,*}, Wei wei^a

^aSchool of Computer Science, Northwestern Polytechnical University, Xi'an 710129, China
^bSchool of Electronic Engineering, Xidian University, Xi'an 710071, China

Abstract

Domain shift is a commonly encountered issue in medical imaging solutions, primarily caused by variations in imaging devices and data sources. To mitigate this problem, unsupervised domain adaptation techniques have been employed. However, concerns regarding patient privacy and potential degradation of image quality have led to an increased focus on source-free domain adaptation. In this study, we address the issue of false labels in self-training based source-free domain adaptive medical image segmentation methods. To correct erroneous pseudo-labels, we propose a novel approach called the local-global pseudo-label correction (LGDA) method for source-free domain adaptive medical image segmentation. Our method consists of two components: An offline local context-based pseudo-label correction method that utilizes local context similarity in image space. And an online global pseudo-label correction method based on class prototypes, which corrects erroneously predicted pseudo-labels by considering the relative distance between pixel-wise feature vectors and prototype vectors. We evaluate the performance of our method on three benchmark fundus image datasets for optic disc and cup segmentation. Our method achieves superior performance compared to the state-of-the-art approaches, even without using of any source data.

Keywords: Source-Free domain adaptation, Pseudo label, Medical image

^{*}Corresponding author

 $^{{\}it Email address:} \; {\it chnatian@xidian.edu.cn} \; ({\it Chunna Tian})$

Yanyu Ye and Zhenxi Zhang contribute equally to this paper and co-share the first authorship of this paper.

1. Introduction

Medical image segmentation plays a crucial role in image-guided surgery systems, computer-aided diagnosis etc. Recent advancements in deep learning promote the development of automatic medical image segmentation methods that exhibit exceptional performance. Most of these methods rely on abundant labeled data. However, the presence of variations in imaging devices and data sources often leads to domain shifts in real-world medical imaging solutions[1]. To address this issue, unsupervised domain adaptation (UDA) techniques have been employed, which leverage labeled source domain data and unlabeled target domain data. Nevertheless, in clinical settings, the disclosure of medical imaging data concerns patient privacy. Besides the transmission of data leads to the deterioration of image quality. Source-free domain adaptation is a possible solutions. Research and exploration on source-free domain adaptation (SFDA) methods in medical image segmentation are just started.

The existing SFDA work can be roughly divided into two categories: source-like domain generation [2] [3] and pseudo-label self-training [4] [5] [6] [7]. The former typically involves two steps. Firstly, the distribution information stored in a pre-trained model of the source domain is utilized to generate source-like images. Subsequently, techniques such as Generative Adversarial Networks (GANs) are employed to align the distribution of target domain images with that of the source domain. This alignment process aims to enhance the performance of segmentation in the target domain, particularly when there is a shift in distribution. Pseudo-label self-training methods utilize the source model to generate pseudo-labels for the target domain to train the segmentation models. However, pseudo-labels are often noisy which leads leading to error accumulation problems and negative impacts on target domain segmentation performance.

In this paper, we present a novel approach for improving the accuracy and reliability of pseudo-labels in source-free domain adaptation for medical image segmentation.

Our method, called local-global pseudo-label correction. To achieve local rectification, we exploit the spatial continuity of images and generate pseudo-labels based on the information ensemble of neighboring pixels, where the label of the current pixel is rectified by the neighbors with high feature similarity. However, relying solely on neighboring pixels within a local region may introduce unreliable pseudo-labels. Therefore, we introduce a global label rectification method based on prototypical feature learning. In our approach, we first divide the target domain dataset into easy and hard samples based on the average entropy at the sample level. We then calculate prototypes for the easy samples and use these prototypes to correct the pseudo-labels by matching pixelwise features with class prototypes. The prototypes, which aggregate pixel features of the same class, contain common information from different samples. This global rectification complements the local rectification process and ensures that the model learns credible knowledge. By combining local and global rectification techniques, we obtain more reliable and accurate pseudo-labels for training of medical image segmentation in target domian.

In summary, our proposed method, called local-global pseudo-label rectification (LGDA), offers a solution for source-free domain adaptation in the context of medical image segmentation. The contributions of our work are as follows:

- We correct pseudo-labels from both local and global perspectives, updating rectified pseudo-labels offline and online during the self-training process, thereby
 gradually adapting the pre-trained model from the source domain to the target
 domain data distribution.
- We propose an offline local denoising moudle to improve the reliability of pseudo-labels, which employs local context similarity in image space to correct erroneous pseudo-labels pixel by pixel.
- We propose an online global denoising moudle to improve the accuracy of pseudo-labels, which split samples into easy ones or hard ones, and corrects erroneously predicted pseudo-labels based on the relative distance between pixel-

wise feature vectors and easy samples class prototype vectors.

 We evaluate our method on three benchmark fundus image datasets for optic disc and cup segmentation. Without any source data, our method achieves better performance than the state-of-the-art methods.

The rest of this paper is organized as follows. In Section 2, we briefly introduce the related work on unsupervised domain adaptation and source-free domain adaptation. In Section 3, we introduce the proposed framework in detail. In Section 4, we report the experiment settings. In Section 5, we analyze the segmentation results. Finally, we summarize this work and give the future work in Section 6.

2. Related Work

We first review related work on unsupervised domain adaptation in subsection 2.1. We then describe some related work on unsupervised domain-adaptive image classification in subsection 2.2. We present related work on source-free domain adaptation for image segmentation in 2.3.

2.1. Unsupervised domain adaptation

Unsupervised domain adaptation (UDA) refers to the task of adapting a model trained on source domain to target domain without labeled data. Many works have focused on UDA for the classification task, which can be broadly divided into two mianstreams. The first category usually uses Max Mean Distance(MMD), Kullback-Leible Divergence(KL), or others metrics to measure the discrepancy between domains. For example, previous works [8] [7] [9] have used maximum mean difference (MMD) to measure domain similarity. The second category utilizes adversarial training [10] [11], where a discriminator is employed to distinguish the source and target domains, and a generator is trained to generate domain-invariant features that fool the discriminator, thus achieves domain-adaptive classification.

Unsupervised domain adaptive image segmentation is an important application of UDA. Existing UDA methods for image segmentation tasks can be maining categorized into three groups: entropy minimization[12][13], adversarial learning[14][15][16], and self-training[17][18]. Entropy minimization aims to encourage unambiguous cluster assignments. For instance, Vu et al.[12] propose an entropy minimization loss to penalize low-confidence predictions on the target domain and further propose an entropybased adversarial training method. However, Chen et al.[13] have found that entropy minimization suffers from a class imbalance problem and propose a maximum square loss to address this issue. Adversarial learning methods typically employ discriminators to reduce the domain discrepancy between the target and source domains. Pan et al.[14] utilize discriminators to minimize inter- and intra-domain gaps, while Yan et al.[15] propose an adversarial learning-based approach to learn domain-invariant features in the segmentation task. Additionally, Zhang et al. [16] develop a noise-adaptive generative adversarial network for eye vessel segmentation, where an image-to-image translation generator is designed to map target domain images to the source domain, and a discriminator enforces content similarity between the generated and real images. The self-training method uses the source model to select the pseudo-label with high confidence to supervise the target model training process. Zou et al.[17] proposed a confidence-regularized self-training framework. They treat pseudo-labels as continuous latent variables. In addition, they propose two confidence regularization methods: label regularization and model regularization. Zou et al.[18] utilize class-normalized confidence scores to select and generate class-balanced pseudo-labels. However, most of these methods utilize threshold-based filtering of pseudo-labels in the target domain without leveraging associated semantic information to generate more reliable pseudolabels. In contrast, our method addresses the suppression of noise in target domain pseudo-labels under source-free domain conditions by incorporating local neighboring pixel information and global class prototype knowledge.

2.2. Source-free domain adaptation for image classication

The previously mentioned UDA methods require the utilization of both source domain data and target domain data simultaneously. However, in practical applications, acquiring the source domain data is challenging due to privacy concerns and data transfer difficulties etc. As a result, the problem of source-free domain adaptation has gained significant attention from medical researchers due to its high clinical value. We review SFDA methods in two applications: classification and segmentation.

Existing source-free domain adaptive image classification works can be broadly divided into two kinds: source-like domain generation and pseudo-label self-training.

Source-like domain generation involves estimating the distribution of source data and generating distribution-similar data of the source domain using target domain data. This allows the SFDA problem to be transformed into a UDA problem. For example, Kurmi et al.[19] proposed a SDDA model. They first utilized a Generative Adversarial Network (GAN) and a pre-trained classifier to learn the underlying data distribution of the source data. During the training process, random noise is used to generate source-like target images, which are then combined with the target domain data for adaptive training. Li et al.[20] proposed the 3CGAN, which is similar to the method of Kurmi et al.[19]. However, in the training process, the generation stage and the adaptive stage iterate and collaborate with each other.

The pseudo-label self-training methods, employed in the works of Kim et al.[21], Liang et al.[22], and Yang et al.[23], use a pre-trained model in the source domain to generate pseudo-labels for the target domain. They adopt an iterative training paradigm, where the pseudo-label quality and the segmentation accuracy is gradually improved.

2.3. Source-free domain adaptive for image segmentation

Image segmentation is a pixel-by-pixel classification task. The segmentation task requires the model to have a full understanding of spatial information and global semantic information. The related work on source-free domain adaptive image segmenta-

tion can be roughly divided into four categories: source-like domain generation, consistency regularization, prior information guidance introduction and pseudo-label self-training.

Specifically, for source-like domain generation method, Yang et al.[2] propose to generate the source-like images by batch normalization constraints through a style loss and a content loss, and further apply mutual Fourier Transform to generate high-quality source-like images. Ye et al.[3] choose the images with high entropy as the virtual source images and align the distribution by adversarial learning.

Secondly, some source-free domain adaptation methods introduce regularization functions. In the literature VS et al.[6], Yang et al.[2], Ye et al.[3], consistency regularization is introduced to the target domain data distribution during the adaptive training of the target domain. In [24], dropout is performed on the decoder parameters to obtain different inputs, and then consistent regularization is applied to multiple predictions to train the network.

Thirdly, other scholars have used the anatomical prior information of the tissue to be segmented to regulate the source-free domain adaptation process. Bateson et al.[25] introduced an auxiliary network to predict the class ratio of objects inspired by anatomical knowledge. In the domain adaptation stage, KL divergence is used to measure the difference between the target class ratio and prior knowledge in the target domain segmentation result. This difference is used to train the network to adapt the source domain pre-trained model to the target domain data distribution.

Last, some methods utilize the source model to generate pseudo labels and finetune the source model. To filter out some noisy components in pseudo-labels, Chen et al.[4] enhance the quality of pseudo-labels by introducing complementary pixel-level and class-level pseudo-label denoising methods, which incorporates uncertainty estimation and prototypical feature learning to reduce noisy components in pseudo-labels. VS et al.[6] propose a two-stage method including a target-specific adaptation stage and a task-specific adaptation stage. The target-specific adaptation employs an ensemble entropy minimization of multiple pseudo-labels. The false-negative predictions in pseudo-labels are filtered out by selective voting. In the task-specific adaptation stage, strong-to-weak consistency learning is introduced by the teacher-student network.

Our method belongs to pseudo-label guided self training. Notablly, we propose an offline local label correction method and an online global pseudo-label correction method in LGDA. Offline local pseudo-label correction can provide reliable initial pseudo-labels for target domain retraining. Online dynamic global pseudo-label correction can further improve the accuracy of pseudo-label supervision information with the improvement of model performance.

3. Methodology

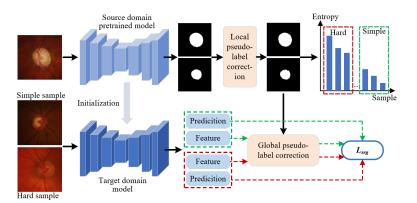


Figure 1: Source-free domain adaptation medical image segmentation method based on the local-global pseudo-label correction.

3.1. Overview

We denote the source domain training set as $D_S = \{(x_i^S, y_i^S) \in (X_S, Y_S)\}_{i=1}^{N_1}$, where x_i^S is the i-th original image in the source domain, $y_i^S \in [0,1]^{H \times W \times C}$ is the ground truth corresponding to the i-th image. H and W are the height and width of the image and ground truth, and C represents the number of categories. Likewise, the unlabeled target domain image is defined as $D_T = \{x_i^T \in (X_T)\}_{i=1}^{N_2}$. The source domain model pre-trained on the source domain dataset is written as $\phi_S(\cdot)$, and the

adaptation target domain model is defined as $\phi_T(\cdot)$. Figure 1 illustrates our SFDA method with local-global pseudo-label correction.

3.2. Segmentation Network

We employed the DeepLabv3+[26] network based on MobileNet v2 [27] as our backbone network. MobileNet v2 incorporates an innovative Inverted Residual structure that effectively transforms the input into the desired output by expanding and subsequently reducing the dimensions. The schematic diagram of the bottleneck inverted residual structure is shown in Figure 2. Table 1 illustrates the network architecture of MobileNet v2, where t represents the channel expansion ratio, i.e., the expansion factor; c denotes the number of channels; n indicates the repetition count of the current bottleneck inverted residual structure, and s represents the stride.

Table 1: MobileNet v2 network

input	Operation	t	c	n	s
$224^2 \times 3$	conv2d	-	32	1	2
$112^2\times32$	bottleneck	1	16	1	1
$112^2\times16$	bottleneck	6	24	2	2
$56^2 \times 24$	bottleneck	6	32	3	2
$28^2 \times 32$	bottleneck	6	64	4	2
$14^2 \times 64$	bottleneck	6	96	3	1
$14^2 \times 96$	bottleneck	6	162	3	2
$7^2 \times 160$	bottleneck	6	320	1	1
$7^2 \times 320$	$\operatorname{conv2d} 1 \times 1$	-	1280	1	1
$7^2\times 1280$	avgpool 7×7	-	-	1	-
$1\times1\times1280$	$\operatorname{conv2d} 1 \times 1$	-	k	-	

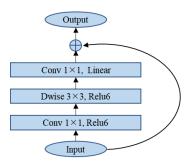


Figure 2: Bottleneck Inverted Residual block

3.3. Local Denoising Module

Due to the domain shift between different datasets, the pseudo-labels generated by the source model are bound to contain erroneous predictions. Incorporating these incorrect pseudo-labels into the model adaptation process can lead to the accumulation of errors, thereby hindering the enhancement of the model's segmentation performance in the target domain. To address this issue, we propose leveraging the local spatial continuity of images to rectify the pseudo-labels assigned to individual pixels. It is important to highlight that the module utilizes an offline mode to rectify pseudo-labels, thereby providing reliable initial pseudo-labels for target domain retraining and effectively reducing error accumulation. We elaborate on the Local denoising module in detail in the following. We elaborate on the local denoising module in the following.

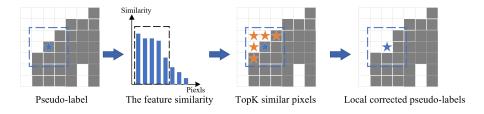


Figure 3: Local pseudo-label correction

First, given an input image from the target domain, we perform dropout sampling with Num forward passes. Then compute the mean of Num predictions as features and pseudo-labels for the target domain images, which is formulated as:

$$f_i^{S \to T}, p_i^{S \to T} = mean(\phi_S(x_i^T)_{Num=1,2,\dots,10})$$
 (1)

where $f_i^{S \to T} \in \mathbb{R}^{H_1 \times W_1 \times C_1}$ is the feature vector of the i-th image in the target domain. $p_i^{S \to T} \in \mathbb{R}^{H_2 \times W_2 \times C_2}$ is the pseudo-label of the i-th image in the target domain. $H_1 = 128, \ W_1 = 128, \ H_2 = 512$ and $W_2 = 512$ are the height and width of the feature matrix and pseudo-label. $C_1 = 512$ is the number of channels of the feature. C_2 is the number of segmentation categories, the size of C_2 is 2 in the optic cup and optic disc segmentation task. The first channel is the predicted probability for the cup class, The second channel is the predicted probability for the disc class. To generate the feature matrix whose size is the same as the pseudo-labels, we recover the feature dimension using the bilinear interpolation method.

We use the feature vector of the pixel to calculate the similarity between pixel j and its surrounding pixels, which is formulated as follows:

$$s_i^{S \to T}(j, j+o) = \langle f_i^{S \to T}(j), f_i^{S \to T}(j+o) \rangle$$
 (2)

where $f_i^{S \to T}(j)$ is the feature vector of pixel j. $f_i^{S \to T}(j+o)$ is the feature vector of pixel j+o, and $o \in N \times N$ is the neighborhood of pixel j. $s_i^{S \to T}(j,j+o)$ is the similarity between j and j+o. $<\cdot>$ is the similarity measure function, and we choose the cosine function as the similarity measure function.

$$s_i^{S->T}(j,j+o) = \frac{f_i^{S->T}(j) \cdot f_i^{S->T}(j+o)}{|f_i^{S->T}(j)| \times |f_i^{S->T}(j+o)|}$$
(3)

After obtaining the similarity, we use the TopK algorithm to select the pseudolabels of the first K pixels to correct the label value of pixel i. The rectified pseudolabel of image i is calculated as,

$$p_i^{S \to T}(j) = \sum_{o \in A} p_i^{S \to T}(o)/K, A = topK(s_i^{S \to T})$$

$$\tag{4}$$

3.4. Global Denoising Module

The local denoising module is designed to utilize the local context information among neighboring pixels within an image to correct the pseudo-labels. However, it is still possible for some pseudo-labels of neighboring pixels to be inaccurately predicted. Consequently, the local denoising module alone is insufficient for effectively filtering out the noisy components in the original pseudo-labels. To overcome this limitation, we propose an online pseudo-label rectification method that incorporates class prototypes containing global contextual information to correct the pseudo-labels during the training process. To obtain accurate class prototype vectors, we first divide the samples in the target domain training set into easy and hard samples. Subsequently, pseudo-label correction is then performed by comparing the relative distances of pixel-wise features to foreground prototypes and background prototypes. Furthermore, to prevent the model from overfitting to the initial pseudo-labels and producing sub-optimal performance, we update the pseudo-labels in an online manner. The proposed global denoising module is described in detail below.

Easy and Hard Sample Division: Firstly, the pixel-level predicted probability $p(x_i^T(j))$ and the feature vector $f_i^T(j)$ of pixel j from the image i in the target domain are generated by the pre-trained source domain model. Then, we use Eq.(5) to calculate the overall entropy of each image.

$$H(x_i^T) = \sum_{i} P(x_i^T(j))I(x_i^T(j)) = \sum_{i} p_i^{S->T}(j)\log_b p_i^{S->T}(j)$$
 (5)

Then, we sort the all images in the target domain based on $H(\boldsymbol{x}_i^T)$, which is formulated as:

$$img_list = sort(H(x_i^T)), x_i^T \in X_T$$
 (6)

We take η as a threshold for dividing easy and hard samples. A list of easy and hard samples is obtained according to Eq.(7) and Eq.(8).

$$hard_list = \{x_i^T | H(x_i^T) > \eta\}$$
 (7)

$$easy_list = \{x_i^T | H(x_i^T) \le \eta\}$$
(8)

Global Pseudo-label Correction Based on Class Prototype: In this step, the pseudo-label is corrected by considering the relative distance between the feature of each pixel and the prototypes of the foreground and background classes.

Firstly, we gain the segmentation prediction $\hat{y}_i^{S \to T}(j)$ and the feature vector $f_i^T(j)$ of each pixel by taking the target image x_i^T as input into the model. Secondly, we calculate the foreground and background class prototype vectors of easy samples according to Eq.(9) and Eq.(10),

$$f_{bck} = \sum_{i=0}^{N_S} \sum_{j} \mathbb{1}[p_i^{S \to T}(j) == 0] * f_i^T(j) / \sum_{i=0}^{N_S} \sum_{j} \mathbb{1}[p_i^{S \to T}(j) == 0]$$
 (9)

$$f_{obj} = \sum_{i=0}^{N_S} \sum_{j} \mathbb{1}[p_i^{S \to T}(j) == 1] * f_i^T(j) / \sum_{i=0}^{N_S} \sum_{j} \mathbb{1}[p_i^{S \to T}(j) == 1]$$
 (10)

where f_{bck} represents the background prototype vector of easy samples, and f_{obj} is the foreground prototype vector of easy samples. N_S is the number of easy samples in a batch. $f_i^T(j)$ is the feature vector of pixel j. Thirdly, we calculate the distance between the feature vector of each pixel j and the prototype vectors f_{bck} and f_{obj} by inner product as Eq.(11) and Eq.(12). Fourthly, we get the mask based on its relative distance as Eq.(13),

$$d_{obj,i}(j) = ||f_{obj}, f_i^T(j)||_2$$
(11)

$$d_{bck,i}(j) = ||f_{bck}, f_i^T(j)||_2$$
(12)

$$mask_{ic} = \begin{cases} 1, d_{obj,i} \ge d_{bck,i} \\ 0, d_{obj,i} < d_{bck,i} \end{cases}$$
 (13)

where $||\cdot||$ means 2-norm. The value of c is 0 or 1, and we get a binary optic cup mask $mask_{i0}$ and a binary optic disc $mask_{i1}$, respectively in our experiment.

Lastly, we initialize a zeros matrix p_i^T with the same size as $p_i^{S \to T}$, namely $p_i^T \in {}^{H_2 \times W_2 \times C_2}$, where H_2 and W_2 represent the height and width of the pseudo-label

respectively; C_2 represents the number of segmentation categories. Set the pseudolabel to 1 only when the corrected pseudo-label is consistent with the mask prediction, that is, the position is a trusted foreground prediction,

$$p_i^T = \begin{cases} 1, p_i^{S \to T} = mask_i \\ 0, else \end{cases}$$
 (14)

In Eq.(14) p_i denotes the local-global corrected pseudo-label. In the process of network adaptive training, with the update of network parameters, f_i^T is constantly updated, and the prototypes of easy samples are also updated online with network training.

As the model gradually adapts to the data distribution of the target domain, the target and background class prototypes of easy samples are gradually updated to more realistic class prototype vectors in the target domain. In the network iteration process, the updated class prototype vector is used for global pseudo-label correction, and the online updated corrected pseudo-label p_i is used as the label of the target domain sample for source-free domain adaptative training.

3.5. loss function

We use pseudo-labels to optimize the target domain training model. The overall loss function of the proposed method (LGDA) is formulated as Eq.(15),

$$L_T = -\sum_{x_i^T \in X_T} p_i^T \log(\phi_T(x_i^T)) + (1 - p_i^T)(1 - \log(\phi_T(x_i^T))$$
 (15)

where $\phi_T(x_i^T)$ is the prediction of the target domain model, and $\phi_T(x_i^T)$ is the pseudo-label after local-global pseudo-label optimization.

4. Experiment Settings

4.1. Dataset and evaluation metric

We use three datasets from different sites to validate the effectiveness of the proposed LGDA. REFUGE challenge[28] dataset is regarded as the source domain and

RIM-ONE-r3[29] and Drishti-GS[30] datasets are treaded as target domains. The source domain includes 400 annotated training images, and the two target domain data are split into 99/60 and 50/51 images for training/testing respectively. The data preprocessing of this paper follows the setting in the literature[4]. The fundus image is cropped into a region of interest (ROI) centered on the optic disc as the network input with the size of 512×512. Additionally, we use the common data augmentation strategies including random rotation, flipping, elastic transformation, contrast adjustment, adding Gaussian noise, and random erasing. For evaluation, we employ two commonly used metrics, including the Dice coefficient for overlap measurement and average surface distance (ASD) for boundary consistency evaluation. Higher Dice and lower ASD indicate better performance.

4.2. Implementation details

In this study, we employ the DeepLabv3+ network as the backbone segmentation network. All methods are implemented in PyTorch and trained on one NVIDIA TITAN RTX GPU. The batch size is set to 8. We use the Adam optimizer in our experiments. We set the fixed learning rate to 0.001. We set N to 3 and K to 5 in local denoising moudel. All experiments follow the same training settings for a fair comparison.

5. Experimental results and analysis

5.1. Comparison with State-of-the-Arts

We compare our method with recent state-of-the-art domain adaptation methods, including BEAL[31],AdvEnt[12],DAE[32],SRDA[25], TT[6], FSM[2] and DPL[4]. Among them, the BEAL[31] and AdvEnt[12] with the access of source domain. DAE[32], SRDA[25], TT [6], FSM[2] and DPL[4] are source-free domain adaptation methods, but DAE[32] and SRDA[25] introduce auxiliary networks to predict prior information during source domain pre-training. In Table 2 and Table 3, "w/o adaptation" represents the results obtained by directly testing the source domain model on the target domain test set, while "Upper bound" denotes the results achieved by training a model

directly on the target domain training set and subsequently testing it. Additionally, we denote the unsupervised domain adaptation method as "U", the source-free domain adaptation method as "F", and the source-free test-time DA as "T".

Table 2 and Table 3 report the quantitative experimental results on the RIM-ONE-r3 dataset and Drishti GS dataset, respectively.

Table 2: Comparison of experimental results of LGDA model in RIM-ONE-r3 dataset

	Optic disc segmentation		Optic cup s	Optic cup segmentation		Avg	
Method	Dice \uparrow [%]	$ASD \downarrow (pixel)$	$\mathrm{Dice}\uparrow [\%]$	ASD ↓(pixel)	$\mathrm{Dice}\uparrow [\%]$	$ASD \downarrow (pixel)$	
w \o adaptation	88.69±0.04	10.63±4.14	75.06±25.09	8.36±5.55	81.87	9.50	
Upper bound	94.74±2.16	4.43±1.70	80.58±20.92	7.06±6.83	87.66	5.75	
BEAL[31](U)	88.70±3.53	16.63±5.58	79.00±12.29	14.49±6.78	83.85	15.56	
AdvEnt[12](U)	89.73±3.66	9.84±3.86	77.99±21.08	7.57±4.24	83.86	8.71	
DAE[32](T)	89.09±3.32	11.63±6.84	79.01±12.82	10.31±8.45	84.05	10.97	
SRDA[25](F)	89.37±2.70	9.91±2.45	77.61±13.58	10.15±5.75	83.49	10.03	
TT[6](F)	65.34±25.91	26.30±24.75	44.87±29.31	22.28±17.74	55.10	24.29	
FSM[2](F)	84.42±4.19	16.53±9.44	80.14±13.28	8.33±4.70	82.28	12.43	
DPL[4](F)	89.47±4.56	6.92±8.24	81.93±14.96	9.56±3.57	85.70	8.24	
LGDA(Ours)(F)	91.15±2.63	7.00 ± 3.34	82.64±10.57	8.33±2.87	86.89	7.66	

According to Table 2, performance on both average Dice coefficient and average ASD in RIM-ONE-r3 dataset are 86.69% and 7.66, respectively. The highest Dice score and lowest ASD on the RIM-ONE-r3 dataset indicate the proposed method can effectively make the segmentation model pre-trained in the source domain adaptive in the RIM-ONE-r3 dataset under the source-free condition. For optic disc segmentation, LGDA achieves 91.15% on Dice, which is 1.68% higher than the second-best DPL's performance. Additionally, for optic cup segmentation, LGDA achieves 82.64% on Dice, which is 0.71% higher than the second-best DPL's performance. These results demonstrate that the local-global pseudo-label correction method effectively corrects the incorrectly predicted pseudo-labels in the target domain so that the model can better

adapt to the data distribution of the target domain. Our method achieves the second-best ASD of 7.00 on the optic disc segmentation task, which has a very close performance with the best ASD of 6.92 achieved by DPL[4]. The ASD achieved by our model for cup segmentation is lower than other comparative methods, indicating that our method can better preserve the boundaries for optic cup segmentation. It is important to highlight that our model's segmentation results outperform BEAL and Advent in terms of Dice and ASD metrics. This indicates that our method achieves superior performance even without access to the source dataset, surpassing some popular UDA methods. The success of our approach attribut to the direct adaptation of the source domain pre-trained model to the target domain data through our proposed LGDA model. This adaptation enables the model to capture more discriminative feature representations, which in turn corrects the noisy components present in the original pseudo labels for source-free domain adaptation segmentation. In contrast, unsupervised domain adaptation methods face challenges in effectively eliminating domain differences and acquiring domain invariant features.

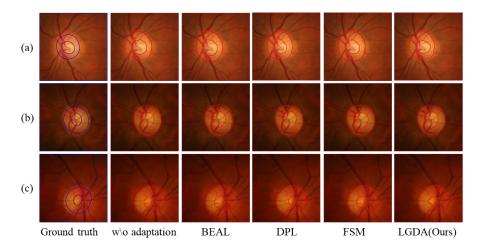


Figure 4: Visual segmentation results for typical samples on the RIM-ONE-r3 dataset

Figure 4 depicts the qualitative segmentation results of some typical samples on the RIM-ONE-r3 dataset. We can observe that our method generates more consistent optic disc segmentation and optic cup segmentation results with ground truth than other methods. The accurate and stable predictions validate the reliability of our method to face source-free adaptation.

Table 3: Comparison of experimental results of LGDA model in Drishti GS dataset

M.d. I	Optic disc s	segmentation	Optic cup s	Optic cup segmentation		Avg	
Method	$\mathrm{Dice}\uparrow [\%]$	$ASD \downarrow (pixel)$	$\mathrm{Dice}\uparrow [\%]$	ASD ↓(pixel)	$\mathrm{Dice}\uparrow [\%]$	$ASD \downarrow (pixel)$	
w \adaptation	96.66±1.12	3.78±1.34	81.55±11.94	11.94±7.86	89.10	7.86	
Upper bound	96.65±1.60	3.60±1.50	89.09±11.23	6.78±3.68	92.87	5.19	
BEAL[31](U)	95.54±2.09	7.78±3.37	85.95±11.44	14.51±8.15	90.75	11.14	
AdvEnt[12](U)	96.16±1.65	4.36±1.83	82.75±11.08	11.36±7.22	89.46	7.86	
DAE[32](T)	94.04±2.85	8.79±7.45	83.11±11.89	11.56±6.32	88.58	10.18	
SRDA[25](F)	96.22±1.30	4.88±3.47	80.67±11.78	13.12±6.48	88.45	9.00	
TT[6](F)	86.59±12.79	14.32±8.53	61.64±25.51	20.14±10.63	74.15	17.23	
FSM[2](F)	95.85±2.36	4.67±2.47	82.24±13.30	12.03±6.56	89.04	8.35	
DPL[4](F)	96.53±1.29	3.92±1.43	83.15±11.78	11.42±6.56	89.84	7.67	
LGDA(Ours)(F)	96.86±1.23	3.50±1.28	86.10±10.67	9.13±5.18	91.48	6.32	

In the Drishti-GS dataset, our LGDA method achieves the best Dice and ASD scores for optic segmentation and optic cup segmentation, as reported in Table 3. Specifically, our method achieves an average Dice score of 91.48%, which is higher than the second-best method by a significant margin, and it performs 6.32% lower than the second-best method on ASD. For optic disc segmentation, LGDA achieves a Dice score of 96.86%, surpassing the second-best method DPL by 0.33%. Additionally, in optic cup segmentation, our method achieves a Dice score of 86.10%, which is 2.95% higher than the performance of DPL. These results clearly demonstrate the effectiveness of our local-global pseudo-label correction approach in rectifying inaccurately predicted pseudo-labels within the target domain. This correction process enhances the model's ability to align with the data distribution of the target domain. Simultaneously, the segmentation results of our model outperform BEAL and Advent on Dice

and ASD, suggesting that our method outperforms some popular UDA methods without access to the source dataset.

Figure 5 shows some qualitative evaluation results on Drishti-GS. Although FSM[12] and DPL[9] outperformed the other existing SFDA methods, they have some visible errors, because they fail to fully leverage the local information present in the target domain images. The results of our method have the highest consistency with the ground truth, indicating the effectiveness and robustness of our method.

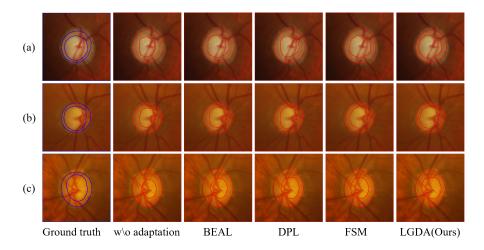


Figure 5: Visual segmentation results for typical samples on the Drishti-GS dataset

5.2. Ablation Study

5.2.1. Ablation Studies on Easy and hard sample division

The method of easy and hard sample division. We conduct experiments utilizing three distinct methods to divide the target domain data into simple and hard samples, and mark them as "TopK," "AVG," and "H > η ," respectively. In the first method, we select the TopK samples in the target domain dataset where K is set to 0.3 times the total number of samples in the training set. In the second method, we calculate the average entropy value across the entire target domain training and then choose samples with entropy values lower than the calculated average as simple samples. As for the

third method, we designate samples with values lower than the threshold η as simple samples. In this case, we assume that the prediction probability of each pixel exceeds 0.9, which is considered more reliable. The entropy value of the entire image is then calculated using Eq.5. In our implementation, we set η to 0.044.

Table 4: The validation of easy and hard sample partitioning of the RIM-ONE-r3 dataset

	Optic disc[%]	Optic cup[%]	Avg[%]
TopK	91.10	81.39	86.24
AVG	93.76	78.84	86.30
$H > \eta$	91.15	82.64	86.89

Table 5: The validation of easy and hard sample partitioning of the Drishti-GS dataset

	Optic disc[%]	Optic cup[%]	Avg[%]
TopK	96.31	82.51	89.41
AVG	94.61	79.46	87.04
$H>\eta$	96.86	86.10	91.48

The experimental results are presented in Table 4 and Table 5. The Dice coefficient is employed as the evaluation metric. By comparing the experimental outcomes on both datasets, we observe that the third method is more suitable for dividing samples into easy and hard categories. Consequently, we select the third method for dividing samples into simple and difficult categories in the final model.

To validate the effectiveness of the selected method for dividing samples into easy and hard categories, we conducte experiments using the three aforementioned methods. Firstly, we apply these methods to obtain two lists: one comprising easy samples and the other consisting of hard samples. Next, we analyze the grayscale distribution of the grayscale images associated with the easy and hard samples, respectively. Finally, we compare these distributions with the grayscale distribution of the source domain data

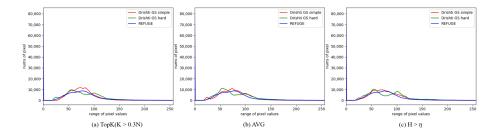


Figure 6: Drishti-GS dataset easy difficult sample segmentation grayscale statistics

image. Figure 6 illustrates the statistical results of the grayscale distribution, where each graph represents the outcomes obtained through the "TopK," "AVG," and "H> η " division methods, respectively. The blue, red, and green lines correspond to the gray distribution of the source domain dataset, easy samples of the target domain dataset, and difficult samples of the target domain dataset, respectively. By examining the statistical results, it is evident that the gray distribution of the simple samples divided using the "H> η " method closely aligns with the gray distribution of the source domain. This observation confirms the rationale behind selecting the third-division method as the more reasonable approach.

Impact of dividing easy and hard sample on global denoising module. We conduct two sets of experiments to verify the impact dividing easy and hard sample on global denoising module. These two sets of experiments are the LGDA method without using easy sample class prototypes for global denoising and the LGDA method using easy samples class prototypes for global denoising. The experimental results are shown in Table 6 and Table 7.

Table 6: Impact of dividing easy and hard sample on global denoising module in RIM-ONE-r3 dataset

Optic disc so		gmentation Optic cup segmentation		Optic cup segmentation Avg		Avg
Wictiod	Dice [%]	ASD(piexl)	Dice [%]	ASD(piexl)	Dice [%]	ASD(piexl)
w \o division	89.93±9.27	9.27±3.04	76.60±7.90	7.90±5.05	83.27	8.58
w division	91.15±2.63	8.33±2.87	82.64±10.57	7.00±3.34	86.89	3.11

Table 7: Impact of dividing easy and hard sample on global denoising module in Drishti-GS dataset

Optic disc		egmentation Optic disc segm		gmentation	1	Avg
Wiethod	Dice [%]	ASD(piexl)	Dice [%]	ASD(piexl)	Dice [%]	ASD(piexl)
w \o division	96.82±1.34	3.57±1.45	83.22±10.16	11.07±6.29	90.02	7.32
w division	96.86±1.23	3.50±1.28	86.10±10.67	9.13±5.18	91.48	6.32

Compared to using a simple easy class prototype for denoising methods, using all sample class prototypes for denoising methods results in a significant decrease in the Dice coefficient for optic disc and optic cup segmentation. This attributs to the presence of biased prototypes. These biased class prototypes cause the optic cup pixels to be closer to the prototype of the background class. By incorporating a global pseudo-label correction method based on easy sample, the performance of optic disc segmentation is enhanced. Specifically, the Dice coefficients for the RIM-ONE-r3 and Drishti-GS datasets show improvements of 1.22% and 0.04% respectively. In terms of optic cup segmentation, our method demonstrates a significant enhancement in Dice by 6.04% and 2.88% compared to the method that does not divide samples into simple ones and hard ones. These results highlight the importance of extracting unbiased prototype vectors in order to improve the accuracy of segmentation.

5.2.2. Ablation Studies on Key Components

We conduct ablation experiments on the REFUGE dataset and Drishti-GS dataset to validate the effectiveness of each module of our proposed model. Concretely, we use the test results generated directly from the source domain pre-trained model as the Baseline. Next, we incorporated a local pseudo-label correction module onto the baseline method, denoted as "Baseline+L". Finally, we intorduce global denoising module based on easy sample class prototype, denoted as "Baseline+LG"(i.e., the final model LGDA). We use the Dice coefficient and the ASD coefficient as evaluation metrics for ablation experiments.

Table 8: Ablation experiment results of LGDA model in RIM-ONE-r3 dataset

Method Optic disc see		gmentation Optic cup seg		gmentation	Avg	
		ASD(piexl)	Dice [%]	ASD(piexl)	Dice [%]	ASD(piexl)
Baseline	88.69±0.04	10.63±4.14	75.06±25.09	8.36±5.55	81.87	9.50
Baseline+L	88.91±10.48	10.48±3.68	80.91±7.71	7.71±6.65	84.94	9.09
Baseline+LG	91.15±2.63	8.33±2.87	82.64±10.57	7.00±3.34	86.89	3.11

Table 9: Ablation experiment results of LGDA model in Drishti-GS dataset

Method	Optic disc segmentation		Optic disc segmentation		Avg	
Method	Dice [%]	ASD(piexl)	Dice [%]	ASD(piexl)	Dice [%]	ASD(piexl)
Baseline	96.66±1.12	3.78±1.34	81.55±11.94	11.94±7.86	89.10	7.86
Baseline+L	96.75±1.40	3.68±1.50	83.37±10.28	10.98±6.26	90.06	7.33
Baseline+LG	96.86±1.23	3.50±1.28	86.10±10.67	9.13±5.18	91.48	6.32

The experimental results are shown in Table 8 and Table 9. Firstly, after incorporating the local denoising module, the optic disc segmentation achieves Dice coefficients of 88.91% and 97.75% on the RIM-ONE-r3 dataset and Drishti-GS dataset, respectively. Compared to the baseline, there is an improvement of 0.22% and 0.09%,

respectively. These results indicate that the local pseudo-label correction module remains effective even in the presence of small segmentation errors. Additionally, for the optic cup segmentation task, the Dice coefficients reach 80.91% and 83.37%, representing an improvement of 5.85% and 1.82% compared to the baseline. The significant enhancement in optic cup segmentation demonstrates that the local denoising module effectively leverages local spatial continuity to refine pixel-level segmentation, thereby providing reliable initial pseudo-labels for adaptive training.

Furthermore, the integration of the class prototype-based global pseudo-label correction brings a significant improvment in optic disc segmentation accuracy across both target domains. Specifically, after introducing the global denoising module, the Dice scores for the optic cup segmentation task on the RIM-ONE-r3 dataset increased by 2.24% and 2.03% respectively. At the same time, the ASD coefficients decreased by 2.15 and 0.71. On the Drishti-GS dataset, the Dice scores for the optic disc segmentation and optic cup segmentation tasks increased by 0.11% and 2.73% respectively, while the ASD scores decreased by 0.18 and 1.85. The global denoising moudle utilizing class prototypes leverages the global image information to rectify pseudo-labels assigned to pixels that exhibit higher similarity to the target class prototype in the feature space but are mistakenly labeled as background.

Ultimately, the optimal performance of our model is achieved through the integration of the local denoising and global denoising modules. The superior performance demonstrates the effectiveness of our approach, which utilizes source domain prior knowledge to effectively leverage the local spatial continuity and global semantic information of target domain images during training. By employing offline correction of the local pseudo-labels, it provides reliable initial pseudo-labels for adaptive training and effectively prevents error accumulation. And by employing online correction of pseudo-labels using the global denoising module, the accuracy of pseudo-labels is further enhanced, enabling the model to successfully adapt to the data distribution of the target domain.

The segmentation error visualization results of some typical samples on the RIM-

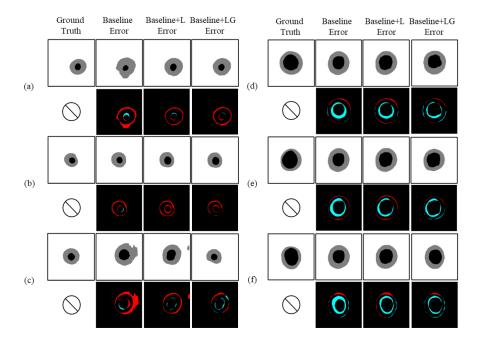


Figure 7: LGDA model ablation experiment error visualization results on RIM-ONE-r3 dataset

ONE-r3 dataset and Drishti-GS dataset are shown in Figure 7. In Figure 7, each sample is represented by two rows. The first row visually presents the segmentation results, while the second row depicts the errors in comparison to the ground truth. The first column represents the ground truth, and the subsequent columns (second to fourth) correspond to the results of the Baseline, Baseline + L, and Baseline + LG, respectively. Within the error maps, the blue regions indicate areas where the predictions fall short of the ground truth, while the red regions highlight areas where the predictions exceed the ground truth.

By comparing the second and fourth columns, we can observe a significant reduction in errors for the LGDA model relative to the baseline model. Specifically, in examples (a) and (c), a distinct decrease in prediction errors is evident when incorporating local pseudo-label correction in the adaptive model, as seen from the comparison between the second and third columns. After partitioning the target domain samples into easy ones and hard ones and subsequently applying global pseudo-label correction,

in example (c), there is a notable enhancement in segmentation quality, surpassing the results obtained from local pseudo-label correction. This observation underscores the effectiveness of the pseudo-label correction method based on prototypes derived from easy samples following the introduction of easy and hard sample classification.

Moreover, through a step-by-step visualization and comparison of the errors introduced by each module, it is apparent that the model's prediction errors steadily decrease. This observation provides compelling evidence for the interplay and complementary nature of the various modules proposed, ultimately resulting in a notable improvement in the model's performance for source-free domain adaptation in the optic cup and disc segmentation tasks.

6. Conclusions

In this paper, we propose a local-global pseudo-label correction method for reliable source-free domain adaptation in the medical image segmentation method. Our LGDA can provide more accurate pseudo-labels for the target domain to improve the segmentation performance of the source domain model in the target domain dataset. We propose a local label correction method that can take the neighborhood pixels into consideration. Second, we propose a global label correction method from the perspective of prototypical learning. To reduce errors in a global pseudo-label correction, we propose an easy-hard sample division method. Extensive experiments on RIM-ONE-r3 and Drishti-GS demonstrate the effectiveness of our method LGDA. In the RIM-ONEr3 dataset, LGDA can improve 2.46% and 7.58% on Dice in the optic disc and cup segmentation compared to the results generated directly from the source domain pretrained model. In the Drishti-GS dataset, LGDA can improve 0.20% and 4.55% on Dice in the optic disc and cup segmentation compared to baseline. Our forthcoming research will concentrate on the challenges of federated learning in the field of medical segmentation. Furthermore, we will explore the issue of federative learning in the context of weak supervision.

References

- [1] V. N. Vapnik, An overview of statistical learning theory, IEEE Transactions on Neural Networks (1999) 988–999.
- [2] C. Yang, X. Guo, Z. Chen, Y. Yuan, Source free domain adaptation for medical image segmentation with fourier style mining, Medical Image Analysis (2022) 102457.
- [3] M. Ye, J. Zhang, J. Ouyang, D. Yuan, Source data-free unsupervised domain adaptation for semantic segmentation, in: Proceedings of the 29th ACM International Conference on Multimedia, 2021, pp. 2233–2242.
- [4] C. Chen, Q. Liu, Y. Jin, Q. Dou, P.-A. Heng, Source-free domain adaptive fundus image segmentation with denoised pseudo-labeling, in: Medical Image Computing and Computer Assisted Intervention, 2021, pp. 225–235.
- [5] Z. Xu, D. Lu, Y. Wang, J. Luo, D. Wei, Y. Zheng, R. K.-y. Tong, Denoising for relaxing: Unsupervised domain adaptive fundus image segmentation without source data, in: Medical Image Computing and Computer Assisted Intervention, 2022, pp. 214–224.
- [6] V. VS, J. M. J. Valanarasu, V. M.Patel, Target and task specific source-free domain adaptive image segmentation, arXiv preprint arXiv:2203.15792 (2022).
- [7] F. You, J. Li, L. Zhu, Z. Chen, Z. Huang, Domain adaptive semantic segmentation without source data, in: Proceedings of the 29th ACM International Conference on Multimedia, 2021, pp. 3293–3302.
- [8] M. Long, Y. Cao, J. Wang, M. Jordan, Learning transferable features with deep adaptation networks, in: International conference on machine learning, 2015, pp. 97–105.
- [9] E. Tzeng, J. Hoffman, N. Zhang, K. Saenko, T. Darrell, Deep domain confusion: Maximizing for domain invariance, arXiv preprint arXiv:1412.3474 (2014).

- [10] Y. Ganin, E. Ustinova, H. Ajakan, P. Germain, H. Larochelle, F. Laviolette, M. Marchand, V. Lempitsky, Domain-adversarial training of neural networks, The journal of machine learning research (2016) 2096–2030.
- [11] E. Tzeng, J. Hoffman, K. Saenko, T. Darrell, Adversarial discriminative domain adaptation, in: Proceedings of the IEEE conference on computer vision and pattern recognition, 2017, pp. 7167–7176.
- [12] T.-H. Vu, H. Jain, M. Bucher, M. Cord, P. Pérez, Advent: Adversarial entropy minimization for domain adaptation in semantic segmentation, in: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, 2019, pp. 2517–2526.
- [13] M. Chen, H. Xue, D. Cai, Domain adaptation for semantic segmentation with maximum squares loss, in: Proceedings of the IEEE/CVF International Conference on Computer Vision, 2019, pp. 2090–2099.
- [14] F. Pan, I. Shin, F. Rameau, S. Lee, I. S. Kweon, Unsupervised intra-domain adaptation for semantic segmentation through self-supervision, in: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, 2020, pp. 3764–3773.
- [15] W. Yan, Y. Wang, M. Xia, Q. Tao, Edge-guided output adaptor: Highly efficient adaptation module for cross-vendor medical image segmentation, IEEE Signal Processing Letters (2019) 1593–1597.
- [16] T. Zhang, J. Cheng, H. Fu, Z. Gu, Y. Xiao, K. Zhou, S. Gao, R. Zheng, J. Liu, Noise adaptation generative adversarial network for medical image analysis, IEEE transactions on medical imaging (2019) 1149–1159.
- [17] Y. Zou, Z. Yu, X. Liu, B. Kumar, J. Wang, Confidence regularized self-training, in: Proceedings of the IEEE/CVF international conference on computer vision, 2019, pp. 5982–5991.

- [18] Y. Zou, Z. Yu, B. Kumar, J. Wang, Unsupervised domain adaptation for semantic segmentation via class-balanced self-training, in: Proceedings of the European conference on computer vision, 2018, pp. 289–305.
- [19] V. K. Kurmi, V. K. Subramanian, V. P. Namboodiri, Domain impression: A source data free domain adaptation method, in: Proceedings of the IEEE/CVF winter conference on applications of computer vision, 2021, pp. 615–625.
- [20] R. Li, Q. Jiao, W. Cao, H.-S. Wong, S. Wu, Model adaptation: Unsupervised domain adaptation without source data, in: Proceedings of the IEEE/CVF conference on computer vision and pattern recognition, 2020, pp. 9641–9650.
- [21] Y. Kim, D. Cho, K. Han, P. Panda, S. Hong, Domain adaptation without source data, IEEE Transactions on Artificial Intelligence (2021) 508–518.
- [22] J. Liang, D. Hu, J. Feng, Do we really need to access the source data? source hypothesis transfer for unsupervised domain adaptation, in: International Conference on Machine Learning, 2020, pp. 6028–6039.
- [23] S. Yang, J. van de Weijer, L. Herranz, S. Jui, et al., Exploiting the intrinsic neighborhood structure for source-free domain adaptation, Advances in neural information processing systems (2021) 29393–29405.
- [24] F. Fleuret, et al., Uncertainty reduction for model adaptation in semantic segmentation, in: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, 2021, pp. 9613–9623.
- [25] M. Bateson, H. Kervadec, J. Dolz, H. Lombaert, I. Ben Ayed, Source-relaxed domain adaptation for image segmentation, in: Medical Image Computing and Computer Assisted Intervention–MICCAI 2020: 23rd International Conference, Lima, Peru, October 4–8, 2020, Proceedings, Part I 23, 2020, pp. 490–499.
- [26] L.-C. Chen, Y. Zhu, G. Papandreou, F. Schroff, H. Adam, Encoder-decoder with

- atrous separable convolution for semantic image segmentation, in: Proceedings of the European Conference on Computer Vision, 2018, pp. 801–818.
- [27] F. Chollet, Xception: Deep learning with depthwise separable convolutions, in: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, 2017, pp. 1251–1258.
- [28] J. I. Orlando, H. Fu, J. B. Breda, K. Van Keer, D. R. Bathula, A. Diaz-Pinto, R. Fang, P.-A. Heng, J. Kim, J. Lee, et al., Refuge challenge: A unified framework for evaluating automated methods for glaucoma assessment from fundus photographs, Medical image analysis (2020) 101570.
- [29] F. Fumero, S. Alayón, J. L. Sanchez, J. Sigut, M. Gonzalez-Hernandez, Rimone: An open retinal image database for optic nerve evaluation, in: 2011 24th international symposium on computer-based medical systems, 2011, pp. 1–6.
- [30] J. Sivaswamy, S. Krishnadas, A. Chakravarty, G. Joshi, A. S. Tabish, et al., A comprehensive retinal image dataset for the assessment of glaucoma from the optic nerve head analysis, JSM Biomedical Imaging Data Papers (2015) 1004.
- [31] S. Wang, L. Yu, K. Li, X. Yang, C.-W. Fu, P.-A. Heng, Boundary and entropy-driven adversarial learning for fundus image segmentation, in: Medical Image Computing and Computer Assisted Intervention, 2019, pp. 102–110.
- [32] N. Karani, E. Erdil, K. Chaitanya, E. Konukoglu, Test-time adaptable neural networks for robust medical image segmentation, Medical Image Analysis (2021) 101907.