Supportive Negatives Spectral Augmentation for Source-Free Cross-Domain Segmentation

Kexin Zheng[1,2], Haifeng Xia[2*], Siyu Xia[1,2*], Ming Shao[3], Zhengming Ding[4]
[1]Advanced Ocean Institute, Southeast University, Nantong, China
[2]School of Automation, Southeast University, Nanjing, China
[3]Department of Computer and Information Science, University of Massachusetts Dartmouth, USA
[4]Department of Computer Science, Tulane University, USA
{zhengkx, hfxia, xsy}@seu.edu.cn, mshao@umassd.edu, zding1@tulane.edu

Abstract

Source-free domain adaptation (SFDA) aims to transfer knowledge from the well-trained source model and optimize it to adapt target data distribution. SFDA methods are suitable for medical image segmentation tasks due to its data-privacy protection and achieve promising performances. However, cross-domain distribution shift makes it difficult for the adapted model to provide accurate decisions on several hard instances and negatively affects model generalization. To overcome this limitation, a novel method 'supportive negatives spectral augmentation' (SNSA) is presented in this work. Concretely, SNSA includes the instance selection mechanism to automatically discover a few hard samples for which the source model produces incorrect predictions. And an active learning strategy is adopted to re-calibrate their predictive masks. Moreover, SNSA deploys the spectral augmentation between hard instances and others to encourage the source model to gradually capture and adapt the attributions of the target distribution. Considerable experimental studies demonstrate that annotating merely 4%∼5% of negative instances from the target domain significantly improves the segmentation performance over previous methods.

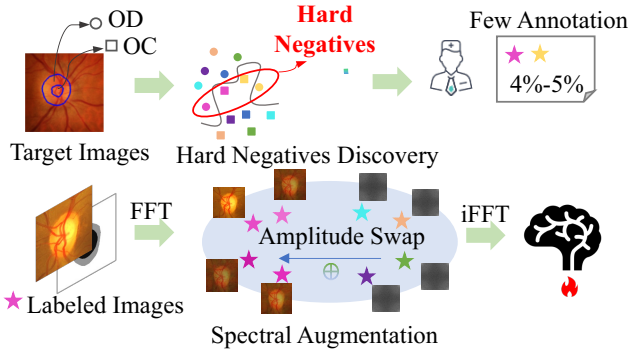


Figure 1: A pipeline overview of our proposed method, *Supportive Negatives Spectral Augmentation*(SNSA). In the context of source-free domain adaptation, an active learning-based approach is proposed in this paper to select representative 'hard negative' instances from the target domain and utilize experts' decisions and signal processing mechanism (spectral augmentation) to optimize the target network.

Introduction

Taking advantage of the rapid development of deep learning, neural network-based methods have recently achieved promising performance in medical image segmentation, especially for the analysis of fundus images (Ronneberger, Fischer, and Brox 2015; Milletari, Navab, and Ahmadi 2016). However, these methods generally suffer from performance degradation when evaluated on different sites. The main reason is that different parameter settings in data acquisition, e.g., illumination and resolution, result in the data distribution shift. For example, when medical image segmentation models are trained on a dataset with abundant labeled data (source) and then applied to a different domain (target) with limited labeled data, the performance is usually suboptimal and low-precision.

In terms of this problem, unsupervised domain adaptation (UDA) solutions shine a light by learning domain-invariant representations from different data silos (source domain & target domain) (Xia and Ding 2020; Ju et al. 2021; Varsavsky et al. 2020). However, these methods are unable to handle learning tasks where source medical data is unavailable due to data privacy and ethical considerations.

To that end, Source-Free Domain Adaptation (SFDA) eliminates the reliance on source data in the transfer process (Liang, Hu, and Feng 2020; Xia, Zhao, and Ding 2021; Ding et al. 2022). Recent efforts have extended SFDA to fundus image segmentation. One common practice is to leverage the prediction results of the source model as pseudo-labels to facilitate the target model's adaptation. However, these strategies are susceptible to label noise, impacting their effectiveness. To mitigate this, DPL (Denoised Pseudo-Labeling) (Chen et al. 2021a) introduces a denoising strategy that filters out pseudo-labels with high uncertainty and employs pixel-level and class-level constraints for pseudo-label refinement. Meanwhile, CPR (Context-Aware Pseudo-label Refinement) (Huai et al. 2023) presents a denoising approach that capitalizes on the contextual relationships between similar features, and CBMT (Class-Balanced Mean Teacher) (Tang et al. 2023) develops loss calibration techniques to address the imbalance between foreground and

*Corresponding author.

background elements. Despite these advancements, SFDA struggles to address difficult samples due to the limited scope of the source model's knowledge. When populated to the target domain, such limitation would give rise to hard negative samples. It is difficult for source model to provide them with accurate predictions due to the significant domain shift. This inspires us to rethink the roles of hard negative samples to improve SFDA for medical image segmentation.

There are conceptually similar active learning works to select decision boundary samples, e.g., BoostMIS (Zhang et al. 2022), SALAD (Kothandaraman et al. 2023a). Nonetheless, these methods do not perform well for the problem of screening hard negative samples in medical image segmentation. Differently, this paper emphasizes the effectiveness of selecting and annotating the most representative hard negatives that demonstrate distribution inconsistencies in the target domain. Fine-tuning the target network with these annotated samples can significantly reduce losses caused by cross-domain divergence. Specifically, as illustrated in Figure 1, this paper presents a strategy to actively select only a few hard negative samples in the target domain and introduces a data augmentation method based on Fourier spectral fusion. The major contributions are summarized as:

- We introduce a novel "Active Hard Negative Discovery" method to identify and annotate the most supportive instances for medical image segmentation within the Source-Free Domain Adaptation (SFDA) context.
- We propose an innovative spectral data augmentation technique to maximize the utility of the inherent knowledge in the hard negative samples, thereby enriching the training dataset with more informative variations.
- We validate the significant impact of hard negative instances on adaptation effectiveness. Our findings underscore the importance of these instances and demonstrate the superior performance of our proposed methods over current SFDA approaches.

Related Work

SFDA in Medical Image Analysis

Source-Free Domain Adaption (SFDA) deals with domain adaption without access to source-domain data (Xia, Xia, and Ding 2024). In medical image segmentation, one pioneering work devised a novel source-free domain adaptation framework with Fourier-style mining, where only a well-trained source segmentation model is available (Yang et al. 2022a). As a successful follow-up, Uncertainty-aware Pseudo Label guided (UPL) was proposed to enhance the diversity of predictions in the target domain by duplicating the pre-trained model's prediction head multiple times with perturbations (Wu et al. 2023). Pseudo labels were also explored in (Chen et al. 2021b) to effectively make use of the source model and unlabeled target data to promote model self-adaptation. Moreover, research in (Yu et al. 2023) investigated a new problem termed "source-free open compound domain adaptation (SF-OCDA)" and studied it in semantic segmentation. To tackle the source data-absent problem, Zhao et al. presented a novel two-stage SFDA framework for medical image segmentation (Zhao et al. 2022), where only a well-trained source segmentation model and unlabeled target data are available during domain adaptation. Context-aware pseudo-label refinement method for SF-UDA was proposed in (Huai et al. 2023), aiming to develop a context-similarity learning module to learn context relations. Despite the significant progress made by SF-UDA, it may lead to distortion in the target domain's distribution (Ning et al. 2021). To prevent this issue and to better align with the reality where different medical institutions typically perform limited expert annotations while keeping the annotation workload minimal, a promising strategy is Active Learning (AL), which we have adopted in this study.

Active Learning for SFDA

Active Learning (AL) aims to select and label a few samples or instances with minimal annotation efforts to achieve optimal performance (Cohn, Ghahramani, and Jordan 1996). It has been extended to domain adaptation (DA) by prioritizing representatives of domains to be annotated to reflect the domain knowledge and features. One representative work, multi-anchor-based active learning, used multiple anchors instead of a single centroid to model the source domain as a multimodal distribution (Ning et al. 2021), and thus allowed more representative and complimentary samples to be selected from the target domain.

AL also plays increasingly important roles in SFDA. It has been found that criteria such as neighbor-chaotic, individual-different, and source dissimilar are keys in sample selection (Wang et al. 2023). These points were defined as the Minimum Happy points that were underexplored by the existing methods. Along this line, SALAD was proposed for the challenging vision task of adapting a pre-trained source domain network to the target domain (Kothandaraman et al. 2023b), with a small budget for annotation in the target domain and a shift in the label space.

Note that most existing work on Active Source Free Domain Adaptation (ASFDA) was applied to classification problems, providing better insights into the degree of dispersion in sample distributions and enabling the selection of sample points that are more inconsistent with the source domain. For medical image segmentation, however, classification is performed at the pixel level. The differences between foreground and background pixels within tissues are significant and require expert knowledge. Additionally, the differences between pixels within a single image and between samples are not as pronounced as in classification problems, highlighting a substantial research gap in this area.

Method

Preliminary & Motivation

In fundus image segmentation under SFDA setting, given a well-trained model $\mathcal{M}_s = \{\mathcal{E}(\cdot), \mathcal{D}(\cdot)\}$ from source data, the challenge is to transfer the domain-invariant knowledge to the unlabeled target domain $\{X_i^t | X_i^t \in \mathbb{R}^{C_x \times H_x \times W_x}\}_{i=1}^{n_t}$, where $\{\mathcal{E}(\cdot), \mathcal{D}(\cdot)\}$ denotes encoder and decoder, respectively, and C_x, H_x, W_x are the channel number, height, and width of the input image. Most existing SFDA methods in fundus segmentation use the prediction

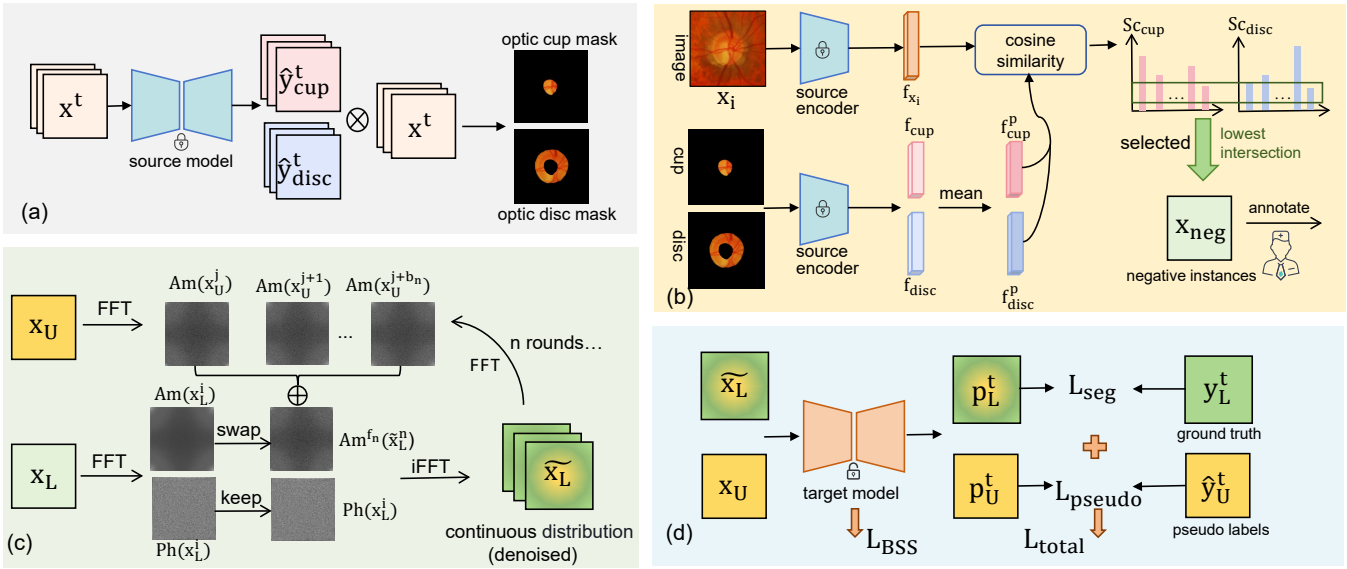


Figure 2: Overview of our framework. (a) Generate pseudo label, (b) Active hard negative instance selection, (c) Spectral Mixed Data augmentation, and (d) Fine-tuning in the target domain.

results of the source model as pseudo-labels to guide the adaptation of the target model. Specifically, given the unlabeled images in the target domain, we denote p_k as the predicted probability of the source model on the k-th pixel. The corresponding pseudo-label can then be generated as: $\hat{y}_{t_k} = 1[p_k \geq \gamma]$, where p_k and y_k are the k-th dimension of p and y, respectively, denoting the prediction and pseudo-label for class k. The target model $\mathcal{M}_t$ is trained using cross-entropy loss. As the domain gap may contain noise, directly using $\hat{y}$ to supervise the network will lead to limited performance. This leads to the presence of hard negative samples, where the model's prediction significantly differs from the ground truth, potentially disrupting the transfer effect. These samples may contain significant domain inconsistencies, motivating further exploration of their role in SFDA.

Inspired by active learning, we propose a novel approach to select a few samples to boost model performance. Our approach involves defining a strategy S to select n_L hard negatives from the total n_t target samples, where $n_L \ll n_t$, showing typical data shift in the source model's performance. Thus, the target-domain data consist of n_t image-label pairs $X_L^t = \{(X_L^t, Y_L^t)\}$ and $n_t - n_L$ unlabeled images $X_U^t = \{(X_U^t)\}$. We aim to optimize the annotation procedure for these hard negatives to improve the model's performance while keeping n_L small.

Active Hard Negative Discovery

To identify representative instances that exhibit poor performance on the source model as negative examples, we consider samples that significantly deviate from the feature space of the learned model. Based on the observation that pixels in the same category naturally cluster in the feature space (Zhang et al. 2019), we expect similar phenomena in the target domain and propose a feature centroid-based hard negative instances selecting method, as illustrated in Algorithm 1 stage 1. For each sample X_i in the target domain, a pseudo-label $\hat{Y}_i^k$ is generated by the source model $\hat{Y}_i^k = \mathcal{M}_s(X_U^t)$ via the prediction for the k-th class. For each image, segmented mask images of the k-th class can be extracted $X_i^k = X_i \odot \hat{Y}_i^k$, where $\odot$ is pixel-wise multiplication. With the pre-trained source model's feature extraction component, we are allowed to extract features of the two types of mask images separately and obtain the mean feature of all samples in the target domain as the centroid feature:

$$\mathbf{f}_i^k = \mathcal{E}(X_i^k), \quad \mathbf{f}_p^k = \frac{1}{n_t} \sum_{i=1}^{n_t} \mathcal{E}(X_i^k), \tag{1}$$

where $\mathcal{E}(\cdot)$ denotes the feature extraction using the encoder of the backbone. $\mathbf{f}_i^k$ represents the feature of the i-th sample of the k-th class extracted by $\mathcal{E}$, and $\mathbf{f}_p^k$ signifies the centroid feature of the k-th class. This process allows us to derive the centroid features of the optic disc and optic cup extracted by the source model, facilitating the computation of the cosine similarity between each instance and the centroid feature as $S_{c_i^k} = \mathbf{cos}(\mathbf{f}_i^k, \mathbf{f}_p^k)$. After that, we calculate the scores $S_{c_i^k}$ for all samples with respect to the k-th class centroid and sort them in ascending order.

We select the top n_L intersection of the k classes' samples with the lowest scores as hard negative instances. As a results, the target domain dataset comprises n_t image-label pairs $X_L^t = \{(x_L^t, y_L^t)\}$ and $n_t - n_L$ unlabeled images $X_U^t = \{x_U^t\}$. Medical experts will annotate the identified hard negative instances, usually a small amount. This strategy not only helps keep labeling costs low but also achieves good transfer effects, even without employing any other task-specific adaptation methods.

Spectral Mixed Augmentation

After annotating a small number of hard negative instances, we designed a data augmentation method based on Fast Fourier Transform (FFT) for multiple rounds of continuous spectrum fusion (as Algorithm 1 stage 2). This method aims to effectively utilize these annotated samples to enhance the model's robustness and generalization by reducing noise from cross-domain inconsistencies. Specifically, for each annotated hard negative instance $X_L \in \mathbb{R}^{H\times W\times C}$ and the remaining unlabeled regular samples $X_U \in \mathbb{R}^{H\times W\times C}$, we perform FFT using the following:

$$\mathcal{F}(X)(u,v) = \sum_{h=0}^{H-1}\sum_{W=0}^{W-1} X(h,w)\mathbf{e}^{-i2\pi\left(\frac{h}{H}u+\frac{w}{W}v\right)}, \quad (2)$$

where $\mathcal{F}(\cdot)$ denotes the Fast Fourier Transform algorithm and obtains the amplitude and phase information of the image reflecting the low-level distribution (e.g., style) and high-level semantics (e.g., object) of the image:

$$\begin{aligned}\mathbf{Am}(u,v) &= |\mathcal{F}(u,v)|,\\ \mathbf{Ph}(u,v) &= \mathbf{arctan}\Big(\frac{\mathrm{Im}(\mathcal{F}(u,v))}{\mathrm{Re}(\mathcal{F}(u,v))}\Big),\end{aligned} \quad (3)$$

where $\mathrm{Re}(\mathcal{F}(u,v))$ and $\mathrm{Im}(\mathcal{F}(u,v))$ denote the real and imaginary parts of $\mathcal{F}(u,v)$, respectively. For each annotated negative instance X_L, the amplitude is $\mathbf{Am}_{L_i}$ and the phase is $\mathbf{Ph}_{L_i}$, while the unlabeled X_U samples have amplitude $\mathbf{Am}_{U_j}$ and phase $\mathbf{Ph}_{U_j}$. Inspired by (Xu et al. 2021), we explore a spectral mixed augmentation strategy for continuous amplitude fusion:

$$\begin{aligned}\mathbf{Am}^{f_n}(\tilde{x}_L^n) =& \frac{1}{b_n}[\mathbf{Am}(x_{U_1}^{n-1}) + \mathbf{Am}(x_{U_2}^{n-1}) + \\ &\ldots + \mathbf{Am}(x_{U_{b_n}}^{n-1})],\end{aligned} \quad (4)$$

where the amplitude of one batch (size of b_n) of X_U samples is taken and averaged to obtain the blended fused amplitude $\mathbf{Am}^{f_n}$, where f_n represents the n-th round of fusion. Denoising is achieved at the amplitude level refer to Theorem 1. Detailed proof can be seen in Proof 1.

The amplitude of X_L, $\mathbf{Am}_L$, is replaced with the fused amplitude $\mathbf{Am}^{f_n}$, and the Fourier inverse transform is performed to obtain the n-th round of the fused amplitude. The Fourier inverse transform is used to get the augmented data:

$$\tilde{X}_L^n = \mathcal{F}^{-1}[\mathbf{Am}^{f_n}(\tilde{X}_L^n), \mathbf{Ph}(X_L)], \quad (5)$$

which yields augmented denoised labeled data: $\{A(X_L)\} = \{\widetilde{X_L^0}, \widetilde{X_L^1}...\widetilde{X_L^n})\}$, where $A(\cdot)$ denotes the data augmentation operation for the above continuous hybrid spectrum fusion. By performing n rounds of spectral fusion, the augmented data will be applied to eliminate cross-domain difference and enhances model robustness.

Theorem 1 *Let the amplitude of noiseless images be* $\mathbf{Am}(u,v)$*, Amplitude of noise be* $\mathbf{Am}^n(u,v)$*, Corrupted amplitude of noisy image be* $\mathbf{Am}'(u,v)$*,* $\mathbf{Am}'(u,v) = \mathbf{Am}(u,v) + \mathbf{Am}^n(u,v)$*, then reducing the noise by adding a set of noisy amplitude of images:* $\{\mathbf{Am}'_i(u,v)\}$*, the mixed amplitude can be* $\overline{\mathbf{Am}}(u,v) = \frac{1}{N}\sum_{i=1}^{N}\mathbf{Am}'_i(u,v)$.

Proof 1

$$\begin{aligned}E\left\{\overline{\mathbf{Am}}(u,v)\right\} &= E\left\{\frac{1}{N}\sum_{i=1}^{N}\mathbf{Am}'_i(u,v)\right\}\\ &= E\left\{\frac{1}{N}\sum_{i=1}^{N}[\mathbf{Am}(u,v) + \mathbf{Am}^{n_i}(u,v)]\right\}\\ &= \mathbf{Am}(u,v) + E\left\{\frac{1}{N}\sum_{i=1}^{N}\mathbf{Am}^{n_i}(u,v)\right\}\\ &= \mathbf{Am}(u,v)\end{aligned}$$

□

Overall Objective and Methodology

In this section, we elaborate on the training objectives and methodologies for the target model. Our training strategy not only utilizes augmented data with their true labels, denoted as $\{A(X_L), Y(X_L)\}$, but also incorporates samples with pseudo-labels $\{X_U, \hat{Y}(X_U)\}$. This dual approach aims to impose concurrent constraints on the network during the training process, thereby enhancing the model's generalization capabilities across various data distributions.

To mitigate the adverse impact of domain-specific information on model performance, we adopt the Batch Spectral Shrinkage (BSS) strategy (Chen et al. 2019). This technique mainly constrains the smallest eigenvalues of the feature matrix and encourages the network to learn domain-invariant knowledge. Specifically, features $\mathbf{f}_i$ extracted from a batch of samples are concatenated to form a matrix Q. We then employ Singular Value Decomposition (SVD) to compute all singular values Σ with $Q = U\Sigma V^\top$.

The Batch Spectral Shrinkage approach penalizes the smallest k singular values on the diagonal of Σ, which are denoted by $[\sigma_1, \sigma_2, ..., \sigma_b]$. The corresponding BSS loss function, $\mathcal{L}_{BSS}(Q)$, is defined as the sum of the squares of these smallest k singular values:

$$\mathcal{L}_{BSS}(Q) = \sum_{i=1}^{k}\sigma_{-i}^2,$$

where σ_{-i} represents the i-th smallest singular value, and k is the number of singular values to be penalized. The overall loss function for our training process is formulated as:

$$\begin{aligned}\mathcal{L}_{total} =& \lambda\mathcal{L}_{seg}(A(X_L), Y(X_L))\\ &+ (1-\lambda)\mathcal{L}_{pseudo}(X_U, \hat{Y}(X_U))\\ &+ \eta\mathcal{L}_{BSS}(Q),\end{aligned} \quad (6)$$

where $\mathcal{L}_{seg}$ and $\mathcal{L}_{pseudo}$ are BCE loss, λ and η are hyperparameters used to balance the three terms of the loss function. By tuning these hyperparameters, we aim to optimize the model's performance while ensuring adaptability and robustness across different domains.

Experiment

Dataset and Implementation Details

Dataset and metrics. We assess our approach on widely-utilized datasets for optic disc and cup segmentation across

Method	S-F	Optic Disc Segmentation		Optic Cup Segmentation	
		Dice[%] ↑	ASSD[pixel] ↓	Dice[%] ↑	ASSD[pixel] ↓
RIM-ONE-r3					
w/o DA (Chen et al. 2021c)		83.18	24.15	74.51	14.44
Oracle (Wang et al. 2019)		96.80	–	85.60	–
BEAL (Wang et al. 2019)	×	89.80	–	81.00	–
AdvEnt (Vu et al. 2019)	×	89.73	9.84	77.99	7.57
SRDA (Bateson et al. 2020)	✓	89.37	9.91	77.61	10.15
DAE (Karani et al. 2021)	✓	89.08	11.63	79.01	10.31
DPL (Chen et al. 2021a)	✓	90.13	9.43	79.78	9.01
TT-SFDA(VS, Valanarasu, and Patel 2022)	✓	85.00	17.05	76.62	10.31
TENT(Wang et al. 2020)	✓	82.92	23.63	72.95	14.00
SFDA-FSM(Yang et al. 2022b)	✓	82.98	23.69	73.56	14.51
CBMT (Tang et al. 2023)	✓	93.36	6.20	81.16	8.37
Ours	✓	**94.76**	**4.55**	**81.53**	8.16
Drishti-GS					
w/o DA (Chen et al. 2021c)		93.84	9.05	83.36	11.39
Oracle (Wang et al. 2019)		97.40	–	90.10	–
BEAL (Wang et al. 2019)	×	96.10	–	86.20	–
AdvEnt (Vu et al. 2019)	×	96.16	4.36	82.75	11.36
SRDA (Bateson et al. 2020)	✓	96.22	4.88	80.67	13.12
DAE (Karani et al. 2021)	✓	94.04	8.79	83.11	11.56
DPL (Chen et al. 2021a)	✓	96.39	4.08	83.53	11.39
TT-SFDA(VS, Valanarasu, and Patel 2022)	✓	95.22	6.00	80.67	13.00
TENT(Wang et al. 2020)	✓	94.06	7.56	80.12	72.17
SFDA-FSM(Yang et al. 2022b)	✓	93.83	7.76	83.19	11.95
CBMT (Tang et al. 2023)	✓	96.61	3.85	84.33	10.30
Ours	✓	**96.92**	**3.48**	**86.54**	**8.78**

Table 1: The quantitative comparisons of results with different Source-Free methods on two datasets. The best scores in each column are highlighted in bold font. "-" indicates the methods without reported results. "S-F" means source-free.

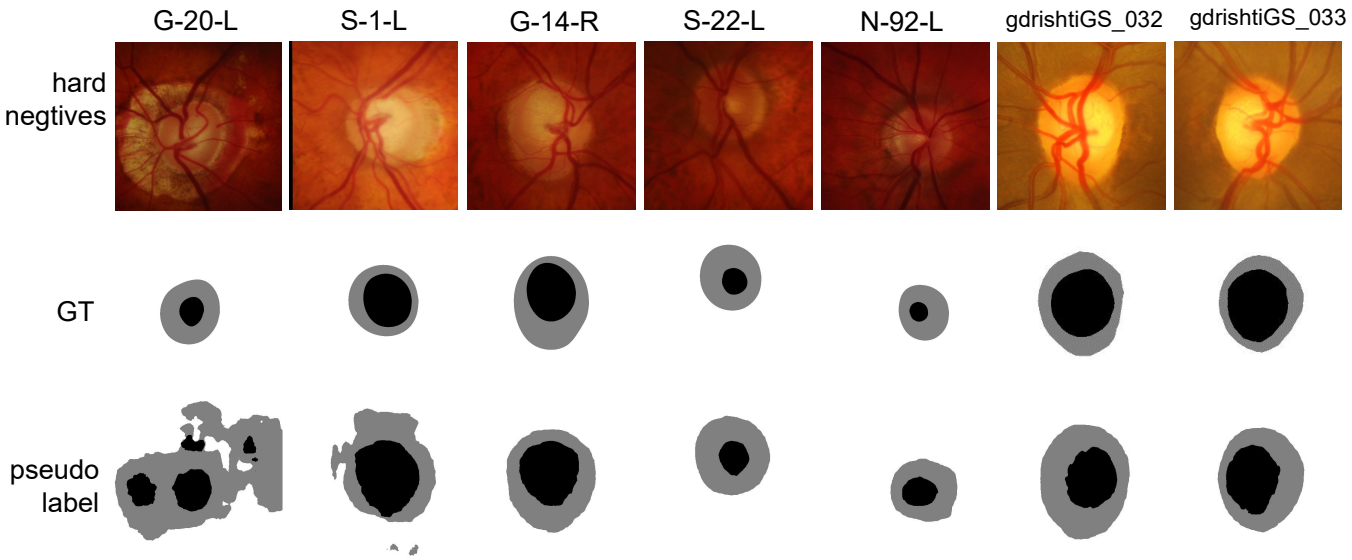


Figure 3: Visualization of selected hard negative instances with poor pseudo segmentation labels.

Method	Optic Disc Segmentation		Optic Cup Segmentation		Avg	
	Dice[%] ↑	ASSD[pixel] ↓	Dice[%] ↑	ASSD[pixel] ↓	Dice[%] ↑	ASSD[pixel] ↓
Random	**94.41**	**5.03**	80.03	8.40	87.22	6.71
SALAD(Kothandaraman et al. 2023a)	92.67	6.67	**80.78**	7.16	86.73	6.92
Ours	94.25	**5.03**	80.35	**6.94**	**87.30**	**5.99**

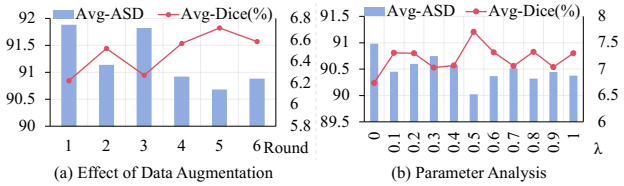
Table 2: The quantitative comparison of results on RIM-ONE-r3 under SFDA settings using different sample selection methods on the two datasets. "Random" means random selection. The best score in each column is highlighted in bold font.

Method	Compound(C)				Open(O)		Avg.			
	RIM-ONE-r3		Drishti-GS		REFUGE val		C		C+O	
	Dice	ASSD	Dice	ASSD	Dice	ASSD	Dice	ASSD	Dice	ASSD
BEAL(Wang et al. 2019)	81.08	16.42	91.10	12.49	68.43	50.14	86.09	14.45	77.26	32.30
AdvEnt(Vu et al. 2019)	80.62	11.90	87.06	9.68	64.33	30.85	83.84	10.79	74.09	20.82
OCDA(Yu et al. 2023)	81.61	13.85	87.68	9.62	82.26	21.95	84.64	11.74	83.45	16.84
DPL(Chen et al. 2021a)	84.96	9.22	89.96	7.74	78.81	11.67	87.46	8.48	83.13	10.07
TT-SFUDA(VS, Valanarasu, and Patel 2022)	80.81	39.95	87.95	9.50	80.23	22.78	84.38	24.72	82.30	23.75
TENT(Wang et al. 2020)	77.94	18.82	87.09	10.54	64.33	30.85	82.51	14.68	73.42	22.76
SFDA-FSM(Yang et al. 2022b)	78.27	19.10	88.51	9.86	79.42	8.62	83.39	14.48	81.40	11.55
PLPB(Li, Zhou, and Yang 2024)	85.42	8.30	90.04	7.56	86.42	6.83	87.73	7.93	87.07	7.38
Ours	**88.15**	**6.36**	**91.73**	**6.13**	**88.02**	**5.63**	**89.94**	**6.24**	**88.98**	**5.94**

Table 3: The quantitative comparisons of results on Open compound settings using different methods on the two datasets. The best score is highlighted in bold font.

Data Aug	BSS-loss	Avg.Dice	Avg.ASSD
×	×	90.47	7.01
✓	×	91.18	6.71
✓	✓	**91.82**	**6.14**

Table 4: Ablation study results of our proposed module on the Drishti-GS dataset.



(a) Effect of Data Augmentation (b) Parameter Analysis

Figure 4: Hyper-parameter analysis on Drishti-GS dataset.

various clinical facilities. Following previous studies, we opted for the REFUGE (Orlando et al. 2020) dataset as the source domain and fine-tuned the model for evaluation on two target domains: the RIM-ONE-r3 (Fumero et al. 2011) and DrishtiGS (Sivaswamy et al. 2015) datasets. The source domain comprises 320/80 fundus images for training/test, accompanied by pixel-wise annotations for optic disc and cup segmentation. For the RIM-ONE-r3 dataset, 99/60 images were included, with 5% (5 images) of samples selected as hard negative instances. For the DrishtiGS dataset, 50/51 images are included, with 4% (2 images) of samples selected. Following (Wang et al. 2019), the fundus images are cropped to 512×512 as Region of Interest (ROI). Common metrics for evaluating model performance in segmentation tasks are used including the Dice coefficient (Dice) and the Average Symmetric Surface Distance (ASSD). The Dice coefficient quantifies pixel-level overlap, where higher values denote superior performance. Conversely, ASSD evaluates the accuracy of prediction boundaries, with lower values indicating better performance.

Implementation details. We utilize MobileNetV2 (Sandler et al. 2018) as the network backbone, adapted from DeepLabv3+ (Chen et al. 2018), following (Wang et al. 2019; Chen et al. 2021c). The output probability threshold γ was set to 0.75. The source model was trained using the Adam optimizer with a learning rate of 2e-3. In the source-free domain adaptation step, data augmentation was carried out for 5 rounds with a hyper-parameter λ of 0.95 for the RIM-ONE-r3 and 0.5 for the DrishtiGS datasets for a total of 20 epochs with a learning rate of 5e-4. The parameter η was set to 1e-4. We implemented our method using PyTorch on an NVIDIA 3080Ti GPU.

Experiment Results

Table 1 presents the quantitative evaluation results of our method under the source-free domain adaptation (SFDA) setting. We included the "w/o adaptation" lower bound and the supervised upper bound "Oracle" in the target domain. We compared our approach with existing methods for this task, using the results without adaptation from (Chen et al. 2021c) as a lower bound and the supervised learning results from (Wang et al. 2019) as an upper bound, following (Chen et al. 2021c). Under two datasets, our method achieves significant improvements in the two numerical metrics for optic disc and optic cup segmentation compared to state-of-the-art

Algorithm 1: Supportive Negatives Spectral Augmentation.

Notation: Feature encoder $\mathcal{E}(\cdot)$ trained on source data, unlabeled target data $X_U^t = \{x_U^t\}$, source model $\mathcal{M}_s(X_U^t)$, feature of k-th class in target domain $\mathbf{f}_i^k$, centroid feature of k-th class in target domain $\mathbf{f}_p^k$, number of target inputsn_t.

Stage 1: Active Hard Negative Discovery

1. Pseudo mask generation: Generate pseudo-labels from pre-trained source model $\hat{Y}_i^k = \mathcal{M}_s(X_U^t)$, acquire segmented masks $X_i^k = X_i \odot \hat{Y}_i^k$ for the k-th class

2. Class-level target mask feature extraction:

for each target data x_U^t **do**

Extract features $\mathbf{f}_i^k = \mathcal{E}(X_i^k)$

Compute centroid $\mathbf{f}_p^k = \frac{1}{n_t}\sum_{i=1}^{n_t} \mathcal{E}(X_i^k)$

end for

3. Support hard negatives mining: Calculate the similarity of each sample's feature with the centroid feature of each class to acquire the score, rank the scores in descending order, and select the n_L with the lowest intersection k classes as hard negatives to be labeled.

Stage 2: Spectral Augmentation

for each selected hard negative **do**

for n rounds **do**

Apply augmentation function $\mathbf{Am}^{f_n}(\tilde{x}_L^n)$

end for

end for

methods. In the Drishti-GS dataset, improved performance is shown in all four evaluation metrics compared to existing methods, with a notable improvement in optic cup segmentation, where Dice and ASSD increased by 2.21% and 1.52%, respectively. On the RIM-ONE-r3 dataset, the optic disc ASSD metric decreased by 1.65 compared to previous methods, indicating better performance. These results demonstrate that calibrating those hard negative samples leads to significant improvements in adaptation results, with only a 4%∼5% increase in labeled data.

To ensure fairness, we compare our method against other sample selection strategies. By maintaining a consistent proportion of labeled data, we separately compare the performance of random selection and SALAD (Kothandaraman et al. 2023a). The results in Table 2 show that our approach outperforms competitive methods in most cases.

In-Depth Analyses

Ablation study. To verify the contribution of each module of our method, we conduct ablation studies on Drishti-GS dataset with data augmentation performed for 5 rounds. Table 4 shows the results of augmented labeled negative samples. The improved performance indicates that the expanded denoised distribution allows the model to learn the better representation. Moreover, using BSS loss allows the model to pay more attention to domain-invariant information.

Open compound study. In order to verify the generalization of our method, following Open Compound Domain Adaptation in SF-OCDA of (Li, Zhou, and Yang 2024), we conduct experiments on unseen open-domain datasets. We use the testing set of REFUGE val (Orlando et al. 2020) as the open domain, which consists of 80 images. In this study, our model was trained on the target domain, using data from an unseen domain to assess its effectiveness in an open compound setting. Results are presented in Table 3. PLPB introduces an innovative, robust pseudo-label and pseudo-boundary method to leverage unlabeled target data to generate pseudo labels and boundaries. Compared to PLPB, our approach achieves a 1.91% improvement in Dice and a 1.44% lower in ASSD on the Compound (C) and Open (C) datasets, based on experimental results using clean samples from their study. Our improvement on the open-set task demonstrates that our method effectively handles unseen data with robust generalization capabilities while maintaining superior performance at the state-of-the-art level.

Effectiveness of data augmentation. We further investigate the impact of the number of data augmentations on model performance by conducting 1-6 rounds of data augmentation on the Drishti-GS dataset. Results in Fig. 4 indicate that model performance increases along with the number of data augmentation rounds. The best performance was achieved after the fifth round.

Hyper-parameter analysis. We further investigate the impact of varying the parameter λ, which controls different loss terms. Experiments are conducted on the Drishti-GS dataset with the results shown in Figure 4 (b). Our model performs best at $\lambda = 0.5$. This suggests that, even with data augmentation, the model still requires pseudo-labeled data to supervise the model due to the limited number of true labels introduced. We need to further explore a reasonable balance between these two supervisions.

Conclusion

This study investigated and demonstrated the effectiveness of employing hard negative samples for calibration in source-free domain adaptation applied to medical image segmentation. A novel active hard negative discovery methodology is designed for identifying these hard negatives and an innovative spectral augmentation technique is developed to enrich the training dataset with a greater variety of data while minimizing the introduction of noise. Experimental results showed that incorporating a small number of hard negatives improves our model's proficiency in segmenting fundus images in both open and compound settings, underscoring their critical role in enhancing model adaptation.

Acknowledgements

This work was supported by National Natural Science Foundation of China (No. 62406068), the Natural Science Foundation of Jiangsu Province (No. BK20241299), U.S. NSF (No. 2144772) and Research Fund for Advanced Ocean Institute of Southeast University, Nantong (No. GP20010002).

References

Bateson, M.; Kervadec, H.; Dolz, J.; Lombaert, H.; and Ben Ayed, I. 2020. 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*, 490–499. Springer.

Chen, C.; Liu, Q.; Jin, Y.; Dou, Q.; and Heng, P.-A. 2021a. Source-free domain adaptive fundus image segmentation with denoised pseudo-labeling. In *Medical Image Computing and Computer Assisted Intervention–MICCAI 2021: 24th International Conference, Strasbourg, France, September 27–October 1, 2021, Proceedings, Part V 24*, 225–235. Springer.

Chen, C.; Liu, Q.; Jin, Y.; Dou, Q.; and Heng, P.-A. 2021b. Source-free domain adaptive fundus image segmentation with denoised pseudo-labeling. In *Medical Image Computing and Computer Assisted Intervention–MICCAI 2021: 24th International Conference, Strasbourg, France, September 27–October 1, 2021, Proceedings, Part V 24*, 225–235. Springer.

Chen, C.; Liu, Q.; Jin, Y.; Dou, Q.; and Heng, P.-A. 2021c. Source-free domain adaptive fundus image segmentation with denoised pseudo-labeling. In *Medical Image Computing and Computer Assisted Intervention–MICCAI 2021: 24th International Conference, Strasbourg, France, September 27–October 1, 2021, Proceedings, Part V 24*, 225–235. Springer.

Chen, L.-C.; Zhu, Y.; Papandreou, G.; Schroff, F.; and Adam, H. 2018. Encoder-decoder with atrous separable convolution for semantic image segmentation. In *Proceedings of the European conference on computer vision (ECCV)*, 801–818.

Chen, X.; Wang, S.; Fu, B.; Long, M.; and Wang, J. 2019. Catastrophic forgetting meets negative transfer: Batch spectral shrinkage for safe transfer learning. *Advances in Neural Information Processing Systems*, 32.

Cohn, D. A.; Ghahramani, Z.; and Jordan, M. I. 1996. Active learning with statistical models. *Journal of artificial intelligence research*, 4: 129–145.

Ding, N.; Xu, Y.; Tang, Y.; Xu, C.; Wang, Y.; and Tao, D. 2022. Source-free domain adaptation via distribution estimation. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, 7212–7222.

Fumero, F.; Alayón, S.; Sanchez, J. L.; Sigut, J.; and Gonzalez-Hernandez, M. 2011. RIM-ONE: An open retinal image database for optic nerve evaluation. In *2011 24th international symposium on computer-based medical systems (CBMS)*, 1–6. IEEE.

Huai, Z.; Ding, X.; Li, Y.; and Li, X. 2023. Context-Aware Pseudo-label Refinement for Source-Free Domain Adaptive Fundus Image Segmentation. In *International Conference on Medical Image Computing and Computer-Assisted Intervention*, 618–628. Springer.

Ju, L.; Wang, X.; Zhao, X.; Bonnington, P.; Drummond, T.; and Ge, Z. 2021. Leveraging regular fundus images for training UWF fundus diagnosis models via adversarial learning and pseudo-labeling. *IEEE Transactions on Medical Imaging*, 40(10): 2911–2925.

Karani, N.; Erdil, E.; Chaitanya, K.; and Konukoglu, E. 2021. Test-time adaptable neural networks for robust medical image segmentation. *Medical Image Analysis*, 68: 101907.

Kothandaraman, D.; Shekhar, S.; Sancheti, A.; Ghuhan, M.; Shukla, T.; and Manocha, D. 2023a. SALAD: Source-free Active Label-Agnostic Domain Adaptation for Classification, Segmentation and Detection. In *Proceedings of the IEEE/CVF Winter Conference on Applications of Computer Vision*, 382–391.

Kothandaraman, D.; Shekhar, S.; Sancheti, A.; Ghuhan, M.; Shukla, T.; and Manocha, D. 2023b. SALAD: Source-free Active Label-Agnostic Domain Adaptation for Classification, Segmentation and Detection. In *Proceedings of the IEEE/CVF Winter Conference on Applications of Computer Vision*, 382–391.

Li, L.; Zhou, Y.; and Yang, G. 2024. Robust source-free domain adaptation for fundus image segmentation. In *Proceedings of the IEEE/CVF Winter Conference on Applications of Computer Vision*, 7840–7849.

Liang, J.; Hu, D.; and Feng, J. 2020. Do we really need to access the source data? source hypothesis transfer for unsupervised domain adaptation. In *International conference on machine learning*, 6028–6039. PMLR.

Milletari, F.; Navab, N.; and Ahmadi, S.-A. 2016. V-net: Fully convolutional neural networks for volumetric medical image segmentation. In *2016 fourth international conference on 3D vision (3DV)*, 565–571. Ieee.

Ning, M.; Lu, D.; Wei, D.; Bian, C.; Yuan, C.; Yu, S.; Ma, K.; and Zheng, Y. 2021. Multi-anchor active domain adaptation for semantic segmentation. In *Proceedings of the IEEE/CVF International Conference on Computer Vision*, 9112–9122.

Orlando, J. I.; Fu, H.; Breda, J. B.; Van Keer, K.; Bathula, D. R.; Diaz-Pinto, A.; Fang, R.; Heng, P.-A.; Kim, J.; Lee, J.; et al. 2020. Refuge challenge: A unified framework for evaluating automated methods for glaucoma assessment from fundus photographs. *Medical image analysis*, 59: 101570.

Ronneberger, O.; Fischer, P.; and Brox, T. 2015. U-net: Convolutional networks for biomedical image segmentation. In *Medical Image Computing and Computer-Assisted Intervention–MICCAI 2015: 18th International Conference, Munich, Germany, October 5-9, 2015, Proceedings, Part III 18*, 234–241. Springer.

Sandler, M.; Howard, A.; Zhu, M.; Zhmoginov, A.; and Chen, L.-C. 2018. Mobilenetv2: Inverted residuals and linear bottlenecks. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, 4510–4520.

Sivaswamy, J.; Krishnadas, S.; Chakravarty, A.; Joshi, G.; Tabish, A. S.; et al. 2015. A comprehensive retinal image dataset for the assessment of glaucoma from the optic nerve head analysis. *JSM Biomedical Imaging Data Papers*, 2(1): 1004.

Tang, L.; Li, K.; He, C.; Zhang, Y.; and Li, X. 2023. Source-free domain adaptive fundus image segmentation with class-balanced mean teacher. In *International Conference on Medical Image Computing and Computer-Assisted Intervention*, 684–694. Springer.

Varsavsky, T.; Orbes-Arteaga, M.; Sudre, C. H.; Graham, M. S.; Nachev, P.; and Cardoso, M. J. 2020. Test-time unsupervised domain adaptation. In *Medical Image Computing and Computer Assisted Intervention–MICCAI 2020: 23rd International Conference, Lima, Peru, October 4–8, 2020, Proceedings, Part I 23*, 428–436. Springer.

VS, V.; Valanarasu, J. M. J.; and Patel, V. M. 2022. Target and task specific source-free domain adaptive image segmentation. *arXiv preprint arXiv:2203.15792*.

Vu, T.-H.; Jain, H.; Bucher, M.; Cord, M.; and Pérez, P. 2019. Advent: Adversarial entropy minimization for domain adaptation in semantic segmentation. In *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*, 2517–2526.

Wang, D.; Shelhamer, E.; Liu, S.; Olshausen, B.; and Darrell, T. 2020. Tent: Fully test-time adaptation by entropy minimization. *arXiv preprint arXiv:2006.10726*.

Wang, F.; Han, Z.; Zhang, Z.; He, R.; and Yin, Y. 2023. Mhpl: Minimum happy points learning for active source free domain adaptation. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, 20008–20018.

Wang, S.; Yu, L.; Li, K.; Yang, X.; Fu, C.-W.; and Heng, P.-A. 2019. Boundary and entropy-driven adversarial learning for fundus image segmentation. In *Medical Image Computing and Computer Assisted Intervention–MICCAI 2019: 22nd International Conference, Shenzhen, China, October 13–17, 2019, Proceedings, Part I 22*, 102–110. Springer.

Wu, J.; Wang, G.; Gu, R.; Lu, T.; Chen, Y.; Zhu, W.; Vercauteren, T.; Ourselin, S.; and Zhang, S. 2023. UPL-SFDA: Uncertainty-aware Pseudo Label Guided Source-Free Domain Adaptation for Medical Image Segmentation. *IEEE transactions on medical imaging*.

Xia, H.; and Ding, Z. 2020. Structure preserving generative cross-domain learning. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, 4364–4373.

Xia, H.; Xia, S.; and Ding, Z. 2024. Discriminative Pattern Calibration Mechanism for Source-Free Domain Adaptation. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, 23648–23658.

Xia, H.; Zhao, H.; and Ding, Z. 2021. Adaptive adversarial network for source-free domain adaptation. In *Proceedings of the IEEE/CVF international conference on computer vision*, 9010–9019.

Xu, Q.; Zhang, R.; Zhang, Y.; Wang, Y.; and Tian, Q. 2021. A fourier-based framework for domain generalization. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, 14383–14392.

Yang, C.; Guo, X.; Chen, Z.; and Yuan, Y. 2022a. Source free domain adaptation for medical image segmentation with fourier style mining. *Medical Image Analysis*, 79: 102457.

Yang, C.; Guo, X.; Chen, Z.; and Yuan, Y. 2022b. Source free domain adaptation for medical image segmentation with fourier style mining. *Medical Image Analysis*, 79: 102457.

Yu, Q.; Xi, N.; Yuan, J.; Zhou, Z.; Dang, K.; and Ding, X. 2023. Source-Free Domain Adaptation for Medical Image Segmentation via Prototype-Anchored Feature Alignment and Contrastive Learning. In *International Conference on Medical Image Computing and Computer-Assisted Intervention*, 3–12. Springer.

Zhang, Q.; Zhang, J.; Liu, W.; and Tao, D. 2019. Category anchor-guided unsupervised domain adaptation for semantic segmentation. *Advances in neural information processing systems*, 32.

Zhang, W.; Zhu, L.; Hallinan, J.; Zhang, S.; Makmur, A.; Cai, Q.; and Ooi, B. C. 2022. Boostmis: Boosting medical image semi-supervised learning with adaptive pseudo labeling and informative active annotation. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, 20666–20676.

Zhao, Y.; Zhong, Z.; Luo, Z.; Lee, G. H.; and Sebe, N. 2022. Source-free open compound domain adaptation in semantic segmentation. *IEEE Transactions on Circuits and Systems for Video Technology*, 32(10): 7019–7032.