Breaking Confirmation Bias: Single-Round Active Manifold Calibration for Source-Free Domain Adaptation in Segmentation

Anonymous ECCV 2026 Submission

Paper ID #10005

Abstract. Driven by privacy regulations, Source-Free Domain Adaptation (SFDA) has emerged as a practical paradigm for medical image segmentation. However, in the presence of severe domain shifts and extreme class imbalance, unsupervised SFDA methods are highly prone to confirmation bias and feature entanglement. While Active Learning (AL) mitigates these issues via expert intervention, existing methods suffer from inaccurate sample selection or rely on clinically impractical multi-round queries. To this end, we propose **Adapt-Label-Adapt ($\mathbf{A^2L}$)**, a single-round active SFDA framework. Unlike direct sampling, A^2L first structures the target manifold via Unsupervised Prototypical Alignment (UPA), thereby inherently reducing sampling bias. This structured manifold enables our Prototypical-Aware Uncertainty Herding (PAUH) strategy to identify a representative sample subset for one-time annotation. These queried samples then serve as reliable centroids to drive Label-Guided Manifold Calibration (LGMC). By reshaping the feature space around these definitive anchors, A^2L globally calibrates the manifold and rectifies confirmation bias. Extensive experiments on fundus segmentation benchmarks demonstrate that A^2L achieves state-of-the-art performance under an extremely low annotation budget.

Keywords: Source-Free Domain Adaptation · Active Learning · Contrastive Learning · Medical Image Analysis

1 Introduction

Medical image segmentation is crucial for computer-aided diagnosis [1, 16, 56]. In fundus image analysis, existing methods [19, 28, 32, 36] show promising results, but often struggle with limited cross-domain generalization due to domain shifts caused by variations in imaging parameters like illumination and resolution.

To mitigate these challenges, Unsupervised Domain Adaptation (UDA) has emerged as an effective paradigm [46, 55]. By minimizing domain discrepancy across heterogeneous data distributions, UDA methods [14, 29, 34, 39] enable joint learning from labeled source data and unlabeled target images. Nevertheless, their practical adoption in clinical settings remains limited due to stringent privacy regulations and ethical constraints. In many real-world medical scenarios, labeled source datasets are highly sensitive and restricted by institutional policies, preventing data sharing or cross-institutional access.

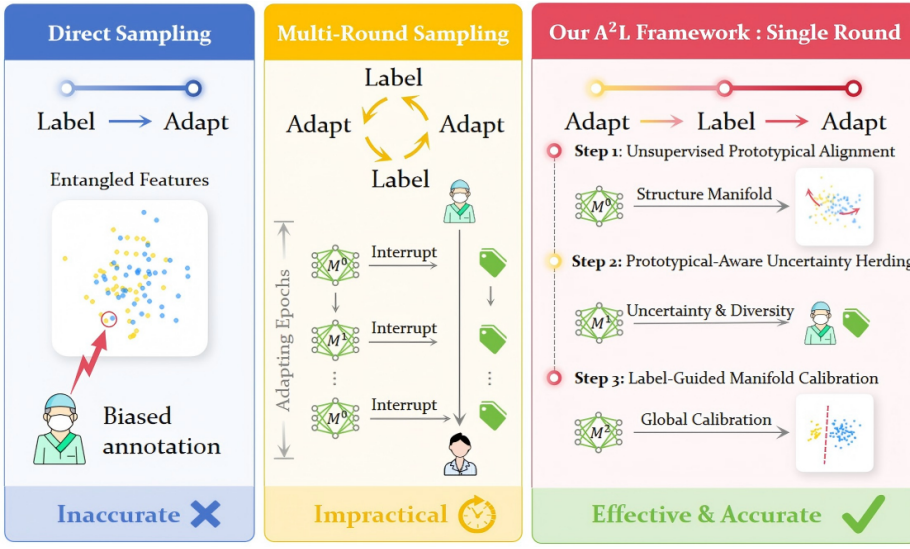


Fig. 1: Comparison of Active SFDA paradigms. Left: Direct sampling leads to biased annotations. **Middle:** Multi-round sampling is clinically impractical. **Right:** A^{2}L performs single-round adaptation via UPA (manifold structuring), PAUH (informative sample selection), and LGMC (global feature calibration).

Source-Free Domain Adaptation (SFDA) is a popular method for adapting a pretrained source model to the target domain using only unlabeled target data [10, 23]. SFDA has recently been applied to fundus image segmentation. Many methods use entropy minimization for target domain adaptation, but they are vulnerable to noise from domain shifts. To address this, Mean Teacher-based frameworks have been used to improve robustness. CBMT [45] uses consistency regularization and global loss calibration to handle label noise and class imbalance, while CCMT [61] improves stability by adjusting the teacher's update rate and aligning output distributions. Despite these improvements, SFDA still faces challenges due to the limited knowledge of the source model. This results in unreliable predictions for uncertain data, which unsupervised adaptation cannot fix. Therefore, a more effective strategy for handling informative samples is needed, motivating the use of Active SFDA.

Active SFDA methods [30, 50, 63] use uncertainty metrics to identify hard samples. However, under extreme domain shifts and class imbalance, these methods often fail during the initial query phase [11, 31, 35, 43, 44, 57]. While multi-round active learning strategies like LFTL [31] alleviate some of these issues, they are impractical for clinical settings that require a single-round process.

To address this challenge, we propose the Adapt-Label-Adapt (A^{2}L) framework, a single-round active learning approach for SFDA. As shown in Figure 1, direct sampling on entangled features leads to biased annotations, and multi-round sampling is impractical due to continuous training interruptions. A^{2}L

solves these issues with a progressive disentanglement strategy. It starts with Unsupervised Prototypical Alignment (UPA) to untangle the target manifold. On this aligned space, Prototypical-Aware Uncertainty Herding (PAUH) queries a highly informative core-set in one round. Finally, Label-Guided Manifold Calibration (LGMC) uses the minimal annotations for global adaptation. By combining unsupervised alignment with active calibration, A^2L offers a one-time solution for clinical workflows. Our contributions are summarized as follows:

1) We propose Adapt-Label-Adapt (A^2L), a single-round SFDA framework that alleviates sampling bias caused by domain shifts through Unsupervised Prototypical Alignment (UPA) and identifies hard samples for active queries.

2) We propose Prototypical-Aware Uncertainty Herding (PAUH), which incorporates class frequency weights and geometric coverage to overcome inaccurate sampling due to domain shifts and class imbalance.

3) We propose Label-Guided Manifold Calibration (LGMC), which uses minimal expert annotations to globally calibrate the feature space and disentangle representations, achieving state-of-the-art performance.

2 Related Work

Source-Free Domain Adaptation. SFDA aims to achieve cross-domain feature transfer relying solely on pre-trained source models and unlabeled target domain data, thereby effectively circumventing data leakage. This paradigm first achieved significant progress in natural image segmentation tasks [22, 27, 59, 62]. Subsequently, it was rapidly introduced into the field of medical image segmentation [4, 20, 58, 60], which is constrained by strict privacy regulations. Among various medical modalities, fundus image segmentation imposes higher demands on SFDA algorithms due to extreme class imbalance and boundary ambiguity. Existing purely unsupervised SFDA methods for fundus image segmentation [6,45,61,63] rely on pseudo-label denoising or Mean Teacher architectures to tackle these challenges. However, limited by the knowledge blind spots of source models, these unsupervised methods are prone to generating high-confidence yet erroneous pseudo-labels on long-tail hard samples with vast morphological variations, thereby falling into confirmation bias. Furthermore, simple self-supervised clustering or denoising mechanisms struggle to fundamentally disentangle feature representations at extremely blurred boundaries.

Active Learning for SFDA. Active Learning (AL) seeks to maximize model performance at minimal cost by heuristically querying the most informative unlabeled samples for expert annotation under a strict budget [9, 13, 38, 40]. To overcome the confirmation bias of unsupervised SFDA, AL is integrated into this framework, termed Active SFDA. Early attempts primarily relied on local uncertainty or feature perturbations for sampling [2,25,26,50]. As this paradigm extends to challenging dense prediction tasks, tackling extreme domain shifts and improving clinical deployment efficiency become pivotal. Consequently, recent studies focus on precise alignment in complex regions, such as the dual-reference sampling in STDR [51] and supportive negative sampling in SNSA [63].

Furthermore, Lyu et al. [31] revealed the long-tail error manifold, where model errors concentrate on scarce hard samples, demonstrating superior feasibility of single-round AL over multi-round interactions. However, existing Active SFDA methods mostly based solely on uncertainty or outlier distances. By neglecting local density and global topological diversity, they easily incur sampling redundancy. Moreover, merely applying simple cross-entropy fine-tuning after obtaining labels fails to explicitly disentangle features in ambiguous regions.

Contrastive Learning. Contrastive learning has revolutionized self-supervised learning by pulling positive pairs closer and pushing negative pairs apart in the representation space [8, 15, 18]. This classic image-level paradigm heavily relies on tailored data augmentations to construct discriminative features. However, since dense prediction tasks demand high-resolution spatial localization and fine-grained semantic classification, traditional image-level contrast often loses critical local context. Consequently, researchers have shifted contrastive mechanisms from the image level to the pixel or region level. Notably, Wang et al. [53] pioneered Cross-Image Pixel Contrast in semantic segmentation, achieving fine-grained alignment from a global perspective by constructing pixel-wise positive and negative pairs across different images. Recently, Dense Contrastive and Prototypical Contrastive approaches have been widely applied to medical image segmentation [3, 5, 54, 60]. For instance, Basak et al. [3] demonstrated the indispensability of contrastive learning for extracting minute lesion features in few-shot medical segmentation, while Yu et al. [60] combined prototype anchors with contrastive learning to explore cross-domain feature alignment. Although these methods significantly enhance representation discriminability, under extreme domain shifts of SFDA [42, 62], the lack of reliable supervision makes models highly prone to forcibly aligning noisy pseudo-labels as pseudo-positives.

3 Theoretic Motivation

In medical image analysis, adapting a model to a target domain faces severe covariate shift and class imbalance. In this section, we highlight the main mathematical challenges of applying active learning under these conditions. By identifying the limitations of standard methods in prototype corruption and topological collapse, we motivate the need for our Adapt-Label-Adapt ($\mathrm{A^2L}$) framework. Due to space constraints, detailed proofs are provided in the Appendix.

3.1 Prototype Corruption

The objective of AL is to query a minimal subset of unlabeled data S^* that maximizes a utility function $A(S)$ balancing informativeness and diversity:

$$A(S) = \sum_{x_i \in S} U(x_i, \{\mu_c\}) - \gamma \sum_{x_i, x_j \in S} K(z_i, z_j), \tag{1}$$

where U measures uncertainty with respect to class prototypes $\{\mu_c\}$ and K promotes diversity. In SFDA, target labels are strictly unavailable, so the uncertainty metric U must rely on feature distances to estimated prototypes derived

from the frozen source model's pseudo-labels. Under severe domain shift, source boundaries overlap with high-density target regions, leading to pseudo-label corruption and prototype bias. To quantify this, we decompose the prototype estimation error into stochastic sampling error and structural bias:

Proposition 1 (Prototype Estimation Error Bound). *Let S_c be the subset of target samples with pseudo-label c and $n_c = |S_c|$. Define the oracle centroid $\mu_c^* := \mathbb{E}_{(x,y)\sim D_T}[\, f(x) \mid y = c]$, and the empirical prototype $\hat{\mu}_c := \frac{1}{n_c}\sum_{x_i \in S_c} f(x_i)$. Further define the misclassification mean $\mu_{err} := \mathbb{E}[\, f(x) \mid y \neq c, \hat{y} = c]$, the false discovery rate $\alpha_c := \mathbb{P}(y \neq c \mid \hat{y} = c)$, and the selection bias vector $\beta_c := \mathbb{E}[\, f(x) \mid y = c, \hat{y} = c] - \mu_c^*$. Then, with probability at least $1 - \delta$, the estimation error satisfies*

$$\|\hat{\mu}_c - \mu_c^*\|_2 \leq M \sqrt{\frac{2d \log(2d/\delta)}{n_c}} + (1 - \alpha_c)\|\beta_c\|_2 + \alpha_c \|\mu_{err} - \mu_c^*\|_2,$$

where the first term reflects sampling variability and the remainder captures structural bias.

This decomposition reveals a deadlock, i.e., while the stochastic sampling error diminishes with larger n_c, the structural bias component remains irreducible whenever $\alpha_c > 0$. This residual bias distorts uncertainty estimates and undermines active sampling.

While Proposition 1 bounds the stochastic error using the worst-case feature norm M, in practice, this sampling variability is directly governed by the empirical intra-class variance, denoted as σ_c^2. Compressing this variance effectively tightens the practical estimation error.

Lemma 1 (Vulnerability of Query Guarantee).
Assume the active learning utility function $U(x, \mu)$ is L-Lipschitz continuous w.r.t. the prototype μ. Let $\Delta := U(x_a, \mu^*) - U(x_n, \mu^*) > 0$ be the true utility margin between a structural anchor x_a and out-of-distribution noise x_n. If the empirical prototype $\hat{\mu}$ suffers from structural bias $B_c = \|\hat{\mu} - \mu^*\|_2 > \frac{\Delta}{2L}$, then there exists a feasible adversarial configuration where the theoretical guarantee to prioritize x_a over x_n is compromised, i.e.,

$$U(x_n, \hat{\mu}) > U(x_a, \hat{\mu}).$$

Lemma 1 implies that directly applying active learning on an uncalibrated source model can bias sampling toward irrelevant noise rather than meaningful boundary structure. To restore query purity ($B_c \to 0$), we introduce Unsupervised Prototypical Alignment (UPA), which uses high-confidence masking to drive $\alpha_c \to 0$ and self-contrastive refinement to compress intra-class variance.

3.2 Topological Collapse

Even if structural bias B_c is reduced through UPA, sampling strategies that focus purely on local uncertainty can misrepresent the global topological structure

of the target distribution. Uncertainty sampling emphasizes boundary instances, while space-covering strategies like Core-set sampling risk wasting the query budget on large, uninformative background regions, especially under extreme class imbalance. In both cases, the distribution of queried samples $\hat{P}_Q$ deviates from the true target distribution D_T. This distributional mismatch impacts generalization, which we bound using the 1-Wasserstein distance between distributions.

Proposition 2 (Generalization Bound via Optimal Transport). *Let $h_Q \in H$ be a hypothesis trained on a queried subset S_Q of size n drawn i.i.d. from the query distribution P_Q. Let D_T be the overall target distribution. Assume covariate shift: $P(y|x)$ is identical across D_T and P_Q. Assume the loss $\ell(y', y)$ is bounded by M_ℓ and is L_ℓ-Lipschitz w.r.t. its prediction argument, and each $h \in H$ is L_h-Lipschitz w.r.t. x. Then with probability at least $1 - \delta$,*

$$R_{D_T}(h_Q) \leq \widehat{R}_Q(h_Q) + 2J_{S_Q}(H) + 3M_\ell\sqrt{\frac{\log(2/\delta)}{2n}} + L_\ell L_h \cdot W_1(D_T^X, P_Q^X), \quad (2)$$

where $\widehat{R}_Q(h_Q)$ is the empirical risk, $J_{S_Q}(H)$ is the Rademacher complexity of H, and $W_1(D_T^X, P_Q^X)$ is the 1-Wasserstein distance between the marginal input distributions.

The Wasserstein term in the above bound penalizes distributional mismatch, reflecting the geometric cost of transporting mass between D_T and $\hat{P}_Q$. When acquisition strategies focus on local uncertainty or sparse coverage, $W_1(D_T, \hat{P}_Q)$ could be increased, leading to weakened generalization guarantees. To control this penalty without succumbing to the background-domination flaw of vanilla Core-sets, the sampling strategy must be topology-aware and class-calibrated. Motivated by this bound, we propose Prototypical-Aware Uncertainty Herding (PAUH), which excludes background noise, applies class frequency weighting, and centers queries around UPA-disentangled prototypes. This ensures the sampled subset preserves the global geometry of D_T while controlling the Wasserstein penalty, promoting balanced and robust target generalization.

4 Methodology

Guided by our theoretical analysis, the Adapt-Label-Adapt (A^2L) framework addresses active domain adaptation bottlenecks through a three-stage pipeline. As illustrated in Fig. 2. A^2L includes: (1) UPA to reduce feature variance and decouple bias; (2) PAUH for single-round target distribution alignment; (3) LGMC to adjust the target manifold using a minimal query budget.

For these stages, we use a weak-strong augmented Mean Teacher architecture [45,61] as the feature extractor. During unsupervised adaptation, the teacher network generates pseudo-labels $\hat{y}$ to guide the student. To avoid manifold distortions from noisy labels, the teacher is updated via an Exponential Moving Average (EMA) of the student's parameters. Let $z_i \in \mathbb{R}^d$ represent the feature embedding of pixel i from the student network. A detailed summary of the framework is provided in the Appendix.

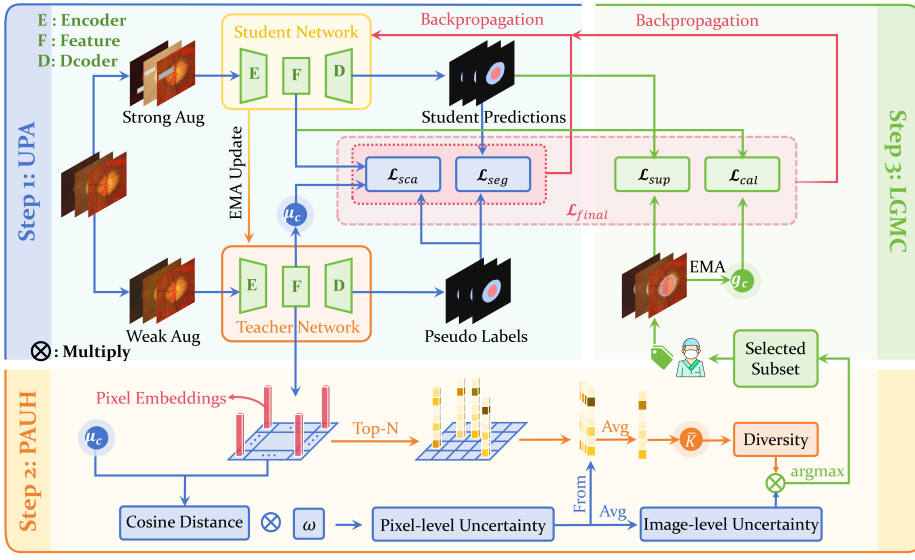


Fig. 2: Overview of the A^{2}L framework. UPA generates pseudo-centroids to reduce structural bias, PAUH selects a core subset via pixel uncertainty and geometric diversity, and LGMC uses annotated samples to calibrate the target feature manifold.

4.1 Unsupervised Prototypical Alignment (UPA)

Directly sampling from the raw target space is flawed due to the structural bias (B_c) from pseudo-label false discovery ($\alpha_c > 0$), as shown in Proposition 1. To restore query purity, UPA constructs a global pseudo-centroid μ_c for each class $c \in C$ using the teacher's stable representations. To reduce structural bias ($B_c \to 0$), we apply a masking strategy that filters uncertain predictions. The centroid update follows an EMA formula:

$$\mu_c^{(t)} = \alpha \mu_c^{(t-1)} + (1-\alpha)\frac{1}{|M^{(c)}|}\sum_{j \in M^{(c)}} z_j, \tag{3}$$

where $M^{(c)}$ is a dynamic high-confidence mask and α is the momentum of the exponential moving average. This ensures that only reliable samples contribute to μ_c, decoupling prototypes from misclassified noise.

While masking mitigates structural bias, the stochastic error (σ_c^2) remains. To compress this intra-class variance ($\sigma_c^2 \to 0$) and enforce inter-class uniformity, we optimize the student via a self-contrastive loss $\mathcal{L}_{sca}$ [53]:

$$\mathcal{L}_{sca} = -\frac{1}{|\Omega|}\sum_{j \in \Omega} \log \frac{\exp(\text{sim}(z_j, \mu_{\hat{y}_j})/\tau)}{\sum_{k \in C} \exp(\text{sim}(z_j, \mu_k)/\tau)}. \tag{4}$$

where Ω is the spatial pixel set, sim denotes cosine similarity, and τ is the temperature hyperparameter.

The warm-up objective integrates this contrastive alignment with the segmentation loss $\mathcal{L}_{seg}$ (Cross-Entropy and Dice losses):

$$\mathcal{L}_{warmup} = \mathcal{L}_{seg} + \beta \mathcal{L}_{sca}, \tag{5}$$

where $\mathcal{L}_{seg}$ is guided by the teacher's pseudo-labels, and β is a balancing factor. This phase disentangles the target space, enabling unbiased sampling.

4.2 Prototypical-Aware Uncertainty Herding (PAUH)

By greedily sampling boundaries, standard AL exacerbates distribution mismatch and violates generalization bounds. In fundus segmentation, background-dominated uncertainty further biases queries against minority structures.

To evaluate structural value without background interference, we propose PAUH. We compute pixel-level uncertainty $u_{i,p}$ for each pixel $p \in \Omega_i$ (the spatial pixel set of image X_i) based on its cosine divergence from the UPA-calibrated centroid:

$$u_{i,p} = \omega_{\hat{y}_{i,p}} \cdot \big(1 - \mathrm{sim}(z_{i,p}, \mu_{\hat{y}_{i,p}})\big), \tag{6}$$

where $\omega_{\hat{y}_{i,p}} = 1/\sqrt{f_c}$ prioritizes minority classes based on their global pixel frequency f_c. Instead of aggregating over the whole image, we focus on localized bottlenecks. For each foreground class $c \in C_{fg}$ in X_i, we collect the top-N most uncertain pixels to form a hard subset $H_{i,c}$, and compute its feature center $v_{i,c}$ as:

$$v_{i,c} = \frac{1}{|H_{i,c}|} \sum_{p \in H_{i,c}} z_{i,p}. \tag{7}$$

The image-level uncertainty $U(X_i)$ is then the macro-average over these hard subsets:

$$U(X_i) = \frac{1}{|C_{fg}|} \sum_{c \in C_{fg}} \left(\frac{1}{|H_{i,c}|} \sum_{p \in H_{i,c}} u_{i,p} \right). \tag{8}$$

To enforce geometric diversity, we measure image-to-image similarity by aligning foreground centers via a similarity kernel $K(\cdot, \cdot)$:

$$\bar{K}(X_i, X_j) = \frac{1}{|C_{fg}|} \sum_{c \in C_{fg}} K(v_{i,c}, v_{j,c}). \tag{9}$$

We use a greedy herding objective to iteratively select the active subset S_t from the unlabeled target pool D_t:

$$\begin{aligned} V(X_i) &= \sum_{X_j \in D_t} \bar{K}(X_i, X_j) - \sum_{X_s \in S_t} \bar{K}(X_i, X_s), \\ X_{t+1} &= \arg \max_{X_i \in D_t \setminus S_t} \big[U(X_i) \cdot V(X_i)\big]. \end{aligned} \tag{10}$$

Here, $V(X_i)$ quantifies the geometric diversity gain. Multiplying $U(X_i)$ and $V(X_i)$ ensures selected samples are both highly uncertain and structurally representative, stopping when the budget $|S_t| = B$ is reached.

4.3 Label-Guided Manifold Calibration (LGMC)

While UPA and PAUH successfully sample an unbiased core-set, unsupervised contrastive alignment can cause blind repulsion, tearing apart semantically similar instances due to the lack of ground truth. Once the subset S^* is annotated with labels Y_{S^*}, we introduce LGMC as a supervised intervention to correct the target manifold.

We improve the metric space by replacing noisy pseudo-centroids (μ_c) with reliable, label-guided anchors g_c. To ensure target stability, we compute these anchors using an EMA over the newly annotated foreground pixels:

$$g_c^{(t)} = \text{EMA}\left(\frac{1}{|\Omega_c^*|}\sum_{X\in S^*}\sum_{j\in\Omega_c^*} z_j\right), \tag{11}$$

where Ω_c^* is the set of labeled pixels for class c. Using these anchors, we apply a supervised contrastive loss $\mathcal{L}_{cal}$ to bind the representations to their true centers:

$$\mathcal{L}_{cal} = -\frac{1}{|\Omega^*|}\sum_{j\in\Omega^*}\log\frac{\exp\left(s_{j,y_j}/\tau\right)}{\exp\left(s_{j,y_j}/\tau\right)+\sum_{k\neq y_j}\exp\left(s_{j,k}/\tau\right)}. \tag{12}$$

where $\Omega^* = \bigcup_{c\in C}\Omega_c^*$ denotes the set of all annotated pixels across all classes in the active subset S^*, y_j is the ground-truth label for pixel j, and $s_{j,k} = \text{sim}(z_j, g_k)$ represents the cosine similarity between the pixel feature z_j and the label-guided anchor g_k.

To prevent overfitting to the sparse annotations of S^* (e.g., $4\% \sim 5\%$ budget), we continue to use the geometry of the unlabeled data. Thus, we preserve the teacher's unsupervised pseudo-segmentation ($\mathcal{L}_{seg}$) and the original prototypical alignment ($\mathcal{L}_{sca}$) on the unlabeled target set ($D_t \setminus S^*$). The final objective is:

$$\mathcal{L}_{final} = \mathcal{L}_{sup}(S^*) + \mathcal{L}_{seg}(D_t \setminus S^*) + \beta\mathcal{L}_{sca}(D_t) + \lambda\mathcal{L}_{cal}(S^*), \tag{13}$$

where $\mathcal{L}_{sup}$ provides standard supervision on S^*, and λ controls calibration strength. This formulation ensures label-guided intervention is balanced with continuous unsupervised manifold maintenance.

5 Experiment

5.1 Dataset and Implementation Details

Datasets and Evaluation Metrics. We evaluate A^{2}L on standard optic disc and cup segmentation benchmarks. Following cross-domain protocols, we use REFUGE [33] (320 train/80 test images) as the fully annotated source. Adaptation is evaluated on two target domains: RIM-ONE-r3 [12] (99 train/60 test) and Drishti-GS [41] (50 train/51 test). For our active SFDA setting, the annotation budget B is strictly limited to 5% (5 images) for RIM-ONE-r3 and 4% (2 images) for Drishti-GS. All images are cropped to a 512×512 ROI [52]. We

Table 1: Quantitative comparisons on the RIM-ONE-r3 and Drishti-GS datasets. The best and second best results are highlighted. "–" denotes methods without reported results. "S-F" indicates source-free. "±" represents the standard deviation across samples. "*" indicates active learning methods.

Method	Venue	S-F	Optic Disc Segmentation		Optic Cup Segmentation	
			Dice[%] ↑	ASSD[pixel] ↓	Dice[%] ↑	ASSD[pixel] ↓
RIM-ONE-r3						
w/o DA [6]	MICCAI'21		83.18 ± 6.46	24.15 ± 15.58	74.51 ± 16.40	14.44 ± 11.27
Oracle [52]	MICCAI'19		96.80	–	85.60	–
BEAL [52]	MICCAI'19	×	89.80	–	81.00	–
AdvEnt [48]	CVPR'19	×	89.73 ± 3.66	9.84 ± 3.86	77.99 ± 21.08	7.57 ± 4.24
SRDA [4]	MICCAI'20	✓	89.37 ± 3.66	9.91 ± 2.45	77.61 ± 13.58	10.15 ± 5.75
DAE [20]	MIA'21	✓	89.08 ± 3.32	11.63 ± 6.84	79.01 ± 12.82	10.31 ± 8.45
DPL [6]	MICCAI'21	✓	90.13 ± 3.06	9.43 ± 3.46	79.78 ± 11.05	9.01 ± 5.59
TT-SFDA [47]	PMLR'24	✓	85.00	17.05	76.62	10.31
TENT [49]	ICLR'21	✓	82.92	23.63	72.95	14.00
SFDA-FSM [58]	MIA'22	✓	82.98	23.69	73.56	14.51
CBMT [45]	MICCAI'23	✓	93.36 ± 4.07	6.20 ± 4.79	81.16 ± 14.71	8.37 ± 6.99
CCMT [61]	Neurocomputing'25	✓	95.45 ± 1.79	3.97 ± 1.56	83.71 ± 12.00	7.33 ± 4.33
SEE [17]	TPAMI'25	✓	95.28 ± 2.15	4.10 ± 1.84	84.07 ± 9.79	6.80 ± 4.20
SNSA* [63]	AAAI'25	✓	94.76	4.55	81.53	8.16
A^2L*	Ours	✓	95.53 ± 2.24	3.88 ± 1.82	85.23 ± 7.66	6.44 ± 3.68
Drishti-GS						
w/o DA [6]	MICCAI'21		93.84 ± 2.91	9.05 ± 7.50	83.36 ± 11.95	11.39 ± 6.30
Oracle [52]	MICCAI'19		97.40	–	90.10	–
BEAL [52]	MICCAI'19	×	96.10	–	86.20	–
AdvEnt [48]	CVPR'19	×	96.16 ± 1.65	4.36 ± 1.83	82.75 ± 11.08	11.36 ± 7.22
SRDA [4]	MICCAI'20	✓	96.22 ± 11.30	4.88 ± 13.47	80.67 ± 11.78	13.12 ± 16.48
DAE [20]	MIA'21	✓	94.04 ± 2.85	8.79 ± 7.45	83.11 ± 11.89	11.56 ± 6.32
DPL [6]	MICCAI'21	✓	96.39 ± 1.33	4.08 ± 1.49	83.53 ± 17.80	11.39 ± 10.18
TT-SFDA [47]	PMLR'24	✓	95.22	6.00	80.67	13.00
TENT [49]	ICLR'21	✓	94.06	7.56	80.12	13.52
SFDA-FSM [58]	MIA'22	✓	93.83	7.76	83.19	11.95
CBMT [45]	MICCAI'23	✓	96.61 ± 1.45	3.85 ± 1.63	84.33 ± 11.70	10.30 ± 5.88
CCMT [61]	Neurocomputing'25	✓	96.91 ± 1.16	3.48 ± 1.35	86.36 ± 12.24	9.03 ± 5.51
SEE [17]	TPAMI'25	✓	96.89 ± 1.12	3.52 ± 1.21	85.33 ± 11.50	9.68 ± 5.25
SNSA* [63]	AAAI'25	✓	96.92	3.48	86.54	8.78
A^2L*	Ours	✓	97.02 ± 1.11	3.35 ± 1.22	86.74 ± 12.26	8.70 ± 5.39

compare A^2L against state-of-the-art UDA [48,52], unsupervised SFDA [4,6,17, 20,45,47,49,58,61], and Active SFDA [63] frameworks. Performance is quantified using the Dice Similarity Coefficient (Dice) for spatial overlap and Average Symmetric Surface Distance (ASSD) for boundary precision.

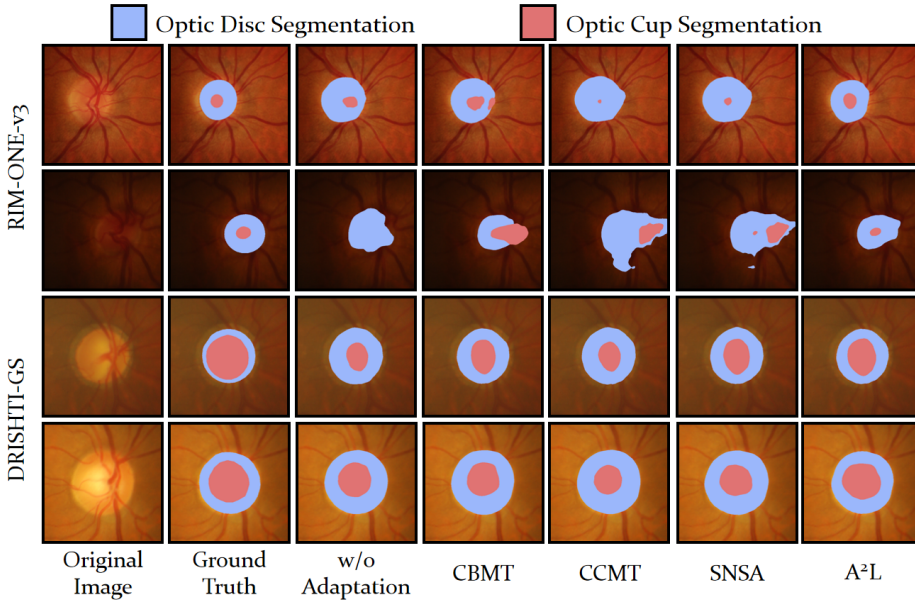


Fig. 3: Quantitative results of different methods on the RIM-ONE-r3 and Drishti-GS datasets.

Implementation details. Consistent with prior studies [6, 45, 52, 61, 63], we adopt DeepLabv3+ [7] equipped with a MobileNetV2 [37] backbone as our segmentation architecture. Optimization across all stages is performed using the Adam optimizer ($\beta_1 = 0.9$, $\beta_2 = 0.99$). For target adaptation, we employ a weak-strong augmented Mean Teacher architecture with a batch size of 8 and a learning rate of 5×10^{-4}. Specifically, the adaptation begins with an initial 10 epochs of unsupervised alignment guided by the self-contrastive loss $\mathcal{L}_{sca}$, where the EMA momentum is 0.98, $\gamma = 0.75$, temperature $\tau = 0.10$, and $\beta = 1.0$. Next, we extract the top-512 uncertain pixels per foreground class for active querying. Finally, the model is calibrated for a subsequent 20 epochs using the supervised contrastive objective $\mathcal{L}_{cal}$, utilizing a temperature of 0.10, a margin scale of 0.20, and $\lambda = 1.0$. The entire framework is implemented in PyTorch on a single NVIDIA 5090 GPU.

5.2 Experiment Results

We quantitatively evaluate the A^{2}L framework on two challenging fundus image segmentation benchmarks and compare it with UDA, SFDA, and the latest active SFDA methods. As shown in Table 1, on the RIM-ONE-r3 dataset, A^{2}L achieves Dice scores of 85.23% and 95.53% for optic cup and optic disc segmentation, respectively. Notably, its optic cup segmentation result significantly outperforms the state-of-the-art (SOTA) active SFDA method, SNSA [63] (81.53%).

Table 2: Ablation study on the core components of our proposed A^{2}L framework. AL denotes Active Learning. $\mathcal{L}_{cal}$ and $\mathcal{L}_{sca}$ represent the two variants of Contrastive Learning. The best and second best results are highlighted.

Baseline	Active	Contrastive Learning		Metrics	
		$\mathcal{L}_{sca}$	$\mathcal{L}_{cal}$	Avg. Dice[%] ↑	Avg. ASSD[pixel] ↓
✓				88.59	6.84
✓		✓		89.70	5.74
✓	✓			89.34	5.95
✓	✓		✓	89.16	6.00
✓	✓	✓		90.08	5.47
✓	✓	✓	✓	90.38	5.16

Furthermore, in terms of the ASSD metric, A^{2}L achieves the minimum boundary deviation, with ASSD values of 3.88 pixels and 6.44 pixels for optic cup and optic disc segmentation, respectively, which significantly surpasses other SOTA methods. On the Drishti-GS dataset, A^{2}L achieves Dice values of 97.02% and 86.74% for optic disc and optic cup segmentation, respectively. Compared with the unsupervised Mean Teacher architectures CBMT [45] and CCMT [61], A^{2}L effectively fills the knowledge gap of the source model in the target domain through precise single-round active intervention, thereby mitigating the confirmation bias problem caused by severe domain shift.

As shown in Figure 3, we present the qualitative segmentation results of various methods on the RIM-ONE-r3 and Drishti-GS datasets. Due to severe domain shift, the baseline model without domain adaptation exhibits significant structural distortions. Although unsupervised SFDA methods [45, 61] and the active learning method [63] improve overall localization, they are still prone to producing irregular contours in regions with blurred boundaries. In contrast, A^{2}L achieves more accurate boundary delineation and demonstrates superior segmentation performance for the optic disc and cup.

5.3 Further Analyses

Ablation Study. To evaluate the contribution of each component in A^{2}L, we report an ablation study in Table 2. The results show that incorporating AL significantly improves performance, which confirms that limited expert supervision effectively corrects pseudo-label errors in source-free adaptation. Introducing UPA ($\mathcal{L}_{sca}$) further enhances performance, which proves that compressing intra-class variance improves feature space organization and sampling efficiency. Additionally, LGMC ($\mathcal{L}_{cal}$) yields additional gains, which demonstrates that using precise anchors and class-adaptive margins enforces wider decision boundaries, redistributing and decoupling the global feature manifold.

Hyper-parameter Analysis. Figure 4 (a) and (b) illustrate the parameter sensitivity of the unsupervised alignment weight β and the label-guided calibration weight λ in Eq. (13). The results indicate that the model achieves optimal

Table 3: Quantitative evaluation of various methods under the open compound setting across the two datasets. The best and second best results are highlighted. "*" indicates active learning methods.

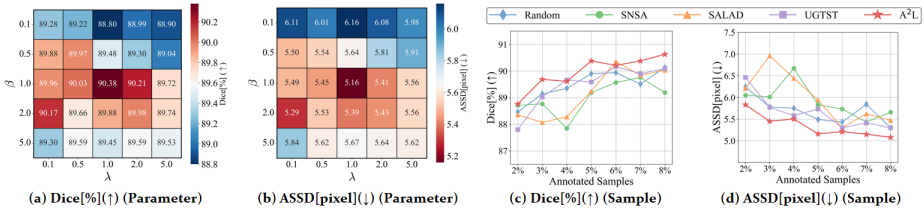
Method	Venue	Compound(C)				Open(O)		Avg.			
		RIM-ONE-r3		Drishti-GS		REFUGE val		C		C and O	
		Dice ↑	ASSD ↓	Dice ↑	ASSD ↓	Dice ↑	ASSD ↓	Dice ↑	ASSD ↓	Dice ↑	ASSD ↓
BEAL [52]	MICCAI'19	81.08	16.42	91.10	12.49	68.43	50.14	86.09	14.45	77.26	32.30
AdvEnt [48]	CVPR'19	80.62	11.90	87.06	9.68	64.33	30.85	83.84	10.79	74.09	20.82
OCDA [60]	MICCAI'23	81.61	13.85	87.68	9.62	82.26	21.95	84.64	11.74	83.45	16.84
DPL [6]	MICCAI'21	84.96	9.22	89.96	7.74	78.81	11.67	87.46	8.48	83.13	10.07
TT-SFUDA [47]	PMLR'24	80.81	39.95	87.95	9.50	80.23	22.78	84.38	24.72	82.30	23.75
TENT [49]	ICLR'21	77.94	18.82	87.09	10.54	64.33	30.85	82.51	14.68	73.42	22.76
SFDA-FSM [58]	MIA'22	78.27	19.10	88.51	9.86	79.42	8.62	83.39	14.48	81.40	11.55
PLPB [24]	WACV'24	85.42	8.30	90.04	7.56	86.42	6.83	87.73	7.93	87.07	7.38
SNSA* [63]	AAAI'25	88.15	6.36	91.73	6.13	88.02	5.63	89.94	6.24	88.98	5.94
A^{2}L*	Ours	90.38	5.16	91.88	6.03	88.64	5.19	91.13	5.60	89.89	5.40

performance on both Dice and ASSD metrics when $\beta = 1.0$ and $\lambda = 1.0$. Excessively small weights result in insufficient contrastive constraints to effectively organize the feature space, while excessively large weights interfere with the primary segmentation loss.

Open Compound Study. We evaluated the generalization of our framework on open-domain datasets, following the Open Compound Domain Adaptation protocol in SF-OCDA [24, 63]. Specifically, we used the REFUGE validation set [33] (80 images) as the open domain for testing after training on the target domain. As shown in Table 3, our approach outperforms SNSA in this open compound setting. By utilizing unlabeled data, our method achieves a 0.91% improvement in Dice and a 0.54 reduction in ASSD compared to SNSA. These results highlight our architecture's strong adaptation potential for handling previously unseen data while maintaining state-of-the-art performance.

Sample Selection Methods. To verify the efficacy of our sample selection mechanism, we conduct a comprehensive comparison against several learning strategies [21, 30, 63]. For a fair assessment, all methods are evaluated under consistent experimental settings across various labeling budgets. As illustrated in Figure 4 (c) and (d), our proposed A^{2}L framework consistently achieves superior performance over all competing strategies across all budget levels. These results validate that the performance gains of A^{2}L are not merely due to the amount of labeled data, but rather our superior ability to curate high-value samples.

Feature Visualization. To show the effectiveness of A^{2}L, we visualize the pixel feature embeddings using t-SNE in Figure 5. After adaptation, features of pixels from the same class are clustered together, while features from different classes are well-separated, confirming that the features align with our expectations.



(a) Dice[%](↑) (Parameter) (b) ASSD[pixel](↓) (Parameter) (c) Dice[%](↑) (Sample) (d) ASSD[pixel](↓) (Sample)

Fig. 4: Further analyses of the proposed A^{2}L framework. (a) and (b) illustrate the parameter sensitivity analysis for the unsupervised alignment weight β and the label-guided calibration weight λ. (c) and (d) present the performance comparison of A^{2}L against various active sampling strategies under varying annotation budgets.

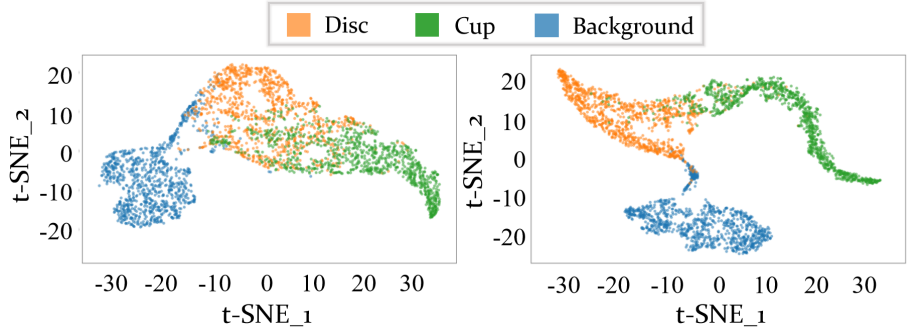


Fig. 5: t-SNE of target features. Left: Before adaptation, features are entangled with blurry boundaries. **Right:** After A^{2}L, features form compact intra-class clusters with clear decision boundaries.

6 Conclusion

To tackle domain shifts and confirmation bias in Source-Free Domain Adaptation (SFDA) for medical image segmentation, we propose Adapt-Label-Adapt (A^{2}L), an efficient single-round active framework. First, Unsupervised Prototypical Alignment (UPA) structures the target manifold, overcoming mode collapse and sampling bias. Next, Prototypical-Aware Uncertainty Herding (PAUH) selects the most representative sample subset for one-time annotation while addressing class imbalance. Finally, Label-Guided Manifold Calibration (LGMC) uses reliable anchors to globally reshape the feature space. Experiments show that A^{2}L outperforms existing methods, offering a strong new paradigm for privacy-restricted clinical image analysis.

References

1. Azad, R., Aghdam, E.K., Rauland, A., Jia, Y., Avval, A.H., Bozorgpour, A., Karimijafarbigloo, S., Cohen, J.P., Adeli, E., Merhof, D.: Medical image segmentation review: The success of u-net. IEEE Trans. Pattern Anal. Mach. Intell. **46**(12), 10076–10095 (2024)
2. Bae, W., Sutherland, D.J., Oliveira, G.L.: Uncertainty herding: One active learning method for all label budgets. In: International Conference on Learning Representations. OpenReview.net (2025)
3. Basak, H., Yin, Z.: Pseudo-label guided contrastive learning for semi-supervised medical image segmentation. In: IEEE/CVF Conference on Computer Vision and Pattern Recognition. pp. 19786–19797 (2023)
4. Bateson, M., Kervadec, H., Dolz, J., Lombaert, H., Ayed, I.B.: Source-relaxed domain adaptation for image segmentation. In: International Conference on Medical Image Computing and Computer Assisted Intervention. vol. 12261, pp. 490–499 (2020)
5. Chaitanya, K., Erdil, E., Karani, N., Konukoglu, E.: Contrastive learning of global and local features for medical image segmentation with limited annotations. In: Annual Conference on Neural Information Processing Systems (2020)
6. Chen, C., Liu, Q., Jin, Y., Dou, Q., Heng, P.: Source-free domain adaptive fundus image segmentation with denoised pseudo-labeling. In: International Conference on Medical Image Computing and Computer Assisted Intervention. vol. 12905, pp. 225–235 (2021)
7. Chen, L., Zhu, Y., Papandreou, G., Schroff, F., Adam, H.: Encoder-decoder with atrous separable convolution for semantic image segmentation. In: European Conference on Computer Vision. vol. 11211, pp. 833–851 (2018)
8. Chen, T., Kornblith, S., Norouzi, M., Hinton, G.E.: A simple framework for contrastive learning of visual representations. In: International Conference on Machine Learning. Proceedings of Machine Learning Research, vol. 119, pp. 1597–1607 (2020)
9. Cohn, D.A., Ghahramani, Z., Jordan, M.I.: Active learning with statistical models. J. Artif. Intell. Res. **4**, 129–145 (1996)
10. Fang, Y., Yap, P., Lin, W., Zhu, H., Liu, M.: Source-free unsupervised domain adaptation: A survey. Neural Networks **174**, 106230 (2024)
11. Fu, B., Cao, Z., Wang, J., Long, M.: Transferable query selection for active domain adaptation. In: IEEE Conference on Computer Vision and Pattern Recognition. pp. 7272–7281 (2021)
12. Fumero, F., Alayón, S., Sánchez, J.L., Sigut, J.F., González-Hernández, M.: RIM-ONE: an open retinal image database for optic nerve evaluation. In: International Symposium on Computer-Based Medical Systems. pp. 1–6 (2011)
13. Gal, Y., Islam, R., Ghahramani, Z.: Deep bayesian active learning with image data. In: International Conference on Machine Learning. Proceedings of Machine Learning Research, vol. 70, pp. 1183–1192 (2017)
14. Ganin, Y., Lempitsky, V.S.: Unsupervised domain adaptation by backpropagation. In: International Conference on Machine Learning. JMLR Workshop and Conference Proceedings, vol. 37, pp. 1180–1189 (2015)
15. Grill, J., Strub, F., Altché, F., Tallec, C., Richemond, P.H., Buchatskaya, E., Doersch, C., Pires, B.Á., Guo, Z., Azar, M.G., Piot, B., Kavukcuoglu, K., Munos, R., Valko, M.: Bootstrap your own latent - A new approach to self-supervised learning. In: Annual Conference on Neural Information Processing Systems (2020)

16. Han, K., Sheng, V.S., Song, Y., Liu, Y., Qiu, C., Ma, S., Liu, Z.: Deep semi-supervised learning for medical image segmentation: A review. Expert Syst. Appl. **245**, 123052 (2024)
17. He, C., Li, K., Zhang, Y., Yang, Z., Pang, Y., Tang, L., Fang, C., Zhang, Y., Kong, L., Li, X., Farsiu, S.: Segment concealed objects with incomplete supervision. IEEE Trans. Pattern Anal. Mach. Intell. **47**(9), 7832–7851 (2025)
18. He, K., Fan, H., Wu, Y., Xie, S., Girshick, R.B.: Momentum contrast for unsupervised visual representation learning. In: IEEE/CVF Conference on Computer Vision and Pattern Recognition. pp. 9726–9735 (2020)
19. Isensee, F., Jaeger, P.F., Kohl, S.A., Petersen, J., Maier-Hein, K.H.: nnu-net: a self-configuring method for deep learning-based biomedical image segmentation. Nat. Methods **18**(2), 203–211 (2021)
20. Karani, N., Erdil, E., Chaitanya, K., Konukoglu, E.: Test-time adaptable neural networks for robust medical image segmentation. Medical Image Anal. **68**, 101907 (2021)
21. Kothandaraman, D., Shekhar, S., Sancheti, A., Ghuhan, M., Shukla, T., Manocha, D.: SALAD : Source-free active label-agnostic domain adaptation for classification, segmentation and detection. In: IEEE/CVF Winter Conference on Applications of Computer Vision. pp. 382–391 (2023)
22. Kundu, J.N., Kulkarni, A.R., Singh, A., Jampani, V., Babu, R.V.: Generalize then adapt: Source-free domain adaptive semantic segmentation. In: IEEE/CVF International Conference on Computer Vision. pp. 7026–7036 (2021)
23. Li, J., Yu, Z., Du, Z., Zhu, L., Shen, H.T.: A comprehensive survey on source-free domain adaptation. IEEE Trans. Pattern Anal. Mach. Intell. **46**(8), 5743–5762 (2024)
24. Li, L., Zhou, Y., Yang, G.: Robust source-free domain adaptation for fundus image segmentation. In: IEEE/CVF Winter Conference on Applications of Computer Vision. pp. 7825–7834 (2024)
25. Li, S., Zhang, R., Gong, K., Xie, M., Ma, W., Gao, G.: Source-free active domain adaptation via augmentation-based sample query and progressive model adaptation. IEEE Trans. Neural Networks Learn. Syst. **36**(2), 2538–2550 (2025)
26. Li, X., Du, Z., Li, J., Zhu, L., Lu, K.: Source-free active domain adaptation via energy-based locality preserving transfer. In: ACM International Conference on Multimedia. pp. 5802–5810 (2022)
27. Liu, Y., Zhang, W., Wang, J.: Source-free domain adaptation for semantic segmentation. In: IEEE Conference on Computer Vision and Pattern Recognition. pp. 1215–1224 (2021)
28. Long, J., Shelhamer, E., Darrell, T.: Fully convolutional networks for semantic segmentation. In: IEEE Conference on Computer Vision and Pattern Recognition. pp. 3431–3440 (2015)
29. Long, M., Zhu, H., Wang, J., Jordan, M.I.: Unsupervised domain adaptation with residual transfer networks. In: Annual Conference on Neural Information Processing Systems. pp. 136–144 (2016)
30. Luo, Z., Luo, X., Gao, Z., Wang, G.: An uncertainty-guided tiered self-training framework for active source-free domain adaptation in prostate segmentation. In: International Conference on Medical Image Computing and Computer Assisted Intervention. vol. 15009, pp. 107–117 (2024)
31. Lyu, M., Hao, T., Xu, X., Chen, H., Lin, Z., Han, J., Ding, G.: Learn from the learnt: Source-free active domain adaptation via contrastive sampling and visual persistence. In: European Conference on Computer Vision. vol. 15059, pp. 228–246 (2024)

32. Ma, J., He, Y., Li, F., Han, L., You, C., Wang, B.: Segment anything in medical images. Nat. Commun. **15**(1), 654 (2024)
33. Orlando, J.I., Fu, H., Breda, J.B., van Keer, K., Bathula, D.R., Diaz-Pinto, A., Fang, R., Heng, P., Kim, J., Lee, J., Lee, J., Li, X., Liu, P., Lu, S., Murugesan, B., Naranjo, V., Phaye, S.S.R., Shankaranarayana, S.M., Sikka, A., Son, J., van den Hengel, A., Wang, S., Wu, J., Wu, Z., Xu, G., Xu, Y., Yin, P., Li, F., Zhang, X., Xu, Y., Bogunovic, H.: REFUGE challenge: A unified framework for evaluating automated methods for glaucoma assessment from fundus photographs. CoRR **abs/1910.03667** (2019)
34. Pinheiro, P.O.: Unsupervised domain adaptation with similarity learning. In: IEEE Conference on Computer Vision and Pattern Recognition. pp. 8004–8013 (2018)
35. Prabhu, V., Chandrasekaran, A., Saenko, K., Hoffman, J.: Active domain adaptation via clustering uncertainty-weighted embeddings. In: International Conference on Computer Vision. pp. 8485–8494 (2021)
36. Ronneberger, O., Fischer, P., Brox, T.: U-net: Convolutional networks for biomedical image segmentation. In: International Conference on Medical Image Computing and Computer Assisted Intervention. vol. 9351, pp. 234–241 (2015)
37. Sandler, M., Howard, A.G., Zhu, M., Zhmoginov, A., Chen, L.: Mobilenetv2: Inverted residuals and linear bottlenecks. In: IEEE Conference on Computer Vision and Pattern Recognition. pp. 4510–4520 (2018)
38. Sener, O., Savarese, S.: Active learning for convolutional neural networks: A core-set approach. In: International Conference on Learning Representations. OpenReview.net (2018)
39. Sener, O., Song, H.O., Saxena, A., Savarese, S.: Learning transferrable representations for unsupervised domain adaptation. In: Annual Conference on Neural Information Processing Systems. pp. 2110–2118 (2016)
40. Sinha, S., Ebrahimi, S., Darrell, T.: Variational adversarial active learning. In: IEEE/CVF International Conference on Computer Vision. pp. 5971–5980 (2019)
41. Sivaswamy, J., Krishnadas, S., Chakravarty, A., Joshi, G., Tabish, A.S., et al.: A comprehensive retinal image dataset for the assessment of glaucoma from the optic nerve head analysis. JSM Biomed. Imaging Data Papers **2**(1), 1004 (2015)
42. Stan, S., Rostami, M.: Unsupervised model adaptation for source-free segmentation of medical images. Medical Image Anal. **95**, 103179 (2024)
43. Su, J., Tsai, Y., Sohn, K., Liu, B., Maji, S., Chandraker, M.: Active adversarial domain adaptation. In: IEEE Conference on Computer Vision and Pattern Recognition Workshops. pp. 1–4 (2019)
44. Sun, T., Lu, C., Ling, H.: Local context-aware active domain adaptation. In: International Conference on Computer Vision. pp. 18588–18597 (2023)
45. Tang, L., Li, K., He, C., Zhang, Y., Li, X.: Source-free domain adaptive fundus image segmentation with class-balanced mean teacher. In: International Conference on Medical Image Computing and Computer Assisted Intervention. vol. 14220, pp. 684–694 (2023)
46. Toldo, M., Maracani, A., Michieli, U., Zanuttigh, P.: Unsupervised domain adaptation in semantic segmentation: a review. CoRR **abs/2005.10876** (2020)
47. VS, V., Valanarasu, J.M.J., Patel, V.M.: Target and task specific source-free domain adaptive image segmentation. In: Medical Imaging with Deep Learning. Proceedings of Machine Learning Research, vol. 250, pp. 1627–1639. PMLR (2024)
48. Vu, T., Jain, H., Bucher, M., Cord, M., Pérez, P.: ADVENT: adversarial entropy minimization for domain adaptation in semantic segmentation. In: IEEE Conference on Computer Vision and Pattern Recognition. pp. 2517–2526 (2019)

49. Wang, D., Shelhamer, E., Liu, S., Olshausen, B.A., Darrell, T.: Tent: Fully test-time adaptation by entropy minimization. In: International Conference on Learning Representations. OpenReview.net (2021)
50. Wang, F., Han, Z., Zhang, Z., He, R., Yin, Y.: MHPL: minimum happy points learning for active source free domain adaptation. In: IEEE/CVF Conference on Computer Vision and Pattern Recognition. pp. 20008–20018 (2023)
51. Wang, H., Chen, J., Zhang, S., He, Y., Xu, J., Wu, M., He, J., Liao, W., Luo, X.: Dual-reference source-free active domain adaptation for nasopharyngeal carcinoma tumor segmentation across multiple hospitals. IEEE Trans. Medical Imaging **43**(12), 4078–4090 (2024)
52. Wang, S., Yu, L., Li, K., Yang, X., Fu, C., Heng, P.: Boundary and entropy-driven adversarial learning for fundus image segmentation. In: International Conference on Medical Image Computing and Computer Assisted Intervention. vol. 11764, pp. 102–110 (2019)
53. Wang, W., Zhou, T., Yu, F., Dai, J., Konukoglu, E., Gool, L.V.: Exploring cross-image pixel contrast for semantic segmentation. In: IEEE/CVF International Conference on Computer Vision. pp. 7283–7293 (2021)
54. Wang, X., Zhang, R., Shen, C., Kong, T., Li, L.: Dense contrastive learning for self-supervised visual pre-training. In: IEEE Conference on Computer Vision and Pattern Recognition. pp. 3024–3033 (2021)
55. Wilson, G., Cook, D.J.: A survey of unsupervised deep domain adaptation. ACM Trans. Intell. Syst. Technol. **11**(5), 51:1–51:46 (2020)
56. Xia, Q., Zheng, H., Zou, H., Luo, D., Tang, H., Li, L., Jiang, B.: A comprehensive review of deep learning for medical image segmentation. Neurocomputing **613**, 128740 (2025)
57. Xie, M., Li, Y., Wang, Y., Luo, Z., Gan, Z., Sun, Z., Chi, M., Wang, C., Wang, P.: Learning distinctive margin toward active domain adaptation. In: IEEE Conference on Computer Vision and Pattern Recognition. pp. 7983–7992 (2022)
58. Yang, C., Guo, X., Chen, Z., Yuan, Y.: Source free domain adaptation for medical image segmentation with fourier style mining. Medical Image Anal. **79**, 102457 (2022)
59. Yin, Y., Hu, W., Liu, Z., Wang, G., Xiang, S., Zimmermann, R.: Crossmatch: Source-free domain adaptive semantic segmentation via cross-modal consistency training. In: IEEE/CVF International Conference on Computer Vision. pp. 21729–21739 (2023)
60. Yu, Q., Xi, N., Yuan, J., Zhou, Z., Dang, K., Ding, X.: 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. vol. 14226, pp. 3–12. Springer (2023)
61. Zhang, Y., Ma, D., Wu, X.: Source-free domain adaptation framework based on confidence constrained mean teacher for fundus image segmentation. Neurocomputing **620**, 129262 (2025)
62. Zhang, Z., Chen, W., Cheng, H., Li, Z., Li, S., Lin, L., Li, G.: Divide and contrast: Source-free domain adaptation via adaptive contrastive learning. In: Annual Conference on Neural Information Processing Systems (2022)
63. Zheng, K., Xia, H., Xia, S., Shao, M., Ding, Z.: Supportive negatives spectral augmentation for source-free cross-domain segmentation. In: AAAI. pp. 10573–10581 (2025)