TrustSyn: Reliable and Divergent Synergy for Mixed-Domain Fundus Segmentation

Baojun Qiao[1], Xinhui Wang[1], Wanjun Zhang[1(✉)], Jason Turuwhenua[2,3], and Mohammad Norouzifard[4]

[1] School of Computer and Information Engineering, Henan University, Kaifeng, China
wjz@henu.edu.cn
[2] Auckland Bioengineering Institute, The University of Auckland, Auckland 1142, New Zealand
[3] School of Optometry and Vision Science, The University of Auckland, Auckland 1142, New Zealand
[4] School of Technology, Yoobee College of Creative Innovation, Auckland 1010, New Zealand

Abstract. Reliable fundus segmentation is crucial for ophthalmic diagnosis, yet label scarcity and cross-center domain shifts severely degrade the generalization of automated models. Although collaborative frameworks incorporating foundation models show promise for Mixed-Domain Semi-Supervised (MiDSS) learning, they suffer from coarse-grained reliability estimation and representational collapse. Representational collapse forces the conventional model to forfeit its unique inductive biases, preventing effective synergistic inference. To address these challenges, we propose TrustSyn, a synergistic Teacher-Student framework for MiDSS learning, which bridges the gap between label scarcity and domain shifts. To rectify reliability estimation, we introduce a Spatial Reliability Assessment (SRA) coupled with an Uncertainty Weighting (UW) strategy. This mechanism dynamically modulates pixel-level supervision to leverage global context priors for reliable pseudo-label generation selectively. To avert representational collapse, we introduce a Feature Discrepancy (FD) constraint between the latent representations of both models. By enforcing feature divergence, FD ensures output consensus is achieved through diverse inference mechanisms rather than redundant representations. Extensive experiments on the Fundus dataset demonstrate that TrustSyn exhibits robustness against domain shifts, achieving state-of-the-art performance with a Dice score of 89.17% and an ASD of 3.20. Code is available at https://github.com/wwwangxinhui/TrustSyn.

Keywords: Semi-supervised Learning · Mixed-Domain Adaptation · Fundus Image Segmentation.

1 Introduction

The cup-to-disc ratio is a pivotal ophthalmic clinical indicator for screening and diagnosis [20]. It is particularly critical for detecting glaucoma, the leading cause of irreversible blindness [29]. However, manual delineation is labor-intensive and requires expert knowledge, limiting widespread deployment [28]. While deep learning offers an automated solution, models trained on single-center data suffer from performance degradation when applied to different domains. This decline is attributed to differences in imaging protocols and patient demographics [30], as well as domain shifts caused by scanner variations [10].

Mixed-Domain Semi-Supervised (MiDSS) segmentation [14] leverages limited labeled source data and abundant unlabeled multi-center data. Early attempts utilizing consistency regularization [19], teacher-student architectures [22], or data augmentation [1] often degrade under significant domain shifts. To bridge these gaps, recent approaches employ intermediate domain construction [14,15], feature alignment [9,27], or dynamic adaptation [31,8,11]. However, conventional models like U-Net [18] often overfit source styles. Alternatively, foundation models (e.g., SAM [7], MedSAM [12]) offer zero-shot generalization, inspiring various decoupled or prompt-based adaptation methods[23,4,2,16]. Yet, direct application risks error accumulation from overconfident mispredictions on specific medical textures [13]. Consequently, collaborative frameworks[6,13] have established state-of-the-art baselines by synergizing global priors with domain-specific efficiency. Despite these advances, current collaborative frameworks face two limitations. (1) **Coarse-grained reliability estimation.** Existing fusion mechanisms rely on image-level metrics or coarse uncertainty [17], overlooking spatial heterogeneity. They fail to capture pixel-level uncertainty in low-contrast regions like the optic cup [5,21], leading to noise propagation. (2) **Representational collapse.** Simply minimizing prediction divergence encourages the conventional model to passively mimic the foundation model. This loss of diversity prevents the framework from leveraging complementary inductive biases to handle domain shifts [25].

To address these, we propose TrustSyn, a collaborative Teacher-Student architecture for domain adaptation. Our contributions are: (1) We propose a Spatial Reliability Assessment (SRA) coupled with an Uncertainty Weighting (UW) strategy. This mechanism captures pixel-level heterogeneity to dynamically modulate supervision, filtering noise in ambiguous zones. (2) We introduce a Feature Discrepancy (FD) constraint to enforce feature divergence between the foundation and conventional models, averting representational collapse. (3) Extensive experiments on the Fundus dataset demonstrate that TrustSyn achieves state-of-the-art performance and robustness against domain shifts.

2 Method

The architecture of TrustSyn is shown in Fig. 1, ϵ and D denote the encoder and decoder modules, respectively. We adopt a synergistic Teacher-Student framework that integrates a pre-trained MedSAM as the foundation model with a U-Net as the conventional model. The Teacher Branch employs SRA and UW to generate reliable pseudo-label and dynamically modulate pixel-level supervision. The Student Branch incorporates an FD constraint to enforce feature divergence, aimed at preventing representational collapse.

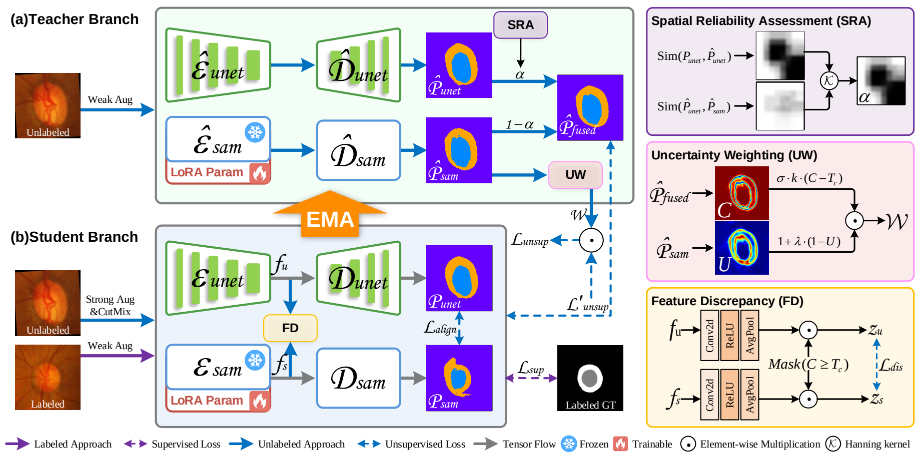

Fig. 1. Overview of TrustSyn. (a) The Teacher Branch utilizes SRA and UW for pseudo-label rectification. (b) The Student Branch employs FD constraint to enforce representational diversity.

2.1 Spatial Reliability Assessment for Fine-Grained Consensus

Domain shifts induce pixel-level uncertainty overlooked by global metrics. To address this, SRA estimates reliability as a fine-grained spatial map, using a sliding window strategy with 50% overlap and Hanning weighting to evaluate local consensus. Unlike SynFoC's [13] coarse image-level confidence, our pixel-level adaptation effectively captures heterogeneity in ambiguous boundaries.

We facilitate reliable pseudo-label generation through two complementary dimensions. Self-Confidence measures the model's internal consistency by comparing predictions between the U-Net student and teacher. Mutual-Confidence measures the semantic alignment between the U-Net and MedSAM by leveraging the foundation model's zero-shot priors.

Let P_{unet} and $\hat{P}_{unet}$ denote the probability maps of the student and teacher U-Net, and let $\hat{P}_{sam}$ denote the probability map of the teacher MedSAM. The Sim operation illustrated in Fig. 1 computes spatial confidence by sliding a local window across global inputs. Specifically, for a local window $\mathcal{W}(i,j)$ centered at pixel (i,j), the Self-Confidence $conf_{self}(i,j)$ is computed locally as:

$$conf_{self}(i,j) = \frac{2\sum_{n\in\mathcal{W}(i,j)} P^n_{unet} \cdot \hat{P}^n_{unet}}{\sum_{n\in\mathcal{W}(i,j)}(P^n_{unet} + \hat{P}^n_{unet})}, \tag{1}$$

and the Mutual-Confidence $conf_{mut}(i,j)$ is computed as:

$$conf_{mut}(i,j) = \frac{2\sum_{n\in\mathcal{W}(i,j)} \hat{P}^n_{unet} \cdot \hat{P}^n_{sam}}{\sum_{n\in\mathcal{W}(i,j)}(\hat{P}^n_{unet} + \hat{P}^n_{sam})}, \tag{2}$$

where n indexes the pixels within $\mathcal{W}(i,j)$. These local estimates are aggregated into spatial reliability maps. To mitigate blocking artifacts caused by window boundaries, we introduce a Hanning window weighting mechanism. Each window's contribution is weighted by a 2D Hanning kernel $\mathcal{K}$, and the final map is normalized by the accumulated weights W_{acc}:

$$\alpha(i,j) = \frac{\sum_{(u,v)} \mathcal{K}(i-u, j-v) \cdot (conf_{self}(u,v) \cdot conf_{mut}(u,v))}{W_{acc}(i,j)}, \tag{3}$$

where (u,v) indexes the spatial locations of overlapping windows. This multiplicative formulation in Eq. (3) ensures that a region is considered reliable only if it exhibits both high internal consistency and semantic alignment. The fused probability map $\hat{P}_{fused}$ is then synthesized via:

$$\hat{P}_{fused} = \alpha(i,j) \cdot \hat{P}_{unet} + (1 - \alpha(i,j)) \cdot \hat{P}_{sam}. \tag{4}$$

Thus, it prioritizes domain-specific predictions in stable regions and adaptively falls back to MedSAM's robust priors in ambiguous zones.

2.2 Uncertainty Weighting for Gradient Rectification

Existing pseudo-labeling methods often suffer from confirmation bias due to misaligned probability calibration. While most approaches rely on hard thresholding, this strategy creates a difficult trade-off in fundus images with ambiguous cup-disc boundaries. High thresholds often lead to under-segmentation by discarding informative transition pixels, thereby hampering the capture of fine-grained anatomical details. Low thresholds risk over-segmentation by misclassifying ambiguous background regions as target tissues. To address these structural inaccuracies, we propose UW, a differentiable pixel-wise mechanism that utilizes a soft, adaptive map to modulate supervisory signals.

We characterize the uncertainty via a pixel-level entropy map derived from the teacher MedSAM. Given the probability map $\hat{P}_{sam}$, we first compute the pixel-wise Shannon entropy $H = -\sum \hat{P}_{sam} \log(\hat{P}_{sam})$. To ensure robustness against variations across different samples, H is normalized via instance-level min-max scaling to generate the final uncertainty map U. Given that MedSAM operates at a different resolution, U is bilinearly upsampled to match the U-Net's spatial dimensions. The gradient rectification weight map $\mathcal{W}(i,j)$ is then computed as:

$$\mathcal{W}(i,j) = (1 + \lambda_h(1 - U(i,j))) \cdot \sigma\left(k \cdot (C(i,j) - T_c)\right), \tag{5}$$

where λ_h is a scaling factor controlling the strength of the entropy-based modulation. Furthermore, $C(i,j) = \max_c \hat{P}_{fused}(i,j,c)$ denotes the pixel-wise confidence derived from the maximum probability across classes, and T_c represents the class-specific threshold for the predicted class at this pixel. The sigmoid function $\sigma(\cdot)$, scaled by a factor k, acts as a differentiable soft gate.

This design mitigates the limitations of binary masking. The first term leverages the distributional certainty of the teacher MedSAM to amplify the contribution of low-entropy predictions. The second term provides a smooth gradient transition around the threshold. Unlike hard thresholding which causes an abrupt loss of supervision signals, this synergistic weighting lets the model progressively learn from challenging samples near decision boundaries. This strategy ensures that the optimization is driven by stable and reliable signals without discarding valuable edge information.

2.3 Feature Discrepancy for Representational Diversity

Unconstrained output alignment risks inducing representational collapse, where the conventional model mimics the foundation model without learning independent discriminatory features. To prevent this, we enforce feature divergence between the two branches within high-confidence regions, ensuring convergence to identical semantics via distinct representational pathways.

We introduce an FD constraint during the optimization of the student branch. The latent representations f_u and f_s, representing the bottleneck features of the U-Net and MedSAM respectively, are first mapped into a shared, normalized metric space via projection heads ϕ. The projection head ϕ comprises a 1×1 convolution for dimension reduction followed by non-linear activation and adaptive pooling. This design aligns the heterogeneous channel dimensions and spatial resolutions of the two branches into a unified embedding space. To prevent optimizing on ambiguous or low-confidence regions, this discrepancy constraint is selectively applied within the reliable region $\mathcal{M}$, defined as:

$$\mathcal{M} = \mathbb{I}(C \geq T_c). \tag{6}$$

Unlike global cosine-similarity minimization in homogeneous sub-nets [25], we address representational collapse across heterogeneous architectures. We minimize the semantic overlap between the two branches by penalizing their cosine similarity:

$$\mathcal{L}_{dis} = \frac{1}{|\mathcal{M}|} \sum_{\mathcal{M}} \left(1 + \cos(\phi(f_u), \phi(f_s))\right). \tag{7}$$

Minimizing $\mathcal{L}_{dis}$ imposes a divergence constraint that encourages the separation of feature directions. While $\mathcal{L}_{dis}$ pushes representations apart, the segmentation objective enforces semantic coherence. This interplay drives a dynamic equilibrium that empirically converges to statistical orthogonality. In this state, the two encoders develop complementary roles as the U-Net branch refines sensitivity to high-frequency local details while the MedSAM branch maintains robust global semantic priors. This synergistic fusion of distinct visual perspectives significantly enhances robustness against domain shifts.

2.4 Optimization Objective for Synergistic Learning

The framework is trained via a composite objective optimizing source supervision and target consistency. The total loss is formulated as:

$$\mathcal{L}_{total} = \mathcal{L}_{sup} + \lambda(t) \left(\mathcal{L}_{unsup} + \alpha\mathcal{L}_{align} + \beta\mathcal{L}_{dis}\right), \tag{8}$$

where $\lambda(t)$ is a time-dependent ramp-up function used to stabilize early-stage training, and α, β are balancing coefficients.

Reliability-Guided Supervision. For labeled data $\mathcal{D}_L$, $\mathcal{L}_{sup}$ minimizes the standard summation of Focal and Dice losses. For unlabeled data $\mathcal{D}_U$, to mitigate confirmation bias, we employ a reliability-weighted consistency constraint:

$$\mathcal{L}_{unsup} = \mathcal{W} \cdot \mathcal{L}_{focal} + \mathcal{L}_{dice}(\mathcal{M}). \tag{9}$$

Here, $\mathcal{W}$ is derived from Eq. (5). It acts as an uncertainty-aware modulator that adaptively attenuates signals from high-entropy regions. Simultaneously, the reliability mask $\mathcal{M}$, derived from Eq. (6), restricts geometric constraints to the reliable manifold. This mechanism effectively filters out noise in ambiguous zones, curbing error accumulation.

Synergistic Regularization. We employ two regularization terms to strike a balance between consistency and diversity. First, $\mathcal{L}_{align}$ enforces output consistency. Following the consensus strategy in SynFoC [13], this term functions as a smoothness regularizer adhering to the semi-supervised cluster assumption, pushing decision boundaries towards low-density regions to improve generalization. Second, to prevent representational collapse, we impose the Feature Discrepancy loss $\mathcal{L}_{dis}$ obtained by Eq. (7). Instead of merely maximizing distance, $\mathcal{L}_{dis}$ promotes feature divergence within reliable regions.

3 Experiments

3.1 Setup and Implementation

Dataset and Protocols. We evaluate TrustSyn on the 1,060-image dataset for Optic Cup and Disc segmentation [24]. This dataset comprises four domains from three public repositories: Drishti-GS (Domain 1), RIM-ONE-r3 (Domain 2), and REFUGE (Domains 3 and 4). These anonymized public datasets feature diverse scanner models such as Nidek AFC-210, Zeiss Visucam 500, and Canon CR-2, posing significant domain shifts. We partition the data into: (i) **Labeled Source** $\mathcal{D}_L$ (20 images sampled from Domain 1 training set), (ii) **Unlabeled Mixture** $\mathcal{D}_U$ (remaining training samples from all domains), and (iii) **Evaluation** $\mathcal{D}_{test}$ (all independent test samples). We use Dice Similarity Coefficient (DSC), Jaccard Coefficient (JC), 95% Hausdorff Distance (HD_{95}), and Average Surface Distance (ASD) for evaluation.

Implementation Details. TrustSyn is implemented with a dual-stream architecture comprising U-Net and MedSAM. For efficient adaptation, LoRA adapters (rank=4) are integrated into the frozen MedSAM encoder. Training is conducted on a single NVIDIA RTX 3090 GPU for 30,000 iterations. We optimize U-Net

Table 1. Comparison with SOTA methods. Best: **bold**; Second: underlined.

Method	Optic Cup & Disc Segmentation (DSC↑)								DSC↑	JC↑	HD_{95}↓	ASD↓
	Domain 1		Domain 2		Domain 3		Domain 4		Avg.			
FixMatch [19]	81.18	91.29	72.04	87.60	80.41	92.95	74.58	87.07	83.39	73.48	11.77	5.60
SS-Net [26]	59.42	78.15	67.32	85.05	45.69	69.91	38.76	61.13	63.18	53.49	44.90	25.73
BCP [1]	71.65	91.10	77.19	92.00	72.63	90.77	77.67	91.42	83.05	73.66	11.05	5.80
ABD [3]	73.92	79.71	65.19	90.96	77.61	86.11	74.79	86.72	79.38	69.28	13.99	8.14
H-SAM [2]	76.97	93.01	79.01	90.47	76.85	91.86	81.03	92.42	85.20	76.26	9.67	4.99
CPC-SAM [16]	75.99	94.34	80.10	**93.08**	83.19	92.81	83.43	93.20	87.02	78.65	8.50	4.24
SymGD [14]	83.71	92.96	80.47	89.93	84.18	92.97	83.71	93.38	87.66	79.10	8.21	3.89
SynFoC [13]	85.44	93.29	**82.50**	90.52	**85.51**	93.56	83.78	94.23	88.60	80.50	6.56	3.47
UST-RUN [15]	83.88	93.34	80.44	90.74	84.49	93.28	84.16	93.72	88.01	79.51	7.70	3.71
HARP [11]	-	-	-	-	-	-	-	-	88.22	80.26	7.81	3.74
TrustSyn	**87.87**	**97.24**	77.66	90.52	84.54	**94.48**	**86.43**	**94.62**	**89.17**	**81.55**	**6.53**	**3.20**

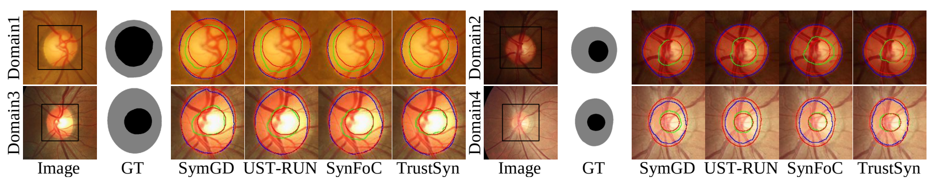

Fig. 2. Visual comparison across target domains. Red: Ground Truth; Green: Optic Cup; Blue: Optic Disc.

using SGD (momentum=0.9, lr=0.03) and MedSAM using AdamW (lr=0.0001), both following a poly decay policy. The teacher U-Net is updated via Exponential Moving Average (EMA) with a decay rate of 0.99. Loss weights are set to $\alpha = 1.0$ and $\beta = 0.1$. Class-specific thresholds T_c are empirically set to 0.98 for background and 0.85 for foreground classes.

3.2 Results and Analysis

As shown in Table 1, TrustSyn achieves a superior average DSC of 89.17% and JC of 81.55%, while significantly reducing geometric errors with an HD_{95} of 6.53 and ASD of 3.20. These results underscore TrustSyn's robust adaptation across heterogeneous conditions, outperforming both traditional SSL methods and SAM-based frameworks. TrustSyn maintains high precision in challenging optic cup segmentation, surpassing previous best results on Domain 1 by 2.43% and Domain 4 by 2.65%. This demonstrates that TrustSyn learns robust discriminatory features rather than merely memorizing source styles. The improvements are further corroborated by visual comparisons in Fig. 2. Unlike existing methods prone to over-segmentation, TrustSyn produces precise contours that closely align with target boundaries, particularly at low-contrast retinal cup edges. This boundary-aware precision validates the effectiveness of our synergistic framework. The SRA and UW modules suppress confirmation bias by generating spatially reliable

Table 2. Ablation study of key components on the Fundus dataset [24].

Base	SRA	UW	FD	DSC ↑	JC ↑	HD_{95} ↓	ASD ↓
✓				88.31	80.46	6.88	3.58
✓	✓			88.58	80.65	6.85	3.40
✓		✓		88.42	80.52	6.99	3.54
✓			✓	88.54	80.67	6.73	3.34
✓	✓	✓		88.93	81.20	6.65	3.29
✓	✓		✓	88.85	81.08	6.70	3.33
✓		✓	✓	88.76	80.95	6.75	3.36
✓	✓	✓	✓	**89.17**	**81.55**	**6.53**	**3.20**

pseudo-labels, while the FD constraint preserves the distinctive inductive biases of both models. By reaching consensus through diverse pathways, TrustSyn prevents representational collapse and ensures stability under severe domain shifts. Failure analysis on degraded samples (e.g., Domain 2) reveals conservative behavior: UW suppresses high-entropy signals and FD remains inactive in low-contrast regions. This limits aggressive adaptation but effectively prevents catastrophic confirmation bias, ensuring superior robustness. Overall, TrustSyn effectively handles cross-center variability and limited annotations, ensuring reliable and robust mixed-domain fundus segmentation.

3.3 Ablation Study

We investigate the contribution of each component in Table 2. The results demonstrate that while individual modules yield incremental improvements, their synergistic integration is essential for handling severe domain shifts. Specifically, introducing SRA improves performance by replacing coarse global metrics with fine-grained spatial reliability maps. This assessment serves as a critical prerequisite for UW, which adaptively modulates supervisory signals. The synergy between SRA and UW yields a substantial performance boost to 88.93% DSC and significantly reduces geometric errors to 3.29 ASD. This confirms that facilitating reliable pseudo-labeling in ambiguous transition zones is the primary driver for boundary alignment. Building on this refined supervision, the FD constraint further sharpens the segmentation results. While SRA and UW ensure the reliability of labels, FD promotes diversity in feature learning. By enforcing feature divergence, FD prevents the student from passively mimicking the foundation model's potential systematic biases, leading to a further reduction in HD_{95} to 6.53 and ASD to 3.20. Paired t-tests confirm statistically significant component gains ($p < 0.05$). Furthermore, TrustSyn exhibits strong parameter robustness: SRA is insensitive to grid partitions, loss weights (α, β) are stable, and soft-gating ensures negligible sensitivity to T_c. The full configuration of TrustSyn achieves the best holistic performance, validating that the combination of high-quality supervision and decoupled representations is indispensable for achieving precise and robust mixed-domain segmentation.

4 Conclusion

In this paper, we propose TrustSyn, a synergistic collaborative framework for mixed-domain semi-supervised fundus segmentation. By harmonizing SRA and UW for spatially reliable pseudo-labeling with an FD constraint for feature divergence, TrustSyn effectively mitigates confirmation bias and prevents representational collapse. Extensive experiments demonstrate that our method achieves state-of-the-art robustness, yielding a Dice score of 89.17% and superior boundary precision across multiple domains. These results confirm that TrustSyn effectively bridges the gap between limited annotations and cross-center variability, offering a reliable tool for automated glaucoma screening in diverse clinical settings. Future work will focus on enhancing generalization across even more heterogeneous healthcare environments.

Acknowledgments. This study was supported by the Key R&D Projects in Henan Province (Grant No. 261111212000).

Disclosure of Interests. The authors have no competing interests to declare that are relevant to the content of this article.

References

1. Bai, Y., Chen, D., Li, Q., Shen, W., Wang, Y.: Bidirectional copy-paste for semi-supervised medical image segmentation. In: CVPR. pp. 11514–11524 (2023)
2. Cheng, Z., Wei, Q., Zhu, H., Wang, Y., Qu, L., Shao, W., Zhou, Y.: Unleashing the potential of SAM for medical adaptation via hierarchical decoding. In: CVPR. pp. 3511–3522 (2024)
3. Chi, H., Pang, J., Zhang, B., Liu, W.: Adaptive bidirectional displacement for semi-supervised medical image segmentation. In: CVPR. pp. 4070–4080 (2024)
4. Gao, Y., Xia, W., Hu, D., Wang, W., Gao, X.: DeSAM: Decoupled Segment Anything Model for generalizable medical image segmentation. In: MICCAI. pp. 509–519 (2024)
5. Guo, E., Wang, Z., Zhao, Z., et al.: Imbalanced medical image segmentation with pixel-dependent noisy labels. IEEE Transactions on Medical Imaging **44**(5), 2016–2027 (2025)
6. Guo, S.: Model cooperation learning for multi-source domain generalization in medical image segmentation. Applied Soft Computing **188**, 114397 (2026)
7. Kirillov, A., Mintun, E., Ravi, N., Mao, H., Rolland, C., Gustafson, L., Xiao, T., Whitehead, S., Berg, A.C., Lo, W.Y., Dollar, P., Girshick, R.: Segment Anything. In: ICCV. pp. 4015–4026 (2023)
8. Li, H., Li, H., Chen, J., Zhong, R., Niu, K., Fu, H., Liu, J.: AIF-SFDA: Autonomous information filter driven source-free domain adaptation for medical image segmentation. In: AAAI. pp. 4716–4724 (2025)
9. Li, Y., Ma, X., Sun, J., Chen, H., Ren, B., Wang, M.: MGC-net: Semi-supervised domain generalization in medical image segmentation via multi-granularity consistency. Neurocomputing **660**, 131919 (2026)

10. Liu, Q., Chen, C., Qin, J., Dou, Q., Heng, P.A.: FedDG: Federated domain generalization on medical image segmentation via episodic learning in continuous frequency space. In: CVPR. pp. 1013–1023 (2021)
11. Liu, Y., Ye, W., Liu, H., Chen, Z., Li, P., Xu, R.X., Sun, M.: HARP: Harmonization and adaptive refinement of pseudo-labels for cross-domain medical image segmentation. In: MICCAI. pp. 311–321 (2025)
12. Ma, J., He, Y., Li, F., Han, L., You, C., Wang, B.: MedSAM: Segment anything in medical images. Nature Communications **15**, 654 (2024)
13. Ma, Q., Zhang, J., Li, Z., Qi, L., Yu, Q., Shi, Y.: Steady Progress Beats Stagnation: Mutual aid of foundation and conventional models in mixed domain semi-supervised medical image segmentation. In: CVPR. pp. 5175–5185 (2025)
14. Ma, Q., Zhang, J., Qi, L., Yu, Q., Shi, Y., Gao, Y.: Constructing and exploring intermediate domains in mixed domain semi-supervised medical image segmentation. In: CVPR. pp. 11642–11651 (2024)
15. Ma, Q., Zhang, J., Qi, L., et al.: Unleashing the power of intermediate domains for mixed domain semi-supervised medical image segmentation. IEEE Transactions on Medical Imaging **44**(11), 4377–4389 (2025)
16. Miao, J., Chen, C., Zhang, K., Chuai, J., Li, Q., Heng, P.A.: Cross prompting consistency with Segment Anything Model for semi-supervised medical image segmentation. In: MICCAI. pp. 167–177 (2024)
17. Philipp, M., Alperovich, A., Gutt-Will, M., Mathis, A., Saur, S., Raabe, A., Mathis-Ullrich, F.: Dynamic CNNs using uncertainty to overcome domain generalization for surgical instrument localization. In: WACV. pp. 3612–3621 (2022)
18. Ronneberger, O., Fischer, P., Brox, T.: U-Net: Convolutional networks for biomedical image segmentation. In: MICCAI. pp. 234–241 (2015)
19. Sohn, K., Berthelot, D., Carlini, N., Zhang, Z., Zhang, H., Raffel, C., Cubuk, E.D., Kurakin, A., Li, C.L.: FixMatch: Simplifying semi-supervised learning with consistency and confidence. In: NeurIPS. pp. 596–608 (2020)
20. Stuart, K.V., de Vries, V.A., Schuster, A.K., et al.: Prevalence of glaucoma in europe and projections to 2050: Findings from the european eye epidemiology consortium. Ophthalmology **132**(10), 1114–1124 (2025)
21. Sun, Y., Liang, D., Li, S.Y., Chen, S., Huang, S.J.: Handling noisy annotation for remote sensing semantic segmentation via boundary-aware knowledge distillation. IEEE Transactions on Geoscience and Remote Sensing **63**, 1–20 (2025)
22. Tarvainen, A., Valpola, H.: Mean teachers are better role models: Weight-averaged consistency targets improve semi-supervised deep learning results. In: NeurIPS (2017)
23. Wang, R., Yang, Z., Song, Y.: OSAM-Fundus: A training-free, one-shot segmentation framework for optic disc and cup in fundus images. Biomedical Signal Processing and Control **100**, 107069 (2025)
24. Wang, S., Yu, L., Li, K., et al.: DoFE: Domain-oriented feature embedding for generalizable fundus image segmentation on unseen datasets. IEEE Transactions on Medical Imaging **39**(12), 4237–4248 (2020)
25. Wang, Z., Zhao, Z., Xing, X., Xu, D., Kong, X., Zhou, L.: Conflict-based cross-view consistency for semi-supervised semantic segmentation. In: CVPR. pp. 19585–19595 (2023)
26. Wu, Y., Wu, Z., Wu, Q., Ge, Z., Cai, J.: Exploring smoothness and class-separation for semi-supervised medical image segmentation. In: MICCAI. pp. 34–43 (2022)
27. Xie, S., Niu, Z., Huang, H., Sun, H., Qin, R., Chen, Y.W., Lin, L.: IS2Net: Intra-domain semantic and inter-domain style enhancement for semi-supervised medical domain generalization. In: ACM MM. pp. 8285–8293 (2023)

28. Xu, Y., Quan, R., Xu, W., Huang, Y., Chen, X., Liu, F.: Advances in Medical Image Segmentation: A comprehensive review of traditional, deep learning and hybrid approaches. Bioengineering **11**, 1034 (2024)
29. Zhang, N., Wang, J., Li, Y., Jiang, B.: Prevalence of primary open angle glaucoma in the last 20 years: a meta-analysis and systematic review. Scientific Reports **11**, 13762 (2021)
30. Zhang, Z., Tong, Z., Tian, C., Ye, Y., Fan, W., Ran, R., Zhou, H., Zhou, F.: Low-frequency amplitude fusion based consistency learning method for multi-source domain adaptation for joint optic disc and cup segmentation. Biomedical Signal Processing and Control **96**, 106481 (2024)
31. Zhang, Z., Chen, Y., Yu, H., et al.: UniAda: Domain unifying and adapting network for generalizable medical image segmentation. IEEE Transactions on Medical Imaging **44**, 1988–2001 (2025)