Lightweight Dual-Task Framework for Semi-Supervised Lesion Segmentation with Knowledge Distillation from SAM

Xuan-Loc Huynh[1(✉)], Huy-Thach Pham[2], Anh Mai Vu[3], Thanh-Minh Nguyen[4], Tran Quang Khai Bui[5], Tat-Bach Nguyen[6], Quan Nguyen[7], Minh Huu Nhat Le[8], and Phat K. Huynh[2(✉)]

[1] Boston University, Boston, MA, USA
xlhuynh@bu.edu
[2] PASSIO Lab, North Carolina A&T State University, Greensboro, NC, USA
pkhuynh@ncat.edu
[3] University of Houston, Houston, TX, USA
[4] Ho Chi Minh City University of Medicine and Pharmacy, Vietnam
[5] University of Science, Viet Nam National University, Ho Chi Minh City, Vietnam
[6] Iowa State University, Ames, IA, USA
[7] Posts and Telecommunications Institute of Technology, Hanoi, Vietnam
[8] Interventional Cardiology Department, Methodist Hospital, Houston, TX, USA

Abstract. In real-world clinical settings, deploying medical AI applications requires lightweight models that can operate under limited computational resources. For skin lesion segmentation, a crucial step in the early detection of skin cancer, the key challenge is to develop models that are not only efficient but also perform reliably with minimal annotated data. To address this, we propose a lightweight and efficient semi-supervised segmentation framework that combines multi-task consistency learning with the representational power of foundation models.Our method is built on three key components: (1) a dual-network co-training framework combining a lightweight MobileNet with a strong ViT-based teacher to balance efficiency and representation power, (2) a fused mask prompt inspired by multi-task consistency, which combines coarse segmentation masks with boundary-aware Signed Distance Function (SDF) maps to guide SAM, and (3) a SAM-guided knowledge distillation strategy, where refined outputs from SAM are used as high-quality pseudo-labels to train the Main Network on unlabeled data. Extensive experiments demonstrate that our approach achieves competitive segmentation performance with significantly reduced annotation effort, offering a practical solution for semi-supervised medical image segmentation in real-world applications.

Keywords: Medical image segmentation · Semi-supervised learning · Segment Anything Model (SAM) · Lightweight models

1 Introduction

Medical image segmentation is a critical component of many clinical applications, but acquiring high-quality annotations remains both expensive and time-consuming as highlighted by Antonelli et al. [1]. Within this domain, skin lesion segmentation is particularly important for the early diagnosis of skin cancer, as it helps extract key visual features like lesion shape and boundary, as demonstrated in Li et al. [9]. To overcome the challenge of limited labeled data, numerous semi-supervised methods for skin lesion segmentation have been proposed, demonstrating strong performance by effectively leveraging both labeled and unlabeled images as reported in Nguyen et al. [13].

One of the core ideas in semi-supervised learning (SSL) is consistency regularization, which encourages models to produce similar predictions when the same input is perturbed. These methods promote smooth decision boundaries in low-density regions of the data space. Data augmentation-based methods include FixMatch [18], which applies weak augmentations to generate pseudo-labels that supervise predictions on strongly augmented versions of the same image, conditioned on a confidence threshold. Many self-training works by Luu et al., Nguyen et al and Yang et al. [10,12,25] build on this by adding sharpening techniques and temporal ensembling to make the pseudo-labels more reliable over time. Other methods, proposed by Pham et al. [16] and Yu et al. [26], also rely on input perturbations with specialized mechanisms, including transformation consistency, uncertainty weighting, or attention-based confidence, to refine how consistency regularization is applied.

Beyond perturbation-based consistency, multi-branch models enforce consistency across complementary task-specific outputs. For example, the method proposed by Li et al. [8] enforces agreement between pixel-wise segmentation and a shape-aware branch to enhance boundary accuracy. Similar multi-branch strategies have been explored in fully supervised settings as well Wan et al. [20], Nguyen et al. [14], where combining pixel-level and structure-level cues has been shown to enhance segmentation quality. Additionally, in semi-supervised domain adaptation, approaches like Ngo et al. [11] extend this idea by promoting consistency across domains to improve generalization under distribution shifts. Besides, Pseudo-labeling frameworks play a major role in SSL segmentation. PseudoSeg [28], CCT [15], CPS [3], and GTA-Seg [7] implement dual-branch co-training or teacher-student setups to improve pseudo-label stability. Unimatch [24] aligns predictions under different augmentation strengths. In the medical domain, DME-FD [13] addresses noise via dual-mask ensemble and feature discrepancy co-training.

Knowledge distillation (KD), first introduced by Hinton et al. [5], is a training approach where a student model learns from the predictions of a stronger teacher model. It has been widely adopted in semi-supervised medical image segmentation to transfer knowledge from well-trained models to weaker ones using unlabeled data. Recent extensions include collaborative teacher teams by Wang et al. [21] and mutual distillation between student networks by Xie et al. [23]. A growing trend is to distill knowledge from powerful foundation models. For

instance, Zhang et al. [27] explores using the Segment Anything Model (SAM) for semi-supervised segmentation by guiding SAM with domain-specific prompts and leveraging its outputs as pseudo-labels for training. Similarly, Huang et al. [6] proposes a Multi-view Co-training framework with a Learnable Prompt Strategy (LPS) and SAM-induced Knowledge Distillation (SKD), enabling sub-networks to adapt and learn from SAM's predictions while reducing noise from incorrect pseudo-labels.

While recent SSL methods have shown strong performance, many rely on large, complex models or struggle to produce accurate boundaries both of which are important for skin lesion segmentation. Newer approaches that use foundation models like SAM are promising, but they often miss the opportunity to guide SAM using rich multi-task cues such as coarse masks and boundary details. To address these challenges, we propose a lightweight co-training framework designed for clinical use. Our method combines a fused mask prompt with SAM to improve boundary accuracy and distills SAM's refined output into a more efficient main network.

In this work, we propose a semi-supervised skin lesion segmentation framework that integrates a lightweight Main Network with a foundation model (SAM) for enhanced supervision. Our main contributions are summarized as follows:

- **Lightweight Co-Training Framework:** We propose a dual-network training setup that pairs a lightweight Main Network (MobileNet) with a more powerful Teacher Network (ViT). This design enables the Main Network to run efficiently during inference, while still receiving strong supervision during training. It makes our framework practical for real-world deployment, particularly in resource-limited clinical settings.
- **Fused Mask Prompt for Boundary-Aware Guidance:** Inspired by consistency regularization across multi-task branches, we design a fused prompt that combines coarse semantic predictions with fine-grained boundary information. Specifically, we fuse the Main Network's initial segmentation mask and its corresponding Signed Distance Function (SDF) map into a single *mask prompt*. This unified representation captures both region-level confidence and structural contour cues, offering SAM a more holistic input for downstream refinement.
- **SAM-Guided Knowledge Distillation:** Leveraging the fused mask prompt, we introduce a SAM-guided distillation strategy. The task-aware prompt is passed into SAM, which produces a refined segmentation output. This output acts as a high-quality pseudo-label for training the Main Network on unlabeled data. By incorporating SAM's structured knowledge, the Main Network learns to generate more accurate and boundary-aligned predictions.

2 Methodology

Overview We adopt a semi-supervised learning framework for lesion segmentation, where the dataset is divided into two subsets: a labeled set $\mathcal{D}_l = \{(x_i, y_i)\}_{i=1}^{N_l}$, in which each image x_i is annotated with a corresponding segmentation mask

y_i; and an unlabeled set $\mathcal{D}_u = \{x_j\}_{j=1}^{N_u}$, which contains only raw input images without annotations.

Our architecture consists of three main components: (1) a lightweight **Main Network** based on MobileNet for efficient segmentation; (2) a powerful **Teacher Network**, built upon a ViT backbone and augmented with SAM, to generate high-quality pseudo-labels; and (3) a structured training pipeline that enables interaction between the two. As shown in Figure 1, the Main Network performs dual-task segmentation by predicting both binary masks and Signed Distance Function (SDF) maps. These outputs are fused to create task-aware prompts that guide the SAM module. SAM then refines the predictions and provides supervision signals back to the Main Network.

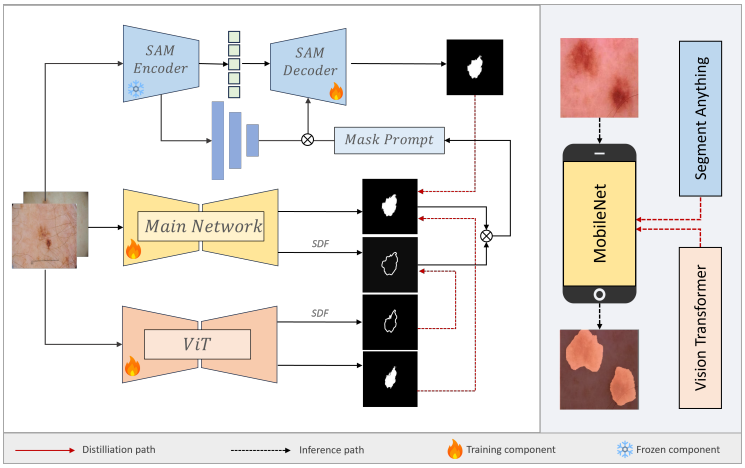

Fig. 1. Overview of our proposed semi-supervised segmentation framework. The Main Network (MobileNet) jointly predicts segmentation masks and Signed Distance Function (SDF) maps. These are fused into a mask prompt that guides the SAM decoder. SAM refines the predictions and provides high-quality pseudo-labels via knowledge distillation to train the Main Network. The Teacher Network (ViT) offers additional supervision, while only the Main Network is used at inference time for lightweight deployment.

2.1 Dual-Output Semi-Supervised Framework

Dual-Output Head with SDF Representation Following [2], both the Main and Teacher networks employ a dual-output prediction head. For each input image, they output:(1) a binary segmentation mask for identifying semantic regions, (2) a Signed Distance Function (SDF) map for capturing fine-grained boundary geometry.

The SDF representation encodes the distance of each pixel to the nearest object boundary. Given a binary mask Y, the SDF at each pixel x is defined as:

$$\phi_Y(x) = \begin{cases} -\min_{y \in \partial V_Y} \|x - y\|, & \text{if } x \in V_Y^{\text{in}}, \\ 0, & \text{if } x \in \partial V_Y, \\ +\min_{y \in \partial V_Y} \|x - y\|, & \text{if } x \in V_Y^{\text{out}}, \end{cases}$$

where ∂V_Y, V_Y^{in}, and V_Y^{out} denote the boundary, interior, and exterior of the segmentation region, respectively. This continuous representation enables the network to model object contours more accurately and reason about spatial structure.

Dual-Task Consistency and Knowledge Distillation To jointly learn semantic and geometric cues, the Main Network is trained with a dual-task objective: it predicts both a segmentation mask and an SDF map. This enforces consistency between region classification and boundary localization, encouraging the network to align its outputs both spatially and structurally. For unlabeled data, both the Main and Teacher networks process the same input image. The Teacher generates high-quality pseudo-labels consisting of a binary mask and an SDF map, which serve as supervision targets for the Main Network. This knowledge distillation process enables the Main Network to mimic the Teacher's predictions, benefiting from both region-based and geometry-aware guidance.

Fused Mask Prompt Generation. Once the Main Network produces its segmentation mask and SDF map, these two outputs are fused into a single *mask prompt*. The binary mask provides region-level confidence, while the SDF introduces detailed boundary information. This fused prompt forms a task-specific, enriched input used to guide the SAM decoder more effectively.

2.2 Learnable Prompt for SAM Using Fused Mask

Following Huang et al. [6] and Zhang et al. [27], the SAM decoder generates segmentation masks conditioned on prompts (e.g., points, boxes, masks) and image features. To enhance this process, we introduce a lightweight decoder $\psi(\cdot)$ that transforms the SAM encoder feature map $Z \in \mathbb{R}^{B \times D \times H \times W}$ into dense learnable prompts: $\mathbf{P}_b = \psi(Z; \Theta_m)$, where Θ_m are the decoder parameters and $\mathbf{P}_b \in \mathbb{R}^{B \times N_b \times L}$ represents the output prompt tokens.

To construct the final prompt for SAM, we fuse the Main Network's predicted binary mask and Signed Distance Function (SDF) map. This fused signal, together with $\mathbf{P}_b$, is fed to the SAM decoder: $\hat{Y}_s = \mathcal{F}_s(\mathbf{P}_b, \hat{Y}_f; \Theta_s)$,
where $\hat{Y}_f$ is the prior prediction and Θ_s denotes the decoder parameters.

To support this prompting strategy, we integrate an adapter module from [22] for decoder fine-tuning. The segmentation loss is defined as:

$$\mathcal{L}_{\text{sam}} = \mathcal{L}_{\text{SEG}}(\hat{Y}_s, Y^l). \tag{1}$$

2.3 SAM-guided knowledge distillation

Inspired by [6], we adopt the knowledge distillation framework from [5] to transfer supervision from SAM to the Main Network. Specifically, we use a temperature-scaled softmax to obtain softened probability maps:

$$\hat{Y}_T^c = \frac{\exp(\hat{q}_c/T)}{\sum_c \exp(\hat{q}_c/T)}, \tag{2}$$

where $\hat{q}_c$ is the SAM logit for class c, and T is a temperature parameter that controls the distribution's smoothness.

Soft predictions from both SAM (teacher) and the Main Network (student) are computed, and the distillation loss is defined via KL divergence:

$$\mathcal{L}_{\text{kd}} = KL(\hat{Y}_T, \hat{Y}_T^{\text{main}}), \tag{3}$$

where $\hat{Y}_T$ and $\hat{Y}_T^{\text{main}}$ are the temperature-softened outputs from SAM and the Main Network, respectively. The SAM model remains frozen, and gradients are only applied to the Main Network.

In summary, our method leverages semantic and geometric cues under a semi-supervised setup. The Main Network jointly predicts binary masks and SDF maps, supervised by pseudo-labels from the SAM-enhanced Teacher. The fused outputs form a prompt that guides SAM refinement, creating a feedback loop that improves learning efficiency. This design supports accurate segmentation with limited labels, while remaining lightweight and deployment-friendly.

3 Experiment

3.1 Dataset

We evaluated our framework on two widely used skin lesion segmentation datasets ISIC-2018 and HAM10000 under a semi-supervised learning setup. To simulate low-annotation scenarios, only 2%, 4%, or 8% of the training data was labeled, with the remainder treated as unlabeled. All experiments followed a 5-fold cross-validation protocol. The ISIC-2018 dataset [4] consists of 3,694 dermoscopic images, of which 2,955 were used for training and 739 for validation. The HAM10000 dataset [19] includes 10,015 images, split into 8,012 training and 2,003 validation samples.

3.2 Methods Under Comparison

To evaluate the effectiveness of our proposed lightweight SAM-assisted segmentation framework, we compare it against six representative state-of-the-art semi-supervised segmentation methods: PseudoSeg [28], CCT [15], CPS [3], GTA-Seg [7], Unimatch [24], and DME-FD [13]. In addition, we include comparisons with methods that incorporate SAM for semi-supervised segmentation, such as SemiSAM [27].

3.3 Evaluation Metrics

We evaluate segmentation performance using Dice, IoU, Sensitivity, and Specificity. Dice and IoU measure the overlap between predicted and ground truth masks. Sensitivity assesses how well lesion areas are detected, while Specificity indicates how well non-lesion areas are excluded.

4 Results

Table 1. Segmentation performance on ISIC-2018 under 2% and 4% labeled data settings. The SupOnly row reports results using fully supervised training.

Method	Data (%)		Metrics			
	Label	Unlabel	Dice (%) ↑	IoU (%) ↑	Sensitivity (%) ↑	Specificity (%) ↑
SupOnly	2%	0%	74.65 ±2.92	60.81 ±2.99	76.16 ±2.44	93.38 ±3.01
	4%	0%	77.23 ±0.48	65.35 ±0.56	79.53 ±1.37	95.10 ±0.59
	8%	0%	82.28 ±0.61	70.66 ±0.85	81.35 ±0.42	96.12 ±0.21
	100%	0%	87.66 ±0.93	78.49 ±1.38	87.11 ±1.05	96.90 ±0.35
PseudoSeg	2%	98%	79.76 ±2.11	67.16 ±2.77	76.65 ±3.72	96.26 ±0.83
CCT			78.66 ±2.02	65.80 ±2.63	77.17 ±4.15	95.56 ±0.69
CPS			79.61 ±1.66	67.04 ±2.28	78.43 ±4.64	95.52 ±1.04
GTA-Seg			77.33 ±2.20	64.21 ±2.59	80.04 ±3.87	93.37 ±2.00
Unimatch			80.03 ±2.04	67.55 ±2.71	78.46 ±4.74	95.84 ±1.50
DME-FD			80.07 ±1.75	67.62 ±2.37	78.97 ±3.51	95.69 ±0.58
Ours			**83.44 ±1.91**	**68.97 ±2.50**	**80.67 ±3.79**	**97.01 ±0.72**
PseudoSeg	4%	96%	81.77 ±0.66	71.18 ±1.03	81.98 ±3.13	96.37 ±0.85
CCT			80.96 ±1.11	68.95 ±1.41	79.75 ±1.68	95.93 ±0.28
CPS			80.89 ±0.91	70.31 ±1.07	82.08 ±2.35	95.67 ±0.96
GTA-Seg			80.83 ±0.80	70.03 ±1.07	82.54 ±2.35	94.64 ±1.46
Unimatch			81.41 ±1.22	69.46 ±1.58	79.50 ±1.76	96.34 ±0.56
DME-FD			82.06 ±0.69	71.54 ±1.04	82.87 ±1.58	96.23 ±0.36
Ours			**85.33 ±0.72**	**73.61 ±0.98**	**84.01 ±1.21**	**96.91 ±0.60**

Quantitative Results on ISIC-2018: Table 1 shows that our method achieves the highest Dice (83.44%, 85.33%) and IoU (68.97%, 73.61%) scores under both 2% and 4% labeled settings, outperforming all semi-supervised. It also leads in sensitivity and specificity, notably surpassing the fully supervised model trained with 8% labels, demonstrating its strength in low-label regimes.

Quantitative Results on HAM10000: As shown in Table 2, our method consistently outperforms semi-supervised baselines across both 2% and 4% label settings, achieving Dice scores of 91.02% and 91.56%, and IoU scores of 83.02% and 83.81%. It also records the highest sensitivity (90.02%) and specificity (97.82%), outperforming fully supervised baselines with more labeled data. Table 3 presents a broader comparison across three key categories: (1) the supervised baseline using Unet [17] trained on 100% labeled data, (2) the semi-supervised baseline DME-FD [13], and (3) the SAM-enhanced semi-supervised method, SemiSAM [27]. Despite using only 4% labeled data, our method achieves the highest Dice (85.33%) and IoU (73.61%) scores. At the same time, it maintains the smallest model size (2.9M parameters) and the fastest inference speed (0.02s). These results highlight that our approach not only delivers strong seg-

Table 2. Segmentation performance on HAM10000 under 2% and 4% labeled data settings. The SupOnly row reports results using fully supervised training.

Method	Data (%)		Metrics			
	Label	Unlabel	Dice (%) ↑	IoU (%) ↑	Sensitivity (%) ↑	Specificity (%) ↑
SupOnly	2%	0%	88.15 ±0.21	78.90 ±0.31	88.37 ±0.72	95.72 ±0.27
	4%	0%	89.59 ±0.07	81.24 ±0.12	88.58 ±0.97	96.80 ±0.40
	8%	0%	91.46 ±0.22	84.33 ±0.37	91.01 ±0.29	97.17 ±0.06
	100%	0%	93.54 ±0.25	87.92 ±0.42	93.30 ±0.07	97.80 ±0.22
PseudoSeg	2%	98%	90.02 ±0.17	81.94 ±0.28	88.18 ±1.32	97.29 ±0.51
CCT			89.93 ±0.10	81.79 ±0.15	88.54 ±0.86	97.09 ±0.30
CPS			89.94 ±0.14	81.81 ±0.23	87.95 ±0.48	97.35 ±0.28
GTA-Seg			89.55 ±0.32	81.17 ±0.54	88.89 ±0.70	96.60 ±0.09
Unimatch			89.66 ±0.15	81.35 ±0.26	87.89 ±0.38	97.15 ±0.31
DME-FD			90.45 ±0.17	82.65 ±0.27	88.74 ±0.83	97.39 ±0.44
Ours			**91.02 ±0.21**	**83.02 ±0.32**	**89.00 ±0.67**	**97.91 ±0.35**
PseudoSeg	4%	96%	90.97 ±0.39	83.21 ±0.64	89.11 ±0.77	97.49 ±0.45
CCT			90.64 ±0.53	82.97 ±0.86	89.08 ±0.94	97.39 ±0.15
CPS			90.76 ±0.51	83.17 ±0.84	89.20 ±0.54	97.44 ±0.19
GTA-Seg			90.86 ±0.19	83.34 ±0.31	89.74 ±0.69	97.24 ±0.32
Unimatch			90.32 ±0.44	82.43 ±0.73	88.93 ±1.29	97.20 ±0.61
DME-FD			91.13 ±0.30	83.79 ±0.50	**90.05 ±0.43**	97.33 ±0.08
Ours			**91.56 ±0.37**	**83.81 ±0.52**	90.02 ±0.60	**97.82 ±0.12**

Table 3. Comparison with SemiSAM methods on ISIC-2018, where SemiSAM denotes approaches that incorporate SAM for semi-supervised segmentation.

Method	Data (%)		Metrics		Params (M)	Speed (s)
	Label	Unlabel	Dice ↑	IoU ↑		
Unet (SupOnly)	100%	0%	0.7723 ± 0.0048	0.6535 ± 0.0056	7.8	0.8
DME-FD	4%	96%	0.8206 ± 0.0069	0.7154 ± 0.0104	41.4	1.2
SemiSAM			0.8412 ± 0.0110	0.7213 ± 0.0149	7.8	0.8
Ours			**0.8533 ± 0.0072**	**0.7361 ± 0.0098**	**2.9**	**0.02**

mentation accuracy, but also meets the requirements of lightweight, real-time deployment in clinical environments.

5 Conclusion

We present a semi-supervised segmentation framework that integrates lightweight networks with SAM for efficient and accurate lesion segmentation. By jointly predicting masks and SDF maps, the Main Network captures both semantic and boundary cues. These are fused into prompts to guide SAM, enabling a closed-loop distillation process that improves learning under limited labels. The design ensures strong performance while remaining practical for clinical deployment.

6 Acknowledgement

We would like to thank AI VIETNAM for financial and computational support.

References

1. Antonelli, M., Reinke, A., Bakas, S., Farahani, K., Kopp-Schneider, A., Landman, B.A., Litjens, G., Menze, B., Ronneberger, O., Summers, R.M., et al.: The medical segmentation decathlon. Nature communications **13**(1), 4128 (2022)
2. Bu, Q., Dong, B., Zhu, Z., Ni, J.: Edge enhancement based semi-supervised medical image segmentation method. In: 2025 2nd International Conference on Digital Image Processing and Computer Applications (DIPCA). pp. 39–43. IEEE (2025)
3. Chen, X., Yuan, Y., Zeng, G., Wang, J.: Semi-supervised semantic segmentation with cross pseudo supervision. In: Proceedings of the IEEE/CVF conference on computer vision and pattern recognition. pp. 2613–2622 (2021)
4. Codella, N.C., Gutman, D., Celebi, M.E., Helba, B., Marchetti, M.A., Dusza, S.W., Kalloo, A., Liopyris, K., Mishra, N., Kittler, H., et al.: Skin lesion analysis toward melanoma detection: A challenge at the 2017 international symposium on biomedical imaging (isbi), hosted by the international skin imaging collaboration (isic). In: 2018 IEEE 15th international symposium on biomedical imaging (ISBI 2018). pp. 168–172. IEEE (2018)
5. Hinton, G., Vinyals, O., Dean, J.: Distilling the knowledge in a neural network. arXiv preprint arXiv:1503.02531 (2015)
6. Huang, K., Zhou, T., Fu, H., Zhang, Y., Zhou, Y., Gong, C., Liang, D.: Learnable prompting sam-induced knowledge distillation for semi-supervised medical image segmentation. IEEE Transactions on Medical Imaging (2025)
7. Jin, Y., Wang, J., Lin, D.: Semi-supervised semantic segmentation via gentle teaching assistant. Advances in Neural Information Processing Systems **35**, 2803–2816 (2022)
8. Li, S., Zhang, C., He, X.: Shape-aware semi-supervised 3d semantic segmentation for medical images. In: Medical Image Computing and Computer Assisted Intervention–MICCAI 2020: 23rd International Conference, Lima, Peru, October 4–8, 2020, Proceedings, Part I 23. pp. 552–561. Springer (2020)
9. Li, X., Yu, L., Chen, H., Fu, C.W., Heng, P.A.: Semi-supervised skin lesion segmentation via transformation consistent self-ensembling model. arXiv preprint arXiv:1808.03887 (2018)
10. Luu, Q.V., Le, K.D., Nguyen, T.H., Nguyen, T.M., Nguyen, Q., Nguyen, T.T., Dinh, Q.V.: Semi-supervised semantic segmentation using redesigned self-training for white blood cells. In: 2025 IEEE 6th International Conference on Image Processing, Applications and Systems (IPAS). pp. 1–6. IEEE (2025)
11. Ngo, B.H., Lam, B.T., Nguyen, T.H., Dinh, Q.V., Choi, T.J.: Dual dynamic consistency regularization for semi-supervised domain adaptation. IEEE Access **12**, 36267–36279 (2024). https://doi.org/10.1109/ACCESS.2024.3374105
12. Nguyen, T.H., Ngo, T.K.N., Vu, M.A., Tu, T.Y.: Blurry-consistency segmentation framework with selective stacking on differential interference contrast 3d breast cancer spheroid. In: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. pp. 5223–5230 (2024)
13. Nguyen, T.H., Nguyen, T., Nguyen, X.B., Vu, N.L.V., Dinh, V.Q., MERIAUDEAU, F.: Semi-supervised skin lesion segmentation under dual mask ensemble with feature discrepancy co-training. In: Medical Imaging with Deep Learning
14. Nguyen, T.H., Vu, N.L.V., Nguyen, H.T., Dinh, Q.V., Li, X., Xu, M.: Semi-supervised histopathology image segmentation with feature diversified collaborative learning. In: AAAI Bridge Program on AI for Medicine and Healthcare. pp. 165–172. PMLR (2025)

15. Ouali, Y., Hudelot, C., Tami, M.: Semi-supervised semantic segmentation with cross-consistency training. In: Proceedings of the IEEE/CVF conference on computer vision and pattern recognition. pp. 12674–12684 (2020)
16. Pham, H.H., Nguyen, H.T., Vu, N.L.V., Dinh, Q.V., Nguyen, T.H., Li, X., Xu, M., et al.: Fetal-bcp: Addressing empirical distribution gap in semi-supervised fetal ultrasound segmentation. In: 2025 IEEE 22nd International Symposium on Biomedical Imaging (ISBI). pp. 1–4. IEEE (2025)
17. Ronneberger, O., Fischer, P., Brox, T.: 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. pp. 234–241. Springer (2015)
18. Sohn, K., Berthelot, D., Carlini, N., Zhang, Z., Zhang, H., Raffel, C.A., Cubuk, E.D., Kurakin, A., Li, C.L.: Fixmatch: Simplifying semi-supervised learning with consistency and confidence. Advances in neural information processing systems **33**, 596–608 (2020)
19. Tschandl, P., Rosendahl, C., Kittler, H.: The ham10000 dataset, a large collection of multi-source dermatoscopic images of common pigmented skin lesions. Scientific data **5**(1), 1–9 (2018)
20. Wang, Y., Wei, X., Liu, F., Chen, J., Zhou, Y., Shen, W., Fishman, E.K., Yuille, A.L.: Deep distance transform for tubular structure segmentation in ct scans. In: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. pp. 3833–3842 (2020)
21. Wang, Y., Cao, P., Hou, Q., Lan, L., Yang, J., Liu, X., Zaiane, O.R.: Progressively correcting soft labels via teacher team for knowledge distillation in medical image segmentation. In: International Conference on Medical Image Computing and Computer-Assisted Intervention. pp. 521–530. Springer (2024)
22. Wu, J., Fu, R., Fang, H., Liu, Y., Wang, Z., Xu, Y., Jin, Y., Arbel, T.: Medical sam adapter: Adapting segment anything model for medical image segmentation. arxiv 2023. arXiv preprint arXiv:2304.12620 (2023)
23. Xie, Y., Yin, Y., Li, Q., Wang, Y.: Deep mutual distillation for semi-supervised medical image segmentation. In: International Conference on Medical Image Computing and Computer-Assisted Intervention. pp. 540–550. Springer (2023)
24. Yang, L., Qi, L., Feng, L., Zhang, W., Shi, Y.: Revisiting weak-to-strong consistency in semi-supervised semantic segmentation. In: Proceedings of the IEEE/CVF conference on computer vision and pattern recognition. pp. 7236–7246 (2023)
25. Yang, L., Zhuo, W., Qi, L., Shi, Y., Gao, Y.: St++: Make self-training work better for semi-supervised semantic segmentation. In: Proceedings of the IEEE/CVF conference on computer vision and pattern recognition. pp. 4268–4277 (2022)
26. Yu, L., Wang, S., Li, X., Fu, C.W., Heng, P.A.: Uncertainty-aware self-ensembling model for semi-supervised 3d left atrium segmentation. In: Medical image computing and computer assisted intervention–MICCAI 2019: 22nd international conference, Shenzhen, China, October 13–17, 2019, proceedings, part II 22. pp. 605–613. Springer (2019)
27. Zhang, Y., Yang, J., Liu, Y., Cheng, Y., Qi, Y.: Semisam: Enhancing semi-supervised medical image segmentation via sam-assisted consistency regularization. In: 2024 IEEE International Conference on Bioinformatics and Biomedicine (BIBM). pp. 3982–3986. IEEE (2024)
28. Zou, Y., Zhang, Z., Zhang, H., Li, C.L., Bian, X., Huang, J.B., Pfister, T.: Pseudoseg: Designing pseudo labels for semantic segmentation. arXiv preprint arXiv:2010.09713 (2020)