DiCoP-SAM: Disagreement-Band Guided Corrective Prompting for Collaborative Semi-Supervised Medical Image Segmentation

Xinyu Wang[1,4], Jiayao Liu[1], Youlin Wang[1], Jiawei Su[2], Zhiming Luo[3], and Sheng Lian[1,4(✉)]

[1] College of Computer and Data Science, Fuzhou University, Fuzhou, China
[2] College of Computer Engineering, Jimei University, Xiamen, China
[3] Department of Artificial Intelligence, Xiamen University, Xiamen, China
[4] Engineering Research Center of Big Data Intelligence, Ministry of Education, Fuzhou, China
✉Correspondences: shenglian@fzu.edu.cn

Abstract. Semi-supervised medical image segmentation (SSMIS) is critical, yet suffers from pseudo-label noise and boundary uncertainty in low-label regimes. While foundation models like SAM offer strong priors, their high-confidence semantic bias can amplify errors when naively integrated into pseudo-supervision loops. We propose *DiCoP-SAM*, a two-stage collaborative framework that closes the loop between a CNN and SAM via automatic, error-aware corrective prompting that requires neither GT masks nor manual prompts for unlabeled images. Stage-I (TriGate-CPS) augments cross pseudo-supervision with a foreground/background/uncertainty tri-gate mechanism to produce reliable fused pseudo-labels and explicitly localize high-risk disagreement bands. Stage-II (DiCoP) harvests these bands to generate such corrective positive/negative point prompts, driving parameter-efficient SAM adaptation; refined predictions are distilled back to the CNN, forming a self-correcting "predict → correct → relearn" loop. On CT, MRI, and ultrasound benchmarks, DiCoP-SAM consistently outperforms semi-supervised and SAM-assisted baselines, approaching fully supervised performance with limited labels (e.g., 83.13 Dice vs. 83.30 for TRFE+ on TN3K with 20% labels).

Keywords: Semi-supervised segmentation · SAM prompt learning · Corrective point prompts.

1 Introduction

Medical image segmentation (MIS) is a cornerstone for computer-aided diagnosis, treatment planning, and quantitative analysis [24, 3]. However, dense pixel-wise annotations are expensive and time-consuming [7]. Real-world medical datasets often follow a "few labeled+many unlabeled" paradigm, further complicated by distribution shifts across imaging modalities, devices, or acquisition protocols. This makes robust segmentation under limited annotation a persistent challenge.

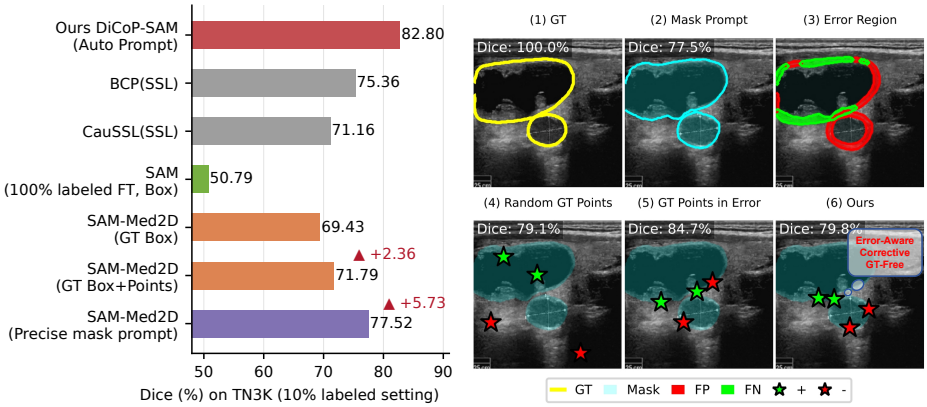

Fig. 1. Motivation for error-aware corrective point prompting. (a) Directly deploying promptable foundation models in SSMIS exhibits a clear performance bottleneck; even medically adapted SAM variants remain highly sensitive to prompt forms. (b) Under same mask prompt, different point strategies lead to markedly different masks and Dice.

Semi-supervised learning (SSL) addresses this through consistency regularization or pseudo-labeling, yet medical images often exhibit fuzzy boundaries and complex anatomy, leading to unreliable pseudo-labels that amplify during training, undermining both training stability and generalization [12, 16, 20].

Recently, promptable segmentation foundation models have opened new possibilities for SSL, motivating "Specialist (CNN) + Generalist (SAM)" collaborative paradigms [11, 1, 22]. Despite their promise, foundation models can still produce *high-confidence yet semantically wrong* predictions on medical images [13]. If such errors enter a closed-loop "prompt refinement → pseudo supervision" pipeline, they may be repeatedly reinforced and propagate through training. Fig. 1(a) shows a typical issue of directly applying SAM in medical workflows: Even with ground truth (GT) bounding-box prompts and full fine-tuning, the performance may still lag behind a strong SSL model [12, 26]. Meanwhile, Fig. 1(a, b) shows that medical-adapted SAM variants can be highly sensitive to prompt types, where multi-point prompts may significantly boost performance compared with box-only prompts [25, 6]. This implies that points are more than "extra localization". When they carry *error-corrective* semantics, they can unlock stronger refinement capabilities. Unfortunately, accurate point/box prompts are hard to obtain in real clinical pipelines, and existing CNN+SAM frameworks often lack a mechanism to automatically generate *corrective prompts* that explicitly target potential errors without relying on GT masks or manual prompts [9, 16, 27]. These issues raise a critical problem: ***How to automatically generate error-aware corrective prompts and unlock SAM's full potential in SSMIS?***

To address this issue, we propose ***DiCoP-SAM:*** a two-stage semi-supervised collaborative framework that enables mutual error correction between CNN and

SAM through automatic, error-aware corrective prompting. In *Stage-I*, *TriGate-CPS* enhances cross pseudo-supervision (CPS) [5] with foreground, background, and uncertainty gating to adaptively fuse branch predictions and explicitly localize *disagreement bands* where prediction conflicts signal high error risk. In *Stage-II*, *DiCoP* harvests these bands to automatically generate corrective positive/negative point prompts; together with mask prompts, these drive parameter-efficient SAM adaptation. Refined predictions are distilled back to the CNN, forming a self-correcting "predict → correct → relearn" loop. The contributions are fourfold:

- Proposing DiCoP-SAM: breaking the error amplification loop in SSMIS by establishing a "CNN ↔ SAM" collaborative paradigm with automatic, error-aware prompting;
- We improve pseudo-label reliability and enable targeted error correction via TriGate-CPS, which explicitly localizes high-risk disagreement regions through tri-gate uncertainty modeling;
- We automate corrective prompting via DiCoP, converting disagreement bands into sparse error-aware points without manual prompts, driving parameter-efficient SAM adaptation with bidirectional distillation;
- Extensive experiments across three modalities establish promising results, outperforming both SSL specialists and SAM variants under low-label regimes.

2 Method

Task Definition. SSMIS aims to learn from a small labeled set $\mathcal{D}_L = \{(\mathbf{x}_i, \mathbf{y}_i)\}_{i=1}^{N_l}$ and a large unlabeled set $\mathcal{D}_U = \{\mathbf{x}_j\}_{j=1}^{N_u}$. Here $\mathbf{x} \in \mathbb{R}^{H\times W\times K}$ denotes the input image and $\mathbf{y} \in \{0,1\}^{H\times W\times C}$ is the C-class one-hot annotation.
Method Overview. DiCoP-SAM establishes a self-correcting "predict → correct → relearn" loop between a lightweight CNN and SAM. *Stage-I (TriGate-CPS)* enhances CPS with foreground/background/uncertainty gating, and localizes *disagreement bands* where dual-branch conflicts signal high error risk. *Stage-II (DiCoP+SAM-APT)* harvests these bands to automatically generate corrective positive/negative point prompts without manual prompts, driving parameter-efficient SAM adaptation; refined predictions are distilled back to the CNN, closing a self-correcting semi-supervised loop.

2.1 Stage-I: TriGate-CPS

Stage-I employs complementary branches A, B to output logits $\mathbf{z}^A, \mathbf{z}^B \in \mathbb{R}^{C\times H\times W}$ and predictions $\mathbf{p}^A, \mathbf{p}^B$. Normalized entropy $\mathcal{H}(\mathbf{p})$ quantifies uncertainty.

Tri-Gated Fusion. Vanilla CPS enforces inter-branch consistency but lacks explicit reliability modeling. We attach lightweight gating heads predicting three pixel-wise gates per branch: foreground $\mathbf{M}_{\text{fg}}$, background $\mathbf{M}_{\text{bg}}$, and uncertainty $\mathbf{M}_{\text{uc}} \in [0,1]^{H\times W}$. Foreground/background gates capture spatial priors, while the uncertainty gate highlights high-risk regions (boundaries, structural ambiguities)

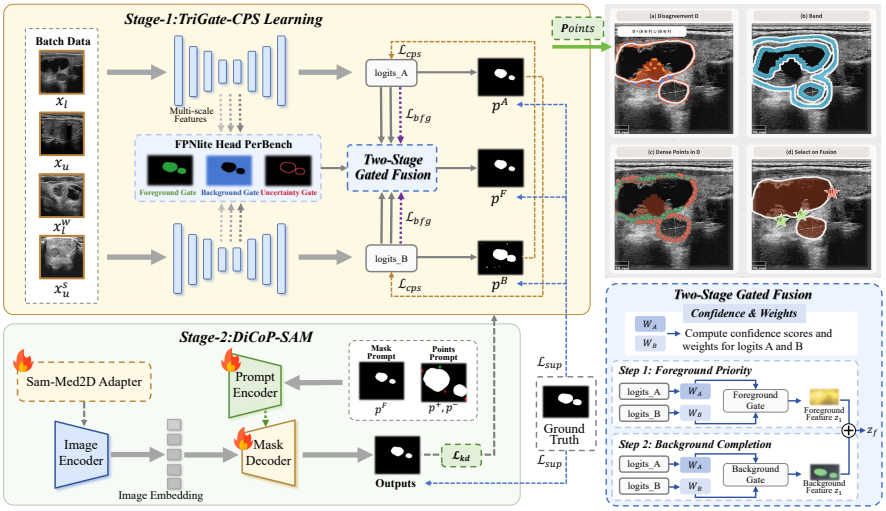

Fig. 2. DiCoP-SAM overview. ***Stage-I (TriGate-CPS):*** dual-branch CNN with tri-gate fusion generates reliable predictions and localizes *disagreement bands*. ***Stage-II (DiCoP+SAM-APT):*** disagreement-guided point prompts drive parameter-efficient SAM adaptation; refined predictions are distilled back to CNN to close self-correcting loop.

to suppress their negative contribution in fusion. Per-pixel fusion weights ($w^A + w^B = 1$) combine prediction entropy and uncertainty:

$$w^i = \frac{(1 - \mathcal{H}(\mathbf{p}^i)) \odot (1 - \mathbf{M}^i_{\text{uc}})}{\sum_j (1 - \mathcal{H}(\mathbf{p}^j)) \odot (1 - \mathbf{M}^j_{\text{uc}}) + \epsilon}, \quad i, j \in \{A, B\}, \tag{1}$$

where $\odot$ is pixel-wise multiplication, ϵ is a stability term. The foreground term is filtered by $\mathbf{M}_{\text{fg}}$ and refined by the channel-mixing operator $\phi(\cdot)$, which is implemented as a lightweight two-layer 1×1 convolutional block with a nonlinear activation to recalibrate channels and refine foreground structures. The background term is completed by $\mathbf{M}_{\text{bg}}$ to suppress uncertain regions and improve background consistency. The fused logits follow Eq. 2:

$$\mathbf{z}^F = \phi\left(\sum_i (w^i \odot \mathbf{M}^i_{\text{fg}} \odot \mathbf{z}^i)\right) + \sum_i (w^i \odot \tilde{\mathbf{M}}^i_{\text{bg}} \odot \mathbf{z}^i), \quad i \in \{A, B\}, \tag{2}$$

where $\tilde{\mathbf{M}}^i_{\text{bg}}$ is the channel-wise background gate ($\mathbf{M}^i_{\text{bg}}$ for background; $(1-\mathbf{M}^i_{\text{fg}}) \odot \mathbf{M}^i_{\text{bg}}$ for foreground classes). This yields fused prediction $\mathbf{p}^F = \text{Softmax}(\mathbf{z}^F)$.

Training Objectives. On labeled data, foreground/background gates are supervised by GT masks $\mathbf{y}_{\text{fg}}, \mathbf{y}_{\text{bg}}$. For the uncertainty gate, we construct a narrow boundary-band target from the GT foreground mask by subtracting its eroded region from its dilated region. This target marks high-risk boundary and structural variation regions, and is used to supervise $\mathbf{M}_{\text{uc}}$. *This tri-gate supervision is*

implemented by a lightweight boundary-focused guidance loss $\mathcal{L}_{\text{bfg}}$: we apply (binary) CE+Dice to $\mathbf{M}_{\text{fg}}$ and $\mathbf{M}_{\text{bg}}$ against $\mathbf{y}_{\text{fg}}$ and $\mathbf{y}_{\text{bg}}$, and use BCE to supervise $\mathbf{M}_{\text{uc}}$ with the boundary-band target. We apply the supervised segmentation loss $\mathcal{L}_{\text{seg}}$ (CE+Dice) to branch A, B, and the fused output; i.e., $\mathcal{L}_{\text{sup}}^{\text{CNN}}$ is the sum of three $\mathcal{L}_{\text{seg}}$ terms computed on these outputs against the GT. For unlabeled data, we adopt standard CPS: each branch uses the other branch's hard prediction as a pseudo-label, yielding the consistency loss $\mathcal{L}_{\text{cps}}$.

$$\mathcal{L}_{\text{Stage1}} = \mathcal{L}_{\text{sup}}^{\text{CNN}} + \lambda_{\text{cps}}\,\mathcal{L}_{\text{cps}} + \lambda_{\text{bfg}}\,\mathcal{L}_{\text{bfg}}. \tag{3}$$

2.2 Stage-II: Disagreement-guided SAM Adaptation

Stage-II adapts SAM to the target domain via parameter-efficient prompt-conditioned fine-tuning (SAM-APT), freezing the image encoder while updating only the prompt encoder and mask decoder. The Stage-I fused prediction serves as the mask prompt, while DiCoP automatically generates corrective point prompts around disagreement bands without using GT masks or manual prompts for unlabeled images.

DiCoP: Corrective Point Generation. Disagreement bands indicate high-risk regions where branch predictions conflict with the fusion. We dilate these conflict zones to obtain a narrow sampling band, then score candidate points by reliability and proximity to disagreements.

Specifically, let $\hat{\mathbf{p}}^A, \hat{\mathbf{p}}^B, \hat{\mathbf{p}}^F \in \{0,1\}^{H\times W}$ denote binary predictions. We mark *disagreement* pixels where any branch $\hat{\mathbf{p}}^A, \hat{\mathbf{p}}^B$ conflicts with the fused decision $\hat{\mathbf{p}}^F$, and dilate the disagreement set to obtain a narrow *disagreement band* as the local candidate space for point prompting. For pixel i with foreground probabilities $p^A(i), p^B(i), p^F(i)$, we define:

Reliable positives (fusion foreground with high confidence in both branches):

$$s^+(i) = \frac{p^A(i) + p^B(i) + p^F(i)}{3} \cdot \varphi\big(d(i,\mathcal{D})\big), \tag{4}$$

Strict negatives (branches confident as "other" while fusion is hesitant):

$$s^-(i) = \frac{(1-p^A(i)) + (1-p^B(i))}{2} \cdot \mathbb{I}\big[p^F(i) \in [\tau_l, \tau_h]\big], \tag{5}$$

where $\varphi(\cdot)$ is a distance-aware weighting function that assigns higher scores to candidates closer to the disagreement set $\mathcal{D}$. During training, we randomly sample the point budget $K \in \{1,3,5,7\}$ and apply spatially-suppressed Top-K selection to avoid clustered prompts. Specifically, candidate points are sorted by their scores in descending order and are greedily selected only when their distance to all previously selected points is larger than the suppression radius r. The selection stops when K points are selected or no valid candidate remains. For labeled images, candidates whose assigned prompt labels conflict with GT labels are discarded before selection. Mask and point prompts are fed into SAM to produce $\mathbf{z}^{\text{SAM}}$.

Training Objectives. On labeled data, SAM is supervised by $\mathcal{L}_{\text{seg}}$ (CE+Dice). On unlabeled data, SAM serves as teacher via two knowledge distillation (KD) paths:
(1) Branch-level KD: Soft targets from $\mathbf{z}^{\text{SAM}}$ distill to both CNN branches:

$$\mathcal{L}_{\text{KD}} = \sum_{t\in\{A,B\}} T^2 \cdot \text{KL}\big(\text{softmax}(\mathbf{z}^{\text{SAM}}/T) \,\|\, \text{softmax}(\mathbf{z}^{t}/T)\big). \tag{6}$$

(2) Fusion-level guidance: SAM predictions regularize the fused output. This enables TriGate to indirectly absorb SAM's structural priors, improving uncertainty modeling on high-risk regions.

Stage-II objective combines supervised SAM loss, distilling, and Stage-I terms:

$$\mathcal{L}_{\text{Stage2}} = \mathcal{L}_{\text{Stage1}} + \lambda_{\text{sam}}\mathcal{L}_{\text{seg}}(\mathbf{p}^{\text{SAM}}, \mathbf{y}) + \lambda_{\text{kd}}\mathcal{L}_{\text{KD}} + \lambda_{\text{fuse}}\mathcal{L}_{\text{seg}}(\mathbf{p}^{F}, \mathbf{p}^{\text{SAM}}). \tag{7}$$

3 Experimental Configurations

Datasets. We evaluate our approach on three representative datasets spanning various imaging modalities: thyroid ultrasound (TN3K / TG3K [8]), cardiac cine-MRI (ACDC [4]), and a 3D cardiac CT dataset (EAT). TN3K contains 3,493 images (2,879 for training and 614 for testing), and TG3K provides 3,585 frames extracted from 16 ultrasound videos. Following the TRFE^{+} protocol [8], we train on TN3K-train∪TG3K and report results on TN3K-test under semi-supervised settings with 10% and 20% labeled data. For ACDC, we adopt the standard patient-wise split (70/10/20) and sample 7 (10%) or 14 (20%) labeled subjects from the training set. EAT is an in-house 3D cardiac CT dataset comprising 97 fully anonymized scans with LV, RV, and peri-atrial Epicardial Adipose Tissue (EAT) annotations, with ethics approval (No: MRCTA, ECFAH of FMU [2021]072) and informed consent obtained. EAT is split into 7:1:2 for train/val/test, corresponding to 4,552/651/1,300 slices, respectively.
Implementations. (1) Architecture. We employ dual CNN branches with U-Net [18] and V-Net [17] backbones equipped with lightweight FPN-lite TriGate heads. SAM-Med2D (ViT-B) [6] is initialized at 256×256 resolution. ***(2) Hyperparameters.*** Point prompts are sampled within the disagreement band ($\text{band}_k = 21$) using a safe-positive/strict-negative scheme with confidence thresholds $\tau_{\text{pos}} = 0.95$ and $\tau_{\text{neg}} = 0.90$. The hesitation interval in Eq. 5 is set to $[\tau_l, \tau_h] = [0.45, 0.75]$; points are selected by *SpatialTopK* with a fixed suppression radius $r = 6$. The CNN branch is optimized with SGD and linear learning-rate decay, while the SAM branch uses AdamW under the same schedule. The consistency regularization weight is warmed up via a sigmoid-rampup strategy, and the distillation temperature is set to $T = 10$. In the late training stage, we apply entropy-guided patch-level CutMix [23] using SAM pseudo labels for robustness on hard regions.
(3) Metrics. We report DSC and IoU (↑) as region metrics and 95HD/ASD (↓) as boundary metrics; ASD is additionally reported for the 3D datasets (ACDC and EAT).

Table 1. Comparison of DiCoP-SAM with representative SSL solutions and recent SAM-based methods on TN3K (ultrasound), ACDC (MRI), and EAT (CT).

Method	#anno.	TN3K			ACDC				EAT			
		DSC↑	IoU↑	95HD↓	DSC↑	IoU↑	95HD↓	ASD↓	DSC↑	IoU↑	95HD↓	ASD↓
TRFE+ [8]/Unet [18]	100%	83.30	71.38	13.23	91.65	84.93	1.89	0.56	69.50	54.02	6.53	1.25
SAM-Med2D [6] (w./ GT)	0%	69.43	56.86	5.43	77.36	64.32	5.40	1.96	56.58	40.27	25.34	6.79
MT [19]	10%	73.09	62.95	4.95	81.11	69.99	8.99	2.70	59.45	43.59	17.63	4.66
CPS [5]	10%	73.76	63.56	4.92	84.24	73.91	8.26	2.45	63.38	47.92	16.31	3.35
CauSSL-CPS [14]	10%	71.16	60.41	5.46	85.25	75.31	6.05	1.97	63.95	48.46	16.02	3.21
MC-Net [21]	10%	71.45	60.72	5.28	86.07	76.58	11.48	3.37	62.18	46.34	18.85	5.70
BCP [2]	10%	75.36	64.65	5.04	88.84	80.62	3.98	**1.17**	65.40	49.54	15.80	3.44
URPC [10]	10%	74.42	64.17	4.88	82.41	71.69	5.83	1.65	60.45	44.66	15.07	3.05
Semi-SAM [27]	10%	66.58	56.03	4.82	84.77	75.05	7.43	2.21	61.83	46.89	16.24	4.32
CPC-SAM [15]	10%	75.02	64.87	4.49	86.28	76.13	4.15	2.50	64.27	48.01	8.79	2.94
DiCoP-SAM(Ours)	10%	**82.80**	**72.14**	**4.43**	**89.02**	**81.14**	**2.57**	1.57	**65.87**	**50.45**	**7.81**	**1.67**
MT [19]	20%	77.12	67.66	4.67	85.46	75.89	8.02	2.39	66.62	51.39	12.72	5.63
CPS [5]	20%	76.97	67.38	4.69	86.85	77.96	5.48	1.64	67.08	51.95	10.85	2.45
CauSSL-CPS [14]	20%	73.48	64.95	5.14	87.24	78.44	5.57	1.73	67.18	52.10	9.88	2.01
MC-Net [21]	20%	76.26	65.47	4.85	86.58	77.68	15.99	4.93	66.87	51.63	6.35	1.89
BCP [2]	20%	75.91	64.21	4.91	89.52	81.62	3.69	1.03	67.43	52.12	10.62	2.51
URPC [10]	20%	76.81	67.09	4.72	85.44	76.36	5.93	1.70	66.95	51.60	9.44	**1.69**
Semi-SAM [27]	20%	68.74	59.65	4.77	86.36	76.99	2.56	1.87	64.53	49.31	8.26	2.32
CPC-SAM [15]	20%	76.71	66.39	**4.64**	88.41	79.86	3.35	1.54	66.79	51.09	8.76	2.86
DiCoP-SAM(Ours)	20%	**83.13**	**73.06**	4.78	**90.28**	**83.67**	**1.26**	**0.35**	**68.86**	**53.77**	**6.24**	1.70

4 Experimental Results

Comparisons with Promising Solutions. Table 1 compares DiCoP-SAM with conventional SSL baselines and recent SAM-based methods on TN3K, ACDC, and EAT. On **TN3K**, DiCoP-SAM achieves SOTA performance under 10%/20% labels (82.80/83.13 DSC) and is close to the fully-supervised TRFE+ [8] result (83.30 DSC), indicating that our prompt construction and incorporated priors can substantially narrow the gap to full supervision. On **ACDC**, DiCoP-SAM attains the best overall performance at 20% labels (90.28 DSC, 1.26 95HD, 0.35 ASD). Under 10% labels, while DiCoP-SAM maintains strong region-level metrics, accurately characterizing complex cine-MRI structures remains challenging with limited supervision. On **EAT**, DiCoP-SAM achieves the best performance at both label ratios and shows more stable region-level results, supporting its robustness across modalities and 3D anatomical structures.

Overall, these results are consistent with our self-correcting loop: Stage-I (Tri-Gate) provides more stable pseudo-labels for constructing prompts; although its gains can be limited under low labels and structure-sensitive cases, incorporating SAM-Med2D in Stage-II injects priors of organ shape and structural regularities, improving prompt quality and assisting mask prediction. Moreover, the refined SAM predictions can be fed back to the CNN branch via pseudo-label supervision or distillation, forming a closed-loop iterative refinement process. The qualitative comparisons in Fig. 3 further corroborate these findings, especially in cases with multiple nodules and in cardiac adhesion regions where structures are ambiguous.

Ablation Studies. We conduct ablation studies on TN3K with 10% labels to validate:

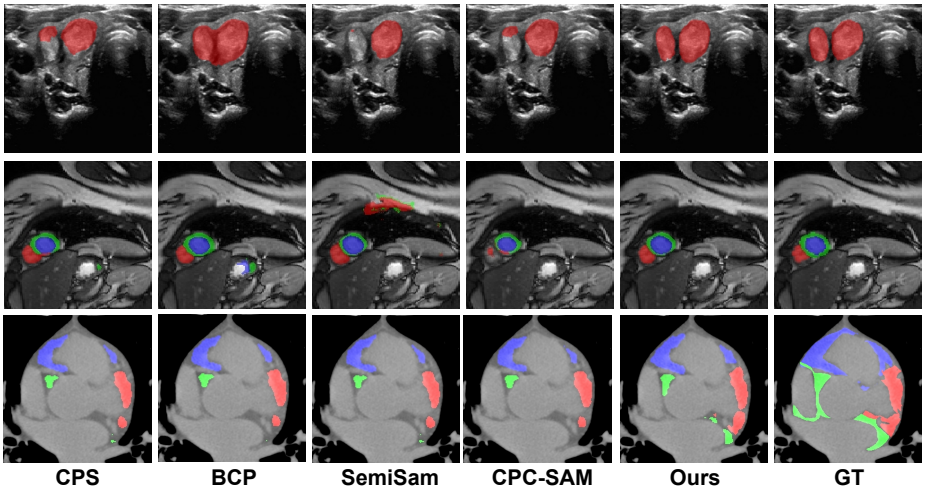

Fig. 3. Qualitative comparisons of different methods on TN3K (top), ACDC (middle), and EAT (bottom), all with 10% labeled data.

Table 2. Component ablation (TN3K, 10%)

Variant	TG	MP	PC	Dice↑	95HD↓
CPS [5]				73.76	4.92
TG-CPS	✓			77.21	4.90
CPS+MP		✓		80.31	5.28
TG-CPS+MP	✓	✓		81.67	4.76
DiCoP-SAM(full)	✓	✓	✓	**82.80**	**4.43**

Table 3. Point choice ablation (TN3K, 10%)

Str.	Band	MinD	Rule	Dice↑	95HD↓
Ent			Ent	80.67	5.23
R@H			Rand	81.68	4.80
DiCoP	✓		S+/S-	82.48	5.15
DiCoP		✓	S+/S-	81.75	4.84
Full	✓	✓	S+/S-	**82.80**	**4.43**

(1) Key component design (Table 2). TG denotes TriGate-enhanced CPS in Stage-I; MP denotes SAM-Med2D integration via mask-prompt distillation (mask only); PC denotes point-based prompt correction.

Observations: TriGate consistently improves over vanilla CPS (73.76→77.21 Dice). Adding SAM priors (MP) further boosts performance (80.31 Dice), but mask-only prompting shows limited boundary refinement (95HD: 5.28). Enabling point correction (PC) achieves the best result (82.80 Dice, 4.43 95HD), validating that sparse corrective points targeting disagreement regions are essential for precise boundary refinement.

(2) Strategy on point prompt choice (Table 3). *Band*: sampling within disagreement band; *MinD*: minimum-distance constraint; *Rule*: point labeling strategy (S+/S-: safe-positive/strict-negative; Ent: entropy-based; Rand: random high-entropy sampling). *R@H*: random positive & negative points with high-entropy.

Observations: Disagreement-band S+/S- sampling (82.48 Dice) outperforms entropy-based (80.67) and random (81.68) strategies, confirming that explicit error-region targeting surpasses uncertainty-only heuristics. Adding the minimum-distance constraint (Full: 82.80 Dice, 4.43 95HD) yields the best trade-off by ensuring spatial diversity of corrective signals.

5 Conclusion

We presented DiCoP-SAM, a two-stage collaborative framework that incorporates SAM into semi-supervised medical image segmentation through automated corrective prompting. By using disagreement-guided prompts, the framework improves pseudo-label quality and boundary refinement under limited annotations. Experiments on ultrasound, MRI, and CT datasets demonstrate consistent gains over representative SSL and SAM-based methods, narrowing the gap to full supervision. A limitation of the current design is that point selection still relies on heuristic thresholds. Future work will explore adaptive prompt selection and extension to fully 3D foundation models.

Acknowledgments. This work is supported by the National Natural Science Foundation of China (No.62501160), the Fujian Provincial Natural Science Foundation of China (No.2023J05117, 2025J08200), the Fujian Provincial Department of Education General Project (No. JAT241056). The authors would also like to acknowledge the support from Fujian Key Laboratory of Network Computing and Intelligent Information Processing (Fuzhou University) and the Key Laboratory of Intelligent Metro of Universities in Fujian, Fuzhou University, Fuzhou, China.

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

References

1. Ali, M., Wu, T., Hu, H., et al.: A review of the segment anything model (sam) for medical image analysis: Accomplishments and perspectives. CMIG (2025)
2. Bai, Y., Chen, D., Li, Q., Shen, W., Wang, Y.: Bidirectional copy-paste for semi-supervised medical image segmentation. In: CVPR (2023)
3. Benedykciuk, E., Denkowski, M., Wójcik, G.M.: Differentiable neural architecture search for medical image segmentation: A systematic review and field audit. CMIG (2026)
4. Bernard, O., Lalande, A., Zotti, C., Cervenansky, F., Yang, X., Heng, P.A., Cetin, I., Lekadir, K., Camara, O., Ballester, M.A.G., et al.: Deep learning techniques for automatic mri cardiac multi-structures segmentation and diagnosis: is the problem solved? IEEE TMI (2018)
5. Chen, X., Yuan, Y., Zeng, G., Wang, J.: Semi-supervised semantic segmentation with cross pseudo supervision. In: CVPR (2021)
6. Cheng, J., Ye, J., Deng, Z., Chen, J., Li, T., Wang, H., Su, Y., Huang, Z., Chen, J., Jiang, L., et al.: Sam-med2d. arXiv:2308.16184 (2023)
7. Dissanayake, T., George, Y., Mahapatra, D., Sridharan, S., Fookes, C., Ge, Z.: Few-shot learning for medical image segmentation: A review and comparative study. ACM Computing Surveys (2025)
8. Gong, H., Chen, J., Chen, G., Li, H., Li, G., Chen, F.: Thyroid region prior guided attention for ultrasound segmentation of thyroid nodules. CIBM (2023)
9. 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 TMI (2025)

10. Luo, X., Chen, J., Song, T., Wang, G.: Semi-supervised medical image segmentation through dual-task consistency. In: AAAI (2021)
11. Ma, J., He, Y., Li, F., Han, L., You, C., Wang, B.: Segment anything in medical images. Nature Communications (2024)
12. 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 (2025)
13. Mazurowski, M.A., Dong, H., Gu, H., Yang, J., Konz, N., Zhang, Y.: Segment anything model for medical image analysis: An experimental study. MedIA (2023)
14. Miao, J., Chen, C., Liu, F., Wei, H., Heng, P.A.: Caussl: Causality-inspired semi-supervised learning for medical image segmentation. In: ICCV (2023)
15. Miao, J., Chen, C., Yuan, Y., Li, Q., Heng, P.A.: Sam-driven cross prompting with adaptive sampling consistency for semi-supervised medical image segmentation. MedIA (2026)
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 (2024)
17. Milletari, F., Navab, N., Ahmadi, S.A.: V-net: Fully convolutional neural networks for volumetric medical image segmentation. In: 3DV (2016)
18. Ronneberger, O., Fischer, P., Brox, T.: U-net: Convolutional networks for biomedical image segmentation. In: MICCAI (2015)
19. Tarvainen, A., Valpola, H.: Mean teachers are better role models: Weight-averaged consistency targets improve semi-supervised deep learning results. In: NeurIPS (2017)
20. Wen, Q., He, J., Wei, X., Lian, S., Lin, X., Luo, Z.: Eutnet: Edge uncertainty-aware consistency and topology-constrained semi-supervised medical image segmentation. BSPC (2026)
21. Wu, Y., Ge, Z., Zhang, D., Xu, M., Zhang, L., Xia, Y., Cai, J.: Mutual consistency learning for semi-supervised medical image segmentation. MedIA (2022)
22. Xiong, S., Wu, L., Zhang, B., Zhang, D., Tao, Y., Tang, Y.: Hl-sam-seg: Complementary high-and low-resolution features based on sam for remote sensing image semantic segmentation. IEEE TGRS (2026)
23. 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: MICCAI (2019)
24. Zhang, J., Chen, X., Yang, B., Guan, Q., Chen, Q., Chen, J., Wu, Q., Xie, Y., Xia, Y.: Advances in attention mechanisms for medical image segmentation. Computer Science Review (2025)
25. Zhang, Y., Hu, S., Xue, L., Ren, S., Hu, Z., Cheng, Y., Qi, Y.: Enhancing the reliability of auto-prompting sam for medical image segmentation with uncertainty estimation and rectification. In: Proceedings of the IEEE/CVF International Conference on Computer Vision (2025)
26. Zhang, Y., Lv, B., Xue, L., Zhang, W., Liu, Y., Fu, Y., Cheng, Y., Qi, Y.: Semisam+: Rethinking semi-supervised medical image segmentation in the era of foundation models. MedIA (2025)
27. Zhang, Y., Yang, J., Liu, Y., Cheng, Y., Qi, Y.: Semisam: Enhancing semi-supervised medical image segmentation via sam-assisted consistency regularization. In: BIBM (2024)