JAM: Unlocking SAM2 without Training and Prompts via Medical-Aware Mask Selection

Jeongeun Kim
Department of Computer Science
Yonsei University
Seoul, Republic of Korea
wjddms2216@yonsei.ac.kr

Youngwan Jo
Department of Computer Science
Yonsei University
Seoul, Republic of Korea
jyy1551@yonsei.ac.kr

Sanghyun Park*
Department of Computer Science
Yonsei University
Seoul, Republic of Korea
sanghyun@yonsei.ac.kr

***Abstract*—Foundation models allow zero-shot transfer, but SAM struggles on medical images where fine anatomy matters. We introduce JAM, a training-free one-shot prototype method that builds prototypes from a single support image and auto-generates optimal prompts at inference. We also propose PSC, a prototype similarity-coverage score that replaces confidence-based mask selection for more reliable results. JAM is plug-and-play and outperforms SAM-based and few-shot baselines on CHAOS-MRI-T2, Synapse-CT, and ETIS in both accuracy and scalability.**

***Index Terms*—Training-free, Foundation Model, Automatic Prompt Generation, Medical Image Segmentation, Few-shot Segmentation**

I. Introduction

Recent advances in deep learning have shown that foundation models trained on large datasets can rapidly transfer to downstream tasks [1]. A representative example is SAM [2], which achieves strong segmentation with minimal user input after training on natural images.

However, SAM underperforms in medical imaging [3], [4] due to (i) fine anatomical details demanding high precision, (ii) limited and costly expert annotations, and (iii) a confidence score calibrated for natural images, not clinical data. This unreliable calibration is problematic, as SAM's greedy decoder [5] selects the highest-confidence mask, thus hindering fine-grained lesion segmentation (Fig 1(a)). Existing remedies are also flawed: domain-specific finetuning (Fig 1(b)) is costly and lacks scalability, while few-shot schemes (Fig 1(c)) degrade on unseen domains and depend on variable manual prompts.

We propose **J**oint prototype-**A**ware **M**ask selection (JAM), a training-free, one-shot prototype framework (Fig 1(d)). JAM builds class prototypes from a single support image and stores them in a memory bank. At test time, it automatically selects optimal prompts by matching these prototypes to encoder features, eliminating extra training and minimizing expert intervention. We also introduce the **P**rototype **S**imilarity–**C**overage (PSC) score. This metric combines candidate coverage with prototype similarity, replacing SAM's unreliable confidence-based selection to robustly choose the best mask for medical images.

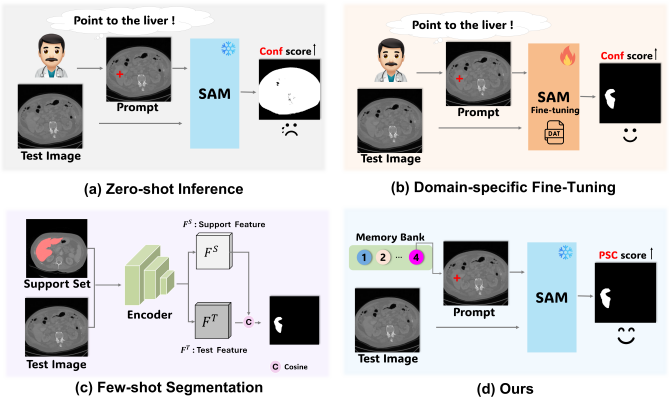

Fig. 1. Comparison of medical image segmentation approaches: (a) Zero-shot: SAM selects a prompt-based mask. (b) Fine-tuning: adapting SAM via additional training. (c) Few-shot: segmenting with support-set features. (d) Ours: training-free, prototype-based optimal mask selection.

As summarized in Table I, JAM requires no retraining, removes manual prompting, reliably selects among multiple masks via PSC, and generalizes across structures/domains by simply updating prototypes.

Our main contributions are:

- **Training-free one-shot framework:** Precise, rapid adaptation from a single support image without finetuning.
- **PSC score:** A coverage–similarity metric that supersedes confidence-based selection for medical images.
- **Automated prompts:** Consistent, reproducible inference without expert-provided prompts.
- **Plug-and-play scalability:** Generalizes to new domains by updating prototypes only.

II. Methods

A. Overview

We propose a training-free, one-shot prototype-based framework to adapt a frozen SAM2 for medical imaging. As illustrated in Fig 2, Our method consists of three key steps:

1) **ProtoBank Construction:** Building a memory bank of class-specific hybrid prototypes (global and local) from a support image.

* Corresponding author.

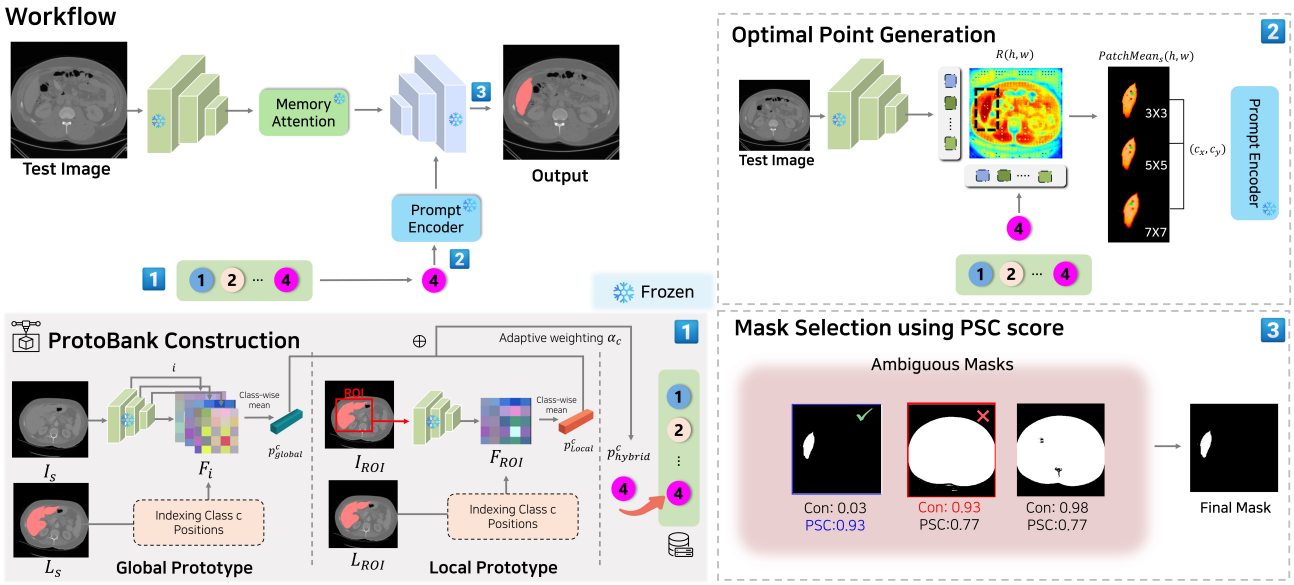

Fig. 2. Overview of JAM: 1 ProtoBank Construction 2 Optimal Point Generation 3 PSC-based Mask Selection.

TABLE I
COMPARISON OF CHARACTERISTICS AMONG ZERO-SHOT SAM, FINE-TUNED SAM, AND OUR METHOD.

Criteria	SAM	Fine-tuned SAM	Ours
1. No training needed?	✓	✗	✓
2. No manual prompts?	✗	✗	✓
3. Reliable mask selection?	✗	✓	✓
4. Generalizes to new medical domains?	✗	✗	✓

2) **Optimal Point Generation:** Automatically generating anchor point prompts for SAM2 based on prototype similarity with test image features.
3) **Mask Selection using PSC Score:** Introducing a PSC Score to select the most reliable mask from SAM2's outputs, mitigating domain discrepancy.

B. ProtoBank Construction

Given a support image $I_s \in \mathbb{R}^{H\times W\times C}$ and label $L_s \in 0, 1, ..., K^{H\times W}$, we build a ProtoBank of class-specific prototypes $p^c \in \mathbb{R}^C$. We utilize the frozen SAM2 encoder's multi-scale features (dimension C) to compute and combine global and local prototypes.

1) Global Prototype: From SAM2's multi-scale feature maps F_i, we identify pixel positions $(y, x) \in \mathcal{P}^c$ for class c. We map label coordinates to feature coordinates (h_i, w_i) at each scale i:

$$(h_i, w_i) = \left(\left\lfloor \frac{y \cdot H_i}{H} \right\rfloor, \left\lfloor \frac{x \cdot W_i}{W} \right\rfloor \right) \tag{1}$$

The global prototype is the mean of all corresponding feature vectors $f \in \mathcal{F}^c$ collected across all scales, capturing robust, multi-resolution context:

$$p^c_{\text{global}} = \frac{1}{|\mathcal{F}^c|} \sum_{f\in\mathcal{F}^c} f \tag{2}$$

2) Local Prototype: To capture fine-grained details, we crop a region of interest (ROI) for class c (I_{ROI}) based on its bounding box. This crop is passed through the encoder to get a single high-resolution feature map F_{ROI}. The local prototype is the mean of class c features within this ROI:

$$p^c_{\text{local}} = \frac{1}{|\mathcal{R}^c|} \sum_{(h,w)\in\mathcal{R}^c} F_{\text{ROI}}[:, h, w] \tag{3}$$

This focuses on local, high-resolution details, unlike the multi-scale global prototype.

3) Hybrid Prototype: The global (context) and local (detail) prototypes are combined via a weighted average to form the final hybrid prototype p^c_{hybrid} stored in the ProtoBank:

$$p^c_{\text{hybrid}} = \alpha_c \cdot p^c_{\text{global}} + (1 - \alpha_c) \cdot p^c_{\text{local}} \tag{4}$$

The weight $\alpha_c \in [0, 1]$ is determined by the size of the class region, balancing focus between large-scale context and fine details.

C. Optimal Point Generation

At test time, we match the stored prototype p^c_{hybrid} with SAM2 encoder features to find an anchor point for class c. We first compute an *anchor mask* from a similarity map, then refine it using multi-scale patches to select a robust point.

1) Anchor Mask via Similarity: We compute the per-pixel cosine similarity $R(h, w)$ between p^c_{hybrid} and encoder features $f_{h,w}$:

$$R(h, w) = \frac{p^c_{\text{hybrid}} \cdot f_{h,w}}{\|p^c_{\text{hybrid}}\|_2 \, \|f_{h,w}\|_2}. \tag{5}$$

Algorithm 1 Multi-scale patch-based anchor selection

Require: Similarity map R, anchor mask M_{class}, scales $\mathcal{S}$
Ensure: Anchor point (c_x, c_y)

$A \leftarrow \text{zeros}(H, W)$
for $s \in \mathcal{S}$ **do**
  $\text{simSum} \leftarrow \text{boxFilter}(R, s)$
  $\text{maskCnt} \leftarrow \text{boxFilter}(M_{\text{class}}, s)$
  for (h, w) **with** $M_{\text{class}}(h, w) = 1$ **do**
    $\text{PatchMean}_s \leftarrow \text{simSum}(h, w)/(\text{maskCnt}(h, w)+\varepsilon)$
    $A(h, w) \leftarrow A(h, w) + \text{PatchMean}_s$
  end for
end for
$(c_x, c_y) \leftarrow \arg\max_{(h,w):M_{\text{class}}(h,w)=1} A(h, w)$; **return** (c_x, c_y) =0

The anchor mask M_{class} is formed by thresholding the top 20% of R.

2) Multi-Scale Patch-Based Point: To mitigate noise, we aggregate similarity over patches. For each pixel within the anchor mask, we compute the PatchMean_s at scale s:

$$\text{PatchMean}_s(h, w) = \frac{1}{|P_s(h, w)|} \sum_{(h',w') \in P_s(h,w)} R(h', w'), \quad (6)$$

where $P_s(h, w)$ is the $s \times s$ neighborhood intersected with the anchor mask M_{class}. The final point is selected by maximizing the sum of patch means across all scales $\mathcal{S} = \{s_1, \ldots, s_K\}$:

$$(c_x, c_y) = \arg \max_{(h,w) \in M_{\text{class}}} \sum_{s \in \mathcal{S}} \text{PatchMean}_s(h, w). \quad (7)$$

The selected (c_x, c_y) is used as a positive point prompt for SAM2. The boxFilter efficiently computes sums for PatchMean_s, and ε ensures numerical stability. The multi-scale patch-based anchor point selection procedure is summarized in Algorithm 1.

D. Mask Selection using PSC Score

Given the anchor point, SAM2 proposes multiple masks $\{A_k\}$. We select the best one using a prototype-guided PSC score, which combines semantic similarity and coverage.

1) Similarity: The similarity sim_k is the cosine similarity between the prototype p^c_{hybrid} and the mask's average feature

$$\bar{f}_k = \frac{1}{|A_k|} \sum_{(h,w) \in A_k} f_{h,w}, \qquad \text{sim}_k = \frac{\bar{f}_k \cdot p^c_{\text{hybrid}}}{\|\bar{f}_k\|_2 \, \|p^c_{\text{hybrid}}\|_2}. \quad (8)$$

2) Coverage: Coverage cov_k measures the overlap between the proposed mask A_k and the anchor mask M_{class}:

$$\text{cov}_k = \frac{|A_k \cap M_{\text{class}}|}{|A_k|}. \quad (9)$$

3) PSC Score: The final score is a weighted sum, where $\mathcal{N}(\cdot)$ is min-max normalization:

$$\text{PSC}_k = \beta \, \text{sim}_k + (1 - \beta) \, \mathcal{N}(\text{cov}_k), \quad \beta \in [0, 1]. \quad (10)$$

We select the mask A_k with the highest PSC_k.

III. Experiments

A. Experimental Setting

1) Datasets: We evaluate JAM on CHAOS-MRI-T2 [6] (20 scans/623 slices), Synapse-CT [7] (30/3,779), and ETIS [8]. For CHAOS/Synapse we follow [9]; for ETIS we adopt the PraNet protocol [10] and compare to in-domain, zero-shot, and fine-tuned baselines. We report Dice and IoU [11], [12]; organs for CHAOS/Synapse are liver, L/R kidney, spleen, and ETIS has a single class.

2) Implementation Details: JAM is implemented in PyTorch and run on a V100 with SAM2.1 (hiera_large). For each class, global/local prototypes are fused by a size-dependent weight α; unless stated, $\alpha \in \{0.3, 0.7\}$ with thresholds 1,500/1,200/1,000 pixels for CHAOS/Synapse/ETIS (set on validation). For new domains, we tune $\alpha \in \{0.1, 0.3, 0.5, 0.7, 0.9\}$ on 5–10 support images (validation only, no retraining). PSC uses β=0.6 (CHAOS/Synapse) and β=0.7 (ETIS), chosen from $\{0.0, 0.3, 0.5, 0.6, 0.7, 1.0\}$. At inference, anchors are obtained by cosine similarity; similarities are averaged over 3×3, 5×5, 7×7 windows, and we keep the top-q% of the smoothed map. Ablations over $q \in [10, 50]$ peak at 20% across benchmarks (and unseen LiTS/KiTS), so we fix q=20.

B. Comparison with State-of-the-Art Methods

We evaluated JAM's segmentation performance against state-of-the-art approaches. We compared few-shot methods on CHAOS-MRI-T2 and Synapse-CT, and analyzed in-domain, zero-shot, and fine-tuning strategies on the ETIS dataset.

Table II compares few-shot methods, SAM2, and JAM on CHAOS-MRI-T2 and Synapse-CT. We used the P2 setting for JAM (two prompt points). Unlike most few-shot models requiring supervised training, SAM2 and JAM are training-free and operate at inference time, offering greater flexibility.

On CHAOS-MRI-T2, JAM improved the average Dice score by 7.83% over SAM2 with one support image, achieving comparable or superior performance for LK, RK, and Spleen. While liver segmentation initially lagged, it improved significantly with more support images (e.g., reaching APSCL-level performance with 10 images). This demonstrates JAM's ability to leverage prototype diversity efficiently without retraining.

Similarly, on Synapse-CT, JAM consistently outperformed SAM2 by 4.46% in average Dice, maintaining strong segmentation across all organs. This confirms our prototype-based approach generalizes effectively across MRI and CT modalities and various anatomical structures.

Table III presents the ETIS dataset results. JAM achieved Dice 61.2% and IoU 54.3% without training, comparable to the in-domain model PraNet [10]. JAM also showed over 6% improvement against SAM-based zero-shot methods (SAM-H, SAM-L) and outperformed fine-tuned medical-domain methods like SAM-Adapter, SAMPath, and SurgicalSAM.

Prompting strategies were also key. While methods like SAM-H and SAM-Adapter require user prompts and others (SAMPath, SurgicalSAM) use fine-tuned generation, JAM

TABLE II

PERFORMANCE COMPARISON OF FEW-SHOT SEGMENTATION METHODS ON THE CHAOS-MRI-T2 AND SYNAPSE-CT DATASETS. *P2 INDICATES THE USE OF TWO POINTS. THE PROPOSED JAM MODEL ACHIEVES SUPERIOR PERFORMANCE COMPARED TO COMPETING METHODS WITHOUT ADDITIONAL TRAINING.

Model	Training	Venue	CHAOS-MRI-T2					Synapse-CT				
			Liver	LK	RK	Spleen	Mean	Liver	LK	RK	Spleen	Mean
SE-Net [13]	🔥	MIA20	28.68	58.95	60.25	50.06	49.49	47.05	41.83	35.02	40.91	41.20
PANet [14]	🔥	ICCV19	47.39	54.29	42.68	50.42	48.70	40.27	33.22	19.61	31.78	31.22
SSL-ALPNet [15]	🔥	ECCV20	71.01	71.94	77.98	63.38	71.08	74.68	62.02	51.38	65.77	63.46
PoissonSeg [16]	🔥	BIBM21	60.06	53.98	59.63	56.83	57.63	56.08	52.83	49.40	53.37	52.92
RP-Net [17]	🔥	ICCV21	67.04	77.39	84.51	74.83	75.94	80.67	70.27	72.82	69.56	73.33
GCN-DE [18]	🔥	CIBM22	53.08	75.05	83.54	65.48	69.29	47.02	69.38	73.48	56.70	61.65
SPRNet [19]	🔥	MICCAI22	76.04	73.70	82.45	70.26	75.61	73.93	66.52	59.71	61.36	65.38
AAS-DCL [20]	🔥	ECCV22	72.78	52.58	83.38	60.93	67.42	72.40	63.80	68.04	67.01	67.81
ADNet [21]	🔥	MIA22	80.69	78.31	87.31	75.85	80.54	75.80	68.26	64.70	60.74	67.38
LVQM [22]	🔥	CVPR23	83.08	80.01	87.54	76.79	81.96	80.43	73.14	76.10	70.81	75.12
APSCL [9]	🔥	MM24	**86.73**	84.66	89.66	80.82	85.47	**87.74**	80.19	78.00	80.05	81.50
SAM2(P2) [23]	❄	ICLR25	71.62	80.82	83.1	78.91	78.61	74.73	81.32	84.54	82.88	80.87
JAM (ours, P2, 1-shot)	❄	-	78.36	88.98	91.85	86.56	86.44	81.30	85.53	88.4	86.1	85.33
JAM (ours, P2, 10-shot)	❄	-	**86.73**	**91.47**	**94.42**	**88.98**	**90.4**	87.71	**89.63**	**90.65**	**87.69**	**88.92**

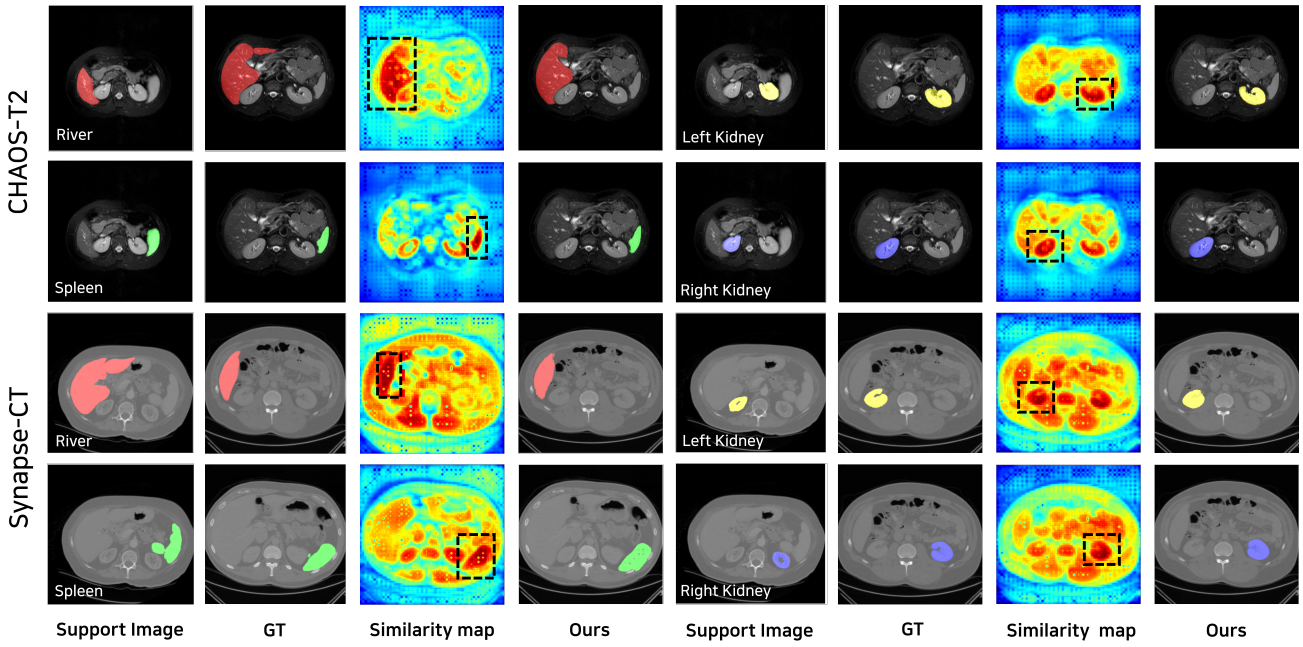

Fig. 3. Visualization of class-wise segmentation results on CHAOS-MRI-T2 and Synapse-CT datasets.

autonomously generates optimal prompts via prototype similarity. Despite being fully automated, JAM achieved higher Dice and IoU scores, demonstrating the practicality of our prompt-free approach.

Fig 3 and 4 offer qualitative comparisons. Fig 3 displays support images, GT, similarity maps, and predictions on CHAOS-MRI-T2 and Synapse-CT, showing how prototype similarity guides point selection. Fig 4 compares SAM-H, SAM-Adapter, and JAM on ETIS. Here, JAM (using the PSC score) consistently produced accurate masks, unlike other methods that missed or over-segmented lesions.

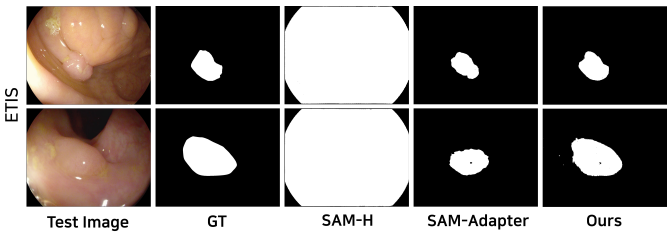

Fig. 4. ETIS segmentation: SAM-H, SAM-Adapter, and JAM.

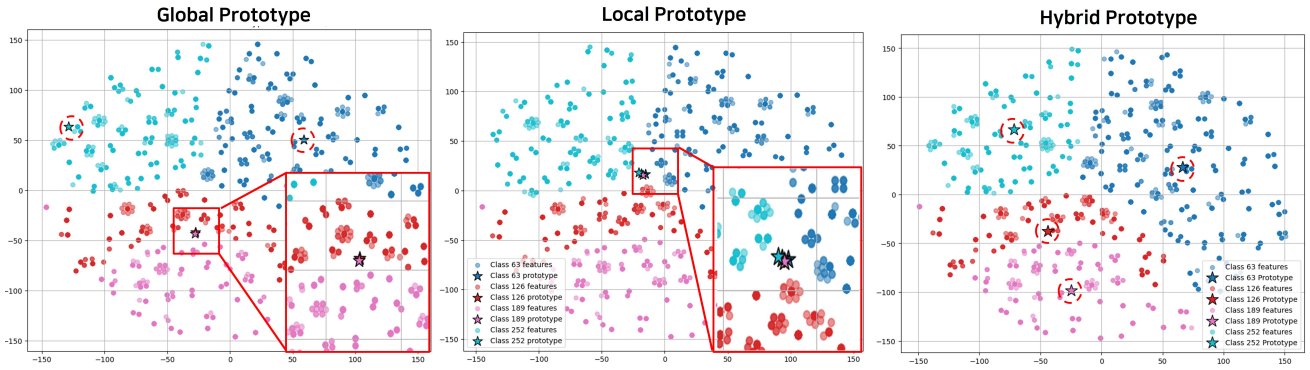

Fig. 5. Prototype effects: class prototypes (★) and features (◯).

TABLE III
PERFORMANCE COMPARISON OF VARIOUS APPROACHES (IN-DOMAIN, ZERO-SHOT, AND FINE-TUNING) ON ETIS DATASET.

Task	No Expert Prompt?	Methods	Venue	ETIS	
				Dice	IoU
In-domain	-	U-Net [24]	MICCAI15	39.8	33.5
	-	PraNet [10]	MICCAI19	**62.8**	**56.7**
Zero-shot	X	SAM-H [2]	ICCV23	51.7	47.7
	X	SAM-L [2]	ICCV23	55.1	50.7
Fine-tuning	X	SAM-Adapter [25]	ICCV23	59	47.6
	O	SAMPath [26]	MICCAI23	55.5	44.2
	O	SurgicalSAM [27]	AAAI24	34.2	23.8
Training-free	O	JAM (ours)	-	61.2	54.3

TABLE IV
PROTOTYPE COMPONENT ABLATION ON CHAOS-MRI-T2.

Model	Liver	LK	RK	Spleen	Mean
Local Prototype	73.75	65.60	79.93	86.91	76.54
Global Prototype	78.57	85.94	85.88	**90.02**	85.24
Hybrid Prototype	**78.36**	**88.98**	**91.85**	86.56	**86.44**

TABLE V
PSC COMPONENT ABLATION ON SYNAPSE-CT.

Model	Liver	LK	RK	Spleen	Mean
Confidence	62.60	77.95	77.95	84.21	75.60
Similarity	79.33	83.39	84.38	83.28	82.58
Coverage	81.06	84.31	**88.60**	84.64	84.65
PSC (sim + cov)	**81.30**	**85.53**	88.40	**86.10**	**85.33**

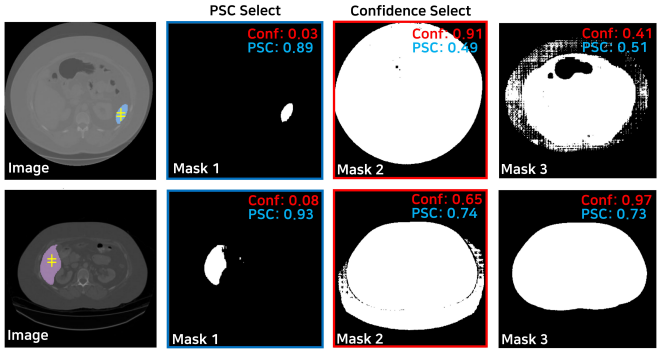

Fig. 6. PSC vs. confidence-based selection on Synapse-CT.

C. Ablation Study

1) Analysis of ProtoBank Effectiveness: Table IV shows that Local prototypes capture details but underperform overall (76.54%). Global prototypes model organ shape better (85.24%) yet miss fine structures. The Hybrid prototype yields the best mean (86.44%), balancing global context and local detail; t-SNE (Fig. 5) also indicates clearer LK/RK separation.

2) Analysis of PSC Score and Its Components: As shown in V, confidence alone is weakest (75.60 mean Dice). Similarity and Coverage each help (82.58/84.65), and their combination (PSC) is best (85.33), especially on difficult kidneys. Qualitatively (Fig. 6), PSC prefers compact, anatomically plausible masks and is robust in low-contrast or ambiguous boundaries, unlike confidence-only selection.

3) Performance Variation with Number of Support Images: As summarized in Table VI, performance scales with the number of supports: even at one-shot, JAM remains competitive, and with 10 support images the CHAOS liver reaches 86.7% Dice while means rise across CHAOS-MRI-T2 and Synapse-CT—evidence that richer, more stable prototypes emerge without retraining.

IV. LIMITATION AND CONCLUSION

Limitation: JAM adds inference overhead for prototype extraction and similarity matching: on one V100 (batch=1) it runs at ∼17 FPS vs. 41 FPS for SAM2, processing a 100-slice CT in ∼6 s and a 150-slice MRI in ∼9 s. This is still clinically practical, but further speed optimizations (e.g., parallelization) are desirable.

Conclusion: We introduced JAM, a training-free, one-shot prototype framework that auto-generates prompts and replaces confidence-based selection with the PSC score. Across CHAOS-MRI-T2, Synapse-CT, and ETIS, JAM matches or surpasses few-shot and fine-tuning baselines without expert-

TABLE VI
DICE VS. NUMBER OF SUPPORT IMAGES (CHAOS-MRI-T2, SYNAPSE-CT, ETIS).

Support images	CHAOS-MRI-T2		Synapse-CT		ETIS
	Liver	Mean	Liver	Mean	Mean
1	78.36	86.44	81.3	85.33	61.2
3	82.5	88.1	83.5	87.2	62.5
5	84.0	89.2	85.0	87.8	63.4
7	85.5	89.8	86.2	88.3	64.1
10	86.7	90.4	87.7	88.9	64.5

crafted prompts, improving deployability. Future work will pursue faster, lighter implementations and broader multi-modality validation.

ACKNOWLEDGMENTS

This work was supported in part by the National Research Foundation of Korea (NRF) grant funded by the Korean government (MSIT) under Grant Nos. RS-2023-00229822 and RS-2025-02312833.

REFERENCES

[1] K. He, X. Zhang, S. Ren, and J. Sun, "Deep residual learning for image recognition," in *Proceedings of the IEEE conference on computer vision and pattern recognition*, 2016, pp. 770–778.

[2] A. Kirillov, E. Mintun, N. Ravi, H. Mao, C. Rolland, L. Gustafson, T. Xiao, S. Whitehead, A. C. Berg, W.-Y. Lo *et al.*, "Segment anything," in *Proceedings of the IEEE/CVF international conference on computer vision*, 2023, pp. 4015–4026.

[3] J. Ma, S. Kim, F. Li, M. Baharoon, R. Asakereh, H. Lyu, and B. Wang, "Segment anything in medical images and videos: Benchmark and deployment," *arXiv preprint arXiv:2408.03322*, 2024.

[4] J. Cheng, J. Ye, Z. Deng, J. Chen, T. Li, H. Wang, Y. Su, Z. Huang, J. Chen, L. Jiang *et al.*, "Sam-med2d," *arXiv preprint arXiv:2308.16184*, 2023.

[5] S. Ding, R. Qian, X. Dong, P. Zhang, Y. Zang, Y. Cao, Y. Guo, D. Lin, and J. Wang, "Sam2long: Enhancing sam 2 for long video segmentation with a training-free memory tree," *arXiv preprint arXiv:2410.16268*, 2024.

[6] A. E. Kavur, N. S. Gezer, M. Barış, S. Aslan, P.-H. Conze, V. Groza, D. D. Pham, S. Chatterjee, P. Ernst, S. Özkan *et al.*, "Chaos challenge-combined (ct-mr) healthy abdominal organ segmentation," *Medical image analysis*, vol. 69, p. 101950, 2021.

[7] B. Landman, Z. Xu, J. Igelsias, M. Styner, T. Langerak, and A. Klein, "Miccai multi-atlas labeling beyond the cranial vault–workshop and challenge," in *Proc. MICCAI multi-atlas labeling beyond cranial vault—workshop challenge*, vol. 5. Munich, Germany, 2015, p. 12.

[8] J. Silva, A. Histace, O. Romain, X. Dray, and B. Granado, "Toward embedded detection of polyps in wce images for early diagnosis of colorectal cancer," *International journal of computer assisted radiology and surgery*, vol. 9, pp. 283–293, 2014.

[9] W. Huang, J. Hu, X. Bi, and B. Xiao, "Anatomical prior guided spatial contrastive learning for few-shot medical image segmentation," in *Proceedings of the 32nd ACM International Conference on Multimedia*, 2024, pp. 5211–5220.

[10] D.-P. Fan, G.-P. Ji, T. Zhou, G. Chen, H. Fu, J. Shen, and L. Shao, "Pranet: Parallel reverse attention network for polyp segmentation," in *International conference on medical image computing and computer-assisted intervention*. Springer, 2020, pp. 263–273.

[11] A. A. Taha and A. Hanbury, "Metrics for evaluating 3d medical image segmentation: analysis, selection, and tool," *BMC medical imaging*, vol. 15, pp. 1–28, 2015.

[12] J. Long, E. Shelhamer, and T. Darrell, "Fully convolutional networks for semantic segmentation," in *Proceedings of the IEEE conference on computer vision and pattern recognition*, 2015, pp. 3431–3440.

[13] A. G. Roy, S. Siddiqui, S. Pölsterl, N. Navab, and C. Wachinger, "'squeeze & excite'guided few-shot segmentation of volumetric images," *Medical image analysis*, vol. 59, p. 101587, 2020.

[14] K. Wang, J. H. Liew, Y. Zou, D. Zhou, and J. Feng, "Panet: Few-shot image semantic segmentation with prototype alignment," in *proceedings of the IEEE/CVF international conference on computer vision*, 2019, pp. 9197–9206.

[15] C. Ouyang, C. Biffi, C. Chen, T. Kart, H. Qiu, and D. Rueckert, "Self-supervision with superpixels: Training few-shot medical image segmentation without annotation," in *Computer Vision–ECCV 2020: 16th European Conference, Glasgow, UK, August 23–28, 2020, Proceedings, Part XXIX 16*. Springer, 2020, pp. 762–780.

[16] X. Shen, G. Zhang, H. Lai, J. Luo, J. Lu, and Y. Luo, "Poisson-seg: semi-supervised few-shot medical image segmentation via poisson learning," in *2021 IEEE international conference on Bioinformatics and biomedicine (BIBM)*. IEEE, 2021, pp. 1513–1518.

[17] H. Tang, X. Liu, S. Sun, X. Yan, and X. Xie, "Recurrent mask refinement for few-shot medical image segmentation," in *Proceedings of the IEEE/CVF international conference on computer vision*, 2021, pp. 3918–3928.

[18] L. Sun, C. Li, X. Ding, Y. Huang, Z. Chen, G. Wang, Y. Yu, and J. Paisley, "Few-shot medical image segmentation using a global correlation network with discriminative embedding," *Computers in biology and medicine*, vol. 140, p. 105067, 2022.

[19] R. Wang, Q. Zhou, and G. Zheng, "Few-shot medical image segmentation regularized with self-reference and contrastive learning," in *International Conference on Medical Image Computing and Computer-Assisted Intervention*. Springer, 2022, pp. 514–523.

[20] H. Wu, F. Xiao, and C. Liang, "Dual contrastive learning with anatomical auxiliary supervision for few-shot medical image segmentation," in *European Conference on Computer Vision*. Springer, 2022, pp. 417–434.

[21] S. Hansen, S. Gautam, R. Jenssen, and M. Kampffmeyer, "Anomaly detection-inspired few-shot medical image segmentation through self-supervision with supervoxels," *Medical Image Analysis*, vol. 78, p. 102385, 2022.

[22] S. Huang, T. Xu, N. Shen, F. Mu, and J. Li, "Rethinking few-shot medical segmentation: a vector quantization view," in *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*, 2023, pp. 3072–3081.

[23] N. Ravi, V. Gabeur, Y.-T. Hu, R. Hu, C. Ryali, T. Ma, H. Khedr, R. Rädle, C. Rolland, L. Gustafson *et al.*, "Sam 2: Segment anything in images and videos," *arXiv preprint arXiv:2408.00714*, 2024.

[24] O. Ronneberger, P. Fischer, and T. Brox, "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*. Springer, 2015, pp. 234–241.

[25] T. Chen, L. Zhu, C. Ding, R. Cao, Y. Wang, Z. Li, L. Sun, P. Mao, and Y. Zang, "Sam fails to segment anything?–sam-adapter: Adapting sam in underperformed scenes: Camouflage, shadow, medical image segmentation, and more," *arXiv preprint arXiv:2304.09148*, 2023.

[26] J. Zhang, K. Ma, S. Kapse, J. Saltz, M. Vakalopoulou, P. Prasanna, and D. Samaras, "Sam-path: A segment anything model for semantic segmentation in digital pathology," in *International Conference on Medical Image Computing and Computer-Assisted Intervention*. Springer, 2023, pp. 161–170.

[27] W. Yue, J. Zhang, K. Hu, Y. Xia, J. Luo, and Z. Wang, "Surgical-sam: Efficient class promptable surgical instrument segmentation," in *Proceedings of the AAAI Conference on Artificial Intelligence*, vol. 38, no. 7, 2024, pp. 6890–6898.