MedPro: Dual-view Propagation and Prototypical Assessment for Training-free Few-shot Medical Segmentation

Jingyi Zhang[1], Mingyi Liu[1], Xiangbo Shu[1], Xu-Yao Zhang[2], and Guo-Sen Xie[1(✉)]

[1] School of Computer Science and Engineering, Nanjing University of Science and Technology, China
gsxiehm@gmail.com
[2] Institute of Automation, Chinese Academy of Sciences, China

Abstract. The emergence of foundation models enables a training-free paradigm for medical image segmentation. While previous frameworks relied on heavy prompt engineering for the original SAM, SAM2 enables direct mask propagation from support sets, obviating manual prompt extraction. However, naïve propagation in a fixed field of view often fails to adapt to subtle anatomical variations and intensity shifts. To fully harness the sequence propagation capability of SAM2, we propose Dual-view **Pro**pagation and **Pro**totypical Assessment for Training-free Few-shot **Med**ical Segmentation (**MedPro**), which integrates three synergistic strategies. First, histogram matching aligns intensity distributions across subjects, improving cross-subject statistical consistency prior to propagation. Building upon this, a focus zoom-in mechanism generates dual-view pseudo-video sequences that incorporate both global contextual information and localized ROI details. These sequences are then processed via SAM2 for bidirectional mask propagation, providing a fine-grained perspective while mitigating spatial drift. Finally, a prototypical assessment is developed for automated semantic confirmation. By generating representative prototypes from DINOv2 features, it evaluates candidate masks via semantic similarity to reliably identify the optimal segmentation. Extensive experiments demonstrate that MedPro achieves competitive results across multiple datasets. Code is available at: https://github.com/CVL-hub/MedPro.

Keywords: Few-shot learning · Medical image segmentation · Foundation model.

1 Introduction

Medical image segmentation is often constrained by the substantial labeling burden, as pixel-wise delineation requires stringent biological expertise and exhaustive manual effort. To address this, few-shot medical image segmentation (FSMIS) aims to distill task-specific knowledge from limited labeled samples to

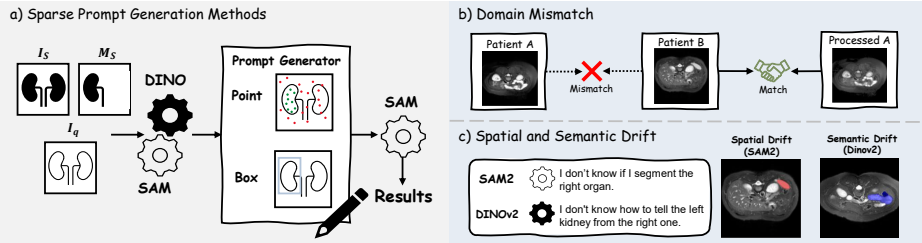

Fig. 1. Challenges in Training-Free FSMIS. (a) Sparse prompt generation methods underutilizing pixel-level morphology. (b) Domain mismatch caused by intensity shifts, hindering propagation stability. (c) Spatial drift (SAM2) and semantic drift (DINOv2) causing failures in anatomical consistency.

segment unseen images of corresponding anatomical structures. Existing FSMIS approaches can be broadly grouped into two categories: **1) prototype-based methods:** Models [3, 12, 16, 20] encode support features into class-specific prototypes for query matching. **2) foundation-model-based methods:** Models adapt large-scale pre-trained architectures through fine-tuning [9,13] and prompt engineering [10, 19] to specialized medical segmentation tasks. Given the exceptional generalization capabilities of these large-scale models, their application in training-free frameworks provides a clinically agile and deployment-friendly alternative. Such approaches [10, 17] avoid expensive retraining procedures and instead exploit strong prior knowledge for high-performance segmentation with minimal deployment overhead.

Although SAM [7] has pioneered training-free workflows, it lacks the mechanism to directly utilize the rich structural priors embedded within support sets. Consequently, as in Fig. 1a, existing research focuses on automated prompting via feature matching to generate sparse guidance points or bounding boxes. However, this paradigm underutilizes the pixel-level morphology of the support mask, risking the loss of critical anatomical details. The advent of SAM2 [14] has partially mitigated this bottleneck by allowing the direct incorporation of support masks as initial states for inter-image propagation [2], effectively modeling the support-query relationship as a pseudo-video sequence. This bypasses tedious prompt extraction but still faces formidable challenges in complex medical scenarios. **First**, domain mismatch remains critical (Fig. 1b); variations in individuals, devices, and protocols cause significant intensity shifts between support and query sets, hindering propagation stability. **Second**, medical scans often exhibit low contrast and high anatomical similarity, where blurred boundaries and subtle visual differences allow surrounding tissues to distract from the target segmentation. **Third**, while SAM2 excels in low-level edge localization, its limited high-level semantic discriminability (Fig. 1c) makes it susceptible to distractors from adjacent similar structures. In anatomically complex regions, this leads to spatial drift, where the mask gradually deviates from the target, resulting in segmentation failure.

To address these challenges, we propose MedPro, integrating three synergistic strategies: Pair Construction and Alignment (PCA), Dual-view Propagation (DP) and Prototypical Assessment (PA). Specifically, PCA employs histogram matching to align intensity distributions, mitigating domain mismatch and introducing a focus zoom-in mechanism for DP, decoupling target structures from surrounding anatomical interference. Building upon the constructed sample pairs, DP propagates features to generate initial candidate masks. Finally, PA leverages the robust high-level semantic representations of DINOv2 [11] to validate candidate masks against class prototypes, effectively rectifying spatial drift and ensuring anatomical consistency with the support prior.

2 Methodology

2.1 Overview

The overall pipeline of MedPro is illustrated in Fig. 2, which consists of three stages: pair construction and alignment (PCA), dual-view propagation (DP), and prototypical assessment (PA). Given a support-query pair $\{I_S, I_Q\}$, we first align their intensity distributions via histogram matching. Subsequently, a focus zoom-in mechanism is employed to construct dual-view pseudo-video sequences comprising both global context and localized ROI details based on the support mask M_S. These sequences are fed into SAM2 to generate a candidate mask set $\mathcal{M}_Q^{candidate} = \{M_Q^{global}, M_Q^{local}\}$. Finally, the PA performs mask average pooling on DINOv2 features to compare the semantic similarity between the candidates and the support prototypes, which automatically selects the optimal segmentation M^*.

2.2 Pair Construction and Alignment

Histogram Matching. Before segmentation, we treat the $\{I_S, I_Q\}$ pair as consecutive video frames. To better emulate temporal continuity of a video, where pixel-level variance between two adjacent frames is minimized and the overall scene remains consistent, we apply histogram matching between the support and query images. This preprocessing step effectively aligns the brightness and contrast distributions of the support and query images, ensuring that the model extracts and propagates features within a consistent intensity domain.

First, normalized histograms are computed as probability density functions. For a query image with intensity levels z, the Cumulative Distribution Function (CDF) $G(z)$ is:

$$G(z) = \int_0^z p_Q(w)dw, \tag{1}$$

where p_Q is the probability density of the respective image.

The support image is then transformed using a mapping function $f : r_k \to z_k$ derived from the CDF of query image:

$$z_k = G^{-1}\left(\sum_{i=0}^{k} \frac{n_{S,i}}{N_S}\right), \tag{2}$$

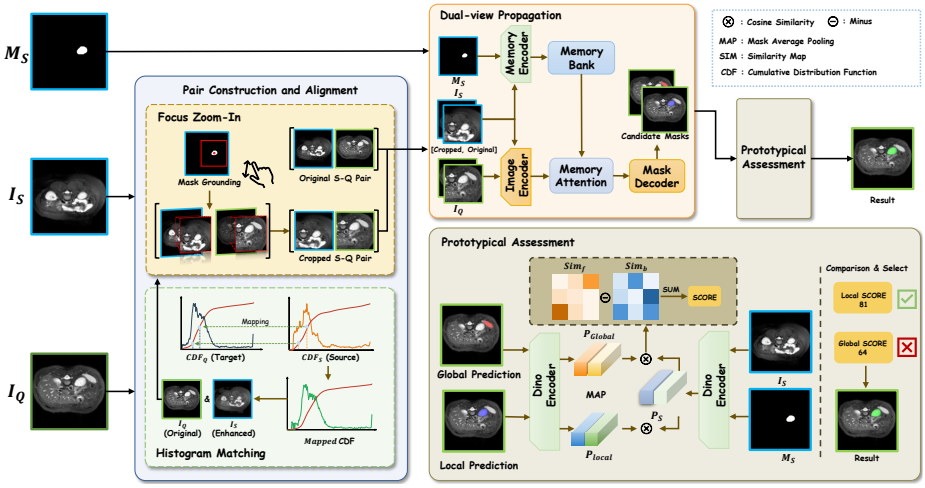

Fig. 2. The overview of the proposed MedPro. It comprises three key stages: (1) Pair construction and alignment via histogram matching of $\{I_S, I_Q\}$; (2) Dual-view sequential propagation using a focus zoom-in mechanism to generate global (M_Q^{global}) and local (M_Q^{local}) candidate masks via SAM2; (3) Prototypical assessment to select the optimal segmentation M^* based on DINOv2 semantic similarity.

where $n_{S,i}$ represents the number of pixels with intensity i in the support image, and N_S is the total pixel count.

Focus Zoom-In. In medical imaging, ambiguous organ boundaries and limited inter-class contrast often cause SAM2 to suffer from global spatial shifts when operating under a fixed view. To address this, we implement a zoom-in strategy to refine the model's focus.

Leveraging the inherent anatomical consistency of medical scans,where the relative positions of organs remain stable across different subjects and environments, we crop a localized Region of Interest (ROI) based on the geometric center of the support mask. By expanding this center into a bounding box and resizing it for the input pair, SAM2 gains a fine-grained perspective, which significantly mitigates segmentation drift.

We first determine the bounding box of the target organ based on the support mask M_S. To ensure sufficient context, a padding δ is added to the boundaries. The boundary coordinates (y_1, y_2, x_1, x_2) are defined as follows:

$$\begin{cases} y_1, y_2 = \text{clip}(y_{min} - \delta, y_{max} + \delta), \\ x_1, x_2 = \text{clip}(x_{min} - \delta, x_{max} + \delta), \end{cases} \tag{3}$$

where the clip function ensures that the coordinates do not exceed the height H and width W of the original image.

Using the coordinates obtained above, we crop the original image pair into localized image patches I_{crop}:

$$I^{crop} = I[y_1 : y_2,\ x_1 : x_2], \quad I \in \{I_S, I_Q\}. \tag{4}$$

Subsequently, we feed both original and cropped $\{I_S, I_Q\}$ pairs into the SAM2 model. This approach benefits from a dual-scale representation: while the original images provide essential global context, the cropped pairs reduce interference from neighboring similar structures. The integration of both global and local information ensures a more precise and robust segmentation.

2.3 Dual-view Propagation

Following the design of FS-MedSAM2 [2], we leverage the temporal propagation capabilities of SAM2 by treating the global image pair $\{I_S, I_Q\}$ and the zoom-in pair $\{I_S^{crop}, I_Q^{crop}\}$ as two independent pseudo-video sequences.

For each sequence, given a support image and its corresponding mask M_S, the memory encoder first generates support memories, which are later stored in the memory bank. During the inference of a query image, the features extracted by the image encoder serve as the query, which undergoes a cross-attention operation with the support features key and value retrieved from the memory bank. This mechanism allows the precise boundary priors from the support mask to guide the segmentation of the query image. Finally, these memory-conditioned features are fed into the mask decoder to generate two candidate masks:

$$\mathcal{M}_Q^{candidate} = \{M_Q^{global}, M_Q^{local}\}, \tag{5}$$

which will be further evaluated by the subsequent selection module.

2.4 Prototypical Assessment

To find the optimal segmentation result from the candidate set $\mathcal{M}_Q^{candidate} = \{M_Q^{global}, M_Q^{local}\}$, we introduce a selection mechanism based on DINOv2 feature prototypes. Leveraging the superior semantic representation of DINOv2, we evaluate the quality of predictions by measuring the feature similarity between the candidate regions and the support set.

We utilize the deep features F extracted directly from the DINOv2 image encoder. For the support image I_S (with its mask M_S) and the query image I_Q (with its candidate mask $M \in \mathcal{M}_Q^{candidate}$), the feature prototypes are generated using Mask Average Pooling:

$$\mathbf{p}_x = \frac{\sum F \cdot M_x}{\sum M_x + \epsilon}, \quad x \in \{fg, bg, m\}, \tag{6}$$

where $\mathbf{p}_{fg}$ and $\mathbf{p}_{bg}$ represent the foreground and background prototypes derived from the support set, respectively. $\mathbf{p}_m$ represents the prototype of the current candidate mask.

The quality score S_{mask} for each candidate is defined by calculating the dot-product similarity between the candidate prototype $\mathbf{p}_m$ and the support prototypes:

$$S_{mask} = (\mathbf{p}_m \cdot \mathbf{p}_{fg}) - (\mathbf{p}_m \cdot \mathbf{p}_{bg}). \tag{7}$$

This score reflects the degree to which a candidate region aligns with the target organ's semantics rather than the background. The final segmentation M^* is determined by selecting the candidate with the highest score:

$$M^* = \arg \max_{M \in \mathcal{M}_Q^{candidate}} S_{mask}(M). \tag{8}$$

This selection process ensures that the final mask is semantically optimal.

3 Experiment

3.1 Experimental Settings

Datasets and Evaluation. As a training-free framework, MedPro is evaluated without parameter updates across three benchmarks: CHAOS-MRI [6] (20 T2-SPIR MRIs), Synapse-CT [8] (30 abdominal CTs), and Card-MRI [21] (35 cardiac MRIs). While CHAOS and Synapse focus on multi-organ segmentation (liver, kidneys, spleen), Card-MRI targets the ventricles and myocardium.

Following the previous work [12], we adopt the mean Dice Similarity Coefficient (mDSC) as the primary evaluation metric. Notably, prior few-shot learning methods typically operate under two experimental settings, distinguished by whether the test-set organs appear in the background during the training phase. Since MedPro is a training-free framework, this distinction is not applicable to our approach. To ensure a fair comparison, we compare our results with the higher performance values reported under Setting 1 in the existing literature.

Implementation Details. Our framework is implemented in PyTorch and executed on a single NVIDIA GeForce RTX 3090 GPU. All input images and masks are resized to 256×256 pixels. We utilize the small variant of SAM2 (sam2-hiera-small) for mask propagation and the base variant of DINOv2 (dinov2-base) for semantic feature extraction. Following the standard few-shot protocol [12] , we conduct experiments under the 1-shot setting. For the focus zoom-in mechanism, a spatial padding of 50 pixels is applied across all three datasets.

3.2 Comparison with SOTA Methods

Performance on CHAOS-MRI. As illustrated in Tab. 1, our method achieves an mDSC of 83.54%, ranking first in the training-free category and performing on par with training-based approaches. Specifically, compared to the recent training-free baseline SynPo, MedPro yields a significant improvement of 3.61% in mean Dice. Notably, for specific organs such as the left kidney (85.22%) and spleen (84.20%), our framework outperforms the best supervised method, SPENet (81.98%, 80.41%), demonstrating the robustness of our DP strategy in

Table 1. Comparison with SOTA methods on two datasets under 1-shot setting. The best and second-best results are highlighted in **bold** and underline, respectively.

Method	CHAOS-MRI					Synapse-CT				
	LK	RK	Spleen	Liver	Mean	LK	RK	Spleen	Liver	Mean
Methods with Training										
SSL-ALPNet [12] ECCV'20	73.63	78.39	67.02	73.05	73.02	63.34	54.82	60.25	73.65	63.02
Q-Net [15] IntelliSys'23	78.36	87.98	75.99	81.74	81.02	75.26	74.79	74.86	71.21	74.03
RPT [20] MICCAI'23	80.72	89.82	76.37	82.86	82.44	77.05	79.13	72.58	82.57	77.83
GMRD [3] TMI'24	83.96	**90.12**	76.09	81.42	82.90	**81.70**	74.46	78.31	79.60	78.52
SPENet [4] MICCAI'25	81.98	89.18	80.41	**83.02**	**83.64**	80.01	73.39	80.74	82.76	79.23
Methods without Training										
PerSAM [18] ICLR'24	64.84	71.36	69.14	42.44	61.12	58.47	60.31	65.03	65.55	62.34
ProtoSAM [1] arXiv'24	71.46	81.43	76.51	81.94	77.83	69.44	71.04	65.50	87.84	73.45
FS_MedSAM2 [2] arXiv'24	69.91	86.66	83.29	71.03	77.72	71.35	80.26	85.49	87.30	81.10
SynPo [10] MICCAI'25	77.32	83.04	80.30	77.32	79.93	75.00	79.63	83.76	81.32	79.91
Ours	**85.22**	88.36	**84.20**	76.37	83.54	73.06	**84.17**	**86.89**	**87.87**	**83.00**

Table 2. Results on Card-MRI dataset under 1-shot setting.

Tr.	Method	Card-MRI			
		BP	MYO	RV	Mean
w/	PGRNet [5] TMI'24	**90.19**	67.65	77.73	78.52
w/	GMRD [3] TMI'24	90.00	67.04	80.29	79.11
w/	SPENet [4] MICCAI'25	90.15	69.99	81.80	80.64
w/o	**Ours**	89.62	**76.01**	**83.62**	**83.08**

Table 3. Ablation study on the effectiveness of each component on CHAOS-MRI dataset.

Component	mDSC	Δ
Baseline	77.72	-
+ Histogram Matching	79.45	+1.73
+ Focus Zoom-in	82.31	+2.86
+ Prototypical Assessment	**83.54**	+1.23

handling complex anatomical structures. This is due to our histogram matching normalizing MRI intensity shifts and DP capturing fine-grained boundaries.

Performance on Synapse-CT. On the Synapse-CT dataset, we achieve an mDSC of 83.00%. This performance not only sets a new record for training-free methods, surpassing FS_MedSAM2 (81.10%) by 1.90%, but also exhibits superior cross-modality generalization. Remarkably, our results consistently exceed those of training-intensive models such as SPENet (79.23%) and GMRD (78.52%), proving that precise semantic alignment and spatial priors can match or even surpass traditional supervised learning accuracy without parameter optimization.

Performance on Card-MRI. To further verify the generalizability across anatomical regions, we conduct experiments on the Card-MRI dataset. As shown in Tab. 2, our training-free MedPro achieves an mDSC of 83.08%, significantly outperforming all compared training-based methods. Compared to the leading supervised method, SPENet (80.64%), our approach provides a 2.44% improvement in overall performance. This significant gain across different anatomical regions demonstrates the generalization of our PA in suppressing spatial drift.

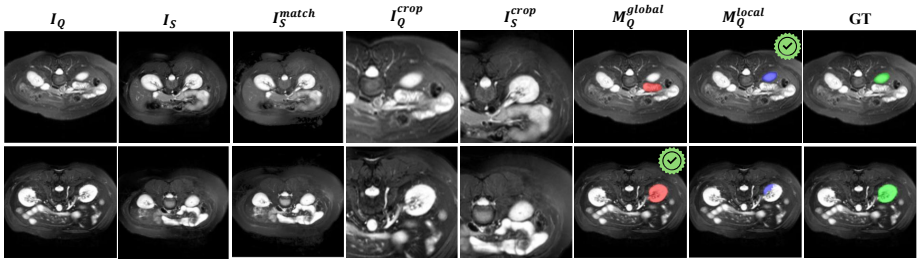

Fig. 3. Qualitative comparison of different components in the proposed framework. The green check-marks denote the final segmentation results selected by our model.

3.3 Ablation Study

Effectiveness of Each Component. Tab. 3 presents the performance impact of each MedPro component on the average segmentation accuracy. Our baseline yields an mDSC of 77.72%. Integrating histogram matching results in a 1.73% improvement, primarily by mitigating the intensity shifts between support and query images to ensure statistical consistency. The focus zoom-in mechanism provides a more substantial increase of 2.86%, as it enhances the representation of target regions by reinforcing localized anatomical details. Notably, the reported focus zoom-in result is obtained by weighted fusion of the segmentation masks predicted from the original and zoomed-in views. Finally, the PA module further boosts the mDSC to 83.54% (+1.23%) by leveraging DINOv2's global semantic features to identify the optimal mask. These results demonstrate the complementary nature of our proposed strategies, confirming that each module effectively contributes to performance gains within our training-free framework.
Qualitative Analysis. To intuitively evaluate the contribution of each component, Fig. 3 presents the intermediate outputs for two representative samples. As illustrated, histogram matching (I_S^{match}) effectively normalizes grayscale variance, ensuring feature consistency by aligning intensity distributions. However, single-scale segmentation is inherently risky: the focus zoom-in (I_Q^{crop}, I_S^{crop}) may lead to under-segmentation due to restricted field-of-view and missing context. Conversely, the global branch (M_Q^{global}), despite its full perspective, is prone to spatial drift caused by interference from adjacent similar tissues.

Regarding the outputs, the global (M_Q^{global}) and local (M_Q^{local}) branches exhibit complementary strengths in overall localization and fine-structure delineation, respectively. Finally, the prototypical assessment intelligently filters the candidate masks to select the semantically optimal result (indicated by check-marks), effectively correcting potential drifts. The high consistency between our results and the GT validates the synergistic effect of all components in enhancing training-free segmentation.

4 Conclusion

This paper presents **MedPro**, a training-free framework for few-shot medical image segmentation. By integrating pair construction and alignment, dual-view propagation and prototypical assessment, MedPro effectively mitigates intensity shifts, detail loss, and spatial drift. Results show that MedPro sets new records for training-free methods across three datasets and even outperforms supervised models on Card-MRI and Synapse-CT. Ablation and qualitative analyzes further confirm its superior generalization and anatomical precision without parameter optimization, offering a robust solution for rapid clinical deployment.

Acknowledgments. This work was supported by the National Natural Science Foundation of China (Grant Nos. 62276134, U25A20442, and 62427808).

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

References

1. Ayzenberg, L., Giryes, R., Greenspan, H.: Protosam: One-shot medical image segmentation with foundational models. arXiv preprint arXiv:2407.07042 (2024)
2. Bai, Y., Yu, Q., Yun, B., Jin, D., Xia, Y., Wang, Y.: Fs-medsam2: Exploring the potential of sam2 for few-shot medical image segmentation without fine-tuning. arXiv e-prints pp. arXiv–2409 (2024)
3. Cheng, Z., Wang, S., Xin, T., Zhou, T., Zhang, H., Shao, L.: Few-shot medical image segmentation via generating multiple representative descriptors. IEEE Transactions on Medical Imaging **43**(6), 2202–2214 (2024)
4. Fan, C., Jia, X., Xiao, A., Yu, H., Yang, Z., Yang, D., Xu, H., Huang, Y., Wang, L.: Spenet: Self-guided prototype enhancement network for few-shot medical image segmentation. In: International Conference on Medical Image Computing and Computer-Assisted Intervention. pp. 584–593. Springer (2025)
5. Huang, W., Hu, J., Xiao, J., Wei, Y., Bi, X., Xiao, B.: Prototype-guided graph reasoning network for few-shot medical image segmentation. IEEE Transactions on Medical Imaging **44**(2), 761–773 (2025)
6. Kavur, A.E., Gezer, N.S., Barış, M., Aslan, S., Conze, P.H., Groza, V., Pham, D.D., Chatterjee, S., Ernst, P., Özkan, S., et al.: Chaos challenge-combined (ct-mr) healthy abdominal organ segmentation. Medical image analysis **69**, 101950 (2021)
7. Kirillov, A., Mintun, E., Ravi, N., Mao, H., Rolland, C., Gustafson, L., Xiao, T., Whitehead, S., Berg, A.C., Lo, W.Y., et al.: Segment anything. In: Proceedings of the IEEE/CVF international conference on computer vision. pp. 4015–4026 (2023)
8. Landman, B., Xu, Z., Igelsias, J., Styner, M., Langerak, T., Klein, A.: Miccai multi-atlas labeling beyond the cranial vault–workshop and challenge. In: Proc. MICCAI multi-atlas labeling beyond cranial vaultworkshop challenge. vol. 5, p. 12. Munich, Germany (2015)
9. Leng, T., Zhang, Y., Han, K., Xie, X.: Self-sampling meta sam: enhancing few-shot medical image segmentation with meta-learning. In: Proceedings of the IEEE/CVF winter conference on applications of computer vision. pp. 7925–7935 (2024)

10. Liu, Y., Xiao, H., Chai, J., Zhang, Y., Wang, R., Meng, Z., Luo, Z.: Synpo: Boosting training-free few-shot medical segmentation via high-quality negative prompts. In: International Conference on Medical Image Computing and Computer-Assisted Intervention. pp. 594–603. Springer (2025)
11. Oquab, M., Darcet, T., Moutakanni, T., Vo, H., Szafraniec, M., Khalidov, V., Fernandez, P., Haziza, D., Massa, F., El-Nouby, A., et al.: Dinov2: Learning robust visual features without supervision. arXiv preprint arXiv:2304.07193 (2023)
12. Ouyang, C., Biffi, C., Chen, C., Kart, T., Qiu, H., Rueckert, D.: Self-supervised learning for few-shot medical image segmentation. IEEE Transactions on Medical Imaging **41**(7), 1837–1848 (2022)
13. Pham, C.M., Nguyen, P.L., Nguyen, T.T., Phan, V.M.H., Nguyen, B.P.: Unleashing sam for few-shot medical image segmentation with dual-encoder and automated prompting. In: International Conference on Medical Image Computing and Computer-Assisted Intervention. pp. 670–680. Springer (2025)
14. Ravi, N., Gabeur, V., Hu, Y.T., Hu, R., Ryali, C., Ma, T., Khedr, H., Rädle, R., Rolland, C., Gustafson, L., et al.: Sam 2: Segment anything in images and videos. arXiv preprint arXiv:2408.00714 (2024)
15. Shen, Q., Li, Y., Jin, J., Liu, B.: Q-net: Query-informed few-shot medical image segmentation. In: Proceedings of SAI Intelligent Systems Conference. pp. 610–628. Springer (2023)
16. Tang, H., Liu, X., Sun, S., Yan, X., Xie, X.: Recurrent mask refinement for few-shot medical image segmentation. In: Proceedings of the IEEE/CVF international conference on computer vision. pp. 3918–3928 (2021)
17. Yu, N., Cai, Z., Huang, Y., Tang, X.: Snapseg: Training-free few-shot medical image segmentation with segment anything model. In: International Workshop on Trustworthy Artificial Intelligence for Healthcare. pp. 109–122. Springer (2024)
18. Zhang, R., Jiang, Z., Guo, Z., Yan, S., Pan, J., Ma, X., Dong, H., Gao, P., Li, H.: Personalize segment anything model with one shot. arXiv preprint arXiv:2305.03048 (2023)
19. Zhou, W., Guan, G., Gao, Y., Si, P., Xu, M., Yan, Q.: Masg-sam: Enhancing few-shot medical image segmentation with multi-scale attention and semantic guidance. IEEE Journal of Biomedical and Health Informatics (2025)
20. Zhu, Y., Wang, S., Xin, T., Zhang, H.: Few-shot medical image segmentation via a region-enhanced prototypical transformer. In: International Conference on Medical Image Computing and Computer-Assisted Intervention. pp. 271–280. Springer (2023)
21. Zhuang, X.: Multivariate mixture model for myocardial segmentation combining multi-source images. IEEE transactions on pattern analysis and machine intelligence **41**(12), 2933–2946 (2018)