RAP: Retrieve, Adapt, and Prompt-Fit for Training-Free Few-Shot Medical Image Segmentation

*Bangpu Chen[1], *,†Zhihao Mao[1]

***Abstract*—Few-shot medical image segmentation (FSMIS) has achieved notable progress, yet most existing methods mainly rely on semantic correspondences from scarce annotations while under-utilizing a key property of medical imagery: anatomical targets exhibit repeatable high-frequency morphology (e.g., boundary geometry and spatial layout) across patients and acquisitions. We propose RAP, a training-free framework that Retrieves, Adapts, and Prompt-Fits Segment Anything Model 2 (SAM2) for FSMIS. First, RAP retrieves morphologically compatible supports from an archive using DINOv3 features to reduce brittleness to single-support choice. Second, it adapts the retrieved support mask to the query by fitting boundary-aware structural cues, yielding an anatomy-consistent pre-mask under domain shifts. Third, RAP converts the pre-mask into prompts by sampling positive points via Voronoi partitioning and negative points via sector-based sampling, and feeds them into SAM2 for final refinement—without any fine-tuning. Extensive experiments on multiple medical segmentation benchmarks show that RAP consistently surpasses prior FSMIS baselines and achieves state-of-the-art performance. Overall, RAP demonstrates that explicit structural fitting combined with retrieval-augmented prompting offers a simple and effective route to robust training-free few-shot medical segmentation.**

***Index Terms*—few-shot segmentation, medical image segmentation, segment anything model, training-free, shape prior**

I. Introduction

Medical image segmentation is fundamental to clinical diagnosis, treatment planning, and disease monitoring [1]. Accurate delineation of anatomical structures such as cardiac chambers, abdominal organs, and tumors enables quantitative analysis that directly impacts patient outcomes. Deep learning methods have achieved remarkable success in this domain, yet they typically require large amounts of pixel-level annotations that are expensive and time-consuming to obtain.

Few-shot segmentation has emerged as a promising paradigm to address the annotation scarcity problem, where models learn to segment novel classes given only one or a few labeled support images [2], [3]. However, existing few-shot methods face critical challenges in medical imaging. First, they often require task-specific training on base classes, limiting their generalization to unseen domains and modalities [4], [5]. Second, the quality of segmentation heavily depends on the visual similarity between support and query images, which is problematic given the high variability in medical scans due to different scanners, protocols, and patient populations. Third, most methods fail to exploit the strong anatomical priors inherent in medical images, where organs exhibit consistent shapes, positions, and topological structures across patients. More broadly, retrieval-oriented systems suggest that external repositories can complement frozen models without retraining [6], [7].

We observe that medical images possess two distinctive properties that can be leveraged for training-free few-shot segmentation. **(1) Boundary-centric characteristics**: The discriminative information in medical images concentrates around object boundaries, where intensity gradients and texture patterns differ significantly between foreground and background regions. **(2) Strong anatomical priors**: Organs of the same category share similar shapes and spatial layouts, providing geometric constraints that can guide the segmentation process.

Based on these observations, we propose **RAP** (**R**etrieve, **A**dapt, and **P**rompt-Fit), a training-free framework for few-shot medical image segmentation. RAP operates in three stages: (1) *Retrieve* leverages DINOv3 [8] features to identify the most semantically similar support sample from a database; (2) *Adapt* employs oriented chamfer matching guided by DINO-based semantic gating to align the support shape onto the query image, producing a preliminary mask; and (3) *Prompt-Fit* generates geometrically-aware point prompts through Voronoi partitioning [9], which are fed into SAM2 [10] to obtain the final segmentation. Unlike existing methods that require task-specific training or fine-tuning, RAP directly leverages pre-trained foundation models (DINOv3 and SAM2) without any gradient updates, requiring only a small database of annotated samples for retrieval.

Our main contributions are summarized as follows:

- We propose RAP, a novel training-free framework for few-shot medical image segmentation that synergistically combines vision foundation models with geometric shape priors.
- We introduce an oriented chamfer matching algorithm with DINO-based semantic gating that enables robust shape adaptation from support to query images, effectively handling appearance variations across different domains.
- We design a Voronoi-based prompt generation strategy that produces geometrically-aware positive and negative points, significantly improving SAM2's segmentation quality in medical imaging scenarios.

[1]School of Computer Science, China University of Geosciences, Wuhan, China. *Equal contribution. †Corresponding author: 20231001111@cug.edu.cn.

- Extensive experiments on three medical imaging datasets demonstrate that RAP achieves competitive performance compared to state-of-the-art methods while requiring no training.

Upon acceptance, we will release our code and pre-built support databases to facilitate reproducibility and future research in training-free medical image segmentation.

II. Related Work

Few-shot Medical Image Segmentation. Few-shot segmentation aims to segment novel classes with limited labeled examples. Prototype-based methods [2], [11] extract class prototypes from support features and match them against query features. SSL-ALPNet [3] introduces self-supervised learning to improve prototype quality, while RPT [4] partitions foreground into multiple regions for comprehensive feature representation. PATNet [5], IFA [12], and FAMNet [13] address cross-domain challenges through domain-invariant features and frequency-aware matching. However, these methods require task-specific training and struggle to generalize across different imaging modalities.

Shape Priors in Medical Segmentation. Incorporating shape priors has been a long-standing approach to improve segmentation robustness [14]. Recent methods integrate shape information through shape-aware loss functions [15] or boundary attention mechanisms. MAUP [16] proposes a training-free approach using multi-center prompting with Voronoi-based region partitioning [9]. However, existing methods either require training to learn shape representations or lack explicit geometric alignment between support and query images.

Foundation Models for Segmentation. SAM [17] and SAM2 [10] have demonstrated impressive zero-shot segmentation capabilities by leveraging large-scale pre-training. Several works have explored adapting SAM for medical imaging through prompt engineering [18], [19]. Self-supervised vision transformers like DINOv2 [20] and DINOv3 [8] provide semantically rich features for retrieval and correspondence. Our work combines DINOv3 for semantic guidance and SAM2 for segmentation, bridging them through geometric shape adaptation without any training.

III. Preliminary

A. Problem Formulation

We consider the one-shot medical image segmentation setting. Given a support set $\mathcal{S} = \{(I_s, M_s)\}$ containing a single image $I_s \in \mathbb{R}^{H\times W}$ with its corresponding binary mask $M_s \in \{0,1\}^{H\times W}$, and a query image $I_q \in \mathbb{R}^{H\times W}$, the goal is to predict the segmentation mask $\hat{M}_q$ for the query image. In practice, we maintain a support database $\mathcal{D} = \{(I_s^{(i)}, M_s^{(i)})\}_{i=1}^{N}$ containing N labeled samples, from which the most suitable support is retrieved for each query.

B. Notations

We denote query and support images as I_q and I_s, with M_s being the ground-truth mask of the support image and $\hat{M}_q$ being the predicted mask for the query. The DINO feature maps are denoted as $\mathbf{F}_q, \mathbf{F}_s \in \mathbb{R}^{h\times w\times d}$, where $h \times w$ is the spatial resolution and d is the feature dimension. We use $E_q \in \mathbb{R}^{H\times W}$ to represent the edge map of the query image, $D_s \in \mathbb{R}^{H\times W}$ for the signed distance field computed from M_s, and $M_{\text{pre}} \in \{0,1\}^{H\times W}$ for the preliminary mask obtained after shape adaptation. The sets of positive and negative prompt points are denoted as $\mathcal{P}^+$ and $\mathcal{P}^-$, respectively.

C. Foundation Models

DINOv3. DINOv3 [8] is a self-supervised vision transformer that extends DINOv2 with improved feature representations. It produces semantically rich patch-level features that capture both local texture and global context. Given an image I, DINOv3 extracts features $\mathbf{F} = \text{DINO}(I) \in \mathbb{R}^{h\times w\times d}$, where each spatial location encodes a d-dimensional descriptor. These features exhibit strong correspondence properties, making them suitable for cross-image matching and retrieval tasks.

Segment Anything Model 2 (SAM2). SAM2 [10] is an upgraded promptable segmentation model that extends SAM with improved architecture and training. SAM2 takes an image and a set of prompts (points, boxes, or masks) as input and produces segmentation masks. Given point prompts $\mathcal{P} = \{(p_i, l_i)\}$ where $p_i \in \mathbb{R}^2$ denotes the 2D coordinates and $l_i \in \{0,1\}$ indicates positive or negative label, SAM2 outputs $\hat{M} = \text{SAM2}(I, \mathcal{P})$. The quality of segmentation critically depends on the accuracy and placement of prompt points, which motivates our geometry-aware prompt generation strategy.

IV. Methodology

A. Overview

Fig. 1 illustrates the overall framework of RAP. Given a query image I_q, our method operates in three stages. **(1) Retrieve**: We extract global DINO features and retrieve the most semantically similar support sample (I_s, M_s) from the database. **(2) Adapt**: We employ oriented chamfer matching guided by DINO-based semantic gating to align the support mask shape onto the query image, yielding a preliminary mask M_{pre}. **(3) Prompt-Fit**: We apply Voronoi partitioning on M_{pre} to generate geometrically-distributed positive points, sample negative points from boundary sectors, and feed these prompts along with a bounding box to SAM2 for final segmentation.

B. Retrieve Stage

The retrieval stage aims to find the support sample whose anatomical structure best matches the query. We leverage DINOv3 to extract semantically meaningful global descriptors. For each image I, we compute the feature map $\mathbf{F} = \text{DINO}(I)$ and obtain the global descriptor by averaging over spatial locations:

$$\mathbf{g} = \frac{1}{hw}\sum_{i,j}\mathbf{F}_{i,j} \tag{1}$$

For support images, we optionally compute a masked descriptor using only foreground regions to better capture target-specific features.

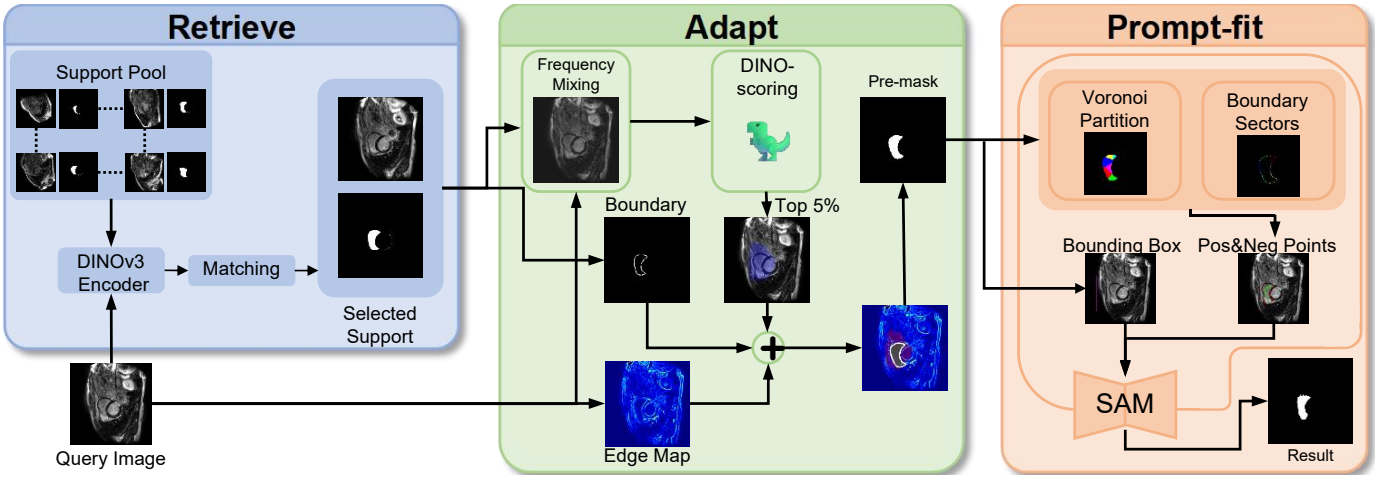

Fig. 1: Overview of the proposed RAP framework. Given a query image, we first retrieve the most similar support from a database using DINOv3 features (Retrieve). Then, we adapt the support shape to the query through oriented chamfer matching with semantic gating (Adapt). Finally, we generate Voronoi-based point prompts and feed them to SAM2 for segmentation (Prompt-Fit).

Given the query descriptor $\mathbf{g}_q$ and a database of support descriptors $\{\mathbf{g}_s^{(i)}\}_{i=1}^{N}$, we compute cosine similarities and retrieve the top-k matches:

$$\text{sim}(q, s^{(i)}) = \frac{\mathbf{g}_q \cdot \mathbf{g}_s^{(i)}}{\|\mathbf{g}_q\| \|\mathbf{g}_s^{(i)}\|} \tag{2}$$

In practice, we select the second-best match (rank-2) to avoid potential self-matching artifacts when query and database share similar sources.

C. Adapt Stage

The adapt stage transfers the support mask shape to the query image while accounting for geometric variations. This stage consists of three components: frequency-domain style adaptation, DINO-based semantic gating, and oriented chamfer matching.

1) Frequency-domain Style Adaptation: To reduce domain shift between support and query images caused by different scanners or imaging protocols, we employ wavelet-based style transfer. As illustrated in Fig. 2, we decompose both images using discrete wavelet transform (DWT) into four subbands: LL (low-frequency approximation), LH, HL, and HH (high-frequency details). The low-frequency subband captures global appearance such as brightness and contrast, while high-frequency subbands preserve edges and textures. We replace the support's LL subband with the query's LL subband and apply inverse wavelet transform to obtain a style-adapted support image:

$$\tilde{I}_s = \text{IDWT}(\text{LL}_q, \text{LH}_s, \text{HL}_s, \text{HH}_s) \tag{3}$$

This operation transfers the query's global appearance to the support while preserving the support's edge structures, which are critical for shape matching.

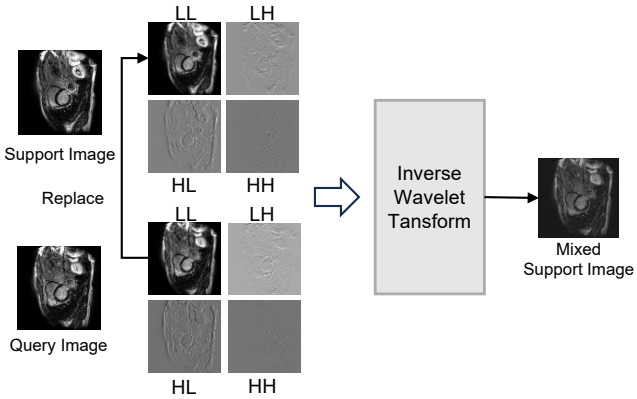

Fig. 2: Frequency-domain style adaptation via wavelet transform. The support's low-frequency (LL) subband is replaced with the query's LL to align global appearance while preserving high-frequency edge details.

2) DINO-based Semantic Gating: To narrow down the search space for shape matching, we compute a semantic similarity map between support and query. We first cluster the support mask into K regions using K-Means on DINO features:

$$\{R_k\}_{k=1}^{K} = \text{KMeans}(\mathbf{F}_s[M_s > 0], K) \tag{4}$$

For each region, we compute a prototype by averaging features within the region and measure its similarity to query features:

$$S_k(i,j) = \frac{\mathbf{F}_q(i,j) \cdot \mathbf{c}_k}{\|\mathbf{F}_q(i,j)\| \|\mathbf{c}_k\|} \tag{5}$$

where $\mathbf{c}_k$ is the prototype of region R_k.

We select the top-K' regions with highest average similarity and aggregate their maps to form a gating mask G:

$$G = \bigcup_{k \in \text{top-}K'} \mathbb{1}[S_k > \tau_s] \tag{6}$$

where τ_s is a threshold determined by the q-th quantile of similarity values. This mask identifies candidate regions in the query where the target structure is likely to appear.

3) Oriented Chamfer Matching: We compute the query edge map E_q by fusing multi-scale Laplacian of Gaussian (LoG) responses with Sobel gradients:

$$E_q = w_{\text{LoG}} \cdot \max_{\sigma} |\text{LoG}_\sigma(I_q)| + w_{\text{grad}} \cdot \|\nabla I_q\| \tag{7}$$

where $\sigma \in \{1, 2, 4, 8\}$ are the LoG scales.

From the support mask M_s, we extract boundary points $\{(x_i, y_i)\}$ along with their normal directions $\{\theta_i\}$ by computing tangent vectors along the contour. These points form the shape template to be matched against the query.

To enable orientation-aware matching, we partition the query edge pixels into K_θ angular bins based on their gradient directions. For each bin k, we compute a distance transform DT_k on the corresponding edge pixels:

$$\text{DT}_k(x, y) = \min_{(x', y') \in E_q^{(k)}} \|(x, y) - (x', y')\| \tag{8}$$

where $E_q^{(k)}$ denotes edge pixels in angular bin k.

We search for the optimal transformation (t_x, t_y, s, r) (translation, scale, rotation) that minimizes the oriented chamfer distance:

$$\mathcal{C}(t_x, t_y, s, r) = \frac{1}{|\mathcal{B}|} \sum_{(x_i, y_i, \theta_i) \in \mathcal{B}} \text{DT}_{k(\theta_i + r)}(T(x_i, y_i)) \tag{9}$$

where $\mathcal{B}$ is the support boundary template, $T(\cdot)$ applies the transformation, and $k(\theta)$ maps angle to bin index. The search is constrained within the gating mask G to improve efficiency and robustness.

The optimal transformation is applied to the support mask to obtain the preliminary mask:

$$M_{\text{pre}} = T^*(M_s) \cap G \tag{10}$$

D. Prompt-Fit Stage

The prompt-fit stage converts the preliminary mask into effective SAM2 prompts through geometry-aware sampling.

1) Voronoi-based Positive Points: Inspired by recent works on region-based prompting [4], [16], we apply Farthest Point Sampling (FPS) on M_{pre} to obtain N_v seed points that are maximally spread:

$$\mathcal{V} = \text{FPS}(M_{\text{pre}}, N_v) \tag{11}$$

These seeds induce a Voronoi partition of the mask region, where each cell V_i contains pixels closest to seed v_i.

For each Voronoi cell, we select its centroid as a positive prompt point:

$$\mathcal{P}^+ = \{\text{centroid}(V_i) \mid i = 1, \ldots, N_v\} \tag{12}$$

This strategy ensures that positive points are geometrically distributed across the entire target region, providing comprehensive coverage for SAM2.

2) Sector-based Negative Points: We partition the boundary of M_{pre} into N_s angular sectors centered at the mask centroid. Within each sector, we sample negative points from the exterior region based on DINO dissimilarity scores:

$$\mathcal{P}^- = \bigcup_{j=1}^{N_s} \arg\min_{p \in \text{Sector}_j \setminus M_{\text{pre}}} S_{\text{DINO}}(p) \tag{13}$$

where $S_{\text{DINO}}(p)$ is the semantic similarity at point p. This ensures negative points are distributed around the object boundary with low similarity to the foreground.

3) SAM2 Inference: We combine the positive points $\mathcal{P}^+$, negative points $\mathcal{P}^-$, and a bounding box $\mathcal{B}_{\text{box}}$ derived from M_{pre} as the complete prompt set:

$$\hat{M}_q = \text{SAM2}(I_q, \mathcal{P}^+, \mathcal{P}^-, \mathcal{B}_{\text{box}}) \tag{14}$$

The bounding box provides a spatial prior that constrains SAM2's attention to the region of interest, while the point prompts offer fine-grained foreground/background guidance.

V. Experiments

A. Datasets

We evaluate our method on three widely-used medical image segmentation benchmarks spanning different modalities and anatomical structures. **Abd-MRI** [21] consists of 20 cases of abdominal MRI scans from the ISBI 2019 Combined Healthy Abdominal Organ Segmentation (CHAOS) challenge, with annotations for four abdominal organs: liver, left kidney (LK), right kidney (RK), and spleen. **Abd-CT** [22] contains 20 cases of abdominal CT scans from the MICCAI 2015 Multi-Atlas Labeling Beyond the Cranial Vault (BTCV) challenge, with the same four organs annotated. **Card-MRI** [23] comprises 45 cases of cardiac MRI scans from the MICCAI 2019 Multi-Sequence Cardiac MRI Segmentation Challenge, with three cardiac structures annotated: left ventricle blood pool (LV-BP), left ventricle myocardium (LV-MYO), and right ventricle (RV).

B. Evaluation Metrics

Following prior works [4], [16], we adopt the Dice Similarity Coefficient (DSC) as the primary evaluation metric: $\text{DSC} = \frac{2|P \cap G|}{|P| + |G|} \times 100\%$, where P and G denote the predicted and ground-truth masks, respectively. All experiments are conducted under the 1-way 1-shot setting to simulate the scarcity of labeled data in medical scenarios.

C. Implementation Details

We employ the pre-trained DINOv3-ViT-L/16 [8] as the feature extractor and SAM2-ViT-H [10] as the segmentation backbone. All images are resized to 512×512 for DINO feature extraction. For oriented chamfer matching, we use $K_\theta = 8$ angular bins, scales $s \in \{0.6, 0.7, \ldots, 1.4\}$, and rotations $r \in \{-20°, -10°, 0°, 10°, 20°\}$. The DINO gating

TABLE I: Quantitative comparison (Dice %) on Abd-MRI and Abd-CT datasets.

Method	Abd-MRI					Abd-CT				
	Liver	LK	RK	Spleen	Mean	Liver	LK	RK	Spleen	Mean
PANet	39.24	26.47	37.35	26.79	32.46	40.29	30.61	26.66	30.21	31.94
SSL-ALPNet	70.74	55.49	67.43	58.39	63.01	71.38	34.48	32.32	51.67	47.46
RPT	49.22	42.45	47.14	48.84	46.91	65.87	40.07	35.97	51.22	48.28
PATNet	57.01	50.23	53.01	51.63	52.97	75.94	46.62	42.68	63.94	57.29
IFA	50.22	35.99	34.00	42.21	40.61	46.62	25.13	26.56	24.85	30.79
FAMNet	73.01	57.28	**74.68**	58.21	65.79	73.57	57.79	61.89	65.78	64.75
MAUP	78.16	58.23	72.34	59.65	67.09	78.25	59.41	**71.80**	60.38	67.46
RAP (Ours)	**79.84**	**60.28**	72.06	**64.46**	**69.16**	**80.16**	**60.62**	69.54	**67.93**	**69.56**

threshold is set to the 90th percentile of similarity values. For Voronoi-based prompt generation, we sample $N_v = 6$ positive points via FPS and $N_s = 8$ negative points from boundary sectors. During retrieval, we select the rank-2 match instead of rank-1 to avoid potential self-matching when query and support slices originate from the same patient volume in leave-one-out evaluation. All experiments are conducted on a single NVIDIA RTX 3090 GPU with 24GB memory using PyTorch.

D. Compared Methods

We compare RAP with several state-of-the-art few-shot medical image segmentation methods. The training-based methods include PANet [2], SSL-ALPNet [3], RPT [4], PATNet [5], IFA [12], and FAMNet [13], all of which require training on source medical domains. For training-free comparison, we include MAUP [16], which leverages DINOv2 and SAM without gradient updates.

VI. Results and Discussion

A. Comparison with State-of-the-art

Table I presents the quantitative comparison on Abd-MRI and Abd-CT datasets. Our RAP achieves the best average Dice scores of 69.16% on Abd-MRI and 69.56% on Abd-CT, outperforming all compared methods. Notably, RAP surpasses the recent MAUP [16] by 2.07% and 2.10% on the two datasets, respectively, validating the effectiveness of our shape-guided approach.

Table II shows the results on the Card-MRI dataset. RAP achieves an average Dice of 75.27%, improving upon MAUP by 2.14%. The improvement is particularly significant on the challenging LV-MYO structure (+1.27%) and RV (+3.58%), where our oriented chamfer matching effectively captures the thin myocardium boundaries.

B. Ablation Study

We conduct ablation studies on the RV structure of Card-MRI to validate the contribution of each component in RAP. As shown in Table III, each module progressively improves the segmentation performance.

Effect of OCM. Adding Oriented Chamfer Matching improves Dice from 78.29% to 80.17% (+1.88%), demonstrating the importance of orientation-aware shape alignment.

Effect of SG. DINO-based Semantic Gating further improves to 81.02% (+0.85%) by constraining the search space and preventing false matches.

TABLE II: Quantitative comparison (Dice %) on Card-MRI dataset.

Method	LV-BP	LV-MYO	RV	Mean
PANet	51.42	25.75	25.75	36.66
SSL-ALPNet	83.47	22.73	66.21	57.47
RPT	60.84	42.28	57.30	53.47
PATNet	65.35	50.63	68.34	61.44
IFA	50.43	31.32	30.74	37.50
FAMNet	86.64	51.82	76.26	71.58
MAUP	88.36	52.74	78.29	73.13
RAP (Ours)	**89.92**	**54.01**	**81.87**	**75.27**

TABLE III: Ablation study on Card-MRI (RV structure). OCM: Oriented Chamfer Matching, SG: Semantic Gating, VP: Voronoi-based Prompts.

OCM	SG	VP	Mean Dice (%)
			78.29
✓			80.17
✓	✓		81.02
✓	✓	✓	**81.87**

Effect of VP. Voronoi-based Prompts achieve 81.87% (+0.85%) through comprehensive geometric coverage of the target region.

C. Qualitative Analysis

Fig. 3 presents qualitative comparisons on the Card-MRI dataset for three cardiac structures (LV-BP, LV-MYO, and RV). Our RAP produces more accurate boundaries compared to baseline methods, especially for the thin myocardium (LV-MYO) where precise boundary delineation is critical. The Voronoi-based prompts effectively guide SAM2 to segment complete structures without missing thin regions or including spurious areas.

D. Discussion

We clarify that "training-free" refers to requiring no gradient-based optimization or weight updates during inference. While RAP utilizes a pre-built support database for retrieval, this database only stores raw images and masks without learned representations, distinguishing our approach from methods that require task-specific training on source domains.

RAP offers several key advantages. First, it enables immediate deployment to new domains without retraining. Second, the shape-guided approach is robust to appearance variations across different scanners and protocols. Third, the oriented chamfer matching effectively handles gradient direction information for subtle boundaries.

Regarding the support database, our experiments use leave-one-out evaluation where each test case retrieves from remain-

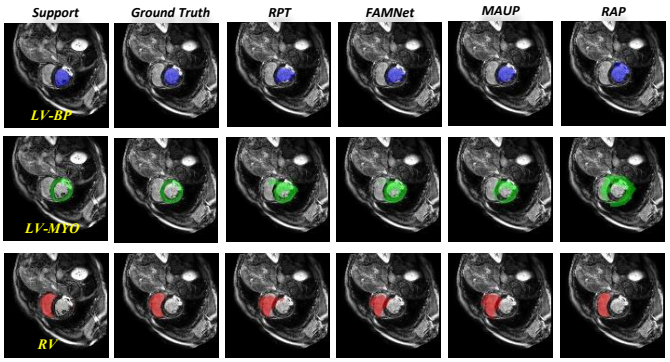

Fig. 3: Qualitative results on Card-MRI for three cardiac structures. Our method produces more accurate boundaries, especially for the thin myocardium (LV-MYO).

ing samples. The retrieval mechanism is robust to database variations: since DINOv3 features capture semantic similarity rather than exact pixel matching, morphologically compatible supports can be retrieved even from small databases (20-45 cases in our benchmarks).

However, performance may degrade for rare anatomical structures or pathological cases that deviate significantly from normal anatomy in the database.

VII. Conclusion

We presented RAP, a training-free framework for few-shot medical image segmentation that synergistically combines vision foundation models with geometric shape priors. Our method introduces oriented chamfer matching with DINO-based semantic gating for robust shape adaptation, and Voronoi-based prompt generation for geometry-aware SAM2 guidance. Experiments on three datasets demonstrate that RAP achieves competitive performance compared to training-based methods while requiring no task-specific optimization. Future work includes extending RAP to 3D volumetric segmentation by leveraging inter-slice consistency.

Acknowledgments

References

[1] R. Azad, E. K. Aghdam, A. Rauland, Y. Jia, A. H. Avval, A. Bozorgpour, S. Karimijafarbigloo, J. P. Cohen, E. Adeli, and D. Merhof, "Medical image segmentation review: The success of u-net," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 46, no. 12, pp. 10 076–10 095, 2024.

[2] 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.

[3] C. Ouyang, C. Biffi, C. Chen, T. Kart, H. Qiu, and D. Rueckert, "Self-supervised learning for few-shot medical image segmentation," *IEEE Transactions on Medical Imaging*, vol. 41, no. 7, pp. 1837–1848, 2022.

[4] Y. Zhu, S. Wang, T. Xin, and H. Zhang, "Few-shot medical image segmentation via a region-enhanced prototypical transformer," in *International Conference on Medical Image Computing and Computer-Assisted Intervention*. Springer, 2023, pp. 271–280.

[5] S. Lei, X. Zhang, J. He, F. Chen, B. Du, and C.-T. Lu, "Cross-domain few-shot semantic segmentation," in *European Conference on Computer Vision*, 2022, pp. 73–90.

[6] M. Li, K. Li, Y. Liu, J. Chen, Z. Zheng, Y. Wu, and X. Chen, "Hierarchical scheduling for multi-vector image retrieval," *arXiv preprint arXiv:2510.08976*, 2025.

[7] Z. Zheng, Z. Mao, S. Tian, M. Li, J. Chen, X. Sun, Z. Zhang, X. Liu, D. Cao, H. Mei *et al.*, "Heisd: Hybrid speculative decoding for embodied vision-language-action models with kinematic awareness," *arXiv preprint arXiv:2603.17573*, 2026.

[8] O. Siméoni, H. V. Vo, M. Seitzer, F. Baldassarre, M. Oquab, C. Jose, V. Khalidov, M. Szafraniec, S. Yi, M. Ramamonjisoa, F. Massa, D. Haziza, L. Wehrstedt, J. Wang, T. Darcet, T. Moutakanni, L. Sentana, C. Roberts, A. Vedaldi, J. Tolan, J. Brandt, C. Couprie, J. Mairal, H. Jégou, P. Labatut, and P. Bojanowski, "Dinov3," 2025. [Online]. Available: https://arxiv.org/abs/2508.10104

[9] F. Aurenhammer, "Voronoi diagrams: a survey of a fundamental geometric data structure," *ACM Computing Surveys (CSUR)*, vol. 23, no. 3, pp. 345–405, 1991.

[10] N. Ravi, V. Gabeur, Y.-T. Hu, R. Hu, C. Ryali, T. Ma, H. Khedr, R. Rädle, C. Rolland, L. Gustafson, E. Mintun, J. Pan, K. V. Alwala, N. Carion, C.-Y. Wu, R. Girshick, P. Dollár, and C. Feichtenhofer, "Sam 2: Segment anything in images and videos," 2024. [Online]. Available: https://arxiv.org/abs/2408.00714

[11] J. Snell, K. Swersky, and R. Zemel, "Prototypical networks for few-shot learning," *Advances in neural information processing systems*, vol. 30, 2017.

[12] J. Nie, Y. Xing, G. Zhang, P. Yan, A. Xiao, Y.-P. Tan, A. C. Kot, and S. Lu, "Cross-domain few-shot segmentation via iterative support-query correspondence mining," *IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pp. 3380–3390, 2024.

[13] Y. Bo, Y. Zhu, L. Li, and H. Zhang, "Famnet: Frequency-aware matching network for cross-domain few-shot medical image segmentation," in *AAAI Conference on Artificial Intelligence*, 2025.

[14] H. Kervadec, J. Bouchtiba, C. Desrosiers, E. Granger, J. Dolz, and I. B. Ayed, "Boundary loss for highly unbalanced segmentation," in *International conference on medical imaging with deep learning*, 2019, pp. 285–296.

[15] J. Ma, J. Chen, M. Ng, R. Huang, Y. Li, C. Li, X. Yang, and A. L. Martel, "Loss odyssey in medical image segmentation," *Medical Image Analysis*, vol. 71, p. 102035, 2021.

[16] Y. Zhu and H. Zhang, " MAUP: Training-free Multi-center Adaptive Uncertainty-aware Prompting for Cross-domain Few-shot Medical Image Segmentation ," in *proceedings of Medical Image Computing and Computer Assisted Intervention – MICCAI 2025*, vol. LNCS 15966. Springer Nature Switzerland, September 2025.

[17] 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.

[18] K. Huang, T. Zhou, H. Fu, Y. Zhang, Y. Zhou, C. Gong, and D. Liang, "Learnable prompting sam-induced knowledge distillation for semi-supervised medical image segmentation," *IEEE Transactions on Medical Imaging*, pp. 1–1, 2025.

[19] R. Wang, Z. Yang, and Y. Song, "Osam-fundus: A training-free, one-shot segmentation framework for optic disc and cup in fundus images," *Biomedical Signal Processing and Control*, vol. 100, p. 107069, 2025.

[20] M. Oquab, T. Darcet, T. Moutakanni, H. Vo, M. Szafraniec, V. Khalidov, P. Fernandez, D. Haziza, F. Massa, A. El-Nouby *et al.*, "Dinov2: Learning robust visual features without supervision," *arXiv preprint arXiv:2304.07193*, 2023.

[21] 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.

[22] 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 *Proceedings of the MICCAI—Workshop Challenge*, vol. 5, 2015, p. 12.

[23] X. Zhuang, "Multivariate mixture model for myocardial segmentation combining multi-source images," *IEEE transactions on pattern analysis and machine intelligence*, vol. 41, no. 12, pp. 2933–2946, 2018.