GenSAM: One-Shot Medical Image Segmentation via Diffusion-Based Test-Time Adaptation

*,†Zhihao Mao[1], *Bangpu Chen[1], *Xuehai Chen[1]

***Abstract*—Medical image segmentation is fundamental for clinical diagnosis, yet developing deep learning models remains challenging due to the scarcity of pixel-level annotations. Few-shot segmentation offers a promising solution by leveraging only a few labeled examples, but existing methods struggle with the support-query misalignment problem—a single support image often fails to represent the diverse appearances of target organs across patients. In this paper, we propose GenSAM, a novel training-free framework that addresses this challenge by leveraging diffusion models for test-time adaptation. Our key insight is that query-guided style injection can synthesize support views that are visually closer to the query domain while maintaining anatomical consistency, effectively bridging the support-query gap without additional annotations. GenSAM comprises three components: (1) Generative Support Expansion (GSE), which synthesizes diverse augmented views using query-guided diffusion-based style transfer; (2) Multi-center Prompting, which generates positive and negative point prompts from similarity maps to guide SAM2; and (3) Consistency-Aware Ensemble, which aggregates predictions weighted by confidence scores to filter unreliable results. Experiments on three medical imaging benchmarks demonstrate that GenSAM significantly outperforms existing training-free methods and achieves competitive performance with training-based approaches, highlighting the potential of combining generative and discriminative foundation models for medical image analysis.**

***Index Terms*—medical image segmentation, few-shot learning, diffusion models, SAM2, foundation models, test-time adaptation**

I. Introduction

Medical image segmentation plays a fundamental role in modern healthcare, enabling precise delineation of anatomical structures for disease diagnosis and treatment planning [1]. However, the development of deep learning based segmentation models heavily relies on extensive pixel-level annotations, which are both costly and time-consuming to acquire in medical imaging scenarios [2]. This challenge is further compounded by the heterogeneity of medical scans, encompassing diverse modalities (e.g., MRI, CT, X-Ray) and anatomical structures, which leads to the difficulty of intra-class generalization [3].

To address the annotation scarcity, few-shot medical image segmentation (FSMIS) has emerged as a promising paradigm that aims to segment target organs with only a few annotated examples [4], [5]. Recently, foundation models such as the Segment Anything Model (SAM) [6] have demonstrated remarkable generalization capabilities, enabling training-free few-shot segmentation with appropriate prompting strategies [7], [8]. However, these methods typically rely on a single support exemplar (one-shot setting), which fails to capture the high intra-class variance inherent in medical images. The support image may exhibit different contrast phases, intensity distributions, or patient-specific variations compared to the query image, leading to the "support-query misalignment" problem that severely degrades segmentation performance.

Existing approaches to mitigate this misalignment can be categorized into two directions. The first direction involves training-based methods [5], [9], [10] that learn to extract domain-invariant features through extensive training on source medical domains. However, these methods sacrifice the universality and ease of deployment, as they require substantial labeled data and computational resources for each new application scenario. The second direction employs retrieval-based strategies that search for additional reference images from external databases [11]. While effective, such methods are constrained by the availability and quality of the retrieval database, and may raise privacy concerns when accessing patient data across institutions.

In this paper, we propose a novel perspective: instead of retrieving real images, we leverage the powerful generative capabilities of diffusion models to synthesize diverse support exemplars guided by the query image. Our key insight is that modern diffusion models can transfer the visual style of the query to the support while preserving its structural integrity. By applying query-guided style injection, we can expand a single support image into multiple augmented views that are visually closer to the query domain, effectively bridging the support-query gap without requiring additional manual annotations.

Based on this insight, we present **GenSAM**, a training-free framework for one-shot medical image segmentation via diffusion-based test-time adaptation. GenSAM consists of three key components: (1) *Generative Support Expansion (GSE)*, which utilizes query-guided diffusion to synthesize diverse augmented views that bridge the visual gap between support and query; (2) *Multi-center Prompting*, which extracts DINOv3 features and generates positive/negative point prompts to guide SAM2 for query segmentation; and (3) *Consistency-Aware Ensemble Inference*, which aggregates predictions from multiple supports based on confidence scores, effectively filtering out hallucinations from the generative model.

Our main contributions can be summarized as follows:

- We propose GenSAM, a training-free framework that

[1]School of Computer Science, China University of Geosciences, Wuhan, China. *Equal contribution. †Corresponding author (Z. Mao): 20231001111@cug.edu.cn. X. Chen: 1486002022@qq.com.

combines generative diffusion models with discriminative foundation models (DINOv3, SAM2) for one-shot medical image segmentation, requiring no domain-specific training or external retrieval databases.
- We introduce Generative Support Expansion (GSE), a query-guided synthesis strategy that effectively bridges the visual gap between support and query images by injecting query style into the support while preserving anatomical structures.
- We design a Consistency-Aware Ensemble mechanism that robustly aggregates predictions from synthesized supports while mitigating the hallucination effects inherent in generative models.
- Extensive experiments on multiple medical segmentation benchmarks demonstrate that GenSAM significantly outperforms existing training-free methods and achieves competitive performance with training-based approaches.

The rest of this paper is organized as follows. Section II reviews related work on few-shot medical image segmentation and diffusion models. Section IV presents our proposed GenSAM framework. Section V describes the experimental setup, and Section VI presents and discusses the results. Finally, Section VII concludes the paper.

Upon acceptance of this paper, we will make our code publicly available to ensure reproducibility and facilitate future research in this area.

II. Related Work

A. Few-shot Medical Image Segmentation

Few-shot medical image segmentation aims to segment target structures with only a few annotated examples [12]. Early approaches adopted prototype-based learning, where PANet [13] and SSL-ALP [4] match support and query features through prototype alignment. Recent works propose multi-prototype representations to capture complex organ structures, such as GMRD [5] and RPT [14], demonstrating that single prototype is insufficient for high intra-class variance in medical images.

Cross-domain few-shot segmentation [15], [16] further addresses domain shift between training and testing. However, these methods still require extensive training on source domains, limiting their practical applicability.

B. Foundation Models for Medical Image Segmentation

The Segment Anything Model (SAM) [6] has revolutionized image segmentation with remarkable zero-shot generalization. While fine-tuning approaches [17], [18] adapt SAM for medical domains, they sacrifice universal applicability and require domain-specific data.

More relevant are training-free adaptations leveraging SAM's prompting mechanism. MAUP [8] introduces multi-center prompting with DINOv2 features, and SynPo [19] improves negative prompt quality. However, these methods are fundamentally limited by the representativeness of the single support image, which our GenSAM addresses through generative expansion.

C. Diffusion Models and Test-Time Adaptation

Diffusion models have shown promise in medical imaging for segmentation [20] and structure-preserving enhancement [21]. For test-time adaptation, Valanarasu et al. [22] adapt batch normalization statistics during inference, but still require pre-trained segmentation models. Our GenSAM differs by leveraging diffusion-based augmentation for test-time adaptation, combining generative priors with discriminative foundation models without any domain-specific training. While the multi-center prompting strategy builds upon MAUP [8], the key contribution of GenSAM lies in the Generative Support Expansion module, which addresses the fundamental limitation of single support representativeness that all existing prompting-based methods suffer from.

III. Preliminary

A. Problem Formulation

We consider the one-shot medical image segmentation task, where the goal is to segment a target anatomical structure in a query image given only a single annotated support example.

Formally, let $I_s \in \mathbb{R}^{H\times W\times C}$ denote the support image and $M_s \in \{0,1\}^{H\times W}$ denote its corresponding binary segmentation mask, where the foreground pixels indicate the target organ. Given a query image $I_q \in \mathbb{R}^{H\times W\times C}$, the objective is to predict its segmentation mask M_q for the same anatomical structure.

The fundamental challenge lies in the *support-query misalignment*: the single support image I_s may exhibit significantly different appearance characteristics compared to I_q due to inter-patient variations, imaging protocols, or acquisition conditions. This distribution shift causes segmentation models to fail when the support exemplar is not representative of the query image.

B. Segment Anything Model 2

The Segment Anything Model 2 (SAM2) [23] is a foundation model consisting of an image encoder $\mathcal{E}_{img}$, a prompt encoder $\mathcal{E}_{prompt}$, and a mask decoder $\mathcal{D}_{mask}$. Given an input image I and point prompts $\mathcal{P} = \{(x_i, y_i, l_i)\}_{i=1}^{N}$ where $l_i \in \{0,1\}$ indicates positive/negative labels, SAM2 produces segmentation mask $M = \mathcal{D}_{mask}(\mathcal{E}_{img}(I), \mathcal{E}_{prompt}(\mathcal{P}))$. In training-free few-shot segmentation, the key challenge is to automatically generate appropriate point prompts from the support pair (I_s, M_s) to guide SAM2 for query segmentation.

C. Diffusion Models for Image-to-Image Translation

Diffusion models learn to reverse a gradual noising process through iterative denoising. For image-to-image translation, instead of starting from pure noise, we add noise to the input image x_0 at an intermediate timestep controlled by the *denoising strength* $\tau \in [0,1]$. A smaller τ preserves more structural information while allowing texture and style variations. This property is crucial for our Generative Support Expansion: we generate diverse appearances while maintaining anatomical structure so that the original mask M_s remains valid for synthesized images.

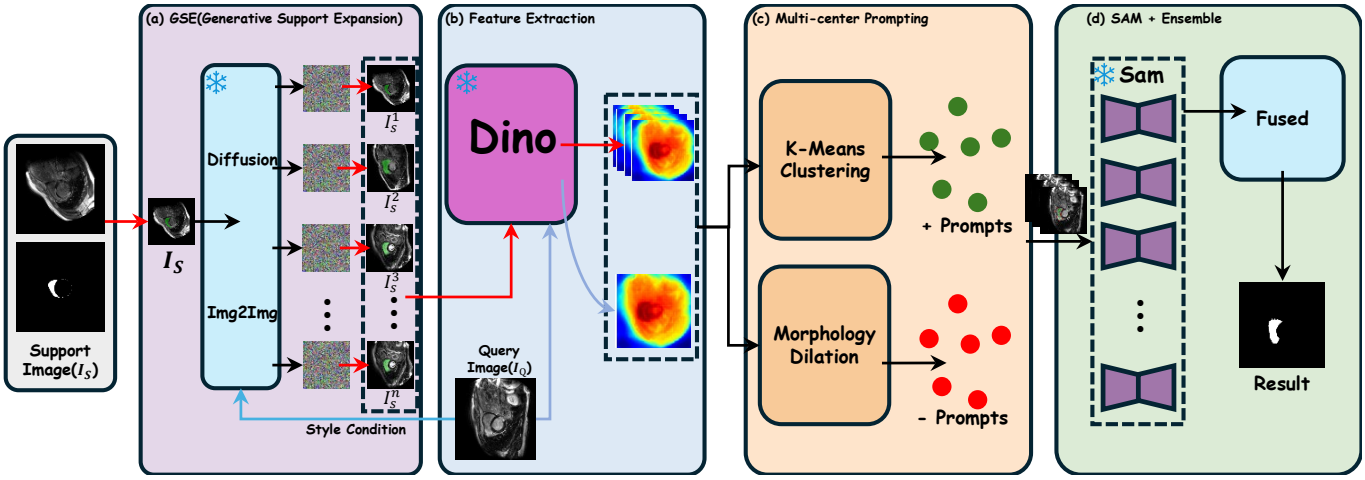

Fig. 1: Overview of the proposed GenSAM framework. Given a support image-mask pair and a query image, we first apply Generative Support Expansion (GSE) to synthesize K diverse augmented views using query-guided style injection while preserving anatomical structures. For each augmented support, we extract DINOv3 features and compute similarity maps with the query image. Multi-center prompting generates positive and negative point prompts, which guide SAM2 to produce segmentation predictions. Finally, consistency-aware ensemble aggregates all predictions weighted by their confidence scores.

IV. Methodology

A. Overall Framework

The overall framework of GenSAM is illustrated in Fig. 1. Given a support image I_s with its mask M_s and a query image I_q, our goal is to segment the target organ in I_q without any training. GenSAM consists of four main stages: (1) Generative Support Expansion synthesizes diverse augmented views of the support image using query-guided diffusion-based style transfer; (2) Feature extraction computes similarity maps between query and each augmented support using DINOv3; (3) Multi-center prompting generates positive and negative point prompts from the similarity maps; (4) Consistency-aware ensemble aggregates predictions from multiple supports based on confidence scores.

B. Generative Support Expansion

The key insight of GenSAM is to expand the support manifold through query-guided generative augmentation. Instead of relying on a single support image that may not represent the diverse appearances of the target organ, we synthesize multiple augmented views that bridge the visual gap between support and query while preserving the anatomical structure. This is motivated by the observation that medical images from different patients or acquisition settings often exhibit significant variations in texture, contrast, and intensity distribution, even for the same anatomical structure.

We employ Stable Diffusion v1.5 with the standard img2img pipeline, enhanced with IP-Adapter for image-conditioned generation. Given the support image I_s and query image I_q, we extract the style representation from the query using the CLIP ViT-H/14 image encoder: $z_q = \mathcal{E}_{\text{CLIP}}(I_q)$. The IP-Adapter injects z_q into the U-Net's cross-attention layers, replacing text conditioning with image-based style guidance. We generate K augmented images $\{I_s^{(k)}\}_{k=1}^K$ through:

$$I_s^{(k)} = \text{SD-img2img}(I_s, z_q, \tau, \epsilon_k) \tag{1}$$

where z_q conditions the generation via cross-attention, $\tau \in [0, 1]$ is the denoising strength, and ϵ_k is the random noise seed providing generation diversity.

The denoising strength τ controls the trade-off between style transfer and structure preservation. With $\tau = 0.2$, only 20% of the diffusion steps are performed, meaning the process starts from a lightly noised version of I_s rather than pure Gaussian noise. This conservative setting constrains generation to local texture and contrast adjustments while preserving the global anatomical structure and organ boundaries from the original support image.

Since the organ boundaries remain unchanged, the original support mask M_s can be directly applied to all augmented images $\{I_s^{(k)}\}_{k=1}^K$, eliminating the need for additional annotations. The augmented support set is defined as:

$$\mathcal{S} = \{(I_s, M_s)\} \cup \{(I_s^{(k)}, M_s)\}_{k=1}^K \tag{2}$$

C. Feature Extraction and Similarity Computation

We employ the pre-trained DINOv3 [24] as our feature extractor due to its strong generalization capability across diverse visual domains. For each support-query pair, we extract dense feature maps:

$$F_s = \mathcal{E}_{\text{DINO}}(I_s), \quad F_q = \mathcal{E}_{\text{DINO}}(I_q) \tag{3}$$

where $F_s, F_q \in \mathbb{R}^{H' \times W' \times C}$ denote the extracted features with spatial dimensions $H' \times W'$ and C channels.

Following [8], we partition the foreground region into N_f sub-regions using Voronoi decomposition [25]. For each sub-

region R_n, we compute a regional prototype via Masked Average Pooling (MAP):

$$\hat{p}_n = \frac{\sum_{(i,j)\in R_n} F_s^{(i,j)}}{|R_n|} \tag{4}$$

where $F_s^{(i,j)}$ denotes the feature vector at spatial location (i, j).

These prototypes are then used to compute similarity maps with the query features. For each prototype $\hat{p}_n$, we compute the cosine similarity map:

$$S_n^{(i,j)} = \frac{F_q^{(i,j)} \cdot \hat{p}_n}{\|F_q^{(i,j)}\| \|\hat{p}_n\|} \tag{5}$$

The mean similarity map $\bar{S} = \frac{1}{N_f} \sum_{n=1}^{N_f} S_n$ aggregates information from all regional prototypes, providing a robust estimate of foreground likelihood. The uncertainty map $U = \frac{1}{N_f} \sum_{n=1}^{N_f} (S_n - \bar{S})^2$ captures regions where different prototypes disagree, which typically correspond to challenging boundary areas.

D. Multi-center Prompting Strategy

Positive Prompts. We employ a two-path approach to generate diverse positive prompts that cover the target organ comprehensively. First, we select candidate pixels where $\bar{S}^{(i,j)} > \theta_s$ (high-similarity threshold) and apply K-means clustering to obtain N_p spatially diverse centroids. This ensures that positive prompts are distributed across the entire organ region rather than concentrated in a single area. Second, we sample additional points from high-uncertainty regions where $U^{(i,j)} > \theta_u$, which typically correspond to challenging boundary areas that benefit from explicit supervision.

Negative Prompts. Negative prompts are crucial for preventing over-segmentation, especially when the target organ has similar intensity to surrounding tissues. We extract the periphery region $\tilde{M}_s = \delta_r(M_s) - M_s$ via morphological dilation with radius r, where $\delta_r(\cdot)$ denotes the dilation operation. A periphery prototype $\hat{p}_{bg}$ is computed by averaging features in this region. We then compute a periphery similarity map $S_{bg} = \cos(F_q, \hat{p}_{bg})$ and select negative prompts from locations with high periphery similarity, effectively marking regions that should be excluded from the segmentation.

E. Consistency-Aware Ensemble Inference

For each augmented support $(I_s^{(k)}, M_s)$ in $\mathcal{S}$, we follow the feature extraction and prompting pipeline described above to generate point prompts, which are then fed to SAM2 [23] to obtain a segmentation prediction $M_q^{(k)}$. SAM2 also outputs a predicted IoU score IoU_k that estimates the quality of each prediction, which we leverage as a confidence measure.

The final mask is computed through confidence-weighted averaging:

$$M_q = \frac{\sum_{k=0}^{K} \text{IoU}_k \cdot M_q^{(k)}}{\sum_{k=0}^{K} \text{IoU}_k} \tag{6}$$

where $k = 0$ corresponds to the original support and $k = 1, \ldots, K$ correspond to the augmented supports. This weighting scheme ensures that high-confidence predictions contribute more to the final result, while unreliable predictions (e.g., caused by diffusion hallucinations) are naturally down-weighted.

The soft prediction is binarized with threshold $\theta = 0.5$ to obtain the final binary mask. To further mitigate hallucinations, we optionally filter out predictions with IoU scores below a confidence threshold τ_{conf} before ensemble, ensuring that only reliable predictions are aggregated.

V. Experiments

A. Datasets

We evaluate GenSAM on three diverse medical imaging datasets that represent different modalities and anatomical structures. All experiments follow the 1-way 1-shot setting to simulate the scarcity of labeled data in medical scenarios.

Abd-MRI [26] consists of 20 cases of abdominal MRI scans from the ISBI 2019 Combined Healthy Organ Segmentation challenge (CHAOS). The target organs include liver, left kidney (LK), right kidney (RK), and spleen.

Abd-CT [27] consists of 30 cases of abdominal CT scans from the MICCAI 2015 Multi-Atlas Labeling Beyond The Cranial Vault challenge (BTCV). The same four abdominal organs are evaluated for segmentation.

Card-MRI [28] contains 45 cases of cardiac MRI scans from the MICCAI 2019 Multi-Sequence Cardiac MRI Segmentation Challenge. The target structures include left ventricle blood pool (LV-BP), left ventricle myocardium (LV-MYO), and right ventricle (RV).

B. Baseline Methods

We compare GenSAM against **training-based methods** including PANet [13], SSL-ALP [4], RPT [14], PATNet [15], IFA [11], and FAMNet [16], as well as **training-free method** MAUP [8].

C. Evaluation Metrics

We adopt the Dice similarity coefficient (DSC) as the primary evaluation metric, which is widely used in medical image segmentation:

$$\text{Dice} = \frac{2|P \cap G|}{|P| + |G|} \tag{7}$$

where P and G denote the predicted and ground truth segmentation masks, respectively. Higher Dice scores indicate better segmentation performance.

D. Implementation Details

We employ DINOv3 ViT-L/14 [24] as the feature extractor for computing similarity maps. For the segmentation component, we utilize SAM2-Large [23] as the promptable segmentation model. For Generative Support Expansion, we use Stable Diffusion v1.5 with IP-Adapter (ViT-H/14) for query-guided style injection, using the DDIM scheduler with 50 inference steps and guidance scale 7.5.

Key hyperparameters: $K = 5$ augmentations, denoising strength $\tau = 0.2$, $N_f = 30$ regional prototypes, $N_p = 5$

TABLE I: Quantitative comparison (Dice %) on Abd-MRI and Abd-CT. Best in **bold**, second-best underlined. † denotes training-free.

Method	Ref.	Abd-MRI					Abd-CT				
		Liver	LK	RK	Spleen	Mean	Liver	LK	RK	Spleen	Mean
PANet	ICCV'19	39.24	26.47	37.35	26.79	32.46	40.29	30.61	26.66	30.21	31.94
SSL-ALP	TMI'22	70.74	55.49	67.43	58.39	63.01	71.38	34.48	32.32	51.67	47.46
RPT	MICCAI'23	49.22	42.45	47.14	48.84	46.91	65.87	40.07	35.97	51.22	48.28
PATNet	ECCV'22	57.01	50.23	53.01	51.63	52.97	75.94	46.62	42.68	63.94	57.29
IFA	CVPR'24	50.22	35.99	34.00	42.21	40.61	46.62	25.13	26.56	24.85	30.79
FAMNet	AAAI'25	73.01	57.28	**75.12**	58.21	65.90	73.57	57.79	61.89	**65.78**	64.75
MAUP†	MICCAI'25	78.16	58.23	72.34	59.65	67.09	**83.15**	59.41	71.80	60.38	68.68
GenSAM†	**Ours**	**80.13**	**67.40**	74.80	**65.70**	**71.97**	82.34	**63.83**	**79.60**	63.32	**72.27**

TABLE II: Quantitative comparison (Dice %) on Card-MRI. † denotes training-free methods.

Method	Ref.	LV-BP	LV-MYO	RV	Mean
PANet	ICCV'19	51.42	25.75	25.75	36.66
SSL-ALP	TMI'22	83.47	22.73	66.21	57.47
RPT	MICCAI'23	60.84	42.28	57.30	53.47
PATNet	ECCV'22	65.35	50.63	68.34	61.44
IFA	CVPR'24	50.43	31.32	30.74	37.50
FAMNet	AAAI'25	86.64	51.82	76.26	71.58
MAUP†	MICCAI'25	**88.36**	52.74	78.29	73.13
GenSAM†	**Ours**	87.40	**65.63**	**83.41**	**78.23**

positive prompt centers, dilation radius $r = 5$, similarity threshold $\theta_s = 0.5$, and confidence threshold $\tau_{\text{conf}} = 0.7$.

Following the standard few-shot medical segmentation protocol [4], [8], we ensure that support and query images are from different patients (case-level separation) to prevent information leakage.

All experiments are conducted on a workstation equipped with an NVIDIA RTX 3090 GPU with 24GB memory.

VI. Results and Discussion

A. Main Results

Table I and Table II present quantitative comparisons on abdominal and cardiac datasets.

GenSAM achieves the highest mean Dice scores on all three datasets, outperforming both training-based and training-free methods. Notably, GenSAM significantly outperforms existing methods despite being completely training-free, highlighting the potential of leveraging diffusion priors for medical image segmentation. Compared to MAUP, the closest training-free baseline, GenSAM achieves consistent improvements across all organ categories, demonstrating the effectiveness of query-guided generative support expansion.

B. Ablation Study

To understand the contribution of each component, we conduct ablation studies on Card-MRI dataset. Table III shows the results.

The results show that removing GSE leads to the most significant performance degradation (5.91% drop in mean Dice on Card-MRI), confirming the critical importance of diverse support augmentation for bridging the support-query gap. Negative prompts effectively prevent over-segmentation, especially for organs with similar intensity to surrounding tissues such as the spleen. Uncertainty-aware prompting helps focus on challenging boundary regions, and confidence-weighted ensemble outperforms simple averaging by filtering out unreliable predictions from occasional diffusion hallucinations.

TABLE III: Ablation study (Dice %) on Card-MRI. GSE: Generative Support Expansion, NP: Negative Prompts, UP: Uncertainty Prompting, CE: Confidence Ensemble.

GSE	NP	UP	CE	LV-BP	LV-MYO	RV	Mean
	✓	✓	✓	82.15	58.42	76.38	72.32
✓		✓	✓	84.67	61.25	79.84	75.25
✓	✓		✓	85.93	63.18	81.26	76.79
✓	✓	✓		86.52	64.37	82.15	77.68
✓	✓	✓	✓	**87.40**	**65.63**	**83.41**	**78.23**

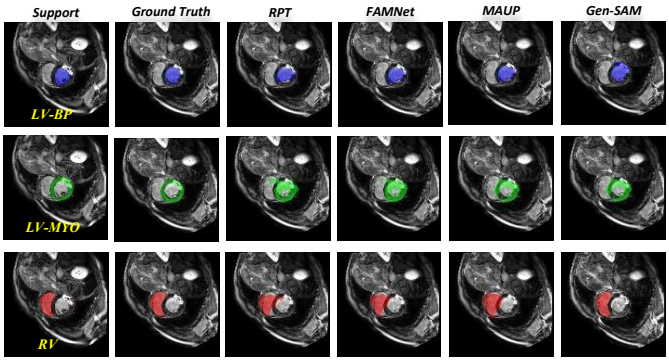

Fig. 2: Qualitative comparison on Card-MRI dataset. GenSAM produces more accurate segmentation boundaries compared to baseline methods, particularly in challenging cases with high appearance variation.

C. Qualitative Results

Fig. 2 presents qualitative comparisons between GenSAM and baseline methods.

The visualization demonstrates that GenSAM effectively captures organ boundaries even when the support and query images exhibit significant appearance differences. In contrast, baseline methods tend to produce incomplete or over-extended segmentations due to the support-query misalignment problem. GenSAM produces smoother boundaries and more complete organ coverage, benefiting from the diverse perspectives provided by multiple augmented supports.

D. Parameter Sensitivity

We analyze the sensitivity of GenSAM to key hyperparameters, including the number of augmentations K and denoising strength τ.

Number of augmentations K: Performance improves as K increases from 1 to 5, then stabilizes. This suggests that 5 augmented supports provide sufficient diversity to cover the appearance variations in query images.

Denoising strength τ: Optimal performance is achieved at $\tau = 0.2$, which provides a good balance between style transfer and structure preservation. Smaller values (e.g., $\tau = 0.1$) preserve too much of the original support appearance, limiting

the diversity of augmented views. Larger values (e.g., $\tau > 0.3$) may distort anatomical structures, causing the original mask to become misaligned with the generated images.

E. Discussion

GenSAM offers training-free deployment, effective handling of support-query misalignment, and robust predictions through ensemble. However, limitations include occasional unrealistic augmentations from the diffusion model, increased inference time, and challenges on extremely low-contrast images.

VII. Conclusion

We presented GenSAM, a training-free framework for one-shot medical image segmentation that addresses the support-query misalignment problem through diffusion-based test-time adaptation. By leveraging structure-preserving image-to-image translation to expand the support manifold, GenSAM effectively transforms the one-shot setting into a multi-shot paradigm without additional annotations. Experiments on three medical imaging benchmarks demonstrate that GenSAM significantly outperforms existing training-free methods while achieving competitive performance with training-based approaches. Future work includes exploring more efficient diffusion sampling strategies and extending GenSAM to 3D volumetric segmentation.

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] S. Zhang and D. Metaxas, "On the challenges and perspectives of foundation models for medical image analysis," *Medical Image Analysis*, vol. 91, p. 102996, 2024.

[3] C. You, W. Dai, F. Liu, Y. Min, N. C. Dvornek, X. Li, D. A. Clifton, L. Staib, and J. S. Duncan, "Mine your own anatomy: Revisiting medical image segmentation with extremely limited labels," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 46, no. 12, pp. 11 136–11 151, 2024.

[4] 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.

[5] Z. Cheng, S. Wang, T. Xin, T. Zhou, H. Zhang, and L. Shao, "Few-shot medical image segmentation via generating multiple representative descriptors," *IEEE Transactions on Medical Imaging*, vol. 43, no. 6, pp. 2202–2214, 2024.

[6] 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.

[7] 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.

[8] 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.

[9] Y. Zhang, H. Li, Y. Gao, H. Duan, Y. Huang, and Y. Zheng, "Prototype correlation matching and class- relation reasoning for few-shot medical image segmentation," *IEEE Transactions on Medical Imaging*, vol. 43, no. 11, pp. 4041–4054, 2024.

[10] S. Tang, S. Yan, X. Qi, J. Gao, M. Ye, J. Zhang, and X. Zhu, "Few-shot medical image segmentation with high-fidelity prototypes," *Medical Image Analysis*, vol. 100, p. 103412, 2025.

[11] 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.

[12] 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 *European Conference on Computer Vision*, 2020, pp. 762–780.

[13] 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.

[14] 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.

[15] 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.

[16] 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.

[17] Y. Yang, X. Fang, X. Li, Y. Han, and Z. Yu, "Cdsg-sam: A cross-domain self-generating prompt few-shot brain tumor segmentation pipeline based on sam," *Biomedical Signal Processing and Control*, vol. 100, p. 106936, 2025.

[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] Y. Liu, H. Xiao, J. Chai, Y. Zhang, R. Wang, Z. Meng, and Z. Luo, "Synpo: Boosting training-free few-shot medical segmentation via high-quality negative prompts," 2025. [Online]. Available: https://arxiv.org/abs/2506.15153

[20] J. Wu, R. Fu, H. Fang, Y. Zhang, Y. Yang, H. Xiong, H. Liu, and Y. Xu, "Medsegdiff: Medical image segmentation with diffusion probabilistic model," 2023. [Online]. Available: https://arxiv.org/abs/2211.00611

[21] Y. Sun, Z. Chen, H. Zheng, Y. Lu, L. Duan, F. Fan, A. Elazab, X. Wan, C. Wang, and R. Ge, "Gl-lcm: Global-local latent consistency models for fast high-resolution bone suppression in chest x-ray images," 2025. [Online]. Available: https://arxiv.org/abs/2508.03357

[22] J. M. J. Valanarasu, P. Guo, V. VS, and V. M. Patel, "On-the-fly test-time adaptation for medical image segmentation," 2022. [Online]. Available: https://arxiv.org/abs/2203.05574

[23] 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

[24] 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

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

[26] 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.

[27] 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.

[28] 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.