SAM-MPA: Applying SAM to Few-shot Medical Image Segmentation using Mask Propagation and Auto-prompting

Jie Xu **Xiaokang Li** **Chengyu Yue** **Chen Ma** **Yuanyuan Wang*** **Yi Guo***
School of Information Science and Engineering, Fudan University, Shanghai, China
xujie23@m.fudan.edu.cn, lixiaokang@fudan.edu.cn, cyyue22@m.fudan.edu.cn
cma24@m.fudan.edu.cn, yywang@fudan.edu.cn, guoyi@fudan.edu.cn

Abstract

Medical image segmentation often faces the challenge of prohibitively expensive annotation costs. While few-shot learning offers a promising solution to alleviate this burden, conventional approaches still rely heavily on pre-training with large volumes of labeled data from known categories. To address this issue, we propose leveraging the Segment Anything Model (SAM), pre-trained on over 1 billion masks, thus circumventing the need for extensive domain-specific annotated data. In light of this, we developed SAM-MPA, an innovative **SAM**-based framework for few-shot medical image segmentation using **M**ask **P**ropagation-based **A**uto-prompting. Initially, we employ k-centroid clustering to select the most representative examples for labelling to construct the support set. These annotated examples are registered to other images yielding deformation fields that facilitate the propagation of the mask knowledge to obtain coarse masks across the dataset. Subsequently, we automatically generate visual prompts based on the region and boundary expansion of the coarse mask, including points, box and a coarse mask. Finally, we can obtain the segmentation predictions by inputting these prompts into SAM and refine the results by post refinement module. We validate the performance of the proposed framework through extensive experiments conducted on two medical image datasets with different modalities. Our method achieves Dices of 74.53%, 94.36% on Breast US, Chest X-ray, respectively. Experimental results substantiate that SAM-MPA yields high-accuracy segmentations within 10 labeled examples, outperforming other state-of-the-art few-shot auto-segmentation methods. Our method enables the customization of SAM for any medical image dataset with a small number of labeled examples.

1 Introduction

Accurate image segmentation plays a pivotal role in medical image analysis and assisted diagnosis [21]. It involves the segmentation of tissues, organs, and lesions in medical images across various imaging modalities. Typically, these segmentation labels are painstakingly acquired through manual outlining by experienced doctors, utilizing specialized annotation tools like ITK-SNAP[25]. Getting a large amount of labeled data is difficult due to the time-consuming and expensive process, as well as medical ethics concerns. Consequently, many deep learning models encounter obstacles in effectively training when confronted with limited data scenarios.

Addressing the challenge of limited labeled data availability, researchers have extensively explored few-shot segmentation (FSS) methods in the context of medical imaging[19]. FSS aims to delineate

*Corresponding author

38th Conference on Neural Information Processing Systems (NeurIPS 2024).

the target in a query image with the assistance of a small number of pixel-wise annotated support images. In general, FSS techniques rely on meta-learning approaches built upon prototype networks[18]. These methods typically involve a two-step process: creating representative prototypes from known classes, then using these to segment new, unseen images. It usually requires substantial labeled training data from known object categories in the first step. However, in medical imaging, such extensive annotations are costly and time-consuming. Therefore, we need to develop new few-shot segmentation methods that do not require large amounts of labeled medical data.

In response to the issue, the Segment Anything Model (SAM[11]) has recently emerged as a promising approach. Trained on over 1 billion masks, SAM demonstrates a powerful generalized few shot segmentation capability, able to segment any subject using visual prompts, even those hasn't seen before. Several studies have already applied SAM to medical imaging, yielding promising segmentation results[2, 12, 13, 24, 26]. They primarily focus on fine-tuning SAM and manually providing prompts. However, there are two significant challenges. (1) Fine-tuning methods typically require anywhere from hundreds[12, 24, 26] to millions[2, 13] of masked training images. In contrast, FSS aims to achieve satisfactory segmentation performance on medical images using only a handful of masked samples (typically 1, 5, or 10 instances). (2) The practice of manually supplying prompts is labor-intensive and highly inefficient. Given these considerations, we propose to explore the application of SAM for FSS using automated prompts in medical image analysis.

Based on existing research, we have identified three key challenges when applying SAM to few-shot medical segmentation, as illustrated in Fig. 1: (1) Selection of appropriate labeled images as support examples. The choice will significantly impact the performance of the foundation visual model on downstream tasks[29]. (2) Effective propagation of mask knowledge from support images to query images. This is the core challenge in few-shot learning scenarios[23], as it demands the model to extract mask-relevant features from a minimal support set and accurately transefer prior knowledge to the diverse query set. (3) Generation of high-quality prompts. Well-crafted prompts are crucial for maximizing SAM's potential[9].

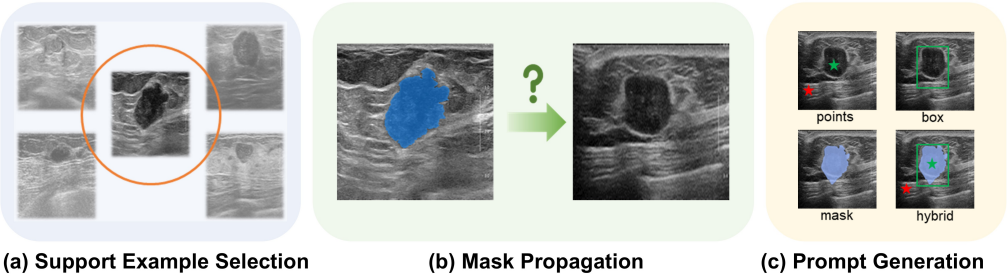

Figure 1: Three key challenges in adapting SAM to few shot medical segmentation

To tackle these problems, we introduce SAM-MPA, a novel framework that enhances the **SAM** for few-shot medical image segmentation by integrating **M**ask **P**ropogation and **A**utomatic prompt generation techniques. First, we cluster samples based on their intrinsic structure and select the most representative instance from each cluster to form the support image set. Second, we perform registration between support and query images, utilizing the resulting deformation field to transfer mask knowledge. Third, leveraging the deformed masks, we automatically generate prompts including foreground points, background points, and bounding boxes.

Our contributions can be summarized as follows:

1. We propose SAM-MPA, an innovative framework that successfully adapts SAM for few-shot medical image segmentation through automatic prompting. It integrates support sample selection, mask propagation, and automatic prompt generation, achieving high-performance segmentation results with only 1-10 labeled samples.
2. To address existing challenges, we introduce three innovative strategies: a clustering-based support sample selection, an unsupervised registration-driven mask propagation, and an automatic high-quality prompt generation. They respectively provide a representative support set, facilitate

efficient knowledge transfer between support and query images, and generate optimal prompts for SAM.

3. Our method achieves Dices of 82.07%, 94.13% on Breast US and Chest X-ray datasets, respectively. Experimental results demonstrate the effectiveness of our method, showcasing superior segmentation accuracy while requiring minimal labeled data across different medical imaging modality.

2 Related Work

2.1 Few-shot Medical Image Segmentation

Few-shot segmentation in natural image typically rely on extensive pre-training using labeled data from known semantic categories. This approach has influenced medical imaging, leading to the development of methods such as SE-Net[17], MRrNet[6], and GCN-DE[20], which utilize known category data to train efficient dual-branch interaction models. However, obtaining large-scale pre-labeled datasets in the medical domain is often impractical. To address this challenge, some researchers have turned to self-supervised learning techniques[3, 8, 14, 22]. These approaches generate pseudo-labels through superpixel[5], enabling self-supervised pre-training. While this method reduces annotation costs, it still faces several limitations: (1) the necessity of a pre-training phase. (2) the potential inaccuracies in superpixel generation. (3) the requirement for large-scale datasets to achieve effective self-supervised pre-training, which is often unfeasible in medical imaging. In contrast, with its publicly available pre-trained weights and extensive training on a vast dataset, SAM eliminates the need for additional domain-specific pre-training or large-scale labeled datasets. Therefore, we can consider SAM as an effective soultion for few-shot medical image segmentation.

2.2 A revisit of SAM

SAM is a universal image segmentation model trained on over 11 million images and 1 billion masks. It consists of three main components: (1) an image encoder, built on a refined Vision Transformer[4], which extracts robust image features. (2) a prompt encoder that converts diverse user prompts—ranging from sparse (points, boxes) to dense (masks)—into a unified feature space. (3) a light mask decoder that synthesizes inputs from two encoders to generate precise segmentations. SAM's primary distinction from traditional models lies in its ability to flexibly accept various forms of prompts. These prompts offer indicative information about the target area, enabling SAM to produce accurate results for new objects, thus exhibiting powerful generalization capabilities. Consequently, when adapting SAM to new domains, providing high-quality prompts becomes one of the most crucial tasks.

2.3 Adapting SAM to Medical Image Segmentation

The remarkable performance of SAM on natural images has led researchers to explore its capabilities in medical image segmentation. Adaptation methods for SAM in this domain can be categorized into fine-tuning and auto-prompting strategies[28]. Fine-tuning approaches include full fine-tuning (e.g., MedSAM[13]) and parameter-efficient techniques (such as Med-SA[24], SAM-Med2D[2], SAMed[26], and AdaptiveSAM[16]). Full fine-tuning updates all model parameters using large annotated datasets, while parameter-efficient methods adjust only some parameters to conserve resources. These fine-tuning approaches have drawbacks: they require substantial computational resources, incur labeling costs(MedSAM used over 1.5 million image-mask pairs), and rely on manually provided prompts in training data. Even parameter-efficient methods need hundreds to thousands of annotated images. Auto-prompting adaptation strategies aim to reduce reliance on manual prompts. For example, Pandey et al. used YOLOv8 to generate boundary boxes as prompts[15], but this still requires 100 masked images for training. PerSAM[27] achieves one-shot segmentation for specific categories in natural images using a single annotated image, employing feature matching to generate point prompts. Such approach may be less effective in medical imaging where target structures and surrounding tissues often have minimal visual differences.

3 Method

3.1 Problem Definition

Let $D = \{d_1, d_2, \cdots, d_N\}$ is an unlabeled dataset which contains N samples. In the few-shot scenario, we define the support set as $D_S = \{(s_1, m_1), (s_2, m_2), \cdots, (c_K, m_K)\}$, where m_i represents the mask corresponding to each image. The query set is denoted as $D_Q = \{q_1, q_2, \cdots, q_{N-K}\}$. Typical few-shot settings use K values of 1, 5, or 10. Our goal is to leverage the mask information in D_S to achieve high-precision segmentation with SAM on D_Q.

As shown in Fig.2, our proposed SAM-MPA framework comprises four primary steps: (1) Sample Selection: Seletcs K samples($K = 1, 5, 10$) from D to form the support set D_S, with the remaining $N - K$ samples consituting the query set D_Q. (2) Mask Propagation: Transefers mask information from D_S to D_Q to generate coarse masks for all unlabeled images in D_Q. (3) Prompt Auto-generating: Automatically generates diverse prompts from the coarse masks. (4) Segmentation and Refining: Utilizes SAM to segment D_Q based on the coarse masks and generated prompts, followed by a Post-Refinement process to further optimize the segmentation results. In the following sections, we will provide a detailed explanation of each step.

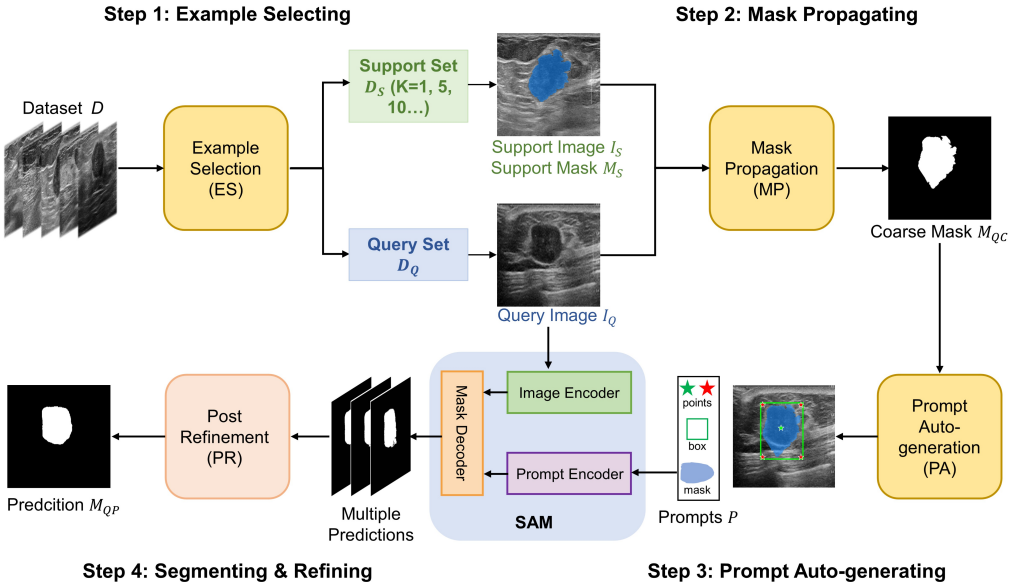

Figure 2: Overview of proposed framework SAM-MPA.

3.2 Support Example Selection

Samples with high similarity typically share more common feature points or structures, enabling better feature matching and precise transformation estimation. Thus our goal is to maximize the similarity between D_S and D_Q. As computing similarity directly from raw images is expensive, we adopt a two-step approach: first extracting image features, then calculating similarity. Leveraging SAM's training on large-scale image datasets, its image encoder $Encoder_I$ possesses powerful representation capabilities. We utilize SAM for this process without the need to train a new model from scratch:

$$z_i = Encoder_I(d_i) \tag{1}$$

Using Eq.1, we project the original dataset D into feature space $Z = \{z_i\}_{i=1}^{N}$. For one-shot tasks ($K = 1$), we select the sample most similar to others. Similarity is computed as:

$$Similarity(d_i) = \left(\frac{1}{N-1} \sum_{j=1, j \neq i}^{N} D(z_i, z_j) \right)^{-1} \tag{2}$$

where $D(z_i, z_j)$ is the cosine distance between features: $D(z_i, z_j) = \frac{||z_i|| \cdot ||z_j||}{||z_i|| \times ||z_j||}$. The final samle is determined as:

$$s_1 = \text{argmax } Similarity(d_i) \tag{3}$$

In few-shot tasks where $K > 1$, simply selecting the K most similar samples may result in a support set D_S that is overly concentrated in the feature space, failing to effectively cover the entire sample distribution. To address this, we employ the K-Centroid algorithm to partition the samples into K clusters. The center of each cluster is designated as a support sample, collectively forming D_S, while the remaining samples within each cluster serve as query samples for their respective support sample. To enhance data coverage and mitigate the randomness inherent in K-Centroid's sensitivity to initial points, we utilize the K-Center-Greedy algorithm to determine the starting points. We initialize the K-Center-Greedy's process with s_1, the sample of highest similarity, rather than a random selection. This approach ensures consistent and representative results by eliminating variability due to random initialization.

3.3 Mask Propagation

After successfully constructing D_S and D_Q, we employ a registration-based mask propagation module to generate coarse masks for samples in D_Q based on the labeled data in D_S. Given the limited number of samples, particularly those with masks, we utilize an unsupervised B-spline elastic registration(BER) algorithm[7]. Specifically, for a support image s_i and its corresponding query image q_j, we compute a deformation field Φ using BER. We then apply Φ to the mask m_i through a spatial transformer(ST) network, resulting in a coarse segmentation mc_j for q_j, as shown in Eq.5:

$$\Phi = BER(s_i, q_j) \tag{4}$$

$$mc_j = ST(m_i, \Phi) = m_i \circ \Phi \tag{5}$$

ST is a non-learnable neural network module designed for spatial transformations. It utilizes Φ to deform m_i into mc_j. Specifically, ST employs a grid generation and sampling process to remap and interpolate each pixel in m_i according to Φ, thereby producing mc_j. This process can be described by Eq. 6 and 7.

$$p' = p + u(p) \tag{6}$$

$$mc_j(p) = m_i \circ \Phi(p) = \sum_{q \in Z(p')} m_i(q) \prod_{d \in \{x,y\}} (1 - |p'_d - q_d|) \tag{7}$$

where $u(p)$ is the displacement of pixel p in the deformation field Φ. $Z(p')$ is the neighboring pixel of pixel p'. d traverses the space in all directions, where x and y represent the horizontal and vertical dimensions of the image, respectively.

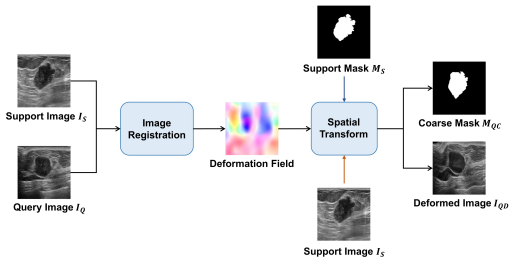

Figure 3: The workflow of the Mask Propagation module.

3.4 Prompts Auto-generation and Post Refinement Strategy

Based on the coarse mask of the query image, we automatically generate various prompts required by SAM. As mentioned in Section 2.2, these prompts include point, box, and mask formats.The generation methods are as follows:

(1) **Point prompts**: Foreground points are located at the centroid of the coarse segmentation, and background points are the four corners of the minimum bounding box surrounding the coarse segmentation.

(2) **Box prompts**: The minimum bounding box of the coarse segmentation.

(3) **Mask prompts**: Apply sigmoid softening to the coarse segmentation.

We then feed the original image along with the generated prompts into SAM to produce an initial segmentation. To further optimize the output, we employ a Post-Refinement strategy that substitutes the preliminary segmentation for the old mask prompt and reprocess this updated input through SAM to yield a more precise delineation of the target object.

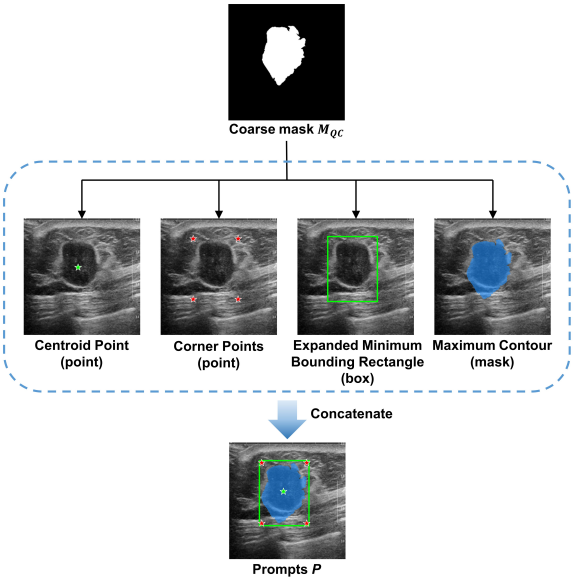

Figure 4: The workflow of the Prompt Auto-generation module.

4 Experiments

4.1 Datasets

We evaluate our method on two medical image datasets with different modalities. Each image in the datasets has accurate pixel-wise segmentation labels that has been precisely annotated by an experienced radiologist as the ground truth. All images were resized to 256 × 256 pixels before segmentation.

- **Breast US:** Breast US is a private breast ultrasound dataset collected from Fudan University Shanghai Cancer Center, Shanghai, China. All images are obtained using the machine of the Aixplorer ultrasound system (SuperSonic Imagine S.A., Aix-en-Provence, France) at 7–15 MHz. It contains 200 images.
- **Chest X-Ray:** Chest X-ray is a standard digital image database for Tuberculosis is created by the National Library of Medicine, Maryland, USA in collaboration with Shenzhen No.3 People's Hospital, Guangdong Medical College, Shenzhen, China[10, 1]. It contains 704 images with labels. The Chest X-rays are from out-patient clinics, and were captured as part of the daily routine using Philips DR Digital Diagnose systems.

4.2 Implementation Details and Comparison Methods

We evaluated the few-shot scenarios for 1, 5 and 10-shot settings. For sample selection and specific segmentation tasks, we employed the vit_b model as SAM's image encoder. During the registration process, we utilized the B-Spline registration function from elastic-5.1.0, setting NumberOfResolutions to 3 while keeping other parameters at their default values.

The Dice coefficient served as our evaluation metric. All methods were implemented on the PyTorch platform, with computations accelerated using two Nvidia V100 32GB GPUs.

We compare the performances the manual prompting methods and automatic prompting methods based on SAM. The manual-prompting methods include using SAM with 1 point, 1 bounding box and the hybrid prompts (1 point and 1 bounding box). The auto-prompting few-shot method is PerSAM.

5 Results

5.1 Ablation Study

In this section, we conduct an ablation study to evaluate the proposed SAM-MPA method in a 10-shot setting, with results presented in Table 1. This systematic experiment assessed the contribution of each key component to the overall performance, specifically examining the Example Selection(ES) module, Mask Propagation(MP) module, Prompt Auto-generation(PA) module, and Post Refinement(PR) strategy. The results show that enabling each of the components leads to different levels of performance improvement, which intuitively reveals the functional importance of each component in the complete process.

In our ablation study, we carefully considered the interdependencies among components, establishing a sequential activation order. Mask Propagation (MP), the cornerstone of the SAM-MPA, was activated first. This component enables effective mask information transfer between support and query sets, without which the entire framework would be inoperable. Activating MP alone achieved a high-precision coarse segmentation with a Dice of 70.67%, validating its efficacy in transmitting prior information. Next, we added Prompt Auto-generation (PA), which leverages the coarse masks produced by MP. The PA module, which automatically generates prompts for SAM based on the coarse masks, yielded a 0.83% increase in Dice, validating its effectiveness. While MP and PA combined already enable SAM-based automatic segmentation, there was clearly room for enhancement. We then introduced the Example Selection (ES) component to assess how support set sample selection impacts segmentation results. The Dice increased by 2.34%, demonstrating that our curated support samples provide superior guidance for query image segmentation. Finally, we activated the Post Refinement module, which replaces initial coarse masks with SAM-generated ones to enhance prompt quality. This step achieves a performance gain of 0.51%, underscoring its role in boosting segmentation prediction performance.

Table 1: Ablation study of the SAM-MPA on Breast US in the setting of 10-shot. ES: Example Selection; MP: Mask Propagation; PA: Prompt Auto-generation; PR: Post Refinement.

ES	MP	PA	PR	Dice (%)	Gain (%)
	✓			70.67	
	✓	✓		71.50	**+0.83**
✓	✓	✓		73.84	**+2.34**
✓	✓	✓	✓	**74.53**	**+0.69**

5.2 Quantitative and Qualitative Results

Table 2 presents a performance comparison between SAM-MPA and other methods. The "Manual" approach refers to providing manual prompts to the original SAM model for direct target image segmentation, without involving few-shot settings such as support and query images. Results show that manually providing both point and bbox prompts yields the best performance, achieving dice scores of 70.32% and 86.73% on the two datasets. However, this method requires tedious manual confirmation of prompt points and bounding box drawing for each test image, such as processing 704 images in the ChestXray dataset. In contrast, our method achieves excellent segmentation performance by providing precise masks for 1, 5, or 10 support images. Using 10 masked support images, SAM-MPA achieves a dice score of 74.53% on ultrasound images, surpassing the manual prompting method by 4.21% and PerSAM by 18.93%. For X-ray images, using just one support image yields a performance of 93.48%, outperforming manual prompting method by 6.75% and PerSAM by 46.55%. These results demonstrate that our method not only avoids the cumbersome manual prompting process but also achieves satisfactory segmentation accuracy.

Table 2: Quantitative comparison of Dice (%) on Breast US and Chest X-ray in different K-shot settings. pt: point; bbox: bounding box.

Method		Breast US			Chest X-ray		
		1-shot	5-shot	10-shot	1-shot	5-shot	10-shot
Manual	SAM (1 point)		31.10			50.42	
	SAM (1 bbox)		67.43			80.79	
	SAM (1 point + 1 bbox)		70.32			86.73	
Automatic	PerSAM	21.70	49.12	55.60	46.93	60.14	76.60
	SAM-MPA(**Ours**)	**66.92**	**68.65**	**74.53**	**93.48**	**94.21**	**94.36**

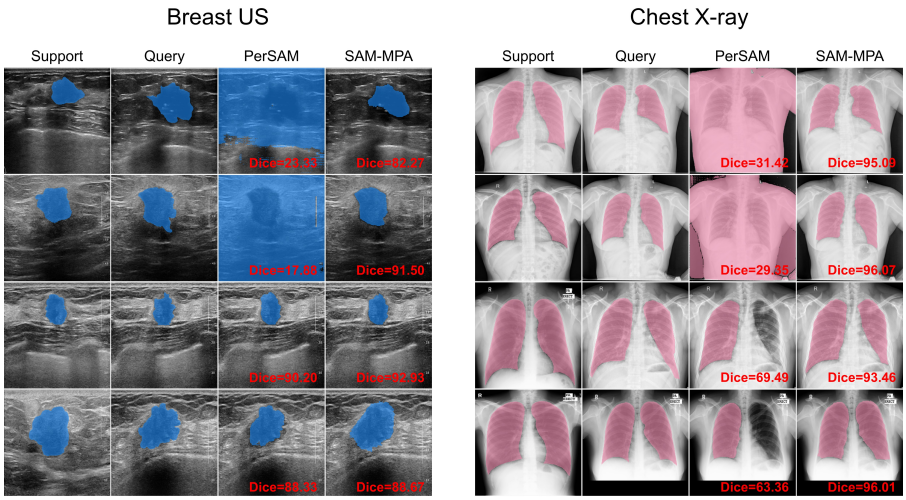

Figure 5: Visual comparison of segmentation on Breast US and Chest X-ray in 10-shot setting.

The experimental results illustrated in Fig.5 demonstrate the segmentation performance of SAM-MPA and PerSAM across various query images. SAM-MPA consistently delivers satisfactory and stable segmentation results, exhibiting minimal performance variance across different images. In contrast, PerSAM's performance is characterized by significant fluctuations. While PerSAM achieves commendable segmentation results on certain images, it struggles considerably with others, sometimes failing to identify the correct target regions entirely. This inconsistency primarily stems from the subtle distinctions between foreground and background in medical images. The prototype similarity strategy employed by PerSAM proves ineffective in such scenarios, leading to instances where large portions of the image are erroneously classified as target regions. SAM-MPA, on the other hand, showcases superior performance. It adeptly identifies and accurately segments target regions, yielding satisfactory results even on images that pose challenges for PerSAM. This stark contrast underscores SAM-MPA's robustness and adaptability in handling complex medical images, highlighting its effectiveness in overcoming the limitations faced by alternative approaches.

6 Discussion

Few-shot methods in medical image segmentation often struggle with the need for large amounts of annotated data for pre-training. To address this issue, we propose the SAM-MPA framework, which leverages SAM for few-shot segmentation of medical images, effectively circumventing the demand for extensive labeled data. By automatically generating high-quality prompts for SAM, we successfully unlock its potential, achieving impressive segmentation results with just a handful of annotated samples (1, 5, or 10 masked images).

First, we will analyze the efficacy of each component in SAM-MPA. Ablation studies in Section 5.1 demonstrate that every element plays an indispensable role. The sample selection module constructs a

high-quality support set by selecting central samples from each cluster in the sample space and yields the most significant improvement(2.34% performance gain). This underscores the importance of representative support samples in transferring label prior knowledge to query instances and highlights the sensitivity of foundation models like SAM to annotated example selection in downstream tasks. Then, the mask propagation module forms the backbone of the framework, facilitating efficient flow of annotation information between support and query samples. As evident from Fig.2, SAM-MPA cannot function properly without this component. Employing unsupervised registration techniques, we achieve label propagation between support and query samples, attaining a substantial baseline segmentation performance of 70.67%. Next, the automatic prompt generation component is equally crucial. Recent foundation models like SAM and GPT have shown that efficient prompts are key to unlocking a model's potential. Experimental results in Table2 reveal the significant impact of different prompts on SAM's performance. Using a single point prompt severely underperforms compared to box prompts, with performance gaps of 36.33% and 30.27% on US and Xray datasets, respectively. Our component automatically generates high-quality point, box, and mask prompts in one go, effectively enhancing segmentation performance. Lastly, the Post Refinement strategy leverages SAM's output mask as input for another round, optimizing segmentation results through SAM's inherent capabilities, proving to be an effective approach.

Second, we explore why our method outperforms others. While manually providing point and box prompts for images can yield decent results, this process becomes tedious when applied to a large quantity of images, requiring an enormous amount of manual labor. In contrast, SAM-MPA only requires manual masking of 1-10 images, with the rest of the process being fully automated. This significantly reduces human effort while achieving high-performance segmentation results, presenting a clear advantage. Compared to the current automated few-shot method PerSAM, our approach demonstrates substantial performance improvements. As shown in Table2, with just one annotated image, our method outperforms PerSAM by 45.22% on ultrasound images and 46.55% on X-ray images. This is primarily due to: (1) The low contrast between foreground and background in medical images, which hinders PerSAM's region similarity-based strategy. (2) PerSAM's limited prompts (one foreground point and one background point) fail to fully harness SAM's potential. In contrast, SAM-MPA provides richer prompts, including one foreground point, four background points, a bounding box, and a coarse segmentation mask, offering SAM more comprehensive prior knowledge and effectively tapping into its capabilities.

Third, we briefly discuss the limitations of our method and future prospects. While SAM exhibits excellent generalization, it was pre-trained on natural images, resulting in a domain gap with medical images. Future work could explore foundation models specifically pre-trained on medical imaging data. Additionally, this study focused on 2D image applications, leaving 3D scenarios unexplored. Future research should investigate the feasibility and generalizability of SAM in segmenting 3D images.

7 Conclusion

Our SAM-MPA framework addresses the key challenges of adapting SAM to few-shot medical image segmentation including selecting representative examples, accurately propagating prior knowledge, and generating suitable prompts through dedicated modules. By leveraging mask propagation-based auto-prompting, we enhance SAM's generalization capabilities for specific medical datasets. SAM-MPA demonstrates superior performance across diverse datasets in few-shot scenarios, offering a practical solution for high- accuracy segmentation with minimal user annotation. This work has the potential to significantly advance medical image analysis and diagnosis, providing a valuable tool for real-world healthcare applications.

Acknowledgments

This work was supported in part by the National Natural Science Foundation of China under Grant 62371139 and Grant 82227813; and in part by the Science and Technology Commission of Shanghai Municipality under Grant 22ZR1404800.

References

[1] Sema Candemir, Stefan Jaeger, Kannappan Palaniappan, Jonathan P Musco, Rahul K Singh, Zhiyun Xue, Alexandros Karargyris, Sameer Antani, George Thoma, and Clement J McDonald. Lung segmentation in chest radiographs using anatomical atlases with nonrigid registration. *IEEE transactions on medical imaging*, 33(2):577–590, 2013.

[2] Junlong Cheng, Jin Ye, Zhongying Deng, Jianpin Chen, Tianbin Li, Haoyu Wang, Yanzhou Su, Ziyan Huang, Jilong Chen, Lei Jiang, et al. Sam-med2d. *arXiv preprint arXiv:2308.16184*, 2023.

[3] Ziming Cheng, Shidong Wang, Tong Xin, Tao Zhou, Haofeng Zhang, and Ling Shao. Few-shot medical image segmentation via generating multiple representative descriptors. *IEEE Transactions on Medical Imaging*, 2024.

[4] Alexey Dosovitskiy, Lucas Beyer, Alexander Kolesnikov, Dirk Weissenborn, Xiaohua Zhai, Thomas Unterthiner, Mostafa Dehghani, Matthias Minderer, Georg Heigold, Sylvain Gelly, et al. An image is worth 16x16 words: Transformers for image recognition at scale. *arXiv preprint arXiv:2010.11929*, 2020.

[5] Pedro F Felzenszwalb and Daniel P Huttenlocher. Efficient graph-based image segmentation. *International journal of computer vision*, 59:167–181, 2004.

[6] Ruiwei Feng, Xiangshang Zheng, Tianxiang Gao, Jintai Chen, Wenzhe Wang, Danny Z Chen, and Jian Wu. Interactive few-shot learning: Limited supervision, better medical image segmentation. *IEEE Transactions on Medical Imaging*, 40(10):2575–2588, 2021.

[7] Suicheng Gu, Xin Meng, Frank C Sciurba, Hongxia Ma, Joseph Leader, Naftali Kaminski, David Gur, and Jiantao Pu. Bidirectional elastic image registration using b-spline affine transformation. *Computerized Medical Imaging and Graphics*, 38(4):306–314, 2014.

[8] Stine Hansen, Srishti Gautam, Robert Jenssen, and Michael Kampffmeyer. Anomaly detection-inspired few-shot medical image segmentation through self-supervision with supervoxels. *Medical Image Analysis*, 78:102385, 2022.

[9] Yuhao Huang, Xin Yang, Lian Liu, Han Zhou, Ao Chang, Xinrui Zhou, Rusi Chen, Junxuan Yu, Jiongquan Chen, Chaoyu Chen, et al. Segment anything model for medical images? *Medical Image Analysis*, 92:103061, 2024.

[10] Stefan Jaeger, Alexandros Karargyris, Sema Candemir, Les Folio, Jenifer Siegelman, Fiona Callaghan, Zhiyun Xue, Kannappan Palaniappan, Rahul K Singh, Sameer Antani, et al. Automatic tuberculosis screening using chest radiographs. *IEEE transactions on medical imaging*, 33(2):233–245, 2013.

[11] Alexander Kirillov, Eric Mintun, Nikhila Ravi, Hanzi Mao, Chloe Rolland, Laura Gustafson, Tete Xiao, Spencer Whitehead, Alexander C Berg, Wan-Yen Lo, et al. Segment anything. In *Proceedings of the IEEE/CVF International Conference on Computer Vision*, pages 4015–4026, 2023.

[12] Xian Lin, Yangyang Xiang, Li Zhang, Xin Yang, Zengqiang Yan, and Li Yu. Samus: Adapting segment anything model for clinically-friendly and generalizable ultrasound image segmentation. *arXiv preprint arXiv:2309.06824*, 2023.

[13] Jun Ma, Yuting He, Feifei Li, Lin Han, Chenyu You, and Bo Wang. Segment anything in medical images. *Nature Communications*, 15(1):654, 2024.

[14] Cheng Ouyang, Carlo Biffi, Chen Chen, Turkay Kart, Huaqi Qiu, and Daniel Rueckert. Self-supervision with superpixels: Training few-shot medical image segmentation without annotation. In *Computer Vision–ECCV 2020: 16th European Conference, Glasgow, UK, August 23–28, 2020, Proceedings, Part XXIX 16*, pages 762–780. Springer, 2020.

[15] Sumit Pandey, Kuan-Fu Chen, and Erik B Dam. Comprehensive multimodal segmentation in medical imaging: Combining yolov8 with sam and hq-sam models. In *Proceedings of the IEEE/CVF international conference on computer vision*, pages 2592–2598, 2023.

[16] Jay N Paranjape, Nithin Gopalakrishnan Nair, Shameema Sikder, S Swaroop Vedula, and Vishal M Patel. Adaptivesam: Towards efficient tuning of sam for surgical scene segmentation. In *Annual Conference on Medical Image Understanding and Analysis*, pages 187–201. Springer, 2024.

[17] Abhijit Guha Roy, Shayan Siddiqui, Sebastian Pölsterl, Nassir Navab, and Christian Wachinger. 'squeeze & excite'guided few-shot segmentation of volumetric images. *Medical image analysis*, 59:101587, 2020.

[18] Jake Snell, Kevin Swersky, and Richard Zemel. Prototypical networks for few-shot learning. *Advances in neural information processing systems*, 30, 2017.

[19] Yisheng Song, Ting Wang, Puyu Cai, Subrota K Mondal, and Jyoti Prakash Sahoo. A comprehensive survey of few-shot learning: Evolution, applications, challenges, and opportunities. *ACM Computing Surveys*, 55(13s):1–40, 2023.

[20] Liyan Sun, Chenxin Li, Xinghao Ding, Yue Huang, Zhong Chen, Guisheng Wang, Yizhou Yu, and John Paisley. Few-shot medical image segmentation using a global correlation network with discriminative embedding. *Computers in biology and medicine*, 140:105067, 2022.

[21] Risheng Wang, Tao Lei, Ruixia Cui, Bingtao Zhang, Hongying Meng, and Asoke K Nandi. Medical image segmentation using deep learning: A survey. *IET image processing*, 16(5):1243–1267, 2022.

[22] Runze Wang, Qin Zhou, and Guoyan Zheng. Few-shot medical image segmentation regularized with self-reference and contrastive learning. In *International Conference on Medical Image Computing and Computer-Assisted Intervention*, pages 514–523. Springer, 2022.

[23] Yaqing Wang, Quanming Yao, James T Kwok, and Lionel M Ni. Generalizing from a few examples: A survey on few-shot learning. *ACM computing surveys (csur)*, 53(3):1–34, 2020.

[24] Junde Wu, Wei Ji, Yuanpei Liu, Huazhu Fu, Min Xu, Yanwu Xu, and Yueming Jin. Medical sam adapter: Adapting segment anything model for medical image segmentation. *arXiv preprint arXiv:2304.12620*, 2023.

[25] Paul A Yushkevich, Yang Gao, and Guido Gerig. Itk-snap: An interactive tool for semi-automatic segmentation of multi-modality biomedical images. In *2016 38th annual international conference of the IEEE engineering in medicine and biology society (EMBC)*, pages 3342–3345. IEEE, 2016.

[26] Kaidong Zhang and Dong Liu. Customized segment anything model for medical image segmentation. *arXiv preprint arXiv:2304.13785*, 2023.

[27] Renrui Zhang, Zhengkai Jiang, Ziyu Guo, Shilin Yan, Junting Pan, Xianzheng Ma, Hao Dong, Peng Gao, and Hongsheng Li. Personalize segment anything model with one shot. *arXiv preprint arXiv:2305.03048*, 2023.

[28] Yichi Zhang, Zhenrong Shen, and Rushi Jiao. Segment anything model for medical image segmentation: Current applications and future directions. *Computers in Biology and Medicine*, page 108238, 2024.

[29] Yuanhan Zhang, Kaiyang Zhou, and Ziwei Liu. What makes good examples for visual in-context learning? *Advances in Neural Information Processing Systems*, 36:17773–17794, 2023.