ELASTIC FLOW-GUIDED FEW-SHOT MEDICAL IMAGE SEGMENTATION WITH SAM-REFINED

Anonymous authors
Paper under double-blind review

ABSTRACT

Few-shot medical image segmentation focuses on generalizing to unseen categories with limited annotations. However, medical images are characterized by widespread non-rigid deformations and significant intra-class differences, leading to a dilemma of support-query foreground ideal matching. This constitutes a critical factor limiting performance of many existing approaches. To overcome this issue, we propose an **E**lastic **F**low-guided few-shot medical image segmentation with SAM-Refined network (EFG-SAM). First, to alleviate deformation gap between support and query foreground features, we design an Elastic Flow-field Deformation Module (EFDM). Building upon deformed support features, we propose a Deformation-aware Generation Module (DGM) to obtain elastic prototypes containing semantic, details, and background information. Subsequently, leveraging semantic elastic prototypes to suppress background noise in query features, we design an Elastic Localization Module (ELM). Then, we propose an Elastic Aggregation Module (EAM) to achieve efficient segmentation of query features under the guidance of these elastic prototypes. Finally, during the inference stage, SAM is seamlessly integrated into EFG as a training-free enhancement component to improve final segmentation results. Extensive experiments on three datasets show that our method outperforms state-of-the-art approaches by 2.45% on average and by 8.30% in cross-domain settings, demonstrating its effectiveness in bridging the deformation gap.

1 INTRODUCTION

Medical image segmentation (MIS) [13; 25; 21; 38] is essential for assisting clinical diagnosis and treatment. Although deep learning has significantly advanced this field, existing models typically require large amounts of annotated data. However, strict privacy constraints, limited data availability, high annotation costs, and poor generalization to unseen categories restrict their practical applications. Fortunately, few-shot learning [1; 39; 32; 7], with its ability to generalize to new categories with only a few labeled examples, provides a promising solution. Based on this characteristic, researchers have begun to introduce it into MIS to overcome challenges of data scarcity and expansion to new categories, leading to emerging field of few-shot medical image segmentation (FSMIS) [24; 30; 27].

Due to complexity of medical image scenarios, a **fundamental fact** exists in FSMIS: *as depicted by green dashed circle of Fig. 1**(a)**, support-query medical goals within same category often exhibit significant non-rigid deformations and intra-class differences*. Some studies named UPRENet [14], GMRD [8], and PAMI [47] ignore this fact and instead focus on designing strategies to extract more effective and representative prototypes from support feature. Other works such as, CoW [15], MSPONet [33], and DGPANet [27], aim to improve segmentation performance by developing modules that emphasizes boundary details and local prototype information. However, the aforementioned approaches are limited by this fact due to their model inherent mechanisms. To bridge distribution gap between support and query features, subsequent studies named QNet [26], CATNet [20], MTPNet [28], DIFD [9], and CART [10] have designed multiple dual-support prototypes, reverse correction mechanisms, and similar strategies. Nevertheless, they still fall into assumption that gap between support and query can be corrected simply by processing with enough modules. In recent years, powerful segmentation capabilities of general visual encoders (*e.g.*, SAM [17], DINO [6]) have brought new ideas to FSMIS. SAM has strong detail capture capability, while DINO has strong global semantic awareness capability. Based on this, methods such as MAUP [45], SynPo [22],

Reviewers: please read the Reviewer Guidelines (iclr.cc/Conferences/2027/ReviewerGuidelines) and the AI Policy for Reviewers (iclr.cc/Conferences/2027/AIPolicyForReviewers). If you used AI to expand, edit, or polish your review, please provide the input text to the LLM. Better still, consider skipping the LLM and submitting your original text: we, and the authors, are much more interested in your unedited thoughts than in what an LLM has to say. AI-assisted or not, you are putting your name and reputation behind your review: LLM-generated falsehoods, hallucinations or misrepresentations are subject to disciplinary action, which may include desk-rejecting all papers you have authored.

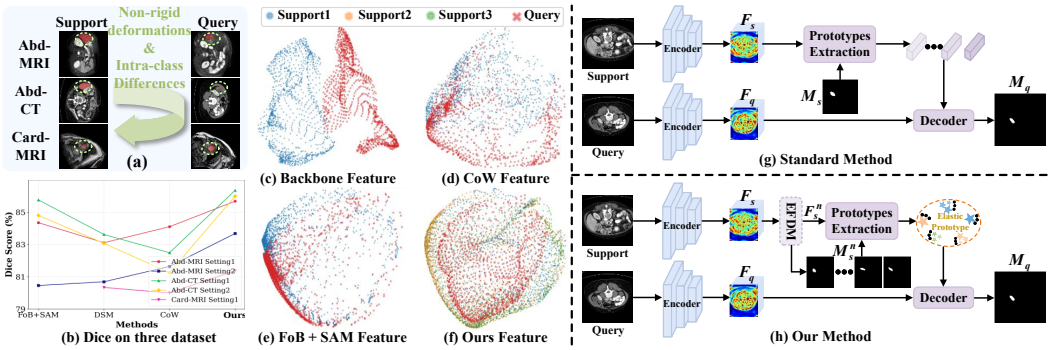


Figure 1: **Comparison of standard [15; 5] and our methods in handling intra-class differences in medical images**. **(a)** Visualizations show obvious support-query intra-class differences (within green dashed circles). **(b)** Dice comparison of different methods. **(c)-(f)** PCA plots indicate our method extracts informative support target feature for query. Support1 is original support features from backbone, CoW, and FoB+SAM, respectively; support2/3 are enriched support features generated by our EFDM. **(g)** Standard methods [15; 9; 4; 48] guide query segmentation using single, unenriched support feature. **(h)** Our method enriches support features to offer comprehensive query guidance.

SSM-SAM [19], FoB+SAM [5], and ProtoSAM [2] utilize these abilities to obtain support and query features with comprehensive semantics and fine boundaries, laying foundation for subsequent segmentation. Methods based on general visual encoders also suffer from this fact. To prove that above methods are affected by this fact, we compare Dice performance of different methods (Fig. 1**(b)**) and perform principal component analysis (PCA) visualization analysis on feature extracted by encoder of different models (Fig. 1**(c)-(f)**). These comparative results reveal a **serious limitation**: *using backbone features alone yields clearly inadequate foreground information from support to effectively guide query; while recent works such as CoW [15] and FoB+SAM [5] have alleviated this, they still fall short of providing sufficient guiding information compared to our method*. This is not only a major bottleneck limiting performance ceiling of current models but also a core challenge that urgently needs to be addressed in FSMIS. Based on above analysis, we can obtain a **conclusion**: *existing methods (Fig. 1**(g)**) typically use only support feature of single morphology to guide query feature, leading to a ideal matching trap between support and query foreground, which affects upper limit of performance*. Therefore, to address this issue, we propose Elastic Flow-guided FSMIS with SAM-Refined network (EFG-SAM), whose core idea is illustrated in Fig. 1**(h)**. Unlike Fig. 1**(g)**, it does not rely solely on support feature under single morphology to guide query feature, but instead addresses current dilemma by enriching support foreground feature.

Specifically, to mitigate inherent deformation gap between support foreground and query foreground features, we propose Elastic Flow-field Deformation Module (EFDM). This module utilizes block and dense field offset mechanisms to expand support feature and its corresponding mask, resulting in enriched support features. Next, relying on adjacent row, column, and global traversal strategies, we design Deformation-aware Generation Module (DGM) to extract elastic prototypes from enriched support feature. These prototypes not only cover global semantic foreground but also capture local foreground details, boundary shape, and background information. Elastic Prototype Generation (EPG) stage is composed of both EFDM and DGM. Furthermore, considering that query feature is susceptible to background noise interference, we propose Elastic Localization Module (ELM), which leverages semantic elastic prototype and query feature to calculate cosine similarity, thereby enhancing query foreground feature. Finally, Elastic Aggregation Module (EAM) matches elastic prototypes and enhanced query features for adaptive multi-scale mask decoding, while seamlessly integrating SAM into EFG during inference for training-free segmentation results enhancement.

Overall, our key contributions are detailed as follows: **1).** To the best of our knowledge, we are the first to systematically mitigate this issue from an elastic flow-field deformation perspective: foreground features of support and query are prone to falling into ideal matching trap. To this end, we develop EFG-SAM, which has been validated on three datasets as well as in unseen domain scenarios. **2).** To mitigate issue of insufficient support information caused by foreground deformation between support and query, we propose EPG stage. The EPG includes EFDM for expanding support feature and DGM for extracting elastic prototypes. **3).** We design ELM to calculate cosine similarity between

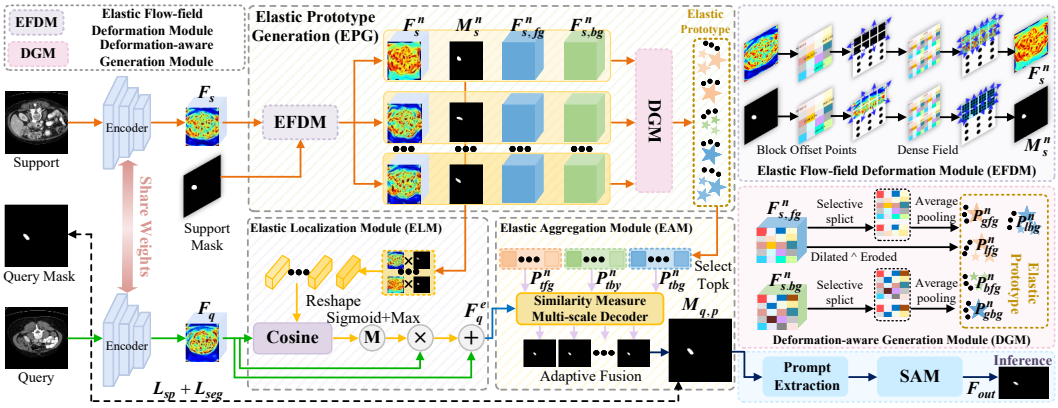


Figure 2: **Overview of the proposed EFG-SAM**. EPG stage generates various elastic prototypes. Then, ELM uses global semantic elastic prototypes to suppress interference in query features. Finally, EAM uses prototypes for query prediction, while training-free SAM integration enhances final results.

semantic elastic prototypes and query feature. This highlights foreground and reduces background interference. EAM is devised to obtain similarity matching results between various elastic prototypes and query feature, and obtain final result by multi-scale decoder and adaptive fusion strategies. During inference, SAM is integrated into EFG for training-free segmentation result enhancement.

2 Related Work

Few-Shot Semantic Segmentation. Driven by progress in deep learning, few-shot semantic segmentation (FSS) [42] has progressed rapidly. Existing methods can be broadly categorized into two groups based on network architecture: those using conventional encoders (*e.g.*, ResNet [12], Transformer [35]) and those leveraging general visual encoders. To improve adaptation to novel classes, conventional encoders-based methods, *e.g.* SFMNet [44], AdaptiveFSS [36], IFA [23], DR-Adapter [29], APM [34], PAT [3], HMNet [40], introduce various fine-tuners and modules. Meanwhile, general encoders-based methods, *e.g.* GFSAM [42], FoB+SAM [5], SSM-SAM [19], ProtoSAM [2], boost performance by exploiting knowledge of pre-trained models. Despite advancing FSS, they suffer from the ideal matching trap regarding support and query foregrounds.

Few-Shot Medical Image Segmentation. Benefiting from rapid development of FSS, researches on FSMIS have also emerged [27; 31; 47]. For example, CoW [15] employs support self-prediction module to refine its focus on boundary features. DIFD [9] proposes dual interspersion module to overcome intra-class variation and non-representative prototypes. To improve model deployment and universality, MAUP [45] designs multi-center adaptive uncertainty aware prompting strategy. DSM [48] utilizes dynamic semantic matching strategy to overcome cross-domain segmentation. FAMNet [4] proposes generalized frequency-aware matching network to recognize novel target domain. Above methods and other existing approaches [26; 8; 14; 33; 24; 30] rely on support feature of single morphology for guiding query segmentation, consequently falling into ideal matching trap.

3 Methodology

3.1 Problem Definition

FSMIS trains a model on training set D_{train} containing base class C_{b}, capable of segmenting new class C_{n} on test set D_{test}, using a few labeled data. Here, $C_{\text{b}} \cap C_{\text{n}} = \emptyset$. We follow the episode way commonly used in FSS[37] to train model. D_{train} and D_{test} are divided into numerous episodes containing support set S and query set Q. For each category of S, there are J supporting image-mask pairs, denoted as $S = \{(\boldsymbol{I}_{\text{s}\cdot i}, \boldsymbol{M}_{\text{s}\cdot i})\}_{i=1}^{J}$, and Q is denoted as $Q = \{(\boldsymbol{I}_{\text{q}}, \boldsymbol{M}_{\text{q}})\}$. In each episode, S and $\boldsymbol{I}_{\text{q}}$ are fed together into model to generate predicted query mask $\boldsymbol{M}_{\text{q,p}}$, while true query mask is denoted as $\boldsymbol{M}_{\text{q}}$. Since vast majority of previous FSMIS methods employ **1-way 1-shot** meta-learning strategy [24; 11], we follow these configurations. The architecture of EFG-SAM is shown in Fig. 2.

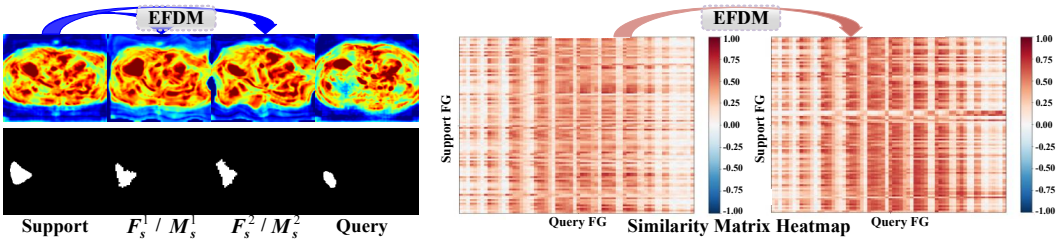


Figure 3: **After EFDM processing**, support backbone features and masks underwent elastic deformation, enhancing support-query foreground similarity matrix values.

3.2 Elastic Prototype Generation Stage

Elastic Flow-field Deformation Module. To address ideal matching trap between support and query foreground features, a reasonable strategy is to generate feature with multiple deformation and their corresponding mask for support feature, thereby reducing gap between support and query foreground features. Based on this, we design EFDM with no learnable parameters, which employs block offset and dense field offset mechanisms. Specifically, its inputs are $\boldsymbol{F}_{\mathrm{s}} \in \mathbb{R}^{B\times C\times H_{\mathrm{s}}\times W_{\mathrm{s}}}$ and $\boldsymbol{M}_{\mathrm{s}} \in \mathbb{R}^{B\times C\times H\times W}$, where $\boldsymbol{F}_{\mathrm{s}}$ is obtained by encoding support image $\boldsymbol{I}_{\mathrm{s}} \in \mathbb{R}^{B\times 3\times H\times W}$ through a shared weight backbone ResNet101. First, we stack $\boldsymbol{F}_{\mathrm{s}}$ and $\boldsymbol{M}_{\mathrm{s}}$ n times to obtain $\boldsymbol{F}_{ns}$ and $\boldsymbol{M}_{ns}$, respectively. The block-wise control field D^n is generated. The each control point d^n in D^n is obtained from uniform distribution $\mathcal{U}$(-α, α). This is used to control local region offset of $\boldsymbol{F}_{ns}$ and $\boldsymbol{M}_{ns}$. Then, for $\boldsymbol{F}_{ns}$ size, ordinate and x-coordinate of each pixel are obtained, and these coordinates are stacked to obtain complete pixel-level basic grid coordinates $G^n(u,v)$, as follows:

$$\begin{aligned} D^n \sim \mathcal{U}(-\alpha,\alpha), \quad D^n \in \mathbb{R}^{B\times N\times N\times 2}, \quad d^n = D^n(x,y) \in \mathbb{R}^{B\times 1\times 1\times 2}, \\ G^n(u,v) = (B, x_u, y_v, 2), \quad x_u, y_v \in [-1,1], \quad G(u,v) \in \mathbb{R}^{B\times H_s\times W_s\times 2}, \end{aligned} \tag{1}$$

where α and n are set 0.1 and 2, see subsection 4.3 for details. Next, to achieve pixel-level offset, we use bicubic interpolation function $\mathcal{B}(\cdot)$ to upsample D^n to resolution of $G^n(u,v)$, thereby obtaining dense offset field $D^n_{\mathcal{B}}(u,v)$. $D^n_{\mathcal{B}}$ is then added to $G^n(u,v)$ to generate deformed grid coordinates $G^n_{def}(u,v)$. Then, we employ differentiable grid sampling with bilinear and nearest-neighbor interpolation to sample $\boldsymbol{F}_{ns}$ and $\boldsymbol{M}_{ns}$, yielding deformed feature $\boldsymbol{F}^n_{\mathrm{s}}$ and mask $\boldsymbol{M}^n_{\mathrm{s}}$, as follows:

$$\begin{aligned} G^n_{def}(u,v) = D^n_{\mathcal{B}} + G^n(u,v), \quad D^n_{\mathcal{B}} \in \mathbb{R}^{B\times H_s\times W_s\times 2}, \quad G^n_{def}(u,v) \in \mathbb{R}^{B\times H_s\times W_s\times 2}, \\ \boldsymbol{F}^n_{\mathrm{s}} = \mathcal{D}_{bilinear}(\boldsymbol{F}_{ns}, G^n_{def}(u,v)), \quad \boldsymbol{M}^n_{\mathrm{s}} = \mathcal{D}_{nearest}(\boldsymbol{M}_{ns}, G^n_{def}(u,v)). \end{aligned} \tag{2}$$

where $\mathcal{D}_{bilinear}(\cdot)$ and $\mathcal{D}_{nearest}(\cdot)$ denote differentiable grid sampling with bilinear and nearest-neighbor interpolation functions. As shown in Fig. 3, EFDM obtains multiple deformed support features and masks, enriches foreground information of support feature, and enhances similarity between support and query foregrounds, reducing gap between them. Theoretical analysis appears in appendix A.1; algorithm pipeline in appendix A.2; deformation-method comparison in appendix C.4.

Deformation-aware Generation Module. High-quality prototypes are crucial for determining performance of FSMIS. Existing methods (*e.g.*, DIFD [9], CoW [15], DSM [48]) obtain prototypes based on local features, pixels, or mean centers of single morphology support feature, but generally neglect inherent spatial orientation prior and deformation gap of medical images. Medical target show continuous vertical changes and horizontal structural consistency. Large rotations are rare in medical imaging due to standardized protocols. Thus, we average adjacent rows and columns of deformed features to capture both vertical texture variations and horizontal structural consistency. Specifically, we sum adjacent rows and columns of foreground feature $\boldsymbol{F}^n_{\mathrm{s,fg}}$ and background feature $\boldsymbol{F}^n_{\mathrm{s,bg}}$ to extract local foreground prototype $\boldsymbol{P}^n_{\mathrm{lfg}}$ and local background prototype $\boldsymbol{P}^n_{\mathrm{lbg}}$. To further capture global context and boundary details, we extract global prototypes ($\boldsymbol{P}^n_{\mathrm{gfg}}$, $\boldsymbol{P}^n_{\mathrm{gbg}}$) and edge prototype $\boldsymbol{P}^n_{\mathrm{bfg}}$ through global average pooling and dilation and erosion operations, respectively. These prototypes, called elastic prototypes, are used to jointly guide subsequent model segmentation, as follows:

$$\begin{aligned} \boldsymbol{P}^n_{\mathrm{lfg}} = Sum_{row}(\boldsymbol{F}^n_{\mathrm{s,fg}}), \quad \boldsymbol{P}^n_{\mathrm{lbg}} = Sum_{col}(\boldsymbol{F}^n_{\mathrm{s,bg}}), \quad \boldsymbol{P}^n_{\mathrm{lfg/lbg}} \in \mathbb{R}^{B\times C\times T_1/T_2}, \\ \boldsymbol{P}^n_{\mathrm{gfg}} = Gap(\boldsymbol{F}^n_{\mathrm{s,fg}}), \quad \boldsymbol{P}^n_{\mathrm{gbg}} = Gap(\boldsymbol{F}^n_{\mathrm{s,bg}}), \quad \boldsymbol{P}^n_{\mathrm{bfg}} = Dil(\boldsymbol{F}^n_{\mathrm{s,fg}}) - Ero(\boldsymbol{F}^n_{\mathrm{s,fg}}). \end{aligned} \tag{3}$$

where $Sum_{row}(\cdot)$ and $Sum_{col}(\cdot)$ denote summation of adjacent rows and columns, respectively. $Gap(\cdot)$ is global average pooling. $Dil(\cdot)$ and $Ero(\cdot)$ denote dilation and erosion operations.

3.3 ELASTIC LOCALIZATION MODULE

Medical images are often accompanied by complex background interference. To suppress impact of background on segmentation performance, we utilize elastic prototypes to explicitly enhance query foreground feature. Specifically, $\boldsymbol{P}^n_{\text{gfg}} \in \mathbb{R}^{B\times C\times 1}$ is first reshaped to size $\mathbb{R}^{B\times C\times 1\times 1}$, and a cosine similarity is calculated with $\boldsymbol{F}_{\text{q}} \in \mathbb{R}^{B\times C\times H_{\text{q}}\times W_{\text{q}}}$ to obtain an initial similarity matrix. Then, the system sequentially passes through *Sigmoid* activation function and *Max* operation to extract maximum similarity response at each spatial location, thereby generating a foreground attention map. Finally, this attention map is multiplied element-wise with query feature to highlight foreground target, and original query feature is fused through residual connections to ensure integrity of feature representation. The above process is as follows:

$$\boldsymbol{F}^{\text{e}}_{\text{q}} = \boldsymbol{F}_{\text{q}}(1 + Max(Sig(Cos(Re(\boldsymbol{P}^n_{\text{gfg}}), \boldsymbol{F}_{\mathbf{q}})))). \tag{4}$$

where Max, Sig, Cos, and Re denote *Max* operation, *Sigmoid* activation function, cosine similarity calculating function, and reshaping operation, respectively. The $\boldsymbol{F}^{\text{e}}_{\text{q}}$ is enhanced query feature.

3.4 ELASTIC AGGREGATION MODULE

Given varied shapes and scales of targets in medical images, we propose EAM based on multi-scale decoder and multiple Top-K elastic prototypes selection strategy. Specifically, we bypass iterative clustering by ranking L_2 norms to extract high-confidence Top-K prototypes from respective $\boldsymbol{P}^n_{\text{gfg}}$, $\boldsymbol{P}^n_{\text{lfg}}$, $\boldsymbol{P}^n_{\text{bfg}}$, $\boldsymbol{P}^n_{\text{gbg}}$, and $\boldsymbol{P}^n_{\text{lbg}}$, thereby integrating them into Top-K foreground elastic prototypes $\boldsymbol{P}^n_{\text{tfg}}$, boundary elastic prototypes $\boldsymbol{P}^n_{\text{tby}}$, and background elastic prototypes $\boldsymbol{P}^n_{\text{tbg}}$. The specific K values for different prototypes will be explained in the experimental section. Then, cosine similarity between these selected prototypes and enhanced query feature $\boldsymbol{F}^{\text{e}}_{\text{q}}$ is calculated to generate multi-dimensional similarity response maps, as follows:

$$\begin{gathered}\boldsymbol{P}^n_{\text{tfg}} = TopK(\boldsymbol{P}^n_{\text{gfg}}, \boldsymbol{P}^n_{\text{lfg}}), \quad \boldsymbol{P}^n_{\text{tby}} = TopK(\boldsymbol{P}^n_{\text{bfg}}), \quad \boldsymbol{P}^n_{\text{tbg}} = TopK(\boldsymbol{P}^n_{\text{gbg}}, \boldsymbol{P}^n_{\text{lbg}}), \\ \boldsymbol{S}^n_{\text{tfg}} = Cos(Cat(\boldsymbol{P}^n_{\text{tfg}}, \boldsymbol{P}^n_{\text{tby}}), \boldsymbol{F}^{\text{e}}_{\text{q}}), \quad \boldsymbol{S}^n_{\text{tbg}} = Cos(\boldsymbol{P}^n_{\text{tbg}}, \boldsymbol{F}^{\text{e}}_{\text{q}}),\end{gathered} \tag{5}$$

where $TopK$, Cos, and Cat denote Top-K selection strategy, cosine similarity calculating function, and stacking operations respectively. $\boldsymbol{S}^n_{\text{tfg}}$ and $\boldsymbol{S}^n_{\text{tbg}}$ are foreground and background similarity response maps. Next, these response maps are input into a multi-scale decoder constructed from convolutions with different receptive fields to fully capture rich local details and generate preliminary segmentation predictions. Finally, *Softmax* function is used to calculate weights assigned to these segmentation results, and adaptive fusion mechanism is used to aggregate these weighted results to obtain $\boldsymbol{M}_{\text{q,p}}$. Inspired by FoB+SAM [5] and ProtoSAM [2], we also use SAM during testing to refine $\boldsymbol{M}_{\text{q,p}}$ and produce final prediction $\boldsymbol{F}_{\text{out}}$. The above process is as follows:

$$\begin{gathered}\boldsymbol{M}_{\text{fg,q}} = Sof(Msd(\boldsymbol{S}^n_{\text{tfg}}))\boldsymbol{S}^n_{\text{tfg}}, \ \boldsymbol{M}_{\text{bg,q}} = Sof(Msd(\boldsymbol{S}^n_{\text{tbg}}))\boldsymbol{S}^n_{\text{tbg}}, \\ \boldsymbol{M}_{\text{q,p}} = Cat(\boldsymbol{M}_{\text{bg,q}}, \boldsymbol{M}_{\text{fg,q}}), \ \boldsymbol{F}_{\text{out}} = SAM(\boldsymbol{M}_{\text{q,p}}, \text{pts}, \text{box}).\end{gathered} \tag{6}$$

where *Sof* and Msd are Softmax activation function and multi-scale convolutional decoder with different receptive fields, respectively. *pts* and *box* denote point and bounding-box prompts, respectively. They are obtained using same prompt-generation method as that employed in ProtoSAM.

3.5 DEFORMATIONS BRIDGE SUPPORT-QUERY GAP

The support-query gap in EFG-SAM is alleviated through deformation space expansion and selective feature matching. *Deformation Space Expansion:* EFDM augments the deterministic support feature F_s by generating diverse deformed features F^n_s. These features construct a candidate manifold under local anatomical constraints, providing richer matching possibilities. *Directional Selective Matching:* EAM adopts a Top-K selection strategy to discard irrelevant candidates and preserve prototypes with higher structural similarity to the query feature F_q. Consequently, the original distance $d(F_s, F_q)$ is transformed into a nearest-neighbor matching objective: $\min_n d(F^n_s, F_q)$. When the expanded feature space covers the latent representation of the query, it satisfies: $\min_n d(F^n_s, F_q) \leq d(F_s, F_q)$. Thus, EFDM improves feature diversity and decreases L2 distance between support and query features as shown in appendix C.5, while EAM provides directional selection, jointly increasing the likelihood of identifying a compatible support representation and narrowing the representation gap between support and query features.

Table 1: Method comparisons on Abd-MRI and Abd-CT under two settings. Results are reported in Mean Dice (%). Best prediction results are in **bold**.

	Methods	Abdominal-MRI					Abdominal-CT				
		LK	RK	Spleen	Liver	Mean	LK	RK	Spleen	Liver	Mean
Setting I	GMRD [8]	83.96	90.12	76.09	81.42	82.90	81.70	74.46	78.31	79.60	78.52
	DGPANet [27]	85.84	86.99	80.64	81.31	83.70	82.67	79.56	83.28	65.59	77.77
	PMCR [43]	81.38	87.87	77.06	78.37	81.17	71.76	76.18	70.43	79.86	74.56
	CoW [15]	85.58	90.27	79.00	81.55	84.10	83.60	80.46	83.56	82.35	82.49
	DIFD [9]	**88.59**	91.19	75.72	80.99	84.12	83.03	78.67	79.41	79.66	80.19
	MAUP [45]	58.23	72.34	59.65	78.16	67.09	59.41	71.80	60.38	78.25	67.46
	DSM [48]	82.63	89.47	78.23	82.12	83.11	83.59	80.41	85.72	84.78	83.62
	FAMNet [4]	78.98	88.76	73.33	80.54	80.40	79.80	78.58	81.93	84.30	81.15
	ProtoSAM [2]	82.75	82.36	77.98	83.14	81.56	71.31	75.67	70.24	84.79	75.50
	RPT [46]	80.72	89.82	76.37	82.86	82.44	77.05	79.13	72.58	82.57	77.83
	FoB+SAM [5]	83.85	89.84	**82.30**	81.42	84.35	85.90	85.71	84.34	**87.10**	85.76
	EFG-SAM	87.78	**92.40**	78.93	**83.62**	**85.68**	**86.85**	**85.77**	**86.88**	85.90	**86.35**
Setting II	GMRD [8]	78.65	86.66	73.25	80.25	79.70	77.40	76.17	75.30	80.39	77.32
	DGPANet [27]	73.76	75.96	69.21	75.93	73.72	74.10	68.06	65.91	65.56	68.41
	PMCR [43]	76.86	82.33	75.08	77.23	77.88	67.32	69.54	70.07	76.42	70.84
	CoW [15]	82.40	89.15	74.55	80.46	81.64	83.36	79.44	79.92	82.26	81.25
	DIFD [9]	85.38	90.54	70.96	80.38	81.82	84.47	75.48	78.67	80.77	79.85
	MAUP [45]	58.23	72.34	59.65	78.16	67.09	59.41	71.80	60.38	78.25	67.46
	DSM [48]	78.64	89.20	73.84	81.10	80.69	83.01	81.25	84.72	83.29	83.07
	FAMNet [4]	74.45	85.95	71.67	81.00	78.27	78.36	79.16	79.13	83.17	79.95
	ProtoSAM [2]	71.46	81.43	**76.51**	81.94	77.83	69.44	71.04	65.50	**87.84**	73.45
	RPT [46]	78.33	86.01	75.46	76.37	79.04	72.99	67.73	70.80	75.24	71.69
	FoB+SAM [5]	78.21	87.91	73.30	**82.43**	80.46	**88.55**	**85.91**	82.43	82.29	84.80
	EFG-SAM	**85.70**	**91.02**	76.29	81.72	**83.68**	87.45	83.12	**86.55**	86.86	**85.99**

Table 2: Comparison with state-of-the-art methods under Setting 1. Results are reported in Dice (%). The best prediction results are highlighted in **bold**.

Methods	Card-MRI				Total Mean Dice	FLOPs ↓	FPS ↑	Params ↓
	LV-BP	LV-MYO	RV	Mean				
GMRD [8]	90.00	67.04	80.29	79.11	79.51	321.92	26.25	100.03
DGPANet [27]	89.82	67.62	80.09	79.18	76.56	264.72	15.71	46.12
PMCR [43]	89.47	65.33	79.33	78.04	76.50	–	–	–
CoW [15]	89.25	68.41	82.33	80.00	81.90	643.92	14.48	132.89
DIFD [9]	90.03	68.72	80.76	79.84	81.16	331.30	7.55	61.85
MAUP [45]	88.36	52.74	78.29	73.13	68.45	2933.44	3.90	945.46
DSM [48]	90.56	**68.45**	82.01	80.34	82.17	333.21	19.03	63.48
FAMNet [4]	**90.73**	62.11	77.80	76.88	79.33	**113.11**	**104.18**	**28.63**
ProtoSAM [2]	–	–	–	–	–	4132.50	1.97	945.46
RPT [46]	89.90	66.91	80.78	79.19	78.04	265.98	25.06	50.33
FoB+SAM [5]	–	–	–	–	–	3005.73	4.25	695.65
EFG-SAM	89.37	67.90	**86.87**	**81.38**	**84.62**	658.22	5.86	142.39

4 EXPERIMENTS

4.1 EXPERIMENTAL SETUP

Datasets and Settings. We use two NVIDIA RTX 4090 GPUs for training. We conduct experiments on three mainstream datasets: Abd-MRI [16], Abd-CT [18], and Card-MRI [49], against 11 state-of-the-art (SOTA) methods. Experiments use Sorensen-Dice coefficient (Dice), a standard metric in FSMIS domain, for evaluation, and strictly adhere to 1-way 1-shot five-fold cross-validation to ensure fairness. To fully evaluate model capabilities, two standard settings are introduced: **setting1** allows test class to appear as background in training set; **setting2** has stricter conditions, completely excluding test class. Due to specific data structure, Card-MRI is evaluated only under setting1. Following CoW [15], 2D slices are cropped to 256 $\times$ 256 for Abd-MRI and Card-MRI, and 257 $\times$ 257 for Abd-CT. We set batch size to 1. The initial learning rate is set to 0.001 (decay rate 0.9), and the number of iterations is set to 50K (5000 iterations/epoch). The number K of $\boldsymbol{P}^n_{\text{tfg}}$, $\boldsymbol{P}^n_{\text{tbg}}$, and $\boldsymbol{P}^n_{\text{tby}}$ are fixed at 250, 600, and 150, respectively. To validate its unseen domain robustness, we evaluate our method on Abd-CT to Card-MRI transfer task. For details, see Appendix B.1 and B.2.

4.2 COMPARISON WITH STATE-OF-THE-ARTS

Quantitative Evaluation. We reproduce and compare FSS methods: GFSAM [42], CoW [15], DIFD [9], FSSAM [41], SFMNet [44], AdaptiveFSS [36], IFA [23], DR-Adapter [29], MAUP [45],

Table 3: Different methods **Cross-domain** comparison under setting1.

	Methods	Liver	RK	LK	Spleen	Mean
CT → MRI	GMRD [8]	63.15	67.69	61.79	56.89	62.38
	DIFD [9]	68.29	79.69	**75.51**	61.05	71.14
	MAUP [45]	58.23	72.34	59.65	78.16	67.09
	DSM [48]	**78.12**	47.52	47.73	53.81	56.79
	FAMNet [4]	73.01	74.68	57.28	58.21	65.79
	RPT [46]	52.58	58.44	42.58	54.94	52.13
	FoB+SAM [5]	75.05	79.57	70.38	**68.21**	73.30
	EFG-SAM	76.04	**79.70**	75.14	67.00	**74.47**
MRI → CT	GMRD [8]	66.12	56.37	57.38	54.56	58.61
	DIFD [9]	73.86	46.88	56.86	62.45	60.01
	MAUP [45]	59.41	71.80	60.38	78.25	67.46
	DSM [48]	68.84	59.75	55.41	53.22	59.31
	FAMNet [4]	73.57	61.89	57.79	65.78	64.75
	RPT [46]	59.47	49.60	40.29	52.04	50.35
	FoB+SAM [5]	81.36	58.81	57.18	70.71	67.02
	EFG-SAM	**82.49**	**70.78**	**70.21**	**77.80**	**75.32**

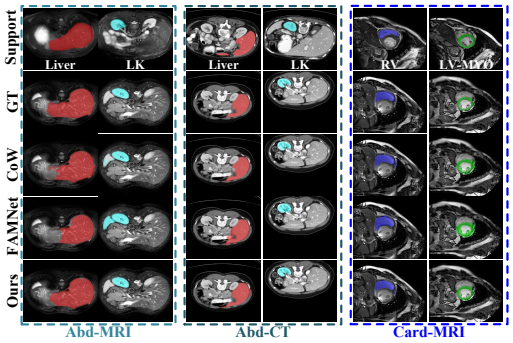


Figure 4: **Qualitative comparison** of different approaches on three datasets.

Table 4: **Ablation of each component** of ours under setting1 on Abd-MRI dataset.

EFDM	DGM	ELM	EAM	Abd-MRI LK	RK	Spleen	Liver	Mean
				76.72	88.69	72.09	80.72	79.55
✓				77.12	90.85	72.70	81.57	80.56
✓	✓			84.17	91.28	**81.82**	79.48	84.19
✓	✓	✓		85.42	92.39	80.36	80.79	84.74
✓	✓	✓	✓	**87.78**	**92.40**	78.93	**83.62**	**85.68**

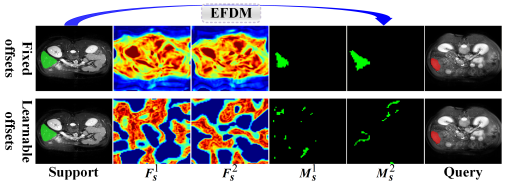


Figure 5: **Effectiveness of different offsets settings in EFDM, Fixed (top row) v.s. learnable (bottom row)**: learnable offsets lead to invalid deformation.

DSM [48], and FAMNet [4]. Table 1 shows experimental results of different methods on Abd-MRI and Abd-CT datasets. Our method outperforms other methods in both setting1 and setting2. Specifically, in setting2 of Abd-MRI and Abd-CT datasets, our method improves five-fold average result of four categories by 1.37% and 1.68% Dice, respectively, compared to the second-best method. This demonstrates superiority of our method. Table 2 shows experimental results on Card-MRI, where we achieve the best performance compared to other methods. We further analyze model's floating point operations (FLOPs), frames per second (FPS), and model parameters (Para.). As shown in Table 2, although ours complexity is not the best, compared to DSM with similar parameters, our method improves performance by 4.23% on Abd-CT setting2, and compared to FSSAM with similar FPS, our method improves performance by 8.31% on Abd-MRI setting2. In summary, our method achieves a good balance between performance and model complexity.

Cross-Domain Quantitative Evaluation. We conduct cross-domain experiments to verify robustness and superiority of ours method. As shown in Table 3, we select several competing models with same backbone for a fair comparison. In this setup, Abd-CT serves as source domain and Card-MRI as target domain. Specifically, models are trained under setting1 of Abd-CT and subsequently evaluated directly on Card-MRI without any further fine-tuning. The results demonstrate that our model possesses a significantly stronger cross-domain generalization capability.

Qualitative Evaluation. We visualize prediction results of several methods on Abd-MRI, Abd-CT, and Card-MRI. As shown in Fig. 4, *Support* and *GT* illustrate ground truth overlaid on support and query images, respectively. Our method provides more complete and accuracy segmentation results for Liver, LK, RV, and LV-MYO categories compared to other methods. This shows effectiveness of EPG, ELM, and EAM. More visualizations are available in appendix material C.6.

4.3 Ablation Studies

Each component of EFG-SAM. As shown in Table 4, we conduct ablation studies on each component. The results indicate that our method performance improves progressively as each module is integrated. Starting from a baseline Dice score of 80.29% (without any modules), the performance rises to 84.13% upon the addition of EPG (EFDM and DGM). Ultimately, our method achieves its peak performance only when all four components are present, demonstrating effectiveness of each individual part. For a detailed visualization analysis of this part, please refer to appendix material C.7.

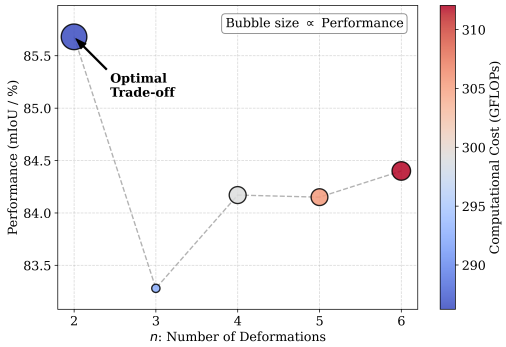


Figure 6: **Ablation on deformation count *n*** in EFDM. $n = 2$ yields optimal trade-off.

Table 5: Ablation on the effect of **deformation count *n*** and **different offsets** in EFDM.

Settings	Abd-MRI setting1				
	LK	RK	Spleen	Liver	Mean
Deformation count n					
3	84.53	91.88	75.85	80.88	83.28
4	83.23	91.90	77.97	83.57	84.17
5	85.80	91.35	76.66	82.79	84.15
6	83.55	91.49	**79.51**	83.05	84.40
2	**87.78**	**92.40**	78.93	**83.62**	**85.68**
Offsets					
learnable	81.14	91.04	77.96	81.88	83.00
1	87.55	90.61	77.16	81.90	84.31
0.5	86.81	92.31	80.41	82.63	85.54
0.1	**87.78**	**92.40**	78.93	**83.62**	**85.68**

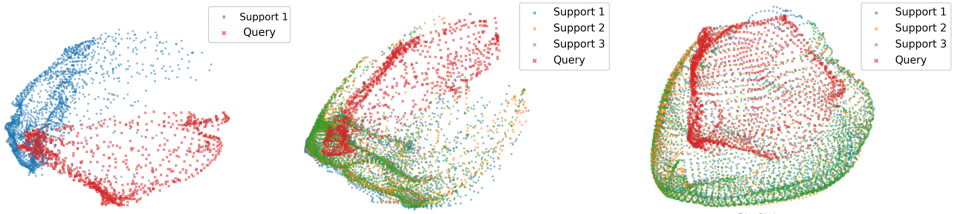


Figure 7: Visualization of the entire process of reducing feature gap between support and query. Support1 holds original features; Support2/3 are the two types of enhanced query features by EFDM with two different groups of control points.

Furthermore, Fig. 5 shows that when using learnable offsets, the deformation feature-mask generated by EFDM is chaotic, likely due to lazy learning, leading to incorrect deformation. Conversely, using a fixed offset forces EFDM to generate deformation feature-mask, thus preventing this problem. To further analyze optimal offset value, we test model with offsets of 0.05, 0.1, 0.5, and 1, finding that **offset of 0.1** yields best result.

Deformation count *n*. To analyze impact of deformation quantity n, within EFDM on overall model performance, we set n to values of 2, 3, 4, 5, and 6. As shown in Table 5, we observe that performance is lower with n = 3 and n = 4 than with n = 2. Although n = 6 shows improvement over n = 3, 4, and 5, its performance is still lower than when n = 2. Overall, our method achieves the best performance with n = 2. Furthermore, to facilitate an intuitive comprehension, we also visualize a bubble chart of deformation quantity. As shown in Fig. 6, larger circles indicate higher Dice, and color closer to red indicates larger FLOPs. When ***n* is 2**, our method performance and FLOPs reach optimal balance. For $n \geq 3$, multi-deformations introduce topological noise, generating anatomically implausible structures that corrupt prototype semantics and mislead matching.

Different offsets of EFDM. To analyze impact of deformation magnitude on model performance, we explore quantitative analysis for learnable and fixed offsets. Table 5 shows that our model with learnable offsets performs worse than model with fixed offsets. In order to analyze entire process of reducing gap between support and query in model, we plot encoder feature PCA diagrams of EFDM and complete model as shown in Fig. 7. It can be clearly seen that as modules are introduced, the gap between them decreases, proving the usefulness of our method.

Number of block-wise control points *N*. We set block-level control points N to 4, 8, 16, and 32 to analyze impact of different deformation densities on model performance. A larger value of N indicates a higher deformation density, and vice versa. We find that N = 8 performs better than N = 4, but as N continued to increase, the model performance actually decreased. This indicates that a higher deformation density does not necessarily lead to better model performance. Table 6 shows that, based on experimental analysis, ***N* = 8** achieved good performance.

Different elastic prototypes. We analyze impact of $\boldsymbol{P}^n_{\text{tfg}}$, $\boldsymbol{P}^n_{\text{tfg}}$ plus $\boldsymbol{P}^n_{\text{tbg}}$ prototypes, $\boldsymbol{P}^n_{\text{tfg}}$ plus $\boldsymbol{P}^n_{\text{tby}}$ prototypes, and $\boldsymbol{P}^n_{\text{tfg}}$ plus $\boldsymbol{P}^n_{\text{tbg}}$ plus $\boldsymbol{P}^n_{\text{tby}}$ prototypes on model performance. As shown in Table 7, we

Table 6: Effect of **block-wise control points number *N*** in EFDM on Abd-MRI setting1.

N	Abd-MRI				
	LK	RK	Spleen	Liver	Mean
4	81.88	89.87	76.84	81.32	82.48
16	83.62	90.18	76.23	81.92	82.99
32	82.91	91.99	77.26	81.38	83.39
8	**87.78**	**92.40**	**78.93**	**83.62**	**85.68**

Table 7: Effectiveness analysis of **different elastic prototypes** on Abd-MRI setting1.

$\boldsymbol{P}^n_{\text{tfg}}$	$\boldsymbol{P}^n_{\text{tbg}}$	$\boldsymbol{P}^n_{\text{tby}}$	Abd-MRI				
			LK	RK	Spleen	Liver	Mean
✓			79.62	92.00	74.88	80.79	81.82
✓	✓		84.23	89.56	78.19	82.11	83.52
✓		✓	83.23	92.13	75.85	81.31	83.13
✓	✓	✓	**87.78**	**92.40**	**78.93**	**83.62**	**85.68**

Table 8: Effectiveness analysis of the **number of prototypes *K*** on setting1.

Dice	Abd-MRI K_{tfg}=250, K_{tbg}=600, K_{tby} =					Abd-MRI K_{tfg}=250, K_{tby}=150, K_{tbg} =					Abd-MRI K_{tbg}=600, K_{tby}=150, K_{tfg} =				
	50	100	150	200	300	200	400	600	800	1000	50	150	250	350	500
LK	87.04	**88.07**	87.78	84.22	79.08	82.77	85.69	**87.78**	85.40	84.78	86.42	82.42	**87.78**	83.43	83.93
RK	92.20	91.80	**92.40**	88.05	89.17	91.63	91.98	**92.40**	92.37	91.52	92.12	91.52	**92.40**	92.14	91.86
Spleen	73.49	77.19	78.93	**80.48**	77.67	75.83	**80.22**	78.93	77.49	78.46	78.28	79.33	78.93	**79.94**	78.71
Liver	81.65	83.11	83.62	**84.06**	82.85	81.77	82.15	**83.62**	82.45	82.69	82.99	81.80	83.62	82.47	**84.35**
Mean	83.59	85.04	**85.68**	84.20	82.19	83.00	85.01	**85.68**	84.43	84.36	84.95	83.77	**85.68**	84.50	84.71

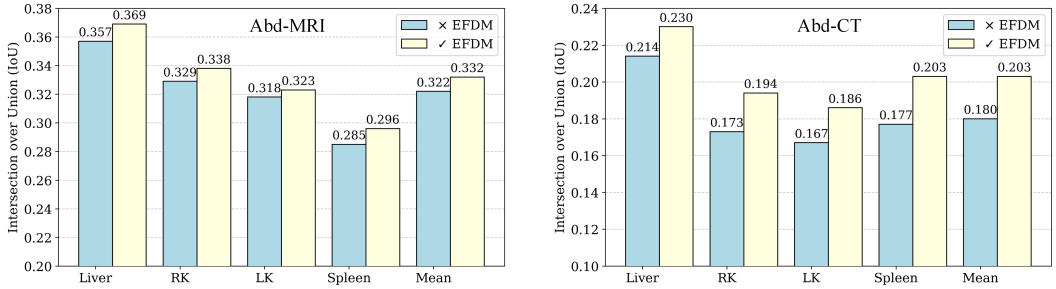


Figure 8: Deformed Support-Query Mask IoU w/o EFDM on two datasets under setting1.

find that boundary prototypes and background prototypes provide certain discriminative information to model, helping it to identify foreground target.

Number of prototypes *K*. We analyze impact of number of $\boldsymbol{P}^n_{\text{tfg}}$, $\boldsymbol{P}^n_{\text{tbg}}$, and $\boldsymbol{P}^n_{\text{tby}}$ on model performance. As indicated in Table 8, given $\boldsymbol{K}_{\text{tfg}} = 250$ and $\boldsymbol{K}_{\text{tbg}} = 600$, the optimal $\boldsymbol{K}_{\text{tby}}$ is 150; given $\boldsymbol{K}_{\text{tfg}} = 250$ and $\boldsymbol{K}_{\text{tby}} = 150$, the optimal $\boldsymbol{K}_{\text{tbg}}$ is 600; and given $\boldsymbol{K}_{\text{tbg}} = 600$ and $\boldsymbol{K}_{\text{tby}} = 150$, the optimal $\boldsymbol{K}_{\text{tfg}}$ is 250. In conclusion, model yields the best result when **$\boldsymbol{K}_{\text{tfg}}$=250, $\boldsymbol{K}_{\text{tbg}}$=600, and $\boldsymbol{K}_{\text{tby}}$=150**.

Effectiveness analysis of EFDM deformation. As shown in Fig. 8, we calculate IoU between support and query masks before and after EFDM deformation. In Abd-MRI setting1 (left), IoU of masks after EFDM is higher than those of undeformed baseline. Furthermore, the generally low baseline IoU scores fully demonstrate existence of a significant spatial deformation gap between support and query masks. Experiments demonstrate that our proposed method mitigates deformation gap, validating module's effectiveness and providing direction for subsequent feature matching.

5 Conclusion and Limitation

To address ideal matching trap in FSMIS caused by non-rigid deformation and intra-class differences, this paper proposes EFG-SAM. To overcome limitation of single morphological features, this model first extracts a resilient prototype encompassing both global semantics and local details through EFDM and DGM strategies in EPG stage, effectively alleviating insufficiency of support information. Subsequently, the model utilizes ELM and EAM to suppress background noise interference, achieving efficient guided segmentation across multiple scales. During inference, SAM is utilized within EFG to refine segmentation results without requiring additional training. Extensive experiments on three benchmark datasets demonstrate that EFG-SAM not only surpasses existing methods, but also exhibits superior cross-domain generalization ability, providing a more robust solution for FSMIS field.

Limitations. Although our method has performance advantages and the number of parameters is within an acceptable range, its computational cost and memory consumption are significant, which is an area that needs to be improved in the future work. Additionally, the performance drop on spleen indicates that EAM over-complicates feature representations for large, homogeneous organs.

REFERENCES

[1] Zhaochong An, Guolei Sun, Yun Liu, Runjia Li, Junlin Han, Ender Konukoglu, and Serge Belongie. Generalized few-shot 3d point cloud segmentation with vision-language model. In *2025 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, pp. 16997–17007, 2025. doi: 10.1109/CVPR52734.2025.01584.

[2] Lev Ayzenberg, Raja Giryes, and Hayit Greenspan. Protosam for automated one shot medical image segmentation using foundational models. *Scientific Reports*, 15(1):41482, 2025.

[3] Hanbo Bi, Yingchao Feng, Wenhui Diao, Peijin Wang, Yongqiang Mao, Kun Fu, Hongqi Wang, and Xian Sun. Prompt-and-transfer: Dynamic class-aware enhancement for few-shot segmentation. *IEEE transactions on pattern analysis and machine intelligence*, 47(1):131–148, 2024.

[4] Yuntian Bo, Yazhou Zhu, Lunbo Li, and Haofeng Zhang. Famnet: Frequency-aware matching network for cross-domain few-shot medical image segmentation. In *Proceedings of the AAAI conference on artificial intelligence*, volume 39, pp. 1889–1897, 2025.

[5] Yuntian Bo, Yazhou Zhu, Piotr Koniusz, and Haofeng Zhang. Focus on background: Exploring sam's potential in few-shot medical image segmentation with background-centric prompting. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, pp. 30032–30041, June 2026.

[6] Mathilde Caron, Hugo Touvron, Ishan Misra, Hervé Jégou, Julien Mairal, Piotr Bojanowski, and Armand Joulin. Emerging properties in self-supervised vision transformers. In *Proceedings of the IEEE/CVF international conference on computer vision*, pp. 9650–9660, 2021.

[7] Yadang Chen, Xinyu Xu, Yuhui Zheng, Zhi-Xin Yang, and Enhua Wu. Refinement on both foreground and background prototype for few–shot segmentation. *IEEE Transactions on Multimedia*, pp. 1–11, 2026. doi: 10.1109/TMM.2026.3668634.

[8] 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*, 43(6):2202–2214, 2024.

[9] Ziming Cheng, Shidong Wang, Yang Long, Tao Zhou, Haofeng Zhang, and Ling Shao. Dual interspersion and flexible deployment for few-shot medical image segmentation. *IEEE Transactions on Medical Imaging*, 2025.

[10] Ziming Cheng, Yazhou Zhu, Shidong Wang, Tong Xin, and Haofeng Zhang. Cascaded alternating refinement transformer for few-shot medical image segmentation. *ACM Transactions on Intelligent Systems and Technology*, 2025.

[11] 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.

[12] Kaiming He, Xiangyu Zhang, Shaoqing Ren, and Jian Sun. Deep residual learning for image recognition. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pp. 770–778, 2016.

[13] Jiankang Hong, Ye Luo, Yinan Liu, and Junsong Yuan. From infusion to assimilation distillation for medical image segmentation. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pp. 20985–20995, 2026.

[14] Junfei Hu, Tao Zhou, Kaiwen Huang, Yi Zhou, Haofeng Zhang, Boqiang Fan, and Huazhu Fu. Uncertainty-guided prototype reliability enhancement network for few-shot medical image segmentation. *IEEE Transactions on Medical Imaging*, 2025.

[15] Jianchao Jiang and Haofeng Zhang. Concentrate on weakness: mining hard prototypes for few-shot medical image segmentation. In *Proceedings of the Thirty-Fourth International Joint Conference on Artificial Intelligence*, pp. 1242–1250, 2025.

[16] A Emre Kavur, N Sinem Gezer, Mustafa Barış, Sinem Aslan, Pierre-Henri Conze, Vladimir Groza, Duc Duy Pham, Soumick Chatterjee, Philipp Ernst, Savaş Özkan, et al. Chaos challenge-combined (ct-mr) healthy abdominal organ segmentation. *Medical image analysis*, 69:101950, 2021.

[17] 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*, pp. 4015–4026, 2023.

[18] Bennett Landman, Zhoubing Xu, Juan Igelsias, Martin Styner, Thomas Langerak, and Arno Klein. Miccai multi-atlas labeling beyond the cranial vault–workshop and challenge. In *Proc. MICCAI multi-atlas labeling beyond cranial vault—workshop challenge*, volume 5, pp. 12. Munich, Germany, 2015.

[19] Tianang Leng, Yiming Zhang, Kun Han, and Xiaohui Xie. Self-sampling meta sam: enhancing few-shot medical image segmentation with meta-learning. In *Proceedings of the IEEE/CVF winter conference on applications of computer vision*, pp. 7925–7935, 2024.

[20] Yi Lin, Yufan Chen, Kwang-Ting Cheng, and Hao Chen. Few shot medical image segmentation with cross attention transformer. In *International Conference on Medical Image Computing and Computer-Assisted Intervention*, pp. 233–243. Springer, 2023.

[21] Xinyu Liu, Zhen Chen, Wuyang Li, Chenxin Li, and Yixuan Yuan. Harnessing lightweight transformer with contextual synergic enhancement for efficient 3d medical image segmentation. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 48(3):3852–3867, 2026. doi: 10.1109/TPAMI.2025.3640233.

[22] Yufei Liu, Haoke Xiao, Jiaxing Chai, Yongcun Zhang, Rong Wang, Zijie Meng, and Zhiming Luo. Synpo: Boosting training-free few-shot medical segmentation via high-quality negative prompts. In *International Conference on Medical Image Computing and Computer-Assisted Intervention*, pp. 594–603. Springer, 2025.

[23] Jiahao Nie, Yun Xing, Gongjie Zhang, Pei Yan, Aoran Xiao, Yap-Peng Tan, Alex C Kot, and Shijian Lu. Cross-domain few-shot segmentation via iterative support-query correspondence mining. In *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*, pp. 3380–3390, 2024.

[24] 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 *European conference on computer vision*, pp. 762–780. Springer, 2020.

[25] Dandan Shan, Zihan Li, Yunxiang Li, Qingde Li, Jie Tian, and Qingqi Hong. Stpnet: Scale-aware text prompt network for medical image segmentation. *IEEE Transactions on Image Processing*, 34:3169–3180, 2025. doi: 10.1109/TIP.2025.3571672.

[26] Qianqian Shen, Yanan Li, Jiyong Jin, and Bin Liu. Q-net: Query-informed few-shot medical image segmentation. In *Proceedings of SAI Intelligent Systems Conference*, pp. 610–628. Springer, 2023.

[27] Yue Shen, Wanshu Fan, Cong Wang, Wenfei Liu, Wei Wang, Qiang Zhang, and Dongsheng Zhou. Dual-guided prototype alignment network for few-shot medical image segmentation. *IEEE Transactions on Instrumentation and Measurement*, 73:1–13, 2024.

[28] Feifan Song, Ziming Cheng, Lunbo Li, and Haofeng Zhang. Mtpnet: Learning multiple twin-support prototypes for few-shot medical image segmentation. In *2024 IEEE International Conference on Bioinformatics and Biomedicine (BIBM)*, pp. 2414–2421. IEEE, 2024.

[29] Jiapeng Su, Qi Fan, Wenjie Pei, Guangming Lu, and Fanglin Chen. Domain-rectifying adapter for cross-domain few-shot segmentation. In *Proceedings of the IEEE/CVF conference on Computer Vision and Pattern Recognition*, pp. 24036–24045, 2024.

[30] Hao Tang, Xingwei Liu, Shanlin Sun, Xiangyi Yan, and Xiaohui Xie. Recurrent mask refinement for few-shot medical image segmentation. In *Proceedings of the IEEE/CVF international conference on computer vision*, pp. 3918–3928, 2021.

[31] Song Tang, Shaxu Yan, Xiaozhi Qi, Jianxin Gao, Mao Ye, Jianwei Zhang, and Xiatian Zhu. Few-shot medical image segmentation with high-fidelity prototypes. *Medical Image Analysis*, 100:103412, 2025.

[32] Pinzhuo Tian, Xiaojie Zhou, Hang Yu, and Shaorong Xie. High specificity guided cross-domain few-shot segmentation. *IEEE Transactions on Multimedia*, 27:9059–9071, 2025. doi: 10.1109/TMM.2025.3613173.

[33] Yuan Tian, Yongquan Liang, Yufeng Chen, Jingjing Zhang, and Hongyang Bian. Multilevel support-assisted prototype optimization network for few-shot medical segmentation of lung lesions. *Scientific Reports*, 15(1):3290, 2025.

[34] Jintao Tong, Yixiong Zou, Yuhua Li, and Ruixuan Li. Lightweight frequency masker for cross-domain few-shot semantic segmentation. *Advances in Neural Information Processing Systems*, 37:96728–96749, 2024.

[35] Ashish Vaswani, Noam Shazeer, Niki Parmar, Jakob Uszkoreit, Llion Jones, Aidan N Gomez, Łukasz Kaiser, and Illia Polosukhin. Attention is all you need. *Advances in neural information processing systems*, 30, 2017.

[36] Jing Wang, Jiangyun Li, Chen Chen, Yisi Zhang, Haoran Shen, and Tianxiang Zhang. Adaptive fss: a novel few-shot segmentation framework via prototype enhancement. In *Proceedings of the AAAI conference on artificial intelligence*, volume 38, pp. 5463–5471, 2024.

[37] Kaixin Wang, Jun Hao Liew, Yingtian Zou, Daquan Zhou, and Jiashi Feng. Panet: Few-shot image semantic segmentation with prototype alignment. In *proceedings of the IEEE/CVF international conference on computer vision*, pp. 9197–9206, 2019.

[38] Jiahong Wei, Bing Xue, and Mengjie Zhang. Multi-objective once-for-all neural architecture search for medical image segmentation. *IEEE Transactions on Image Processing*, 35:2981–2996, 2026. doi: 10.1109/TIP.2026.3671022.

[39] Lili Wei, Congyan Lang, Zheming Xu, Liqian Liang, and Jun Liu. Few-shot 3d point cloud segmentation via relation consistency-guided heterogeneous prototypes. *IEEE Transactions on Multimedia*, 27:3158–3170, 2025. doi: 10.1109/TMM.2025.3557699.

[40] Qianxiong Xu, Xuanyi Liu, Lanyun Zhu, Guosheng Lin, Cheng Long, Ziyue Li, and Rui Zhao. Hybrid mamba for few-shot segmentation. *Advances in Neural Information Processing Systems*, 37:73858–73883, 2024.

[41] Qianxiong Xu, Lanyun Zhu, Xuanyi Liu, Guosheng Lin, Cheng Long, Ziyue Li, and Rui Zhao. Unlocking the power of sam 2 for few-shot segmentation. In *International Conference on Machine Learning*, pp. 69950–69966. PMLR, 2025.

[42] Anqi Zhang, Guangyu Gao, Jianbo Jiao, Chi H Liu, and Yunchao Wei. Bridge the points: Graph-based few-shot segment anything semantically. *Advances in Neural Information Processing Systems*, 37:33232–33261, 2024.

[43] Yumin Zhang, Hongliu Li, Yajun Gao, Haoran Duan, Yawen Huang, and Yefeng Zheng. Prototype correlation matching and class-relation reasoning for few-shot medical image segmentation. *IEEE Transactions on Medical Imaging*, 43(11):4041–4054, 2024.

[44] Xiaofei Zhou, Jia Lin, Dongmei Chen, Deyang Liu, Jiyong Zhang, and Runmin Cong. Scale-invariant feature matching network for vdt few-shot semantic segmentation. *IEEE Transactions on Image Processing*, 2026.

[45] Yazhou Zhu and Haofeng Zhang. Maup: Training-free multi-center adaptive uncertainty-aware prompting for cross-domain few-shot medical image segmentation. In *International Conference on Medical Image Computing and Computer-Assisted Intervention*, pp. 326–336. Springer, 2025.

[46] Yazhou Zhu, Shidong Wang, Tong Xin, and Haofeng Zhang. Few-shot medical image segmentation via a region-enhanced prototypical transformer. In *International conference on medical image computing and computer-assisted intervention*, pp. 271–280. Springer, 2023.

[47] Yazhou Zhu, Shidong Wang, Tong Xin, Zheng Zhang, and Haofeng Zhang. Partition-a-medical-image: Extracting multiple representative subregions for few-shot medical image segmentation. *IEEE Transactions on Instrumentation and Measurement*, 73:1–12, 2024.

[48] Yazhou Zhu, Shidong Wang, Tao Zhou, Zechao Li, Haofeng Zhang, and Ling Shao. Cross-domain few-shot medical image segmentation via dynamic semantic matching. *IEEE Transactions on Image Processing*, 2025.

[49] Xiahai Zhuang. Multivariate mixture model for myocardial segmentation combining multi-source images. *IEEE transactions on pattern analysis and machine intelligence*, 41(12): 2933–2946, 2018.

A Details of Our Method

A.1 Theoretical Perspective of EFDM

We provide a theoretical perspective on why EFDM enriches support foreground features to provide more informative guidance for query segmentation. Let P_{s} and P_{q} denote the distributions of support and query foreground features, respectively. Rather than relying entirely on the original support features, EFDM introduces a family of spatial transformations $\{\mathcal{T}_D\}$ parameterized by a random flow field $D \sim \mathcal{U}(-\alpha, \alpha)$. Each transformation is defined as:

$$\mathcal{T}_D(\boldsymbol{F}_{\mathrm{s}}) = \boldsymbol{F}_{\mathrm{s}} \circ (G + D_{\mathcal{B}}),$$

where G denotes the basic grid coordinates. Consequently, the support features are expanded from a single discrete representation into a semantically rich distribution induced by $\mathcal{T}_D(\boldsymbol{F}_{\mathrm{s}})$, denoted as $P_{\mathrm{s}}^{\mathrm{aug}}$. The corresponding augmented support prototype can be formulated as the expectation of this distribution: $\boldsymbol{p}_{\mathrm{s}}^{\mathrm{aug}} = \mathbb{E}_D[\mathcal{T}_D(\boldsymbol{F}_{\mathrm{s}})]$.

Proposition 1 (Information Enrichment via Deformation Augmentation). Assuming that the spatial variations of query features can be approximated by a set of local deformations, *i.e.*, for $\boldsymbol{F}_{\mathrm{q}} \sim P_{\mathrm{q}}$, there exists an implicit deformation field D_{q} such that $\boldsymbol{F}_{\mathrm{q}} \approx \mathcal{T}_{D_{\mathrm{q}}}(\boldsymbol{F}_{\mathrm{s}})$. EFDM effectively expands the information of the support set, ensuring that compared to the original support features, the augmented distribution covers the valid spatial variations of the query set with a strictly higher probability:

$$\mathcal{P}(\boldsymbol{F}_{\mathrm{q}} \in \mathrm{supp}(P_{\mathrm{s}}^{\mathrm{aug}})) \geq \mathcal{P}(\boldsymbol{F}_{\mathrm{q}} \in \mathrm{supp}(P_{\mathrm{s}})).$$

Theoretical Explanation. EFDM primarily enhances the richness of the support guidance through the following three mechanisms:

(1) Expansion of the Mathematical Support. Without augmentation, the mathematical support of P_{s} is confined to a minimal region within the original feature space, which typically provides inadequate guidance. Since $D \sim \mathcal{U}(-\alpha, \alpha)$ and the bicubic interpolation guarantees the spatial smoothness of the flow field, EFDM maps the features into a family of continuous and locally smooth deformations, yielding:

$$\mathrm{supp}(P_{\mathrm{s}}) \subset \mathrm{supp}(P_{\mathrm{s}}^{\mathrm{aug}}),$$

This expansion directly increases the information capacity of the feature space, providing a richer semantic subspace for the query set to reference.

(2) Informative Manifold Covering. Since the transformation $\mathcal{T}_D$ is continuous with respect to D, as long as the implicit deformation D_{q} of the query feature falls within the predefined range $\mathcal{U}(-\alpha, \alpha)$, there always exists a sampled D such that:

$$\|\mathcal{T}_D(\boldsymbol{F}_{\mathrm{s}}) - \boldsymbol{F}_{\mathrm{q}}\| \leq \epsilon,$$

This indicates that EFDM generates samples that successfully capture the spatial variations of the query set with a non-zero probability in the probability space. Instead of enforcing a rigid alignment, it enriches the support manifold to ensure that adequate guidance information is consistently provided.

(3) Robust Prototype Aggregation. In the prototype-guided paradigm, let $\boldsymbol{p}_{\mathrm{s}}$ and $\boldsymbol{p}_{\mathrm{s}}^{\mathrm{aug}}$ denote the original and augmented feature prototypes, respectively. The augmented prototype performs an expected integration over local deformations:

$$\boldsymbol{p}_{\mathrm{s}}^{\mathrm{aug}} = \mathbb{E}_D[\mathcal{T}_D(\boldsymbol{F}_{\mathrm{s}})],$$

This operation is equivalent to applying a low-pass filter to spatial variations at the feature level. It smooths out high-frequency structural noise caused by rigid spatial constraints while preserving semantically coherent foreground details. Compared to the unaugmented prototype $\boldsymbol{p}_{\mathrm{s}}$, the integrated $\boldsymbol{p}_{\mathrm{s}}^{\mathrm{aug}}$ incorporates higher spatial diversity and structural robustness, acting as a highly informative anchor for activating the query set.

Discussion. EFDM does not demand a strict deterministic alignment between the support and query sets. Instead, through random flow field sampling based on a uniform distribution, it multiplicatively enriches the foreground representations. This strategy of exchanging rigid constraints for information diversity significantly broadens the semantic coverage of the support features, effectively alleviating the scarcity of guidance information caused by severe spatial deformations in few-shot segmentation.

A.2 EFDM ALGORITHM PROCESS

Algorithm 1 Elastic Flow-field Deformation Module (EFDM)

Input: Feature $\boldsymbol{F}_{\mathrm{s}} \in \mathbb{R}^{B\times C\times H_{\mathrm{s}}\times W_{\mathrm{s}}}$, Mask $\boldsymbol{M}_{\mathrm{s}} \in \mathbb{R}^{B\times C\times H\times W}$, deformation count n, number of blocks N_b, offset s
Output: deformed feature $\mathbf{F}_{\mathrm{s}}^{n}$, deformed masks $\mathbf{M}_{\mathrm{s}}^{n}$
$\boldsymbol{M}_{\mathrm{s}} \leftarrow \text{Interpolate}(\boldsymbol{M}_{\mathrm{s}}, \text{size} = (H, W), \text{mode} = \text{'nearest'})$
$\boldsymbol{F}_{n\mathrm{s}} \leftarrow \text{Repeat}(\boldsymbol{F}_{\mathrm{s}}, \text{repeats} = n)$
$\boldsymbol{M}_{n\mathrm{s}} \leftarrow \text{Repeat}(\boldsymbol{M}_{\mathrm{s}}, \text{repeats} = n)$
Basic grid coordinates $G^{n}(u, v) \in [-1, 1]^{B\times H_{\mathrm{s}}\times W_{\mathrm{s}}\times 2}$
Block-wise control filed $D^{n} \sim \mathcal{U}(-\alpha, \alpha)$ of shape $[B, N, N, 2]$
$D_{\mathcal{B}}^{n} \leftarrow \text{Interpolate}(D^{n}, \text{size} = (H, W), \text{mode} = \text{'bicubic'})$
deformed grid coordinates: $G_{def}^{n}(u, v) \leftarrow D_{\mathcal{B}}^{n} + G^{n}(u, v)$
$\boldsymbol{F}_{\mathrm{s}}^{n} \leftarrow \text{GridSample}(\boldsymbol{F}_{n\mathrm{s}}, G_{def}^{n}(u, v), \text{mode} = \text{'bilinear'})$
$\boldsymbol{M}_{\mathrm{s}}^{n} \leftarrow \text{GridSample}(\boldsymbol{M}_{n\mathrm{s}}, G_{def}^{n}(u, v), \text{mode} = \text{'nearest'})$

B DETAILS OF DATASET AND EXPERIMENT

B.1 DATASET INTRODUCTION

To fully validate generalization ability of proposed model in medical image analysis across multiple modalities, we select three publicly available medical image datasets for our experiments:

Abd-MRI: This abdominal magnetic resonance imaging (MRI) image is derived from ISBI 2019 Healthy Abdominal Organ Segmentation Challenge [16]. During model evaluation phase, we primarily extract and analyze four target organs: liver, spleen, and left and right kidneys (LK/RK).

Abd-CT: Abdominal CT data is provided by MICCAI 2015 Multi-Atlas Abdominal Annotation Competition [18]. To ensure consistency across modalities, the four organ categories tested in this dataset are identical to those in Abd-MRI dataset.

Card-MRI: This cardiac MRI dataset is extracted from MICCAI 2019 Multi-Sequence Cardiac MRI Segmentation Challenge [49]. The evaluated anatomical structures cover three core categories: left ventricular blood pool (LV-BP), left ventricular myocardium (LV-MYO), and right ventricle (RV).

B.2 EXPERIMENTAL SETTINGS

To ensure fairness and effectiveness of model evaluation, we adopt experimental specifications proposed in SSL-ALPNet, which is also currently accepted evaluation standard for mainstream few-shot medical image segmentation (FSMIS) work. The specific experimental settings are divided into the following two types:

Setting1: Allows test category targets to exist in image slices of training set as background, but does not apply any ground truth labels to these categories.

Setting2: Performs a more stringent filtering, that is, directly deletes all samples containing test category in background during training phase, thereby ensuring that test category is absolutely *unseen* for model. However, when processing Card-MRI data, since all test categories will inevitably appear in same 2D slice simultaneously, setting2 cannot be applied to Card-MRI dataset.

C ANALYSIS OF DETAIL ABLATION EXPERIMENT

C.1 EFFECTIVENESS ANALYSIS OF DIFFERENT PROMPTS

We also analyzed how different types of prompts affect model performance. As shown in Table 9, *Foreg-pt*, *Backg-pt*, and *box* represent foreground prompt points, background prompt points, and bounding box prompts, respectively. We found that model performance was optimal only when both *Foreg-pt* and *box* were present, compared to other configurations.

Table 9: **Ablation of different prompts** under setting1 on Abd-MRI and Abd-CT datasets.

Foreg-pt	Backg-pt	box	Abd-MRI					Abd-CT				
			LK	RK	Spleen	Liver	Mean	LK	RK	Spleen	Liver	Mean
✓			78.87	86.97	75.21	55.76	74.20	83.92	79.43	73.66	75.78	78.20
	✓		27.90	22.47	19.66	34.56	26.15	30.49	23.35	22.35	28.53	26.18
		✓	85.37	91.16	77.81	81.48	83.96	84.58	83.65	84.90	83.69	84.20
✓	✓		79.51	87.82	77.91	63.38	77.16	84.82	82.74	81.95	81.92	82.86
✓		✓	**87.78**	**92.40**	**78.93**	**83.62**	**85.68**	86.85	**85.77**	**86.88**	**85.90**	**86.35**
	✓	✓	86.95	92.27	78.82	82.95	85.25	**87.13**	85.88	86.58	85.53	86.28
✓	✓	✓	87.31	92.31	78.81	82.72	85.28	86.88	85.70	86.77	85.83	86.29

C.2 Performance Comparison Across Different Category Complexities

We investigated boundary complexities across various organ categories and evaluated the segmentation performance of some methods. Table 10 demonstrates that, despite variations in patients or imaging modalities (e.g., abdominal MRI and CT), the liver and LV-MYO consistently exhibit the most intricate boundaries. Conversely, structures such as the left kidney, right kidney, and spleen feature relatively regular contours. To measure this quantitatively, we apply the formula $C = \frac{P^2}{4\pi A}$. In this equation, A represents the total foreground pixels for a given category, while P denotes the perimeter of the outline; naturally, a larger C value reflects greater morphological complexity. Notably, our proposed method achieves the best performance among SAM-based frameworks on the liver. This superior performance is primarily attributed to the EFDM, which effectively models the deformation gap through block-wise control points and dense offset fields, resulting in more significant improvements.

Table 10: Performance comparison (mean Dice across all configurations) and boundary complexity metrics evaluated on the Abd-MRI, Card-MRI, and Abd-CT datasets.

Dataset	Organ	A	P	C	Rank	ProtoSAM	RPT	FoB+SAM	**Ours**
Abd-MRI	Liver	5937.7	347.9	**1.894**	**1**	82.54	79.62	81.93	**82.67**
	Spleen	1734.8	189.4	1.844	2	77.25	75.92	**77.80**	77.61
	LK	1041.5	125.0	1.314	3	77.11	79.53	81.03	**86.74**
	RK	1096.4	129.1	1.311	4	81.90	87.92	88.88	**91.71**
Card-MRI	LV-MYO	1001.5	292.2	**7.142**	**1**	–	66.91	–	**67.90**
	RV	1313.4	169.6	2.001	2	–	80.78	–	**86.87**
	LV-BP	1149.0	124.3	1.137	3	–	**89.90**	–	89.37
Abd-CT	Liver	5054.5	353.3	**2.290**	**1**	86.32	78.91	84.70	**86.38**
	RK	798.9	133.0	1.962	2	73.36	73.43	**85.81**	84.45
	LK	798.3	132.6	1.940	3	70.38	75.02	**87.23**	87.15
	Spleen	1692.1	178.3	1.784	4	67.87	71.69	83.39	**86.72**

C.3 Analysis of Support Features Semantic Fidelity.

We analyze whether the semantic fidelity of supporting features is affected after processing by EFDM. As shown in Table 11, experiments using Setting 1 are conducted on Abd-MRI dataset. The results indicate that after EFDM deformation processing, average Dice score of the supporting features increases from baseline value of 86.24 to 88.39. This directly demonstrates that elastic deformation bridges spatial gaps without compromising semantic integrity.

Table 11: Analysis of Support Features Semantic Fidelity on Abd-MRI Setting1.

	Support	Deformation Support
Spleen	**89.09**	82.94
RK	87.10	**90.60**
LK	85.33	**89.80**
Liver	83.44	**90.21**
Mean Dice	86.24	**88.39**

C.4 Validity Analysis of Different Deformation Methods.

We demonstrate the advantages of our method by comparing it with several common deformation-enhanced techniques: Standard Elastic Augmentation (SEA), Deformable Feature Transformation (DFT), and Registration-based Deformation (RID). SEA and DFT are limited to image-level data augmentation and local feature encoding, respectively; neither can directly generate diverse supporting representations to address the issue of missing prototypes in few-shot scenarios. Meanwhile, RID strategies heavily rely on explicit correspondences, which are computationally expensive and susceptible to errors, making it difficult to effectively handle complex intra-class anatomical variations.

To overcome these limitations, EFDM spontaneously constructs a support representation space rich in deformation diversity within the feature space without requiring additional parameters or enforcing a fixed support-query relationship, thereby guiding the model to learn more robust anatomical prior information through a generative approach.

Table 12: Effectiveness Analysis of Different Deformation Methods on Abd-MRI.

Methods	LK	RK	Spleen	Liver	Mean
SEA	88.06	90.78	78.80	81.88	84.88
DFT	87.64	90.22	**79.50**	**84.13**	85.37
RID	84.35	92.03	77.81	83.41	84.40
Ours	**87.78**	**92.40**	78.93	83.62	**85.68**

C.5 L2 Distance Analysis of Support-Query.

To clarify that EFDM can reduce distance between support and query features, we have included an analysis of distances between support and query features before and after EFDM processing. Specifically, we measured their L2 distances, as detailed in Table 13, and found that after EFDM processing, the distances between support and query features decreased across all three datasets.

Table 13: L2 Distance Analysis of support-query on three datasets.

Dataset	L2 before EFDM	L2 after EFDM	**ΔL2**
Abd-CT	11.75	11.34	**-0.42**
Abd-MRI	9.20	9.12	**-0.08**
Card-MRI	8.81	7.86	**-0.96**

C.6 Detailed Qualitative Analysis

We analyze visualization results compared to other methods. As shown in Fig. 9, on Card-MRI dataset, across three classes RV, LV-MYO, and LV-BP, our method provides more complete predictions and a lower probability of mistaking background for foreground. As shown in Fig. 10, on Abd-MRI and Abd-CT datasets, across four classes LK, RK, liver, and splee, our method exhibits a lower segmentation error rate compared to other methods. In summary, our approach exhibits significant superiority when compared to other state-of-the-art methods.

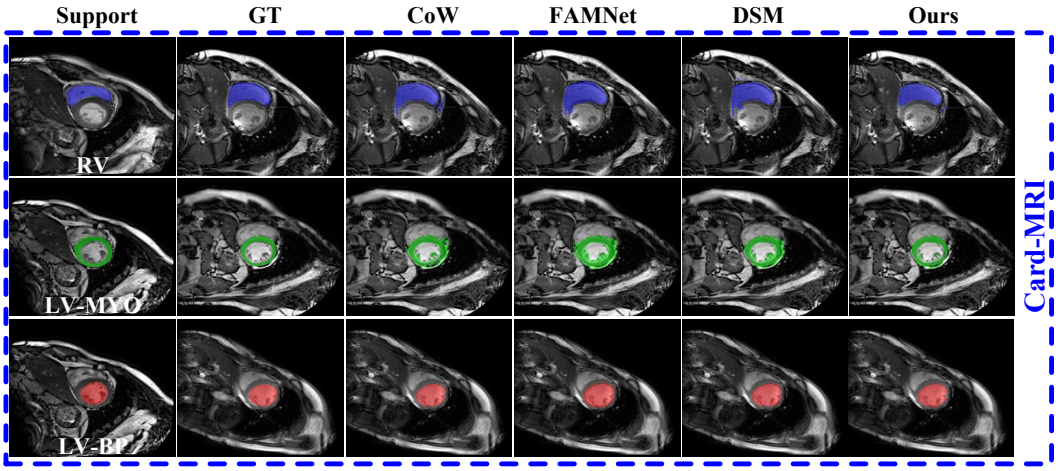


Figure 9: **Qualitative comparison of different approaches on Card-MRI datasets.**

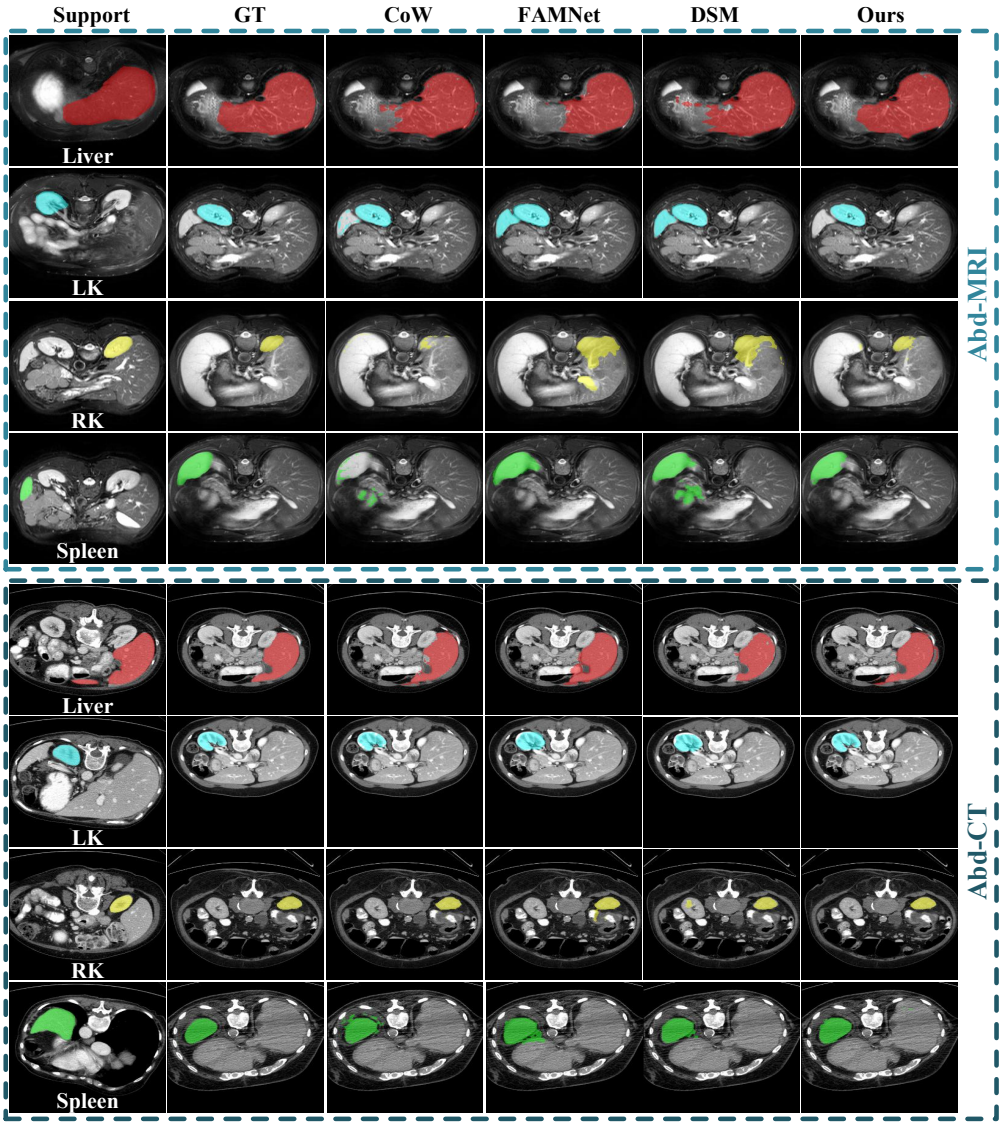


Figure 10: **Qualitative comparison of different approaches on Abd-MRI and Abd-CT datasets.**

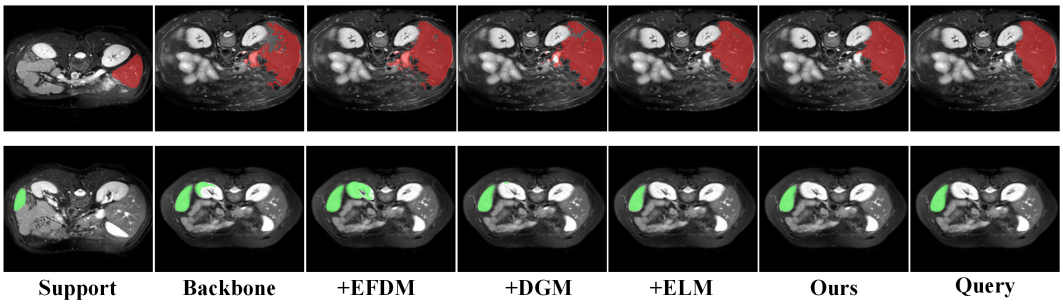
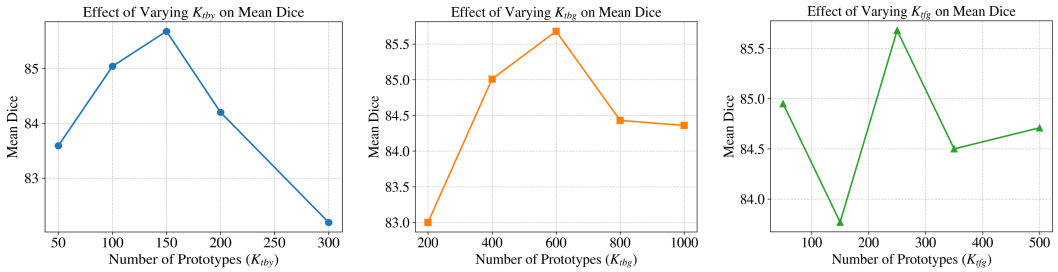


Figure 11: **Qualitative comparison of each component of EFG-SAM on Abd-MRI dataset.**



(a) K_{tfg}=250, K_{tbg}=600 (b) K_{tfg}=250, K_{tby}=150 (c) K_{tbg}=600, K_{tby}=150

Figure 12: Line chart showing effect of prototypes number on model performance.

C.7 Qualitative Analysis of Each Component of EFG-SAM

As illustrated in Fig. 11 for the Abd-MRI dataset, it is obvious that the model's predictions are highly incomplete and even contain missegmentations when only backbone is utilized. However, with the incremental integration of our proposed components including EFDM, DGM, ELM, and EAM, the predictive performance improves significantly compared to backbone alone. The model achieves its optimal performance only when all components are incorporated.

C.8 Detailed Training Loss

We jointly train our proposed model using multi-loss strategy. Specifically, we compute the Negative Log-Likelihood (NLL) loss between support-guided query segmentation prediction and query ground truth to obtain L_{query}. Similarly, the NLL loss between query-guided support segmentation prediction and support ground truth is calculated to derive L_{support}. Additionally, a Boundary F1 loss is applied between query segmentation result and corresponding boundary mask, denoted as $L_{\text{query_by}}$. The total loss, L_{total}, is formulated as sum of these three components, with a ratio of 1:1:1.

$$L_{\text{total}} = L_{\text{query}} + L_{\text{support}} + L_{\text{query_by}}. \tag{7}$$

C.9 Visualization of Prototypes Number K

As shown in Fig. 12, we plot line graphs showing how changes in number of different prototypes affect model performance. We can clearly see that when K_{tfg} = 250, K_{tbg} = 600, and K_{tby} = 150, model performs well; when K_{tfg} = 250 and K_{tby} = 150, model performs well with K_{tbg} = 600; and when K_{tbg} = 600 and K_{tby} = 150, the model performs well with K_{tfg} = 250. Therefore, when K_{tfg} = 250, K_{tbg} = 600, and K_{tby} = 150, model performance is optimal.