GEOMETRY-AWARE AND PERIPHERAL-BACKGROUND SYNERGY FOR CROSS-DOMAIN FEW-SHOT MEDICAL IMAGE SEGMENTATION

Anonymous authors
Paper under double-blind review

ABSTRACT

Few-shot medical image segmentation (FSMIS) aims to generalize to novel classes with minimal annotations, yet its effectiveness is severely hindered by the challenge of domain shift, where substantial inter-domain discrepancies (e.g., MRI to CT) often lead to significant performance degradation. Current cross-domain FSMIS methods primarily rely on volatile appearance-based features or computationally expensive graph-matching, failing to capture stable anatomical representations. In this paper, we observe a crucial phenomenon: while domain-specific textures vary drastically, the geometric topology of organs and their surrounding peripheral-background distributions remain remarkably consistent across divergent domains. Based on this insight, we propose a novel framework termed Geometry-aware and Peripheral-background Synergy (**GPS**) for cross-domain FSMIS. GPS leverages these stable invariants as "Anatomical Anchors" to steer the model beyond domain-sensitive biases. Specifically, GPS consists of three synergistic modules: (1) a Geometry-Aware Injector module (GAI) that modulates feature responses based on geometric shape priors to distill domain-invariant representations; (2) a Peripheral-aware Prototype Mining (PPM) module that constructs multi-prototypes for both the target organ and its surrounding tissues; and (3) a Prototype Synergy Refinement (PSR) module that exploits foreground-background complementarity for fine-grained segmentation. Extensive experiments on three popular medical image benchmarks demonstrate that our GPS significantly outperforms state-of-the-art methods, exhibiting superior generalization across diverse domains. Code is available at `https://anonymous.4open.science/r/GPS-DDCD`.

1 INTRODUCTION

Automated medical image segmentation is vital in computer-aided diagnosis (Tsochatzidis et al., 2021) and treatment (Chen et al., 2021), but its performance relies heavily on large-scale, high-quality annotated data which is challenging in medical images due to the high cost of expert annotation, the time-consuming process, and privacy constraints. Therefore, Few-Shot Medical Image Segmentation (FSMIS) (Xu et al., 2026; Bo et al., 2026) has emerged to alleviate this data scarcity, aiming to achieve rapid generalization to novel classes using only a few annotated samples through paradigms such as meta-learning (Dhir et al., 2025). However, in practical clinical deployment, FSMIS faces a challenge known as domain shift (Guo et al., 2022). Specifically, when a model is trained on MRI data but deployed for inference on CT images, the generalization ability (Huo et al., 2025) degrades severely due to the significant distribution discrepancy in image appearance.

To tackle severe domain shifts, research has pivoted toward Cross-Domain Few-Shot Medical Image Segmentation (CD-FSMIS). Robust-EMD (Zhu et al., 2025a) employs Earth Mover's Distance (Pele & Werman, 2009) for local matching; however, its reliance on volatile appearance features makes it prone to failure under drastic appearance changes (e.g., MRI to CT). In the frequency domain, FAMNet (Bo et al., 2025b) swaps amplitude spectra for style alignment, yet this "amplitude-as-style" assumption often fails across complex medical modalities, confusing invariant anatomical details with noise. In the semantic space, DSM (Zhu et al., 2025b) introduces dynamic channel weighting to filter robust components. Nevertheless, lacking explicit geometric constraints, DSM

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.

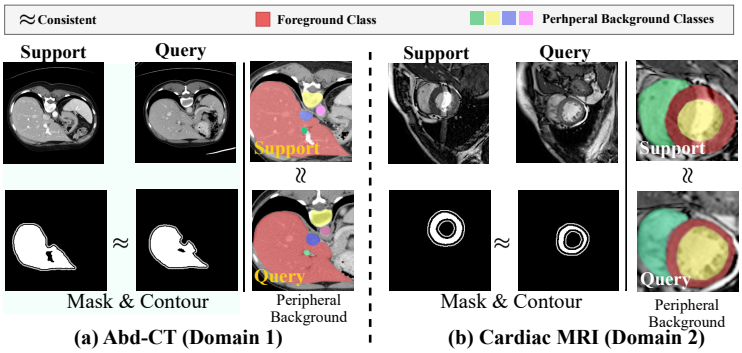

Figure 1: Illustration of the motivation. Despite significant appearance variations across imaging domains (e.g., Abd-CT and Cardiac-LGE MRI), both the geometric topology of target organs and the distribution of surrounding tissues remain highly consistent between support and query images, serving as reliable cross-domain anatomical invariants.

remains vulnerable to unseen artifacts and extreme texture shifts. Although C-Graph (Bo et al., 2025a) introduces a structural graph-matching paradigm via GCN (Lee et al., 2025), its reliance on dense pixel-to-node mapping leads to high computational overhead. Such high complexity is ill-suited for data-limited few-shot tasks and hinders efficient clinical inference.

Notably, we observe that even in cross-domain scenarios (e.g., CT to MRI) involving unseen classes, there exists strong anatomical consistency between the support and query sets. Specifically, both the geometric topology of the target organ and the distribution of its surrounding background remain highly stable across domains, as illustrated in Fig. 1. This key observation motivates us to leverage these "cross-domain invariants" as anatomical anchors, guiding the model to transcend volatile texture biases and instead capture intrinsic, domain-agnostic representations.

In this paper, we proposed a novel framework termed Geometry-aware and Peripheral-background Synergy (**GPS**) for CD-FSMIS. The GPS leverages geometric consistency and peripheral-background distribution consistency as anatomical anchors to alleviate cross-domain generalization challenge (Yanzhen et al., 2024). Specifically, our GPS consists of three complementary modules: (1) the Geometry-Aware Injector (**GAI**); (2) the Peripheral-aware Prototype Mining (**PPM**); and (3) the Prototype Synergy Refinement (**PSR**). First, the GAI module encodes the shape prior from the support set and adaptively enhances geometry-consistent channel responses while suppressing irrelevant ones, thereby learning domain-invariant representations. Subsequently, the PPM module builds multi-prototype representations of the foreground while mining background prototypes from surrounding tissues, improving the model's ability to capture complex anatomy categories. Finally, the PSR module exploits the complementarity between foreground and peripheral-background prototypes to refine segmentation predictions. In summary, our contributions are as follows:

- We observe that anatomical geometry and peripheral-background distributions exhibit strong cross-domain invariance, which can serve as reliable anchors for improving generalization in CD-FSMIS task.
- We propose a unified framework (**GPS**) consisting of three modules: **GAI** for geometry-aware feature modulation, **PPM** for multi-prototype construction of foreground and peripheral background, and **PSR** for segmentation refinement via multi-prototypes synergy.
- Extensive experiments conducted on four popular medical image datasets reveal that our proposed method outperforms other existing state-of-the-art (SOTA) methods and exhibits significant performance.

2 Related Works

2.1 Few-Shot Medical Image Segmentation

FSMIS (Tang et al., 2025; Zhao et al., 2025) tackles medical data scarcity in medical imaging, enabling models to rapidly generalize to novel classes using only a few labeled samples. Existing FSMIS methods (Zhu et al., 2023; Cheng et al., 2025) are predominantly based on prototype metric

learning. PANet (Wang et al., 2019) employs Masked Average Pooling to obtain a single prototype for the target class; SSL-ALP (Ouyang et al., 2022) introduces superpixel mechanisms to improve model generalization; QNet (Shen et al., 2023) dynamically adapts target representations through a dual-path architecture; GMRD (Cheng et al., 2024) enhances the representation of complex anatomical structures through a multi-prototype strategy; Cow (Jiang & Zhang, 2025) deeply explores hard prototypes from uncertain regions; BePMI (Xu et al., 2026) strengthens boundary awareness and further performs slice-level alignment; and Fob-SAM (Bo et al., 2026) improves segmentation performance by incorporating negative points together with SAM (Kirillov et al., 2023).

Although these methods achieve promising performance in single-domain scenarios (Su et al., 2023), they largely rely on domain-specific appearance features, leading to severe performance degradation under cross-domain transfer.

2.2 Cross-domain FSMIS

CD-FSMIS (Zhu & Zhang, 2025; Zhou et al., 2025; Shi et al., 2025) has recently attracted increasing attention for addressing severe domain shifts across imaging modalities. Existing CD-FSMIS methods mainly focus on appearance alignment in different representation spaces. Robust-EMD (Zhu et al., 2025a) performs local feature matching via Earth Mover's Distance (Rubner et al., 2000), while FAMNet (Bo et al., 2025b) aligns domains through frequency-spectrum adaptation. DSM (Zhu et al., 2025b) further introduces dynamic channel weighting to enhance robust semantic representations. In addition, C-Graph (Bo et al., 2025a) explores structural correspondence using graph matching. However, existing methods fail to explicitly model cross-domain geometric-topology consistency. In addition, they largely overlook the consistency of surrounding tissues and organ distributions around the target class between support and query images across domains, which is crucial for achieving robust cross-domain generalization.

3 Methodology

3.1 Problem Definition

In the CD-FSMIS tasks, the goal is to train a segmentation model Θ on a source domain dataset $\mathcal{D}^{\mathrm{s}}$, which contains a set of base categories $\mathcal{C}_{\mathrm{base}}$, and enable it to generalize to an unseen target domain $\mathcal{D}^{\mathrm{t}}$ for segmenting novel categories $\mathcal{C}_{\mathrm{target}}$. The two label spaces are disjoint, (i.e., $\mathcal{C}_{\mathrm{base}} \cap \mathcal{C}_{\mathrm{target}} = \emptyset$), and there exists a substantial domain shift between $\mathcal{D}^{\mathrm{s}}$ and $\mathcal{D}^{\mathrm{t}}$. To formulate the learning process, we follow an episodic meta-learning paradigm and construct tasks in the form of $(\mathcal{S}, \mathcal{Q})$. For each N-way K-shot episode, N classes are randomly selected from the dataset. During training, all episodes are drawn exclusively from the source domain $\mathcal{D}^{s}$. Each support set $\mathcal{S} = \{(\mathbf{I}_{\mathrm{s}}^{j}, \mathbf{M}_{\mathrm{s}}^{j})\}_{j=1}^{N\times K}$ consists of K annotated examples per class, while the corresponding query set $\mathcal{Q} = \{(\mathbf{I}_{\mathrm{q}}^{i}, \mathbf{M}_{\mathrm{q}}^{i})\}_{i=1}^{N_q}$ includes unlabeled samples that need to be segmented. Here, $\mathbf{I} \in \mathbb{R}^{3\times H\times W}$ denotes an input medical image and $\mathbf{M} \in \{0, 1, \ldots, N\}^{H\times W}$ represents the pixel-wise segmentation annotation over the sampled categories and background. At inference time, the learned model Θ is evaluated on the target domain $\mathcal{D}^{\mathrm{t}}$ using the same episodic sampling strategy, allowing it to adapt and generalize to unseen classes under domain shift.

3.2 Network Overview

Initially, the support and query images, $\mathbf{I}_{\mathrm{s}}, \mathbf{I}_{\mathrm{q}} \in \mathbb{R}^{3\times H\times W}$, are processed by a shared-weight feature extractor $f_\theta(\cdot)$. Following common practice in FSMIS, we adopt the first three residual stages of a ResNet-50 (Lin et al., 2014b) backbone, pre-trained on MS-COCO (Lin et al., 2014a), to serve as our encoder. This yields the deep feature representations $\mathbf{F}_{\mathrm{s}} = f_\theta(\mathbf{I}_{\mathrm{s}})$ and $\mathbf{F}_{\mathrm{q}} = f_\theta(\mathbf{I}_{\mathrm{q}})$, where $\mathbf{F}_{\mathrm{s}}, \mathbf{F}_{\mathrm{q}} \in \mathbb{R}^{C\times H\times W}$. Here, C, H, and W denote the channel depth, spatial height, and width, respectively. Subsequently, guided by the GAI module, we distill domain-agnostic support and query features, denoted as $\mathbf{F}_{\mathrm{s}}^{'}$ and $\mathbf{F}_{\mathrm{q}}^{'}$, where $\mathbf{F}_{\mathrm{s}}^{'}, \mathbf{F}_{\mathrm{q}}^{'} \in \mathbb{R}^{C\times H\times W}$. These rectified features are then processed by the PPM module to harvest a set of discriminative local prototypes, $\mathbf{P}_{\mathrm{local}}^{\mathrm{fg}} \in \mathbb{R}^{N_{\mathrm{fg}}\times C}$ for the foreground of target class and $\mathbf{P}_{\mathrm{local}}^{\mathrm{bg}} \in \mathbb{R}^{N_{\mathrm{bg}}\times C}$ for the peripheral background. Finally,

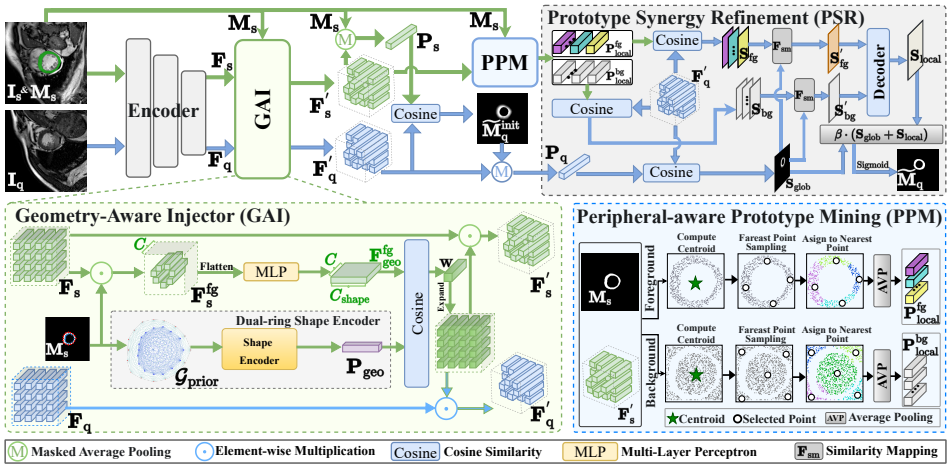

Figure 2: Overview of our proposed GPS method: (1) the GAI module for domain-agnostic feature extraction via geometric prior injection; (2) the PPM module for multi-prototype construction of foreground and peripheral background; (3) the PSR module for synergizing foreground and peripheral-background prototypes to refine segmentation.

the PSR module leverages these bipartite prototypes to execute fine-grained refinement, synergizing both global semantics with local cues for the final prediction $\widetilde{\mathbf{M}}_{\mathrm{q}} \in \mathbb{R}^{H \times W}$.

3.3 Geometry-Aware Injector (GAI)

GAI extracts domain-robust geometric priors from the support mask and uses them to recalibrate support and query features. It consists of a **Dual-ring Shape Encoder (DSE)** for geometry encoding and a **Geometric Feature Enhancement (GFE)** module for feature modulation.

Dual-ring Shape Encoder (DSE). Given the support mask $\mathbf{M}_s$, we compute its foreground centroid C_{cent} and cast N_{dir} uniformly distributed rays. For each direction i, the first and last foreground intersections are taken as the inner- and outer-ring points, $\mathbf{C}_i^{\mathrm{in}}, \mathbf{C}_i^{\mathrm{out}} \in \mathbb{R}^2$:

$$\mathbf{C}^{\mathrm{in/out}} = [\mathbf{C}_1^{\mathrm{in/out}}, \ldots, \mathbf{C}_{N_{\mathrm{dir}}}^{\mathrm{in/out}}]^\top \in \mathbb{R}^{N_{\mathrm{dir}} \times 2}. \tag{1}$$

The dual-ring design naturally captures both hollow and solid anatomical structures. We construct translation-invariant geometric matrices using pairwise offsets. In addition, to remove scale variation, both matrices are normalized by the maximum pairwise distance $d_{\max}$ of the outer ring. Finally, the inner- and outer-ring geometric matrices $\mathbf{A}_{\mathrm{in}}, \mathbf{A}_{\mathrm{out}} \in \mathbb{R}^{N_{\mathrm{dir}} \times N_{\mathrm{dir}} \times 2}$ are formulated as:

$$\mathbf{A}_{\mathrm{in/out}} = \left[\mathbf{C}_i^{\mathrm{in/out}} - \mathbf{C}_j^{\mathrm{in/out}}\right]_{i,j=1}^{N_{\mathrm{dir}}}, \quad \mathbf{A}_{\mathrm{in/out}} = \mathbf{A}_{\mathrm{in/out}} / d_{\max}. \tag{2}$$

The normalized inner- and outer-ring geometry $\mathbf{F}_{\mathrm{geo}} \in \mathbb{R}^{N_{\mathrm{dir}} \times C_{\mathrm{shape}}}$ is embedded by two MLPs:

$$\mathbf{F}_{\mathrm{in/out}}^{\mathrm{shape}} = \left[\mathrm{MLP}_{\theta_1}\left(\mathrm{vec}(\mathbf{A}_{\mathrm{in/out}}^{(i)})\right)\right]_{i=1}^{N_{\mathrm{dir}}} \in \mathbb{R}^{N_{\mathrm{dir}} \times C_{\mathrm{shape}}/2}, \quad \mathbf{F}_{\mathrm{geo}} = \mathrm{MLP}_{\theta_2}\left(\mathbf{F}_{\mathrm{out}}^{\mathrm{shape}} \oplus \mathbf{F}_{\mathrm{in}}^{\mathrm{shape}}\right), \tag{3}$$

where $\mathrm{vec}(\cdot)$ denotes the vectorization operation that flattens the i-th pairwise offset matrix $\mathbf{A}_{\mathrm{in/out}}^{(i)}$ into a one-dimensional feature vector. $\mathrm{MLP}_{\theta 1}(\cdot)$ denotes a two-layer MLP mapping each flattened pairwise offset vector of dimension $2N_{\mathrm{dir}}$ to a $C_{\mathrm{shape}}/2$ dimensional feature. $\oplus$ denotes the concatenation along the last dimension. $\mathrm{MLP}_{\theta 2}(\cdot)$ is a two-layer MLP mapping the concatenated feature of dimension C_{shape} to C_{shape}.

The final geometry descriptor is obtained by directional averaging:

$$\mathbf{P}_{\mathrm{geo}} = \frac{1}{N_{\mathrm{dir}}} \sum_{i=1}^{N_{\mathrm{dir}}} \mathbf{F}_{\mathrm{geo}}^{(i)}. \tag{4}$$

A rotation-consistency loss $\mathcal{L}_{\text{rotate}}$ in the subsequent loss function section further encourages geometry embeddings from different directions to be aligned.

Geometric Feature Enhancement (GFE). Each channel of the support feature $\mathbf{F}_{\text{s}}$ is first projected into the same feature space as the geometric representation $\mathbf{P}_{\text{geo}}$, generating the foreground geometric feature $\mathbf{F}_{\text{geo}}^{\text{fg}} \in \mathbb{R}^{C \times C_{\text{shape}}}$:

$$\mathbf{F}_{\text{geo}}^{\text{fg}} = \text{MLP}(\text{Flatten}(\mathbf{F}_{\text{s}} \odot \mathbf{M}_{\text{s}})), \tag{5}$$

where $\odot$ denotes the operation of element-wise multiplication, $\text{Flatten}(\cdot)$ denotes the flatten operation along spatial dimensions, $\text{MLP}(\cdot)$ denotes a multi-layer perceptron projecting the spatial features of dimension hw to the shape descriptor dimension C_{shape}.

We then compute the similarity score $\mathbf{W} \in \mathbb{R}^{1 \times C}$ between each channel and the foreground geometric feature. A higher similarity score indicates that the corresponding channel is more capable of representing the current geometric feature. $\mathbf{W}$ is formulated as:

$$\mathbf{W} = \text{Cosine}(\mathbf{P}_{\text{geo}}, \mathbf{F}_{\text{geo}}^{\text{fg}}) \in \mathbb{R}^{1 \times C}, \tag{6}$$

where $\text{Cosine}(\cdot, \cdot)$ denotes computing cosine similarity along the last dimension.

Finally, the obtained weights are used to enhance both the support features and query features, the enhanced support and query features $\mathbf{F}_{\text{s}}, \mathbf{F}_{\text{q}} \in \mathbb{R}^{C \times H \times W}$ are formulated as:

$$\mathbf{F}_{k}^{'} = \mathbf{F}_{k} \odot \text{Extend}(\mathbf{W}), \quad k \in \{\text{s}, \text{q}\} \tag{7}$$

where $\text{Extend}(\cdot)$ denotes broadcasting along spatial dimensions to $\mathbb{R}^{C \times H \times W}$.

Before PPM and PSR, we first compute the support prototype $\mathbf{P}_{\text{s}} \in \mathbb{R}^{1 \times C}$ using support feature $\mathbf{F}_{\text{s}}^{'}$ and support mask $\mathbf{M}_{\text{s}}$. Then, the coarse query prediction $\widetilde{\mathbf{M}}_{\text{q}}^{\text{init}} \in \mathbb{R}^{H \times W}$ is obtained by computing the cosine similarity between query feature $\mathbf{F}_{\text{q}}^{'}$ and support global prototype $\mathbf{P}_{\text{s}}$:

$$\mathbf{P}_{\text{s}} = \frac{\sum_p \mathbf{F}_{\text{s}}^{'}(p) \cdot \mathbf{M}_{\text{s}}(p)}{\sum_p \mathbf{M}_{\text{s}}}, \quad \widetilde{\mathbf{M}}_{\text{q}}^{\text{init}}(p) = \sigma\big(\text{Cos}(\mathbf{F}_{\text{q}}^{'}(p), \mathbf{P}_{\text{s}}) \cdot \beta\big), \tag{8}$$

where p denotes a spatial location; $\text{Cos}(\cdot, \cdot)$ denotes the operation of computing cosine similarity; β represents a learnable parameter; and $\sigma(\cdot)$ is a Sigmoid function.

3.4 Peripheral-aware Prototype Mining (PPM)

We design PPM to simultaneously obtain multiple foreground prototypes of the target organ and multiple background prototypes from its surrounding regions.

First, we construct a box mask $\mathbf{M}_{\text{box}} \in \mathbb{R}^{H \times W}$, where the pixels inside the minimum bounding rectangle that completely covers the foreground region of $\mathbf{M}_{\text{s}}$ are set to 1, while all other positions are set to 0. Subsequently, we obtain the foreground mask $\mathbf{M}_{\text{fg}}$ and the surrounding background mask $\mathbf{M}_{\text{bg}}$. $\mathbf{M}_{\text{fg/bg}} \in \mathbb{R}^{H \times W}$ are formulated as:

$$\mathbf{M}_{\text{bg}} = (1 - \mathbf{M}_{\text{s}}) \odot \mathbf{M}_{\text{box}}, \quad \mathbf{M}_{\text{fg}} = \mathbf{M}_{\text{s}}, \tag{9}$$

where $\odot$ denotes element-wise multiplication.

Next, we denote the foreground pixel set and surrounding background pixel as $\mathcal{P}_{\text{fg/bg}}$ extracted from $\mathbf{M}_{\text{fg/bg}}$. The $\mathcal{P}_{\text{fg/bg}}$ are formulated as: $\mathcal{P}_{\text{fg/bg}} = \{(h, w) | \mathbf{M}_{\text{fg/bg}}^{(h,w)} = 1\}$. Then, we obtain the foreground centroid C_{fg} and the background centroid C_{bg}. $C_{\text{fg/bg}} \in \mathbb{R}^2$ is formulated as:

$$C_k = \frac{1}{|\mathcal{P}_k|} \sum_{p \in \mathcal{P}_k} p, \quad k \in \{\text{fg}, \text{bg}\}. \tag{10}$$

To extract representative local prototypes, we identify N_{fg} foreground reference centers $\{S_i^{\text{fg}}\}_{i=1}^{N_{\text{fg}}}$ and N_{bg} background reference centers $\{S_i^{\text{bg}}\}_{i=1}^{N_{\text{bg}}}$ from the pixel set $\mathcal{P}_{\text{fg/bg}}$. Unlike the conventional

Farthest Point Sampling (FPS) selecting the first point at random, we select the pixel closest to the centroid as the first point.

$$\begin{cases} S_1^* = \arg\min\limits_{p \in \mathcal{P}_*} \|p - C_*\|_2, & * \in \{\mathrm{fg}, \mathrm{bg}\} \\ S_i^* = \arg\max\limits_{p \in \mathcal{P}_*} \min\limits_{j<i} \|p - S_j^*\|_2, & i \in [2, N_*] \end{cases}. \tag{11}$$

Subsequently, each pixel in $\mathcal{P}_{\mathrm{fg}}$ (or $\mathcal{P}_{\mathrm{bg}}$) is assigned to its nearest center, partitioning the sets into N_{fg} and N_{bg} disjoint regions, denoted as $\{\Omega_i^{\mathrm{fg}}\}_{i=1}^{N_{fg}}$ and $\{\Omega_i^{\mathrm{bg}}\}_{i=1}^{N_{bg}}$, respectively.

$$\begin{cases} \Omega_i^{\mathrm{fg}} = \{p \in \mathcal{P}_{\mathrm{fg}} | \arg\min\limits_j \|p - S_j^{\mathrm{fg}}\|_2 = i\} \\ \Omega_i^{\mathrm{bg}} = \{p \in \mathcal{P}_{\mathrm{bg}} | \arg\min\limits_j \|p - S_j^{\mathrm{bg}}\|_2 = i\} \end{cases}. \tag{12}$$

Finally, given the support feature $\mathbf{F}_{\mathrm{s}}^{'} \in \mathbb{R}^{C \times H \times W}$, local foreground and background prototypes $\mathbf{P}_{\mathrm{local}}^{\mathrm{fg/bg}} \in \mathbb{R}^{N_{\mathrm{fg/bg}} \times C}$ are computed by average pooling within each region:

$$\begin{cases} \mathbf{P}_{\mathrm{local}}^{\mathrm{fg}} = \dfrac{1}{|\Omega_i^{\mathrm{fg}}|} \sum\limits_{p \in \Omega_i^{\mathrm{fg}}} \mathbf{F}_{\mathrm{s}}^{'}(p), & i = 1, \cdots, N_{\mathrm{fg}} \\ \mathbf{P}_{\mathrm{local}}^{\mathrm{bg}} = \dfrac{1}{|\Omega_i^{\mathrm{bg}}|} \sum\limits_{p \in \Omega_i^{\mathrm{bg}}} \mathbf{F}_{\mathrm{s}}^{'}(p), & i = 1, \cdots, N_{\mathrm{bg}} \end{cases}. \tag{13}$$

3.5 Prototype Synergy Refinement (PSR)

The distribution of tissues or organs around the target organ exhibits domain-invariance. Thus, PSR module utilizes the surrounding background prototypes $\mathbf{P}_{\mathrm{local}}^{\mathrm{bg}}$ for stable spatial contextual constraints and foreground prototypes $\mathbf{P}_{\mathrm{local}}^{\mathrm{fg}}$ to capture structural heterogeneity.

Specifically, we first extract a query-aligned global prototype $\mathbf{P}_{\mathrm{q}} \in \mathbb{R}^{1 \times C}$ from query feature space. Then, by computing the pixel-wise similarity between $\mathbf{P}_{\mathrm{q}}$ and query feature $\mathbf{F}_{\mathrm{q}}^{'}$, we generate the global similarity map $\mathbf{S}_{\mathrm{glob}} \in \mathbb{R}^{1 \times H \times W}$:

$$\begin{cases} \mathbf{P}_{\mathrm{q}} = \dfrac{\sum_p \mathbf{F}_{\mathrm{q}}^{'}(p) \cdot \widetilde{\mathbf{M}}_{\mathrm{q}}^{\mathrm{init}}(p)}{\sum_p \widetilde{\mathbf{M}}_{\mathrm{q}}^{\mathrm{init}}(p)} \\ \mathbf{S}_{\mathrm{glob}}(p) = \mathrm{Cos}(\mathbf{F}_{\mathrm{q}}^{'}(p), \mathbf{P}_{\mathrm{q}}) \end{cases}, \tag{14}$$

where where p denotes a spatial location; $\mathrm{Cos}(\cdot, \cdot)$ denotes the operation of computing cosine similarity.

Subsequently, we compute the foreground and surrounding background similarity map $\mathbf{S}_{\mathrm{fg}} \in \mathbb{R}^{N_{\mathrm{fg}} \times H \times W}$ and $\mathbf{S}_{\mathrm{bg}} \in \mathbb{R}^{N_{\mathrm{bg}} \times H \times W}$ between $\mathbf{F}_{\mathrm{q}}^{'}$ and $\mathbf{P}_{\mathrm{local}}^{\mathrm{fg}}$ or $\mathbf{P}_{\mathrm{local}}^{\mathrm{bg}}$:

$$\begin{cases} \mathbf{S}_{\mathrm{fg}}(i,p) = \mathrm{Cos}(\mathbf{F}_{\mathrm{q}}^{'}(p), \mathbf{P}_{\mathrm{local}}^{\mathrm{fg}}(i)) \\ \mathbf{S}_{\mathrm{bg}}(i,p) = \mathrm{Cos}(\mathbf{F}_{\mathrm{q}}^{'}(p), \mathbf{P}_{\mathrm{local}}^{\mathrm{bg}}(i)) \end{cases}, \tag{15}$$

where $\mathrm{Cos}(\cdot, \cdot)$ denotes the operation of computing cosine similarity; i refers to the i-th local prototype and similarity map; p denotes a spatial location in the query feature map.

Next, guided by the global similarity map $\mathbf{S}_{\mathrm{glob}}$, we exploit the foreground similarity maps $\mathbf{S}_{\mathrm{fg}}$ and the surround background similarity maps $\mathbf{S}_{\mathrm{bg}}$ to obtain a more fine-grained similarity map $\mathbf{S}_{\mathrm{local}}$:

$$\begin{cases} \mathbf{S}_{\mathrm{fg}}^{'} = \mathrm{F}_{\theta_1}^{\mathrm{sm}}(\mathbf{S}_{\mathrm{fg}} \oplus \mathbf{S}_{\mathrm{glob}}) \in \mathbb{R}^{1 \times H \times W} \\ \mathbf{S}_{\mathrm{bg}}^{'} = \mathrm{F}_{\theta_2}^{\mathrm{sm}}(\mathbf{S}_{\mathrm{bg}} \oplus (1 - \mathbf{S}_{\mathrm{glob}})) \in \mathbb{R}^{1 \times H \times W} \\ \mathbf{S}_{\mathrm{local}} = \mathrm{D}_{\theta}(\mathbf{S}_{\mathrm{fg}}^{'} \oplus \mathbf{S}_{\mathrm{bg}}^{'}) \in \mathbb{R}^{H \times W} \end{cases}, \tag{16}$$

where $\oplus$ denotes the concatenation along the first dimension. $\mathrm{F}^{\mathrm{sm}}_{\theta_1}(\cdot)$ and $\mathrm{F}^{\mathrm{sm}}_{\theta_2}(\cdot)$ denote two refinement blocks composed of two convolution layers, which map the input channels from $N_{\mathrm{fg}}+1$ and $N_{\mathrm{bg}}+1$ to 1, respectively. $\mathrm{D}_{\theta}(\cdot)$ denotes a decoder that maps the concatenated similarity maps from 2 channels to 1 channel.

The final prediction $\widetilde{\mathbf{M}}_{\mathrm{q}}$ is obtained by integrating the global semantic localization from $\mathbf{S}_{\mathrm{glob}}$ and the fine-grained details from $\mathbf{S}_{\mathrm{local}}$, achieving a coarse-to-fine refinement:

$$\widetilde{\mathbf{M}}_{\mathrm{q}} = \sigma((\mathbf{S}_{\mathrm{glob}} + \mathbf{S}_{\mathrm{local}}) \cdot \beta) \in \mathbb{R}^{H \times W}, \tag{17}$$

where β represents a learnable parameter that modulates the fused responses before they are projected into the probability space $[0, 1]$ via Sigmoid function $\sigma(\cdot)$.

3.6 Loss Function

To supervise each training episode, we leverage the pixel-wise Cross-Entropy loss $\mathcal{L}_{\text{prim}}$ as primary loss between the predicted mask $\widetilde{\mathbf{M}}_{\mathrm{q}}$ and ground truth $\mathbf{M}_{\mathrm{q}} \in \{0, 1\}^{H \times W}$:

$$\mathcal{L}_{\text{prim}} = -\frac{1}{HW} \sum \left[\mathbf{M}_{\mathrm{q}} \log \widetilde{\mathbf{M}}_{\mathrm{q}} + (1 - \mathbf{M}_{\mathrm{q}}) \log(1 - \widetilde{\mathbf{M}}_{\mathrm{q}}) \right]. \tag{18}$$

To facilitate rapid convergence and provide explicit semantic guidance, we impose an auxiliary coarse loss $\mathcal{L}_{\mathrm{cor}}$ on the initial prediction $\widetilde{\mathbf{M}}^{\mathrm{init}}_{\mathrm{q}}$ in Eq. 8:

$$\mathcal{L}_{\mathrm{cor}} = -\frac{1}{HW} \sum \left[\mathbf{M}_{\mathrm{q}} \log \widetilde{\mathbf{M}}^{\mathrm{init}}_{\mathrm{q}} + (1 - \mathbf{M}_{\mathrm{q}}) \log(1 - \widetilde{\mathbf{M}}^{\mathrm{init}}_{\mathrm{q}}) \right]. \tag{19}$$

In GAI module, we employ a rotation loss $\mathcal{L}_{\text{rotate}}$ to regularize $\mathbf{F}_{\mathrm{geo}} \in \mathbb{R}^{N_{\text{dir}} \times C_{\text{shape}}}$ (Eq. 3), forcing the model to learn orientation-agnostic geometric descriptors that maintain consistency across all N_{dir} directions. By using cosine similarity $\mathrm{Cos}(\cdot, \cdot)$, $\mathcal{L}_{\text{rotate}}$ is formulated as:

$$\mathcal{L}_{\text{rotate}} = \frac{1}{N^2_{\text{dir}}} \sum_{i,j=1}^{N_{\text{dir}}} \left(1 - \mathrm{Cos}(\mathbf{F}\mathrm{geo}^{(i)}, \mathbf{F}\mathrm{geo}^{(j)}) \right). \tag{20}$$

In summary, the loss function for each training is formulated as: $\mathcal{L}_{\text{total}} = \mathcal{L}_{\text{prim}} + \mathcal{L}_{\mathrm{cor}} + \mathcal{L}_{\text{rotate}}$.

4 Experiment

4.1 Datasets and Setting

Datasets. (1) **The Cross-Modality dataset** comprises Abdominal MRI (Abd-MR) dataset (Kavur et al., 2021) with 20 3D T2-SPIR volumes and Abdominal CT (Abd-CT) dataset (Landman et al., 2015) with 20 3D CT volumes. Four organs are utilized: liver, spleen, left kidney (LK), and right kidney (RK). (2) **The Cross-Sequence dataset** is a cardiac dataset (Zhuang et al., 2022) including 45 3D LGE MRI volumes and 45 b-SSFP MRI volumes with three distinct labels: the left ventricle blood pool (LV-BP), left ventricle myocardium (LV-MYO), and right ventricle (RV).

Setting. Following the evaluation protocol established in FAMNet (Bo et al., 2025b), we conduct four challenging cross-domain tasks: (1) Abd-CT $\rightarrow$ Abd-MR, (2) Abd-MR $\rightarrow$ Abd-CT, (3) Cardiac-LGE $\rightarrow$ Cardiac-bssFP, and (4) Cardiac-bssFP $\rightarrow$ Cardiac-LGE. In each task, the notation $\mathcal{S} \rightarrow \mathcal{T}$ indicates that the model is trained on the source domain $\mathcal{S}$ and evaluated on the target domain $\mathcal{T}$, where a significant modality shift exists.

4.2 Implementation Details

Our GPS is trained for 10 epochs, with each epoch comprising 5,000 iterations, using SGD (learning rate 0.001, momentum 0.9), with a decay factor of 0.9 every 1,000 iterations. We adopt a 5-fold cross-validation protocol. In each fold, the third patient is designated as the support set, and the remaining patients constitute the query set. For the support set, under the 1-shot setting, the middle

Table 1: Quantitative comparison on the Cross-Modality tasks in terms of Dice score (%, mean ± std). The best and second-best results are highlighted in **bold** and underlined, respectively.

Method	Reference	Abdominal CT → MRI					Abdominal MRI → CT				
		Liver	LK	RK	Spleen	Mean	Liver	LK	RK	Spleen	Mean
SSL-ALP	TMI'22	73.35 ± 04.22	62.01 ± 05.77	70.70 ± 04.43	57.59 ± 10.06	65.91 ± 03.64	76.47 ± 05.52	63.86 ± 13.38	57.51 ± 12.51	67.79 ± 08.39	66.41 ± 08.86
PATNet	ECCV'22	67.39 ± 04.41	58.37 ± 07.54	64.81 ± 04.95	55.01 ± 07.30	61.39 ± 04.60	67.60 ± 06.41	45.37 ± 10.33	50.92 ± 10.84	50.07 ± 11.29	53.49 ± 07.14
QNet	IntelliSys'23	48.95 ± 13.30	53.23 ± 08.98	62.11 ± 11.74	56.62 ± 07.88	55.23 ± 04.47	74.10 ± 03.84	64.22 ± 06.92	59.23 ± 14.41	65.75 ± 07.62	65.82 ± 04.92
RPT	MICCAI'23	66.93 ± 09.21	56.99 ± 10.51	57.18 ± 16.16	56.61 ± 07.35	59.43 ± 06.37	**79.87 ± 01.26**	58.62 ± 03.57	60.37 ± 14.08	68.89 ± 09.25	66.94 ± 04.75
IFA	CVPR'24	72.69 ± 03.24	53.24 ± 08.03	51.28 ± 07.46	54.38 ± 08.59	57.90 ± 04.36	71.36 ± 05.94	56.72 ± 16.45	49.15 ± 21.78	60.63 ± 08.06	59.46 ± 11.39
TPM	MICCAI'25	65.58 ± 05.03	61.46 ± 07.72	68.99 ± 08.62	57.03 ± 07.22	63.26 ± 05.39	72.74 ± 07.58	53.78 ± 15.17	54.54 ± 13.58	62.52 ± 05.02	60.89 ± 05.35
FAMNet	AAAI'25	**75.23 ± 04.73**	62.64 ± 08.87	68.19 ± 09.12	62.32 ± 10.29	67.10 ± 06.29	74.79 ± 05.54	57.76 ± 09.70	56.19 ± 18.84	63.38 ± 06.93	63.03 ± 08.01
DSM	TIP'25	69.19 ± 08.26	61.01 ± 11.23	71.16 ± 10.01	54.69 ± 11.05	64.01 ± 07.33	77.69 ± 03.94	59.18 ± 04.62	59.53 ± 21.75	**69.90 ± 10.27**	66.57 ± 08.35
C-Graph	TMI'26	68.73 ± 04.07	69.99 ± 08.85	74.66 ± 11.60	62.46 ± 11.02	68.96 ± 07.11	71.74 ± 08.67	66.14 ± 04.29	64.63 ± 06.75	67.36 ± 05.95	67.47 ± 04.55
GPS (Ours)	—	70.13 ± 03.26	**75.87 ± 07.25**	**84.34 ± 04.99**	**64.91 ± 11.33**	**73.81 ± 05.52**	70.52 ± 03.94	**71.37 ± 06.07**	**66.85 ± 09.10**	68.90 ± 04.55	**69.41 ± 05.33**

Table 2: Quantitative comparison on the Cross-Sequence tasks in terms of Dice score (%, mean ± std). The best and second-best results are highlighted in **bold** and underlined, respectively.

Method	Reference	Cardiac LGE → b-SSFP				Cardiac b-SSFP → LGE			
		LV-BP	LV-MYO	RV	Mean	LV-BP	LV-MYO	RV	Mean
SSL-ALP	TMI'22	84.89 ± 02.89	52.92 ± 06.19	74.97 ± 04.94	70.93 ± 04.97	70.15 ± 05.02	50.25 ± 03.68	62.24 ± 03.64	60.88 ± 03.79
PATNet	ECCV'22	76.29 ± 08.11	24.77 ± 05.22	56.88 ± 12.80	52.65 ± 08.07	51.58 ± 16.03	27.63 ± 08.22	46.60 ± 04.53	41.94 ± 08.36
QNet	IntelliSys'23	85.08 ± 02.27	40.38 ± 05.71	71.90 ± 01.84	65.78 ± 01.39	68.10 ± 07.32	41.93 ± 07.21	56.65 ± 07.48	55.56 ± 05.86
RPT	MICCAI'23	83.06 ± 02.43	48.55 ± 05.16	71.76 ± 05.89	67.79 ± 03.58	75.50 ± 05.60	49.05 ± 04.01	60.18 ± 05.50	61.58 ± 03.29
IFA	CVPR'24	78.03 ± 10.50	44.69 ± 06.61	65.96 ± 10.58	62.89 ± 08.72	71.15 ± 11.02	46.54 ± 03.99	54.01 ± 12.77	57.23 ± 08.37
TPM	MICCAI'25	85.31 ± 02.70	40.41 ± 10.59	74.60 ± 03.56	66.77 ± 04.60	64.98 ± 10.70	45.37 ± 11.51	61.00 ± 06.97	57.12 ± 08.01
FAMNet	AAAI'25	87.34 ± 02.04	49.76 ± 01.23	74.57 ± 03.26	70.56 ± 01.96	**76.76 ± 06.36**	47.54 ± 03.99	56.92 ± 05.98	60.40 ± 03.75
DSM	TIP'25	87.85 ± 03.65	49.79 ± 03.33	73.53 ± 09.31	70.39 ± 05.18	67.93 ± 01.86	45.62 ± 04.71	50.90 ± 11.38	54.82 ± 04.72
C-Graph	TMI'26	86.62 ± 04.80	52.69 ± 04.41	75.88 ± 04.05	71.73 ± 03.37	70.82 ± 06.63	52.21 ± 05.54	62.57 ± 07.36	61.87 ± 05.27
GPS (Ours)	—	**89.94 ± 02.23**	**62.58 ± 02.86**	**83.12 ± 04.43**	**78.55 ± 01.78**	76.35 ± 05.17	**57.16 ± 05.13**	**64.11 ± 01.83**	**65.87 ± 05.58**

Table 3: Effect of each loss in GPS (mean Dice(%)).

$\mathcal{L}_{\text{prim}}$	$\mathcal{L}_{\text{cor}}$	$\mathcal{L}_{\text{rotate}}$	Abd-CT → Abd-MR				
			Liver	RK	LK	Spleen	Mean
✓			62.90	69.96	57.32	57.03	61.80
✓		✓	64.22	71.16	58.65	58.79	63.20
✓	✓		69.00	82.16	73.96	63.70	72.20
✓	✓	✓	**70.13**	**84.34**	**75.87**	**64.91**	**73.81**

Table 4: Effect of key modules in GPS in terms of mean Dice Score (%).

GPS-base	GAI	PPM&PSR	Abd-CT → Abd-MR				
			Liver	RK	LK	Spleen	Mean
✓			62.74	61.36	56.04	55.79	58.98
✓	✓		67.75	80.12	74.00	60.09	70.49
✓	✓	✓	**70.13**	**84.34**	**75.87**	**64.91**	**73.81**

slice containing the target organ and its corresponding segmentation mask are selected as the support pair. For a fair comparison, all methods use the same ResNet-50 encoder pretrained on MS-COCO and are retrained and evaluated under identical data splits and evaluation protocols.

We set $C = C_{\text{shape}} = 512$ (no relationship of them), $N_{\text{dir}} = 16$, and $N_{\text{fg}} = N_{\text{bg}} = 10$, while initializing the learnable scaling parameter β in PSR to 10. We follow the 5-fold evaluation protocol of FAMNet (Bo et al., 2025b), using the third patient volume in each fold as support and the remaining volumes as queries. Performance is measured by the Sørensen-Dice coefficient (Setiawan, 2020).

4.3 Comparison with SOTA Methods

To evaluate the efficacy of GPS, we conduct comprehensive comparisons against various state-of-the-art (SOTA) methods, including SSL-ALP (Ouyang et al., 2022), PATNet (Wang et al., 2019), QNet (Shen et al., 2023), RPT (Zhu et al., 2023), IFA (Nie et al., 2024), TPM (Kim et al., 2025), FAMNet (Bo et al., 2025b), DSM (Zhu et al., 2025b), and C-Graph (Bo et al., 2025a).

As shown in Table. 1 and Table. 2. Our GPS achieves the sota performance in the mean Dice score of four task. Notably, GPS delivers a significant performance leap in the LV-MYO category, surpassing the second-best method in the Cardiac LGE ↔ bssFP tasks, confirming the unique capability of GAI's dual-ring structure to preserve the geometric integrity of target class.

4.4 Ablation Studies

Validation of Geometric and peripheral background Consistency. We evaluate the consistency of foreground shapes and surrounding-background distributions between support and query samples using the Bhattacharyya coefficient (BC), where higher values indicate greater similarity. As shown in Fig. 3, both foreground geometry and peripheral-background distributions remain highly consistent across Abd-MR, Abd-CT, Cardiac-LGE, and Cardiac-bSSFP. This suggests that anatomical shape and surrounding context provide complementary domain-robust cues, motivating their joint exploitation in our method.

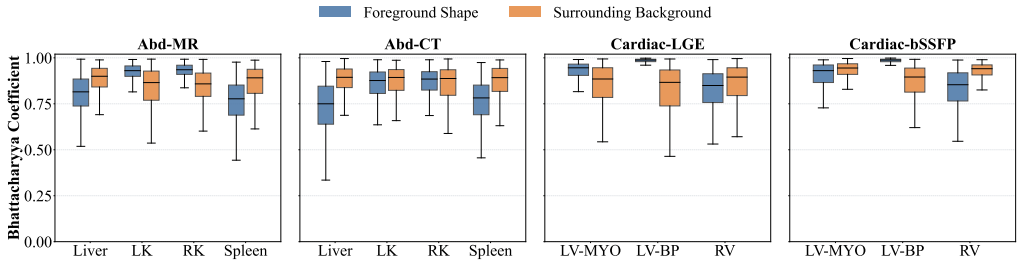

Figure 3: Bhattacharyya coefficient distributions for foreground shapes and surrounding backgrounds between support-query samples. Higher values indicate stronger consistency.

Table 5: Effect of Dual / Single structure in the GAI module in terms of mean Dice score(%).

Structure	Cardiac LGE → Cardiac bssFP			
	LV-MYO (Hollow)	LV-BP (Solid)	RV (Solid)	Mean
Single-Ring	60.39	89.85	83.06	77.76
Dual-Ring	**62.58**	**89.94**	**83.12**	**78.55**

Table 6: Effect of the foreground/peripheral background prototypes in terms of mean Dice Score (%).

	Abd-CT → Abd-MR				
	Liver	RK	LK	Spleen	Mean
GPS w/o PSR	67.75	80.12	74.00	60.09	70.49
PSR w/o bg	69.55	<u>82.61</u>	<u>74.14</u>	<u>63.68</u>	<u>72.49</u>
PSR w/o fg	<u>69.61</u>	82.36	73.28	61.75	71.75
PSR	**70.13**	**84.34**	**75.87**	**64.91**	**73.81**

Effect of each loss function. In Table. 3, we independently evaluate the effect of $\mathcal{L}_{\text{cor}}$ and $\mathcal{L}_{\text{rotate}}$, alongside their synergistic relationship. Building upon the primary loss $\mathcal{L}_{\text{prim}}$ (Mean Dice score 61.80%), the separate integration of either $\mathcal{L}_{\text{rotate}}$ or $\mathcal{L}_{\text{cor}}$ yields consistent performance gains by 1.40% and 10.40%, respectively. Notably, their joint supervision achieves the highest result (73.81%), demonstrating a strong synergistic effect. These results confirm that $\mathcal{L}_{\text{cor}}$ is essential for consolidating global semantics, while $\mathcal{L}_{\text{rotate}}$ ensures the rotational invariance of geometric features, both of which are indispensable for robust cross-domain alignment.

Effect of each module. We evaluate the impact of GPS by incrementally integrating its modules into the GPS-base (global prototype matching). In Table 4, GAI significantly boosts performance (from 58.98% to 70.49%) by injecting geometric priors. Jointly adding PPM and PSR further improves results by 3.32%, confirming that synergizing stable peripheral background is vital for refining predictions.

Effect of dual-ring structure in GAI module. To validate the adaptive capability of GAI's dual-ring structure, we compare it against a single-ring structure (outer-ring only) on the Cardiac dataset, which contains both hollow and solid classes. As shown in Table. 5, when addressing hollow structure (LV-MYO), the dual-ring design demonstrates superior representational capacity compared to the single-ring structure, with the Dice score improving from 60.39% to 62.58%. Conversely, for solid organs (e.g., LV-BP and RV), the dual-ring design naturally degenerates into its single-ring equivalent, maintaining consistent performance. This confirms the robust topological adaptability of our dual-ring GAI module across diverse and complex target classes.

Impact of bipartite prototypes in PSR module. We evaluate the individual contributions of foreground and background prototypes by comparing various PSR configurations against the baseline (GPS w/o PSR). As shown in Table 6, incorporating foreground multi-prototypes (PSR w/o BG) improves the mean Dice score from 70.49% to 72.49%, while utilizing only background prototypes (PSR w/o FG) also yields a notable gain (71.75%) by effectively suppressing false positives. Notably, the full PSR module achieves the best result (73.81%), demonstrating the importance of combining foreground structure and peripheral context for robust refinement.

5 Conclusion

In this paper, we proposed GPS, a novel CD-FSMIS framework that leverages invariant geometric and peripheral-background priors as "Anatomical Anchors". By synergizing the GAI, PPM, and PSR modules, GPS effectively transcends domain-specific biases and achieves robust feature alignment. Extensive experiments confirm its superiority, highlighting the critical value of GPS in CD-FSMIS.

6 AI USE STATEMENT

In this work, we used generative AI tools solely for language polishing, including improving grammar, clarity, fluency, and academic style. We did not use generative AI tools to generate research ideas, experimental results, data, figures, or scientific claims. All AI-assisted text was carefully reviewed and revised by the authors to ensure technical accuracy and consistency with the original content. We take full responsibility for the final content of this work, including all text, claims, and artifacts produced with the aid of generative AI.

REFERENCES

Yuntian Bo, Tao Zhou, Zechao Li, Haofeng Zhang, and Ling Shao. Contrastive graph modeling for cross-domain few-shot medical image segmentation. *IEEE Transactions on Medical Imaging*, 45: 2173–2186, 2025a.

Yuntian Bo, Yazhou Zhu, Lunbo Li, and Haofeng Zhang. Famnet: Frequency-aware matching network for cross-domain few-shot medical image segmentation. In *AAAI*, volume 39, pp. 1889–1897, 2025b.

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 *CVPR*, pp. 30032–30041, 2026.

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.

Xuming Chen, Shanlin Sun, Narisu Bai, Kun Han, Qianqian Liu, et al. A deep learning-based auto-segmentation system for organs-at-risk on whole-body computed tomography images for radiation therapy. *Radiotherapy and Oncology*, 160:175–184, 2021.

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.

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*, 44(6):2732–2744, 2025.

P. Choyke, B. Turkbey, P. Pinto, M. Merino, and B. Wood. Data From PROSTATE-MRI, 2016. URL `https://doi.org/10.7937/K9/TCIA.2016.6046GUDV`.

Anish Dhir, Matthew Ashman, James Requeima, and Mark van der Wilk. A meta-learning approach to bayesian causal discovery. In *ICLR*, volume 2025, pp. 14158–14178, 2025.

Lin Lawrence Guo, Stephen R Pfohl, Jason Fries, Alistair EW Johnson, Jose Posada, Catherine Aftandilian, Nigam Shah, and Lillian Sung. Evaluation of domain generalization and adaptation on improving model robustness to temporal dataset shift in clinical medicine. *Scientific reports*, 12(1):2726, 2022.

Jiayu Huo, Xi Ouyang, Sébastien Ourselin, and Rachel Sparks. Generative medical segmentation. In *AAAI*, volume 39, pp. 3851–3859, 2025.

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.

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

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.

Hyeongji Kim, Stine Hansen, and Michael Kampffmeyer. Tied prototype model for few-shot medical image segmentation. In *MICCAI*, pp. 651–661, 2025.

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 *ICCV*, pp. 4015–4026, 2023.

Bennett Landman, Zhoubing Xu, J Igelsias, Martin Styner, Thomas Langerak, and Arno Klein. Miccai multi-atlas labeling beyond the cranial vault–workshop and challenge. In *MICCAI Workshop Challenge*, volume 5, pp. 12, 2015.

Sang-Woong Lee, Jawad Tanveer, Amir Masoud Rahmani, Hamid Alinejad-Rokny, Parisa Khoshvaght, Gholamreza Zare, Pegah Malekpour Alamdari, and Mehdi Hosseinzadeh. Sfgcn: Synergetic fusion-based graph convolutional networks approach for link prediction in social networks. *Information Fusion*, 114:102684, 2025.

Shuo Lei, Xuchao Zhang, Jianfeng He, Fanglan Chen, Bowen Du, and Chang-Tien Lu. Cross-domain few-shot semantic segmentation. In *ECCV*, pp. 73–90, 2022.

Yiwen Li, Yunguan Fu, Iani JMB Gayo, Qianye Yang, Zhe Min, Shaheer U Saeed, Wen Yan, Yipei Wang, J Alison Noble, Mark Emberton, et al. Prototypical few-shot segmentation for cross-institution male pelvic structures with spatial registration. *Medical Image Analysis*, 90:102935, 2023.

Tsung-Yi Lin, Michael Maire, Serge Belongie, James Hays, Pietro Perona, Deva Ramanan, Piotr Dollár, and C Lawrence Zitnick. Microsoft coco: Common objects in context. In *ECCV*, pp. 740–755, 2014a.

Tsung-Yi Lin, Michael Maire, Serge Belongie, James Hays, Pietro Perona, Deva Ramanan, Piotr Dollár, and C Lawrence Zitnick. Microsoft coco: Common objects in context. In *ECCV*, pp. 740–755, 2014b.

Yi Lin, Yufan Chen, Kwang-Ting Cheng, and Hao Chen. Few shot medical image segmentation with cross attention transformer. In *MICCAI*, pp. 233–243, 2023.

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 *CVPR*, pp. 3380–3390, 2024.

Cheng Ouyang, Carlo Biffi, Chen Chen, Turkay Kart, Huaqi Qiu, and Daniel Rueckert. Self-supervised learning for few-shot medical image segmentation. *IEEE Transactions on Medical Imaging*, 41(7):1837–1848, 2022.

Ofir Pele and Michael Werman. Fast and robust earth mover's distances. In *ICCV*, pp. 460–467. IEEE, 2009.

Yossi Rubner, Carlo Tomasi, and Leonidas J Guibas. The earth mover's distance as a metric for image retrieval. *International journal of computer vision*, 40(2):99–121, 2000.

Agung W Setiawan. Image segmentation metrics in skin lesion: accuracy, sensitivity, specificity, dice coefficient, jaccard index, and matthews correlation coefficient. In *2020 International Conference on Computer Engineering, Network, and Intelligent Multimedia (CENIM)*, pp. 97–102. IEEE, 2020.

Qianqian Shen, Yanan Li, Jiyong Jin, and Bin Liu. Q-net: Query-informed few-shot medical image segmentation. In *IntelliSys*, pp. 610–628, 2023.

Weili Shi, Penglong Zhang, Yuqin Li, and Zhengang Jiang. Segment anything model for few-shot medical image segmentation with domain tuning. *Complex & Intelligent Systems*, 11(1):37, 2025.

Zixian Su, Kai Yao, Xi Yang, Kaizhu Huang, Qiufeng Wang, and Jie Sun. Rethinking data augmentation for single-source domain generalization in medical image segmentation. In *AAAI*, volume 37, pp. 2366–2374, 2023.

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.

Philipp Tschandl, Cliff Rosendahl, and Harald Kittler. The ham10000 dataset, a large collection of multi-source dermatoscopic images of common pigmented skin lesions. *Scientific data*, 5(1): 180161, 2018.

Lazaros Tsochatzidis, Panagiota Koutla, Lena Costaridou, and Ioannis Pratikakis. Integrating segmentation information into cnn for breast cancer diagnosis of mammographic masses. *Computer Methods and Programs in Biomedicine*, 200:105913, 2021.

Kaixin Wang, Jun Hao Liew, Yingtian Zou, Daquan Zhou, and Jiashi Feng. Panet: Few-shot image semantic segmentation with prototype alignment. In *ICCV*, pp. 9197–9206, 2019.

Bin Xu, Yazhou Zhu, Shidong Wang, Yang Long, and Haofeng Zhang. Few-shot medical image segmentation via boundary-extended prototypes and momentum inference. *Computer Vision and Image Understanding*, 263:104571, 2026.

Ming Yanzhen, Chen Song, Li Wanping, Yang Zufang, and Alan Wang. Exploring approaches to tackle cross-domain challenges in brain medical image segmentation: a systematic review. *Frontiers in Neuroscience*, 18:1401329, 2024.

Lin Zhao, Xiao Chen, Eric Z Chen, Yikang Liu, Terrence Chen, and Shanhui Sun. Retrieval-augmented few-shot medical image segmentation with foundation models. *IEEE Transactions on Neural Networks and Learning Systems*, 2025.

Wei Zhou, Guilin Guan, Qifeng Yan, and Yitian Zhao. Adaptsam: Adaptive sam for cross-domain few-shot medical image segmentation. In *2025 IEEE International Conference on Bioinformatics and Biomedicine (BIBM)*, pp. 3389–3396. IEEE, 2025.

Yazhou Zhu and Haofeng Zhang. Maup: Training-free multi-center adaptive uncertainty-aware prompting for cross-domain few-shot medical image segmentation. In *MICCAI*, pp. 326–336. Springer, 2025.

Yazhou Zhu, Shidong Wang, Tong Xin, and Haofeng Zhang. Few-shot medical image segmentation via a region-enhanced prototypical transformer. In *MICCAI*, pp. 271–280, 2023.

Yazhou Zhu, Minxian Li, Qiaolin Ye, Shidong Wang, Tong Xin, and Haofeng Zhang. Robustemd: Domain robust matching for cross-domain few-shot medical image segmentation. *Artificial Intelligence in Medicine*, pp. 103197, 2025a.

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*, 34:6669–6682, 2025b.

Xiahai Zhuang, Jiahang Xu, Xinzhe Luo, Chen Chen, Cheng Ouyang, et al. Cardiac segmentation on late gadolinium enhancement mri: a benchmark study from multi-sequence cardiac mr segmentation challenge. *Medical Image Analysis*, 81:102528, 2022.

APPENDIX

.1 IMPLEMENTATION DETAILS

Computational Setup. The proposed GPS framework is implemented in PyTorch 2.4.1 and executed on an NVIDIA RTX 3090 GPU (24GB).

Hyperparameter Settings. With the batch size of 1, the model is trained for 10 epochs, with each epoch comprising 5,000 iterations. We utilize the Stochastic Gradient Descent (SGD) optimizer with an initial learning rate of 1×10^{-3} and a momentum of 0.9. An exponential decay factor of 0.9 is applied every 1,000 iterations to ensure stable convergence. Both the feature dimension C and the geometric shape dimension C_{shape} are set to 512 (no relationship of them); The number of geometric sampling directions N_{dir} is fixed at 16. In the PPM module, both the number of foreground prototypes N_{fg} and peripheral-background prototypes N_{bg} are both set to 10. The learnable scaling parameter β in PSR module is initialized to 10.

Evaluation Protocol. To ensure a fair and rigorous comparison, we strictly follow the protocols established by baseline methods (Bo et al., 2025b; Zhu et al., 2025a;b; Bo et al., 2025a): All experiments are conducted in a 1-way 1-shot setting using 2D slice-level segmentation. During inference, we perform 5-fold cross-validation, where the third patient volume in each task is designated as the support set, while the remaining volumes serve as queries. Notably, for each support volume, we select the median slice containing the target organ as the representative support image.

Evaluation Metric. To assess the segmentation performance, we employ the Sorensen-Dice Score (Dice score) coefficient (Setiawan, 2020) commonly used in FSMIS tasks. The DSC measures the spatial overlap between the predicted mask $\mathbf{X}$ and the ground truth $\mathbf{Y}$, defined as:

$$\text{DSC}(\mathbf{X}, \mathbf{Y}) = \frac{2|\mathbf{X} \cap \mathbf{Y}|}{|\mathbf{X}| + |\mathbf{Y}|}, \tag{21}$$

where $|\cdot|$ denotes the number of pixels (or voxels) in the corresponding region; X, Y denote the two masks respectively, and DSC denotes the Dice score range from 0 to 1.

Computation of Centroid. In the GAI module, we obtain the centroid $C_{\text{cent}} \in \mathbb{R}^2$ of foreground region in support mask $\mathbf{M}_{\text{s}} \in \mathbb{R}^{H \times W}$. The centroid C_{cent} is formulated as:

$$C_{\text{cent}} = \frac{1}{|\Omega_s|} \sum_{p \in \Omega_s} p, \tag{22}$$

where Ω_s denotes the set of spatial coordinates (h, w) where $\mathbf{M}_s^{(h,w)} = 1$, and $|\cdot|$ represents the cardinality of the set.

.2 ABLATION

Effect of different pre-trained encoders. To comprehensively investigate the impact of feature initialization and network capacity on cross-domain generalization, we evaluate different pre-trained encoders including ResNet50-Imagenet, ResNet101-Imagenet, ResNet50-COCO, ResNet101-COCO on Abd-CT $\rightarrow$ Abd-MR task. Notably, Table 7 shows that a deeper pre-trained encoder always improves performance, as ResNet101 gets higher scores than ResNet50 in both settings, with mean Dice score increases of 1.7% (Resnet50-Imagenet vs. Resnet101-Imagenet) and 1.09% (Resnet50-COCO vs. Resnet101-COCO); however, COCO-pretrained models work much better than ImageNet-pretrained ones. This can be attributed to the fact that the COCO dataset is designed for instance-level tasks like segmentation, which provides more relevant spatial and structural feature representations than the image-level classification objective of ImageNet.

Effect of the number of geometric directions. To investigate how the number of geometric directions N_{dir} in the GAI module affects performance, we conduct experiments by varying N_{dir} from 8 to 40. In Table. 8, the mean Dice scores initially improves as N_{dir} increases, indicating that a larger N_{dir} helps represent complex target classes more effectively. However, when N_{dir} continues to grow, the performance begins to decline, This suggests that an excessive value of N_{dir} makes the model over-sensitive to minor geometric variations, thereby reducing its overall robustness.

Table 7: Ablation study on the Abd-CT → Abd-MR for different pre-trained encoders on the performance of GPS in terms of the Dice Score(%).

Encoder	Abd-CT → Abd-MR				
	Liver	RK	LK	Spleen	Mean
ResNet50-ImageNet	66.59	59.29	51.43	55.30	58.15
ResNet101-ImageNet	66.65	60.09	56.40	56.27	59.85
ResNet50-COCO	70.13	84.34	75.87	64.91	73.81
ResNet101-COCO	71.11	85.45	76.85	66.20	74.90

Table 8: Effect of the number of geometric directions N_{dir} in GPS in terms of mean Dice Score (%).

N_{dir}	Abd-CT → Abd-MR				
	Liver	RK	LK	Spleen	Mean
8	69.79	78.96	70.62	61.56	70.23
16	70.13	84.34	**75.87**	64.91	73.81
24	**71.42**	**84.48**	74.05	**65.97**	**73.98**
32	70.03	84.44	73.27	62.18	72.48
40	69.60	83.49	71.55	61.62	71.56

Table 9: Effect of the initial value of learnable scaling parameter β in terms of mean Dice Score (%).

β	Abd-CT → Abd-MR				
	Liver	RK	LK	Spleen	Mean
5	70.39	82.69	74.42	63.41	72.73
10	70.13	84.34	**75.87**	**64.91**	**73.81**
15	70.33	84.04	73.71	64.24	73.08
20	**70.48**	**84.54**	74.07	62.95	73.01

Table 10: Effect of the number of geometric dimension C_{shape} in GPS in terms of mean Dice Score (%).

C_{shape}	Abd-CT → Abd-MR				
	Liver	RK	LK	Spleen	Mean
64	69.33	77.55	68.72	61.27	69.21
128	69.89	79.84	69.92	61.73	70.34
256	69.80	82.48	71.30	62.50	71.53
512	70.13	**84.34**	**75.87**	**64.91**	**73.81**
1024	**70.84**	83.46	74.70	63.98	73.24

Table 11: Effect of the number of foreground prototypes N_{fg} in terms of mean Dice Score (%).

N_{fg}	N_{bg}	Abd-CT → Abd-MR				
		Liver	RK	LK	Spleen	Mean
5	10	69.70	80.87	73.67	62.88	71.78
10	10	70.13	84.34	75.87	**64.91**	73.81
15	10	**70.91**	**84.63**	**75.91**	64.37	**73.95**
20	10	70.53	82.57	75.14	63.89	73.03
25	10	70.75	81.95	72.21	64.31	72.31

Table 12: Effect of the number of peripheral background prototypes N_{bg} in terms of mean Dice Score (%).

N_{fg}	N_{bg}	Abd-CT → Abd-MR				
		Liver	RK	LK	Spleen	Mean
10	5	69.84	83.27	73.60	63.66	72.59
10	10	70.13	84.34	75.87	64.91	73.81
10	15	70.69	84.67	75.63	64.98	73.99
10	20	70.70	83.38	73.98	64.45	73.13
10	25	70.63	83.09	73.73	63.66	72.78

Impact of the scaling factor β. To validate the sensibility of the initial value of the learnable scaling factor β in the PSR module on segmentation performance, we conduct a sensitivity analysis by evaluating the model across a range of different β values (5 to 20) on task Abd-CT → Abd-MR. Table. 9 demonstrates that our GPS is robust to the initialization of β. While different initial values lead to slight variations in mean Dice scores, the performance remains consistently high, confirming the stability of the learning process in the PSR module.

Effect of geometric dimension C_{shape}. To investigate the impact of the geometric feature dimensionality C_{shape} on segmentation performance, we conduct a sensitivity analysis by evaluating the model across a range of different C_{shape} values (from 64 to 1024) on the Abd-CT → Abd-MR task. As shown in Table. 10, the mean Dice score first rises and then declines with increasing C_{shape}. This trend indicates that an insufficient dimensionality fails to capture enough geometric information for feature alignment, while an overly large C_{shape} triggers overfitting due to the limited supervision in CD-FSMIS task.

Impact of foreground prototypes N_{fg}. In order to investigate the effect of the number of the foreground prototypes N_{fg} in our GPS framework, we conduct experiments on task Abd-CT → Abd-MR by changing the value of N_{fg} (5 to 25) while keeping the number of peripheral background prototypes N_{bg} consistent. As illustrated in Table. 11, a insufficient number of foreground prototypes (N_{fg}=5) fails to adequately represent structures of complex target class. While increasing N_{fg} initially enhances the model's capacity to capture structural heterogeneity of the foreground, an excessive N_{fg} lead to a decline in segmentation performance. This degradation can be accounted for the inclusion of spurious noise in the feature space, which compromises the discriminative power of the foreground prototypes.

Impact of peripheral background prototypes N_{bg}. In order to validate the effect of the number of the peripheral background prototypes N_{bg} in our GPS framework, we conduct experiments on task Abd-CT → Abd-MR by changing the value of N_{bg} (5 to 25) while keeping the number N_{fg} of foreground prototypes consistent. As shown in Table. 12, the performance of GPS initially exhibits

Table 13: Efficiency's comparison of different methods on task Abd-CT → Abd-MR. The metrics measured are the mean Dice Score (%) for accuracy, the number of parameters, FLOPs for computational complexity, and Average Slice Processing Time (ASPT) for inference speed.

Method	mDice ↑	Params ↓	FLOPs ↓	ASPT ↓
FAMNet	67.10	2.86M	11.31G	41.74ms
DSM	64.01	6.35M	33.32G	68.05ms
C-Graph	68.96	7.35M	36.66G	163.52ms
GPS(Ours)	73.81	6.52M	16.58G	48.72ms

Table 14: Comparisons between our GAI and Direct Geometric Attention Activation (DGAA) method in terms of mean Dice Score (%) on the Abd-CT → Abd-MR task.

Method	Abd-CT → Abd-MR				
	Liver	RK	LK	Spleen	Mean
DGAA	67.91	69.71	60.93	63.51	65.51
GAI	70.13	84.34	75.87	64.91	73.81

an upward trend. This improvement is attributed to the fact that peripheral prototypes provide stable spatial constraints, effectively suppressing the activation of non-target pixels that correlate with the background context, thereby refining the segmentation. However, as N_{bg} continues to increase, the performance begins to decline. This suggests that an excessive number of background prototypes may capture domain-specific noise, which compromises the discriminative purity of the anatomical anchors and hinders robust generalization.

Efficiency comparison of other CD-FSMIS methods. To assess the practical deployability of our GPS, we compare the mean dice score (mDice), model complexity (Parameters and FLOPs) and average slice processing time (ASPT) of GPS against other SOTA methods, including FAMNet (Bo et al., 2025b), DSM (Zhu et al., 2025b) and C-Graph (Bo et al., 2025a) on a single NVIDIA RTX 3090 (24GB) GPU.

As shown in Table 13, while FAMNet maintains the lowest parameter count and FLOPs, its segmentation accuracy is significantly lower than our GPS. Notably, compared to C-Graph, GPS achieves superior accuracy (73.81% vs. 68.96%) with a much lighter computational burden. Specifically, GPS reduces FLOPs by 54.8% (16.58G vs. 36.66G) and accelerates the inference speed by approximately 3.4× (48.72ms vs. 163.52ms in ASPT). This dramatic reduction in complexity demonstrates that our GPS are far more efficient than dense graph-matching approaches, making GPS highly suitable for time-sensitive clinical applications.

Effectiveness of Dynamic Channel Activation. In the GAI module, we utilize the geometric feature vector $\mathbf{P}_{\text{geo}} \in \mathbb{R}^{1\times C_{\text{shape}}}$ to dynamically modulate domain-agnostic feature channels via cosine similarity, as formulated in Eqs. (6)–(7). A primary challenge in cross-domain scenarios is that the semantic mapping of feature channels often shifts; for instance, the i-th channel may represent geometric cues in the source domain but capture irrelevant noise in the target domain. By employing a similarity-based matching mechanism, GAI can adaptively identify and score the most relevant channels regardless of such semantic misalignments. To verify this, we compare GAI with **Direct Geometric Attention Activation (DGAA)**, which statically modulates channels using weights directly derived from the geometric vector without dynamic matching. As shown in Table. 14, GAI consistently outperforms DGAA. This superiority validates that our dynamic scoring mechanism provides crucial cross-domain adaptability, ensuring robust feature alignment even when channel distributions vary.

.3 More Experiment Results

External datasets introduction. We conducted external experiments in this section. Consequently, we first provide a detailed description of the extra utilized datasets.

Table 15: Quantitative comparison of Dice scores (%) of different methods on the Cross-Institution Dataset. The best results are highlighted in **bold**, and the second best are underlined.

Method	Reference	Prostate UCLH → NCI				Prostate NCI → UCLH			
		Bladder	CG	Rectum	Mean	Bladder	CG	Rectum	Mean
SSL-ALP	TMI'22	54.86	42.59	41.65	46.37	66.60	43.12	56.94	55.55
PATNet	ECCV'22	50.04	52.71	44.57	49.10	69.91	53.02	62.79	61.91
CATNet	MICCAI'23	47.61	45.29	42.51	45.14	51.98	48.81	55.01	51.93
RPT	MICCAI'23	52.17	41.22	46.93	46.77	62.35	51.66	64.63	59.55
IFA	CVPR'24	41.02	39.15	36.45	38.87	52.54	39.85	34.27	42.22
RobustEMD	AIIM'25	43.87	41.29	40.89	42.01	57.88	47.32	50.76	51.98
FAMNet	AAAI'25	53.06	48.48	52.57	51.37	77.64	48.73	68.91	65.09
DSM	TIP'25	56.72	49.56	51.97	52.75	75.43	59.26	75.43	70.04
C-Graph	TMI'26	40.67	50.43	48.10	46.40	68.79	52.94	60.11	60.61
GPS(Ours)	—	**73.42**	**58.07**	**52.71**	**61.40**	**80.19**	**60.92**	**76.33**	**72.48**

Table 16: Comparison of different methods on task Abd-CT → Chest-X-Ray in terms of Mean Dice(%).

	FAMNet	DSM	C-Grpah	GPS(Ours)
Dice	69.80	59.51	73.92	**75.36**

Table 17: Comparison of different methods on task Abd-CT → ISIC2018 in terms of Mean Dice Scores (%).

	FAMNet	DSM	C-Grpah	GPS(Ours)
Dice	45.31	48.56	50.19	**51.88**

(1) The Cross-Institution Dataset. This prostate dataset comprises 82 MRI scans from the National Cancer Institute (NCI) (Choyke et al., 2016) and 321 scans from the University College London Hospital (UCLH) (Li et al., 2023). The significant variations in scanning protocols and patient populations between these institutions provide a rigorous test for model generalization. Three categories are evaluated: bladder, central gland (CG), and rectum.

(2) The Chest X-ray Dataset. The Chest X-ray dataset Candemir et al. (2013); Jaeger et al. (2013), originally curated for Tuberculosis, comprises 704 X-ray images with pixel-level lung masks. We leverage lung tissue regions for experiments.

(3) The ISIC2018 Dataset. This dataset (Tschandl et al., 2018) contains 2594 lesion images with pixel-level masks for skin cancer. We leverage skin tissue regions for subsequent experiments.

Quantitative Results. For further validation, we compare our GPS with state-of-the-art methods including SSL-ALP (Ouyang et al., 2022), RPT (Zhu et al., 2023), PatNet (Lei et al., 2022), CATNet (Lin et al., 2023), IFA (Nie et al., 2024), RobustEMD (Zhu et al., 2025a), FAMNet (Bo et al., 2025b), DSM (Zhu et al., 2025b) and C-Graph (Bo et al., 2025a). The comparisons are conducted on Cross-Institution datasets (Prostate NCI ↔ Prostate UCLH tasks). Results of detailed experiments are listed in Table. 15.

In addition, we also conduct comparisons between our GPS and existing CD-FSMIS methods including FAMNet (Bo et al., 2025b), DSM (Zhu et al., 2025b) and C-Graph (Bo et al., 2025a) on the Abd-CT → Chest-X-ray and Abd-CT → ISIC2018 tasks. Detailed results are recorded in Table. 16 and Table. 17.

As shown in Table. 15, Table. 16, and Table. 17, GPS consistently achieves the highest mean Dice scores across all benchmark tasks. In the cross-institution prostate tasks (Table. 15), GPS outperforms the suboptimal DSM by a large margin, particularly in the UCLH → NCI setting (61.40% vs. 52.75%). Furthermore, GPS demonstrates remarkable adaptability to extreme shifts, such as Abd-CT → Chest-X-Ray task and Abd-CT → ISIC2018 task, where it continues to surpass graph-based C-Graph. This consistent superiority across disparate modalities and anatomical regions validates that our "Anatomical Anchors" provide a universal and reliable guidance mechanism for cross-domain few-shot medical image segmentation.

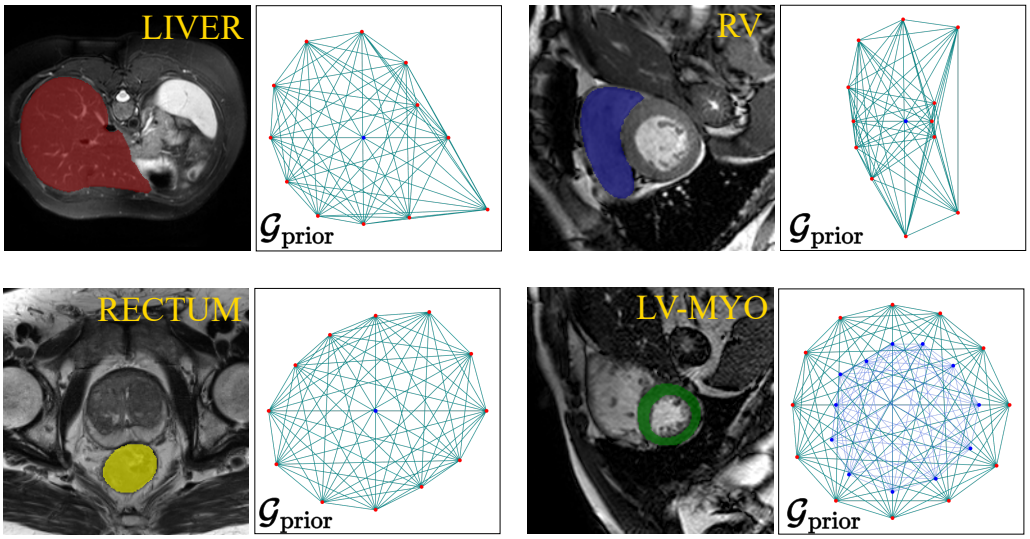

Figure 4: Visualization of the construction of geometric prior $\mathcal{G}_{\text{prior}}$ in GAI module from diverse target classes including liver in abdomen, right ventricle (RV) in cardiac, rectum in prostate, and left ventricle myocardium (LV-MYO) in cardiac. Beside the last class (LV-MYO) is hollow, the other classes are solid.

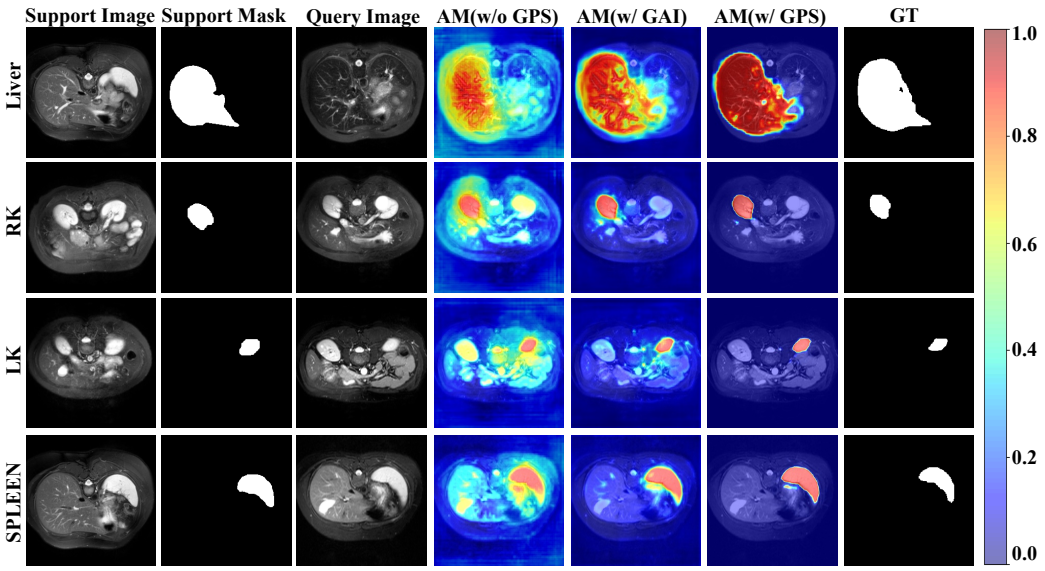

Figure 5: Qualitative visualization of activation maps (AMs) of GPS on task Abd-CT → Abd-MR. The values in the AM are normalized to a range of 0-1. From the left to right: support image, support mask, query image, AM generated by baselined (w/o GPS), AM generated by GAI module, AM generated by GPS and the ground truth of query image.

.4 VISUALIZATION

Activation Maps. To qualitatively evaluate the efficacy of GPS, we visualize the activation maps (AMs) derived from query features at different stages. From Fig. 5. we observe that the baseline method (a standard encoder with single prototype matching), is severely hampered by domain-specific noise under domain shift. This leads to redundant and scattered activations in non-target regions (fourth column). Upon integrating the GAI module, the activation regions become more concentrated and aligned with the explicit geometric priors.Finally, with the full GPS (incorporating PPM and PSR), the model synergizes foreground multi-prototypes to refine the target content and leverages peripheral-background prototypes to suppress false positives. This synergistic mecha-

nism yields fine-grained activation maps, validating the superior robustness and performance of our proposed GPS in cross-domain scenarios.

Visualization of geometric priors constructing. In order to verify the accurate construction of geometric priors in the GAI module, we visualize $\mathcal{G}_{\text{prior}}$ across diverse target classes including the liver, right ventricle (RV), rectum, and left ventricle myocardium (LV-MYO). As illustrated in Fig. 4, GAI generates high-fidelity geometric representations, establishing a robust foundation for subsequent domain-invariant channel modulation. Notably, this representational efficacy stems from the module's adaptability to both solid and hollow anatomical structures. Specifically, for hollow structures like the LV-MYO, the geometric prior is synergistically formed by both inner and outer rings. In contrast, for solid organs (e.g., liver and rectum), the outer ring sufficiently captures the morphology, while the inner ring naturally degenerates into a single point without interfering with the construction of $\mathcal{G}_{\text{prior}}$. These results further confirm the topological versatility of our dual-ring GAI module.

Visualization of segmentation results. To qualitatively evaluate the performance of GPS, we conduct comprehensive visual comparisons across multiple scenarios. Specifically, we first visualize the segmentation results of GPS and competing methods (RobustEMD, FAMNet, DSM, and C-Graph) on standard cross-domain benchmarks, including Abdominal (CT $\leftrightarrow$ MRI), Cardiac (LGE $\leftrightarrow$ bssFP), and Prostate (NCI $\leftrightarrow$ UCLH) tasks, as shown from Fig. (6-11).

Visualization Wide Cross-Domain Gap. Furthermore, to assess the extreme generalization capability of our framework with computing methods (FAMNet, DSM, C-Graph), we perform inference on datasets (Chest-X-Ray and ISIC2018) with significant modality discrepancies using the model pre-trained on Abd-CT. As shown in Fig. 12, our GPS outperforms previous methods on these hard cross-domain tasks. The core reason lies in that out GPS robustly captures the consistent geometric shape and peripheral-background distribution across disparate domains, providing intuitive evidence of the superior robustness and adaptability of GPS under diverse and challenging domain shifts.

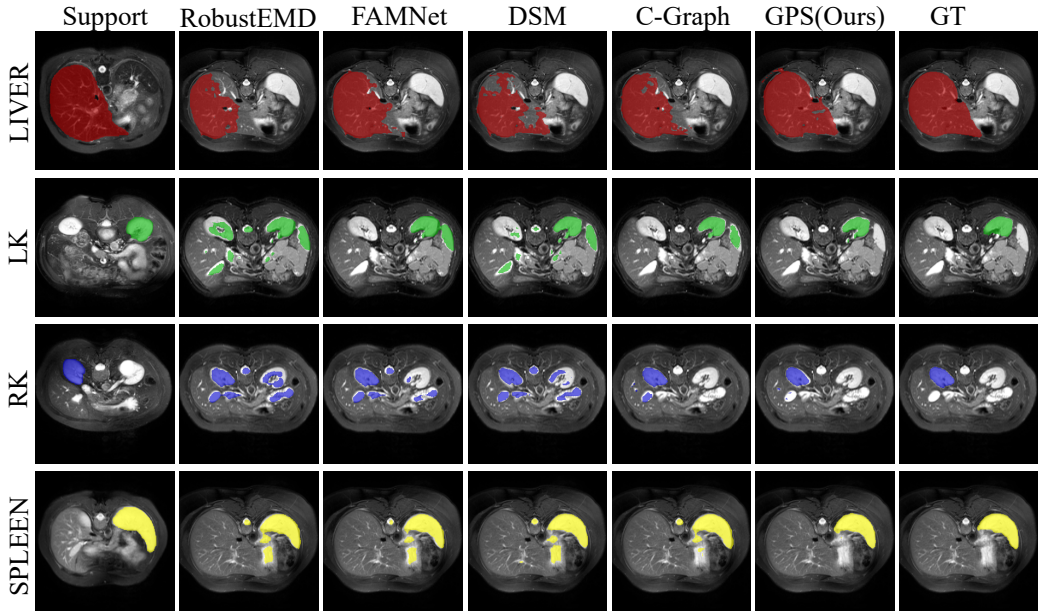

Figure 6: Qualitative comparison between our GPS and other methods on the Cross-Modality dataset (Abdomen CT $\rightarrow$ MRI). From top to bottom, the rows correspond to Liver, Left Kidney (LK), Right Kidney (RK) and Spleen.

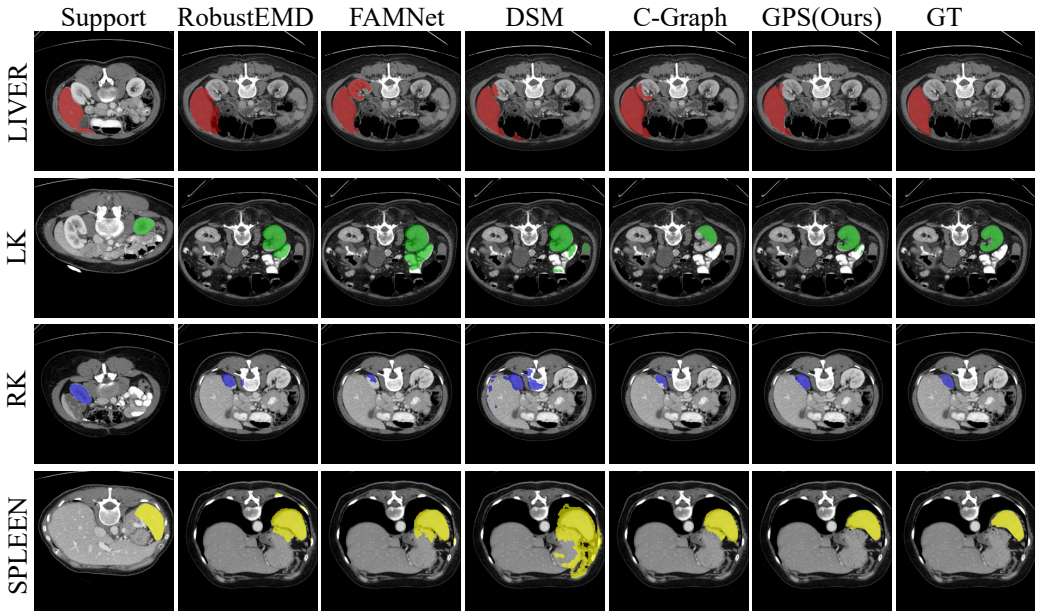

Figure 7: Qualitative comparison between our GPS and other methods on the Cross-Modality dataset (Abdomen MRI $\rightarrow$ CT). From top to bottom, the rows correspond to Liver, Left Kidney (LK), Right Kidney (RK) and Spleen.

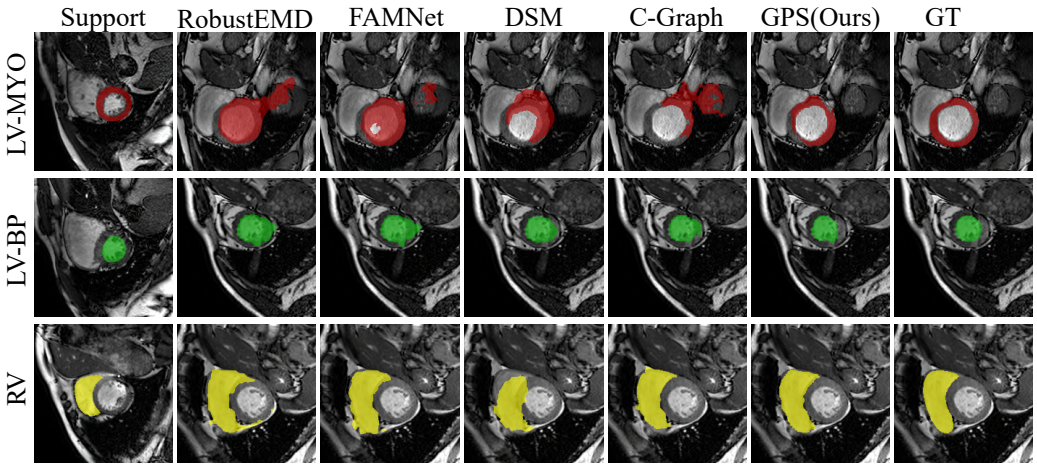

Figure 8: Qualitative comparison between our GPS and other methods on the Cross-Sequence dataset (Cardiac LGE $\rightarrow$ b-SSFP). From top to bottom, the rows correspond to the Left Ventricle Myocardium (LV-MYO), Left Ventricle Blood Pool (LV-BP) and Right Ventricle (RV).

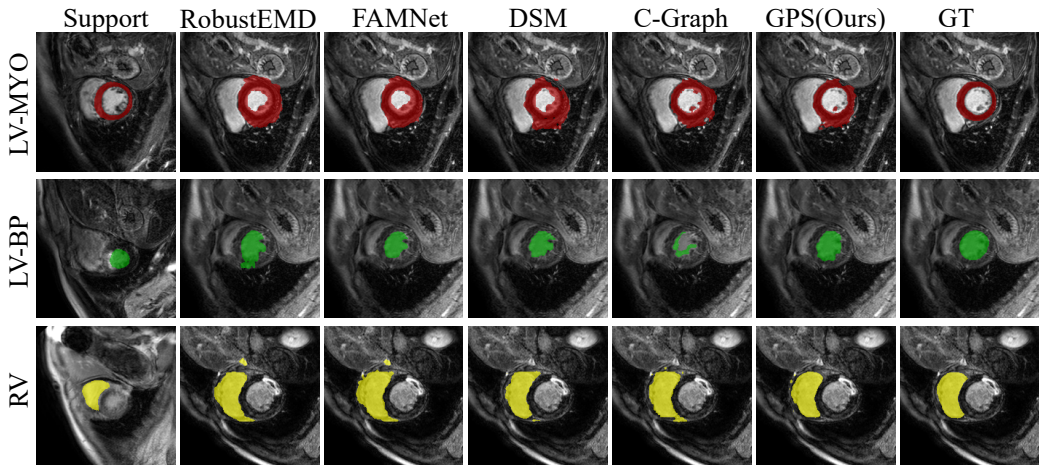

Figure 9: Qualitative comparison between our GPS and other methods on the Cross-Sequence dataset (Cardiac b-SSFP $\rightarrow$ LGE). From top to bottom, the rows correspond to the Left Ventricle Myocardium (LV-MYO), Left Ventricle Blood Pool (LV-BP) and Right Ventricle (RV).

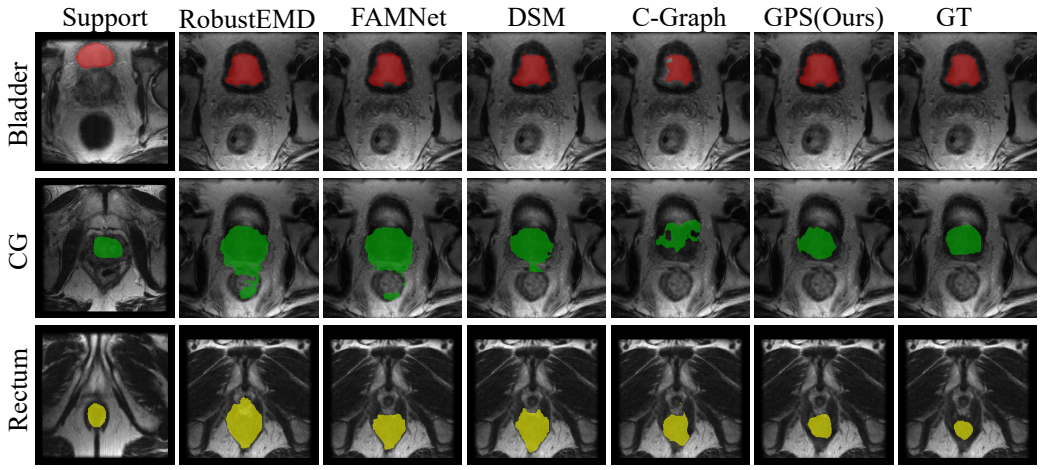

Figure 10: Qualitative comparison between our GPS and other methods on the Cross-Institution dataset (Prostate UCLH $\rightarrow$ NCI). From top to bottom, the rows correspond to Bladder, Central Gland, and Rectum.

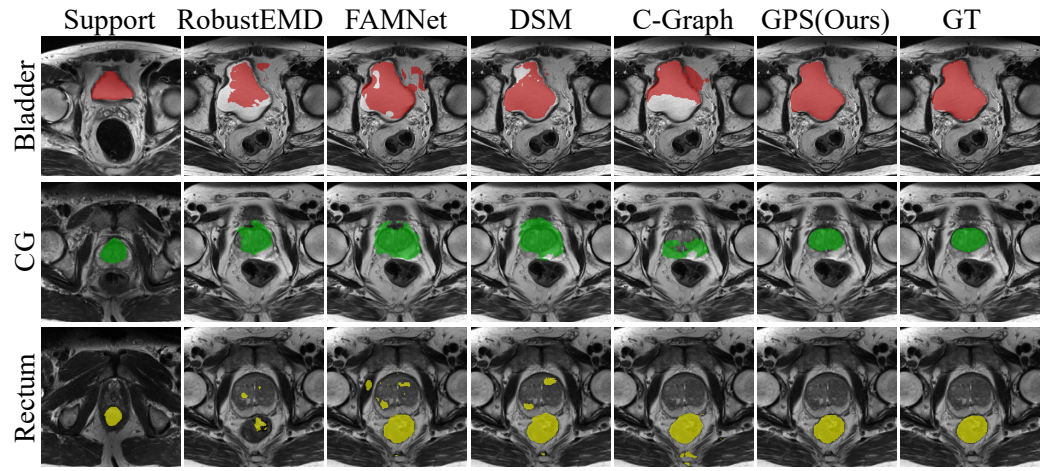

Figure 11: Qualitative comparison between our GPS and other methods on the Cross-Institution dataset (Prostate NCI → UCLH). From top to bottom, the rows correspond to Bladder, Central Gland, and Rectum.

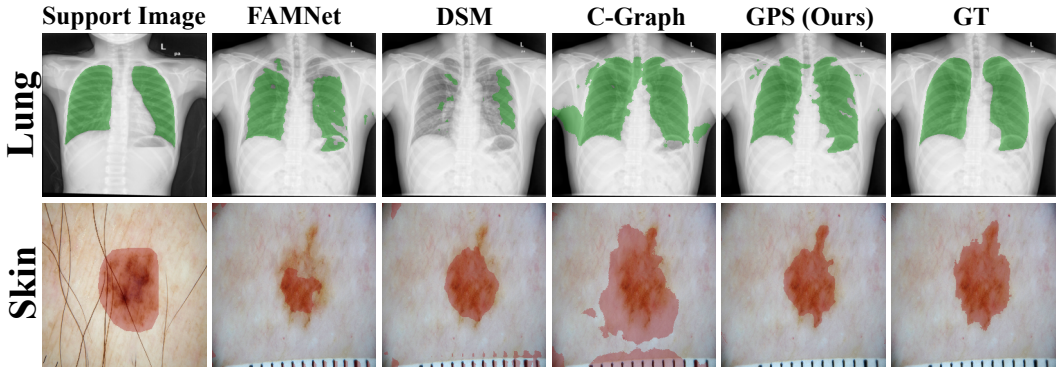

Figure 12: Visual qualitative comparison between our proposed GPS and other methods for the Abd-CT → Chest-X-ray and Abd-CT → ISIC2018 cross-modal translation tasks. Top to bottom: lung examples from Chest-X-ray and skin examples from ISIC2018. Left to right: labeled support image, predicted outputs generated by FAMNet, DSM, C-Graph, our GPS, and ground truth (GT).