PSP: Harnessing Position and Shape Priors for Cross-Domain Few-Shot Medical Image Segmentation

Bin Xu[0009−0009−2667−0663], Yazhou Zhu[0000−0002−7537−5945], and Haofeng Zhang(✉)[0000−0002−4039−7618]

School of Computer Science and Engineering, Nanjing University of Science and Technology, Nanjing 210094, China.
{xubin_work,zyz_nj,zhanghf}@njust.edu.cn

Abstract. Few-Shot Medical Image Segmentation (FSMIS) offers a powerful solution to data scarcity but struggles to generalize across different imaging modalities. This performance collapse stems primarily from the drastic texture discrepancies between domains, which mislead models trained on source-specific intensity distributions. While existing methods attempt to align frequency or local texture features, they often fail to decouple semantic structure from domain-specific appearance. To address this, we identify a critical invariance: despite distinct imaging physics, the position and geometric shape of organs remain robustly consistent across modalities. Therefore, we propose a novel framework that harnesses Position and Shape Priors (PSP) for cross-domain FSMIS. Specifically, PSP first introduces a Position Coordinate Embedding (PCE) module to inject relative spatial coordinates for rapid organ localization. Subsequently, a Shape Prototype Modulation (SPM) module constructs domain-invariant structural prototypes via explicit shape priors, effectively filtering out texture noise. Furthermore, the Hybrid-Prototype Prediction (HPP) module adaptively calibrates the support prototype to the query feature distribution, mitigating feature misalignment. Extensive experiments on two public medical imaging datasets demonstrate that PSP significantly outperforms state-of-the-art methods. Code is available at https://github.com/xubin471/PSP.

Keywords: Medical Image Segmentation · Cross-domain · Few-shot Learning · Position and Shape Priors.

1 Introduction

Automated medical image segmentation has significantly improved the advancement of computer-aided diagnosis[17] and therapy [2]. However, it relies heavily on annotated data, which is scarce due to high costs and privacy concerns. Few-Shot Medical Image Segmentation (FSMIS) [6, 19, 3, 8, 24, 20] alleviates this by generalizing from limited samples. However, practical deployment faces a critical challenge: the domain shift [4] (e.g., training on MRI vs. inferring on CT). Drastic texture discrepancies between modalities inevitably degrade performance.

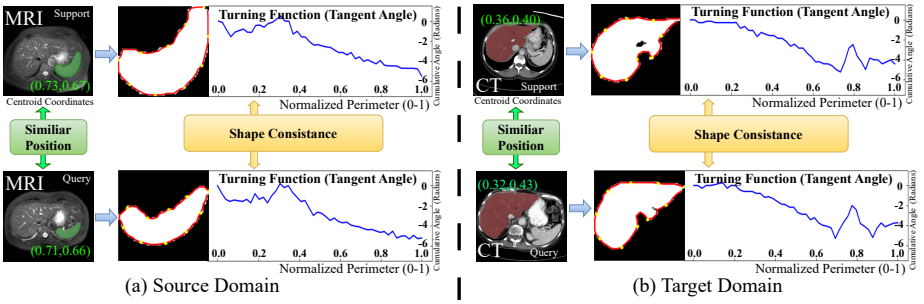

Fig. 1. Regardless of the domain, the support and query images exhibit high consistency in position (centroid coordinates) and shape (visualized by Turning Functions).

To address this domain shift, Cross-Domain FSMIS (CD-FSMIS) methods [22, 25, 23] have emerged to enhance generalization across heterogeneous imaging scenarios. Among these methods: FAMNet [1] utilizes frequency domain generalization but often incurs high-frequency noise and semantic loss. RobustEMD [23] introduces Earth Mover's Distance [16] for local matching. But its reliance on low-level texture similarity makes it vulnerable to drastic domain shifts (e.g., MRI to CT), leading to matching failure and performance degradation.

Notably, we observe a key medical characteristic: although imaging textures vary drastically across modalities, the anatomical position and geometric shape of the same organ remain highly consistent between support and query images, as illustrated by the Spleen in MRI and Liver in CT in Fig. 1. This indicates that position and shape serve as ideal "cross-domain invariants". Effectively harnessing these anatomical priors can guide the model to break free from excessive reliance on domain-specific features, thereby achieving robust segmentation. Inspired by this, we propose the **PSP** (Position and Shape Priors) framework to explicitly capture these invariant features. Specifically, PSP consists of three core modules: (1) the Position Coordinate Embedding (**PCE**) module for embedding rapid positional information to endow features with spatial awareness. (2) the Shape Prototype Modulation (**SPM**) module for generating shape-aware prototypes using explicit shape priors. (3) the Hybrid-Prototype Prediction (**HPP**) module for mitigating feature distribution misalignment between support and query images. In summary, our contributions are as follows:

- We observe that while imaging textures vary drastically across modalities, the position and geometric shape of organs exhibit robust consistency.
- We propose a novel framework, named Harnessing Position and Shape Priors (**PSP**) for CD-FSMIS, which consists of three core modules: the **PCE** module, the **SPM** module , and the **HPP** module.
- Extensive experiments conducted on two popular medical image datasets reveal that our proposed method outperforms other existing state-of-the-art (SOTA) methods and exhibits significant performance.

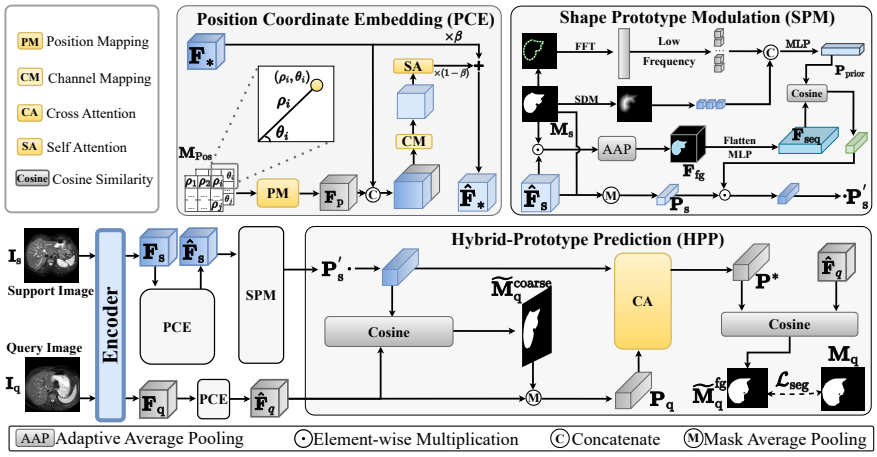

Fig. 2. Overview of our proposed CD-FSMIS method **PSP**.

2 Methodology

2.1 Overview

Given support and query images $\mathbf{I}_\mathrm{s}, \mathbf{I}_\mathrm{q}$, a weight-sharing ResNet-50 encoder $f_\theta(\cdot)$ pretrained on MS-COCO [12] extracts features $\mathbf{F}_\mathrm{s}, \mathbf{F}_\mathrm{q} \in \mathbb{R}^{C\times H\times W}$. First, the **PCE** module injects relative positions into these features via self-attention to discriminate spatially distinct features. Next, **SPM** utilizes the support mask $\mathbf{M}_\mathrm{s} \in \mathbb{R}^{H\times W}$ to construct shape-aware prototypes. Finally, **HPP** calibrates the prototype to the query distribution via a coarse-to-fine strategy, generating the final segmentation $\widetilde{\mathbf{M}}_\mathrm{q}^{\mathrm{fg}} \in \mathbb{R}^{H\times W}$ through cosine similarity. Fig. 2 illustrates the overall architecture.

2.2 Position Coordinate Embedding (PCE)

The PCE module injects explicit spatial positions into the feature representation to mitigate texture reliance. Taking the feature $\mathbf{F}_* \in \mathbb{R}^{C\times H\times W}$ as input, where $\mathbf{F}_*$ denotes either the support feature $\mathbf{F}_\mathrm{s}$ or the query feature $\mathbf{F}_\mathrm{q}$, we first construct a center-based polar coordinate system. For each pixel (h, w), we compute the normalized radial distance ρ and polar angle θ:

$$\rho_{h,w} = \sqrt{\bar{h}^2 + \bar{w}^2}, \quad \theta_{h,w} = \arctan2(\bar{w}, \bar{h}), \tag{1}$$

where $(\bar{h}, \bar{w})$ represent relative coordinates centered at the image. These are stacked to form a coordinate map $\mathbf{M}_\mathrm{pos} \in \mathbb{R}^{2\times H\times W}$, which is projected into the high-dimensional feature space via a Position Mapping (PM) module (a 2-layer MLP), yielding the position embedding $\mathbf{F}_\mathrm{p} \in \mathbb{R}^{C\times H\times W}$.

Subsequently, we concatenate $\mathbf{F}_*$ and $\mathbf{F}_\mathrm{p}$ and fuse them using a Channel Mapping (CM) module (1×1 convolution) to restore the channel dimension

C. The fused feature is processed by a Self-Attention (SA) module to discriminate features that are semantically similar but spatially distinct. Finally, we can obtain the position-enhanced feature $\hat{\mathbf{F}}_* \in \mathbb{R}^{C\times H\times W}$ by employing a dynamic residual connection to balance semantic and spatial information:

$$\hat{\mathbf{F}}_* = (1-\beta)\cdot \mathbf{F}_* + \beta \cdot \mathrm{SA}(\mathrm{CM}(\mathrm{Concat}(\mathbf{F}_*, \mathbf{F}_\mathrm{p}))), \quad \mathbf{F}_* \in \{\mathbf{F}_\mathrm{s}, \mathbf{F}_\mathrm{q}\}, \tag{2}$$

where $\mathrm{Concat}(\cdot)$ denotes the concatenation operation along the channel dimension; $\mathrm{CM}(\cdot)$ denotes the 1×1 convolution; $\mathrm{SA}(\cdot)$ denotes the self-attention operation; β is a learnable parameter initialized to 0.1;

2.3 Shape Prototype Modulation (SPM)

SPM incorporates cross-domain invariant shape priors to modulate prototypes, effectively decoupling structure from texture noise. Taking the enhanced support feature $\hat{\mathbf{F}}_\mathrm{s}$ and the support mask $\mathbf{M}_\mathrm{s}$ as inputs, the process consists of Explicit Shape Prior Encoding, Feature-Shape Alignment, and Prototype Modulation.

Explicit Shape Prior Encoding. We construct a high-dimensional shape representation $\mathbf{P}_\mathrm{prior} \in \mathbb{R}^{1\times C_s}$ by combining geometric statistics and spectral descriptors. First, regarding Geometric Statistics, we compute the Signed Distance Map [21] (SDM) on $\mathbf{M}_\mathrm{s}$, where each pixel value represents the Euclidean distance to the nearest background. We extract the maximum, mean, and variance values from the foreground region of the SDM to form a geometric vector $v_{geo} \in \mathbb{R}^3$. Second, for Spectral Shape Descriptors, we employ Fourier Descriptors [7] to capture global topology. We extract boundary coordinates $\{(x_m, y_m)\}_{m=0}^{M-1}$ from $\mathbf{M}_\mathrm{s}$, convert them into complex signals $z_m = x_m + iy_m$, and apply the Discrete Fourier Transform [11] (DFT) to obtain the sequence of coefficients $\{Z_u \in \mathbb{C}\}_{u=0}^{M-1}$. To filter high-frequency noise and ensure geometric invariance, the spectral vector $v_{spec} \in \mathbb{R}^{N_{spec}}$ is formulated as follows:

$$v_{spec}(i) = \frac{|Z_{i+1}|}{|Z_1|}, \quad i = 0, \ldots, N_{spec} - 1, \tag{3}$$

where $|\cdot|$ denotes the magnitude of the complex coefficient. In this formulation, omitting the direct current component Z_0 ensures translation invariance, normalizing by the fundamental component $|Z_1|$ achieves scale invariance, and using the magnitude guarantees rotation invariance. N_{spec} denotes the number of retained low-frequency components. Finally, the $\mathbf{P}_\mathrm{prior}$ is formulated as:

$$\mathbf{P}_\mathrm{prior} = \mathrm{MLP}([\mathrm{Concat}(v_{geo}, v_{spec})]) \in \mathbb{R}^{1\times C_s} \tag{4}$$

where $\mathrm{Concat}(\cdot)$ denotes the concatenation operation; $\mathrm{MLP}(\cdot)$ denotes a multilayer perceptron mapping $\mathbb{R}^{3+N_{spec}} \to \mathbb{R}^{C_s}$.

Feature-Shape Alignment. Next, we estimate the relevance of each feature channel to the explicit shape prior. We mask $\hat{\mathbf{F}}_\mathrm{s}$ with $\mathbf{M}_\mathrm{s}$ to isolate the foreground and apply Adaptive Average Pooling (AAP) to reduce spatial dimensions, obtaining $\mathbf{F}_\mathrm{fg} \in \mathbb{R}^{C\times h\times w}$. This is flattened and projected to generate a sequence

of channel descriptors $\mathbf{F}_{\text{seq}} \in \mathbb{R}^{C \times C_s}$. The $\mathbf{F}_{\text{seq}}$ is formulated as follows:

$$\mathbf{F}_{\text{seq}} = \text{MLP}(\text{Flatten}(\mathbf{F}_{\text{fg}})), \quad \mathbf{F}_{\text{fg}} = \text{AAP}(\hat{\mathbf{F}}_{\text{s}} \odot \mathbf{M}_{\text{s}}) \tag{5}$$

where $\odot$ denotes the operation of element-wise multiplication, $\text{AAP}(\cdot)$ denotes the adaptive average pooling operation, $\text{Flatten}(\cdot)$ 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_s.

Subsequently, we compute the channel-wise attention weights $\mathbf{w} \in \mathbb{R}^C$ by measuring the cosine similarity between each channel descriptor and the shape prior. To quantify the probability that the i-th channel encodes shape-relevant information, we apply a Sigmoid activation to the similarity score:

$$\mathbf{w}_i = \text{Relu}\left(\frac{\mathbf{F}_{\text{seq}}^{(i)} \cdot \mathbf{P}_{\text{prior}}}{\|\mathbf{F}_{\text{seq}}^{(i)}\|_2 \|\mathbf{P}_{\text{prior}}\|_2}\right), \tag{6}$$

where $\text{Relu}(\cdot)$ is the Rectified Linear Unit, and $\|\cdot\|_2$ represents the L_2 norm.

Prototype Modulation. Next, we obtain the initial global prototype $\mathbf{P}_{\text{s}} \in \mathbb{R}^{1 \times C}$ via Masked Average Pooling on the support features. Finally, we modulate the shape prototype $\mathbf{P}_{\text{s}}^{'} \in \mathbb{R}^{1 \times C}$ using the learned attention weights $\mathbf{w}$:

$$\mathbf{P}_{\text{s}}^{'} = \mathbf{P}_{\text{s}} \odot \mathbf{w}, \quad \mathbf{P}_{\text{s}} = \text{MAP}(\hat{\mathbf{F}}_{\text{s}}, \mathbf{M}_{\text{s}}), \tag{7}$$

where $\odot$ denotes element-wise multiplication and MAP is mask average pooling.

2.4 Hybrid-Prototype Prediction (HPP)

To mitigate feature misalignment caused by intra-class variance between support and query images, HPP employs a coarse-to-fine strategy to calibrate the support prototype towards the query distribution.

First, we generate a coarse similarity map $\widetilde{\mathbf{M}}_{\text{q}}^{\text{coar}}$ using the shape-modulated prototype $\mathbf{P}'_{\text{s}}$. This map guides Masked Average Pooling (MAP) on the enhanced query features $\hat{\mathbf{F}}_{\text{q}}$ to extract a query-specific prototype $\mathbf{P}_{\text{q}} \in \mathbb{R}^{1 \times C}$:

$$\mathbf{P}_{\text{q}} = \text{MAP}(\hat{\mathbf{F}}_{\text{q}}, \widetilde{\mathbf{M}}_{\text{q}}^{\text{coar}}), \quad \widetilde{\mathbf{M}}_{\text{q}}^{\text{coar}} = \text{Cosine}(\hat{\mathbf{F}}_{\text{q}}, \mathbf{P}'_{\text{s}}), \tag{8}$$

where $\text{Cosine}(\cdot, \cdot)$ denotes computing cosine similarity along the channel dimension. $\text{MAP}(\cdot)$ denotes mask average pooling operation.

Next, we fuse the shape prototype $\mathbf{P}'_{\text{s}}$ with the query-specific prototype $\mathbf{P}_{\text{q}}$ via Cross-Attention (CA) with a residual connection to obtain calibrated hybrid prototype $\mathbf{p}^*$. Here, $\mathbf{P}'_{\text{s}}$ acts as the query, while $\mathbf{P}_{\text{q}}$ serves as the key and value:

$$\mathbf{P}^* = \mathbf{P}'_{\text{s}} + \text{CA}(\mathbf{P}'_{\text{s}}, \mathbf{P}_{\text{q}}, \mathbf{P}_{\text{q}}), \quad \mathbf{P}^* \in \mathbb{R}^{1 \times C}. \tag{9}$$

Finally, the foreground query prediction $\widetilde{\mathbf{M}}_{\text{q}} \in \mathbb{R}^{H \times W}$ is computed as follows:

$$\widetilde{\mathbf{M}}_{\text{q}}^{(i,j)} = 1 - \sigma\left(-\alpha \, \text{Cosine}(\hat{\mathbf{F}}_{\text{q}}^{(i,j)}, \mathbf{P}^*) - \tau\right), \tag{10}$$

where $\text{Cosine}(\cdot, \cdot)$ denotes cosine similarity; $\sigma(\cdot)$ denotes the sigmoid activation; We set scaling factor $\alpha = 20$ to amplify the similarity distinction between foreground and background, and τ is an adaptive threshold predicted by two fully-connected layers $\text{FC}(\cdot)$ based on $\hat{\mathbf{F}}_{\text{q}}$, $\tau = \text{FC}(\hat{\mathbf{F}}_{\text{q}})$.

Table 1. Quantitative comparison (in Dice scores %) of different methods on the Cross-Modality Dataset. Best: **bold**, second best: underlined.

Method	Ref.	Abd CT → MRI					Abd MRI → CT				
		Liver	LK	RK	Spleen	Mean	Liver	LK	RK	Spleen	Mean
SSL-ALP	TMI'22	70.74	55.49	67.43	58.39	63.01	71.38	34.48	32.32	51.67	47.46
ADNet	MIA'22	50.33	39.36	37.88	39.37	41.73	64.25	37.39	25.62	42.94	42.55
PATNet	ECCV'22	57.01	50.23	53.01	51.63	52.97	75.94	46.62	42.68	63.94	57.29
CATNet	MICCAI'23	44.58	43.67	50.27	46.34	46.21	54.52	41.73	40.24	45.84	45.60
RPT	MICCAI'23	49.22	42.45	47.14	48.84	46.91	65.87	40.07	35.97	51.22	48.28
IFA	CVPR'24	48.81	45.79	51.46	51.42	49.37	50.05	36.45	32.69	43.08	40.57
RobustEMD	AIIM'25	60.16	66.34	70.26	53.71	62.61	69.82	63.79	50.34	59.88	60.95
FAMNet	AAAI'25	**73.01**	57.28	74.68	58.21	65.79	73.57	57.79	61.89	65.78	64.75
DSM	TIP'25	72.94	61.59	69.52	**59.00**	65.76	**77.69**	56.60	56.45	59.63	62.59
PSP(Ours)		70.24	**69.96**	**78.70**	58.56	**69.36**	73.44	**64.48**	**69.17**	**65.91**	**68.25**

2.5 Loss Function

We quantify the segmentation dissimilarity using the Cross-Entropy loss $\mathcal{L}_{\text{prim}}$ between the predicted mask $\widetilde{\mathbf{M}}_{\text{q}}$ and ground truth $\mathbf{M}_{\text{q}} \in \mathbb{R}^{H\times W}$:

$$\mathcal{L}_{\text{prim}} = -\frac{1}{HW}\sum\nolimits_{i,j}\left[\mathbf{M}_{\text{q}}^{(i,j)}\log\widetilde{\mathbf{M}}_{\text{q}}^{(i,j)} + (1-\mathbf{M}_{\text{q}}^{(i,j)})\log(1-\widetilde{\mathbf{M}}_{\text{q}}^{(i,j)})\right]. \quad (11)$$

Additionally, we adapt the Alignment Loss [18] to infer the support prediction $\widetilde{\mathbf{M}}_{\text{s}} \in \mathbb{R}^{H\times W}$ using the query features and predicted masks. This reverse inference is regularized by the Cross-Entropy loss:

$$\mathcal{L}_{\text{align}} = -\frac{1}{HW}\sum\nolimits_{i,j}\left[\mathbf{M}_{\text{s}}^{(i,j)}\log\widetilde{\mathbf{M}}_{\text{s}}^{(i,j)} + (1-\mathbf{M}_{\text{s}}^{(i,j)})\log(1-\widetilde{\mathbf{M}}_{\text{s}}^{(i,j)})\right]. \quad (12)$$

Finally, the total training loss is set as $\mathcal{L} = \mathcal{L}_{\text{prim}} + \mathcal{L}_{\text{align}}$.

3 Experiment

3.1 Datasets and Implementation Details

Datasets. We evaluate our PSP on two public datasets under domain shift scenarios. Specifically, **Cross-Modality**: We utilize the Abdominal dataset comprising 20 3D T2-SPIR MRI volumes [9] and 20 CT scans [10]. We perform segmentation on four organs: liver, spleen, right kidney (RK), and left kidney (LK). **Cross-Sequence**: this dataset [26] contains 45 3D b-SSFP and 45 LGE MRI scans. We perform segmentation on three organs: the left ventricle myocardium (LV-MYO), left ventricle blood pool (LV-BP), and right ventricle (RV).

Implementation Details. To evaluate cross-domain generalization, we conduct bidirectional transfer experiments for all datasets (e.g., Abd-MRI ↔ Abd-CT). The model is trained for 6 epochs (5,000 iterations per epoch) using an SGD optimizer with an initial learning rate of 10^{-3}, momentum of 0.9, and a decay of 0.98 every 1,000 iterations. For **PSP**-specific hyperparameters, the number of low-frequency components N_{spec} in SPM module is set to 15. The scaling factor in HPP module α is set to 20.

Table 2. Quantitative comparison of Dice scores (%) of different methods on the Cross-Sequence Dataset. Best: **bold**, second best: underlined.

Method	Ref.	Cardiac LGE → b-SSFP				Cardiac b-SSFP → LGE			
		LV-BP	LV-MYO	RV	Mean	LV-BP	LV-MYO	RV	Mean
SSL-ALP	TMI'22	83.47	22.73	66.21	57.47	65.81	25.64	51.24	47.56
ADNet	MIA'22	58.75	36.94	51.37	49.02	40.36	37.22	43.66	40.41
PATNet	ECCV'22	65.35	50.63	68.34	61.44	66.82	53.64	59.74	60.06
CATNet	MICCAI'23	64.63	42.41	56.13	54.39	45.77	43.51	46.02	45.10
RPT	MICCAI'23	60.84	42.28	57.30	53.47	50.39	40.13	50.50	47.00
IFA	CVPR'24	64.04	43.22	74.58	62.28	68.07	36.07	60.42	54.85
RobustEMD	AIIM'25	75.32	51.32	72.86	66.50	73.19	50.02	60.29	61.16
FAMNet	AAAI'25	86.64	51.84	76.26	71.58	**77.37**	52.05	54.75	61.39
DSM	TIP'25	85.27	50.74	73.20	69.74	71.27	53.62	**63.65**	62.85
PSP(Ours)		**90.26**	**61.30**	**84.33**	**78.63**	74.51	**56.41**	62.10	**64.34**

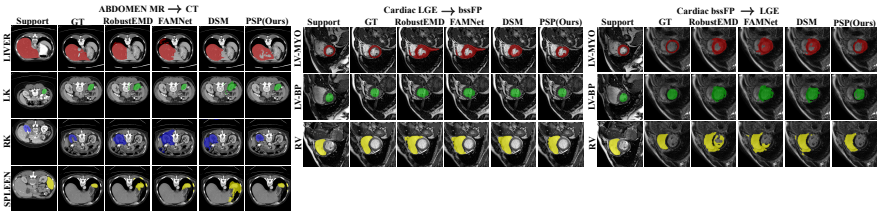

Fig. 3. The qualitative results of our model and compared models in tasks: Abd-MR → Abd-CT, Cardiac-LGE → Cardiac-bssFP and Cardiac bssFP → Cardiac-LGE.

3.2 Quantitative and Qualitative Results

Quantitative Results. To demonstrate the effectiveness of our proposed method, We benchmark PSP against extensive existing methods: SSL-ALP [15], ADNet [5], PATNet [18], CATNet [13], RPT [24], IFA [14], RobustEMD [23], FAMNet [1] and DSM [25]. The results are recorded in Table 1 and Table 2. As shown in Table 1, PSP establishes a new state-of-the-art on the Cross-Modality dataset. Specifically, our method outperforms the second-best approach by a substantial margin of 3.75% in the Abd-CT → Abd-MRI task and 3.50% in the Abd-MRI → Abd-CT task. This superiority extends to the Cross-Sequence dataset (Table 2), where PSP achieves consistent improvements of 1.49% and 7.05% over the nearest competitor on bSSFP → LGE and LGE → bSSFP tasks, respectively. These significant performance gains, strongly validate the effectiveness of our method: harnessing explicit position and shape priors provides significantly more robust guidance than relying on unstable texture features for CD-FSMIS.

Qualitative Results. Fig. 3 presents visual comparisons across cross-domain tasks. In contrast to the fragmented or noisy masks produced by RobustEMD and DSM, PSP generates significantly more accurate segmentation maps with high structural integrity. These results demonstrate that our proposed position and shape priors serve as reliable anchors, effectively guiding the model to alleviate the texture gap for robust generalization.

Table 3. Ablation study of key modules in the proposed PSP on task Abd-MR → CT in terms of mean Dice Scores (%).

PCE	SPM	HPP	Dice Score (%)
			62.08
✓			63.45
✓	✓		66.71
✓	✓	✓	**68.25**

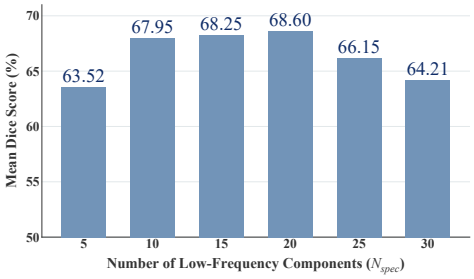

Fig. 4. Effect of the number of low-frequency components N_{spec} on task Abd-MR → CT.

3.3 Ablation Studies

Effect of each module. We validate the effectiveness of individual modules within our PSP framework, with results summarized in Table 3. Table 3 reports the incremental gains. The baseline yields 62.08% (mean Dice Score). Introducing PCE module improves the score to 63.45% by providing positional anchors. Notably, SPM module delivers the most significant gain of 3.26% (reaching 66.71%), validating the importance of shape priors in filtering texture noise. Finally, HPP module further boosts performance to 68.25% by mitigating feature misalignment between support and query images.

Effect of the number of low frequency components. We further investigate the sensitivity of the number of low-frequency components N_{spec} in the SPM module, as illustrated in Fig. 4. The performance improves as N_{spec} increases to 20, confirming that low-frequency components capture essential anatomical shape topology. Conversely, the performance drops when $N_{spec} > 20$, implying that excessive high-frequency components lead to overfitting on local boundary variations. This suggests that a compact shape representation is more robust for transferring knowledge across heterogeneous modalities.

4 Conclusion

In this paper, we proposed the **PSP** framework to address the generalization bottleneck in CD-FSMIS caused by drastic texture discrepancies. Diverging from existing approaches, PSP explicitly harnesses anatomical position and shape priors which remain invariant across domains. Specifically, the **PCE** module endows the model with spatial awareness for rapid anatomical localization; the **SPM** module constructs shape-aware prototypes via geometric descriptors to decouple structural information from domain-specific texture noise; and the **HPP** module bridges the feature distribution gap through a coarse-to-fine calibration strategy. Extensive experiments on abdominal and cardiac datasets demonstrate that PSP significantly outperforms state-of-the-art methods, validating the efficacy of explicit anatomical position and shape priors in mitigating domain shifts.

Acknowledgments. This work was supported in part by the National Natural Science Foundation of China under the Grants No. 62371235 and U25A20444, in part by the Key Research and Development Plan of Jiangsu Province (Industry Foresight and Key Core Technology Project) under Grant No. BE2023008-2. Thanks to the KinaMind society for their inspiring environment and unwavering support.

Disclosure of Interests. All authors declare that they have no conflicts of interest.

References

1. Bo, Y., Zhu, Y., Li, L., Zhang, H.: Famnet: Frequency-aware matching network for cross-domain few-shot medical image segmentation. In: AAAI. vol. 39, pp. 1889–1897 (2025)
2. Chen, X., Sun, S., Bai, N., Han, K., Liu, Q., 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)
3. Cheng, Z., Wang, S., Long, Y., Zhou, T., Zhang, H., Shao, L.: Dual interspersion and flexible deployment for few-shot medical image segmentation. IEEE Transactions on Medical Imaging **44**(6), 2732–2744 (2025)
4. Guo, L.L., Pfohl, S.R., Fries, J., Johnson, A.E., Posada, J., Aftandilian, C., Shah, N., Sung, L.: Evaluation of domain generalization and adaptation on improving model robustness to temporal dataset shift in clinical medicine. Scientific reports **12**(1), 2726 (2022)
5. Hansen, S., Gautam, S., Jenssen, R., Kampffmeyer, M.: Anomaly detection-inspired few-shot medical image segmentation through self-supervision with supervoxels. Medical Image Analysis **78**, 102385 (2022)
6. Huang, W., Xiao, B., Hu, J., Bi, X.: Location-aware transformer network for few-shot medical image segmentation. In: 2023 IEEE International Conference on Bioinformatics and Biomedicine (BIBM). pp. 1150–1157 (2023)
7. Jeon, Y.S., Yang, H., Feng, M.: Fcsn: Global context aware segmentation by learning the fourier coefficients of objects in medical images. IEEE Journal of Biomedical and Health Informatics **28**(3), 1195–1206 (2022)
8. Jiang, J., Zhang, H.: 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)
9. Kavur, A.E., Gezer, N.S., Barış, M., Aslan, S., Conze, P.H., Groza, V., Pham, D.D., Chatterjee, S., Ernst, P., Özkan, S., et al.: Chaos challenge-combined (ct-mr) healthy abdominal organ segmentation. Medical image analysis **69**, 101950 (2021)
10. Landman, B., Xu, Z., Igelsias, J., Styner, M., Langerak, T., Klein, A.: Miccai multi-atlas labeling beyond the cranial vault–workshop and challenge. In: MICCAI Workshop Challenge. vol. 5, p. 12 (2015)
11. Li, P., Zhou, R., He, J., Zhao, S., Tian, Y.: A global-frequency-domain network for medical image segmentation. Computers in Biology and Medicine **164**, 107290 (2023)
12. Lin, T.Y., Maire, M., Belongie, S., Hays, J., Perona, P., Ramanan, D., Dollár, P., Zitnick, C.L.: Microsoft coco: Common objects in context. In: ECCV. pp. 740–755 (2014)

13. Lin, Y., Chen, Y., Cheng, K.T., Chen, H.: Few shot medical image segmentation with cross attention transformer. In: MICCAI. pp. 233–243 (2023)
14. Nie, J., Xing, Y., Zhang, G., Yan, P., Xiao, A., Tan, Y.P., Kot, A.C., Lu, S.: Cross-domain few-shot segmentation via iterative support-query correspondence mining. In: CVPR. pp. 3380–3390 (2024)
15. Ouyang, C., Biffi, C., Chen, C., Kart, T., Qiu, H., Rueckert, D.: Self-supervised learning for few-shot medical image segmentation. IEEE Transactions on Medical Imaging **41**(7), 1837–1848 (2022)
16. Rubner, Y., Tomasi, C., Guibas, L.J.: The earth mover's distance as a metric for image retrieval. International journal of computer vision **40**(2), 99–121 (2000)
17. Tsochatzidis, L., Koutla, P., Costaridou, L., Pratikakis, I.: Integrating segmentation information into cnn for breast cancer diagnosis of mammographic masses. Computer Methods and Programs in Biomedicine **200**, 105913 (2021)
18. Wang, K., Liew, J.H., Zou, Y., Zhou, D., Feng, J.: Panet: Few-shot image semantic segmentation with prototype alignment. In: ICCV. pp. 9197–9206 (2019)
19. Xiao, B., Hu, J., Li, W., Pun, C.M., Bi, X.: Ctnet: Contrastive transformer network for polyp segmentation. IEEE Transactions on Cybernetics **54**(9), 5040–5053 (2024)
20. Xu, B., Zhu, Y., Wang, S., Long, Y., Zhang, H.: Few-shot medical image segmentation via boundary-extended prototypes and momentum inference. Computer Vision and Image Understanding **263**, 104571 (2026)
21. Xue, Y., Tang, H., Qiao, Z., Gong, G., Yin, Y., Qian, Z., Huang, C., Fan, W., Huang, X.: Shape-aware organ segmentation by predicting signed distance maps. In: Proceedings of the AAAI conference on artificial intelligence. vol. 34, pp. 12565–12572 (2020)
22. Zhou, W., Guan, G., Yan, Q., Zhao, Y.: 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)
23. Zhu, Y., Li, M., Ye, Q., Wang, S., Xin, T., Zhang, H.: Robustemd: Domain robust matching for cross-domain few-shot medical image segmentation. Artificial Intelligence in Medicine p. 103197 (2025)
24. Zhu, Y., Wang, S., Xin, T., Zhang, H.: Few-shot medical image segmentation via a region-enhanced prototypical transformer. In: MICCAI. pp. 271–280 (2023)
25. Zhu, Y., Wang, S., Zhou, T., Li, Z., Zhang, H., Shao, L.: Cross-domain few-shot medical image segmentation via dynamic semantic matching. IEEE Transactions on Image Processing **34**, 6669–6682 (2025)
26. Zhuang, X., Xu, J., Luo, X., Chen, C., Ouyang, C., 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)