Energy-guided Dual Domain-invariant Prompting Framework with Fourier Regularization for Generalized Few-shot Medical Segmentation

Shaolei Liu[1], Yuting Wu[12], Dongchen Zhu[12*], Jiamao Li[12]

[1] Bio-vision System Laboratory, Science and Technology on Micro-system Laboratory, Shanghai Institute of Microsystem and Information Technology, Chinese Academy of Sciences, Shanghai, 200050, China
[2]University of Chinese Academy of Sciences, Beijing, 101408, China
{slliu@mail.sim.ac.cn, dczhu@mail.sim.ac.cn}

Abstract

Precise segmentation of organ and tissue lesions is essential for clinical diagnosis and treatment. Despite the progress of deep learning and foundation segmentation models, their domain generalization capability remains limited particularly when dealing with cross-domain scenarios or unseen data, leading to significant performance degradation. Current medical SAM-based generalization methods face two primary challenges: First, existing prompt-tuning strategies inadequately capture key domain-invariant features; Second, the reliance on fully labeled source domain data is unrealistic in clinical practice. To address these challenges, we propose a novel Dual domain-Invariant Prompt Optimization (DIPO) enhanced by energy-guided augmentation and frequency consistency regularization for few-shot medical image segmentation generalization. Our approach introduces a multi-band momentum enhancement strategy to dynamically augment source data by leveraging diverse frequency bands of the Fourier amplitude spectrum. Furthermore, we integrate multi-scale geometric representation-based non-subsampled shearlet transform and text prompts to strengthen the extraction of shape- and texture-related domain-invariant features. Finally, we employ frequency consistency regularization to refine model robustness using predictions from unlabeled data. Experimental results in prostate and fundus datasets demonstrate that our method significantly outperforms current state-of-the-art methods.

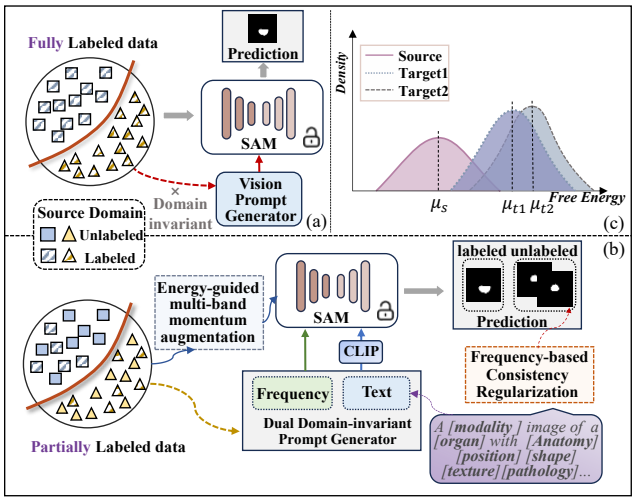

Figure 1: Conceptual overview of our motivation. (a) The previous approach lies in how to fine-tune the SAM model in fully labeled source domain. (b) Our energy-guided multi-band dynamic enhancement and a dual domain-invariant optimization strategy in more practical partially labeled scenarios. (c) Energy distribution and bias among different domains.

Introduction

Precise segmentation of organs and diseased tissues stands as a core task in medical image analysis, playing a crucial role in improving the accuracy of clinical diagnosis, formulating personalized treatment plans and implementing precise radiotherapy (Gao et al. 2025; Niu et al. 2024a). In recent years, deep learning-based models have made remarkable progress in this field, with segmentation performance approaching or even surpassing that of traditional methods (Rayed et al. 2024; Liu et al. 2023b). However, existing models commonly face a critical challenge: when significant distributional shifts exist between training data (source domain) and test data (target domain), or when encountering unseen out-of-distribution data, their segmentation performance degrades significantly due to insufficient model generalization capability, which hinders their robust application in real and diverse clinical practice (Liu, Dou, and Heng 2020; Liu et al. 2023a).

To address the challenge of insufficient model generalization, Domain Generalization (DG) methods have emerged (Liu, Dou, and Heng 2020). Their goal is to learn a robust model capable of performing well on unseen target domains, using data only from one or multiple source domains during training. A particularly realistic and challenging setting is Single-Domain Generalization (SDG), where models must generalize robustly across multiple unseen target domains using data from only a single source domain (Liu, Dou, and Heng 2020). SDG not only substantially alleviates the burden of collecting labeled data, but also addresses practical constraints in data collection and privacy protection.

Recently, the foundation segmentation model (Segment

*Corresponding author.

Anything Model, SAM) (Kirillov et al. 2023; Ali et al. 2025) has drawn significant attention due to its remarkable zero-shot generalization capabilities demonstrated on natural images. Naturally, researchers are exploring ways to transfer its impressive capabilities to medical image segmentation by prompt tuning and model fine-tuning strategies. For instance, COSAM (Fu et al. 2024) proposed a self-correcting SAM that refines coarse masks using preliminary prompts and introduced a universal error decoder to simulate manual corrections, thereby enhancing its adaptability to unseen data. DAPSAM (Wei et al. 2024) leveraged a prototype-based memory bank learned from source domain images to propose a domain-adaptive prompt tuning strategy. Moreover, a new adapter was further designed by integrating cross-layer features to improve the robustness of the SAM. MoSE (Wei et al. 2025) employed a Mixed Shape Expert framework and dictionary learning, effectively capturing diverse and robust shape priors through a dynamic gating mechanism to enhance generalization capability. These works have successfully introduced SAM to medical image processing and demonstrated improved performance.

However, as shown in Fig.1, two critical issues remain unresolved in current SAM-based medical segmentation generalization research. **1. Insufficient domain-invariant feature learning.** The key objective of SDG is to learn domain-invariant feature representations that exhibit consistency and robustness in different domains. However, existing prompt-tuning strategies primarily focus on leveraging prompts to guide the model to learn target-related information, lacking direct design specifically to facilitate learning of these crucial domain-invariant features. **2. Impractical assumption of fully labeled source domain data.** Current methods assume that the source domain training data is fully annotated. However, obtaining a large amount of pixel-level segmentation annotations is extremely challenging in real clinical practice due to the time and expert knowledge, which severely limits their practical applicability.

To address the above problems, we propose a novel energy-guided few-shot medical image segmentation generalization framework based on Dual domain-Invariant Prompt Optimization (DIPO), multi-band momentum enhancement and frequency consistency regularization. To enable the source domain to incorporate more diverse variations of the target domain and effectively utilize unlabeled data, we propose an energy-guided multi-band momentum enhancement strategy to dynamically augment the source data at low, medium and high frequency bands of the Fourier amplitude spectrum, respectively, reducing the domain gap between the source domain and the unseen target domain. Furthermore, to encourage the model to learn domain-invariant features related to organ shapes and textures, we introduce a dual domain-invariant prompt optimization strategy based on Non-Subsampled Shearlet Transform (NSST) (Easley, Labate, and Lim 2008) and text prompts. We leverage the multi-scale geometric representation capability of NSST and combine with frequency self-attention mechanisms (Deng et al. 2025) to extract cross-domain invariant features, while text prompts further enhance generalization. Finally, to address the issue of insufficient labeling of source domain data, we impose frequency consistency regularization on the unlabeled source domain data after energy-guided momentum enhancement, further improving the model robustness.

The main contributions are as follows:

- From the perspective of energy-guided enhancement, we propose a novel generalization framework for few-shot medical image segmentation based on dual domain-invariant prompt optimization and Fourier consistency regularization.
- We propose a multi-band momentum enhancement strategy to dynamically enhance the source domain data in different frequency bands, enabling the source domain data to incorporate various variations of the target domain, thereby reducing the domain gap.
- To enable the model to learn domain-invariant features related to organ shapes and textures, we introduce dual domain-invariant prompts based on the multi-scale geometric structure representation ability of NSST and text prompts.

Related Works

Existing DG methods can be broadly categorized into non-SAM-based approaches and SAM-based approaches.

Non-SAM-based methods typically address DG through three primary strategies:(1) Data-level methods enhance robustness by increasing data diversity to simulate potential domain shifts. Techniques mainly include extensive randomization (Zhou et al. 2022), adversarial augmentation (Zhang et al. 2023) and normalization (Xu et al. 2022). Generative models are increasingly employed to synthesize new and domain-varying training samples (Li et al. 2020; Zhang et al. 2022). (2) Feature-level approaches attempt to learn representations insensitive to domain variations. Key strategies encompass feature normalization(Yu et al. 2023), explicit feature alignment (Li et al. 2020), dual consistency constraints (Niu et al. 2024b) and feature disentanglement frameworks designed to separate domain-invariant factors from domain-specific ones (Wang et al. 2020; Zhang et al. 2022). (3) Model-level methods enhance generalization through specialized learning paradigms (such as Meta-learning and ensemble learning) and optimization. Optimization focuses on robust regularization and loss functions (Robey, Pappas, and Hassani 2021; Stolte et al. 2023). Although these non-SAM approaches demonstrate significant progress, they often struggle to meet the stringent demands of highly heterogeneous real-world clinical deployment due to inherent limitations in generalization capability and adaptability.

SAM-based methods leverage the powerful pre-trained SAM for enhanced generalization. For example, COSAM (Fu et al. 2024) introduces a self-correcting SAM to update the coarse mask with prior prompts and a generalized error decoder to simulate manual correction, improving the adaptability to unseen scanner variations. DAPSAM (Wei et al. 2024) proposes a domain-adaptive prompt tuning strategy for SAM using a prototype-based memory bank learned from source domain images and designs a new adapter by

integrating cross-level features to improve robustness. MedSA (Wu et al. 2025) proposes a light yet effective adaptation technique by incorporating domain-specific medical knowledge and also proposes Hyper-Prompting Adapter (HyP-Adpt) to achieve prompt-conditioned adaptation. MoSE (Wei et al. 2025) employs a mixture of the shape experts framework with dictionary learning, efficiently capturing diverse and robust shape priors by a dynamic gating mechanism for improved generalization. However, current SAM-based DG methods assume fully annotated source domain data, which is impractical in medical settings due to high annotation costs. Furthermore, they lack explicit mechanisms for robust extraction of domain-invariant information, limiting their effectiveness in complex clinical scenarios.

Methods

In the few-shot DG setting, we are provided with a single source domain dataset $\mathcal{D}_l^s$ including labeled data $\mathcal{D}_l^s = \{(x_i^s, y_i^s)\}_{i=1}^{n_l}$ consisting of n_l source domain images $x_i^s \in \mathbb{R}^{H\times W}$ with their corresponding segmentation labels $y_i^s \in \mathbb{R}^{H\times W\times C}$ and unlabeled data $\mathcal{D}_u^s = \{(x_i^{su})\}_{i=1}^{n_u}$ of n_u unlabeled images. H, $W$and C represent the height, width and segmentation class, respectively. Then, the trained model is tested on unseen target domains $\mathcal{D}^t = \{\mathcal{D}_1^t, \mathcal{D}_2^t, \ldots, \mathcal{D}_n^t\}$. Following DAPSAM (Wei et al. 2024), we freeze the encoder. The proposed adapter and the decoder are set to fully trainable. The objective is to train a model on the limited annotated source domain $\mathcal{D}^s$ so that the trained model generalizes well on the unseen target domains $\mathcal{D}^t$.

Energy-guided Multi-band Momentum Enhancement

In practical clinical applications, we need to apply the trained model to several unseen target domains containing different imaging modalities, devices and parameters, which will lead to significant domain discrepancy between the source domain and the target domain. To effectively improve the model generalization performance in a few-shot source labeled application, it is necessary to make the source domain data contain as many different changes of the target domain as possible, that is, to be more in line with the real distribution of target domain data. According to the reference (Xie et al. 2022), there is a free energy difference between the distributions of the real source domain and the target domain which both basically follow the Gaussian distribution, as shown in Fig.1 (c). By Parseval's theorem (Kelkar, Grigsby, and Langsner 2007), we can infer that the energy distribution of an image is determined by the central spectrum of the Fourier amplitude. Therefore, we propose an Energy-guided Multi-band Momentum Enhancement (EMME) strategy to make the energy of the enhanced source domain data as consistent as possible with the energy distribution of the real target domain.

Specifically, we first perform Fourier transform on all source domain images x_i to obtain their amplitude spectrum $A_{x_i} = \mathcal{F}(x_i)$. Assuming the center coordinates of the image size are $(c_h, c_w) = (H/2, W/2)$, we define square masks for different frequency bands (low, medium, and high frequencies):

$$
\begin{aligned}
M_{\text{low}}(a,b) &= \begin{cases} 1 & \text{if } L_a \le r_1 \text{ and } L_b \le r_1 \\ 0 & \text{otherwise} \end{cases}, \\
M_{\text{mid}}(a,b) &= \begin{cases} 1 & \text{if } r_1 < \max(L_a, L_b) \le r_2 \\ 0 & \text{otherwise} \end{cases}, \\
M_{\text{high}} &= \mathbf{1} - M_{\text{low}} - M_{\text{mid}}.
\end{aligned} \tag{1}
$$

where $L_a = |a - c_h|$, $L_b = |b - c_w|$, $r_1 = min(H, W)/R_1$ and $r_2 = min(H, W)/R_2$ are the division radii, and then the global mean $\mu_{g,k}$ and variance $\sigma_{g,k}^2 (k \in \{low, mid, high\})$ of the amplitude spectrum of each frequency band are calculated as

$$
\mu_{\text{g},k} = \mathbb{E}_{x_i \in \mathcal{D}_\text{s}} \left[\frac{\sum_{x,y} A_k(x_i)}{N_k} \right], \tag{2}
$$

$$
\sigma_{\text{g},k}^2 = \mathbb{E}_{x_i \in \mathcal{D}_\text{s}} \left[\frac{\sum_{x,y} \left(A_k(a,b) - \mu_{x_i,k}\right)^2}{N_k} \right]. \tag{3}
$$

Among them, $A_k = A_{x_i} \odot M_k$ represents the amplitude spectrum of different frequency bands, $N_k = \sum_{a,b} M_k(a,b)$ is the number of pixels in the current frequency band, and $\mu_{x_i,k} = \frac{1}{N_k} \sum_{a,b} A_k(a,b)$ is the average value of the current frequency band k.

Then, for the images in the current batch $\mathcal{B}_m$, we calculate the mean $\mu_{\mathcal{B}_m,k}$ and variance $\sigma_{\mathcal{B}_m,k}^2$ of each frequency band, and perform a momentum update on the global statistics.

$$
\begin{aligned}
\mu_{\mathcal{B}_m,k} &= \frac{1}{|\mathcal{B}_m|} \sum \mu_{x_{\mathcal{B}_{mi}},k}, \\
\sigma_{\mathcal{B}_m,k}^2 &= \frac{1}{|\mathcal{B}_m|} \sum \sigma_{x_{\mathcal{B}_{mi}},k}^2.
\end{aligned} \tag{4}
$$

$$
\begin{aligned}
\mu_{\text{g},k} &\leftarrow \beta \cdot \mu_{\text{g},k} + (1-\beta) \cdot \mu_{\mathcal{B}_m,k}, \\
\sigma_{\text{g},k}^2 &\leftarrow \beta \cdot \sigma_{\text{g},k}^2 + (1-\beta) \cdot \sigma_{\mathcal{B}_m,k}^2,
\end{aligned} \tag{5}
$$

where $\beta = 0.9$ is the momentum coefficient. Then, we generate the perturbed statistics for the target domain:

$$
\begin{aligned}
\mu_{\text{trgt},k} &= \mu_{\text{g},k} + \lambda_\mu \cdot \delta_\mu, \\
\sigma_{\text{trgt},k}^2 &= \sigma_{\text{g},k}^2 + \lambda_\sigma \cdot \delta_\sigma,
\end{aligned} \tag{6}
$$

where λ_μ, $\lambda_\sigma \sim \mathcal{U}(0,1)$, δ_μ, δ_σ are generated by the Box-Muller transformation.

$$
\begin{aligned}
\delta_\mu &= \sqrt{-2 \ln u_1} \cos(2\pi u_2), \\
\delta_\sigma &= \sqrt{-2 \ln u_1} \sin(2\pi u_2), \\
u_1, u_2 &\sim \mathcal{U}(0,1).
\end{aligned} \tag{7}
$$

Next, the amplitude spectrum $A_k(x_{\mathcal{B}_{mi}})$ of the current source domain image frequency band is standardized and projected to the statistical space of the target domain.

$$
A_{\text{norm},k} = \frac{A_k(x_{\mathcal{B}_{mi}}) - \mu_{x_{\mathcal{B}_{mi}},k}}{\sqrt{\sigma_{x_{\mathcal{B}_{mi}},k}^2 + \epsilon}}, \tag{8}
$$

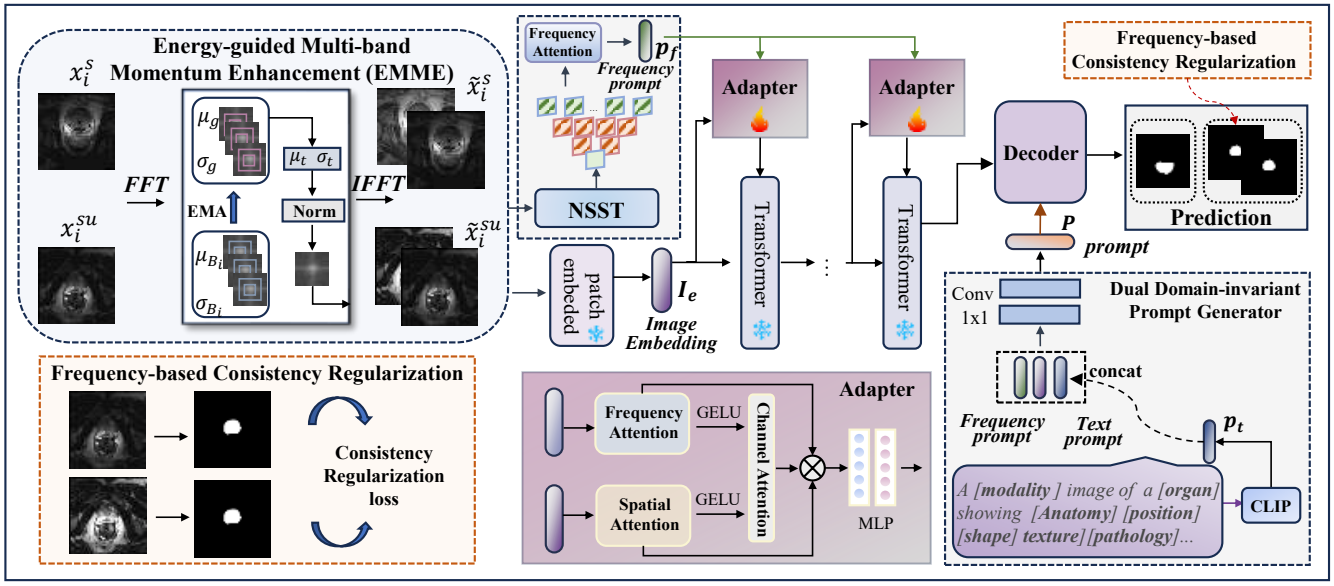

Figure 2: The overall framework of our proposed method. We first dynamically augment source data by leveraging diverse frequency bands of the Fourier amplitude spectrum. Then, we propose a novel Dual domain-Invariant Prompt Optimization strategy by integrating NSST and text prompts to strengthen the extraction of domain-invariant features. Finally, frequency consistency regularization is employed to refine model robustness using predictions from unlabeled data.

$$A_{\text{trgt},k} = A_{\text{norm},k} \cdot \sqrt{\sigma^2_{\text{trgt},k}} + \mu_{\text{trgt},k}. \tag{9}$$

Finally, the target amplitude spectrum of all frequency bands are merged while keeping the source domain phase spectrum unchanged, and the inverse Fourier transform is performed to obtain the enhanced image $\tilde{x}_i$.

$$A_{\text{target}} = \sum_{k \in \{\text{low, mid, high}\}} A_{\text{target},k}, \tag{10}$$

$$F_{\text{target}} = A_{\text{target}} \odot e^{j \cdot P_{\text{src}}}, \tag{11}$$

$$\tilde{x}_i = \mathcal{F}^{-1}(F_{\text{target}}). \tag{12}$$

Dual Domain-invariant Prompt Optimization

To achieve robust and stable single-source medical segmentation generalization, the ability to learn effective domain-invariant information is a critical prerequisite. It is difficult to achieve stable generalization ability in complex clinical scenarios, especially when there are few annotations in the source domain. For this reason, we propose a Dual domain-Invariant Prompt Optimization (DIPO) strategy, which combines the Non-subsampled Shearlet Transform (NSST) and text prompt for the first time.

For medical images, the contours and structural shapes of organs tend to remain relatively stable across different domains (Wei et al. 2025). Therefore, the design of prompts is required to retain more information related to contours and semantic structures. Regarding the frequency-domain invariance prompt, we introduce an effective multi-scale geometric analysis tool, NSST. It has developed on the basis of synthetic wavelets and ridge waves, and achieves directional sensitivity by introducing a shear matrix, thereby effectively capturing geometric features such as edges and textures in images. Firstly, the Non-subsampled Pyramid Filter (NSPF) is utilized to perform multi-scale decomposition on the source domain image $\mathcal{D}^s$.Then, the high-frequency sub-bands obtained through decomposition are further decomposed in multiple directions by using the Shearlet Filter (SF). Finally, after completing the processing of the corresponding sub-bands, the image is reconstructed through the inverse NSST operation.

After L-level NSST decomposition, we can obtain a total of $\sum_{l=1}^{L} 2^l + 1$ Frequency Components (FCs), including $\zeta(0)$ corresponding to the low-frequency sub-band LS_L, and band-pass FCs $\zeta(1)$-$\zeta(\sum_{l=1}^{L} 2^l)$ corresponding to L high frequency sub-bands $BS_l(l = 1, 2, \ldots, L)$. Figure 3 shows the three-level NSST decomposition. The support base of NSST is a pair of trapezoidal intervals, and each trapezoidal interval has a width of 2^{2j} and a height of 2^j. The low frequency sub-band FC preserves most of the semantic content of the image, while the high frequency sub-band FCs capture the structure and texture in different directions. Taking a prostate MRI slice x_i^s as an example, we obtain its FCs by three-layer NSCT decomposition:

$$\zeta_i^s = \text{NSST}(x_i^s) = \{\zeta_i^s(0), \zeta_i^s(1), \ldots, \zeta_i^s(14)\} \tag{13}$$

Specifically, the source domain image and the enhanced image are decomposed into low-frequency (global structure) $\zeta_i^s(0)$ and high-frequency (organ contour) sub-bands of different scales $\zeta_i^s(1), \ldots, \zeta_i^s(14)$, using the frequency self-attention mechanism (Deng et al. 2025) to extract domain-invariant features as the frequency structure prompt p_f;

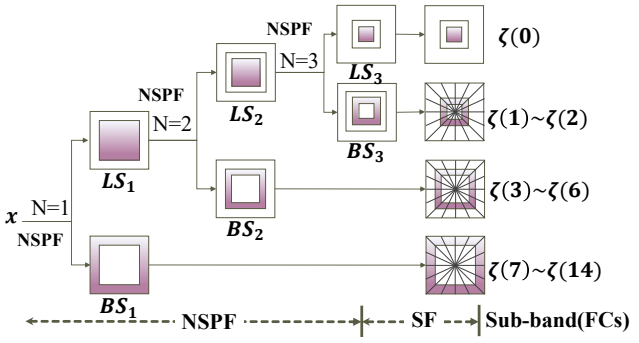

Figure 3: The details of the three-level NSST.

For text domain-invariant prompt, we combine the domain-invariant descriptions automatically generated by ChatGPT (such as organ location, anatomical structure, and pathological features, etc.) with domain-specific descriptions (such as device-related gray range), and extract the domain-invariant text feature p_t that fuses semantic *a priori* through the CLIP encoder. Then, we concatenate p_f, p_t and image embedding I_e. The final frequency-text domain-invariant prompt P is generated through two layers of convolution 1×1.

Frequency Consistency Regularization

To make full use of the unlabeled data x_i^{su} in the source domain, we propose a Frequency Consistency Regularization (FCR) strategy based on the frequency domain. In the EMME module, we obtained source domain images with different degrees of amplitude spectrum dynamic enhancement. Therefore, we require the model to make consistent predictions for the original images and the enhanced images, thereby improving the model's prediction robustness against different input variations and further enhancing its generalization ability.

$$\mathcal{L}_u = \frac{1}{n_u} \sum_{i=1}^{n_u} KL\left(f_\theta\left(x_i^{su}\right), f_\theta\left(\tilde{x}_i^{su}\right)\right) \tag{14}$$

We give the overall loss function $\mathcal{L}$, where λ is the weight of the consistency loss. We adopt the ramp-up strategy and λ gradually increases with training.

$$\mathcal{L}_s = \frac{1}{n_l} \sum_{i=1}^{n_l} \Big\{ \mathrm{CE}\big(f_\theta(x_i^s), y_i^s\big) + \mathrm{Dice}\big(f_\theta(x_i^s), y_i^s\big) \Big\} \tag{15}$$

$$\mathcal{L} = \mathcal{L}_s + \lambda \mathcal{L}_u \tag{16}$$

Experiments and Results

Experimental Settings

Dataset: We evaluated our proposed method on two commonly used medical image segmentation tasks: prostate segmentation and joint segmentation of the optic disc (OD) and optic cup (OC). The prostate dataset (Xu et al. 2022) comprises 116 T2-weighted MRI from 6 different domains with distribution shift. Following DAPSAM, we resized each sample to the resolution of 384×384. The joint OC/OD segmentation dataset RIGA + (Niu et al. 2024a; Liu et al. 2021) consists of multi-domain annotated fundus images from five distinct domains: BinRushed, Magrabia, Messidor BASE1, BASE2 and BASE3. We select BinRushed as the source domains for training and then evaluate the model's performance on three target domains, Messidor BASE1-3. For these two datasets, we employ the Dice Similarity Coefficient (DSC) to evaluate the segmentation performance.

Implementation Details: We employed SAM of the 'ViT-B' version as the segmentation backbone. Following Adaptformer (Chen et al. 2022b), we introduce two trainable MLP-structured adapters into each encoder layer to train our baseline model. Our network was trained on an NVIDIA RTX 3090 GPU with a batch size of 8 and 150 epochs, using an ADAM optimizer and a cosine annealing learning rate adjustment strategy with a learning rate of 1e-4 and a weight decay of 5e-4. We also adopted the warm-up strategy following SAMed (Zhang and Liu 2023), with warm-up periods set to 250 and 25 for the prostate and RIGA+ datasets, respectively, due to different data training settings. We apply early stop at 160 epochs, with a maximum of 200 epochs. The hyperparameter λ in Equation (16) is set to 0.8.

Comparison Results

Prostate Segmentation The quantitative results of prostate segmentation are shown in Table 1. Our proposed method achieves state-of-the-art performance under both 100% and 25% source domain annotation settings. In the 100% annotation scenario, it attains an average score of 83.56%, surpassing all other competitive methods. Additionally, it ranks first in 5/6 domains, with particularly significant advantages in Domain A (87.43%, +1.09% over Mamba-Sea) and Domain D (86.22%, +1.86%). More critically, under the challenging 25% annotation setting, our method maintains robust performance, outperforming the strongest competitor DAPSAM by 1.70% and exhibiting leading results across all six domains. Notably, it achieves a remarkable 70.35% in Domain C and outperforms Mamba-Sea by 7.0% and DAPSAM by 5.18%, demonstrating exceptional cross-domain generalization with limited supervision. These results validate that our proposed framework reduces annotation dependency while maintaining generalization. The visual results are shown in Figure 4, our method maintains more organ texture and edges details compared to other methods.

Fundus OD/OC Segmentation The quantitative results of joint OD/OC segmentation in Table 2 show that our proposed method achieves satisfactory performance under both 100% and 25% source-domain annotation ratios. In the full annotation setting, our method achieves average OD/OC 92.95%, surpassing all competitive methods, which exceeds the DAPSAM (92.02%) by 0.93%. Under limited annotation (25%), our method achieves average OD and OC (82.48%), outperforming DAPSAM by 1.83%. This shows that our proposed method obtains both high accuracy and annotation efficiency in few-shot medical image segmentation.

Method	Model	Ratio	A	B	C	D	E	F	Average
Supervised (Zhou et al. 2022)	UNet	100%	85.38	83.68	82.15	85.21	87.04	84.29	84.63
CSDG (Ouyang et al. 2022) (IEEE TMI 2022)	UNet	100%	80.72	68.00	59.78	72.40	68.67	70.78	70.06
MaxStyle (Chen et al. 2022a) (MICCAI 2022)	UNet	100%	81.25	70.27	62.09	58.18	70.04	67.77	68.27
D-Norm (Zhou et al. 2022) (CVPR 2022)	UNet	100%	81.69	71.36	67.22	63.21	72.32	71.36	71.19
SLAug (Su et al. 2023) (AAAI 2023)	UNet	100%	80.95	69.38	68.51	67.66	73.63	69.47	71.60
CCSDG (Liu et al. 2020) (MICCAI 2023)	UNet	100%	80.72	68.00	59.78	72.40	68.67	70.78	70.06
Mamba-Sea (Cheng et al. 2025) (TMI 2025)	Mamba	100%	86.34	82.47	**78.69**	84.36	80.64	84.31	82.80
SAMed (Zhang and Liu 2023) (Arxiv 2023)	ViT	100%	80.42	81.44	66.75	82.09	80.19	80.17	78.51
SAMMed (Wang et al. 2024) (Arxiv 2024)	ViT	100%	84.25	80.63	68.25	84.69	80.23	81.78	79.97
DeSAM (Gao et al. 2024) (MICCAI 2024)	ViT	100%	82.80	80.61	64.77	83.41	80.36	82.17	79.02
DAPSAM (Wei et al. 2024) (MICCAI 2024)	ViT	100%	86.21	81.05	70.81	85.28	82.91	81.48	81.31
Med-SA (Wu et al. 2025) (MIA 2025)	ViT	100%	85.36	80.64	69.82	84.36	83.65	82.47	81.05
DIPO (Ours)	ViT	100%	**87.43**	**82.67**	75.35	**86.22**	**84.56**	**85.12**	**83.56**
CSDG (Ouyang et al. 2022) (IEEE TMI 2022)	UNet	25%	65.35	59.76	45.81	60.23	43.65	44.37	53.20
MaxStyle (Chen et al. 2022a) (MICCAI 2022)	UNet	25%	64.25	58.47	46.98	59.14	45.21	49.65	53.95
D-Norm (Zhou et al. 2022) (CVPR 2022)	UNet	25%	65.52	57.23	44.21	61.27	49.98	54.21	57.07
SLAug (Su et al. 2023) (AAAI 2023)	UNet	25%	66.93	58.64	43.85	64.58	52.68	55.14	56.97
CCSDG (Liu et al. 2020) (MICCAI 2023)	UNet	25%	67.25	57.23	45.81	63.28	51.54	53.95	56.51
Mamba-Sea (Cheng et al. 2025) (MIA 2025)	Mamba	25%	73.24	70.62	63.35	71.84	70.58	70.25	69.98
SAMed (Zhang and Liu 2023) (Arxiv 2023)	ViT	25%	71.03	72.15	59.47	71.31	72.87	72.76	69.93
SAMMed (Wang et al. 2024) (Arxiv 2024)	ViT	25%	72.36	70.94	60.24	72.68	73.68	72.44	70.39
DeSAM (Gao et al. 2024) (Arxiv 2023)	ViT	25%	72.87	71.54	57.65	72.43	72.56	74.43	70.25
DAPSAM (Wei et al. 2024) (MICCAI 2024)	ViT	25%	77.24	71.76	65.17	74.58	75.63	72.61	72.83
Med-SA (Wu et al. 2025) (MIA 2025)	ViT	25%	76.85	70.65	63.56	73.68	74.66	71.25	71.78
DIPO (Ours)	ViT	25%	**78.81**	**73.86**	**70.35**	**77.43**	**78.21**	**76.39**	**75.84**

Table 1: Quantitative comparison results on prostate segmentation with other state-of-the-art methods. The best performance results are shown in bold and the second ones are underlined.

Method	Ratio	A	B	C	Average
CSDG	100%	87.28	89.09	88.81	88.39
MaxStyle	100%	88.45	80.68	87.35	85.49
SLAug	100%	89.30	88.43	89.98	89.23
D-Norm	100%	88.19	86.42	89.25	87.95
CCSDG	100%	90.93	91.01	91.11	91.02
Mamba-Sea	100%	90.06	90.24	89.45	89.92
DeSAM	100%	86.33	87.76	88.93	87.67
SAMed	100%	89.76	87.16	88.72	88.55
SAMMed	100%	90.03	88.59	90.29	89.63
DAPSAM	100%	92.29	91.21	92.56	92.02
Med-SA	100%	90.76	90.08	91.20	90.68
DIPO(Ours)	100%	**92.77**	**91.86**	**94.24**	**92.95**
CSDG	25%	77.90	79.79	79.13	78.94
MaxStyle	25%	77.97	72.65	79.38	76.66
SLAug	25%	79.08	79.62	81.10	79.93
D-Norm	25%	78.71	75.95	80.08	78.24
CCSDG	25%	79.51	77.63	80.31	79.15
Mamba-Sea	25%	80.23	79.02	80.03	79.76
DeSAM	25%	78.34	77.25	79.78	78.46
SAMed	25%	80.46	79.46	80.22	80.04
SAMMed	25%	79.31	78.55	79.41	79.09
DAPSAM	25%	80.32	79.74	81.88	80.65
Med-SA	25%	79.07	78.55	80.21	79.28
DIPO(Ours)	25%	**82.10**	**81.41**	**83.95**	**82.48**

Table 2: Quantitative comparison results on OD/OC segmentation with other state-of-the-art methods. The best performance results are shown in bold and the second ones are underlined.

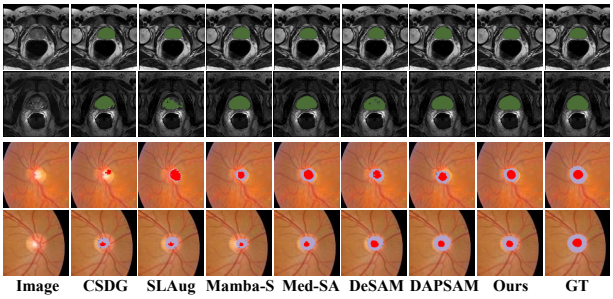

Figure 4: The visualization of the segmentation results.

Ablation Studies

Effect of Different Key Components In this ablation study, we conducted exhaustive experiments to assess the contribution of different key components to enhancing model performance. From Table 3, we can see that by applying EMME, the performance rises to 71.75%, indicating a noticeable improvement over the baseline model. Further incorporation of Frequency prompt and Text prompt increases the performance to 73.12% and 72.57%, demonstrating the effectiveness of combining domain-invariant prompt with momentum augmentation. Subsequently, the performance improves to 73.76% by applying FCR, confirming the con-

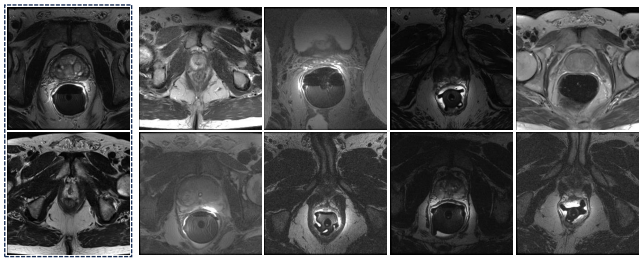

Figure 5: Visualization of enhanced source data.

tribution of frequency consistency to segmentation performance. Finally, the combination of all components achieves an outstanding performance of 75.84%. In addition, we also added our EMME and FCR strategies for the dapsam as well as the Mamba-sea methods, and as can be seen in Table 4, the segmentation performance is also improved for all of them, which further illustrates the effectiveness and generalizability of our proposed method.

EMME	Frequency prompt	Text prompt	FCR	Average
				70.65
✓				71.75
✓	✓			73.12
✓		✓		72.57
	✓	✓		73.24
✓	✓	✓		73.89
	✓	✓	✓	73.76
✓	✓		✓	74.67
✓		✓	✓	74.43
✓	✓	✓	✓	**75.84**

Table 3: Quantitative comparison results on the effect of different key components.

Method	EMME	FCR	prostate
DAPSAM			72.83
	✓		73.17
		✓	73.25
	✓	✓	**73.68**
Med-SA			71.78
	✓		72.36
		✓	72.23
	✓	✓	**72.94**

Table 4: Quantitative comparison results on the effect of EMME and FCR in other competitive methods.

Effect of the NSST Decomposition Level The number of decomposition layers is a key parameter, which directly affects the transformation's ability to represent the image contour and details. As can be seen in Figure 6, when the number of decomposition layers is 3, the shape and contour information of the image can be well represented, and the best segmentation performance is achieved in both datasets.

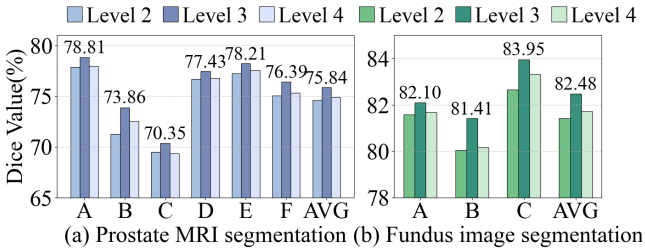

Figure 6: Ablation analysis of different NSST decomposition level.

Effect of Different Band Division Radius Figure 5 visualizes the enhanced source domain image (right) that contains more diverse potential target domain variations compared with the original image (left). We conducted extensive ablation experiments on different frequency band division radii (R_1,R_2). It can be seen that when R_1=4 and R_2=2, the best consistent segmentation results are obtained in the two datasets.

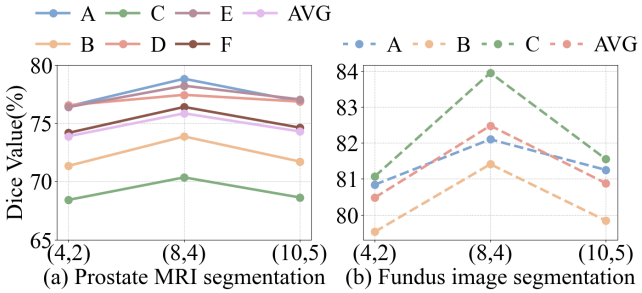

Figure 7: Effect of different frequency band division (R_1,R_2) on the generalization performance.

Conclusion

In this paper, we propose a novel energy-guided framework for few-shot medical image segmentation generalization. In order to make the source domain adaptable to more complex changes in the unseen target domains, we propose a multi-band momentum enhancement strategy by dynamically enhancing the source domain data in different bands of the Fourier amplitude spectrum. Meanwhile, a dual domain-invariant prompt optimization is proposed by combining NSST, frequency attention and text prompt, so that the model learns domain-invariant features related to organ shapes and textures. Finally, the robustness of the model on the augmented unlabeled data is further enhanced by the frequency-domain consistency regularization. Quantitative and qualitative results in prostate and fundus datasets consistently show that we substantially outperform currently competitive methods.

Acknowledgments

This work was supported by the National Natural Science Foundation of China under Grant 62303441, Shanghai National Natural Science Foundation (25ZR1402548), Youth Innovation Promotion Association, Chinese Academy of Sciences (2021233) and Shanghai Academic Research Leader (22XD1424500).

References

Ali, M.; Wu, T.; Hu, H.; Luo, Q.; Xu, D.; Zheng, W.; Jin, N.; Yang, C.; and Yao, J. 2025. A review of the segment anything model (sam) for medical image analysis: Accomplishments and perspectives. *Computerized Medical Imaging and Graphics*, 119: 102473.

Chen, C.; Li, Z.; Ouyang, C.; Sinclair, M.; Bai, W.; and Rueckert, D. 2022a. Maxstyle: Adversarial style composition for robust medical image segmentation. In *International Conference on Medical Image Computing and Computer-Assisted Intervention*, 151–161. Springer.

Chen, S.; Ge, C.; Tong, Z.; Wang, J.; Song, Y.; Wang, J.; and Luo, P. 2022b. Adaptformer: Adapting vision transformers for scalable visual recognition. *Advances in Neural Information Processing Systems*, 35: 16664–16678.

Cheng, Z.; Guo, J.; Zhang, J.; Qi, L.; Zhou, L.; Shi, Y.; and Gao, Y. 2025. Mamba-Sea: A Mamba-based Framework with Global-to-Local Sequence Augmentation for Generalizable Medical Image Segmentation. *IEEE Transactions on Medical Imaging*.

Deng, M.; Sun, S.; Li, Z.; Hu, X.; and Wu, X. 2025. FMNet: Frequency-Assisted Mamba-Like Linear Attention Network for Camouflaged Object Detection. *arXiv preprint arXiv:2503.11030*.

Easley, G.; Labate, D.; and Lim, W.-Q. 2008. Sparse directional image representations using the discrete shearlet transform. *Applied and Computational Harmonic Analysis*, 25(1): 25–46.

Fu, Y.; Chen, Z.; Ye, Y.; Lei, X.; Wang, Z.; and Xia, Y. 2024. CoSAM: Self-correcting SAM for domain generalization in 2D medical image segmentation. *arXiv preprint arXiv:2411.10136*.

Gao, Y.; Jiang, Y.; Peng, Y.; Yuan, F.; Zhang, X.; and Wang, J. 2025. Medical Image Segmentation: A Comprehensive Review of Deep Learning-Based Methods. *Tomography*, 11(5): 52.

Gao, Y.; Xia, W.; Hu, D.; Wang, W.; and Gao, X. 2024. Desam: Decoupled segment anything model for generalizable medical image segmentation. In *International Conference on Medical Image Computing and Computer-Assisted Intervention*, 509–519. Springer.

Kelkar, S.; Grigsby, L.; and Langsner, J. 2007. An extension of Parseval's theorem and its use in calculating transient energy in the frequency domain. *IEEE Transactions on Industrial Electronics*, (1): 42–45.

Kirillov, A.; Mintun, E.; Ravi, N.; Mao, H.; Rolland, C.; Gustafson, L.; Xiao, T.; Whitehead, S.; Berg, A. C.; Lo, W.-Y.; et al. 2023. Segment anything. In *Proceedings of the IEEE/CVF international conference on computer vision*, 4015–4026.

Li, H.; Wang, Y.; Wan, R.; Wang, S.; Li, T.-Q.; and Kot, A. 2020. Domain generalization for medical imaging classification with linear-dependency regularization. *Advances in neural information processing systems*, 33: 3118–3129.

Liu, Q.; Chen, C.; Qin, J.; Dou, Q.; and Heng, P.-A. 2021. Feddg: Federated domain generalization on medical image segmentation via episodic learning in continuous frequency space. In *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*, 1013–1023.

Liu, Q.; Dou, Q.; and Heng, P.-A. 2020. Shape-aware meta-learning for generalizing prostate MRI segmentation to unseen domains. In *International conference on medical image computing and computer-assisted intervention*, 475–485. Springer.

Liu, Q.; Dou, Q.; Yu, L.; and Heng, P. A. 2020. MS-Net: multi-site network for improving prostate segmentation with heterogeneous MRI data. *IEEE transactions on medical imaging*, 39(9): 2713–2724.

Liu, S.; Yin, S.; Qu, L.; and Wang, M. 2023a. Reducing domain gap in frequency and spatial domain for cross-modality domain adaptation on medical image segmentation. In *Proceedings of the AAAI conference on artificial intelligence*, volume 37, 1719–1727.

Liu, S.; Yin, S.; Qu, L.; Wang, M.; and Song, Z. 2023b. A structure-aware framework of unsupervised cross-modality domain adaptation via frequency and spatial knowledge distillation. *IEEE Transactions on Medical Imaging*, 42(12): 3919–3931.

Niu, Z.; Ouyang, S.; Xie, S.; Chen, Y.-w.; and Lin, L. 2024a. A survey on domain generalization for medical image analysis. *arXiv preprint arXiv:2402.05035*.

Niu, Z.; Sun, H.; Ouyang, S.; Xie, S.; Chen, Y.-w.; Tong, R.; and Lin, L. 2024b. IRLSG: Invariant Representation Learning for Single-Domain Generalization in Medical Image Segmentation. In *ICASSP 2024-2024 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, 5585–5589. IEEE.

Ouyang, C.; Chen, C.; Li, S.; Li, Z.; Qin, C.; Bai, W.; and Rueckert, D. 2022. Causality-inspired single-source domain generalization for medical image segmentation. *IEEE Transactions on Medical Imaging*, 42(4): 1095–1106.

Rayed, M. E.; Islam, S. S.; Niha, S. I.; Jim, J. R.; Kabir, M. M.; and Mridha, M. 2024. Deep learning for medical image segmentation: State-of-the-art advancements and challenges. *Informatics in medicine unlocked*, 47: 101504.

Robey, A.; Pappas, G. J.; and Hassani, H. 2021. Model-based domain generalization. *Advances in Neural Information Processing Systems*, 34: 20210–20229.

Stolte, S. E.; Volle, K.; Indahlastari, A.; Albizu, A.; Woods, A. J.; Brink, K.; Hale, M.; and Fang, R. 2023. DOMINO++: domain-aware loss regularization for deep learning generalizability. In *International Conference on Medical Image Computing and Computer-Assisted Intervention*, 713–723. Springer.

Su, Z.; Yao, K.; Yang, X.; Huang, K.; Wang, Q.; and Sun, J. 2023. Rethinking data augmentation for single-source domain generalization in medical image segmentation. In *Proceedings of the AAAI conference on artificial intelligence*, volume 37, 2366–2374.

Wang, H.; Ye, H.; Xia, Y.; and Zhang, X. 2024. Leveraging sam for single-source domain generalization in medical image segmentation. *arXiv preprint arXiv:2401.02076*.

Wang, S.; Yu, L.; Li, K.; Yang, X.; Fu, C.-W.; and Heng, P.-A. 2020. Dofe: Domain-oriented feature embedding for generalizable fundus image segmentation on unseen datasets. *IEEE Transactions on Medical Imaging*, 39(12): 4237–4248.

Wei, J.; Zhao, X.; Woo, J.; Ouyang, J.; El Fakhri, G.; Chen, Q.; and Liu, X. 2025. Mixture-of-Shape-Experts (MoSE): End-to-End Shape Dictionary Framework to Prompt SAM for Generalizable Medical Segmentation. In *Proceedings of the Computer Vision and Pattern Recognition Conference*, 6448–6458.

Wei, Z.; Dong, W.; Zhou, P.; Gu, Y.; Zhao, Z.; and Xu, Y. 2024. Prompting segment anything model with domain-adaptive prototype for generalizable medical image segmentation. In *International Conference on Medical Image Computing and Computer-Assisted Intervention*, 533–543. Springer.

Wu, J.; Wang, Z.; Hong, M.; Ji, W.; Fu, H.; Xu, Y.; Xu, M.; and Jin, Y. 2025. Medical sam adapter: Adapting segment anything model for medical image segmentation. *Medical image analysis*, 102: 103547.

Xie, B.; Yuan, L.; Li, S.; Liu, C. H.; Cheng, X.; and Wang, G. 2022. Active learning for domain adaptation: An energy-based approach. In *Proceedings of the AAAI conference on artificial intelligence*, volume 36, 8708–8716.

Xu, C.; Wen, Z.; Liu, Z.; and Ye, C. 2022. Improved domain generalization for cell detection in histopathology images via test-time stain augmentation. In *International Conference on Medical Image Computing and Computer-Assisted Intervention*, 150–159. Springer.

Yu, W.; Huang, Z.; Zhang, J.; and Shan, H. 2023. SAN-Net: Learning generalization to unseen sites for stroke lesion segmentation with self-adaptive normalization. *Computers in Biology and Medicine*, 156: 106717.

Zhang, H.; Zhang, Y.-F.; Liu, W.; Weller, A.; Schölkopf, B.; and Xing, E. P. 2022. Towards principled disentanglement for domain generalization. In *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*, 8024–8034.

Zhang, K.; and Liu, D. 2023. Customized segment anything model for medical image segmentation. *arXiv preprint arXiv:2304.13785*.

Zhang, Z.; Wang, B.; Yao, L.; Demir, U.; Jha, D.; Turkbey, I. B.; Gong, B.; and Bagci, U. 2023. Domain generalization with adversarial intensity attack for medical image segmentation. *arXiv preprint arXiv:2304.02720*.

Zhou, Z.; Qi, L.; Yang, X.; Ni, D.; and Shi, Y. 2022. Generalizable cross-modality medical image segmentation via style augmentation and dual normalization. In *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*, 20856–20865.