

Segment Anything Model with Symmetric Spectral Uncertainty Gating for Generalizable Medical Image Segmentation

He Li¹, Ran Gu¹, Jia Fu¹, Ping Ye¹, Litingyu Wang¹, Kang Li¹,
Guotai Wang^{1,2}(✉), and Shaoting Zhang^{1,2}

¹ School of Mechanical and Electrical Engineering, University of Electronic Science and Technology of China, Chengdu, China

² Shanghai Artificial Intelligence Laboratory, Shanghai, China
guotai.wang@uestc.edu.cn

Abstract. Parameter-Efficient Fine-Tuning (PEFT) enables foundation models like the Segment Anything Model (SAM) to be adapted for medical image segmentation with limited supervision. However, limited supervision from a single domain tends to collapse the inherent generalization capacity of SAM toward source domain-specific styles, leading to performance drops on unseen clinical centers. To address this, we propose **Symmetric Spectral Uncertainty Gating (SSUG)** for SAM-based Single-Domain Generalization (SDG), which explicitly disentangles anatomical representation from styles through a synergistic closed-loop design. First, a Spectral Manifold Extrapolation (SME) module proactively probes the statistical boundaries of the source spectrum to simulate various possible stylistic shifts. Second, an Uncertainty-aware Symmetric Channel Gating (USCG) module measures response divergence to softly calibrate style-sensitive channels. Finally, we introduce Hierarchical Decoupled Consistency (HDC) constraints to reinforce domain-invariant anatomical representations across multiple feature granularities. Experimental results on two challenging domain generalization benchmarks for Fundus and Prostate segmentation proved the significant superiority of SSUG compared to state-of-the-art SDG methods. Code is available at <https://github.com/HiLab-git/SSUG>.

Keywords: Single-Domain Generalization · Segment Anything Model · Uncertainty Gating · Medical Image Segmentation.

1 Introduction

Deep learning (DL) models have made great achievements in medical image segmentation [16]. However, the well-trained deep networks on the source domain struggle to generalize to unseen target domains due to the domain shifts arising from differences in imaging modalities, equipment manufacturers, and scanning protocols [25, 6, 12]. Although Multi-Domain Generalization (MDG) mitigates

this issue by leveraging data from multiple institutions [19,27], its practical adoption is hindered by data-privacy protocols and the high cost of expert annotations. Consequently, Single-Domain Generalization (SDG), which aims to learn domain-robust representations from a single labeled source domain [11,9], has emerged as a more practical yet challenging setting.

Previous SDG efforts primarily relied on data augmentation or style randomization. Classic methods perturb images via low-level intensity transforms and Bézier Curves augmentation [25,28] or manipulate feature statistics (e.g., MixStyle [26] and MaxStyle [1]) to simulate stylistic diversity. Despite their effectiveness, conventional architectures are constrained by limited representational capacity and single-domain supervision, hindering robust anatomical prior learning and causing generalization bottlenecks under severe domain shifts. This limitation motivates the exploration of large-scale foundation models.

The emergence of foundation models, particularly the Segment Anything Model (SAM) [10], has introduced new opportunities for SDG. However, directly applying SAM to specialized medical tasks remains suboptimal due to the significant domain gap between natural and medical images [14,8]. To this end, Parameter-Efficient Fine-Tuning (PEFT) strategies, such as Adapters and LoRA, have been proposed to adapt SAM to medical tasks [24,22]. Despite their promise, PEFT-adapted SAM models suffer from a critical limitation in the SDG setting. The high-capacity architecture of SAM makes it inherently sensitive to appearance statistics, and fine-tuning further exacerbates overfitting to source domain-specific imaging styles, leading to poor generalization on unseen domains. Recent SAM-based methods, DAPSAM [20] and Nora [21], attempt to mitigate this issue through prototypes or stochastic noise injection, but they largely rely on passive perturbations and ignore the heterogeneous channel-wise sensitivity to style variations, limiting their robustness under domain shifts.

To counteract the unpredictable style variations in unseen target domains, we propose the Symmetric Spectral Uncertainty Gating (SSUG) for SAM-based SDG medical image segmentation. To the best of our knowledge, this is the first synergistic closed-loop framework designed to alleviate style sensitivity and source-domain overfitting within the foundation model architecture for SDG. Specifically, SSUG comprises three synergistic components: 1) Spectral Manifold Extrapolation (SME), which acts as an active probe to proactively induce distribution bias along the statistical boundaries of the source spectrum; 2) Uncertainty-aware Symmetric Channel Gating (USCG), a dynamic filter that identifies and suppresses style-corrupted channels based on their response divergence; and 3) Hierarchical Decoupled Consistency (HDC), which consolidates domain-invariant anatomical representations across multiple feature granularities. Extensive experiments on two widely used generalizable medical image segmentation benchmarks demonstrate that SSUG effectively improves model generalizability on unseen testing domains.

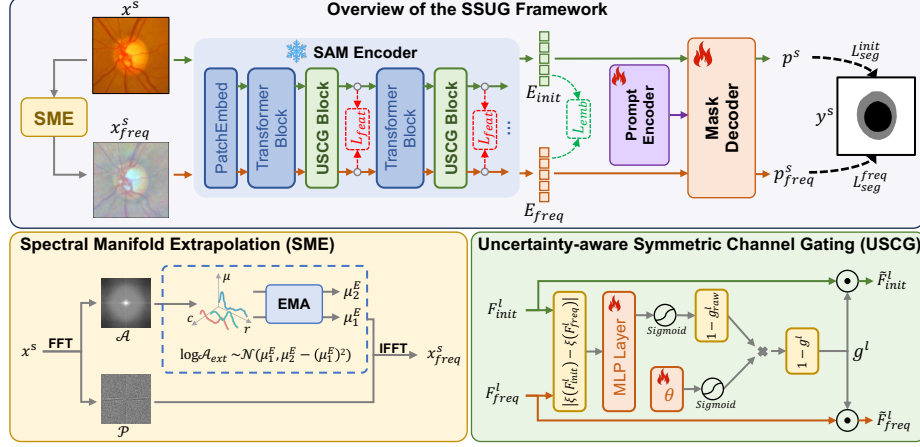


Fig. 1. Overview of the proposed SSUG framework: (Bottom Left) SME probes the statistical boundaries of the source spectrum to simulate various possible styles, (Bottom Right) USCG measures response divergence between branches to softly calibrate style-sensitive channels, and HDC constraints (L_{feat} and L_{emb}) enforce consistency across multiple feature granularities for domain-invariant anatomical representations.

2 Method

Assuming that a source domain dataset $\mathcal{S} = \{x_i^s, y_i^s\}_{i=1}^{N_s}$ is collected from a single domain, where x_i^s represents the i -th training sample along with its corresponding ground-truth annotation y_i^s . In the Single-Domain Generalization (SDG) setting, the model is trained exclusively on $\mathcal{S}$ and evaluated on multiple unseen target domains $\mathcal{T} = \{x_i^t\}_{i=1}^{N_t}$. Following the SAMed [24] protocol, we freeze the image encoder and fully fine-tune the prompt encoder with default embeddings and the mask decoder. As shown in Fig. 1, the proposed SSUG adopts a dual-branch architecture, comprising a standard branch for the original data and an augmented branch driven by the Spectral Manifold Extrapolation (SME) module. In the SAM encoder, an Uncertainty-aware Symmetric Channel Gating (USCG) module and a Hierarchical Decoupled Consistency (HDC) constraint are synergistically integrated for domain-invariant feature learning.

2.1 Spectral Manifold Extrapolation (SME)

Domain shifts in medical imaging, largely driven by scanner and protocol variations, primarily manifest as fluctuations in low-frequency spectral components. To anticipate unseen shifts, our proposed SME module proactively probes the statistical boundaries of the source domain.

Given an input x^s , we apply the 2D Fast Fourier Transform (FFT) to decouple its amplitude $\mathcal{A}$ and phase $\mathcal{P}$. To efficiently characterize the global style

distribution of the source domain, we map the amplitude into a log-radial space. For each channel c and radial distance r , we compute the mean $\mu_{c,r}$ of the amplitude $\mathcal{A}$. To maintain the global radial statistics, we track the first-order moment $\mu_{1,c,r}^E$ and the uncentered second moment $\mu_{2,c,r}^E$ through Exponential Moving Average (EMA) with $\mu_{c,r}$ and $\mu_{c,r}^2$. The corresponding standard deviation is then estimated as $\sigma_{c,r} = \sqrt{\mu_{2,c,r}^E - (\mu_{1,c,r}^E)^2}$. Based on these estimated spectral statistics, SME simulates challenging out-of-distribution styles by performing manifold extrapolation along the statistical envelope:

$$\log \mathcal{A}_{ext}(c, r) = \mu_{1,c,r}^E + \zeta_{c,r} \cdot \sigma_{c,r}, \quad (1)$$

where $\zeta \sim \mathcal{N}(0, I)$ represents a standard normal perturbation to ensure spatial coherence. The final augmented image x_{freq} is reconstructed via Inverse FFT (IFFT) by combining the original phase $\mathcal{P}$ with the extrapolated amplitude $\mathcal{A}_{ext}$, where extrapolation is applied only within the low-to-mid FFT amplitude band with normalized radius $\rho \in [0.25, 0.45]$. SME acts as an active probe that induces parameter-level response uncertainty by simulating extreme stylistic shifts.

2.2 Uncertainty-aware Symmetric Channel Gating (USCG)

Building upon the induced domain-specific bias, we propose the USCG module to identify and suppress style-sensitive channels. USCG is a lightweight plug-in module integrated into each Transformer block of the SAM image encoder.

Channel Uncertainty Probing. Let F_{init}^l and F_{freq}^l denote the intermediate features for images x^s and x_{freq} , respectively. To quantify channel-wise sensitivity to spectral perturbations, we extract both first- and second-order statistics from each feature map. The inclusion of the standard deviation allows for a finer perception of texture and contrast shifts. The response divergence $\Delta^l \in \mathbb{R}^{B \times 2C}$ between two branches is then computed as:

$$\Delta^l = |\xi(F_{init}^l) - \xi(F_{freq}^l)|, \quad (2)$$

where B is the batch size, and C is the total number of feature channels. $\xi(F) = [\text{Mean}(F), \text{Std}(F)] \in \mathbb{R}^{B \times 2C}$ denotes the channel-wise statistics. A larger Δ^l indicates that the corresponding channel exhibits higher sensitivity to domain-specific style variations.

Warm-start Symmetric Calibration. To convert the estimated uncertainty into gating weights, we employ a bottleneck MLP to generate raw weights $g_{raw}^l = \text{Sig}(\text{MLP}(\Delta^l))$, where $\text{Sig}(\cdot)$ denotes the Sigmoid function. To preserve SAM’s pre-trained structural priors during early fine-tuning, we introduce a Warm-start Gating strategy controlled by a learnable parameter θ :

$$g^l = \mathbf{1} - \text{Sig}(\theta) \cdot (\mathbf{1} - g_{raw}^l), \quad (3)$$

where θ is initialized to a large negative value such that $\text{Sig}(\theta) \approx 0$. As training progresses, $\text{Sig}(\theta)$ adaptively increases, enabling the module to suppress style-related channels progressively. We then apply symmetric calibration to both

branches: $\tilde{F}_{init}^l = F_{init}^l \odot g^l$ and $\tilde{F}_{freq}^l = F_{freq}^l \odot g^l$. This creates an information bottleneck that prunes domain-specific attributes before they propagate to deeper layers.

2.3 Hierarchical Decoupled Consistency (HDC)

To solidify the invariant anatomical representations extracted by the USCG module, we further introduce Hierarchical Decoupled Consistency (HDC) constraints across the SAM architecture. Recognizing that shallower layers capture fine-grained boundary details while deeper layers are responsible for high-level semantic alignment, HDC enforces consistency within the calibrated feature space at multiple granularities. This design ensures that anatomical priors remain robust against spectral perturbations throughout the entire encoding process.

The primary consistency term $\mathcal{L}_{emb}$ is applied to the final image embeddings E , which directly interface with the mask decoder. Simultaneously, an auxiliary consistency term $\mathcal{L}_{feat}$ permeates the intermediate features across L transformer blocks. To account for the non-directional nature of style drift, we employ a symmetric L1 formulation:

$$\mathcal{L}_{emb} = \|E_{init} - E_{freq}\|_1, \quad \mathcal{L}_{feat} = \frac{1}{L} \sum_{l=1}^L \left(\|\tilde{F}_{init}^l - \tilde{F}_{freq}^l\|_1 \right), \quad (4)$$

where $\tilde{F}^l$ denotes the calibrated feature map from the USCG module at the l -th encoder block.

Total Objective. The SSUG framework is optimized in an end-to-end manner. The total loss $\mathcal{L}_{total}$ integrates the task-specific segmentation losses (Dice and Cross-Entropy) for both the initial and SME-augmented images with the hierarchical consistency terms:

$$\mathcal{L}_{total} = \mathcal{L}_{seg}^{init} + \mathcal{L}_{seg}^{freq} + \lambda_1 (\mathcal{L}_{emb} + \mathcal{L}_{feat}), \quad (5)$$

where $\lambda_1 = 0.1$ is a trade-off parameter to balance the influence of the hierarchical consistency constraints against the segmentation tasks.

3 Experiments

3.1 Experimental Settings

Fundus OC/OD Segmentation. We evaluated our method on a DG benchmark for Optic Cup (OC) and Disc (OD) segmentation¹, comprising Drishti-GS [17] (Domain 1), RIM-ONE [4] (Domain 2), and REFUGE [15] training and validation sets (Domains 3 and 4). In SDG settings, Domain 1 (50 train, 51 test) served as the source domain, with others as unseen targets. We follow DoFE [19] to use some basic data augmentation, and all images are resized to 512×512 .

¹ <https://github.com/emma-sjwang/DoFe>

Table 1. Quantitative comparison of different methods on Single-Domain Generalizable OC/OD segmentation in Dice score (%) and ASD (pixel). Underscores highlight results that are significantly better than the second-place method (p-value < 0.05).

Methods	Domain 2		Domain 3		Domain 4		Avg
	Cup	Disc	Cup	Disc	Cup	Disc	
Dice (%) $\uparrow$							
Upper bound	81.53 $\pm$ 9.42	94.62 $\pm$ 3.01	87.57 $\pm$ 7.59	95.91 $\pm$ 1.85	88.88 $\pm$ 7.10	95.58 $\pm$ 1.98	90.44
Cutout [3]	61.98 $\pm$ 22.23	84.39 $\pm$ 10.27	66.29 $\pm$ 10.91	92.12 $\pm$ 11.58	47.52 $\pm$ 23.43	65.62 $\pm$ 28.55	69.67
AdvBias [2]	63.79 $\pm$ 23.50	82.21 $\pm$ 17.45	64.02 $\pm$ 15.76	91.01 $\pm$ 12.56	43.76 $\pm$ 18.50	69.20 $\pm$ 17.03	69.01
MixStyle [26]	60.04 $\pm$ 23.67	82.15 $\pm$ 16.24	63.10 $\pm$ 13.95	91.15 $\pm$ 12.65	45.31 $\pm$ 18.73	68.23 $\pm$ 19.87	68.33
MaxStyle [1]	60.93 $\pm$ 21.31	83.62$\pm$12.69	71.80 $\pm$ 16.71	91.10 $\pm$ 13.17	54.55 $\pm$ 18.17	82.39 $\pm$ 10.91	74.07
SLAug [18]	57.15 $\pm$ 25.98	80.01 $\pm$ 15.31	71.33 $\pm$ 11.29	92.44$\pm$5.93	62.96 $\pm$ 16.07	78.43 $\pm$ 11.17	73.72
CCSDG [7]	64.39 $\pm$ 21.91	82.58 $\pm$ 11.35	75.86 $\pm$ 14.12	89.28 $\pm$ 7.95	52.84 $\pm$ 16.04	70.86 $\pm$ 12.78	72.64
DeSAM [5][w]	65.25 $\pm$ 20.70	80.72 $\pm$ 14.52	67.42 $\pm$ 15.14	90.89 $\pm$ 8.81	45.24 $\pm$ 15.60	64.98 $\pm$ 15.55	69.08
DeSAM [5][g]	63.44 $\pm$ 21.06	80.58 $\pm$ 14.89	69.29 $\pm$ 15.18	89.86 $\pm$ 13.20	47.15 $\pm$ 17.15	63.75 $\pm$ 17.55	69.01
DAPSAM [20]	63.28 $\pm$ 22.12	81.65 $\pm$ 10.96	74.06 $\pm$ 14.28	90.09 $\pm$ 6.44	63.65 $\pm$ 16.37	77.70 $\pm$ 13.68	75.07
Nora [21]	57.98 $\pm$ 21.26	76.91 $\pm$ 15.69	76.31 $\pm$ 12.61	89.80 $\pm$ 8.67	61.70 $\pm$ 17.40	62.41 $\pm$ 17.42	70.85
Baseline	57.92 $\pm$ 21.89	76.74 $\pm$ 15.41	77.73 $\pm$ 11.91	88.92 $\pm$ 9.55	54.31 $\pm$ 19.72	61.59 $\pm$ 17.14	69.54
SSUG (Ours)	66.60$\pm$22.51	82.77 $\pm$ 13.85	79.92$\pm$10.75	91.51 $\pm$ 5.65	66.97$\pm$15.98	84.52$\pm$11.02	78.72
ASD (pixel) $\downarrow$							
Upper bound	13.10 $\pm$ 7.68	8.55 $\pm$ 5.80	8.41 $\pm$ 5.02	6.32 $\pm$ 4.02	6.07 $\pm$ 3.41	5.46 $\pm$ 2.48	8.98
Cutout [3]	39.14 $\pm$ 35.70	28.84 $\pm$ 131.80	31.16 $\pm$ 10.69	12.62 $\pm$ 22.30	70.26 $\pm$ 60.35	81.69 $\pm$ 76.17	43.95
AdvBias [2]	33.91 $\pm$ 31.93	30.81 $\pm$ 38.89	38.09 $\pm$ 27.34	15.33 $\pm$ 23.88	67.33 $\pm$ 39.52	64.03 $\pm$ 38.70	41.58
MixStyle [26]	38.41 $\pm$ 36.87	31.26 $\pm$ 41.74	37.61 $\pm$ 23.61	15.67 $\pm$ 25.16	65.67 $\pm$ 39.49	73.63 $\pm$ 55.71	43.71
MaxStyle [1]	32.28 $\pm$ 17.63	21.75$\pm$12.24	20.02 $\pm$ 12.07	11.54 $\pm$ 12.69	42.20 $\pm$ 25.68	28.81 $\pm$ 23.31	26.11
SLAug [18]	38.39 $\pm$ 26.48	26.91 $\pm$ 14.10	25.89 $\pm$ 10.38	11.48$\pm$9.73	29.27 $\pm$ 18.68	32.64 $\pm$ 23.86	27.43
CCSDG [7]	33.24 $\pm$ 29.58	28.09 $\pm$ 26.45	20.43 $\pm$ 11.96	18.78 $\pm$ 19.47	35.30 $\pm$ 19.86	49.97 $\pm$ 35.48	30.97
DeSAM [5][w]	34.92 $\pm$ 37.73	36.08 $\pm$ 41.23	33.06 $\pm$ 18.93	16.85 $\pm$ 19.19	60.23 $\pm$ 50.10	67.21 $\pm$ 43.58	41.39
DeSAM [5][g]	36.57 $\pm$ 39.38	36.26 $\pm$ 41.17	31.69 $\pm$ 22.32	17.01 $\pm$ 26.30	60.11 $\pm$ 46.66	61.13 $\pm$ 38.04	40.46
DAPSAM [20]	30.49 $\pm$ 23.97	29.94 $\pm$ 21.08	18.14 $\pm$ 12.70	13.62 $\pm$ 12.59	23.14 $\pm$ 18.12	30.77 $\pm$ 25.69	24.35
Nora [21]	36.40 $\pm$ 21.89	40.28 $\pm$ 41.65	21.60 $\pm$ 11.08	17.87 $\pm$ 19.34	28.15 $\pm$ 21.88	54.28 $\pm$ 34.83	33.10
Baseline	37.54 $\pm$ 38.30	35.14 $\pm$ 40.09	22.49 $\pm$ 12.77	18.03 $\pm$ 19.62	36.67 $\pm$ 27.12	54.38 $\pm$ 39.85	34.04
SSUG (Ours)	28.47$\pm$20.03	29.41 $\pm$ 28.06	14.54$\pm$8.46	13.07 $\pm$ 10.42	18.20$\pm$16.37	24.92$\pm$26.84	21.44

Prostate Segmentation. This benchmark [13] includes 116 MRI cases collected from six clinical sites with distinct imaging protocols: Domain A (RUNMC), B (BMC), C (I2CVB), D (UCL), E (BIDMC), and F (HK). We follow the leave-one-out setting where the model is trained on a single source domain and evaluated on the others. All slices are resampled to 384×384 .

Implementation Details. Our framework is implemented based on the SAM ‘ViT-B’ backbone, and SAM is used with its everything mode. The EMA rate in the SME module is 0.99. We use the AdamW optimizer with a unified initial learning rate of 10^{-4} and a batch size of 8. We set the training duration to 400 epochs for the Fundus OC/OD task and 200 epochs for the Prostate task. We adopt a warm-up strategy with 25 iterations for Fundus and 250 iterations for Prostate to ensure stable optimization. All experiments are performed on a single NVIDIA RTX3090 GPU using the PyTorch framework. Dice score (Dice) and Average Surface Distance (ASD) are used for evaluation. Statistical significance is assessed by a two-sided paired t-test, with p-value < 0.05 considered significant.

Table 2. Quantitative comparison of different methods on Single-domain Generalizable prostate segmentation in Dice score (%) and ASD (mm). Underscores highlight results that are significantly better than the second-place method (p-value < 0.05).

Methods	Type	A	B	C	D	E	F	Avg Dice	Avg ASD
Upper bound	U-Net	87.80	83.75	82.60	86.35	87.10	84.50	85.35	2.92
Cutout [3]	U-Net	75.10	57.85	44.80	61.10	50.30	49.55	56.45	11.46
AdvBias [2]		77.10	62.55	51.60	70.65	51.40	52.05	60.89	9.83
MixStyle [26]		73.60	59.10	43.55	62.70	52.80	51.35	57.18	10.12
MaxStyle [1]		81.00	70.80	61.70	58.55	69.60	65.80	67.91	6.88
SLAug [18]		81.20	68.40	60.30	72.10	69.05	74.20	70.88	5.62
CCSDG [7]		80.30	70.10	65.60	68.40	59.20	65.00	68.10	5.96
DeSAM [5][w]	SAM	82.55	77.80	66.20	82.45	77.10	79.50	77.60	4.73
DeSAM [5][g]		83.10	80.20	65.10	83.85	80.05	83.00	79.22	4.62
DAPSAM [20]		85.28	80.75	70.97	85.01	82.34	82.14	81.08	4.38
Nora [21]		84.43	78.64	70.47	85.57	82.13	80.05	80.22	4.23
Baseline	SAM	80.90	81.10	67.20	82.40	79.80	81.10	78.75	4.89
SSUG (Ours)		<u>86.64</u>	<u>82.49</u>	<u>78.97</u>	85.46	82.01	81.20	<u>82.79</u>	<u>3.85</u>

3.2 Comparison with State-of-the-Art Methods

We comprehensively compare our proposed SSUG with a wide range of state-of-the-art (SOTA) SDG methods. These include traditional DL-based methods (Cutout [3], AdvBias [2], MixStyle [26], MaxStyle [1], SLAug [18], and CCSDG [7]) as well as recent SAM-based frameworks (DeSAM [5], DAPSAM [20], and Nora [21]). Additionally, we reported the performance of SAMed [24], which serves as the baseline.

Results on Fundus OC/OD. Quantitative results on the Fundus benchmark are presented in Table 1. Overall, the Baseline model achieves an average Dice of 69.54%. Our SSUG establishes a new SOTA performance by a large margin, reaching an average Dice score of 78.72% and the lowest average ASD of 21.44 pixels. It outperforms the leading U-Net-based method (MaxStyle: 74.07%) and SAM-based method (DAPSAM: 75.07%). Particularly, on the most challenging target domain (Domain 4), our method achieves Dice scores of 66.97% and 84.52% for the optic cup and disc, yielding absolute improvements of 3.32% and 6.82% over the second-place DAPSAM, respectively. Although SSUG achieves the lowest average ASD, its optic disc ASD on Domain 2 is less favorable, mainly due to a few distant false-positive boundary predictions, as ASD is sensitive to outlier contours.

Results on Prostate. Table 2 summarizes the cross-site performance on 3D MRI Prostate data. Our method consistently yields superior performance across all six diverse clinical sites, achieving the highest average Dice of 82.79% and the lowest average ASD of 3.85 mm. Notably, in sites with high variability such as B and C, SSUG demonstrates remarkable stability with Dice of 82.49% and 78.97%, surpassing the noise-injection method Nora (78.64% and 70.47%).

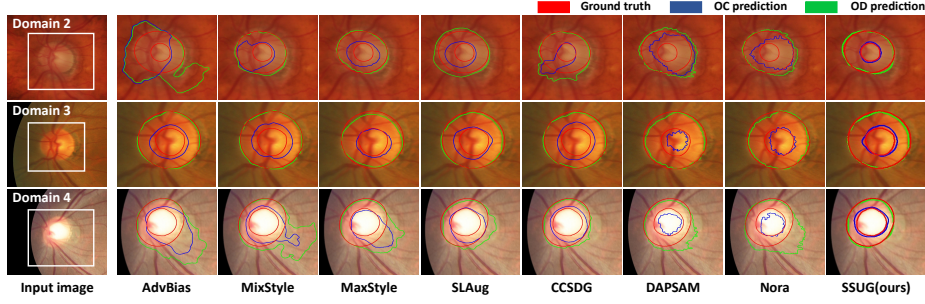


Fig. 2. Visual comparisons on Fundus OC/OD Segmentation.

Table 3. Ablation study of different components on Fundus OC/OD Segmentation. The best values are highlighted in bold.

Baseline	SME	USCG	HDC	Dice $\uparrow$	ASD $\downarrow$
✓				69.54	34.04
✓	✓			72.69	30.02
✓	✓	✓		74.29	23.64
✓	✓		✓	74.99	25.02
✓	✓	✓	✓	78.72	21.44

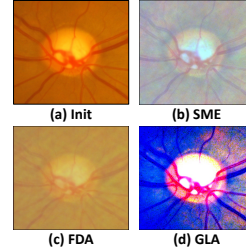


Fig. 3. Visual comparison of different augmented strategies.

Qualitative Comparison. Fig. 2 provides a visual comparison of different methods on the Fundus dataset. Compared to other SDG methods that suffer from severe false positives or fragmented boundaries, the predictions of SSUG are much closer to the ground truth, especially for the extremely blurred optic cup boundaries in Domain 4.

3.3 Ablation Study

We conduct ablation studies on the Fundus benchmark to evaluate the effectiveness of each component. As shown in Table 3, integrating the SME module improves the average Dice from 69.54% to 72.69%, demonstrating that inducing distribution shifts enables the model to anticipate unseen domain shifts. Then, introducing USCG and HDC improves the average Dice to 74.29% and 74.99%, respectively. Combining SME, USCG, and HDC achieves the highest performance, validating their closed-loop cooperation in style probing, response calibration, and anatomical consistency. Fig. 3 visually demonstrates the superiority of SME over existing augmentation strategies. Unlike Global Location-scale Augmentation (GLA) [18], which yields excessive stylistic variation, and Fourier Domain Adaptation (FDA) [23], which risks anatomical blur due to low-frequency replacement, our SME expands stylistic diversity while rigorously preserving anatomical structure.

4 Conclusion

In this work, we proposed the Symmetric Spectral Uncertainty Gating (SSUG) for SAM-based SDG medical image segmentation. The Spectral Manifold Extrapolation (SME) module proactively probes the statistical boundaries of the source spectrum to simulate various possible stylistic shifts. Subsequently, the Uncertainty-aware Symmetric Channel Gating (USCG) softly calibrates style-sensitive channels based on response divergence, while the Hierarchical Decoupled Consistency (HDC) constraint solidifies domain-invariant anatomical representations across multiple feature granularities. Extensive experiments on two widely used DG benchmarks demonstrated the superiority of SSUG over SOTA methods. Future work could further validate our method on broader datasets, incorporate boundary-aware refinement to mitigate ASD degradation in specific domains, and extend it to source-free domain adaptation.

Acknowledgements. This work was supported by the Natural Science Foundation of Sichuan Province under grant 2025ZNSFSC0455.

Disclosure of Interests. The authors have no competing interests to declare that are relevant to the content of this article.

References

1. Chen, C., Li, Z., Ouyang, C., Sinclair, M., Bai, W., Rueckert, D.: MaxStyle: Adversarial style composition for robust medical image segmentation. In: MICCAI. pp. 151–161 (2022)
2. Chen, C., Qin, C., Qiu, H., Ouyang, C., Wang, S., Chen, L., Tarroni, G., Bai, W., Rueckert, D.: Realistic adversarial data augmentation for MR image segmentation. In: MICCAI. pp. 667–677 (2020)
3. DeVries, T., Taylor, G.W.: Improved regularization of convolutional neural networks with cutout. arXiv preprint arXiv:1708.04552 (2017)
4. Fumero, F., Alayón, S., Sanchez, J.L., Sigut, J., Gonzalez-Hernandez, M.: RIM-ONE: An open retinal image database for optic nerve evaluation. In: CBMS. pp. 1–6 (2011)
5. Gao, Y., Xia, W., Hu, D., Wang, W., Gao, X.: DeSAM: Decoupled segment anything model for generalizable medical image segmentation. In: MICCAI. pp. 509–519 (2024)
6. Gu, R., Zhang, J., Wang, G., Lei, W., Song, T., Zhang, X., Li, K., Zhang, S.: Contrastive semi-supervised learning for domain adaptive segmentation across similar anatomical structures. IEEE Transactions on Medical Imaging **42**(1), 245–256 (2022)
7. Hu, S., Liao, Z., Xia, Y.: Devil is in channels: Contrastive single domain generalization for medical image segmentation. In: MICCAI. pp. 14–23 (2023)
8. Huang, Y., Yang, X., Liu, L., Zhou, H., Chang, A., Zhou, X., Chen, R., Yu, J., Chen, J., Chen, C., et al.: Segment anything model for medical images? Medical Image Analysis **92**, 103061 (2024)

9. Huang, Y., Han, L., Dou, H.: Generative feature style augmentation for domain generalization in medical image segmentation. *Pattern Recognition* **162**, 111416 (2025)
10. Kirillov, A., Mintun, E., Ravi, N., Mao, H., Rolland, C., Gustafson, L., Xiao, T., Whitehead, S., Berg, A.C., Lo, W.Y., et al.: Segment anything. In: *ICCV*. pp. 4015–4026 (2023)
11. Li, H., Li, H., Zhao, W., Fu, H., Su, X., Hu, Y., Liu, J.: Frequency-mixed single-source domain generalization for medical image segmentation. In: *MICCAI*. pp. 127–136 (2023)
12. Liu, C., Cao, Y., Su, X., Zhu, H.: Universal frequency domain perturbation for single-source domain generalization. In: *ACM MM*. pp. 6250–6259 (2024)
13. Liu, Q., Dou, Q., Heng, P.A.: Shape-aware meta-learning for generalizing prostate mri segmentation to unseen domains. In: *MICCAI*. pp. 475–485 (2020)
14. Mazurowski, M.A., Dong, H., Gu, H., Yang, J., Konz, N., Zhang, Y.: Segment anything model for medical image analysis: an experimental study. *Medical Image Analysis* **89**, 102918 (2023)
15. Orlando, J.I., Fu, H., Breda, J.B., van Keer, K., Bathula, D.R., Diaz-Pinto, A., Fang, R., Heng, P.A., Kim, J., Lee, J., et al.: REFUGE Challenge: A unified framework for evaluating automated methods for glaucoma assessment from fundus photographs. *Medical Image Analysis* **59**, 101570 (2020)
16. Shen, D., Wu, G., Suk, H.I.: Deep learning in medical image analysis. *Annual Review of Biomedical Engineering* **19**, 221–248 (2017)
17. Sivaswamy, J., Krishnadas, S., Chakravarty, A., Joshi, G., Tabish, A.S., et al.: A comprehensive retinal image dataset for the assessment of glaucoma from the optic nerve head analysis. *JSM Biomedical Imaging Data Papers* **2**(1), 1004 (2015)
18. Su, Z., Yao, K., Yang, X., Wang, Q., Sun, J., Huang, K.: Rethinking data augmentation for single-source domain generalization in medical image segmentation. In: *AAAI* (2023)
19. Wang, S., Yu, L., Li, K., Yang, X., Fu, C.W., Heng, P.A.: DoFE: Domain-oriented feature embedding for generalizable fundus image segmentation on unseen datasets. *IEEE Transactions on Medical Imaging* (2020)
20. Wei, Z., Dong, W., Zhou, P., Gu, Y., Zhao, Z., Xu, Y.: Prompting segment anything model with domain-adaptive prototype for generalizable medical image segmentation. In: *MICCAI*. pp. 533–543 (2024)
21. Wei, Z., Wu, C., Du, H., Yu, R., Du, B., Xu, Y.: Noise-robust tuning of SAM for domain generalized ultrasound image segmentation. In: *MICCAI*. pp. 476–486 (2025)
22. Wu, J., Wang, Z., Hong, M., Ji, W., Fu, H., Xu, Y., Xu, M., Jin, Y.: Medical SAM adapter: Adapting segment anything model for medical image segmentation. *Medical image analysis* **102**, 103547 (2025)
23. Yang, Y., Soatto, S.: FDA: Fourier domain adaptation for semantic segmentation. In: *CVPR*. pp. 4085–4095 (2020)
24. Zhang, K., Liu, D.: Customized segment anything model for medical image segmentation. *arXiv preprint arXiv:2304.13785* (2023)
25. Zhang, L., Wang, X., Yang, D., Sanford, T., Harmon, S., Turkbey, B., Wood, B.J., Roth, H., Myronenko, A., Xu, D., et al.: Generalizing deep learning for medical image segmentation to unseen domains via deep stacked transformation. *IEEE Transactions on Medical Imaging* **39**(7), 2531–2540 (2020)
26. Zhou, K., Yang, Y., Qiao, Y., Xiang, T.: Domain generalization with MixStyle. In: *ICLR* (2021)

27. Zhou, Z., Qi, L., Shi, Y.: Generalizable medical image segmentation via random amplitude mixup and domain specific image restoration. In: ECCV (2022)
28. Zhou, Z., Qi, L., Yang, X., Ni, D., Shi, Y.: Generalizable cross-modality medical image segmentation via style augmentation and dual normalization. In: CVPR. pp. 20856–20865 (2022)