Prompting Segment Anything Model with Domain-Adaptive Prototype for Generalizable Medical Image Segmentation

Zhikai Wei[1], Wenhui Dong[1], Peilin Zhou[1], Yuliang Gu[1], Zhou Zhao[2], and Yongchao Xu [1(✉)]

[1] National Engineering Research Center for Multimedia Software, Institute of Artificial Intelligence, School of Computer Science, Medical Artificial Intelligence Research Institute of Renmin Hospital, Wuhan University, Wuhan, China
yongchao.xu@whu.edu.cn
[2] School of Computer Science, Central China Normal University, Wuhan, China

Abstract. Deep learning based methods often suffer from performance degradation caused by domain shift. In recent years, many sophisticated network structures have been designed to tackle this problem. However, the advent of large model trained on massive data, with its exceptional segmentation capability, introduces a new perspective for solving medical segmentation problems. In this paper, we propose a novel **D**omain-**A**daptive **P**rompt framework for fine-tuning the **S**egment **A**nything **M**odel (termed as **DAPSAM**) to address single-source domain generalization (SDG) in segmenting medical images. DAPSAM not only utilizes a more generalization-friendly adapter to fine-tune the large model, but also introduces a self-learning prototype-based prompt generator to enhance model's generalization ability. Specifically, we first merge the important low-level features into intermediate features before feeding to each adapter, followed by an attention filter to remove redundant information. This yields more robust image embeddings. Then, we propose using a learnable memory bank to construct domain-adaptive prototypes for prompt generation, helping to achieve generalizable medical image segmentation. Extensive experimental results demonstrate that our DAPSAM achieves state-of-the-art performance on two SDG medical image segmentation tasks with different modalities. The code is available at https://github.com/wkklavis/DAPSAM.

Keywords: Single domain generalization · Medical image segmentation · Segment Anything Model · Prompt learning.

1 Introduction

The advancement of deep neural networks has led to significant progress in the field of medical image segmentation. Most methods have shown notable performance when the training and testing data share the same distribution. However, distribution shift (also known as domain shift [2]) leads to a decline

in performance, hindering the practical application of deep learning methods in real-world scenarios. In medical image segmentation tasks, this shift occurs more frequently due to discrepancies in imaging distribution caused by non-uniform characteristics of imaging equipment, varying operator skills, and factors such as patient radiation exposure and imaging time. Unlike unsupervised domain adaptation [27] and multi-source domain generalization [6,11], single domain generalization (SDG) is a more practical but challenging setting, under which only the labeled data from one source domain is used to train the model.

Traditional CNNs mainly focus on style augmentation at the image [25,28,31] or feature level [3,19,14] against domain shifts. CCSDG [14] incorporates contrastive feature disentanglement into a segmentation backbone. Recently, Vision Transformers have been shown to be significantly more robust in the out-of-distribution generalization [12,17]. In particular, the Segment Anything Model (SAM) [18], trained on more than 1 billion masks, has achieved unprecedented generalization capabilities on a variety of natural images. Some works have shown favorable results when applying SAM to medical image segmentation [33,9,23,24]. DeSAM [9] modifies SAM's decoder to decouple mask generation and prompt embeddings while leveraging pretrained weights, but without fully utilizing the capability and adaptability of the encoder. These developments showcase the potential of a robust huge segmentation model by leveraging a pre-trained SAM, eliminating the necessity for crafting a complex data augmentation method.

The accuracy of SAM heavily relies on the design of prompt information, such as dots and boxes. However, these suitable prompts often require interaction with humans. This type of prompt generation relies on subjective human judgments, often requiring several attempts to find the right prompt.

We introduce a novel prototype-based prompt generation module capable of automatically generating prompts specifically suited for the current image segmentation, which are weakly domain-specific. We aim to generate domain-adaptive prompts by leveraging features learned from the source domain. When confronted with unseen images, we utilize stored feature knowledge to generate instance-level strongly correlated and domain-adaptive prompts that guide the mask decoder in the segmentation process. We implement memory and storage functionality using a parameterized memory bank, taking inspiration from [10]. Similar to few-shot learning, we aspire to have the module serve as guiding support features when encountering target query features.

To further ensure that the feature information stored in the memory bank is more robust, we redesign a fine-tuning structure to fully harness the model's generalization capability. We use the vanilla adapter structure [5,30] to fine-tune the encoder as the basic model. Low-level features contain more contour information [22,35], which are crucial for medical image segmentation [26]. Motivated by this, we propose a new generalized adapter structure, in which low-level information is first mixed with intermediate features. Then, we further introduce a selective attention mechanism [29,20] to suppress information that is detrimental to generalization. After fine-tuning each layer with the generalized adapters

in the encoder, we obtain more robust features, further assisting the prompt generation module.

We evaluate the proposed method termed DAPSAM on two widely used generalizable medical image segmentation benchmarks. Experiments on different types of datasets show that DAPSAM significantly/consistently outperforms previous CNN-based and some other SAM-based methods for single out-of-distribution generalization in medical image segmentation.

Our main contributions are summarized: **1)**We propose a novel domain-adaptive prompt generator using prototype-based memory bank learned from source domain images. This generates domain-weakly-correlated but instance-strongly-correlated prompt, making use of the rich prior knowledge from pre-trained large model for generalization. **2)**We propose to redesign the adapters in each transformer block by integrating low-level features into intermediate features, followed by a channel attention filter to improve the robustness of image embeddings. **3)** Extensive experiments show that our DAPSAM outperforms previous state-of-the-art methods on two different types of SDG medical image segmentation tasks.

2 Method

The single source domain problem is defined as training on a single source domain $\mathcal{D}^s = \{x_i^s, y_i^s\}_{i=1}^{N_s}$, where x_i^s and y_i^s denote the source image and corresponding label, and then testing model performance on unseen test domains $\mathcal{D}^t = \{D_1^t, D_2^t, \ldots, D_n^t\}$. We use SAM's encoder and decoder as the baseline model. Specifically, we freeze the encoder and adopt two trainable MLP-structured adapters for each layer of the encoder following [5,30] for its efficiency and scalability. The decoder is set to fully trainable. Following SAMed [33], we change the original prediction of SAM to semantic segmentation output.

2.1 Generalized Adapter

For an image I of dimension $H \times W$, we first get the initial image embedding e_0 through the frozen Patch Embedding layer of ViT. Then we obtain low-level feature F_{low} from e_0 through a simple trainable linear layer.

In medical images, low-level information such as contours is crucial for the final image segmentation, especially in segmenting organ structures and areas of pathology. We provide low-level feature F_{low} to each intermediate feature F in every adapter. We add F_{low} and F to obtain a mixed feature F_{fuse}.

Then, we use a selective attention mechanism along the channel dimension to filter out information not conducive to generalization [20], obtaining more robust features. Channel filtering first involves separately applying global average pooling and global max pooling to the fused feature along the spatial dimensions. The results are then added together, and a sigmoid function is followed to generate a mask which is applied to the fused features. The filtering process

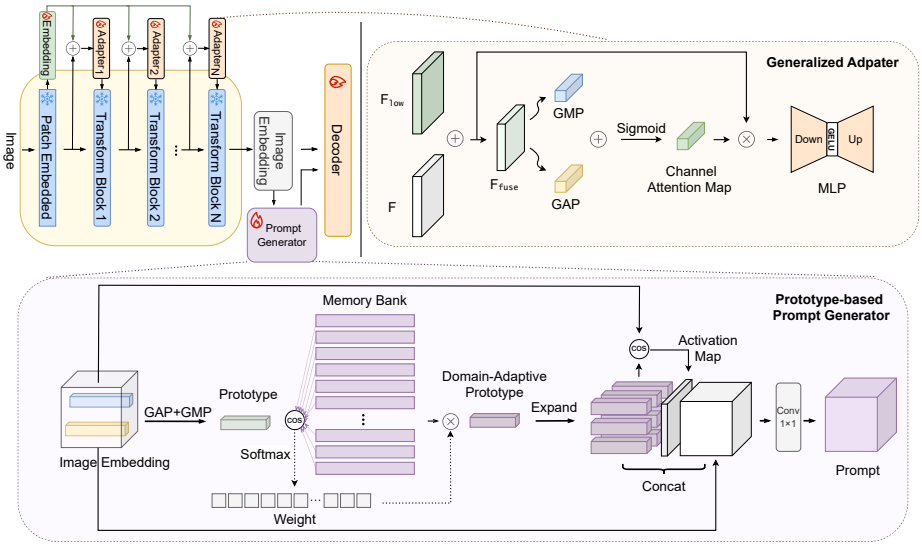

Fig. 1. The pipeline of the proposed DAPSAM. We design a generalization framework to fine-tune SAM, with generalized adapters (top right) to obtain robust features and a prompt generation module (bottom) to generate instance-related source domain prototypes for target image segmentation.

can be formulated as:

$$F_{filtered} = F_{fuse} \otimes \sigma(\mathrm{GAP}(F_{fuse}) + \mathrm{GMP}(F_{fuse})), \tag{1}$$

where σ denotes the sigmoid function and $\otimes$ denotes element-wise multiplication. $\mathrm{GAP}(\cdot)$ and $\mathrm{GMP}(\cdot)$ respectively denote the global average and max pooling operations along the spatial dimension.

Spatial dimension filtering can disrupt the spatial structure of features, which is often crucial for segmentation. Therefore, different from [29,20], we only employ channel-dimension filtering in our adapters.

Filtered features are then passed through the vanilla adapter structure, which efficiently and effectively performs adaptation across all layers:

$$F^{'} = F + \mathrm{MLP}_{\mathrm{up}}(\mathrm{GELU}(\mathrm{MLP}_{\mathrm{down}}(F_{filtered}))), \tag{2}$$

where F represents the original intermediate feature and $F^{'}$ represents the features after adapter. $\mathrm{GELU}(\cdot)$ stands for the GELU activation function, and $\mathrm{MLP}_{\mathrm{down}}(\cdot)$ and $\mathrm{MLP}_{\mathrm{up}}(\cdot)$ denote the linear layers for downward and upward projection, respectively.

2.2 Prototype-based Prompt Generator

After fine-tuning the image through the SAM encoder, the resulting image embedding is input into the subsequent mask decoder. Inspired by prompt learn-

ing [16,18], we propose to utilize a learnable memory bank to store robust information. Since prototype vector can capture global information from the feature map and form a continuous semantic space [8], we adopt prototype generated from embedding to interact with the memory bank.

Firstly, we employ global average and max pooling on the embedding to obtain instance-level prototype p_i. For each image x_i embedded into e_i, p_i is given by:

$$p_i = \mathrm{GAP}(e_i) + \mathrm{GMP}(e_i), \tag{3}$$

The memory bank is designed as a parameterized matrix $M \in \mathbb{R}^{N \times C}$ with random initialization, where N represents the number of prototypes in the memory bank, and C represents their dimension. Given a prototype vector $p_i \in \mathbb{R}^{1 \times C}$ of an image embedding, the memory bank module utilizes stored knowledge to generate a domain-adaptive and robust prototype $\widehat{p_i}$:

$$\widehat{p_i} = p_i \cdot M = \sum_{j=1}^{N} w_{i,j} m_j, \tag{4}$$

where m_j represents the j-th prototype in the memory bank, and $w_{i,j}$ represents the similarity weight between the prototype p_i of the image and $\mathrm{m_j}$.

We compute each weight $w_{i,j}$ via a softmax operation:

$$w_{i,j} = \frac{\exp(\mathrm{Sim}(p_i, m_j))}{\sum_{j=1}^{N} \exp(\mathrm{Sim}(p_i, m_j))}, \quad \mathrm{Sim}(p_i, m_j) = \frac{p_i \cdot {m_j}^{\mathsf{T}}}{\|p_i\| \|m_j\|}, \tag{5}$$

where $\mathrm{Sim}(\cdot, \cdot)$ denotes the cosine similarity operation.

For each image, $\widehat{p_i}$ is the adjusted and more robust prototype feature after being updated through the memory bank. To better guide the embedding, we first compute the cosine similarity between $\widehat{p_i}$ and e_i to generate an activation map A_i as the guidance information:

$$A_i = \mathrm{Sim}(\zeta_{h \times w}(\widehat{p_i}), e_i), \tag{6}$$

where $\zeta_{h \times w}(\cdot)$ expands the given vector to the same spatial size $h \times w$ as e_i.

Then, we concatenate $\widehat{p_i}$, e_i, and the map A_i and input them into a 1×1 convolution to generate a specific prompt for image embedding:

$$Prompt_i = \mathrm{Conv}_{1 \times 1}([\widehat{p_i}, A_i, e_i]), \tag{7}$$

where $\mathrm{Conv}_{1 \times 1}(\cdot)$ refers to a convolution layer with a kernel size of 1, which is used to perform dimension reduction.

Overall, we introduce a novel module to store learned information, compute instance-level difference and generate domain-adaptive prompt. We project all target domain knowledge into the latent space and use source domain knowledge of the memory bank to represent them, which helps to align the source and target domain. This helps to improve the model's generalization ability.

Table 1. Quantitative comparison of our DAPSAM and some state-of-the-art single-source domain generalization methods on prostate dataset. The best and second-best are **bolded** and underlined, respectively. Each column represents leave-one-out results for the model trained on the corresponding domain while testing on the other domains.

Method	Model	A	B	C	D	E	F	Average
Upper bound [15]	U-Net	85.38	83.68	82.15	85.21	87.04	84.29	84.63
AdvBias [4]	U-Net	77.45	62.12	51.09	70.20	51.12	50.69	60.45
RandConv [32]		75.52	57.23	44.21	61.27	49.98	54.21	57.07
MixStyle [34]		73.04	59.29	43.00	62.17	53.12	50.03	56.78
MaxStyle [3]		81.25	70.27	62.09	58.18	70.04	67.77	68.27
CSDG [25]		80.72	68.00	59.78	72.40	68.67	70.78	70.06
CCSDG [14]		80.62	69.52	65.18	67.89	58.99	63.27	67.58
DeSAM [9][whole]	ViT	82.30	78.06	66.65	82.87	77.58	79.05	77.75
DeSAM [9][grid]		82.80	80.61	64.77	83.41	80.36	**82.17**	79.02
SAMed [33]		80.42	**81.44**	66.75	82.09	80.19	80.17	78.51
Baseline	ViT	84.42	79.79	64.83	83.49	80.50	80.18	78.87
DAPSAM (Ours)		**86.34**	81.05	**70.81**	**85.28**	**82.91**	81.48	**81.31**

2.3 Training objective

Following SAMed [33] and TriD [6], we combine cross entropy loss and dice loss to supervise the entire training process on the source domain:

$$\mathcal{L} = (1 - \lambda)\mathcal{L}_{CE} + \lambda\mathcal{L}_{Dice}, \tag{8}$$

where λ denotes the weight to balance these two loss terms.

3 Experiments

3.1 Experimental Settings

The prostate dataset. The prostate dataset [21] comprises 116 MRI cases from six different domains, namely A: RUNMC, B: BMC, C: I2CVB, D: UCL, E: BIDMC, and F: HK. These cases were collected from three distinct public datasets used for the purpose of prostate segmentation. The slices are resized to a uniform 384×384 resolution with consistent voxel spacing. We employ the Dice Similarity Coefficient (DSC) for the evaluation.

The RIGA+ dataset. The multi-domain joint OC/OD segmentation dataset RIGA+ [1,7,13] is used in this paper. This dataset encompasses annotated fundus images from five distinct domains: BinRushed, Magrabia, BASE1, BASE2 and BASE3. For our segmentation model, we select BinRushed and Magrabia as the source domains for training and subsequently evaluate the model's performance on the remaining three domains regarded as target domains. The DSC is also employed as the metric to quantify the segmentation quality.

Implementation Details: The rank of the adapter is set to 4 for both efficiency and performance optimization. All training is conducted using the 'ViT-B' version of SAM. The initial learning rate is set to $5e^{-4}$, and the weight decay for

Table 2. Quantitative comparison of our DAPSAM and some state-of-the-art domain generalization methods on RIGA+ dataset. The best and second-best are **bolded** and underlined, respectively. In the upper part and lower part, BinRushed [Rows 3-13] and Magrabia [Rows 14-24] are used as the corresponding source domain, respectively. We run the proposed DAPSAM three times and report the mean and standard deviations.

Method	BASE1		BASE2		BASE3		Average	
	D_{OD}	D_{OC}	D_{OD}	D_{OC}	D_{OD}	D_{OC}	D_{OD}	D_{OC}
CSDG [25]	93.56±0.13	81.00±1.01	94.38±0.23	83.79±0.58	93.87±0.03	83.75±0.89	93.93	82.85
ADS [31]	94.07±0.29	79.60±5.06	94.29±0.38	81.17±3.72	93.64±0.28	81.08±4.97	94.00	80.62
MaxStyle [3]	94.28±0.14	82.61±0.67	86.65±0.76	74.71±2.07	92.36±0.39	82.33±1.24	91.09	79.88
SLAug [28]	95.28±0.12	83.31±1.10	95.49±0.16	81.36±2.51	95.57±0.06	84.38±1.39	95.45	83.02
D-Norm [36]	94.57±0.10	81.81±0.76	93.67±0.11	79.16±1.80	94.82±0.28	83.67±0.60	94.35	81.55
CCSDG [14]	95.73±0.08	86.13±0.07	95.73±0.09	86.28±0.58	95.45±0.04	86.77±0.19	95.64	86.57
DeSAM [9][w]	89.33±2.53	79.68±2.42	93.44±0.89	82.97±0.01	91.51±1.79	82.70±1.34	91.42	81.78
DeSAM [9][g]	91.79±1.62	80.87±0.11	92.57±2.04	82.95±1.62	93.66±0.07	84.19±1.79	92.67	82.67
SAMed [33]	95.28±0.07	84.24±0.10	94.11±0.10	80.21±0.64	94.84±0.08	82.60±0.32	94.74	82.35
Baseline	95.86±0.18	86.30±0.53	95.96±0.26	80.90±0.30	96.32±0.23	86.33±0.34	96.05	84.51
DAPSAM	**96.34**±0.17	**88.24**±0.16	**96.10**±0.10	**86.31**±0.13	**96.34**±0.14	**88.77**±0.21	**96.26**	**87.87**
CSDG [25]	89.67±0.76	75.39±3.22	87.97±1.04	76.44±3.48	89.91±0.64	81.35±2.81	89.18	77.73
ADS [31]	90.75±2.42	77.78±4.23	90.37±2.07	79.60±3.34	90.34±2.93	79.99±4.02	90.48	79.12
MaxStyle [3]	91.63±0.12	78.74±1.95	90.61±0.45	80.12±0.90	91.22±0.07	81.90±1.14	91.15	80.25
SLAug [28]	93.08±0.17	80.70±0.35	92.70±0.12	80.15±0.43	92.23±0.16	80.89±0.14	92.67	80.58
D-Norm [36]	92.35±0.37	79.02±0.39	91.23±0.29	80.06±0.26	92.09±0.28	79.87±0.25	91.89	79.65
CCSDG [14]	94.78±0.03	84.94±0.36	95.16±0.09	85.68±0.28	95.00±0.09	85.98±0.29	94.98	85.53
DeSAM [9][w]	82.45±2.61	69.66±2.94	84.97±0.32	75.75±1.36	83.86±2.99	74.74±2.54	83.76	73.38
DeSAM [9][g]	81.39±3.29	67.88±3.02	83.95±0.93	76.33±0.11	79.99±1.65	73.05±1.45	84.50	72.42
SAMed [33]	95.41±0.10	85.26±0.38	95.36±0.13	84.25±0.27	95.38±0.10	84.76±0.28	95.38	84.76
Baseline	95.50±0.33	86.63±0.22	95.88±0.24	88.29±0.35	**96.37**±0.23	87.61±0.30	95.92	87.51
DAPSAM	**96.22**±0.18	**86.74**±0.36	**96.32**±0.16	**89.59**±0.24	96.35±0.20	**88.12**±0.22	**96.30**	**88.15**

the AdamW optimizer is set to 0.1. We also adopt the warm-up strategy following SAMed [33], 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 Eq. (8) is set to 0.8. The baseline of our method is described at the begining of Section 2.

3.2 Comparison with SOTA Methods

Results on prostate are presented in Table 1. Compared to the traditional CNN-based U-Net structure, the ViT-based methods designed on SAM show superior performance. Our method outperforms the best CNN-based method and some recent SAM-based methods on the prostate dataset. Specifically, compared to the baseline, our method achieves a 2.44% improvement. Moreover, compared to the recently proposed SAM-based SDG medical segmentation method, DeSAM [9], our approach exhibits a notable 2.29% enhancement.

Results on RIGA+ are presented in Table 2. Our DAPSAM still achieves superior performance. When using BinRushed as the source domain, DAPSAM surpasses the CNN-based state-of-the-art CCSDG [14] by 0.62% (96.26% vs.

Table 3. Ablation study on the effect of different components on prostate. Our adapter component consists of two parts: low-level feature integration (LLFI) and filtering (Filter). PPG: Prototype-based Prompt Generator.

Baseline	LLFI	Filter	PPG	Average
✓				78.87
✓	✓			79.29
✓		✓		79.52
✓	✓	✓		79.97
✓			✓	80.31
✓	✓	✓	✓	81.31

Table 4. Ablation study of the memory bank size on prostate segmentation. We vary the number of prototypes stored in the memory bank. The first line is the baseline model without domain-adaptive prompt generator.

Num	Params(K)	FLOPs(M)	Averaged
0	0	0	78.87
64	16	75.67	79.15
128	32	75.71	79.42
256	**64**	75.77	**80.31**
512	128	75.90	79.34
1024	256	76.16	79.13
2048	512	76.69	78.92

95.64%) and 1.30% (87.87% vs. 86.57%). Our method also outperforms other SAM-based methods and baseline. With Magrabia as the source domain, while the baseline shows impressive results, DAPSAM further improves upon this performance. These SOTA results further justify the robustness and competitiveness of our DAPSAM.

More additional experimental results can be found in the **supplementary**.

3.3 Ablation Studies

We conduct extensive ablation studies of our method on the prostate dataset.

Effect of different components. We first assess the impact of the generalized adapter. As demonstrated in the first to fourth rows of Table 3, when supplementing only low-level features, the model shows a slight improvement. Notably, using filtering mechanisms to remove redundant information leads to further enhancement. These results reveal that our design not only supplements crucial low-level information in medical image segmentation but also enhances the robustness of intermediate features.

We further explore the role of the Prototype-based Prompt Generator module (PPG). The results presented in the fifth row of Table 3 confirm that incorporating PPG yields a 1.44% boost in the average Dice score relative to the baseline. This improvement distinctly highlights the PPG module's ability to proficiently utilize the knowledge acquired, thereby significantly enhancing the network's generalization capacity and robustness.

Effect of the memory bank size. We evaluate the impact of the hyperparameter N involved in Eq. (4). As depicted in Table 4, when N is set to a lower value, suboptimal results implies that a smaller memory bank cannot fully learn all the information. Conversely, an excessively high value of N leads to a decline in performance, since a too large memory bank tends to overfit the source domain information. Optimal performance is achieved for N is set to 256.

4 Conclusion

In this paper, we first analyze the performance of fine-tuning large model SAM for domain generalization, and find its excellent potential for generalizable medical image segmentation. We then propose a novel prototype-based domain-adaptive prompt generator to mine such potential of SAM in SDG medical image segmentation. We also propose a more generalization-friendly adapter that improves the robustness of image embedding, further boosting the model's generalization ability. The proposed method termed DAPSAM outperforms some state-of-the-art CNN-based and SAM-based methods on two widely used benchmarks for generalizable medical image segmentation.

Acknowledgments. This work was supported in part by the National Key Research and Development Program of China (2023YFC2705700), NSFC 62222112 and 62176186, the Postdoctoral Fellowship Program of CPSF (GZC20230924), the NSF of Hubei Province of China (2024AFB245), and CAAI Huawei MindSpore Open Fund.

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

References

1. Almazroa, A., Alodhayb, S., Osman, E., Ramadan, E., Hummadi, M., Dlaim, M., Alkatee, M., Raahemifar, K., Lakshminarayanan, V.: Retinal fundus images for glaucoma analysis: the riga dataset. In: Medical Imaging 2018: Imaging Informatics for Healthcare, Research, and Applications. vol. 10579, pp. 55–62 (2018)
2. Carlucci, F.M., D'Innocente, A., Bucci, S., Caputo, B., Tommasi, T.: Domain generalization by solving jigsaw puzzles. In: CVPR. pp. 2229–2238 (2019)
3. 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)
4. 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)
5. Chen, S., Ge, C., Tong, Z., Wang, J., Song, Y., Wang, J., Luo, P.: Adaptformer: Adapting vision transformers for scalable visual recognition. In: NeurIPS (2022)
6. Chen, Z., Pan, Y., Ye, Y., Cui, H., Xia, Y.: Treasure in distribution: A domain randomization based multi-source domain generalization for 2d medical image segmentation. In: MICCAI. vol. 14223, pp. 89–99 (2023)
7. Decencière, E., Zhang, X., Cazuguel, G., Lay, B., Cochener, B., Trone, C., Gain, P., Ordonez, R., Massin, P., Erginay, A., et al.: Feedback on a publicly distributed image database: the messidor database. Image Analysis & Stereology **33**(3), 231–234 (2014)
8. Dong, N., Xing, E.P.: Few-shot semantic segmentation with prototype learning. In: BMVC. vol. 3 (2018)
9. Gao, Y., Xia, W., Hu, D., Gao, X.: Desam: Decoupling segment anything model for generalizable medical image segmentation. arXiv preprint arXiv:2306.00499 (2023)

10. Gong, D., Liu, L., Le, V., Saha, B., Mansour, M.R., Venkatesh, S., Hengel, A.v.d.: Memorizing normality to detect anomaly: Memory-augmented deep autoencoder for unsupervised anomaly detection. In: ICCV. pp. 1705–1714 (2019)
11. Gu, R., Wang, G., Lu, J., Zhang, J., Lei, W., Chen, Y., Liao, W., Zhang, S., Li, K., Metaxas, D.N., et al.: Cddsa: Contrastive domain disentanglement and style augmentation for generalizable medical image segmentation. Medical Image Analysis **89**, 102904 (2023)
12. Guo, Y., Stutz, D., Schiele, B.: Improving robustness of vision transformers by reducing sensitivity to patch corruptions. In: CVPR. pp. 4108–4118 (2023)
13. Hu, S., Liao, Z., Xia, Y.: Domain specific convolution and high frequency reconstruction based unsupervised domain adaptation for medical image segmentation. In: MICCAI. pp. 650–659 (2022)
14. Hu, S., Liao, Z., Xia, Y.: Devil is in channels: Contrastive single domain generalization for medical image segmentation. In: MICCAI. vol. 14223, pp. 14–23 (2023)
15. Isensee, F., Jaeger, P.F., Kohl, S.A., Petersen, J., Maier-Hein, K.H.: nnu-net: a self-configuring method for deep learning-based biomedical image segmentation. Nature Methods **18**(2), 203–211 (2021)
16. Jia, M., Tang, L., Chen, B.C., Cardie, C., Belongie, S., Hariharan, B., Lim, S.N.: Visual prompt tuning. In: ECCV. pp. 709–727 (2022)
17. Kim, H., Shin, Y., Hwang, D.: Dimix: Disentangle-and-mix based domain generalizable medical image segmentation. In: MICCAI. pp. 242–251 (2023)
18. Kirillov, A., Mintun, E., Ravi, N., Mao, H., Rolland, C., Gustafson, L., Xiao, T., Whitehead, S., Berg, A.C., Lo, W.Y., Dollar, P., Girshick, R.: Segment anything. In: ICCV. pp. 3992–4003 (2023)
19. 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)
20. Lin, S., Zhang, Z., Huang, Z., Lu, Y., Lan, C., Chu, P., You, Q., Wang, J., Liu, Z., Parulkar, A., et al.: Deep frequency filtering for domain generalization. In: CVPR. pp. 11797–11807 (2023)
21. 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)
22. Liu, W., Shen, X., Pun, C., Cun, X.: Explicit visual prompting for low-level structure segmentations. In: CVPR. pp. 19434–19445 (2023)
23. Ma, J., He, Y., Li, F., Han, L., You, C., Wang, B.: Segment anything in medical images. Nature Communications **15**, 1–9 (2024)
24. 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)
25. Ouyang, C., Chen, C., Li, S., Li, Z., Qin, C., Bai, W., Rueckert, D.: Causality-inspired single-source domain generalization for medical image segmentation. IEEE Transactions on Medical Imaging **42**(4), 1095–1106 (2022)
26. Ronneberger, O., Fischer, P., Brox, T.: U-net: Convolutional networks for biomedical image segmentation. In: MICCAI. pp. 234–241 (2015)
27. Shin, H., Kim, H., Kim, S., Jun, Y., Eo, T., Hwang, D.: Sdc-uda: Volumetric unsupervised domain adaptation framework for slice-direction continuous cross-modality medical image segmentation. In: CVPR. pp. 7412–7421 (2023)
28. Su, Z., Yao, K., Yang, X., Huang, K., Wang, Q., Sun, J.: Rethinking data augmentation for single-source domain generalization in medical image segmentation. In: AAAI. vol. 37, pp. 2366–2374 (2023)

29. Woo, S., Park, J., Lee, J.Y., Kweon, I.S.: Cbam: Convolutional block attention module. In: ECCV. pp. 3–19 (2018)
30. Wu, J., Fu, R., Fang, H., Liu, Y., Wang, Z., Xu, Y., Jin, Y., Arbel, T.: Medical sam adapter: Adapting segment anything model for medical image segmentation. arXiv preprint arXiv:2304.12620 (2023)
31. Xu, Y., Xie, S., Reynolds, M., Ragoza, M., Gong, M., Batmanghelich, K.: Adversarial consistency for single domain generalization in medical image segmentation. In: MICCAI. pp. 671–681 (2022)
32. Xu, Z., Liu, D., Yang, J., Raffel, C., Niethammer, M.: Robust and generalizable visual representation learning via random convolutions. In: ICLR (2021)
33. Zhang, K., Liu, D.: Customized segment anything model for medical image segmentation. arXiv preprint arXiv:2304.13785 (2023)
34. Zhou, K., Yang, Y., Qiao, Y., Xiang, T.: Domain generalization with mixstyle. In: ICLR (2021)
35. Zhou, Y., Lu, R., Xue, F., Gao, Y.: Occlusion relationship reasoning with a feature separation and interaction network. Visual Intelligence **1**(1), 23 (2023)
36. 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)