


SR-SAM: Subspace Regularization for Domain Generalization of Segment Anything Model

Xixi Jiang¹, Chen Yang¹, Liang Zhang², Tim Kwang-Ting CHENG¹, and Xin Yang²

¹ Department of Electronic and Computer Engineering, The Hong Kong University of Science and Technology

² School of Electronic Information and Communications, Huazhong University of Science and Technology
xinyang2014@hust.edu.cn

Abstract. Parameter Efficient Fine-Tuning (PEFT) methods have been widely used to adapt foundation models like the Segment Anything Model (SAM) for better generalization in unseen domains. Despite their widespread use, PEFT often suffers from overfitting to the source training domain, which limits their generalization performance. To address this limitation, we propose a novel subspace regularization (SR) method for robust fine-tuning. Our approach iteratively removes the knowledge of task-specific directions, as identified by LoRA parameters learned from the source domain, from the subspace of pre-trained weights. This strategy effectively encourages the LoRA parameters to acquire a more diverse range of knowledge. In addition, we introduce an exponential moving average (EMA) LoRA module that aggregates historical updates of the LoRA parameters throughout the fine-tuning process. This aggregation enhances stability and the generalizability of the learned features by smoothing the trajectory of parameter updates. Our enhanced framework, SR-SAM, incorporates both subspace regularization and the EMA LoRA module to fine-tune the popular SAM model effectively. Experimental results on two widely used domain generalization benchmarks demonstrate that SR-SAM outperforms existing state-of-the-art methods, underscoring the effectiveness of our method. The source code is available at <https://github.com/xjiangmed/SR-SAM>.

Keywords: Parameter-efficient fine-tuning · Domain generalization · Segment Anything Model.

1 Introduction

Variations in imaging protocols and scanners across different institutions often lead to domain shifts in medical images, causing performance degradation in deep learning models and hindering their practical deployment. Domain generalization (DG) techniques address this challenge by enabling models to generalize from known source domains to unknown target domains. Traditional DG methods [18,9,32,4] primarily focus on extracting domain-invariant features or

designing various style transformations to augment source domain images. However, the data-driven nature of deep learning suggests that training generalized models on large-scale, diverse datasets might be a more straightforward and effective strategy compared to these complex DG methods. Recent advancements in foundation models, such as the Segment Anything model (SAM)[12], which is trained on over one billion masks, have shown outstanding zero-shot performance. Nonetheless, due to the substantial domain gap between medical images and natural images, SAM requires fine-tuning in medical scenarios to fully unleash its capabilities. The workflow of recent DG methods [5,24] has gradually shifted to using a pre-trained model as initialization, then fine-tuning it through one or more source domains, and finally evaluating the generalization performance on target domains different from the source domains.

To tailor SAM’s capabilities for various domains in specific medical downstream tasks, it is crucial to apply appropriate fine-tuning methods. Two common adaptation techniques include full fine-tuning and only fine-tuning the mask decoder [16,5,21]. However, full fine-tuning is storage-intensive, and training only the mask decoder leads to poor adaptation performance. Parameter-Efficient Fine-Tuning (PEFT) [7,6,11] methods address these limitations by freezing the pretrained model and optimizing lightweight trainable modules, significantly reducing training costs. This benefit has popularized PEFT and driven the development of various PEFT approaches to adapt SAM for downstream tasks [25,24,29,3]. Low-Rank Adaptation (LoRA) [7] stands out as a widely adopted PEFT method, injecting trainable low-rank decomposition matrices into specific layers of the transformer architecture. We observed that applying LoRA to fine-tune the SAM yields superior generalization performance on out-of-distribution (OOD) domains compared to conventional DG methods.

In this work, we address a more challenging and realistic scenario, single-source domain generalization, in which only one source domain is available at training. This demanding scenario necessitates that the model learns new knowledge from limited source data while simultaneously achieving effective generalization across diverse domains. Despite its notable advantages, employing LoRA in domain generalization scenarios presents certain challenges. A primary challenge is the tendency of LoRA modules to overfit to the training source domain, a phenomenon that becomes particularly pronounced when training on smaller datasets. While LoRA facilitates task-specific adaptation, it also heightens the risk of incorporating source domain-specific noise or biases into the model. Moreover, a single low-rank module often lacks the necessary stability and robustness to handle the variability encountered in diverse, unseen target domains.

To mitigate the overfitting problem, we analyzed the subspace similarities of LoRA parameters across different domains and discovered that domain shifts diminish the alignment of top singular directions. Based on this insight, we propose a subspace regularization method called SR-SAM, which enhances robust fine-tuning by regulating the subspace of the pre-trained model. Utilizing the task-specific update directions of the source domain discovered by the LoRA module, we iteratively truncate the corresponding knowledge from the pre-trained

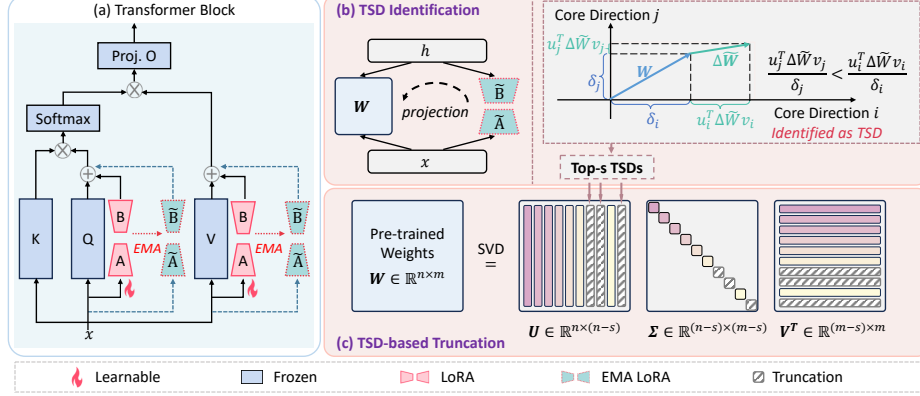


Fig. 1. Framework of SR-SAM: (a) We insert a set of LoRA and EMA LoRA modules in the query and value layers of each transformer block in the image encoder. The subspace regularization involves two main steps: (b) projecting the EMA LoRA weight onto the subspace of pretrained weight to discern the top s TSDs; (c) removing the TSDs components from the pretrained weights.

weights. This approach effectively broadens the diversity and scope of LoRA’s update directions, mitigating overfitting to the source domain. Furthermore, to ensure robust generalization across different OOD domains, we employ two sets of low-rank adapters: a standard LoRA adapter for learning task-specific information, and an exponential moving average (EMA) LoRA adapter designed to capture a more stable and generalizable representation through temporal ensembles. The results on two DG benchmarks show that SR-SAM consistently outperforms both traditional DG methods and SAM-based fine-tuning methods.

2 Method

The single-source DG task involves training on a single source domain $D_s = \{x_s^k, y_s^k\}_{k=1}^{N_s}$, where x_s^k and y_s^k denote the source image and corresponding ground truth mask. We evaluate the generalization performance on OOD target domains $D_t = \{D_t^1, D_t^2, \dots, D_t^T\}$. Our baseline model follows the SAMed [29] setup, with the image and prompt encoders frozen and the mask decoder fully trainable. We insert the trainable LoRA module into each transformer block of the image encoder to fine-tune SAM. Fig. 1 gives an overview of our proposed method. In this section, we first review LoRA and explore the correlation of LoRA parameters across domains, and then detail the SR-SAM components.

2.1 Preliminary

LoRA. Based on the observation that the updated weights typically have low intrinsic rank [13], for a pretrained weight matrix $W \in \mathbb{R}^{n \times m}$, the LoRA [7]

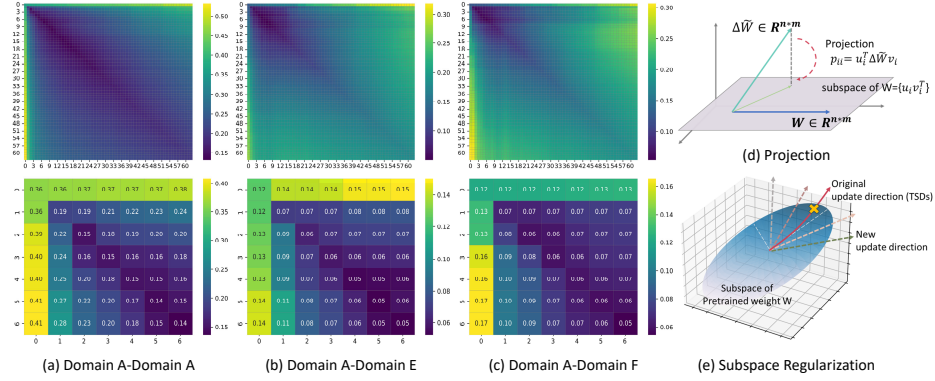


Fig. 2. Normalized subspace similarity is measured (a) between two randomly seeded runs for domain A, and (b) between domain A and E, as well as (c) between domain A and F of the Prostate dataset [15]; the second-row zooms in on the upper left corner of the first-row. (d) Visualization of projecting EMA LoRA weight ΔW to the subspace of pre-trained weight W . (e) Illustration depicting the concept behind our subspace regularization approach.

method models the change of the model as $\Delta W = AB$, where $A \in \mathbb{R}^{n \times r}$ and $B \in \mathbb{R}^{r \times m}$ with the rank of r . During training, only the injected low-rank weights A and B are updated, with W remaining frozen. A forward pass for LoRA is represented as $h = Wx + \Delta Wx = (W + AB)x$.

Pretrained weight W can be seen as a matrix in a subspace spanned by a set of linearly independent bases $\{u_i v_i^T\}$, where u_i and v_i are the left and right singular vectors of W obtained by SVD. Each singular value σ_i corresponds to a direction in the subspace. LoRA emphasizes that ΔW enhances some directions in W that are not critical for pre-training but are essential for specific downstream tasks, which LoRA calls “task-specific directions” (TSDs). TSDs refer to the directions in which model parameters need to be adjusted during the adaptation process, transitioning from the pre-trained state W to the optimal parameters W^* for a specific downstream task. LoRA-Dash [20] considers these directions as an intuitive representation of low-dimensional manifolds, highlighting the importance of task-specific directions for successful fine-tuning. However, despite the importance of these directions pointed out by LoRA and LoRA-Dash, the exploitation of TSDs remains unexplored in the DG task.

Subspace similarity. Inspired by previous work [7,20,22], LoRA’s parameters serve as holders for distinct subspaces of task gradients. For the DG task, a natural question arises: *What is the connection between the LoRA weight ΔW of different domains?* To answer this question, we fine-tune SAM via LoRA on different domains and then examine the subspace similarity [7] between ΔW in different domains. We measure the subspace similarity based on the Grassmann

distance:

$$\phi(\Delta W1, \Delta W2, i, j) = \frac{\|U_{\Delta W1}^i U_{\Delta W2}^j\|_F^2}{\min(i, j)} \in [0, 1], \quad (1)$$

where $U_{\Delta W1}^i$ represents the top- i left-singular vectors of $\Delta W1$. Greater values of $\phi(\cdot)$ indicate a larger overlap of the subspaces.

In Fig. 2 (a-c), we analyze the subspace similarity of LoRA parameters between two seeds within the same domain and across different domains. We can draw two observations from Fig. 2: (1) There is a substantial overlap in the directions of the top singular vectors when comparing runs within the same domain, indicating that these directions are crucial for capturing task-specific knowledge. Meanwhile, the less significant singular vectors may primarily capture noise. (2) Comparing across domains, the overlap in the top singular vector directions is notably reduced, suggesting that while there are inherent similarities in these essential directions, domain shifts diminish their congruence.

Motivated by these findings, we introduce a novel subspace regularization (SR) strategy specifically designed to reduce overfitting and bias during training on in-distribution source data. As shown in Fig. 2 (e), this strategy involves identifying the task-specific directions (TSDs) of the source training data and iteratively removing these directions from the pre-trained model. In this way, we limit LoRA’s excessive updates along these critical TSDs, thereby enabling a broader exploration of the optimization subspace.

2.2 Subspace Regularization

Our subspace regularization strategy consists of two steps, TSD identification and TSD-based truncation.

TSD Identification. In LoRA-Dash [20], TSDs are defined as the core directions of the pre-trained weight matrix W that undergo the largest relative changes when transitioning to W^* for a specific downstream task. Since both the optimal weight W^* and the optimal weight change ΔW^* are unknown before fine-tuning, LoRA-Dash uses the parameters learned by LoRA as an estimate of ΔW^* . Following this approach, we first train low-rank matrices (A and B) for a predefined number of steps t . Then we project the learned weight changes ΔW onto the core bases of W . As shown in Fig. 2 (d), the projection of ΔW indicates the direction in which W needs to evolve. $u_i^T \Delta W v_i$ denotes the projection operation. Subsequently, the top s directions exhibiting the highest change rates are identified as the TSDs. The change rate δ_i for each core direction σ_i is calculated as follows:

$$\delta_i = \frac{u_i^T \Delta W v_i}{\sigma_i + \epsilon}, \quad (2)$$

where u_i and v_i are the i -th left and right singular vectors of W , and σ_i is the corresponding singular value. ϵ is a small constant to avoid division by zero.

TSD-based Truncation. Contrary to enhancing focus on TSDs as in LoRA-Dash, our strategy restricts updates along these directions by modifying the base

model W . Specifically, we truncate the subspace corresponding to the TSDs, recalibrating W to:

$$W' = \sum_{i=1, i \notin TSDs}^{\min(n,m)} u_i \sigma_i v_i^T, \quad (3)$$

where W' becomes the new weight matrix. We iteratively (*i.e.*, every t steps) identify TSDs and remove these top-direction subspaces from the pre-trained weight. This truncation encourages learning from other, less dominant directions of downstream tasks, thereby preventing overfitting to the source domain and fostering better generalization against OOD domains. Notably, the truncation operation does not lose essential information. TSDs are typically minor singular value components in the pre-trained weight [20]. Additionally, prior to truncation, LoRA has already amplified and updated the truncated knowledge.

Temporal Ensemble. Capitalizing on ensembling’s strengths to improve generalization over diverse distributions [14], we devise a temporal ensemble method within the LoRA framework called EMA LoRA. This approach leverages temporal aggregation by accumulating historical LoRA weights during training, gaining a wider comprehension of the feature space. The accumulated weights $\Delta\tilde{A}, \Delta\tilde{B}$ are updated as $\Delta\tilde{A} = \alpha\Delta\tilde{A} + (1 - \alpha)\Delta A$ and $\Delta\tilde{B} = \alpha\Delta\tilde{B} + (1 - \alpha)\Delta B$, where $\alpha \in [0,1]$ is the update rate. To enhance the accuracy of TSD identification, we project the EMA LoRA weight $\Delta\tilde{W} = \Delta\tilde{A}\Delta\tilde{B}$ onto the subspace of W to identify the TSDs. Additionally, by acting as a teacher model, EMA LoRA guides the training of the LoRA student model using a distillation loss defined as

$$\mathcal{L}_{distill} = \frac{1}{N} \sum_{n \in N} KL(\tilde{p}_n || p_n), \quad (4)$$

where $KL(\cdot)$ denotes the Kullback-Leibler divergence, \tilde{p}_n and p_n represent the predicted probabilities of the n -th pixel from the teacher and student models, respectively. This distillation mechanism enhances the LoRA’s ability to generalize by learning robust predictions from the EMA LoRA. The final training object is $\mathcal{L} = \mathcal{L}_{seg} + \lambda\mathcal{L}_{distill}$, where \mathcal{L}_{seg} is a combination of cross entropy loss and dice loss.

3 Experiments

Datasets. Our method is evaluated on two cross-domain segmentation datasets: (1) **Prostate dataset** [15]: consists of MRI images from six sources: Domain A: RUNMC, B: BMC, C: I2CVB, D: UCL, E: BIDMC, and F: HK. All MRI images are resampled to a uniform spacing and resized to 384 x 384. The number of slices in each domain is 261, 384, 158, 468, 421, and 175, respectively. (2) **Polyp dataset** [31]: consists of images from Domain A: CVC-ClinicDB, B: CVC-ColonDB, C: ETIS, and D: Kvasir, with totals of 612, 380, 196, and 1,000 images respectively. The images are resized to 384 x 384.

Table 1. Comparison of Dice scores between our SR-SAM with SOTA methods on the polyp dataset.

Method	Model	A	B	C	D	Average
Upper bound [19]	U-Net	95.26	94.21	93.07	96.31	94.71
CutMix [27]	U-Net	50.72	45.42	58.17	69.12	55.86
Mixup [28]		59.37	41.57	60.17	71.25	58.09
BigAug [30]		56.79	42.10	60.26	69.56	57.18
Randaugment [4]		59.11	50.41	58.78	66.73	58.76
MCC [23]		67.88	48.38	59.74	63.89	59.98
RGIA [31]		64.64	47.74	61.73	69.17	60.82
DeSAM [5][whole]	Decoder	51.73	44.39	50.55	31.53	44.55
DeSAM [5][grid]		51.39	44.47	52.34	31.50	44.93
SAM4Med [21]		48.38	64.31	57.35	66.70	59.19
Med-SA [25]	Adapter	<u>82.81</u>	78.75	<u>80.81</u>	<u>79.88</u>	80.56
DAPSAM [24]		82.26	78.49	79.63	78.76	79.79
SAMed [29]	LoRA	82.46	79.54	79.68	79.38	80.27
H-SAM [3]		82.08	78.82	80.56	79.74	80.30
PACE [17]		81.16	79.77	79.79	78.14	79.72
PEGO [8]		82.18	<u>80.28</u>	80.76	79.33	<u>80.64</u>
LoRA-Dash [3]		81.68	79.91	80.77	79.76	<u>80.53</u>
SR-SAM (Ours)	LoRA	82.83	81.23	81.21	80.57	81.46

Implementations. All SAM-based methods use the ViT-B backbone. Our experiments are conducted on an NVIDIA RTX3090 GPU, with the training process spanning 160 epochs. The initial learning rate is configured to $5e-4$, the batch size is set to 8, and the weight decay for the AdamW optimizer is 0.1. A warm-up period of 250 iterations is implemented. The Dice Similarity Coefficient is used as the evaluation metric. For optimal baseline performance, the rank of LoRA is configured to 64. The weight λ is set to $1e-7$ and $1e-6$ for the polyp and prostate dataset. The EMA rate α is 0.999. TSD identification and truncation are performed every 4 epochs, with the truncation size fixed at 96.

Comparison with SOTA Methods. We compare SR-SAM with traditional CNN-based DG methods and SAM-based methods, with results shown in Table 1 and Table 2 for polyp and prostate datasets. SAM-based methods are categorized into: 1) Decoder-based: DeSAM decouples prompt encoding and mask prediction, while SAM4Med introduces an automatic prompt generator. 2) Adapter-based PEFT: Med-SA utilizes Adapter [6] structures, and DAPSAM introduces domain-adaptive prompts. 3) LoRA-based PEFT, with SAMed using LoRA structures, H-SAM enhancing hierarchical decoding and LoRA-Dash maximizing the impact of TSDs. Additionally, PACE and PEGO enhance generalization through consistency and orthogonal regularization, respectively. Compared to traditional DG methods, most SAM-based approaches demonstrate superior generalization performance. Adapter-based and LoRA-based PEFT methods exhibit superior performance compared to decoder-based approaches. Notably, SR-SAM surpasses both the leading CNN-based methods and recent SAM-based techniques on both datasets. Specifically, on the polyp dataset, SR-SAM achieved the best performance in all four domains and achieved a 0.82% improvement in average dice compared with the PEGO method. On the prostate dataset, SR-

Table 2. Comparison of Dice scores between our SR-SAM with SOTA methods on the prostate dataset.

Method	Type	A	B	C	D	E	F	Average
Upper bound [10]	U-Net	85.38	83.68	82.15	85.21	87.04	84.29	84.63
AdvBias [2]	U-Net	77.45	62.12	51.09	70.20	51.12	50.69	60.45
RandConv [26]		75.52	57.23	44.21	61.27	49.98	54.21	57.07
MixStyle [32]		73.04	59.29	43.00	62.17	53.12	50.03	56.78
MaxStyle [1]		81.25	70.27	62.09	58.18	70.04	67.77	68.27
CSDG [18]		80.72	68.00	59.78	72.40	68.67	70.78	70.06
CCSDG [9]		80.62	69.52	65.18	67.89	58.99	63.27	67.58
DeSAM [5] ^[whole]	Decoder	82.30	78.06	66.65	82.87	77.58	79.05	77.75
DeSAM [5] ^[grid]		82.80	80.61	64.77	83.41	80.36	82.17	79.02
SAM4Med [21]		84.08	77.29	73.98	82.40	80.47	79.04	79.54
Med-SA [25]	Adapter	84.46	83.06	68.23	85.82	83.69	80.10	80.89
DAPSAM [24]		86.34	81.05	70.81	85.28	82.91	81.48	81.31
SAMed [29]	LoRA	85.07	82.02	75.16	85.81	82.87	81.53	82.08
H-SAM [3]		85.63	83.34	76.71	84.44	83.17	81.94	82.54
PACE [17]		85.07	83.33	71.55	85.47	84.42	81.93	81.96
PEGO [8]		84.19	81.10	74.88	83.88	83.67	81.17	81.48
LoRA-Dash [3]		84.81	82.60	74.63	85.37	82.67	81.63	81.95
SR-SAM (Ours)	LoRA	87.07	83.06	77.95	86.91	84.51	82.36	83.64

Table 3. Ablation of different components.

Baseline	EMA	Trun	A	B	C	D	Average
✓			82.46	79.54	79.68	79.38	80.27
✓	✓		82.45	80.01	80.38	79.48	80.58
✓	✓	✓	82.83	81.23	81.21	80.57	81.46

Table 4. Effect of truncation size.

Size	A	B	C	D	Average
8	82.74	80.43	80.21	79.77	80.79
16	82.68	80.50	80.97	80.09	81.06
32	82.61	80.28	80.52	79.99	80.85
64	83.03	80.91	81.27	79.85	81.27
96	82.83	81.23	81.21	80.57	81.46
128	82.43	80.30	80.82	79.81	80.84

SAM outperforms the H-SAM method by 1.1% in average dice. We consistently deliver state-of-the-art results across most domains in both datasets, showcasing robust generalization capabilities. It is worth noting that our subspace regularization method significantly outperforms the PEGO method [8], which enforces orthogonal regularization between multiple, specifically four LoRA modules to encourage LoRA to learn diverse knowledge. Compared with LoRA-Dash [3], which enhances learning in the TSDs, our subspace regularization weakens learning in the TSDs of the source domain, showing superior performance.

Effect of different components. We conduct ablation experiments on the components of SR-SAM including EMA LoRA (EMA) and TSD Truncation (Trun) on the polyp dataset. The results, as shown in Table 3, reveal that both modules contribute to the performance improvement of the SR-SAM. Compared to the baseline, our method shows consistent superiority, with an average improvement of 1.19% dice.

Effect of truncation size. We analyze the influence of truncation size s on polyp dataset in Table 4. While increasing s allows for learning a broader range of knowledge, overly large values may remove essential pretrained information. We observe that the model performs best with a truncation size of 96.

4 Conclusion

In this paper, we propose to enhance the generalization ability of SAM on DG tasks from a new perspective of subspace regularization. We establish two sets of LoRA adapters, enabling the identification of the update direction of the source domain. By iteratively removing task-specific knowledge from the pre-training model, we constrain the LoRA adapters to acquire more comprehensive and diverse knowledge representations. Our experimental results validate the effectiveness of subspace regularization, highlighting the benefits of appropriate utilization of TSD in achieving excellent generalization performance.

Acknowledgments. This research was partially supported by HKSAR RGC General Research Fund (GRF) 16208823, in part by the National Science Foundation of China under Grant 62472184, and in part by the Fundamental Research Funds for the Central Universities.

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. Cheng, Z., Wei, Q., Zhu, H., Wang, Y., Qu, L., Shao, W., Zhou, Y.: Unleashing the potential of sam for medical adaptation via hierarchical decoding. In: Computer Vision and Pattern Recognition. pp. 3511–3522 (2024)
4. Cubuk, E.D., Zoph, B., Shlens, J., Le, Q.V.: Randaugment: Practical automated data augmentation with a reduced search space. In: Computer Vision and Pattern Recognition Workshops. pp. 702–703 (2020)
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. Houlsby, N., Giurugu, A., Jastrzebski, S., Morrone, B., De Laroussilhe, Q., Gesmundo, A., Attariyan, M., Gelly, S.: Parameter-efficient transfer learning for nlp. In: International Conference on Machine Learning. pp. 2790–2799 (2019)
7. Hu, E.J., Shen, Y., Wallis, P., Allen-Zhu, Z., Li, Y., Wang, S., Wang, L., Chen, W.: Lora: Low-rank adaptation of large language models. International Conference on Learning Representations (2022)
8. Hu, J., Zhang, J., Qi, L., Shi, Y., Gao, Y.: Learn to preserve and diversify: Parameter-efficient group with orthogonal regularization for domain generalization. In: European Conference on Computer Vision. pp. 198–216. Springer (2024)
9. 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)

10. 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)
11. Jia, M., Tang, L., Chen, B.C., Cardie, C., Belongie, S., Hariharan, B., Lim, S.N.: Visual prompt tuning. In: *European Conference on Computer Vision*. pp. 709–727. Springer (2022)
12. 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: *International Conference on Computer Vision*. pp. 4015–4026 (2023)
13. Li, C., Farkhoor, H., Liu, R., Yosinski, J.: Measuring the intrinsic dimension of objective landscapes. In: *International Conference on Learning Representations* (2018)
14. Lin, Y., Tan, L., Hao, Y., Wong, H.N., Dong, H., Zhang, W., Yang, Y., Zhang, T.: Spurious feature diversification improves out-of-distribution generalization. In: *International Conference on Learning Representations* (2024)
15. 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)
16. Ma, J., He, Y., Li, F., Han, L., You, C., Wang, B.: Segment anything in medical images. *Nature Communications* **15**(1), 654 (2024)
17. Ni, Y., Zhang, S., Koniusz, P.: Pace: marrying generalization in parameter-efficient fine-tuning with consistency regularization. *Advances in Neural Information Processing Systems* **37**, 61238–61266 (2025)
18. 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)
19. Ronneberger, O., Fischer, P., Brox, T.: U-net: Convolutional networks for biomedical image segmentation. In: *MICCAI*. pp. 234–241 (2015)
20. Si, C., Shi, Z., Zhang, S., Yang, X., Pfister, H., Shen, W.: Unleashing the power of task-specific directions in parameter efficient fine-tuning. *International Conference on Learning Representations* (2024)
21. Wang, H., Ye, H., Xia, Y., Zhang, X.: Leveraging sam for single-source domain generalization in medical image segmentation. *arXiv preprint arXiv:2401.02076* (2024)
22. Wang, X., Chen, T., Ge, Q., Xia, H., Bao, R., Zheng, R., Zhang, Q., Gui, T., Huang, X.: Orthogonal subspace learning for language model continual learning. In: *EMNLP (Findings)* (2023)
23. Wei, Y., Ma, J., Jiang, Z., Xiao, B.: Mixed color channels (mcc): A universal module for mixed sample data augmentation methods. In: *IEEE International Conference on Multimedia and Expo (ICME)*. pp. 1–6 (2022)
24. 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. Springer (2024)
25. 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)
26. Xu, Z., Liu, D., Yang, J., Raffel, C., Niethammer, M.: Robust and generalizable visual representation learning via random convolutions. In: *International Conference on Learning Representations* (2021)
27. Yun, S., Han, D., Oh, S.J., Chun, S., Choe, J., Yoo, Y.: Cutmix: Regularization strategy to train strong classifiers with localizable features. In: *International Conference on Computer Vision*. pp. 6023–6032 (2019)

28. Zhang, H., Cisse, M., Dauphin, Y.N., Lopez-Paz, D.: mixup: Beyond empirical risk minimization. In: International Conference on Learning Representations (2018)
29. Zhang, K., Liu, D.: Customized segment anything model for medical image segmentation. arXiv preprint arXiv:2304.13785 (2023)
30. 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)
31. Zhang, Z., Li, Y., Shin, B.S.: Generalizable polyp segmentation via randomized global illumination augmentation. *IEEE Journal of Biomedical and Health Informatics* (2024)
32. Zhou, K., Yang, Y., Qiao, Y., Xiang, T.: Domain generalization with mixstyle. In: International Conference on Learning Representations (2021)