Unleashing SAM for Few-Shot Medical Image Segmentation with Dual-Encoder and Automated Prompting

Cuong M. Pham[1][0009-0002-7359-034X], Phi Le Nguyen[1⋆][0000-0001-6547-7641], Thanh Trung Nguyen[2], Minh Hieu Phan[3][0000-0003-3861-0296], and Binh P. Nguyen[4⋆][0000-0001-6203-6664]

[1] Hanoi University of Science and Technology
lenp@soict.hust.edu.vn
[2] 108 Military Central Hospital
[3] Australian Institute for Machine Learning
[4] Victoria University of Wellington
binh.p.nguyen@vuw.ac.nz

Abstract. Deep learning has made significant progress in natural image segmentation but faces challenges in medical imaging due to the limited availability of annotated data. Few-shot learning offers a solution by enabling segmentation with only a few labeled samples, yet generalization remains a challenge when data is scarce. In this work, we investigate the potential of the Segment Anything Model (SAM), a foundation model trained on over one billion annotated images, for few-shot medical image segmentation. However, SAM faces two key challenges: (1) the domain gap between natural and medical images, leading to suboptimal performance, and (2) prompt dependency, as SAM requires user-defined prompts, limiting automation. To address these issues, we propose a novel framework, named AM-SAM, that adapts SAM for few-shot medical image segmentation. Our approach introduces a medical image-specific augmentation strategy and a dual-encoder architecture to bridge the domain gap. Additionally, we develop an automated dual-prompt mechanism to eliminate prompt dependency, generating point and mask prompts from support images. Extensive experiments show that AM-SAM outperforms existing approaches by up to 3.8% on ABD-MRI and 4.0% on ABD-30 in terms of dice score metric.

Keywords: Medical image segmentation · Segment Anything Model · Few-shot learning

1 Introduction

Medical image segmentation is a fundamental task in healthcare, playing a crucial role in diagnosis, treatment planning, and quantitative tissue analysis [26]. By precisely delineating anatomical structures and abnormalities, segmentation

[⋆] Phi Le Nguyen and Binh P. Nguyen are corresponding authors.

empowers clinicians to make accurate and informed medical decisions. However, unlike natural image segmentation, medical imaging presents unique challenges, primarily due to the scarcity of annotated data [2, 3]. Producing high-quality annotations requires expertise from trained medical professionals and is both time-consuming and labor-intensive. This limitation significantly hinders the effectiveness of supervised learning models, making it difficult to develop robust and generalizable segmentation algorithms.

To tackle this issue, few-shot learning has emerged as a promising approach, allowing models to segment new anatomical structures with only a small number of labeled samples [22, 27]. In a typical few-shot learning framework, a base model is pre-trained on a dataset containing known anatomies. During inference, a small support set of annotated novel anatomies is provided, and the model leverages this information to accurately segment corresponding structures in query images. The performance of few-shot segmentation models depends on two key factors: the ability of the base model to learn generalizable features and the effectiveness of utilizing support data for query image segmentation. Among existing techniques, two dominant approaches have gained traction: prototypical methods [21, 9] and meta-learning [24, 1]. Prototypical methods create representative prototypes from annotated support masks and use them to segment query images [25], while meta-learning focuses on training a model that can quickly adapt to unseen data [17]. Regardless of the approach, developing a highly generalizable base model is essential and achieving this requires training on a large-scale dataset.

The Segment Anything Model (SAM) [12] is a state-of-the-art foundation model for image segmentation, trained on a massive dataset of over one billion annotated images. Due to its strong generalization capabilities, an intriguing direction is to use SAM as the base model for few-shot medical image segmentation. However, integrating SAM into a few-shot framework presents two key challenges. First, SAM is pre-trained on natural images, yielding subpar performance on the medical imaging [28].

Although fine-tuning SAM for medical imaging tasks has been explored [10], these methods typically require large amounts of labeled data, which contradicts the constraints of few-shot learning. Second, SAM requires user-specified prompts for interactive inferencing. This prompt dependency limits its its use in a prompt-free segmentation scenario.

To overcome these challenges, we propose a novel framework that incorporates SAM for few-shot medical image segmentation while addressing both domain gap and prompt dependency issues. To mitigate the domain gap, we introduce a medical image-specific augmentation strategy that enhances the visibility of anatomical structures. Additionally, we develop a dual-encoder mechanism with two separate encoders: one extracting features from the original image and the other from the augmented image. This approach improves feature extraction and enhances segmentation performance. To resolve the prompt dependency issue, we introduce a dual-prompt mechanism that automatically generates two types of prompts—point prompts and mask prompts—directly from the sup-

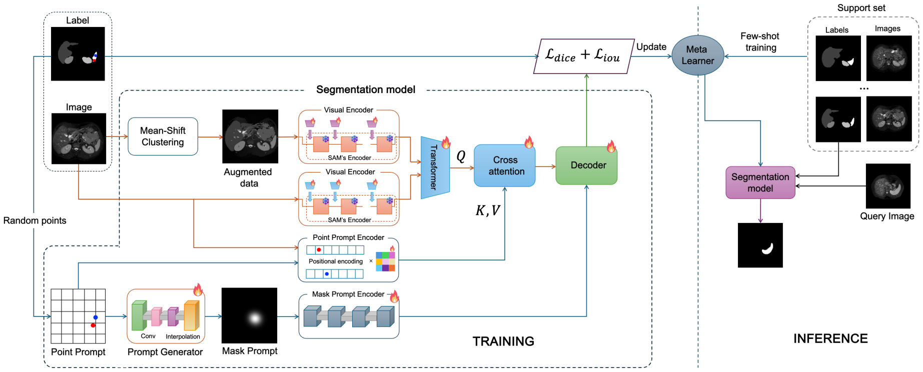

Fig. 1: Illustration of the training and inference pipeline of AM-SAM.

port samples. These prompts are then fused to guide the segmentation decoder, enabling prompt-free inference.

Our contributions can be summarized as follows:

- We introduce a novel approach, named AM-SAM, for leveraging SAM in few-shot medical image segmentation, focusing on enhancing the encoder's ability to extract useful information from query images while maximizing the effective use of the support data.
- We propose a data augmentation technique based on the Mean-Shift Clustering algorithm [6], which highlights anatomical boundaries and improves segmentation accuracy.
- We develop an automatic dual-prompt generation method that extracts both point prompt and mask prompt from a support image, effectively guiding the segmentation decoder.
- Extensive experiments demonstrate the superiority of our proposed method over existing approaches, significantly improving segmentation performance in few-shot medical imaging scenarios.

2 Proposed Method

2.1 Overview

Figure 1 illustrates the overall pipeline of AM-SAM, which consists of three main components: the Visual Encoder, the Prompt Encoder, and the Decoder. The Visual Encoder (Sec. 2.2) learns from both the original and augmented images, extracting essential features for segmentation. The Prompt Encoder (Sec.2.3) generates prompts from query images and learns their representation to guide the model in identifying the relevant regions for segmentation. Finally, the Decoder integrates the information from both the Visual and Prompt Encoders to generate the segmentation output. In our approach, we utilized a lightweight decoder [14], which allows for time-efficient execution.

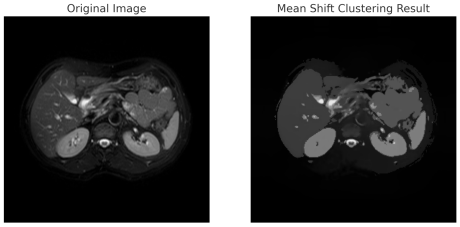

Fig. 2: Comparison between the original image (left) and the Mean Shift Clustering result (right). The clustering process groups pixels with similar intensities, effectively enhancing the segmentation of distinct anatomical regions in the medical image.

2.2 Visual Encoder

The Visual Encoder is crucial to the model's performance. While existing approaches fine-tune the pretrained SAM, we argue that in few-shot learning, fine-tuning alone is insufficient to bridge the gap between unseen medical images and training data. To address this, we enhance the encoder by incorporating augmented data that highlights anatomical structures, improving segmentation accuracy. Our Visual Encoder comprises two key components: image augmentation and a dual-encoder mechanism. We apply the Mean-Shift Clustering, an unsupervised algorithm that enhances contrast between anatomical structures and the background. This generates augmented images with clearer anatomical boundaries, allowing the encoder to extract more informative features for improved segmentation. Specifically, the algorithm takes the original image as input and iteratively shifts pixels toward regions of higher density using the following formula:

$$m(x) = \frac{\sum_{x_i \in S} K(x_i - x) x_i}{\sum_{x_i \in S} K(x_i - x)}; \quad K(x) = \exp\left(-\frac{\|x\|^2}{2h^2}\right), \tag{1}$$

where x represents the current pixel, x_i are neighboring pixels within the search window S, and $K(x)$ is a kernel function. The pixel positions are iteratively updated as $x_{t+1} = m(x_t)$, shifting each pixel toward the mean of its weighted neighbors until convergence. This process systematically clusters pixels into high-density regions by grouping those with similar characteristics. Consequently, at image boundaries where intensity variations are pronounced, data points become more concentrated, leading to improved delineation of anatomical structures. Figure 2 illustrates the input image alongside the results of the clustering process.

We adopt a dual-encoder architecture, where one encoder processes the original image while the other handles the augmented data. The extracted embeddings from both encoders are then fused through a Transformer-based fusion block to enhance feature representation. We leverage the pretrained Vision Transformer (ViT) from SAM's image encoder and introduce a lightweight Adapter at the beginning of each ViT block. Unlike existing methods, our approach benefits from augmented images, enabling the use of a more computationally efficient lightweight Adapter. Each adapter comprises two components, with four inner MLP layers per adapter and a shared outer layer across the adapters. The empirical results in Section 3 shows that our Adapter yields superior performance compared to existing approaches.

2.3 Automated Prompt Generation

We generate two distinct types of prompts from a support image: the point prompt and the mask-prompt. The point prompt is derived by randomly selecting a set of points within the annotated label region of the support image. On the other hand, the mask prompt is produced using a mask generator, which leverages convolutional layers in combination with bilinear interpolation to create a mask from the point prompt. The mask generator is trained during the model's training phase and remains freezing during inference. Both prompts are processed through separate encoders to learn their respective representations. The representation of the point prompt is fused with the visual embeddings, while that of the mask prompt is directly fed into the decoder to guide the segmentation process.

2.4 Training Objective

Our loss function is composed of two components: Dice Loss and IoU Loss. The Dice Loss minimizes the discrepancy between the predicted mask and the ground truth, while the IoU Loss maximizes the overlap between the predicted and ground truth regions. By combining these two losses, we effectively guide the model to generate accurate segmentation predictions. We employ a common meta-learning approach named MAML [7] to enable the model to adapt to the support samples during the inference phase.

3 Experiments and Evaluations

3.1 Experimental Settings

We conduct experiments on two datasets, ABD-MRI [11] and ABD-30 [13], with the following two settings:

- **Setting I:** allows for the presence of unseen anatomies in the training data (although they remain unlabeled).
- **Setting II:** ensures that all images in the training data do not contain the unseen anatomies.

We use the dice score as the primary metric to evaluation and compare the performance of our method with six popular medical images few-shot learning baselines, namely SE-Net [19], RP-Net [23], SSL-ALPNet [18], Q-Net [20], CAT-Net [15] and SSM-SAM [14]. For each task, we allocate 125 images: 5 images serve as support images for training the inner model, while the remaining 120 images are reserved for evaluating the outer model. During testing, we adopt a 1-way, 5-shot learning approach, using five images as the support set. To leverage the pre-trained weights of SAM, all images in PNG format are resized to 1024×1024 pixels. The general configuration for the parameters of SAM's image encoder is retained, ensuring that the model remains frozen during training. The resized 1024×1024 images are split into patches of size 16×16 and embedded into a latent vector space. For all experiments, we train the AM-SAM network for 50 epochs, with 3 inner training rounds per epoch. During inner training, we use the Stochastic Gradient Descent optimizer [4] with a learning rate of 0.0001. For optimizing the meta-learner, we utilize the AdamW optimizer [29] with a learning rate of 0.0002. Additionally, we apply Linear Learning Decay [8] with a decay rate of $1e^{-2}$, and Cosine Annealing [16] with a minimum learning rate of $1e^{-7}$, to dynamically adjust the learning rate during training. Our models are implemented in PyTorch, and the AM-SAM network is trained on an NVIDIA GeForce RTX 3090 with 24GB VRAM.

3.2 Comparison with Baseline Models

We compare the top-1 dice score achieved by AM-SAM with state-of-the-art (SOTA) methods in medical few-shot image segmentation (Table 1). As demonstrated, AM-SAM outperforms the other benchmarks in nearly all cases, achieving the highest results in four tasks. Specifically, our method shows improvements of approximately 3.8% in Setting I and 2.3% in Setting II on average for the ABD-MRI dataset, and 4.0% and 1.6% for the ABD-30 dataset, compared to the second-best performing method. Notably, our approach excels in tasks involving the left kidney, right kidney, and spleen. However, the results for the liver task are relatively less impressive.

3.3 Ablation Studies

We conducted a series of experiments to assess the effectiveness of two key components in our approach: the dual encoder and automated prompting. The results of these experiments are provided in Table 2.

Impacts of the Dual-encoder. We compare the model's performance with and without the use of the augmented image (denoted as "w/o MSC"). The findings reveal that integrating Mean Shift-based augmented images results in a dice score improvement of approximately 6.5% to 13.6% across four tasks in Setting I, with an average increase of 7.7% in Setting II on the ABD-MRI dataset. For the ABD-30 dataset, the improvements are 15.6% in Setting I and 11.9% in Setting II, on average across the four tasks.

Impacts of the Automated Prompts. To evaluate the impact of the automated mask prompts, we compared the performance of AM-SAM with a variant that does not use any prompt (denoted as "w/o MP"). The results demonstrate that incorporating our automated masks leads to notable improvements in three tasks, excluding the liver task. Specifically, there were improvements of around 3.4% to 3.6% in ABD-MRI and 5.5% to 6.5% in ABD-30 on average across the two settings. Additionally, we compare our proposed mask-prompt with the

Table 1: Comparison of AM-SAM with state-of-the-art methods based on the dice score. The highest-performing results are highlighted in **red**, while the second-best results are marked in blue. The values within (.) indicate the **relative** performance gap between AM-SAM and the best competing method.

Settings	**Methods**	**Tasks**				
		L Kidney ↑	R Kidney ↑	Liver ↑	Spleen ↑	Mean ↑
Setting I **ABD-MRI**	SE-Net	45.78	47.96	29.02	47.30	42.51
	ALPNet	70.17	77.05	72.45	67.71	71.85
	Q-Net	73.96	81.07	72.36	65.39	73.20
	CAT-Net	75.31	**83.23**	75.02	67.31	75.22
	SSM-SAM	74.13	81.22	76.16	74.97	76.62
	Ours	**78.28** (+3.9%)	81.46 (-2.1%)	**79.70** (+4.6%)	**78.80** (+5.1%)	**79.56** (+3.8%)
Setting II **ABD-MRI**	SE-Net	62.11	61.32	27.43	51.80	50.66
	RP-Net	79.30	84.66	71.51	75.69	77.79
	ALPNet	73.63	78.39	73.05	67.02	73.02
	Q-Net	74.05	77.52	78.71	67.43	74.43
	CAT-Net	74.01	78.90	**78.98**	68.83	75.18
	SSM-SAM	81.70	80.38	77.50	78.81	79.59
	Ours	**84.17** (+3.0%)	**84.95** (+0.3%)	76.12 (-3.6%)	**80.36** (+1.7%)	**81.40** (+2.3%)
Setting I **ABD-30**	SE-Net	24.42	12.51	35.42	43.66	29.00
	ALPNet	72.36	71.81	78.29	70.96	73.35
	Q-Net	68.55	63.47	71.12	68.95	68.02
	CAT-Net	72.34	69.91	75.90	73.39	72.89
	SSM-SAM	79.12	80.03	81.36	82.92	80.86
	Ours	**82.78** (+4.6%)	**84.02** (+5.0%)	**85.40** (+5.0%)	**83.97** (+1.3%)	**84.04** (+4.0%)
Setting II **ABD-30**	SE-Net	32.83	14.34	0.27	0.23	11.91
	RP-Net	70.48	70.00	79.62	69.85	72.48
	ALPNet	63.34	54.82	73.65	60.25	63.02
	Q-Net	63.26	58.37	74.36	63.36	64.83
	CAT-Net	63.36	60.05	75.31	67.65	66.59
	SSM-SAM	80.96	84.47	87.12	86.95	84.87
	Ours	**84.37** (+4.2%)	**86.01** (+1.8%)	**87.28** (+0.2%)	**87.11** (+0.2%)	**86.19** (+1.6%)

Gaussian support mask mechanism used in SSM-SAM. The results, presented in Table 4, demonstrate that our mask prompt consistently outperforms the Gaussian support mask across all experimental scenarios. Specifically, our approach achieves an average accuracy improvement ranging from 0.9% to 2.3%.

Impacts of the Adapter. We conducted experiments to evaluate the effectiveness of our proposed adapter mechanism by comparing it with the adapter used in Adapter-SAM [5]. The results, presented in Table 3, indicate that our adapter mechanism consistently outperforms Adapter-SAM across all scenarios, achieving a substantial performance gap, with an average relative improvement ranging from 1.3% to 2.3%.

In summary, the result indicates that both the dual encoder and the Mask-guided prompting mechanism make substantial contributions to the model's performance. Specifically, the dual encoder enhances the model's ability to refine spatial information and feature representation, while the mask-guide prompting effectively captures and leverages intricate details within the data. Together, these components synergistically improve the overall effectiveness and accuracy of the model in the experimental results.

4 Conclusion

In this study, we proposed a novel approach for leveraging SAM in few-shot medical image segmentation. Our method is built upon two key components:

Table 2: Comparison of AM-SAM with its variants—one without augmented data (w/o MSC) and one without automated prompts (w/o MP)—across different settings. The best results are highlighted in **red** and the second-best results in blue. The numbers in (.) indicate the **relative** performance gaps between AM-SAM and the second-best performing model.

Settings	**Methods**	**Tasks**				
		L Kidney ↑	R Kidney ↑	Liver ↑	Spleen ↑	Mean ↑
Setting I **ABD-MRI**	w/o MSC	71.99	76.52	72.13	69.36	72.50
	w/o MP	75.73	79.68	76.93	75.31	76.91
	AM-SAM	**78.28** (+3.9%)	**81.46** (+2.2%)	**79.70** (+3.6%)	**78.80** (+4.6%)	**79.56** (+3.4%)
Setting II **ABD-MRI**	w/o MSC	75.91	77.97	75.15	73.27	75.58
	w/o MP	79.40	80.64	**76.87**	77.49	78.59
	AM-SAM	**84.17** (+6.0%)	**84.95** (+5.3%)	76.12 (-1.0%)	**80.36** (+3.7%)	**81.40** (+3.6%)
Setting I **ABD-30**	w/o MSC	74.55	73.86	70.08	72.42	72.73
	w/o MP	78.94	79.22	79.01	78.36	78.88
	AM-SAM	**82.78** (+4.9%)	**84.02** (+6.1%)	**85.40** (+8.1%)	**83.97** (+7.2%)	**84.04** (+6.5%)
Setting II **ABD-30**	w/o MSC	76.83	75.58	78.20	77.46	77.02
	w/o MP	80.16	82.11	81.39	83.25	81.73
	AM-SAM	**84.37** (+5.3%)	**86.01** (+4.7%)	**87.28** (+7.2%)	**87.11** (+4.6%)	**86.19** (+5.5%)

Table 3: Comparison of our lightweight Adapter and Adapter-SAM's adapter.

Settings	**Methods**	**Tasks**				
		L Kidney ↑	R Kidney ↑	Liver ↑	Spleen ↑	Mean ↑
Setting I **ABD-MRI**	AS Adapter	77.12	80.36	79.42	75.09	78.00
	Ours	78.28	81.46	79.70	78.80	79.56 (+1.3%)
Setting II **ABD-MRI**	AS Adapter	83.33	81.46	74.34	80.21	79.84
	Ours	84.17	84.95	76.12	80.36	81.40 (+2.0%)
Setting I **ABD-30**	AS Adapter	80.18	82.86	82.13	83.60	82.19
	Ours	82.78	84.02	85.40	83.97	84.04 (+2.3%)
Setting II **ABD-30**	AS Adapter	83.12	84.24	86.33	85.17	84.72
	Ours	84.37	86.01	87.28	87.11	86.19 (+1.7%)

(*) AS Adapter means we replace our adapter with those of Adapter-SAM,
(**) Better results are underlined.

a dual-encoder framework, which enriches feature representation by incorporating both the original and augmented images; and an automated prompting mechanism, where point-prompt and mask-prompt are generated from support images to effectively utilize support sample information and guide the decoder for enhanced accuracy. Experimental results demonstrated that our approach significantly outperforms existing state-of-the-art methods, achieving superior segmentation accuracy across various tasks. Specifically, AM-SAM outperforms existing approaches by up to 3.8% on ABD-MRI and 4.0% on ABD-30 datasets.

Acknowledgments. This research is partially funded by Hanoi University of Science and Technology (HUST) under grant number T2024-TĐ-002. The work of BPN was supported in part by the Faculty Strategic Research Grant (FSRG No. 411494) at Victoria University of Wellington, the Endeavour Fund – Smart Ideas initiative from the New Zealand Ministry of Business, Innovation and Employment (MBIE) under contract RTVU2301, the MBIE Strategic Science Investment Fund for Data Science under contract RTVU1914, and the Explorer Grant 25/910/A from the Health Research Council of New Zealand.

Table 4: Comparison of our proposed masks an the mask used in SSM-SAM.

Settings	Methods	Tasks				
		L Kidney ↑	R Kidney ↑	Liver ↑	Spleen ↑	Mean ↑
Setting I	SSM-SAM	75.45	81.25	77.60	78.13	78.10
ABD-MRI	**Ours**	78.28	81.46	79.70	78.80	79.56(+0.9%)
Setting II	SSM-SAM	83.21	83.18	76.61	79.12	80.53
ABD-MRI	**Ours**	84.17	84.95	76.12	80.36	81.40(+1.1%)
Setting I	SSM-SAM	80.15	83.26	83.35	81.99	82.19
ABD-30	**Ours**	82.78	84.02	85.40	83.97	84.04(2.3%)
Setting II	SSM-SAM	82.49	84.88	86.32	84.79	84.62
ABD-30	**Ours**	84.37	86.01	87.28	87.11	86.19(+1.9%)

(*) Better results are underlined.

Disclosure of Interests. The authors declare no competing interests.

References

1. Aguiar, G.J., Mantovani, R.G., Mastelini, S.M., De Carvalho, A.C., Campos, G.F., Junior, S.B.: A meta-learning approach for selecting image segmentation algorithm. Pattern Recognition Letters **128**, 480–487 (2019)
2. Alzubaidi, L., Bai, J., Al-Sabaawi, A., Santamaría, J., Albahri, A.S., Al-Dabbagh, B.S.N., Fadhel, M.A., Manoufali, M., Zhang, J., Al-Timemy, A.H., et al.: A survey on deep learning tools dealing with data scarcity: definitions, challenges, solutions, tips, and applications. Journal of Big Data **10**(1), 46 (2023)
3. Bansal, M.A., Sharma, D.R., Kathuria, D.M.: A systematic review on data scarcity problem in deep learning: solution and applications. ACM Computing Surveys **54**(10s), 1–29 (2022)
4. Bottou, L.: Large-scale machine learning with stochastic gradient descent. In: Proceedings of COMPSTAT'2010: 19th International Conference on Computational Statistics. pp. 177–186. Springer (2010)
5. Chen, T., Zhu, L., Deng, C., Cao, R., Wang, Y., Zhang, S., Li, Z., Sun, L., Zang, Y., Mao, P.: Sam-adapter: Adapting segment anything in underperformed scenes. In: Proceedings of the IEEE/CVF International Conference on Computer Vision. pp. 3367–3375 (2023)
6. Cheng, Y.: Mean shift, mode seeking, and clustering. IEEE Transactions on Pattern Analysis and Machine Intelligence **17**(8), 790–799 (1995)
7. Finn, C., Abbeel, P., Levine, S.: Model-agnostic meta-learning for fast adaptation of deep networks. In: International conference on machine learning. pp. 1126–1135. PMLR (2017)
8. Glorot, X., Bengio, Y.: Understanding the difficulty of training deep feedforward neural networks. In: Proceedings of the thirteenth international conference on artificial intelligence and statistics. pp. 249–256. JMLR Workshop and Conference Proceedings (2010)
9. Houde, S., Hill, C.: What do prototypes prototype? In: Handbook of human-computer interaction, pp. 367–381. Elsevier (1997)
10. Hu, X., Xu, X., Shi, Y.: How to efficiently adapt large segmentation model (SAM) to medical images. arXiv preprint arXiv:2306.13731 (2023)

11. Kavur, A.E., Gezer, N.S., Barış, M., Aslan, S., Conze, P.H., Groza, V., Pham, D.D., Chatterjee, S., Ernst, P., Özkan, S., et al.: CHAOS challenge-combined (CT-MR) healthy abdominal organ segmentation. Medical Image Analysis **69**, 101950 (2021)
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: Proceedings of the IEEE/CVF international conference on computer vision. pp. 4015–4026 (2023)
13. Landman, B., Xu, Z., Igelsias, J., Styner, M., Langerak, T., Klein, A.: Multi-atlas labeling beyond the cranial vault–workshop and challenge. In: Proc. MICCAI Multi-atlas Labeling Beyond Cranial Vault–Workshop Challenge. vol. 5, p. 12. Munich, Germany (2015)
14. Leng, T., Zhang, Y., Han, K., Xie, X.: Self-sampling meta sam: enhancing few-shot medical image segmentation with meta-learning. In: Proceedings of the IEEE/CVF winter conference on applications of computer vision. pp. 7925–7935 (2024)
15. Lin, Y., Chen, Y., Cheng, K.T., Chen, H.: Few shot medical image segmentation with cross attention transformer. In: International Conference on Medical Image Computing and Computer-Assisted Intervention. pp. 233–243. Springer (2023)
16. Loshchilov, I., Hutter, F.: Sgdr: Stochastic gradient descent with warm restarts. arXiv preprint arXiv:1608.03983 (2016)
17. Luo, S., Li, Y., Gao, P., Wang, Y., Serikawa, S.: Meta-seg: A survey of meta-learning for image segmentation. Pattern Recognition **126**, 108586 (2022)
18. Ouyang, C., Biffi, C., Chen, C., Kart, T., Qiu, H., Rueckert, D.: Self-supervision with superpixels: Training few-shot medical image segmentation without annotation. In: Computer Vision–ECCV 2020: 16th European Conference, Glasgow, UK, August 23–28, 2020, Proceedings, Part XXIX 16. pp. 762–780. Springer (2020)
19. Roy, A.G., Siddiqui, S., Pölsterl, S., Navab, N., Wachinger, C.: 'squeeze & excite'guided few-shot segmentation of volumetric images. Medical image analysis **59**, 101587 (2020)
20. Shen, Q., Li, Y., Jin, J., Liu, B.: Q-net: Query-informed few-shot medical image segmentation. In: Proceedings of SAI Intelligent Systems Conference. pp. 610–628. Springer (2023)
21. Snell, J., Swersky, K., Zemel, R.: Prototypical networks for few-shot learning. Advances in Neural Information Processing Systems **30** (2017)
22. Sung, F., Yang, Y., Zhang, L., Xiang, T., Torr, P.H., Hospedales, T.M.: Learning to compare: Relation network for few-shot learning. In: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. pp. 1199–1208 (2018)
23. Tang, H., Liu, X., Sun, S., Yan, X., Xie, X.: Recurrent mask refinement for few-shot medical image segmentation. In: Proceedings of the IEEE/CVF international conference on computer vision. pp. 3918–3928 (2021)
24. Vanschoren, J.: Meta-learning. Automated machine learning: methods, systems, challenges pp. 35–61 (2019)
25. Wang, K., Liew, J.H., Zou, Y., Zhou, D., Feng, J.: PANet: Few-shot image semantic segmentation with prototype alignment. In: Proceedings of the IEEE/CVF International Conference on Computer Vision. pp. 9197–9206 (2019)
26. Wang, R., Lei, T., Cui, R., Zhang, B., Meng, H., Nandi, A.K.: Medical image segmentation using deep learning: A survey. IET Image Processing **16**(5), 1243–1267 (2022)
27. Wang, Y., Yao, Q., Kwok, J.T., Ni, L.M.: Generalizing from a few examples: A survey on few-shot learning. ACM Computing Surveys **53**(3), 1–34 (2020)

28. Zhang, L., Deng, X., Lu, Y.: Segment anything model (SAM) for medical image segmentation: A preliminary review. In: 2023 IEEE International Conference on Bioinformatics and Biomedicine (BIBM). pp. 4187–4194. IEEE (2023)
29. Zhou, P., Xie, X., Lin, Z., Yan, S.: Towards understanding convergence and generalization of AdamW. IEEE Transactions on Pattern Analysis and Machine Intelligence (2024)