This MICCAI paper is the Open Access version, provided by the MICCAI Society. It is identical to the accepted version, except for the format and this watermark; the final published version is available on SpringerLink.

Iterative Foundation-Dedicated Learning: Optimized Key Frames, Prompts and Memories for Semi-Supervised Segmentation

Ziman Yin[1], Dong Nie[2], Shuo Li[3], Junjun Pan[1], and Zhenyu Tang[1](✉)

[1] Beihang University, Beijing 100191, China
[2] Tower Cloud Labs
[3] Case Western Reserve University, Cleveland, OH 44106, USA
tangzhenyu@buaa.edu.cn

Abstract. Semi-supervised learning (SSL) can effectively reduce the labor-intensive labeling required for deep learning based medical image segmentation. The emergence of visual foundation models show zero-shot capability, offering a new way of SSL. In this paper, a novel SSL framework that combines foundation and dedicated models is proposed. Unlike most existing SSL methods, where the foundation model is manually prompted to generate pseudo-labels from unlabeled images for training the dedicated model in a one-way strategy without further refinement. In our framework, foundation (SAM2) and dedicated (UNet) models are in an iterative pipeline. Specifically, in each iteration, prompts from coarse segmentation results using UNet are calculated for SAM2 to generate pseudo-labels which are used to further train the UNet for better prompts in next iteration. In this way, the pseudo-labels and UNet can be mutually improved until convergence. To enhance the performance of SAM2 in medical image segmentation, a new uncertainty-aware module using historical cues is presented to optimize key frames selection and prompts generation for SAM2. Furthermore, a new semantic-aware memory bank is introduced, where memories in the memory bank of SAM2 are divided into semantic groups. In this way, anatomical prior knowledge can be leveraged by SAM2. In the experiment, our framework is evaluated using public and in-house datasets in the context of multi-label segmentation, and the experimental results demonstrate that our framework outperforms state-of-the-art SSL methods in both datasets.

Keywords: SSL segmentation· SAM2 · prompt generation · semantic memories.

1 Introduction

Deep learning (DL) based medical image segmentation has been widely studied [3,6,10]. To achieve accurate segmentation, large scale of training data is needed. However, manual annotation requires experienced experts and is labor-intensive. To address this issue, semi-supervise learning (SSL) becomes a hot topic in

medical image segmentation [15,8,2]. The basic idea of SSL is to generate pseudo-labels from unlabeled data using a dedicated segmentation model pre-trained by a small set of labeled data. The pseudo-labels can then be used to further train the dedicated segmentation model. However, since the low capacity of dedicated segmentation models, pseudo-labels are often of poor quality.

Recently, the emergence of visual foundation models, e.g., SAM [9], has demonstrated exceptional capacity of general and zero-shot segmentation, offering a new way of SSL. Most existing foundation model based SSL methods simply use the foundation model to generate pseudo-labels from unlabeled data. Since medical image segmentation requires strong domain knowledge, high-quality prompts for foundation models are needed and usually delineated manually [4,11,12]. Some methods try to automate the prompts generation [5], and the basic idea is to use pre-trained dedicated segmentation models to get coarse segmentation results which are used as prompts. However, most of them are in a one-way strategy, making the resulting prompts of low quality and cannot get further refinement. Moreover, SAM is usually adopted as the foundation model which are unsuitable to handle 3D multi-label medical image.

Based on the above observations, a novel SSL segmentation framework is proposed in this paper, where foundation and dedicated segmentation models are combined in an iterative pipeline. Specifically, SAM2 is used as the foundation model for 3D medical image segmentation whose slices are spatially continuous and analogous to videos. In each iteration, prompts are generated from segmentation results of unlabeled data using the dedicated segmentation model pre-trained by a small set of labeled data. Based on the prompts, pseudo-labels of unlabeled data can be obtained by SAM2, which are used to further train the dedicated segmentation model for good prompts in next iteration. Therefore, as the iteration goes on, the pseudo-labels and dedicated segmentation model can be mutually improved. To enhance the performance of SAM2 in medical image segmentation, a new uncertainty-aware module (UAM) using historical cues is presented in the iterative pipeline to ensure stable and accurate prompts. Moreover, based on the UAM, key frames in SAM2, which follows a fixed selection rule (e.g., every three frames), can be adaptively determined. Additionally, a new semantic-aware memory bank (SA-MB) is introduced, where memories in the memory bank of SAM2 are divided into different semantic groups (e.g., bones and cartilages). In this way, anatomical prior knowledge can be leveraged by SAM2 to achieve accurate multi-label medical image segmentation. Public and in-house datasets containing MR images of knee joint are used to evaluate our framework. Experimental results demonstrate that our framework outperforms the state-of-the-art (SOTA) methods in both datasets. Our main contributions are listed below:

(1) The novel iterative pipeline of foundation and dedicated models is proposed for accurate 3D multi-label medical image segmentation.
(2) The new uncertainty-aware module is presented to optimize key frames selection and prompts generation for SAM2.

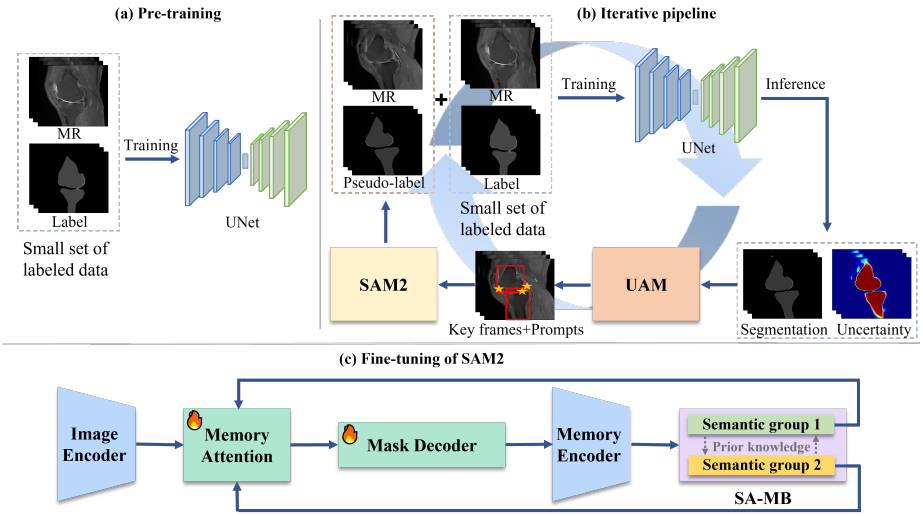

Fig. 1. Overview of our framework. (a) a dedicated segmentation model (UNet) is first pre-trained with a small set of labeled data; (b) in the iterative pipeline, segmentation results with pixel-wise uncertainty are produced from unlabeled data by the UNet, based on which key frames and prompts can be determined by the uncertainty-aware module (UAM) and used by SAM2 to get pseudo-labels for further training the UNet. This process is iteratively proceeded until convergence; (c) SAM2 is fine-tuned with the semantic-aware memory bank (SA-MB) where memories are divided into different semantic groups. In this way, anatomical prior knowledge can be leveraged by SAM2.

(3) The new semantic-aware memory bank in SAM2 is introduced, based on which anatomical prior knowledge can be leveraged.
(4) Our framework achieves the best performance of all SOTA methods under evaluation on both public and in-house datasets.

2 Method

The overview of our framework is shown in Fig. 1. Specifically, UNet is adopted as the dedicated segmentation model, which can be replaced by any existing segmentation models, and the foundation model is SAM2. At the beginning, the UNet is pre-trained using a small set of labeled data (see Fig. 1a). Then in the iterative pipeline, segmentation results with pixel-wise uncertainty of unlabeled data are produced by the pre-trained UNet, based on which key frames and prompts can be determined by the uncertainty-aware module (UAM) and used by SAM2 to get pseudo-labels of unlabeled data. The pseudo-labels are used to extend the training data to improve the UNet for better segmentation results of unlabeled data as well as better prompts for SAM2 in next iteration (see Fig.

1b). The iteration goes until the pseudo-labels and UNet reach convergence. Moreover, a fine-tuning of SAM2 with the proposed semantic-aware memory bank (SA-MB) is performed (see Fig. 1c). In the SA-MB, memories are divided into different semantic groups, in this way, anatomical prior knowledge (e.g., spatial relations of anatomical structures) can be leveraged by SAM2, and multi-label segmentation can also be achieved.

2.1 The Dedicated Segmentation Model

The UNet is adopted as the dedicated segmentation model in our framework. In the initial stage, it is pre-trained with a small set of labeled data and used to produce coarse segmentation results of unlabeled data. Moreover, the Monte Carlo dropout (mc-dropout) [13] is applied to the UNet to produce pixel-wise uncertainty for the coarse segmentation results. Specifically, in the training and inference phases of the UNet, neurons are randomly deactivate with a certain probability. In this way, different segmentation results can be produced. Assume that $I \in \mathbb{R}^{H\times W\times S}$ is the input data (H, W, and S stand for the height, width and number of slices) and $Y_1, ..., Y_N$ are the segmentation results using the mc-dropout, then the corresponding uncertainty map $U \in \mathbb{R}^{H\times W\times S}$ is defined as:

$$U = \frac{1}{N}\sum_{n=1}^{N}(Y_n - \mu)^2,\ \mu = \frac{1}{N}\sum_{n=1}^{N}Y_n, \tag{1}$$

where $\mu \in \mathbb{R}^{H\times W\times S}$ is the average of all segmentation results. Therefore, big values in U indicate high uncertainty at corresponding pixels.

2.2 The Uncertainty-aware Module for Key Frames and Prompts Optimization

In most applications of SAM2, prompts are usually generated manually. Moreover, key frames are also selected in a fixed rule, e.g., every three frames. To optimize and automate key frames selection and prompts generation, an uncertainty-aware module (UAM) is presented in our framework. Specifically, based on the uncertainty map U of each segmentation result produced by the dedicated segmentation model, the corresponding confidence map $C = \mathbf{1} - U$ is calculated, where $\mathbf{1} \in \mathbb{R}^{H\times W\times S}$. The slice I^s with the highest sum of confidence of every M slices in I is selected as the key frame, which is defined as:

$$I^s = \arg\max_{s}\sum_{h=1}^{H}\sum_{w=1}^{W}C(h, w, s), \tag{2}$$

where s stands for each one of the M slices, and M is set to 3 in our framework.

After the selection of key frames, prompts for each key frame are generated in the UAM using historical cues (see Fig. 2). Specifically, assume that C_T is the confidence map of I at iteration T, an accumulated confidence map $\overline{C}_T$ is

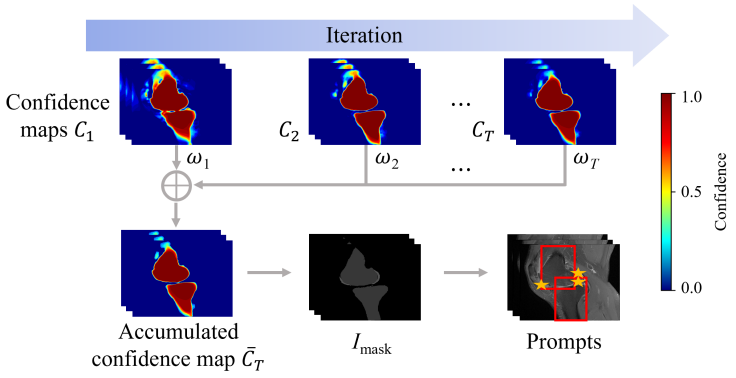

Fig. 2. Prompts generation in the UAM. According to the confidence maps $C_1, ..., C_T$ in current and previous iterations, the accumulated uncertainty map at current iteration $\overline{C}_T$ can be obtained, from which mask and prompts can be generated.

calculated by weighted averaging of historical confidence maps from previous iterations, which is defined as:

$$\overline{C}_T = \sum_{t=1}^{T} \omega_t C_t, \;\; w_t = \frac{(1-\alpha)\alpha^{T-t}}{1-\alpha^T} \tag{3}$$

where ω_t is weighting factor, α is a constant value controlling the weight changing rate and is set to 0.5. Clearly, the closer to the current iteration, the greater the weight assigned. After getting the accumulated confidence map $\overline{C}_T$, thresholding with 0.5 is applied to $\overline{C}_T$ to generate a mask I_{mask} defined as:

$$I_{\text{mask}}(x,y) = \begin{cases} 1 & \overline{C}_T(x,y) \geq 0.5 \\ 0 & \text{otherwise} \end{cases} \tag{4}$$

Based on the resulting I_{mask}, prompts can be calculated for every connective areas and used by SAM2. It is clear that the accumulated confidence map can relieve the influence of erroneous segmentation in previous iterations, leading to stable and accurate prompts.

2.3 The Foundation Model with Semantic-aware Memory Bank

Our current study is conducted in the context of multi-label segmentation of knee joint, for each semantic label (i.e., bones or cartilages), a corresponding mask $I_{\text{mask},l}$ is available after the UAM, where l indicates the semantic label. Connected components are derived from $I_{\text{mask},l}$, based on which prompts for l can be generated. It is worth noting that for bones (e.g., femur and tibia), bounding boxes are chosen as the prompts, while for cartilages, which are of small and irregular regions, foreground and background points are used as prompts.

In medical image segmentation, anatomical prior knowledge is important and can be adopted to enhance the segmentation performance. For example, in the knee joint segmentation, cartilages are attached to the surface of bones. However, in original SAM2, such prior knowledge cannot be leveraged. To address this issue, a semantic-aware memory bank (SA-MB) is presented. Specifically, memories in the memory bank of SAM2 are divided into different semantic groups according to the segmentation results, e.g., groups for bones and cartilages in our study. Assume that $I_{\text{seg},l}$ is segmentation results of $l \in \{\text{bone}, \text{cart}\}$ for the input image I after the mask decoder of SAM2. In the inference propagating phase, memories in SA-MB are updated following the same FIFO strategy as the original memory bank in SAM2. According to $I_{\text{seg},l}$, memories in SA-MB are divided into different semantic groups, i.e., F_{bone} and F_{cart}, defined as:

$$\begin{aligned} F_{\text{bone}} =& \text{CxBlock}(\text{Conv}(I_{\text{seg,bone}}) + \text{Conv}(F_I)) \\ F_{\text{cart}} =& \text{CxBlock}(\text{Conv}(\hat{I}_{\text{seg,bone}} \cap \hat{I}_{\text{seg,cart}} \cup I_{\text{seg,cart}}) + \text{Conv}(F_I)), \end{aligned} \tag{5}$$

where $\hat{I}_{\text{seg},l}$ is the dilatation of $\hat{I}_{\text{seg},l}$, and $\hat{I}_{\text{seg,bone}} \cap \hat{I}_{\text{seg,cart}}$ is the intersection region of bones and cartilages. It is known that cartilages are thin tissues attached to the surface of bones, the resulting intersection region can be used to enhance the quality of F_{cart} as well as the subsequent memory attention process, finally improving the segmentation accuracy. Moreover, based on the SA-MB, multi-label image segmentation can be achieved.

Cross entropy and Dice losses are adopted in the training of UNet and the fine-tuning of SAM2 with SA-MB. In each iteration, the UNet is trained for 100 epochs (batch size 4) using Adam optimizer with learning rate of 1×10^{-4}. SAM2 with SA-MB is fine tuned for 40 epochs at initial stage using AdamW optimizer with learning rate of 5×10^{-5}, and the training data are the same as used in the pre-training of UNet. After the training phase, the dedicated segmentation model in the iterative pipeline is used as the segmentation model for testing.

3 Experiments

In the experiment, public (SKI10 [7]) and in-house datasets containing 3D MR images of the knee joint from 100 and 136 patients are used to evaluate our framework, respectively. In both datasets, label images of bones and cartilages are available for each patient (see Fig. 3).

Besides our framework, UNet and five state-of-the-art (SOTA) SSL methods, including UCMT [14], SCP-Net [16], SS-Net [15], BCP [1], and SemiMedSAM [5], are evaluated. 3D MR images of 80% patients are randomly selected from the public and in-house datasets as the training data, respectively, and the rest 20% patients are for testing. Moreover, in the training data, 10 patients are randomly selected to form a small set of labeled data for the public and in-house datasets, respectively. Dice score, Hausdorff distance (HD), and sensitivity (Sen), are adopted to evaluate the segmentation results.

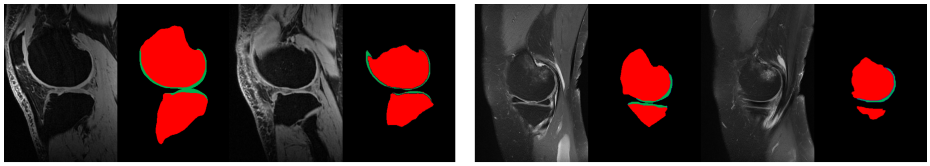

Fig. 3. Examples of MR and label images in the public (left) and in-house (right) datasets. Each label image contains labels of bones (red) and cartilages (green).

3.1 Evaluation of Image Segmentation

Table 1 shows the evaluation results. Except for the sensitivity of bones on the two datasets and HD of bones on the in-house dataset, our framework achieves the best performance of all methods. Since the small set of labeled data (10 patients), the UNet show inferior performance compared to most of the SSL methods. For both datasets, our framework is convergent after four iterations.

Table 1. Evaluation of segmentation results using each method in testing phase.

Methods		SKI10			In-house		
		Dice	HD	Sen	Dice	HD	Sen
UNet	bone	0.841±0.300	10.656±14.122	0.841±0.303	0.832±0.169	27.061±19.636	0.760±0.176
	cart	0.559±0.191	20.851±15.993	0.638±0.229	0.431±0.238	40.767±35.355	0.412±0.259
UCMT	bone	0.843±0.260	14.427±16.070	**0.951±0.167**	0.898±0.144	**9.565±7.828**	**0.960±0.108**
	cart	0.666±0.215	19.536±27.357	0.613±0.221	0.218±0.162	76.407±18.334	0.109±0.087
SCP-Net	bone	0.844±0.296	17.913±28.712	0.856±0.302	0.869±0.186	21.920±27.912	0.846±0.204
	cart	0.628±0.196	16.529±18.220	0.691±0.232	0.405±0.222	48.516±31.881	0.366±0.226
SS-Net	bone	0.919±0.168	24.559±35.172	0.929±0.178	0.869±0.167	27.970±32.003	0.869±0.185
	cart	0.642±0.189	14.460±16.860	0.624±0.203	0.560±0.168	42.974±33.117	**0.584±0.189**
BCP	bone	0.837±0.299	17.636±25.578	0.834±0.305	0.846±0.160	27.763±24.912	0.806±0.171
	cart	0.627±0.203	16.114±17.240	0.648±0.238	0.395±0.211	42.406±33.622	0.307±0.188
Semi-MedSAM	bone	0.833±0.301	11.993±13.608	0.817±0.301	0.834±0.170	26.651±18.980	0.765±0.175
	cart	0.558±0.189	19.965±15.261	0.654±0.221	0.432±0.242	40.303±37.072	0.415±0.262
Ours	bone	**0.949±0.070**	**9.962±12.414**	0.936±0.089	**0.908±0.160**	12.019±13.167	0.885±0.179
	cart	**0.670±0.178**	**12.194±14.437**	**0.729±0.190**	**0.569±0.214**	**21.791±17.443**	0.538±0.233

Examples of segmentation results using all methods under evaluation are shown in Fig. 4, and the segmentation results of our framework are more consistent with the ground truth (GT) than the other methods.

3.2 Ablation Study

In the ablation study, our framework without the UAM and SA-MB is evaluated, respectively. For our framework without the UAM, confidence maps at current iteration are used to generate prompts, and the key frames selection follows the original rule in SAM2, e.g., every three slices. Table 2 shows the evaluation results which demonstrate that both UAM and SA-MB play positive roles in

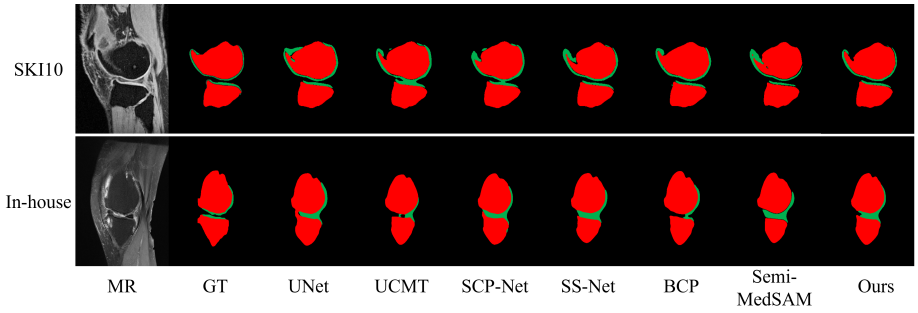

Fig. 4. Examples of segmentation results using all methods under evaluation on SKI10 (top) and in-house (bottom) datasets.

the segmentation. Moreover, to show the effect of iterative pipeline, Dice scores of pseudo-labels of unlabeled data after each iteration are calculated, and it is clear that the quality of pseudo-labels are iteratively improved.

Table 2. Dice scores of segmented unlabeled images after the UNet and SAM2 in each iteration in training phase and segmentation results in testing phase.

Dataset	UAM	SA-MB	Model	label	Iter-1	Iter-2	Iter-3	Iter-4	Testing
SKI10			UNet	bone	0.841±0.300	0.851±0.300	0.857±0.296	0.857±0.292	0.852±0.294
				cart	0.560±0.191	0.427±0.221	0.410±0.231	0.406±0.242	0.560±0.189
			SAM2	bone	0.850±0.292	0.855±0.293	0.858±0.288	0.857±0.288	-
				cart	0.444±0.216	0.407±0.178	0.396±0.241	0.393±0.238	
	✓		UNet	bone	0.841±0.300	0.855±0.296	0.859±0.290	0.860±0.289	0.867±0.291
				cart	0.560±0.191	0.444±0.216	0.427±0.232	0.403±0.229	0.599±0.188
			SAM2	bone	0.859±0.294	0.857±0.288	0.857±0.293	0.860±0.284	-
				cart	0.506±0.215	0.434±0.226	0.414±0.190	0.397±0.180	
	✓	✓	UNet	bone	0.841±0.300	0.856±0.295	0.863±0.291	0.860±0.296	0.949±0.070
				cart	0.560±0.191	0.600±0.188	0.614±0.190	0.627±0.182	0.670±0.178
			SAM2	bone	0.873±0.270	0.860±0.294	0.861±0.293	0.865±0.291	-
				cart	0.607±0.176	0.604±0.184	0.624±0.177	0.631±0.183	
In-house			UNet	bone	0.843±0.156	0.869±0.140	0.876±0.132	0.884±0.129	0.854±0.301
				cart	0.345±0.230	0.367±0.233	0.386±0.231	0.348±0.238	0.361±0.282
			SAM2	bone	0.845±0.218	0.871±0.222	0.833±0.175	0.871±0.097	-
				cart	0.287±0.227	0.335±0.215	0.335±0.214	0.303±0.224	
	✓		UNet	bone	0.843±0.156	0.874±0.134	0.878±0.125	0.877±0.128	0.857±0.301
				cart	0.345±0.230	0.389±0.232	0.391±0.215	0.411±0.231	0.474±0.208
			SAM2	bone	0.858±0.146	0.881±0.142	0.888±0.122	0.888±0.129	-
				cart	0.303±0.249	0.385±0.257	0.397±0.254	0.398±0.251	
	✓	✓	UNet	bone	0.843±0.156	0.881±0.142	0.909±0.122	0.916±0.120	0.908±0.160
				cart	0.345±0.230	0.384±0.232	0.429±0.247	0.450±0.244	0.569±0.214
			SAM2	bone	0.869±0.200	0.887±0.122	0.916±0.107	0.916±0.106	-
				cart	0.362±0.235	0.393±0.228	0.432±0.247	0.453±0.245	

4 Conclusion

A novel foundation-dedicated model based semi-supervised learning (SSL) framework is proposed for medical image segmentation. An iterative pipeline was adopted in our framework, where prompts, pseudo-labels and dedicated model can be iteratively and mutually enhanced. Moreover, a new uncertainty-aware module (UAM) was presented to optimize key frames selection and prompts generation for SAM2. A new semantic-aware memory bank (SA-MB) was proposed, where memories are divided into different semantic groups, in this way, anatomical prior knowledge can be leveraged by SAM2. Experimental results on public and in-house datasets demonstrated that our framework outperformed the state-of-the-art SSL methods, and the UAM and SA-MB are effective.

Acknowledgments. This work is supported by National Key Research and Development Program of China (No. 2023YFC2411001, 2023YFC2411002).

Disclosure of Interests. The authors have no competing interests.

References

1. Bai, Y., Chen, D., Li, Q., Shen, W., Wang, Y.: Bidirectional copy-paste for semi-supervised medical image segmentation. In: Proceedings of the IEEE/CVF conference on computer vision and pattern recognition. pp. 11514–11524 (2023)
2. Cai, H., Qi, L., Yu, Q., Shi, Y., Gao, Y.: 3d medical image segmentation with sparse annotation via cross-teaching between 3d and 2d networks. In: International Conference on Medical Image Computing and Computer-Assisted Intervention. pp. 614–624. Springer (2023)
3. Cao, H., Wang, Y., Chen, J., Jiang, D., Zhang, X., Tian, Q., Wang, M.: Swin-unet: Unet-like pure transformer for medical image segmentation. In: European conference on computer vision. pp. 205–218. Springer (2022)
4. Cheng, D., Qin, Z., Jiang, Z., Zhang, S., Lao, Q., Li, K.: Sam on medical images: A comprehensive study on three prompt modes. arXiv preprint arXiv:2305.00035 (2023)
5. Deshpande, T., Prakash, E., Ross, E.G., Langlotz, C., Ng, A., Valanarasu, J.M.J.: Auto-generating weak labels for real & synthetic data to improve label-scarce medical image segmentation. In: Medical Imaging with Deep Learning (2024)
6. He, X., Zhou, Y., Zhao, J., Zhang, D., Yao, R., Xue, Y.: Swin transformer embedding unet for remote sensing image semantic segmentation. IEEE Transactions on Geoscience and Remote Sensing **60**, 1–15 (2022)
7. Heimann, T., Morrison, B.J., Styner, M.A., Niethammer, M., Warfield, S.: Segmentation of knee images: a grand challenge. In: Proc. MICCAI Workshop on Medical Image Analysis for the Clinic. vol. 1. Beijing, China (2010)
8. Hu, X., Zeng, D., Xu, X., Shi, Y.: Semi-supervised contrastive learning for label-efficient medical image segmentation. In: Medical Image Computing and Computer Assisted Intervention–MICCAI 2021: 24th International Conference, Strasbourg, France, September 27–October 1, 2021, Proceedings, Part II 24. pp. 481–490. Springer (2021)

9. 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)
10. Liu, H., Ni, Z., Nie, D., Shen, D., Wang, J., Tang, Z.: Multimodal brain tumor segmentation boosted by monomodal normal brain images. IEEE Transactions on Image Processing **33**, 1199–1210 (2024). https://doi.org/10.1109/TIP.2024.3359815
11. Putz, F., Grigo, J., Weissmann, T., Schubert, P., Hoefler, D., Gomaa, A., Tkhayat, H.B., Hagag, A., Lettmaier, S., Frey, B., et al.: The segment anything foundation model achieves favorable brain tumor autosegmentation accuracy on mri to support radiotherapy treatment planning. arXiv preprint arXiv:2304.07875 (2023)
12. Ranem, A., Babendererde, N., Fuchs, M., Mukhopadhyay, A.: Exploring sam ablations for enhancing medical segmentation in radiology and pathology. arXiv preprint arXiv:2310.00504 (2023)
13. Seoh, R.: Qualitative analysis of monte carlo dropout. arXiv preprint arXiv:2007.01720 (2020)
14. Shen, Z., Cao, P., Yang, H., Liu, X., Yang, J., Zaiane, O.R.: Co-training with high-confidence pseudo labels for semi-supervised medical image segmentation. arXiv preprint arXiv:2301.04465 (2023)
15. Wu, Y., Wu, Z., Wu, Q., Ge, Z., Cai, J.: Exploring smoothness and class-separation for semi-supervised medical image segmentation. In: International conference on medical image computing and computer-assisted intervention. pp. 34–43. Springer (2022)
16. Zhang, Z., Ran, R., Tian, C., Zhou, H., Li, X., Yang, F., Jiao, Z.: Self-aware and cross-sample prototypical learning for semi-supervised medical image segmentation. In: International Conference on Medical Image Computing and Computer-Assisted Intervention. pp. 192–201. Springer (2023)