Cross Prompting Consistency with Segment Anything Model for Semi-supervised Medical Image Segmentation

Juzheng Miao[1], Cheng Chen[2(✉)], Keli Zhang[3], Jie Chuai[3], Quanzheng Li[2,4], and Pheng-Ann Heng[1,5]

[1] Department of Computer Science and Engineering, The Chinese University of Hong Kong, Hong Kong, China
[2] Center for Advanced Medical Computing and Analysis, Massachusetts General Hospital and Harvard Medical School, Boston, MA, USA
cchen101@mgh.harvard.edu
[3] Huawei Noah's Ark Lab, Shenzhen, China
[4] Data Science Office, Massachusetts General Brigham, Boston, MA, USA
[5] Institute of Medical Intelligence and XR, The Chinese University of Hong Kong, Hong Kong, China

Abstract. Semi-supervised learning (SSL) has achieved notable progress in medical image segmentation. To achieve effective SSL, a model needs to be able to efficiently learn from limited labeled data and effectively exploit knowledge from abundant unlabeled data. Recent developments in visual foundation models, such as the Segment Anything Model (SAM), have demonstrated remarkable adaptability with improved sample efficiency. To harness the power of foundation models for application in SSL, we propose a cross prompting consistency method with segment anything model (CPC-SAM) for semi-supervised medical image segmentation. Our method employs SAM's unique prompt design and innovates a cross-prompting strategy within a dual-branch framework to automatically generate prompts and supervisions across two decoder branches, enabling effectively learning from both scarce labeled and valuable unlabeled data. We further design a novel prompt consistency regularization, to reduce the prompt position sensitivity and to enhance the output invariance under different prompts. We validate our method on two medical image segmentation tasks. The extensive experiments with different labeled-data ratios and modalities demonstrate the superiority of our proposed method over the state-of-the-art SSL methods, with more than 9% Dice improvement on the breast cancer segmentation task. Code is available at: https://github.com/JuzhengMiao/CPC-SAM.

Keywords: Semi-supervised Segmentation · Segment Anything Model · Prompt Consistency.

1 Introduction

Segmentation is an essential step for accurate disease diagnosis and treatment planning [4,21]. Although deep learning methods have obtained impressive re-

sults in various organ or lesion segmentation tasks [16], a large scale of labeled data is required, which are extremely expensive and time-consuming to collect. Given that unlabeled data is typically plentiful in practice, semi-supervised learning (SSL) emerges as a compelling approach by efficiently leveraging both the limited labeled data and the extensive amounts of unlabeled data [2,11,26,27,28].

We consider the keys to the success of SSL methods are two folds. First, the model must be capable of quickly learning sufficiently general discriminative information from a limited amount of labeled data. On the other hand, once it has acquired this discriminative information, the model should effectively leverage the unlabeled data for further optimization. Current SSL methods mainly focus on the latter aspect, devising strategies to more effectively utilize unlabeled data, such as utilizing the predictions as pseudo labels for supervision [2,11], and imposing a consistency regularization on the predictions of different models or branches [26,27,28]. However, the first key aspect of rapid learning from limited labeled data is often overlooked. To overcome this limitation, we draw our attention to the general segmentation foundation model, i.e., the segment anything model (SAM), which is pre-trained on a large-scale natural datasets and has the potential of transferring to a new task by using only limited labeled data with the impressive few-shot learning capabilities demonstrated in prior research [5].

Current methods adapting SAM to medical image segmentation tend to train SAM in a fully supervised way with plenty of labeled data [25,29]. Very recently, only a limited number of works attempt to adapt SAM in the SSL setting. For example, Samdsk [31] leverages SAM to produce pseudo labels and select reliable ones into the labeled set to train a traditional segmentation network, i.e., a convolutional neural network (CNN). Li et al. [19] generate prompts from the prediction of a CNN and then choose outputs with a high consistency between the CNN and SAM as pseudo labels. SemiSAM [30] produces prompts in a similar way by using the CNN trained in a Mean Teacher framework and uses SAM's output as an additional supervision signal. In these methods, SAM is simply leveraged as a static and standalone component to generate pseudo labels on medical images, which may not yield desired performance due to the significant domain gap between natural and medical images [10,13]. Chen et al. [7] include the fine-tuning of SAM into the loop of SSL and thus enhance the adaptation ability of SAM to medical images. However, this work only fine-tunes SAM with a small number of labeled data whereas information contained in the large number of unlabeled data is not fully explored.

In this paper, we aim to leverage the few-shot learning capabilities of the SAM model to bolster our SSL framework for rapid learning from a limited amount of labeled data. Building on this foundation, we then leverage the unique advantage of SAM's prompting mechanism [18], to develop effective strategies for learning from unlabeled data in SSL. We propose a semi-supervised medical image segmentation framework which is driven by **c**ross **p**rompting **c**onsistency with **s**egment **a**nything **m**odel (CPC-SAM). Our method innovates a SAM enabled cross-prompting strategy within a dual-branch framework, which uses the unprompted output from one branch to generate prompts for the other branch.

Then the prompted output from the second branch is employed to guide the training of the first branch. Such a cross prompting and supervision strategy enhances the learning process, effectively leveraging the unlabeled data. Nonetheless, without ground truth for the unlabeled inputs, the prompts generated from unprompted outputs can be inherently unreliable and noisy. The prompted output is thus probably to be unreliable as well due to SAM's high sensitivity to prompt positions [9,12]. To address this issue, we design a novel prompt consistency regularization strategy aimed at improving the consistency of outputs across varying prompts. This strategy reduces SAM's sensitivity to different prompts and enhances the invariance of the output, ensuring more reliable and stable results even when derived from less dependable prompts. Our method has been extensively evaluated on two public datasets for breast cancer segmentation and cardiac structure segmentation, showing superiority over existing methods, especially when the labeled data are extremely limited. Specifically, using only 10 labeled ultrasound images, our method obtains an improvement of over 9% Dice than various strong baselines on the breast cancer segmentation task.

2 Method

Fig. 1 gives an overview of our cross prompting consistency framework with SAM for SSL medical image segmentation, called CPC-SAM. Considering the few-shot learning capabilities of the SAM model, we directly fine-tune SAM in the SSL pipeline to achieve the rapid learning from a limited amount of labeled data. Building on this foundation, a cross prompting dual-branch framework is developed based on the promptable property of SAM to make full use of the large scale of unlabeled data. Moreover, considering the potential harmfulness of SAM's sensitivity to prompts' positions for SSL, we further propose the prompt consistency regularization to enhance output invariance under various prompts.

2.1 Problem Formulation and Architecture

SSL segmentation aims to obtain a satisfactory performance using a small number of labeled data $\mathcal{L} = \{(\boldsymbol{x}_i, \boldsymbol{y}_i)\}_{i=1}^{N}$ and a large scale of unlabeled data $\mathcal{U} = \{(\boldsymbol{x}_i)\}_{i=N+1}^{N+M}$, where $\boldsymbol{x}_i \in \mathcal{R}^{H \times W}$ indicates an $H \times W$ image and $\boldsymbol{y}_i \in \{0,1\}^{H \times W \times C}$ denotes the corresponding annotation for labeled data with C semantic classes. To improve the prediction quality of SAM on the target dataset, we directly fine-tune SAM in the SSL setting using all the available data. The fine-tuned SAM also functions as the final segmentation model. To enable the better use of the unlabeled data, we propose a cross prompting strategy introduced later and adapt the original architecture of SAM to a dual-branch SAM with one shared image encoder $\mathcal{E}$ and prompt encoder $\mathcal{P}$, on top of which two decoders $\mathcal{D}_1, \mathcal{D}_2$ with the same structure but different weight initializations are used to encourage the output diversity. The function of each module remains the same as the original SAM [18], with the image encoder to extract feature embeddings from the image, the prompt encoder to output prompt embeddings for given

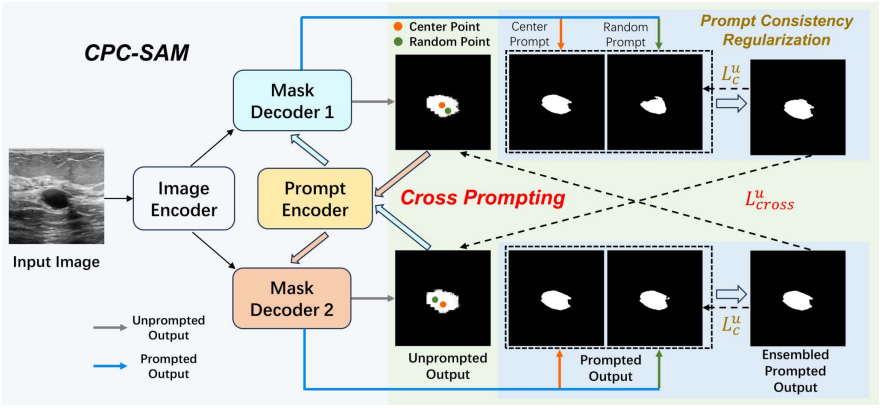

Fig. 1: The overview of our proposed method. The adapted dual-branch SAM is fine-tuned by the cross prompting loss L^u_{cross} with a prompt consistency regularization L^u_c on the unlabeled data in addition to the supervised loss L_s (L_s is not illustrated for a more concise figure). L^u_c is used to reduce SAM's sensitivity to prompt positions. L_s uses annotations to supervise both prompted and unprompted outputs for the labeled data.

prompts, and the mask decoder to produce segmentation results based on the feature embeddings and prompt embeddings. As done in [29], when no explicit prompts are given, the default dense prompt embedding is used and fine-tuned during training for automatic segmentation, which is also used during inference.

2.2 SAM-enabled Cross Prompting

Although fine-tuning on the target dataset can effectively integrate domain knowledge of specific medical images to SAM, current methods only use the small labeled set for fine-tuning [7], neglecting the potential of the large number of unlabeled images in the SSL setting and thus limiting the fine-tuning performance on target datasets. Therefore, we propose a cross prompting scheme based on the promptable property of SAM to make full use of the unlabeled data and integrate more domain knowledge on the target task to SAM.

First, we generate prompts for each other from the unprompted outputs under our dual-branch framework. Here, point prompt is used following [7,15,19,30] for its simplicity and flexibility. Take the first branch $\mathcal{D}_1$ for prompt generation as an example. Given an unlabeled image $\boldsymbol{x} \in \mathcal{U}$, we first obtain the output p^u_1 from $\mathcal{D}_1$ using the feature embedding and the default prompt embedding since no explicit prompt is provided: $p^u_1 = \mathcal{D}_1(\mathcal{E}(x), \mathcal{P}(None))$. Then, we generate the point prompt Pt_2 from p^u_1 by selecting the center or a random point of the largest connected component of the object of interest. After that, we can obtain the prompted output of $\mathcal{D}_2$: $\hat{p}^u_2 = \mathcal{D}_2(\mathcal{E}(x), \mathcal{P}(Pt_2))$. Similarly, we can obtain the unprompted output p^u_2 when using $\mathcal{D}_2$ to produce prompts and the prompted

prediction $\hat{p}_1^u$ output by $\mathcal{D}_1$. Second, we use the prompted outputs to guide the optimization of unprompted predictions based on the intuition that the output of SAM given an appropriate prompt $\hat{p}^u$ should be more accurate and reliable compared to the output without any explicit prompts p^u, since the prompt offers position information of the target area. To alleviate the confirmation bias of using the same branch to supervise itself, the prompted output $\hat{p}_1^u$ of $\mathcal{D}_1$ is used to supervise the unprompted prediction p_2^u of $\mathcal{D}_2$, and vice versa. Therefore, our cross prompting loss is a symmetrical constraint with a combination of the Dice loss and the cross-entropy loss:

$$L_{cross}^u = \frac{1}{2}\left[L_{dice}(p_1^u, \hat{p}_2^u) + L_{ce}(p_1^u, \hat{p}_2^u)\right] + \frac{1}{2}\left[L_{dice}(p_2^u, \hat{p}_1^u) + L_{ce}(p_2^u, \hat{p}_1^u)\right] \quad (1)$$

Compared with the vanilla cross pseudo supervision method that directly uses the output to supervise each other [8], our cross prompting scheme makes full use of the promptable property of SAM as a refinement step to obtain a better pseudo label as a better guidance than the unprompted output. The final SSL performance is thus improved as shown in the ablation studies.

2.3 Prompt Consistency Regularization

The cross prompting scheme solves how to utilize unlabeled images to improve the fine-tuning performance. The key to its success is generating reliable predictions for unlabeled data in the dilemma where SAM's output is sensitive to the prompt locations whereas the point prompts generated from the noisy coarse mask in SSL tend to have a high variance and a low accuracy in terms of positions. The core of alleviating this dilemma is to enhance the output invariance under various prompts. In the ideal case, predictions under two different prompts on the target area should be the same and approach the ground truth as close as possible in the meanwhile. Based on this motivation, we propose a novel prompt consistency regularization (PCR) loss to enhance the output invariance of SAM under various prompts. Take the prompted outputs of $\mathcal{D}_1$ as an example, where $\mathcal{D}_2$ is used to generate prompts, we first find the largest connected component of each semantic class of the unprompted prediction p_2^u as a post-processing step for the noisy output to increase the chance of selecting a point on the hidden ground truth area. After that, a center point and a random point are selected from the the largest connected component simultaneously, based on which two seperate predictions are obtained by taking each as the prompt, denoted as $\hat{p}_{1,c}^u, \hat{p}_{1,r}^u$, respectively. The center point is selected since it can provide a stable prediction empirically, while the random point prompt is chosen to simulate the potential prompt variance in SSL segmentation. Then, $\hat{p}_{1,r}^u$ should be similar to $\hat{p}_{1,c}^u$. Since the ensemble of multiple predictions can usually obtain a more robust result, we use the ensemble of both predictions $\hat{p}_1^u = 1/2(\hat{p}_{1,r}^u + \hat{p}_{1,c}^u)$ as a more reliable guidance for the randomly prompted prediction $\hat{p}_{1,r}^u$. Symmetrically, we can obtain the prompted outputs $\hat{p}_{2,c}^u, \hat{p}_{2,r}^u$ and the ensemble result $\hat{p}_2^u$ of $\mathcal{D}_2$. Finally, the prompt consistency regularization is applied to each decoder seperately:

$$L_c^u = \frac{1}{2}\left[L_{dice}(\hat{p}_{1,r}^u, \hat{p}_1^u) + L_{ce}(\hat{p}_{1,r}^u, \hat{p}_1^u)\right] + \frac{1}{2}\left[L_{dice}(\hat{p}_{2,r}^u, \hat{p}_2^u) + L_{ce}(\hat{p}_{2,r}^u, \hat{p}_2^u)\right] \quad (2)$$

The ensemble results $\hat{p}_1^u, \hat{p}_2^u$ are also used to be the supervision signals in Equ. 1 as a by-product of PCR. Also, since the output using the center prompt tends to be more stable, the PCR is only applied to $p_{1,r}^u, p_{2,r}^u$. The efficacy of such a design will be proved in ablation studies.

Meanwhile, to ensure the prompted outputs can approach the ground truth well, we supervise the prompted outputs $p_{1,c}^l, p_{2,c}^l, p_{1,r}^l, p_{2,r}^l$ with annotations on the labeled set besides the supervised loss on unprompted outputs p_1^l, p_2^l in SSL:

$$L_s = L_s^l(p_1^l, \boldsymbol{y}) + L_s^l(p_2^l, \boldsymbol{y}) + L_{s,p}^l(p_{1,c}^l, \boldsymbol{y}) + L_{s,p}^l(p_{2,c}^l, \boldsymbol{y}) + L_{s,p}^l(p_{1,r}^l, \boldsymbol{y}) + L_{s,p}^l(p_{2,r}^l, \boldsymbol{y}) \tag{3}$$

where $L_s^l = 0.8L_{dice} + 0.2L_{ce}$ following SAMed [29] and $L_{s,p}^l = 0.5L_{dice} + 0.5L_{ce}$. Finally, the total loss for training is the combination of the supervised loss L_s on labeled data, the cross prompting loss L_{cross}^u and the prompt consistency regularization loss L_c^u on unlabeled data: $L_{total} = L_s + \lambda_1 L_{cross}^u + \lambda_2 L_c^u$.

3 Experimental Results

Datasets. We evaluate our proposed method on two publicly available datasets: the BUSI dataset [1] and the ACDC dataset [4]. The BUSI dataset [1] consists of 647 ultrasound images for breast cancer segmentation, with 437 benign cases and 210 malignant ones. We randomly split the data on each category and finally obtain 431, 86, and 130 images for training, validation, and testing, respectively. The ACDC dataset [4] contains 200 cine MRI scans from 100 patients with three regions of interest, i.e., the right ventricle cavity, the myocardium, and the left ventricle cavity. Following [3,6], the dataset is randomly split on the patient level, with 70 patients for training, 10 for validation, and 20 for testing.

Implementation Details and Evaluation Metrics. Our method is implemented by Pytorch and trained on an NVIDIA A40 GPU. Most training settings are the same on both datasets. Specifically, the ViT_B version of SAM is employed. Following [29], we apply LoRA [14] to the query and value heads in each transformer block of $\mathcal{E}$ with $r = 4$ and optimize all the parameters in $\mathcal{P}$ and $\mathcal{D}_1, \mathcal{D}_2$ through the normal back propagation. We load the pre-trained weights for the image encoder and prompt encoder, whereas two decoders are initialized randomly. We expand the number of point prompt embeddings in the prompt encoder to the number of semantic classes for multi-class segmentation. Also, we increase the output resolution of mask decoders by the progressive upsampling strategy in [5]. The input images are resized to 512×512 and normalized to [0,1]. Data augmentation used in training include random rotation between [-20°, 20°] and random flips. Our adapted SAM is optimized by an AdamW optimizer for 10000 epochs. The same warmup and exponential learning rate decay strategy as [29] are adopted, setting the maximum learning rate as 0.001 and the warmup period as 5000 iterations. λ_1 and λ_2 are empirically set as 0.4 and 0.05. The batch size is 6 and 12 for the BUSI and ACDC dataset, respectively, each containing half labeled data. As done in [3,6], four evaluation metrics are taken, including the Dice similarity coefficient (DSC), Jaccard (JC), 95% Hausdorff Distance

Table 1: Comparisons with SOTA methods on the BUSI and ACDC dataset. Column "#Lab" denotes the number of labeled data and the number of all training data, respectively.

Method	BUSI					ACDC				
	#Lab	DSC↑	JC↑	HD95↓	ASD↓	#Lab	DSC↑	JC↑	HD95↓	ASD↓
U-Net [24]	431/431	77.19	68.29	75.03	31.21	70/70	91.53	84.77	4.23	1.11
SAM-point(MIA'23) [22]	0/431	52.99	44.51	168.26	91.78	0/70	62.88	49.53	20.46	7.07
U-Net [24]	10/431	31.63	24.52	159.49	63.43	1/70	29.37	20.53	107.51	52.84
SAMed [29]	10/431	65.09	54.78	119.75	47.84	1/70	75.01	61.53	28.99	9.13
UAMT(MICCAI'19) [28]	10/431	40.93	30.96	175.31	76.51	1/70	29.14	20.14	107.69	53.58
CPS(CVPR'21) [8]	10/431	32.92	25.70	144.92	50.54	1/70	30.46	21.00	95.74	45.48
URPC(MIA'22) [20]	10/431	32.16	24.75	151.59	64.97	1/70	31.00	20.81	123.03	59.94
MC-Net+(MIA'22) [26]	10/431	36.24	27.45	167.91	71.80	1/70	38.84	28.58	62.21	30.67
DCNet(MICCAI'23) [6]	10/431	42.14	32.11	154.39	64.21	1/70	41.13	31.61	56.16	24.71
BCP(CVPR'23) [3]	10/431	61.81	51.12	112.91	38.15	1/70	68.39	56.8	50.9	21.99
UniMatch(CVPR'23) [27]	10/431	60.98	49.85	109.79	47.50	1/70	84.47	74.25	15.36	4.57
SemiSAM [30]	10/431	43.43	32.48	177.30	84.46	1/70	34.18	23.96	100.75	47.03
CPC-SAM (ours)	10/431	**71.20**	**61.15**	**100.22**	**37.86**	1/70	**85.56**	**75.74**	**9.19**	**2.84**
U-Net [24]	20/431	44.22	34.73	160.04	69.52	3/70	45.95	35.96	71.11	32.47
SAMed [29]	20/431	67.28	57.55	107.31	49.70	3/70	83.04	71.98	14.93	4.05
UAMT(MICCAI'19) [28]	20/431	45.83	35.84	163.53	80.92	3/70	56.67	45.93	15.06	45.24
CPS(CVPR'21) [8]	20/431	46.74	37.61	142.73	56.70	3/70	56.87	46.88	20.18	2.91
URPC(MIA'22) [20]	20/431	45.26	35.51	173.11	73.47	3/70	55.98	44.75	40.47	14.13
MC-Net+(MIA'22) [26]	20/431	47.29	33.00	183.14	84.53	3/70	65.37	54.18	27.64	6.32
DCNet(MICCAI'23) [6]	20/431	56.87	46.60	130.31	56.14	3/70	72.21	62.27	26.50	10.59
BCP(CVPR'23) [3]	20/431	65.54	56.05	**93.07**	**39.09**	3/70	87.57	78.58	8.68	2.30
UniMatch(CVPR'23) [27]	20/431	62.47	51.48	100.73	45.88	3/70	87.31	78.20	8.62	2.74
SemiSAM [30]	20/431	50.09	38.63	170.42	77.85	3/70	51.01	39.45	70.13	28.26
CPC-SAM (ours)	20/431	**72.41**	**62.72**	96.26	40.93	3/70	**87.95**	**79.01**	**5.80**	**1.54**

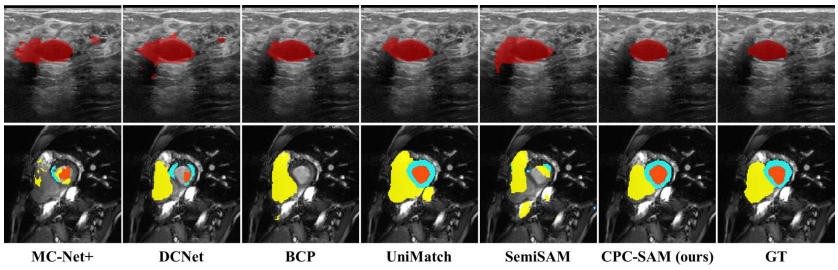

Fig. 2: Visualizations of different methods on the BUSI dataset with 10 labeled images (top) and on the ACDC dataset with 1 labeled patient (bottom).

(95HD), and the average surface distance (ASD). The unit of DSC and JC is %. The unit of HD95 and ASD is pixel and mm on the BUSI and ACDC dataset, respectively, since the resolution is not provided on the BUSI dataset.

Comparisons with the State-of-the-arts. We compare our method with the state-of-the-arts (SOTA) SSL methods, including UAMT [28], CPS [8], URPC [20], MC-Net+ [26], DCNet [6], BCP [3], and UniMatch [27], and a representative SAM-based SSL method SemiSAM [30]. We also compare with the supervised counterparts trained with labeled data alone, i.e., U-Net [24] and SAMed [29]. The zero-shot performance of SAM is also included using the center point prompt generated from the ground truth labels following [22], denoted as

SSL	Cross Prompting	PCR	DSC↑	JC↑	HD95↓	ASD↓
			77.26	64.88	15.72	5.22
✓			80.20	68.44	19.48	6.52
✓	✓		84.75	74.39	13.41	3.84
✓	✓	✓	**85.56**	**75.74**	**9.19**	**2.84**

Table 2: Ablation studies of different components of our method on the ACDC dataset.

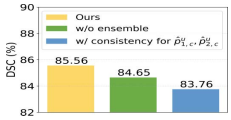

Fig. 3: Effects of various consistency constraints.

# Center	# Random	DSC↑	JC↑	HD95↓	ASD↓
0	2	84.95	74.66	12.53	3.64
1	1	**85.56**	**75.74**	**9.19**	**2.84**
1	5	84.46	74.10	13.19	3.63
1	10	84.23	73.61	12.75	3.95

Table 3: Effects of different numbers of center and random points in the PCR on the ACDC dataset.

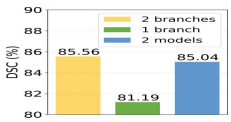

Fig. 4: Effects of different architectures.

"SAM-point". As shown in Table 1, our method outperforms other SSL methods on the BUSI dataset by a large margin of over 9.3% and 6.8% DSC when different numbers of labeled data are used. On the ACDC dataset, our method obtains the best results on all the metrics across the two labeled-data ratios. It is worth noting that although SAMed only using the labeled data can obtain comparable and even better results to some SSL methods, our proposed method obtains considerably better results on both datasets thanks to the effective use of unlabeled data. Moreover, by using SAM to generate predictions for unlabeled data, SemiSAM obtains a superior result to most SSL methods on the BUSI dataset. However, its performance on the ACDC dataset degrades steeply, probably due to the difficulty of segmenting the ring structure of the myocardium without fine-tuning as shown in [23]. The robustness and superior results of our method validate the necessity of fine-tuning on the target dataset and effectiveness of our proposed method. The ensemble results of two branches is reported in this work, but the difference between each branch and the ensembled result is marginal, e.g., 85.59%, 85.47% and 85.56% DSC for $\mathcal{D}_1, \mathcal{D}_2$, and ensemble, respectively on the ACDC dataset with 1 labeled patient. The visualization comparisons in Fig. 2 further validate the superiority of our method.

Ablation Studies. Table 2 shows ablation studies on the key components of our method with the ACDC dataset. Obviously, introducing unlabeled data into fine-tuning (Row 2-4) outperforms training on the labeled data alone (Row 1) with a margin of over 2.9% DSC. Also, using only the center point prompt (Row 3) for cross prompting, we can obtain superior results over the method using the unprompted outputs to supervise each other (Row 2), proving the efficacy of the cross prompting scheme. With the help of PCR, the result is further improved by 0.81% DSC. Also, we show the efficacy of specific designs of the PCR strategy in Fig. 3, proving the necessity of using ensemble results and dropping the regularization on the center prompted outputs. We further validate the efficacy of our center-random prompt selection strategy (Row 2) in Table 3. Introducing more random points leads to a lower performance, possibly because

more random point prompts can include some points outside the hidden ground truth for unlabeled data. Moreover, we apply our method to other architectures, such as the original SAM with only one mask decoder and 2 models with two SAMs (See Fig. 4). The inferior result using 1 branch might be caused by the more serious confirmation bias problem in such a self-training framework [8,17].

4 Conclusion

This paper proposes a cross prompting framework with a prompt consistency regularization to adapt SAM for SSL medical image segmentation. Comparisons on two datasets demonstrate the efficacy of our method, especially when labeled data are extremely limited. Since our method is general, it can be easily extended to medical-specific foundation models such as SAM-Med2D beyond the original SAM by changing the backbone and lead to a potential performance improvement with more domain knowledge. In the future, we'll explore more strategies to select appropriate prompts for reliable outputs.

Acknowledgments. The work described in this paper was supported in part by the Research Grants Council of the Hong Kong Special Administrative Region, China, under Project T45-401/22-N; and by the Hong Kong Innovation and Technology Fund (Project No. GHP/080/20SZ).

Disclosure of Interests. The authors have no competing interests to declare.

References

1. Al-Dhabyani, W., Gomaa, M., Khaled, H., Fahmy, A.: Dataset of breast ultrasound images. Data in brief **28**, 104863 (2020)
2. Bai, W., Oktay, O., Sinclair, M., Suzuki, H., Rajchl, M., Tarroni, G., Glocker, B., King, A., Matthews, P.M., Rueckert, D.: Semi-supervised learning for network-based cardiac mr image segmentation. In: Medical Image Computing and Computer-Assisted Intervention- MICCAI 2017: 20th International Conference, Quebec City, QC, Canada, September 11-13, 2017, Proceedings, Part II 20. pp. 253–260. Springer (2017)
3. 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)
4. Bernard, O., Lalande, A., Zotti, C., Cervenansky, F., Yang, X., Heng, P.A., Cetin, I., Lekadir, K., Camara, O., Ballester, M.A.G., et al.: Deep learning techniques for automatic mri cardiac multi-structures segmentation and diagnosis: is the problem solved? IEEE transactions on medical imaging **37**(11), 2514–2525 (2018)
5. Chen, C., Miao, J., Wu, D., Yan, Z., Kim, S., Hu, J., Zhong, A., Liu, Z., Sun, L., Li, X., et al.: Ma-sam: Modality-agnostic sam adaptation for 3d medical image segmentation. arXiv preprint arXiv:2309.08842 (2023)
6. Chen, F., Fei, J., Chen, Y., Huang, C.: Decoupled consistency for semi-supervised medical image segmentation. In: International Conference on Medical Image Computing and Computer-Assisted Intervention. pp. 551–561. Springer (2023)

7. Chen, S., Lin, L., Cheng, P., Tang, X.: Aslseg: Adapting sam in the loop for semi-supervised liver tumor segmentation. arXiv preprint arXiv:2312.07969 (2023)
8. Chen, X., Yuan, Y., Zeng, G., Wang, J.: Semi-supervised semantic segmentation with cross pseudo supervision. In: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. pp. 2613–2622 (2021)
9. Dai, H., Ma, C., Liu, Z., Li, Y., Shu, P., Wei, X., Zhao, L., Wu, Z., Zhu, D., Liu, W., et al.: Samaug: Point prompt augmentation for segment anything model. arXiv preprint arXiv:2307.01187 (2023)
10. Deng, R., Cui, C., Liu, Q., Yao, T., Remedios, L.W., Bao, S., Landman, B.A., Wheless, L.E., Coburn, L.A., Wilson, K.T., et al.: Segment anything model (sam) for digital pathology: Assess zero-shot segmentation on whole slide imaging. arXiv preprint arXiv:2304.04155 (2023)
11. Fan, D.P., Zhou, T., Ji, G.P., Zhou, Y., Chen, G., Fu, H., Shen, J., Shao, L.: Inf-net: Automatic covid-19 lung infection segmentation from ct images. IEEE transactions on medical imaging **39**(8), 2626–2637 (2020)
12. Gao, Y., Xia, W., Hu, D., Gao, X.: Desam: Decoupling segment anything model for generalizable medical image segmentation. arXiv preprint arXiv:2306.00499 (2023)
13. He, S., Bao, R., Li, J., Grant, P.E., Ou, Y.: Accuracy of segment-anything model (sam) in medical image segmentation tasks. arXiv preprint arXiv:2304.09324 (2023)
14. 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. arXiv preprint arXiv:2106.09685 (2021)
15. Huang, Z., Liu, H., Zhang, H., Xing, F., Laine, A., Angelini, E., Hendon, C., Gan, Y.: Push the boundary of sam: A pseudo-label correction framework for medical segmentation. arXiv preprint arXiv:2308.00883 (2023)
16. 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)
17. Ke, Z., Wang, D., Yan, Q., Ren, J., Lau, R.W.: Dual student: Breaking the limits of the teacher in semi-supervised learning. In: Proceedings of the IEEE/CVF international conference on computer vision. pp. 6728–6736 (2019)
18. 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. arXiv preprint arXiv:2304.02643 (2023)
19. Li, N., Xiong, L., Qiu, W., Pan, Y., Luo, Y., Zhang, Y.: Segment anything model for semi-supervised medical image segmentation via selecting reliable pseudo-labels. Available at SSRN 4477443 (2023)
20. Luo, X., Wang, G., Liao, W., Chen, J., Song, T., Chen, Y., Zhang, S., Metaxas, D.N., Zhang, S.: Semi-supervised medical image segmentation via uncertainty rectified pyramid consistency. Medical Image Analysis **80**, 102517 (2022)
21. Ma, J., He, Y., Li, F., Han, L., You, C., Wang, B.: Segment anything in medical images. Nature Communications **15**(1), 654 (2024)
22. 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)
23. Miao, J., Zhou, S.P., Zhou, G.Q., Wang, K.N., Yang, M., Zhou, S., Chen, Y.: Sc-ssl: Self-correcting collaborative and contrastive co-training model for semi-supervised medical image segmentation. IEEE Transactions on Medical Imaging (2023)

24. Ronneberger, O., Fischer, P., Brox, T.: U-net: Convolutional networks for biomedical image segmentation. In: Medical Image Computing and Computer-Assisted Intervention–MICCAI 2015: 18th International Conference, Munich, Germany, October 5-9, 2015, Proceedings, Part III 18. pp. 234–241. Springer (2015)
25. 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)
26. Wu, Y., Ge, Z., Zhang, D., Xu, M., Zhang, L., Xia, Y., Cai, J.: Mutual consistency learning for semi-supervised medical image segmentation. Medical Image Analysis **81**, 102530 (2022)
27. Yang, L., Qi, L., Feng, L., Zhang, W., Shi, Y.: Revisiting weak-to-strong consistency in semi-supervised semantic segmentation. In: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. pp. 7236–7246 (2023)
28. Yu, L., Wang, S., Li, X., Fu, C.W., Heng, P.A.: Uncertainty-aware self-ensembling model for semi-supervised 3d left atrium segmentation. In: Medical Image Computing and Computer Assisted Intervention–MICCAI 2019: 22nd International Conference, Shenzhen, China, October 13–17, 2019, Proceedings, Part II 22. pp. 605–613. Springer (2019)
29. Zhang, K., Liu, D.: Customized segment anything model for medical image segmentation. arXiv preprint arXiv:2304.13785 (2023)
30. Zhang, Y., Cheng, Y., Qi, Y.: Semisam: Exploring sam for enhancing semi-supervised medical image segmentation with extremely limited annotations. arXiv preprint arXiv:2312.06316 (2023)
31. Zhang, Y., Zhou, T., Wang, S., Wu, Y., Gu, P., Chen, D.Z.: Samdsk: Combining segment anything model with domain-specific knowledge for semi-supervised learning in medical image segmentation. arXiv preprint arXiv:2308.13759 (2023)