Self-Paced Sample Selection for Barely-Supervised Medical Image Segmentation

Junming Su[1,2,†], Zhiqiang Shen[1,2,†], Peng Cao[1,2,3(✉)], Jinzhu Yang[1,2,3], and Osmar R. Zaiane[4]

[1] Computer Science and Engineering, Northeastern University, Shenyang, China
[2] Key Laboratory of Intelligent Computing in Medical Image of Ministry of Education, Northeastern University, Shenyang, China
[3] National Frontiers Science Center for Industrial Intelligence and Systems Optimization, Shenyang, China
`caopeng@mail.neu.edu.cn`
[4] Alberta Machine Intelligence Institute, University of Alberta, Edmonton, Canada

Abstract. The existing barely-supervised medical image segmentation (BSS) methods, adopting a registration-segmentation paradigm, aim to learn from data with very few annotations to mitigate the extreme label scarcity problem. However, this paradigm poses a challenge: pseudo-labels generated by image registration come with significant noise. To address this issue, we propose a self-paced sample selection framework (SPSS) for BSS. Specifically, SPSS comprises two main components: 1) self-paced uncertainty sample selection (SU) for explicitly improving the quality of pseudo labels in the image space, and 2) self-paced bidirectional feature contrastive learning (SC) for implicitly improving the quality of pseudo labels through enhancing the separability between class semantics in the feature space. Both SU and SC are trained collaboratively in a self-paced learning manner, ensuring that SPSS can learn from high-quality pseudo labels for BSS. Extensive experiments on two public medical image segmentation datasets demonstrate the effectiveness and superiority of SPSS over the state-of-the-art. Our code is release at *https://github.com/SuuuJM/SPSS*.

Keywords: Barely-Supervised Learning · Self-Paced Learning · Contrast Learning.

1 Introduction

With the continuous development of deep learning-based methods, medical image segmentation has been significantly advanced [3,12], relying on large amounts of labeled data. However, annotating medical images at the pixel level is laborious and requires professional knowledge, making large-scale labeled data expensive or even unavailable. Semi-supervised learning (SSL) [15] primarily builds

---

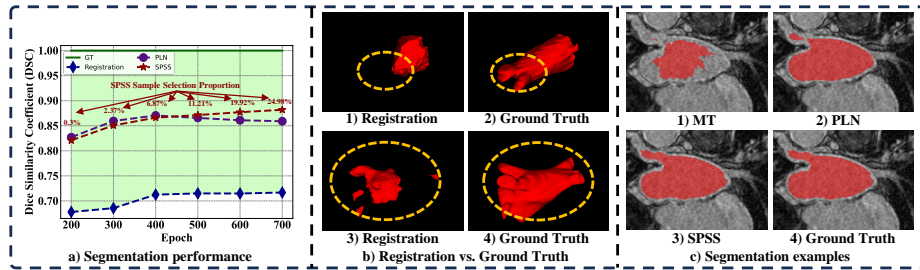† Junming Su and Zhiqiang Shen contributed equally to this work.

Fig. 1. Illustration of a) segmentation performance of PLN and SPSS, b) registration noise, and c) qualitative results of MT, PLN, and SPSS. Note that the green region in a) is the performance drop from ground truth caused by registration noise.

upon the pseudo labeling [9] and consistency regularization [14] techniques, achieving state-of-the-art (SOTA) performance on common label scarcity scenarios, e.g., with 10% labeled data, for medical image segmentation [18]. However, this line of methods faces a critical limitation when encountering the extreme label scarcity problem where the training set includes a barely-annotated labeled set with single-slice annotations and an unlabeled set with numerous images. As depicted in Fig. 1c.1), for instance, the representative SSL method, Mean-Teacher [14], yields unsatisfactory performance on the 3D left atrium segmentation task [16] with barely-annotated data.

Barely-supervised learning (BSL) has emerged to solve such extreme label scarcity for further alleviating the annotation burden [10,1,2]. SOTA barely-supervised medical image segmentation (BSS) methods, such as PLN [10], adopt a registration-segmentation paradigm to construct 3D pseudo-labels from single-slice annotations. This paradigm mainly includes the following procedure: 1) a registration module constructs volumetric pseudo labels using single-slice annotations of barely labeled data, 2) a teacher segmentation model generates pseudo labels for both barely labeled data and unlabeled data, and 3) a student segmentation model is trained based on the fused pseudo labels produced by the first two steps. However, the pseudo labels generated by image registration are unreliable and noisy. This registration noise inevitably degrades the performance of the segmentation module, resulting in inferior performance. To illustrate it, we conducted a preliminary experiment on the left atrial segmentation task for BSS. Quantitatively, as depicted in Fig. 1a), the registration performance only reaches 60%-70% in terms of dice similarity coefficient (DSC), indicating a significant gap between the registered 3D pseudo labels and the ground truth annotations. Qualitatively, the pseudo labels also differ substantially from ground truth [Fig. 1b)]. Based on the registration-segmentation paradigm, the state-of-the-art BSS approach, PLN [10], shows a performance drop in the latter epochs of the training stage [Fig. 1a)] and fails in the challenging boundary regions [Fig. 1c.2)]. *A core problem in the registration-segmentation paradigm naturally arises: how to*

improve the quality of pseudo labels during training for barely-supervised medical image segmentation.

To this end, we propose a self-paced sample selection framework (SPSS), tailored for BSS. As shown in Fig. 1a), the DSC curve of our SPSS is higher than that of PLN [10] during the latter epochs of the training phase, suggesting that the segmentation performance of SPSS is consistently improved due to the selected high-quality pseudo labels. Considering the difficulty in directly enhancing the capability of the registration module with a limited number of labeled slices[10], we provide an alternative that aims at strengthening the segmentation capability to alleviate the impact of registration noise by improving the quality of pseudo labels. *The main idea of SPSS lies in improving the quality of pseudo labels in both the image and feature spaces by self-paced sample selection.* To realize this idea, SPSS consists of two components: 1) self-paced uncertainty sample selection (SU), and 2) self-paced bidirectional feature contrastive learning (SC). On the one hand, SU explicitly selects high-quality pseudo labels in the image space at the voxel level. On the other hand, SC enhances the separability of class semantics in the feature space through bidirectional feature contrastive learning. Both SU and SC are trained collaboratively in a self-paced learning manner, guaranteeing that SPSS can learn from high-quality pseudo labels for BSS. We evaluate the proposed method on two benchmark datasets for BSS. Experimental results show that our SPSS has significant improvements over the state-of-the-art. For instance, compared with the state-of-the-art BSS method, i.e. PLN [10], our SPSS obtains DSC gains of 2.18% and 1.25%; it also outperforms the representative semi-supervised segmentation method UA-MT [18] by a large margin of 23.08% and 13.98% in terms of DSC on the LA and KiTS datasets with 20% barely-labeled data, respectively.

In a nutshell, our contributions mainly include:

- We design SU and SC to effectively improve the quality of pseudo labels in both the image and feature spaces.
- We further train SU and SC collaboratively through a self-paced learning scheme, to construct a novel BSS framework SPSS.
- Extensive experiments on left atrial and kidney segmentation tasks validate the effectiveness of SPSS and suggest that our SPSS provides an alternative aiming to strengthen segmentation capability, alleviating the impact of registration noise by improving the quality of pseudo-labels.

2   Methodology

The training set $\mathcal{D} = \{\mathcal{D}_L, \mathcal{D}_U\}$ contains a labeled dataset $\mathcal{D}_L = \{X_i^l, y_{k_i}^l\}_{i=1,...,N_L}$ and an unlabeled dataset $\mathcal{D}_U = \{X_j^u\}_{j=1,...,N_U}$, where X_i^l represents the i_{th} labeled image with the corresponding single-slice annotation y^{k_i} (only the k_{th} slice in X_i^l has annotation), X_j^u denotes the i_{th} unlabeled image.

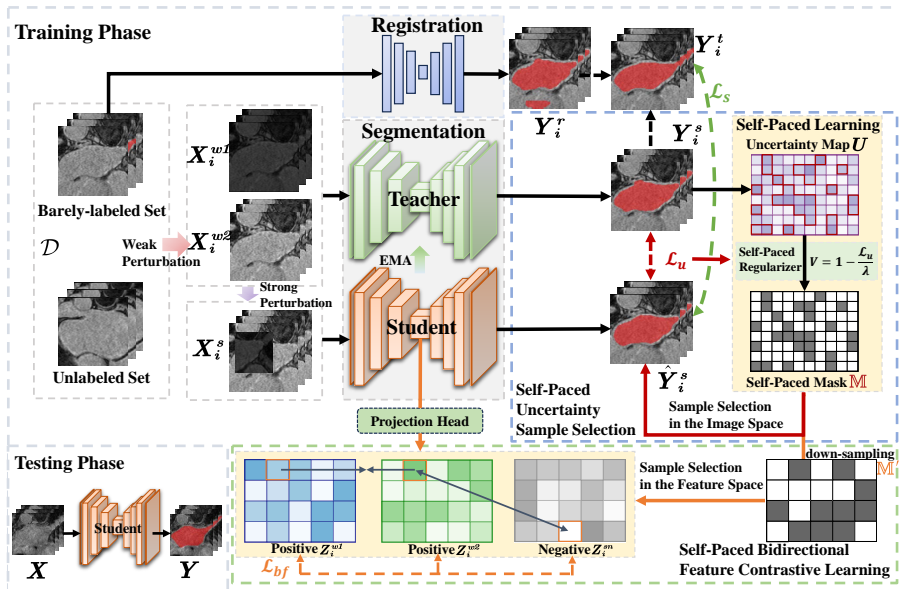Samples refer to pixels in both the image and feature spaces.

Fig. 2. Overview of the proposed self-paced sample selection framework (SPSS). SPSS includes: 1) a self-paced uncertainty sample selection strategy (SU) for explicitly pseudo labels selection in the image space and 2) a self-paced bidirectional feature contrastive learning scheme (SC) for class semantics discrimination in the feature space.

2.1   Overview

To address the issue of the noisy pseudo labels generated by registration model in the registration-segmentation paradigm, we propose a self-paced sample selection framework (SPSS) [Fig. 2] to improve the quality of pseudo labels in both the image and feature spaces. Specifically, SPSS involves 1) a self-paced uncertainty sample selection strategy (SU) to explicitly select high-quality pseudo-labels in the image space and 2) a self-paced bidirectional feature contrastive learning scheme (SC) to enhance the discrimination of class semantics in the feature space. SU and SC are trained collaboratively in a self-paced learning manner.

2.2   Self-Paced Uncertainty Sample Selection

Thresholding is a common strategy for pseudo label selection [13]. A fixed threshold is arbitrary for pixel/voxel-level sample selection of an image with numerous pixels. However, it is challenging to determine the appropriate thresholds manually. Motivated by the concept of self-paced learning [11,8], instead of setting the pixel-level thresholds manually, we propose SU to gradually select high-quality pseudo labels from easy to hard at the voxel level. During the self-paced learning, a course of difficulty ranking and an age parameter λ are required. In this study, we determine the self-paced course and the age parameters λ based on the

model's learning status by considering two aspects: model uncertainty and loss. Formally, SU calculates a self-paced mask $\mathbb{M}$ to select K most certain voxels with high-quality pseudo labels in the current iteration. The self-paced mask $\mathbb{M}$ is defined as:

$$\mathbb{M} = \mathbb{I}\{U < Sort(U)[K]\}, \tag{1}$$

where $\mathbb{I}\{\cdot\}$ represents the indicator function, $Sort(\cdot)$ is an ascending sorting function, K indicates the index of the K_{th} most certain voxel, and U denotes the uncertainty map of the model for the current input. We employ Monte Carlo dropout [6] to estimate the uncertainty U: $U' = \frac{1}{T} \sum_t p_t^c$ and $U = -\sum_c U' log U'$, where p_t^c is the probability of the c_{th} class in the t_{th} prediction. The number of certain voxels K is determined by: $K = R_{conf} \cdot \mathcal{H} \cdot \mathcal{W} \cdot \mathcal{D}$, where $\mathcal{H}$, $\mathcal{W}$, and $\mathcal{D}$ represent the height, width, and depth of the input image, respectively. R_{conf} denotes the confident ratio depending on the model's status:

$$R_{conf} = \begin{cases} 0.1 * min(\xi(t) \cdot \tau, 1), \text{ if } \mathcal{L}_u \geq \lambda \\ v \cdot min(\xi(t) \cdot \tau, 1), \quad \text{otherwise } \mathcal{L}_u < \lambda \end{cases} \tag{2}$$

where τ, $\xi(t)$, and λ are the temperature coefficient, the warm-up function, and the age parameters, respectively. $\mathcal{L}_u$ refers to the unsupervised loss that indicates the model's status. We define the self-paced weight v as: $v = (1 - \frac{\mathcal{L}_u}{\lambda})$. In the initial training stage, i.e., $\mathcal{L}_u \geq \lambda$, the model may consider all voxels as hard samples, leading to a scenario where no samples are selected. To address this, we introduce a warm-up in the early stage of self-paced learning. The warm-up function $\xi(t)$ increases as the number of training iterations increases, controlling the confident proportion R_{conf} for gradually including samples during training. As training progresses, since the model becomes more reliable, i.e., $\mathcal{L}_u < \lambda$, more high-quality samples can be involved for training.

2.3   Self-Paced Bidirectional Feature Contrastive Learning

Enhancing the separability between different class semantics in the feature space can further improve the quality of pseudo labels [4,20]. Nevertheless, it is crucial to consider how to select appropriate positive and negative pixels in the feature space while avoiding false negatives. To this end, we introduce a novel self-paced bidirectional feature contrastive learning scheme (SC), where the contrastive samples are selected in a self-paced learning manner [Fig. 2].

Specifically, SC utilizes the self-paced mask $\mathbb{M}$ generated in SU to screen positive and negative samples for contrastive learning. We consider the corresponding positions in the two weakly perturbed feature maps Z_i^{w1} and Z_i^{w2} that have consistent predictions as positive samples. Meanwhile, the samples in the strong-perturbed feature map Z_i^{sn} with different predictions from the positive samples are regarded as negative ones. Due to the significantly larger number of negative samples compared to positive samples, we further select the top K most confident negative samples to prevent an overwhelming imbalance of negatives.

Based on the selected positive and negative samples, we define the bidirectional feature contrast loss $\mathcal{L}_{bf}$ as:

$$\mathcal{L}_{bf} = \sum_{z_i^{w1} \in Z_i^{w1}, z_i^{w2} \in Z_i^{w2}, z_i^{sn} \in Z_i^{sn}} \mathcal{L}_f(z_i^{w1}, z_i^{w2}, z_i^{sn}) + \mathcal{L}_f(z_i^{w2}, z_i^{w1}, z_i^{sn}) \quad (3)$$

where $Z_i^{w[1/2]}$ and Z_i^{sn} are the selected positive and negative feature maps. These feature maps are obtained by multiplying the feature maps $F_i^{w1/2}$ and F_i^{sn} from the projection head with the self-paced mask $\mathbb{M}$. Note that we down-sample $\mathbb{M}$ to match the size of the feature maps. z_i^{w1}, z_i^{w2}, z_i^{sn} are the corresponding samples in these feature maps. $\mathcal{L}_f$ is the feature contrast loss, defined in Eq. 4, for pulling together the positive samples and pushing away the negative samples.

$$\mathcal{L}_f(z_i^{w1}, z_i^{w2}) = -log \frac{exp\{cos(z_i^{w1}, z_i^{w2})/\tau\}}{exp\{cos(z_i^{w1}, z_i^{w2})/\tau\} + \sum_{j=1}^{K} exp\{cos(z_i^{w1}, z_{ij}^{sn})/\tau\}} \quad (4)$$

where $cos(\cdot)$ is the cosine similarity function.

2.4   Loss Function

The overall loss for training SPSS is defined as: $\mathcal{L} = \mathcal{L}_s + \mathcal{L}_u + \mathcal{L}_{bf}$, where $\mathcal{L}_{bf}$, $\mathcal{L}_s$, and $\mathcal{L}_u$ represent the aforementioned contrastive loss, the supervised loss, and the unsupervised loss, respectively. The supervised loss $\mathcal{L}_s$ is calculated by: $\mathcal{L}_s = \mathcal{L}_{Dice}(\hat{Y}_i^s, Y_i^t) + \mathcal{L}_{CE}(\hat{Y}_i^s, Y_i^t)$, where Y_i^t is the weighted fusion of the registration pseudo labels Y_i^r and the segmentation pseudo labels Y_i^s. Based on the self-paced mask $\mathbb{M}$, the unsupervised loss $\mathcal{L}_u$ is formulated as:

$$\mathcal{L}_u = \mathcal{L}_{Dice}(\mathbb{M} \cdot \hat{Y}_i^s, \mathbb{M} \cdot Y_i^s) + \mathcal{L}_{CE}(\mathbb{M} \cdot \hat{Y}_i^s, \mathbb{M} \cdot Y_i^s) \quad (5)$$

where X_i^s, Y_i^s, and $\hat{Y}_i^s$ denote the strong-perturbed image, the corresponding pseudo label, and the prediction of the student model, respectively. We apply the CutMix operation [19] on the two weak-perturbed images X_i^{w1} and X_i^{w2}, as well as their corresponding pseudo labels Y_i^{w1} and Y_i^{w2}, to generate X_i^s and Y_i^s.

3   Experiments and Results

Dataset We evaluate the proposed SPSS on the 2018 Left Atrial Segmentation Challenge (LA) dataset [16] and the 2019 Kiney Segmentation Challenge (KiTS) dataset [7]. For the LA dataset, we randomly divide the 100 scans into 80 training samples and 20 testing samples. For the KiTS dataset, we randomly split the total 210 images into 190 and 20 samples for training and testing, respectively. **Implementation Details** We implement our method on an NVIDIA A30 Tensor Core GPU using the PyTorch framework. We utilize the Adam optimizer with a fixed learning rate of 1e-4 for the registration module. We leverage

Table 1. Comparison with state-of-the-art methods on the LA dataset

Method		Scans Used		Metrics			
		L/U	Labeled Slices	DSC(%)	Jaccard(%)	ASD(voxel)	HD(voxel)
Barely-supervised	MT [14]	16/64	16	59.80	42.97	13.44	34.55
	UA-MT [18]	16/64	16	63.11	46.78	17.01	44.15
	CPS [5]	16/64	16	59.19	42.79	20.00	50.83
	FixMatch [13]	16/64	16	61.75	44.94	11.98	29.87
	UniMatch [17]	16/64	16	60.74	44.13	13.21	31.30
	PLN [10]	16/64	16	84.01	72.76	5.62	22.07
	SPSS (ours)	16/64	16	**86.19**	**75.89**	**3.49**	**13.54**
Semi-supervised	MT [14]	16/64	1280	88.12	79.03	2.65	10.92

Table 2. Comparison with state-of-the-art methods on the KiTS dataset

Method		Scans Used		Metrics			
		L/U	Labeled Slices	DSC(%)	Jaccard(%)	ASD(voxel)	HD(voxel)
Barely-supervised	MT [14]	38/152	38	73.58	58.21	18.76	58.21
	UA-MT [18]	38/152	38	76.24	61.61	3.02	8.00
	CPS [5]	38/152	38	76.02	63.32	4.96	20.83
	FixMatch [13]	38/152	38	74.32	61.10	**1.47**	8.83
	UniMatch [17]	38/152	38	74.81	61.55	1.58	8.85
	PLN [10]	38/152	38	88.97	81.02	1.83	**5.98**
	SPSS (ours)	38/152	38	**90.22**	**83.55**	1.72	8.48
Semi-supervised	MT [14]	38/152	2432	92.56	87.67	1.70	6.26

MT [14] as the baseline framework of our SPSS, where the teacher model is employed to generate pseudo labels and updated by an exponential moving average (EMA) of the student model. The student model is trained using the SGD optimizer for 6000 iterations, where the initial learning rate is 0.01 and gradually decays by 0.1 every 2500 iterations. Following PLN [10], we randomly select the middle slice in a volume for image registration. The warm-up function is $\xi(t) = min(0.1 \times exp[-5(1 - t/t_{max})^2], 1)$. The age parameter is set as $\lambda = \alpha \cdot \delta$, which is an initial value $\alpha = 0.1$ growing with a factor $\delta = 1.01$. Please refer to the supplementary materials for the investigation of these hyperparameters.

3.1   Comparison with SOTA

We compare SPSS with SOTA semi-supervised and barely-supervised methods: MT [14], UA-MT [18], CPS [5], FixMatch [13], UniMatch [17], PLN [10].

Results on LA In Table 1 and Fig. 3, we present the quantitative and qualitative results on the LA dataset. Specifically, the SSL methods show inferior performance for barely-supervised left atrial segmentation. This result suggests that the SSL approaches cannot handle the barely-supervised learning problem. When compared to the barely-supervised SOTA, i.e., PLN, our method achieves an improvement of 2.18% in terms of DSC. Remarkably, our SPSS, utilizing only 16 labeled slices, shows comparable performance against the SSL method that employs 1280 labeled slices.
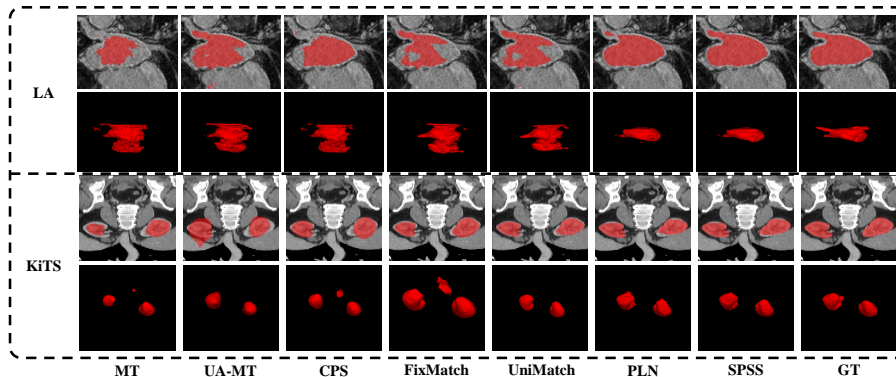
Fig. 3. Qualitative results on the LA dataset and KiTS dataset.

Table 3. Ablation study on the LA dataset with 20% barely-annotated labeled data.

Method	DSC(%)	Jaccard(%)	ASD(voxel)	HD(voxel)
Baseline	84.01	72.76	5.62	22.07
Baseline + SU	84.73	73.71	4.05	14.64
Baseline + SC	84.45	73.32	5.13	20.01
Baseline + SU + SC (SPSS)	86.19	75.89	3.49	13.54

Results on KiTS We further evaluate our SPSS on the KiTS dataset. Table 2 and Fig. 3 report the quantitative and qualitative results. Quantitatively, SPSS obtains the best performance among all the compared methods, with 90.22% DSC, 83.55% Jaccard, 1.72 ASD, and 8.48 HD. Surprisingly, SPSS, trained using only 38 labeled slices, attains performance close to MT trained with 1280 labeled slices. Similar to the situation on the LA dataset, our SPSS outperforms the SSL methods by a large margin in the barely-supervised scenario. For example, SPSS surpasses UniMatch [17] by 15.41% DSC under 38 labeled slices. Furthermore, compared with PLN [10], our method obtains a DSC gain of 1.22%. Qualitatively, SPSS shows more accurate and smoother segmentation results, especially on the segmentation boundaries. These results further demonstrate the effectiveness and advantages of SPSS for barely-supervised medical image segmentation.

3.2   Ablation Study

Table 3 reports the results on the LA dataset with 20% barely-annotated labeled data. Following PLN [10], we employ the registration-segmentation paradigm built upon MT [14] as the baseline. The results suggest that SU and SC contribute significantly to the performance improvement, which can be attributed to two main reasons: 1) SU allows for the selection of voxels with high-quality pseudo labels, and 2) SC enhances the separability between inter-class semantics in the feature space, further implicitly improving the quality of pseudo-labels.

4   Conclusion

We propose a novel framework, called SPSS, for barely-supervised medical image segmentation. We pinpoint that the limitation of the registration-segmentation paradigm lies in the noisy pseudo labels generated by image registration. Motivated by this, our main idea is to improve the quality of pseudo labels in both the image and feature spaces guided by self-paced sample selection. Extensive experiments on two public datasets, including left atrial and kidney segmentation tasks, demonstrate the effectiveness of our SPSS and suggest its capability to achieve state-of-the-art performance.

Acknowledgment

Disclosure of Interests

The authors declare that they have no conflict of interest.

References

1. Cai, H., Li, S., Qi, L., Yu, Q., Shi, Y., Gao, Y.: Orthogonal annotation benefits barely-supervised medical image segmentation. In: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. pp. 3302–3311 (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. Chen, L.C., Zhu, Y., Papandreou, G., Schroff, F., Adam, H.: Encoder-decoder with atrous separable convolution for semantic image segmentation. In: Proceedings of the European conference on computer vision (ECCV). pp. 801–818 (2018)
4. Chen, T., Kornblith, S., Norouzi, M., Hinton, G.: A simple framework for contrastive learning of visual representations. In: International conference on machine learning. pp. 1597–1607. PMLR (2020)
5. 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)
6. Gal, Y., Ghahramani, Z.: Dropout as a bayesian approximation: Representing model uncertainty in deep learning. In: international conference on machine learning. pp. 1050–1059. PMLR (2016)
7. Heller, N., Sathianathen, N., Kalapara, A., Walczak, E., Moore, K., Kaluzniak, H., Rosenberg, J., Blake, P., Rengel, Z., Oestreich, M., et al.: The kits19 challenge data: 300 kidney tumor cases with clinical context, ct semantic segmentations, and surgical outcomes. arXiv preprint arXiv:1904.00445 (2019)

8. J, S., Y, M.D., B, X.Z.: Meta self-paced learning. Sci Sin Inform **50**(6), 781–793 (2020)

9. Lee, D.H., et al.: Pseudo-label: The simple and efficient semi-supervised learning method for deep neural networks. In: Workshop on challenges in representation learning, ICML. vol. 3, p. 896. Atlanta (2013)

10. Li, S., Cai, H., Qi, L., Yu, Q., Shi, Y., Gao, Y.: Pln: Parasitic-like network for barely supervised medical image segmentation. IEEE Transactions on Medical Imaging **42**(3), 582–593 (2022)

11. Meng, D., Zhao, Q., Jiang, L.: A theoretical understanding of self-paced learning. Information Sciences **414**, 319–328 (2017)

12. 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)

13. Sohn, K., Berthelot, D., Carlini, N., Zhang, Z., Zhang, H., Raffel, C.A., Cubuk, E.D., Kurakin, A., Li, C.L.: Fixmatch: Simplifying semi-supervised learning with consistency and confidence. Advances in neural information processing systems **33**, 596–608 (2020)

14. Tarvainen, A., Valpola, H.: Mean teachers are better role models: Weight-averaged consistency targets improve semi-supervised deep learning results. Advances in neural information processing systems **30** (2017)

15. Van Engelen, J.E., Hoos, H.H.: A survey on semi-supervised learning. Machine learning **109**(2), 373–440 (2020)

16. Xiong, Z., Xia, Q., Hu, Z., Huang, N., Bian, C., Zheng, Y., Vesal, S., Ravikumar, N., Maier, A., Yang, X., et al.: A global benchmark of algorithms for segmenting the left atrium from late gadolinium-enhanced cardiac magnetic resonance imaging. Medical image analysis **67**, 101832 (2021)

17. 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)

18. 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)

19. Yun, S., Han, D., Oh, S.J., Chun, S., Choe, J., Yoo, Y.: Cutmix: Regularization strategy to train strong classifiers with localizable features. In: Proceedings of the IEEE/CVF international conference on computer vision. pp. 6023–6032 (2019)

20. Zhao, X., Qi, Z., Wang, S., Wang, Q., Wu, X., Mao, Y., Zhang, L.: Rcps: Rectified contrastive pseudo supervision for semi-supervised medical image segmentation. arXiv preprint arXiv:2301.05500 (2023)