Decoupled Consistency for Semi-supervised Medical Image Segmentation

Faquan Chen[1], Jingjing Fei[2], Yaqi Chen[1], and Chenxi Huang[1(✉)]

[1] Schoor of Informatics, Xiamen University, Xiamen, China
supermonkeyxi@xmu.edu.cn
[2] SenseTime Research, Shanghai, China

Abstract. By fully utilizing unlabeled data, the semi-supervised learning (SSL) technique has recently produced promising results in the segmentation of medical images. Pseudo labeling and consistency regularization are two effective strategies for using unlabeled data. Yet, the traditional pseudo labeling method will filter out low-confidence pixels. The advantages of both high- and low-confidence data are not fully exploited by consistency regularization. Therefore, neither of these two methods can make full use of unlabeled data. We proposed a novel decoupled consistency semi-supervised medical image segmentation framework. First, the dynamic threshold is utilized to decouple the prediction data into consistent and inconsistent parts. For the consistent part, we use the method of cross pseudo supervision to optimize it. For the inconsistent part, we further decouple it into unreliable data that is likely to occur close to the decision boundary and guidance data that is more likely to emerge near the high-density area. Unreliable data will be optimized in the direction of guidance data. We refer to this action as directional consistency. Furthermore, in order to fully utilize the data, we incorporate feature maps into the training process and calculate the loss of feature consistency. A significant number of experiments have demonstrated the superiority of our proposed method. The code is available at https://github.com/wxfaaaaa/DCNet.

Keywords: Semi-supervised Learning · Image Segementation · Consistency Regularization

1 Introduction

Deep learning technology can significantly assist clinicians in clinical diagnosis through accurate and robust segmentation of lesions or organs from medical

Supplementary Information The online version contains supplementary material available at https://doi.org/10.1007/978-3-031-43907-0_53.

H. Greenspan et al. (Eds.): MICCAI 2023, LNCS 14220, pp. 551–561, 2023.
https://doi.org/10.1007/978-3-031-43907-0_53

images [1]. In comparison to natural images, acquiring pixel-level labels of medical images involves input from clinical professionals, making such labels expensive to produce. In order to effectively alleviate the problem of data labeling, many attempts have been made towards semi-supervised medical image segmentation [2,3,5,6,11]. How to fully utilize unlabeled data in this situation becomes crucial.

Pseudo labeling [15,16] and consistency regularization [19–22] are two powerful techniques for using unlabeled data. The traditional pseudo labeling approach employs a high and fixed threshold to select the predicted pixels with high confidence as the pseudo label of unlabeled data [15]. Nevertheless, with this strategy, only a few samples can exceed the chosen threshold at the start of training. As a result, Zhang et al. [23]. present the Curriculum Pseudo Labeling (CPL) strategy, which adjusts the flexible threshold of each category dynamically during the training process. Despite the positive outcomes, low-confidence pixels will still be removed. Wang et al. [12]. demonstrated the importance of low-confidence pixels in model training. Given that consistency regularization appears to utilize data more, this research will focus more on this method.

Despite using all available prediction data, consistency regularization focuses more on how to get two similar predictions, such as data perturbation [9,10], model perturbation [25,28], feature perturbation [7,8], etc. Yet, when it comes to prediction optimization, all data is processed uniformly using a consistency loss function, such as L2 Loss. In light of this, we consider if we may decouple the prediction data into data with distinct functions, and optimize the prediction data by playing to each advantage's strengths. The benefit of this is that we can utilize all the data and, more significantly, fully utilize the advantages of various prediction data.

To sum up, this paper proposed a novel decoupled consistency regularization strategy. Specifically, inspired by CPL, we first designed a consistency threshold related to pixel confidence (Wang et al. [24] proved that the ideal dynamic threshold should be related to pixel confidence) to distinguish the consistent part from the inconsistent part, and the threshold increased as the training progressed to the maximum threshold. For the consistent part with high confidence, we use the method of cross pseudo supervision [18] to optimize. For the inconsistent part, we further decouple it into unreliable data that is likely to appear close to the decision boundary and guidance data that is more likely to be present near the high-density area. The guidance data plays the role of guiding the direction, and We don't propagate its gradient back. We refer to this action as directional consistency strategy. Additionally, we incorporate the feature map into the training process, suggest a feature consistency approach, and compute its loss in order to make better use of the data. We evaluated our method on the public PROMISE12 [14] and ACDC [13] datasets. Several experimental findings demonstrate that our strategy can significantly boost performance.

Overall, our contributions are four-fold: (1) we proposed a novel decoupled consistent semi-supervised medical image segmentation framework. The framework fully exploits prediction data, decoupled prediction data into data for

various functions, and maximizes each function's advantages. (2) A dynamic threshold is proposed that can separate the prediction data into consistent and inconsistent parts. This threshold can effectively reflect the model's learning state and encourage more diversity of data at the start of training. (3) A novel direction consistency strategy is proposed to optimize the inconsistent part. This strategy focuses on optimizing the data around the decision boundary. The results of the experiment demonstrate that the direction consistency strategy is superior to the conventional consistent regularization. (4) The feature consistency approach is suggested and the feature map is incorporated into the training, allowing for better data utilization.

2 Method

Figure 1 shows the overall pipeline of our method. To describe this work precisely, we first define some mathematical terms. Let $D = D_l \cup D_u$ be the whole provided dataset. We denote an unlabeled image as $x_i \in D_u$ and a labeled image pair as $(x_i, y_i) \in D_l$, where y_i is ground-truth. θ_{dA} and θ_{dB} represent decoder A and decoder B, respectively. θ_{dA} employs bi-linear interpolation for up-sampling, and θ_{dB} employs original transpose convolution for up-sampling. o^A and o^B represent the outputs of θ_{dA} and θ_{dB}, respectively. $p^A(p^B)$ is the segmentation confidence map, which is the network output $o^A(o^B)$ after softmax normalization.

2.1 Dynamic Consistency Threshold

In recent years, the pseudo labeling method based on threshold has achieved great success. The sampling strategy of this method can be defined as follows:

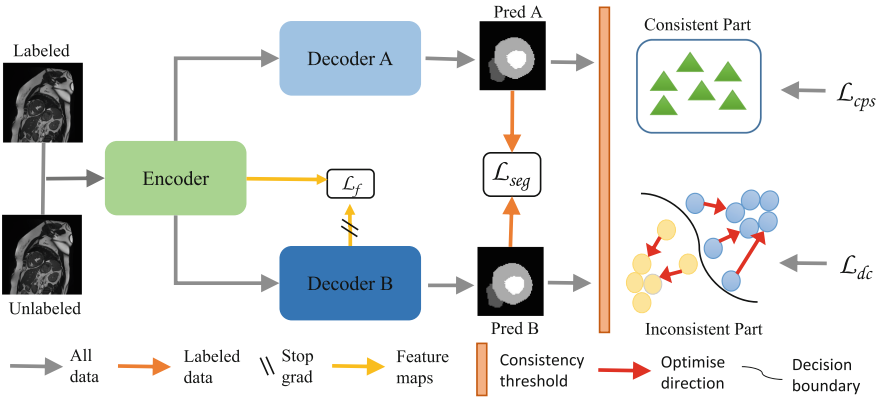

Fig. 1. Pipeline of our proposed DC-Net. DC-Net contains two decoders θ_{dA} and θ_{dB}, where θ_{dA} uses bi-linear interpolation for up-sampling and θ_{dB} uses transposed convolution for up-sampling. For the labeled data we calculate the loss $\mathcal{L}_{seg}$ between them and ground-truth, for the consistent part we calculate cross pseudo supervision loss $\mathcal{L}_{cps}$, for the inconsistent part we calculate directional consistency loss $\mathcal{L}_{dc}$, for feature maps, we calculate the feature consistency loss $\mathcal{L}_f$.

$$\tilde{y} = \mathbb{1}\left[p > \gamma\right] \tag{1}$$

where $\gamma \in (0, 1)$ is a threshold used to select pseudo labels. p is the segmentation confidence map. FlexMatch [23] has proved that in the early stage of training, in order to improve the utilization of unlabeled data and promote the diversification of pseudo labels, γ should be relatively small. As the training progresses, γ should maintain a stable proportion of pseudo labels. Therefore, our consistency threshold is defined as follows:

$$\gamma_t = (1 - \lambda)\gamma_{t-1} + \lambda \frac{1}{B} \sum_{b=1}^{B} max(p) \tag{2}$$

where B is the batch size, λ is the weight coefficient that increases with the training and we set $\lambda = \frac{t}{t_{max}}$. In order to sample more unlabeled data, we conduct threshold evaluation on p^A and p^B and select a smaller threshold as our consistency threshold. We initialize λ_t as $\frac{1}{C}$, where C indicates the number of classes. The consistency threshold γ_t is finally defined and adjusted as:

$$\gamma_t = \begin{cases} \frac{1}{C} & , if \quad t = 0 \\ min(\gamma_t^A, \gamma_t^B) & , otherwise \end{cases} \tag{3}$$

where γ_t^A and γ_t^B represent the threshold of p^A and p^B.

2.2 Decoupled Consistency

Inconsistent Part. We decouple the inconsistent part into unreliable data, which is probably going to appear at the decision boundary, and guidance data, which is more likely to occur near the high-density area. These two parts have the same index information. The distinction is that guidance data is more confident than unreliable data. Based on the smoothing assumption, the output of these two parts should be consistent and located in the high-density area. As a result, we should concentrate on optimizing the pixels around the decision boundary in order to bring them closer to the high-density region. We initially sharpen the confidence of these pixels to bring the high-confidence pixels closer to the high-density area. These are the sharpening processes:

$$Sp = softmax(o/T) \tag{4}$$

where o represents the output of the model and $T \in (0, 1)$ represents the sharpening temperature. In the experiment, we set $T = 0.5$. By comparing Sp^A and Sp^B, the high-confidence parts hSp^A, hSp^B (See Appendix Algorithm 2.) and low-confidence parts lSp^A, lSp^B can be obtained. We take $L2$ loss as the loss function of this part. Note that we only optimize the low-confidence part and do not back-propagate the gradient of the high-confidence part. Therefore, our loss of directional consistency can be written as:

$$\mathcal{L}_{dc} = L2(lSp^B, detach(hSp^A)) + L2(lSp^A, detach(hSp^B)) \tag{5}$$

Consistent Part. Similar to CPS [18], the method of cross pseudo supervision is adopted for optimization. The details are as follows:

$$\mathcal{L}_{cps} = CE(o^B, PL^A) + CE(o^A, PL^B) \tag{6}$$

where PL^A and PL^B represent corresponding pseudo labels.

Feature Part. We incorporate the feature map into the training process to further utilize the data. In order to reduce the amount of computation, we have carried out an average mapping of the feature map to reduce its dimension($\mathbb{R}^{C_m \times H_m \times W_m} \longrightarrow \mathbb{R}^{H_m \times W_m}$). The mapping process is as follows:

$$\bar{f}_m = \frac{1}{C} \sum_i^C |f_{mi}|^p \tag{7}$$

where, $p > 1$, f_m represents the feature map of $m-$th layer, and f_{mi} denotes the $i-$th slice of f_m in the channel dimension, $\bar{f}_m$ represents the corresponding mapping result. In the experiment, we set $p = 2$. Our feature consistency loss is as follows:

$$\mathcal{L}_f = \sum_{m=1}^{n} \sum_{i=1}^{N} ||\bar{f}_{mi}^e - \bar{f}_{mi}^d||_2^2 \tag{8}$$

where N is the number of pixels of $\bar{f}_{mi}$, n is the number of network layers, $\bar{f}_{mi}^e$ and $\bar{f}_{mi}^d$ represent $i-$th pixels of $m-$th feature map of decoder and encoder respectively. In this paper, only feature maps of decoder B were used to calculate the loss.

Total Loss. The total loss is a weighted sum of the segmentation loss $\mathcal{L}_{seg}$ and the other three losses:

$$\mathcal{L}_{total} = \mathcal{L}_{seg} + \alpha \mathcal{L}_{cps} + \beta \mathcal{L}_{dc} + \lambda \mathcal{L}_f \tag{9}$$

where $\mathcal{L}_{seg}$ is a Dice loss, which is applied for the few labeled data. And hyperparameter β is set as an iteration-dependent warming-up function [29], $\beta = e^{(-5(1-\frac{t}{t_{max}})^2)}$, $\lambda = 1 - \beta$, $\alpha = 0.3$ in the experiment.

3 Experiment and Results

3.1 Dataset

We evaluated our methods on ACDC dataset [13] and PROMERE12 dataset [14]. For the ACDC dataset, we randomly selected 140 scans from 70 subjects, 20 scans from 10 subjects, and 40 scans from 20 subjects as training, validation, and test sets, ensuring that each set contains data from different subjects. For the PROMISE12 dataset, we randomly divided the data into 35, 5, and 10 cases as training, validation, and test sets. And according to the previous work on datasets [27], these two data sets are segmented in 2D (piece by piece). Each slice was resized to 256×256, and the intensities of pixels are normalized to the $[0, 1]$ range.

Table 1. Comparison with five recent methods on the ACDC dataset. All results are reproduced in the form of [18,25–28] in the same experimental environment for a fair comparison.

Method	#Scans	used	Metrics			
	Labeled	Unlabeled	Dice(%)↑	Jaccard(%)↑	95HD(voxel)↓	ASD(voxel)↓
SupOnly	3 (5%)	67 (95%)	49.33	39.19	22.57	9.66
URPC [28]			55.87	44.64	13.60	3.74
CPS [18]			56.59	46.51	6.43	1.16
DTC [25]			56.90	45.67	23.36	7.39
MCNet [26]			62.85	52.29	7.62	2.33
SSNet [27]			65.82	55.38	6.67	2.28
Ours			**70.36**	**60.78**	**3.94**	**0.86**
SupOnly	7 (10%)	63 (90%)	79.37	67.78	10.73	3.12
URPC [28]			83.10	72.41	4.84	1.53
CPS [18]			84.78	74.69	7.56	2.39
DTC [25]			84.29	73.92	12.81	4.01
MCNet [26]			86.44	77.04	5.50	1.84
SSNet [27]			86.78	77.67	6.07	1.40
Ours			**89.42**	**81.37**	**1.28**	**0.38**

3.2 Implementation Details

All comparisons and ablation experiments are performed using the same experimental setting for a fair comparison. They are conducted on PyTorch using an Intel(R) Xeon(R) CPU and NVIDIA GeForce RTX 1080 Ti GPU. We adopt U-net [4] as our base network. And we use SGD as an optimizer, with a weight attenuation of 0.0005 and momentum of 0.9. The learning rate is 0.01. The batch size is set to 24, in which 12 images are labeled. All methods performed 30000 iterations during training. Moreover, a data augmentation strategy including random flipping and random rotation is exploited to alleviate overfitting.

3.3 Results

Comparison with Other Semi-supervised Methods. We use the metrics of Dice, Jaccard, 95% Hausdorff Distance (95HD), and Average Surface Distance (ASD) to evaluate the results. Table 1 gives the averaged performance of three-class segmentation including the myocardium, left and right ventricles on the ACDC dataset. Table 2 shows the segmentation results on the PROMISE12 dataset. It can be seen from the table that our method is superior to other methods. The visualized results in Fig. The visualization result in Fig. 2 shows that the segmentation result of our model is closer to the ground-truth and effectively eliminates most false-positive predictions on the ACDC (highlighted by yellow boxes at the bottom line).

Table 2. Comparison with five recent methods on the PROMISE12 dataset. All results are reproduced in the form of [18,25–28] in the same experimental environment for a fair comparison.

Method	#Scans	used	Metrics			
	Labeled	Unlabeled	Dice(%)↑	Jaccard(%)↑	95HD(voxel)↓	ASD(voxel)↓
SupOnly	4 (10%)	31 (90%)	46.08	35.40	35.86	11.23
URPC [28]			52.96	39.93	37.53	11.43
CPS [18]			49.19	37.27	39.96	11.97
DTC [25]			57.87	43.80	81.54	25.38
MCNet [26]			56.91	43.82	23.44	5.91
SSNet [27]			61.10	47.07	23.73	7.44
Ours			**68.89**	**54.88**	**12.93**	**3.75**
SupOnly	7 (20%)	28 (80%)	62.28	50.42	16.55	3.56
URPC [28]			67.04	54.01	11.54	**2.11**
CPS [18]			64.50	50.71	12.07	3.09
DTC [25]			72.03	58.32	11.48	2.65
MCNet [26]			71.77	59.07	10.76	2.85
SSNet [27]			71.56	59.35	14.38	3.03
Ours			**78.68**	**65.44**	**10.65**	2.53

Ablation Study. In order to verify the effectiveness of different loss functions of the model, we conducted experiments on the ACDC dataset with 10% labeled data, and the experimental results are shown in Table 3. It can be seen from the table that the dice scores of $\mathcal{L}_{cps}$, $\mathcal{L}_f$ and $\mathcal{L}_{dc}$ are improved by 2.93%, 3.33%, 3.60%, respectively, compared with only using $\mathcal{L}_{seg}$. On the basis of $\mathcal{L}_{seg}$ and $\mathcal{L}_f$, $\mathcal{L}_{dc}$ is also better than $\mathcal{L}_{con}$, even close to our final results. To investigate the usefulness of directional consistency further, we do not employ a threshold and compare it to classical consistency regularization and cross pseudo supervision. The experimental results are shown in Table 4. It can be seen from the table that the $\mathcal{L}_{dc}$ is obviously superior to the other two methods, which fully proves the importance of the division of different pixel functions. Also, we discovered that utilizing threshold would improve the effect of $\mathcal{L}_{cps}$. This further demonstrates the significance of our dynamic threshold.

Table 3. Ablation studies on ACDC dataset with 10% labeled data.Where $\mathcal{L}_{cps}$ represents cross pseudo supervision, $\mathcal{L}_{dc}$ represents direction consistency, and $\mathcal{L}_{f}$ represents feature consistency.

$\mathcal{L}_{seg}$	$\mathcal{L}_{cps}$	$\mathcal{L}_{dc}$	$\mathcal{L}_{f}$	Dice(%)↑	Jaccard(%)↑	95HD(voxel)↓	ASD(voxel)↓
✓				84.98	75.00	12.03	3.31
✓	✓			87.91	79.13	5.09	1.47
✓		✓		88.58	80.10	5.35	1.49
✓	✓	✓		89.29	81.21	3.07	0.92
✓			✓	88.31	79.67	3.18	1.04
✓	✓		✓	89.12	80.82	2.63	0.96
✓		✓	✓	89.39	81.36	2.32	0.82
✓	✓	✓	✓	**89.42**	**81.37**	**1.28**	**0.38**

Table 4. Comparison of different optimization methods when there is no threshold. Here $\mathcal{L}_{mse}$ represents the classic consistent regularization using the L2 loss function, and $\mathcal{L}_{cps}$ represents the cross pseudo supervision.

$\mathcal{L}_{seg}$	$\mathcal{L}_{mse}$	$\mathcal{L}_{cps}$	$\mathcal{L}_{dc}$	Dice(%)↑	Jaccard(%)↑	95HD(voxel)↓	ASD(voxel)↓
✓				84.98	75.00	12.03	3.31
✓	✓			86.47	77.18	7.31	2.01
✓		✓		87.61	78.72	7.48	1.87
✓			✓	88.71	80.31	4.05	0.98
✓		✓	✓	**88.99**	**80.68**	**3.09**	**0.92**

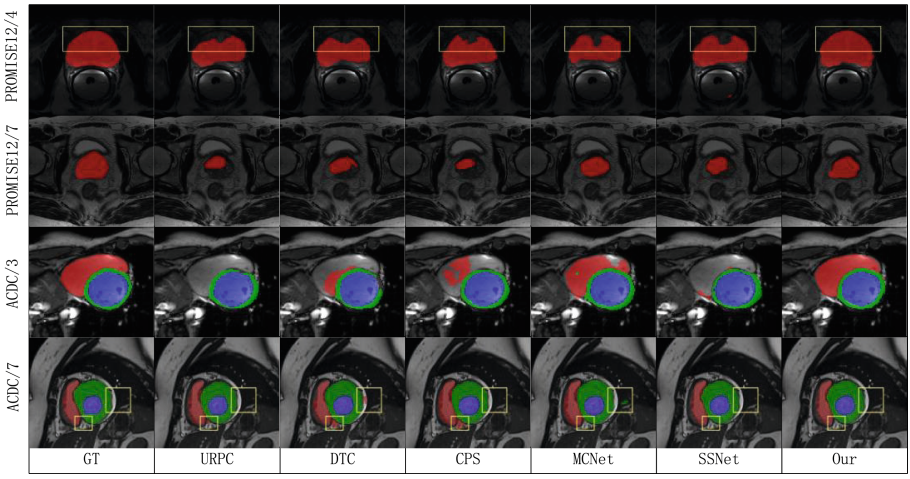

Fig. 2. Exemplar results of several semi-supervised segmentation methods and corresponding ground truth (GT) on PROMISE12 dataset (Top two rows) and ACDC dataset (Bottom two rows).

4 Conclusion

This paper proposed a framework DC-Net for semi-supervised medical image segmentation. In view of the current problem of insufficient utilization of unlabeled data, our fundamental concept is to fully exploit the benefits of data with various functionalities. Based on this, we decouple the prediction data into consistent and inconsistent parts through a dynamic threshold. Furthermore, the inconsistent part is further decoupled into guidance data and unreliable data, and optimized by a novel directional consistency strategy. Our method yielded excellent outcomes on both the ACDC and PROMISE12 datasets. In addition, directional consistency shows promising potential in the experiment, and future research will further explore the selection and treatment of directions.

Acknowledgement. This work was supported by the National Natural Science Foundation of China (Grant No. 62002304), and also by Anhui Province KevLaboratory of Translational Cancer Research (KFKT 202308), China.

References

1. Masood, S., Sharif, M., Masood, A., Yasmin, M., Raza, M.: A survey on medical image segmentation. Curr. Med. Imaging **11**(1), 3–14 (2015)
2. Li, X., Yu, L., Chen, H., Fu, C.W., Xing, L., Heng, P.A.: Transformation-consistent self-ensembling model for semisupervised medical image segmentation. IEEE Trans. Neural Netw. Learn. Syst. **32**(2), 523–534 (2020)
3. Qiao, S., Shen, W., Zhang, Z., Wang, B., Yuille, A.: Deep co-training for semi-supervised image recognition. In: Proceedings of the European Conference on Computer Vision (ECCV), pp. 135–152 (2018)
4. Ronneberger, O., Fischer, P., Brox, T.: U-Net: convolutional networks for biomedical image segmentation. In: Navab, N., Hornegger, J., Wells, W.M., Frangi, A.F. (eds.) MICCAI 2015. LNCS, vol. 9351, pp. 234–241. Springer, Cham (2015). https://doi.org/10.1007/978-3-319-24574-4_28
5. Zhou, Y., et al.: Semi-supervised 3d abdominal multi-organ segmentation via deep multi-planar co-training. In: 2019 IEEE Winter Conference on Applications of Computer Vision (WACV), pp. 121–140. IEEE (2019)
6. Xia, Y., et al.: 3D semi-supervised learning with uncertainty-aware multi-view co-training. In: Proceedings of the IEEE/CVF Winter Conference on Applications of Computer Vision, pp. 3646–3655 (2020)
7. Ouali, Y., Hudelot, C., Tami, M.: Semi-supervised semantic segmentation with cross-consistency training. In: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, pp. 12674–12684 (2020)
8. Ke, Z., Qiu, D., Li, K., Yan, Q., Lau, R.W.H.: Guided collaborative training for pixel-wise semi-supervised learning. In: Vedaldi, A., Bischof, H., Brox, T., Frahm, J.-M. (eds.) ECCV 2020. LNCS, vol. 12358, pp. 429–445. Springer, Cham (2020). https://doi.org/10.1007/978-3-030-58601-0_26
9. Kim, J., Jang, J., Park, H.: Structured consistency loss for semi-supervised semantic segmentation. arXiv preprint arXiv:2001.04647 (2020)
10. French, G., Aila, T., Laine, S., Mackiewicz, M., Finlayson, G.: Semi-supervised semantic segmentation needs strong, high-dimensional perturbations (2019)

11. Masood, S., et al.: Automatic choroid layer segmentation from optical coherence tomography images using deep learning. Sci. Rep. **9**(1), 1–18 (2019)
12. Wang, Y., et al.: Semi-supervised semantic segmentation using unreliable pseudo-labels supplementary material (2022)
13. Bernard, O., et al.: Deep learning techniques for automatic MRI cardiac multi-structures segmentation and diagnosis: is the problem solved? IEEE Trans. Med. Imaging **37**(11), 2514–2525 (2018)
14. Litjens, G., et al.: Evaluation of prostate segmentation algorithms for MRI: the promise12 challenge. Med. Image Anal. **18**(2), 359–373 (2014)
15. 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 (2013)
16. Arazo, E., Ortego, D., Albert, P., O'Connor, N.E., McGuinness, K.: Pseudo-labeling and confirmation bias in deep semi-supervised learning. In: 2020 International Joint Conference on Neural Networks (IJCNN), pp. 1–8. IEEE (2020)
17. Rizve, M.N., Duarte, K., Rawat, Y.S., Shah, M.: In defense of pseudo-labeling: an uncertainty-aware pseudo-label selection framework for semi-supervised learning. arXiv preprint arXiv:2101.06329 (2021)
18. 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)
19. Bachman, P., Alsharif, O., Precup, D.: Learning with pseudo-ensembles. In: Advances in Neural Information Processing Systems 27 (2014)
20. Sajjadi, M., Javanmardi, M., Tasdizen, T.: Regularization with stochastic transformations and perturbations for deep semi-supervised learning. In: Advances in Neural Information Processing Systems 29 (2016)
21. Bortsova, G., Dubost, F., Hogeweg, L., Katramados, I., de Bruijne, M.: Semi-supervised medical image segmentation via learning consistency under transformations. In: Shen, D., et al. (eds.) MICCAI 2019. LNCS, vol. 11769, pp. 810–818. Springer, Cham (2019). https://doi.org/10.1007/978-3-030-32226-7_90
22. Xu, Y., et al.: Dash: Semi-supervised learning with dynamic thresholding. In: International Conference on Machine Learning, pp. 11525–11536. PMLR (2021)
23. Zhang, B., et al.: FlexMatch: boosting semi-supervised learning with curriculum pseudo labeling. Adv. Neural. Inf. Process. Syst. **34**, 18408–18419 (2021)
24. Wang, Y., et al.: FreeMatch: self-adaptive thresholding for semi-supervised learning. arXiv preprint arXiv:2205.07246 (2022)
25. Luo, X., Chen, J., Song, T., Wang, G.: Semi-supervised medical image segmentation through dual-task consistency. In: Proceedings of the AAAI Conference on Artificial Intelligence, vol. 35, pp. 8801–8809 (2021)
26. Wu, Y., Xu, M., Ge, Z., Cai, J., Zhang, L.: Semi-supervised left atrium segmentation with mutual consistency training. In: de Bruijne, M., et al. (eds.) MICCAI 2021. LNCS, vol. 12902, pp. 297–306. Springer, Cham (2021). https://doi.org/10.1007/978-3-030-87196-3_28
27. Wu, Y., Wu, Z., Wu, Q., Ge, Z., Cai, J.: Exploring smoothness and class-separation for semi-supervised medical image segmentation. arXiv preprint arXiv:2203.01324 (2022)

28. Luo, X., et al.: Efficient semi-supervised gross target volume of nasopharyngeal carcinoma segmentation via uncertainty rectified pyramid consistency. In: de Bruijne, M., et al. (eds.) MICCAI 2021. LNCS, vol. 12902, pp. 318–329. Springer, Cham (2021). https://doi.org/10.1007/978-3-030-87196-3_30
29. Laine, S., Aila, T.: Temporal ensembling for semi-supervised learning. arXiv preprint arXiv:1610.02242 (2016)