Overlay Mantle-Free for Semi-Supervised Medical Image Segmentation

Jiacheng Liu[1], Wenhua Qian[1](✉), Jinde Cao[2,3], and Peng Liu[1]

[1] Yunnan University, Kunming, China
[2] Southeast University, Nanjing, China.
[3] Yonsei Frontier Lab, Yonsei Universiy, Seoul, South Korea
{liujiacheng@stu., whqian@, liuupeng0606@mail.}ynu.edu.cn, jdcao@seu.edu.cn

Abstract. Semi-supervised medical image segmentation, crucial for medical research, enhances model generalization using unlabeled data with minimal labeled data. Current methods face edge uncertainty and struggle to learn specific shapes from pixel classification alone. To address these issues, we proposed two-stage knowledge distillation approach employs a teacher model to distill information from labeled data, enhancing the student model with unlabeled data. In the first stage, we use true labels to augment data and sharpen target edges to make teacher predictions more confident. In the second stage, we freeze the teacher model parameters to generate pseudo labels for unlabeled data and guide the student model to learn. By feeding the original background image to the teacher and the enhanced image to the student, The student model learns the information hidden under the mantle and the overall shape of hidden information of the segmented target. Experimental results on the Left Atrium dataset surpass existing methods. Our Overlay Mantle-Free training method enables segmentation based on learned shape information even in data loss scenarios, exhibiting improved edge segmentation accuracy.The code is available at https://github.com/vigilliu/OMF.

Keywords: Semi-supervised learning · Segmentation · Knowledge distillation · Data augmentation.

1 Introduction

In medical image segmentation, the substantial cost of manual annotation has led to the adoption of semi-supervised learning [12,10,20,1,2,3,7,6,18,15,8], which utilizes a small amount of labeled data alongside unlabeled data to improve generalization, significantly advancing medical research. The magnetic resonance images(MRI) segmentation like Left Atrium (LA) [17] serves as a representative task for semi-supervised medical image segmentation [18,3,7,2,6,15,8] .

Several recent semi-supervised medical image methods have emerged. For instance, UA-MT [18] primarily employs Monte Carlo dropout [5] and *exponential moving average* (EMA) [13] to assess uncertainty in predictions of the teacher

model, as illustrated in (a) of Figure 1, and applies uncertainty mask to enhance learning reliability. DTC [7] utilizes a dual-task deep network to jointly predict pixel-level segmentation maps of targets and geometrically perceive level set representations. SASSnet [6] introduces adversarial loss between predictions of sdm for labeled and unlabeled data, enabling more effective capture of shape-aware features. MC-Net [16] consists of an encoder and two slightly different decoders, implementing a cyclic pseudo-labeling scheme. BCP [2] adopts a bi-directional cut-and-paste strategy similar to CutMix [19], as shown in (b) of Figure 1's top row, to reduce distribution discrepancy between labeled and unlabeled data. Inspired by UCC [4] and ClassMix [11], another data augmentation method combines pseudo-labeling [1] with consistency regularization. It applies the edge-cropped copy-and-paste technique across different images and classes in multi-class segmentation tasks to sharpen the edges of each class, thereby alleviating label contamination issues. Drawing from these principles, we utilize the concept to sharpen the segmentation targets of left atrium against other tissues, treating it as a binary classification problem based on the most certain edges of the true labels. This approach aims to mitigate edge uncertainty issues.

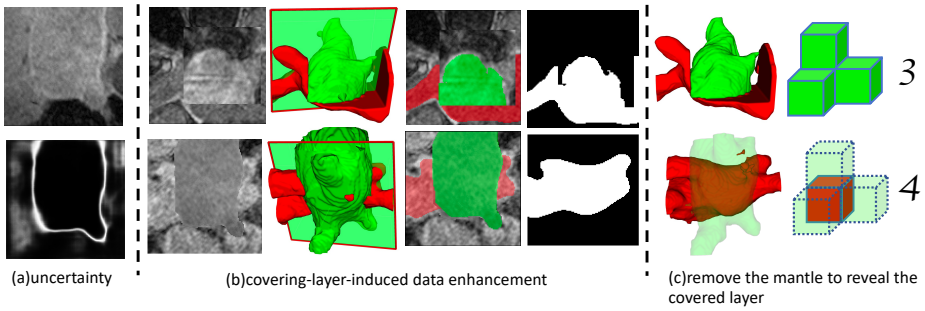

Fig. 1. Illustration of the problem we solve and the idea of the approach.(a) The first row displays an MRI example, while the second row shows the uncertainty map predicted via UA-MT. (b) The top row in showcases BCP's bidirectional copy-paste method, and the bottom row illustrates our data augmentation method and ground truth. (c) Our approach aims to eliminate data augmentation's mantle and predict the underlying information of lower layers to learn the overall shape of segmentation targets.

In semi-supervised learning, it is necessary first to learn generalizable knowledge from labeled data. The edges of segmentation targets pose a significant challenge due to high uncertainty, as depicted in Figure 1(a). While uncertainty masks are utilized in UA-MT [18] to address this issue, more confident predictions are required at the segmentation edges.Perceiving the shape of segmentation targets plays a crucial role in the accuracy of segmentation tasks [6]. However, previous methods have typically classified input images pixel by pixel without imposing constraints on the overall segmentation shape.This is evident

when input data are randomly occluded, causing the network to refrain from predicting the occluded parts, which stems from a lack of perception of the overall shape of segmentation targets.

Therefore, in this work, we propose the Overlay Left Atrium Mantle-Free for Semi-Supervised Medical Image Segmentation (OMF) method, which consists of two stages, corresponding to labeled and unlabeled data, respectively. In the first stage, labeled data are cropped and concatenated based on their true labels for data augmentation, pre-training a teacher model. This data augmentation based on true labels enhances the network's confidence in predictions. In the second stage, the parameters of the pre-trained teacher model are fixed, and the student model is initialized with these parameters. The teacher model is then used to generate pseudo-labels [1,14] for unlabeled data, which are further augmented based on these pseudo-labels. Notably, during the knowledge distillation process, background images are fed into the teacher model while mixup images are fed into the student model. Consistency loss is employed to train the student model, enabling it to remove the mantle during training and learn hidden information obscured by the mantle, as depicted in Figure 1(c), as well as the overall shape of segmentation targets. We compare our method with six recent approaches on two semi-supervised learning ratios to demonstrate its effectiveness. Additionally, we conduct ablation experiments to verify the effectiveness of data augmentation and differentiated input knowledge distillation, revealing the network's ability to accurately segment data with varying degrees of occlusion.

The main innovations of this paper are summarized as follows:(1)We propose a novel data augmentation method that crops and concatenates images along segmentation edges based on labels, boosting confidence in edge predictions and addressing uncertainty.(2)A method was developed to design differentiated inputs and fix the parameters of the teacher model during knowledge distillation, thereby allowing the student to perceive the underlying shape of segmentation targets by removing the mantle.(3)Our network achieves state-of-the-art performance in semi-supervised segmentation tasks on the LA database. Moreover, owing to its unique training approach, our method consistently outperforms others in accurately segmenting inputs with various degrees of missing data during testing.

2 Method

2.1 Model Architecture

As shown in Figure 2, we propose a semi-supervised learning method for medical image segmentation on the LA dataset, based on label-edge-based data augmentation and a two-stage knowledge distillation approach corresponding to labeled and unlabeled data, respectively. We use V-net [9] as the backbone network.Stage one and stage two correspond to labeled and unlabeled data. In stage one, we use the true labels of labeled data to extract foreground from the original images, forming a mantle, which is then pasted onto another background image for data augmentation. The training labels for the augmented teacher model use the true

labels of the foreground and background images multiplied together. In stage two, we fix the parameters of the teacher model and load it onto the student model. The pseudo labels predicted by the teacher model for unlabeled data are used for data augmentation of foreground and background images. However, during the subsequent knowledge distillation process, we use complete background images as input to the teacher model and augmented images as input to the student model for consistency loss. During training, the student model gradually tends to predict the target covered by the mantle and learns the overall shape of the target, enabling the learned model to predict segmentation targets based on the overall shape even in the absence of image content.

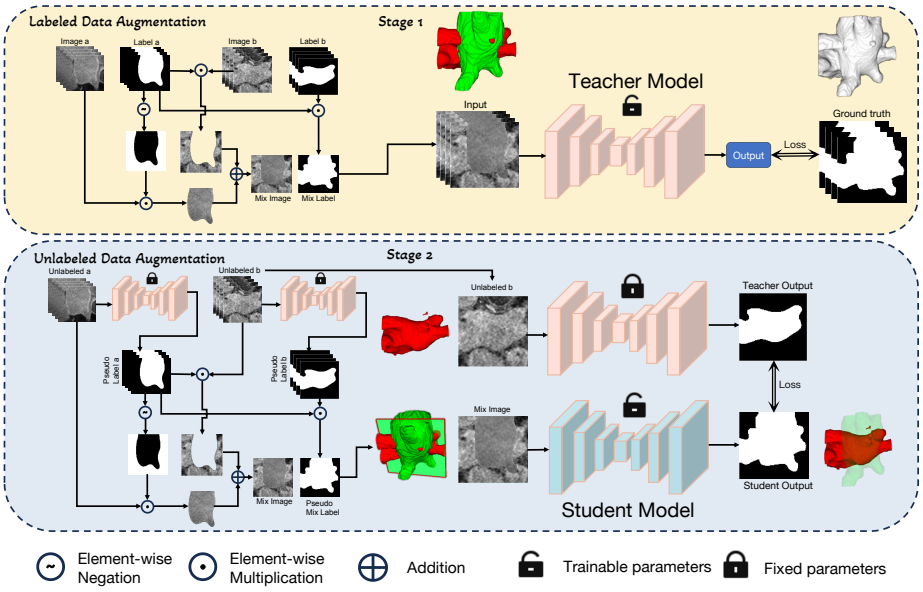

Fig. 2. Overview of our Overlay Mantle-Free method.

2.2 Pre-train via Overlay Left Atrium Data Augmentation

Inspired by previous work, we pre-train a teacher model using only labeled data in the first stage to generate pseudo-labels for subsequent use. Previous studies have shown that uncertainty mainly occurs at the edges of segmentation targets. Inspired by classmix, we segment the images based on the true labels of the labeled data to create a mantle and overlay it onto another image (background image) for data augmentation. This enables the network to make more confident predictions at the edges, thus mitigating the impact of uncertainty at the edges on training.

$\mathcal{D}_L$ represents the labeled set represents the unlabeled set, $\mathbf{X}_i$ represents the labeled elements in the set, and i ranges from 1 to N, $\mathcal{D}_U$ represents the unlabeled set, $\mathbf{X}_p$ represents the elements in the set, and p ranges from $N+1$ to $N+M$. $\mathbf{X}_i \in \mathbb{R}^{H\times W\times D}$ is the input volumes and $\mathbf{Y}_i \in \{0,1\}^{H\times W\times D}$ is the ground-truth annotations.

$\mathbf{X}_i^l, \mathbf{X}_j^l \in \mathcal{D}_L, i \neq j, \mathcal{D}_L = \{(\mathbf{X}_i^l, \mathbf{Y}_i^l)\}_{i=1}^N$. $\mathbf{X}_i^{S1mix}$ is the mixup image after our data augmentation. The i-th image is taken as the foreground image. The mantle is obtained by multiplying the inverse of its label $\mathbf{Y}_i^l$ with the original image $\mathbf{X}_i^l$. Additionally, the j-th image is taken as the background, and it is obtained by multiplying it with $\mathbf{Y}_i^l$. The resulting mixture is denoted as $\mathbf{X}_i^{S1mix}$, given by the sum of the mantle and background images.The data augmentation process for labeled data can be expressed as:

$$\mathbf{X}_i^{S1mix} = \mathbf{X}_j^l \odot \mathbf{Y}_i^l + \mathbf{X}_i^l \odot (1 - \mathbf{Y}_i^l), \;\; \mathbf{Y}_i^{S1mix} = \mathbf{Y}_i^l \odot \mathbf{Y}_j^l \tag{1}$$

For the training labels of data augmentation, we multiply the labels of image i and image j to obtain $\mathbf{Y}_i^{S1mix}$, which serves as the label for the mixup image $\mathbf{X}_i^{S1mix}$.

The total data used in stage 1 consists of $\mathcal{D}_{S1} = (\mathbf{X}_i^{S1mix}, \mathbf{Y}_i^{S1mix})_{i=1}^N \cup \mathcal{D}_L$.The data used here are discussed in table 2 in the ablation studies. The segmentation training loss of pre-training in the first stage is $\mathcal{L}_{seg}$:

$$\begin{aligned}\mathcal{L}_{seg} = \frac{1}{N}\Bigg(&\sum_{i=1}^{N} \mathcal{L}_{CE}(f_{seg}(\{\mathbf{X}_i, \mathbf{X}_i^{S1mix}\};\theta), \{\mathbf{Y}_i, \mathbf{Y}_i^{S1mix}\}) \\ &+ \sum_{i=1}^{N} \mathcal{L}_{dice}(f_{seg}(\{\mathbf{X}_i, \mathbf{X}_i^{S1mix}\};\theta), \{\mathbf{Y}_i, \mathbf{Y}_i^{S1mix}\})\Bigg)\end{aligned} \tag{2}$$

$\mathcal{L}_{CE}$ denotes the cross-entropy loss, while $\mathcal{L}_{dice}$ signifies the Dice loss.The prediction result of the network for the input x_i and $\mathbf{X}_i^{S1mix}$ under the θ parameter is $f_{seg}(\{\mathbf{X}_i, \mathbf{X}_i^{S1mix}\};\theta)$.

2.3 Mantle-Free Knowledge Distillation

For the unlabeled data $\mathcal{D}_U = \{\mathbf{X}_p^u\}_{p=N+1}^{N+M}$ to be used in the second stage, we will utilize the pre-trained parameters θ from the first stage to generate pseudo-labels $\widetilde{\mathbf{Y}}_p$ for $\mathbf{X}_p^u$:

$$\widetilde{\mathbf{Y}}_p = f_{seg}(\mathbf{X}_p^u;\theta) \tag{3}$$

$\mathbf{X}_p^u, \mathbf{X}_q^u \in \mathcal{D}_U, p \neq q$.Similar to the labeled data, data augmentation operations will be performed on the unlabeled data.

$$\mathbf{X}_p^{S2mix} = \mathbf{X}_q^u \odot \widetilde{\mathbf{Y}}_p^u + \mathbf{X}_p^u \odot (1 - \widetilde{\mathbf{Y}}_p^u) \tag{4}$$

$\mathbf{X}_p^{S2mix}$ represent the mixup images after data augmentation.

In contrast to previous approaches where we discontinued the use of EMA between the teacher and student models, we instead initialize the pre-trained parameters separately for the teacher and student models. During the knowledge distillation process, we freeze the parameters of the teacher model θ and only train the student model ζ. Throughout the distillation process, we feed the background original images $\mathbf{X}_q^u$ into the teacher model and the mixup images $\mathbf{X}_p^{S2mix}$ into the student. $\mathcal{L}_{MSE}$ is mean squared error loss.

$$\min_{\zeta} \sum_{p=N+1}^{N+M} \mathcal{L}_{MSE}(f_{seg}(\mathbf{X}_p^{S2mix}; \zeta), f_{seg}(\mathbf{X}_q^u; \theta)) \tag{5}$$

3 Experiments and Results

Implementation Details and Evaluation Metrics In our study, the segmentation backbone network employed is V-net. Following [18],we divide them into 80 scans for training and 20 scans for validation,Our OMF utilizes the SGD optimizer, the learning rate (lr) initialization was set to 0.1 with a fixed seed on an NVIDIA 3090 GPU, and decayed by 64% every 2.5k iterations. To ensure uniformity in training data size, random patches of size $112 \times 112 \times 80$ were cropped during training as representatives. Both stage 1 and stage 2 training iterations were set to 15k.

During the testing phase, Our evaluation utilizes commonly employed metrics such as the Dice coefficient (Dice), Jaccard Index (Jaccard), 95% Hausdorff Distance (95HD) and Average Symmetric Surface Distance (ASD) to measure performance.we employ non-maximum suppression (NMS) as a post-processing step to eliminate isolated extraneous regions.

Compare with Sate-of-the-Art Methods. In the 80 training scans, we respectively utilize 10% and 20% (i.e., 8 and 16) of the scans as labeled data, with the remaining 72 and 64 scans serving as unlabeled data. Table 1 presents the segmentation performance of V-Net trained solely on labeled data (first two rows) and our semi-supervised approach (OMF) along with other SOTA methods on the testing dataset. Our approach outperforms others, achieving an average Dice of 88.14% and Jaccard of 79.10% (10%), and an average Dice of 88.59% and Jaccard of 79.76% (20%) when using only labeled training data (refer to Table 2). Leveraging unlabeled data, our semi-supervised framework further enhances segmentation performance, with Jaccard increasing to 90.23% and Dice to 82.34% (10%), and Jaccard increasing to 90.30% and Dice to 82.43% (20%).

Due to special training methods in the second stage, the trained student model possesses perceptual awareness of the overall shape of the segmentation targets. It can still make segmentation predictions based on the inertia of the shape even in the absence of missing parts of the image. We randomly mask square regions of varying proportions in the test images and to obtain segmentation accuracy as shown in Figure 3. By observing the predicted segmentation

Table 1. Comparison between our method and various methods on the LA dataset.

Method	#Scans used		Metrics			
	Labeled	Unlabeled	Dice[%]↑	Jaccard[%]↑	95HD[voxel]↓	ASD[voxel]↓
V-Net	8(10%)	0	82.74	71.72	13.35	3.26
V-Net	16(20%)	0	86.03	76.06	14.26	3.51
V-Net	80(All)	0	91.47	84.36	5.48	1.51
UA-MT [18]	8	72	87.79	78.39	8.68	2.12
SASSNet [6]	8	72	87.54	78.05	9.84	2.59
DTC [7]	8	72	87.51	78.17	8.23	2.36
SS-Net [15]	8	72	88.55	79.62	7.49	1.90
BCP [2]	8	72	89.62	81.31	6.81	1.76
OMF(Ours)	8	72	**90.23**	**82.34**	**5.95**	**1.63**
UA-MT [18]	16	64	88.88	80.21	7.32	2.26
SASSNet [6]	16	64	89.42	80.98	7.32	2.10
DTC [7]	16	64	89.42	80.98	7.32	2.10
SS-Net [15]	16	64	89.57	81.25	6.99	1.77
BCP [2]	16	64	88.69	79.96	8.46	2.09
OMF(Ours)	16	64	**90.30**	**82.43**	**6.52**	**1.65**

images, we find that even in the absence of parts of the image, we can still make accurate predictions, whereas other methods fail to predict missing parts. As the masking proportion continues to decrease, our method consistently outperforms others, suggesting that our method may outperform other methods at the edges of the predicted targets.

Table 2. Ablation studies of our OMF method on the LA dataset.

Method	#Scans used		Metrics			
	Labeled	Unlabeled	Dice[%]↑	Jaccard[%]↑	95HD[voxel]↓	ASD[voxel]↓
V-Net	8	0	82.74	71.72	13.35	3.26
S1	8	0	87.09	77.48	11.47	2.64
S1+orgloss	8	0	88.14	79.10	10.50	2.31
S1+S2	8	72	89.98	81.90	6.97	1.58
S1+orgloss+S2	8	72	**90.23**	**82.34**	**5.95**	**1.63**
V-Net	16	0	86.03	76.06	14.26	3.51
S1	16	0	88.07	79.37	8.50	1.96
S1+orgloss	16	0	88.59	79.76	9.63	2.24
S1+S2	16	64	89.60	81.30	6.78	1.97
S1+orgloss+S2	16	64	**90.30**	**82.43**	**6.52**	**1.65**

Ablation Studies. In the ablation experiments, as shown in Table 2, we validated the effectiveness of our data augmentation methods and the knowledge distillation approach using differentiated inputs. The notation "S1" represents

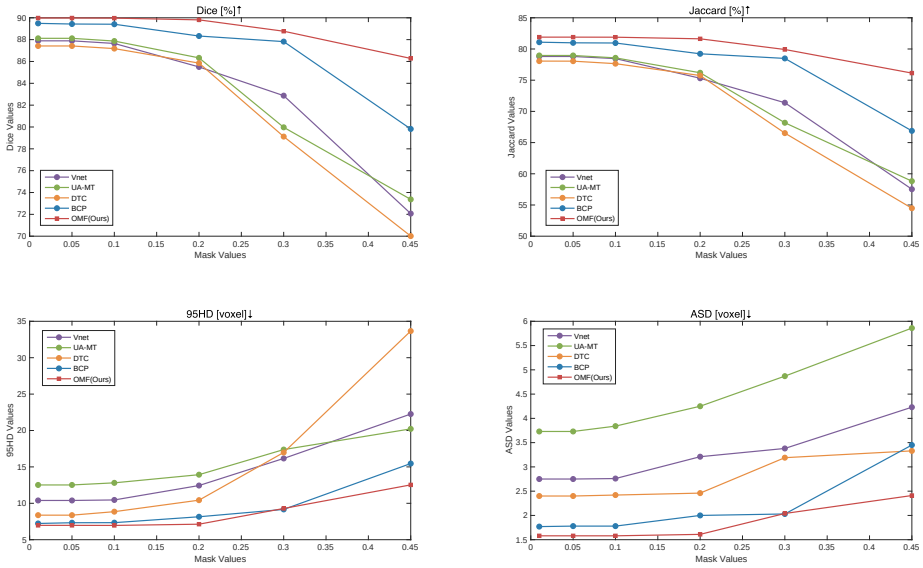

Fig. 3. Comparison of other methods under different missing proportions.

only the mixup images generated from our Overlay augmentation methods using labeled data are utilized in the process without using the original input images. Notably, "S1" in Table 2 shows that V-net using 10% labeled data with Overlay data augmentation methods achieves a 4.35% Dice improvement compared to using original images and labels (first row in Table 2), and 1.06% greater improvement than the regular method even with 20% labeled data (Vnet has a dice of 86.03 for the case of 16 labeled and 0 unlabeled). The term "orgloss" denotes the incorporation of original images in training. In stage 2 (S2), we introduced unlabeled data and the student model obtained through the differentiated input-based Mantle-Free knowledge distillation approach achieved state-of-the-art results. Additionally, we observed that incorporating the original image loss (orgloss) in stage 1 to enhance the accuracy of pre-trained models benefited the segmentation performance of the student model in stage 2.

4 Conclusion.

In this paper, we introduce a novel Overlay data augmentation method and integrate it into semi-supervised segmentation with Mantle-Free knowledge distillation, specifically targeting the Left Atrium task. By employing knowledge distillation with a fixed teacher model and a student model, we harness consistency loss and differentiated input strategies. This enables the student model to enhance its perception of segmentation targets' underlying shapes. Overall, our contribution lies in advancing semi-supervised medical image segmentation

methodologies, providing promising avenues for continued exploration and application in clinical settings.

Acknowledgments. This work was supported by the National Natural Science Foundation of China under Grant 62162065, Joint Special Project Research Foundation of Yunnan Province (202401BF070001-023) and Yunnan Fundamental Research Projects (202201AT070167).And thanks for the establishment of LA dataset [17].

Disclosure of Interests. The authors have no competing interests to declare that are relevant to the content of this article.

References

1. 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: MICCAI 2017. pp. 253–260. Springer (2017)
2. Bai, Y., Chen, D., Li, Q., Shen, W., Wang, Y.: Bidirectional copy-paste for semi-supervised medical image segmentation. In: CVPR. pp. 11514–11524 (2023)
3. Bortsova, G., Dubost, F., Hogeweg, L., Katramados, I., De Bruijne, M.: Semi-supervised medical image segmentation via learning consistency under transformations. In: MICCAI 2019. pp. 810–818. Springer (2019)
4. Fan, J., Gao, B., Jin, H., Jiang, L.: Ucc: Uncertainty guided cross-head co-training for semi-supervised semantic segmentation. In: CVPR. pp. 9947–9956 (2022)
5. Kendall, A., Gal, Y.: What uncertainties do we need in bayesian deep learning for computer vision? Advances in neural information processing systems **30** (2017)
6. Li, S., Zhang, C., He, X.: Shape-aware semi-supervised 3d semantic segmentation for medical images. In: MICCAI 2020. pp. 552–561. Springer (2020)
7. Luo, X., Chen, J., Song, T., Wang, G.: Semi-supervised medical image segmentation through dual-task consistency. In: AAAI. vol. 35, pp. 8801–8809 (2021)
8. Luo, X., Liao, W., Chen, J., Song, T., Chen, Y., Zhang, S., Chen, N., Wang, G., Zhang, S.: Efficient semi-supervised gross target volume of nasopharyngeal carcinoma segmentation via uncertainty rectified pyramid consistency. In: MICCAI 2021. pp. 318–329. Springer (2021)
9. Milletari, F., Navab, N., Ahmadi, S.A.: V-net: Fully convolutional neural networks for volumetric medical image segmentation. In: 2016 fourth international conference on 3D vision (3DV). pp. 565–571. Ieee (2016)
10. Nie, D., Gao, Y., Wang, L., Shen, D.: Asdnet: Attention based semi-supervised deep networks for medical image segmentation. In: MICCAI 2018. pp. 370–378. Springer (2018)
11. Olsson, V., Tranheden, W., Pinto, J., Svensson, L.: Classmix: Segmentation-based data augmentation for semi-supervised learning. In: CVPR. pp. 1369–1378 (2021)
12. Ouali, Y., Hudelot, C., Tami, M.: Semi-supervised semantic segmentation with cross-consistency training. In: CVPR. pp. 12674–12684 (2020)
13. 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)
14. Wang, Y., Wang, H., Shen, Y., Fei, J., Li, W., Jin, G., Wu, L., Zhao, R., Le, X.: Semi-supervised semantic segmentation using unreliable pseudo-labels. In: CVPR. pp. 4248–4257 (2022)

15. Wu, Y., Wu, Z., Wu, Q., Ge, Z., Cai, J.: Exploring smoothness and class-separation for semi-supervised medical image segmentation. In: CoRR. pp. 34–43. Springer (2022)
16. Wu, Y., Xu, M., Ge, Z., Cai, J., Zhang, L.: Semi-supervised left atrium segmentation with mutual consistency training. In: MICCAI 2021. pp. 297–306. Springer (2021)
17. 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)
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: MICCAI 2019. 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: CVPR. pp. 6023–6032 (2019)
20. Zheng, H., Lin, L., Hu, H., Zhang, Q., Chen, Q., Iwamoto, Y., Han, X., Chen, Y.W., Tong, R., Wu, J.: Semi-supervised segmentation of liver using adversarial learning with deep atlas prior. In: MICCAI 2019. pp. 148–156. Springer (2019)