DOUBLE NOISE MEAN TEACHER SELF-ENSEMBLING MODEL FOR SEMI-SUPERVISED TUMOR SEGMENTATION

Ke Zheng[1] *Junhai Xu*[1,⋆] *Jianguo Wei*[1]

[1]College of Intelligence and Computing, Tianjin University, Tianjin, China
{zhengke, jhxu}@tju.edu.cn

ABSTRACT

Accurate tumor segmentation of tumor images can assist doctors to diagnose diseases. However, achieving very high precision in tumor segmentation requires a large amount of annotated data, which is not easy for medical image data. In this paper, we present a novel double noise mean teacher self-ensembling model for semi-supervised 2D tumor segmentation. Concretely, the network is serialized by two groups of student-teacher networks. We design an auxiliary student-teacher module to learn the consistency regularity between the unlabeled image feature maps. In order to improve the robustness of the network, we add the random Gaussian noise to the student model every time the teacher model is updated. We test our model on the small cell lung tumor dataset and CVC-ClinicDB, and our model achieves the performance of nearly fully supervised segmentation. Moreover, the performance of our method outperforms the existing semi-supervised methods in four indicators.

Index Terms— Semi-supervised learning, consistency regularity, self-ensembling, tumor segmentation

1. INTRODUCTION

Accurate segmentation of tumor regions plays an important role in designing computer-aided diagnose or detection systems [1, 2]. Recently, the convolutional neural network based on the full supervision has been able to achieve very good performance in medical image segmentation field [3, 4]. However, an important factor of this advantage is that this algorithm requires a large amount of high-quality labeled training data, and high-quality labeled data requires a lot of time and energy. In order to utilize the limited labeled data and abundant unlabeled data, we focus on researching semi-supervised methods in the field of medical image segmentation.

Recently, a few semi-supervised methods have been proposed to utilize unlabeled data to improve the segmentation performance. Bai et al. [5] introduced an iterative framework that utilized the trained model to generate soft labels for the unlabeled data, and then added the soft labeled data to the training set to adjust the model. There are also many adversarial learning methods [6, 7] have been used in semi-supervised learning. Temporal ensembling [8] was proposed to complete semi-supervised learning, training the network by minimizing the loss between the source unlabeled data and unlabeled data with Gaussian noise added. Tarvainen et al. [9] proposed a mean teacher framework to make full use of the unlabeled data, where teacher model parameters are obtained by the moving average of parameters of the student model, and many subsequent methods [10, 11, 12] use this network as the backbone network. Li et al. [11] introduced more data perturbations and model perturbations on the basis of mean teacher to build the consistency of the same input under different perturbations. Yu et al. [12] added uncertainty-aware to the mean teacher to make the model learn more reliable goals. Besides, multi-task network has been used in semi-supervised learning [13, 14]. Li et al. [13] introduced a shape-aware semi-supervised method to enforce a geometric shape constraint on the segmentation output. Luo et al. [14] utilized a multi-task network to build the consistency from the task level for semi-supervised learning.

Although many methods have tried to use consistency of the same input under different perturbations, these methods did not make full use of the consistency relationship between the feature maps under different disturbances and the parameter disturbances between the network layers.

In this work, we propose a novel double noise mean teacher self-ensembling model for tumor segmentation. We design an auxiliary convolutional network, which is trained by labeled feature map and ground truth. Then we use the unlabeled feature map to pass through the network. The difference between the output of the student model and teacher model is used as a part of the unsupervised loss to train the main module. Moreover, self-ensembling in mean teacher can also be regarded as a kind of regularization, so we also add disturbances to the two mean teacher networks to improve the performance of the network. Compared with other methods, our method can utilize segmentation results and feature maps at the same time, and it can also enhance the perturbation in the mean teacher model. The experiments show that our method can get a better segmentation result compared with other existing semi-supervised methods.

2. PROPOSED METHODOLOGY

In this section, we introduce the proposed method in detail. The pipeline of our double noise mean teacher framework is shown in Fig. 1. We firstly introduce the general semi-supervised learning task to better understand the proposed method. We then present our auxiliary convolutional module. Lastly, we introduce the Gaussian noise disturbance in the mean teacher framework.

2.1. Temporal ensembling model for segmentation

In our problem setting, we are given $N + M$ training samples with N labeled and M unlabeled samples. We denote the labeled sets as $D_l = \{(x_i, y_i)\}_{i=1}^{N}$ and unlabeled set as $D_u = \{x_i\}_{i=N+1}^{N+M}$, where $x_i \in \mathbb{R}^{H\times W}$ is the input image and $y_i \in \mathbb{R}^{H\times W}$ is the segmentation label of x_i, H and W are the image height and width, respectively. The goal of our semi-supervised segmentation framework is to minimize the following combined objective function:

$$\mathcal{L}_{total} = \mathcal{L}_s + \lambda\mathcal{L}_u \tag{1}$$

Thanks to the National Natural Science Foundation of China (No. 62176181) and China Postdoctoral Science Foundation (No. 2020M680905).

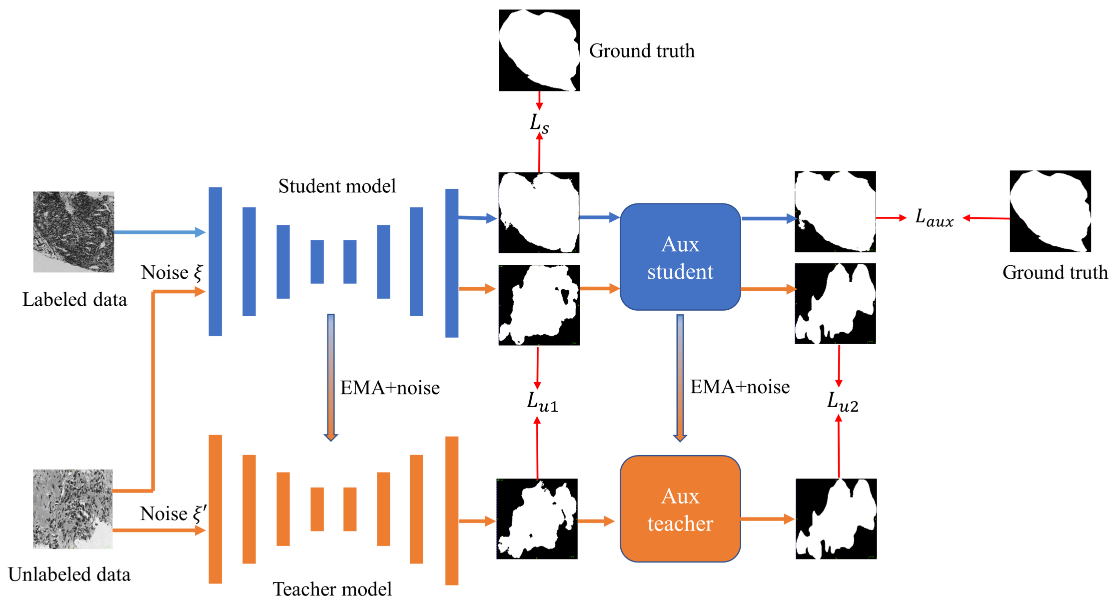


Fig. 1: Overview of the proposed double noise mean teacher self-ensembling (DNMT) network for semi-supervised medical image segmentation. The network is composed of two groups of student-teacher networks. The first group of student-teacher networks use UNet [15] as the backbone, and the second group of student-teacher networks consists of a small number of convolution operations. In addition, we have added Gaussian noise to the EMA calculation formula [9] to obtain a more disturbed teacher network, which increased the disturbance at the network level and improved the performance of networks.

where $\mathcal{L}_s$ and $\mathcal{L}_u$ are the supervised loss and unsupervised loss, respectively. $\mathcal{L}_s$ denotes the supervised loss to evaluate the quality of the network output on labeled inputs, which can be written as:

$$\mathcal{L}_s = \sum_{i \in D_l} \mathcal{L}_{seg}(f_{seg}(x_i; \theta_{seg}), y_i) \tag{2}$$

where we use the cross-entropy loss as supervised loss $\mathcal{L}_{seg}$.

The unsupervised loss in the main module is $\mathcal{L}_{u_1}$, as shown in the following formula:

$$\mathcal{L}_{u_1} = \sum_{i \in D_u} \mathcal{L}_{cons}(f_{seg}(x_i; \theta'_{seg}, \xi'_{seg}), f_{seg}(x_i; \theta_{seg}, \xi_{seg})) \tag{3}$$

where $\mathcal{L}_{u_1}$ measures the consistency of the teacher model and student model in the main module for the same input x_i under different disturbances. Here, $f(\cdot)$ denotes the segmentation neural networks; (θ', ξ') and (θ, ξ) represents the weights of model and different perturbation operations(e.g., adding noise to input data, network dropout) of the teacher and student models, respectively. The subscript seg represents the student and teacher network in the main module. λ is a ramp-up weighting coefficient that controls the trade-off between the supervised and unsupervised loss.

In the mean teacher framework, the teacher model is an exponential moving average (EMA) of the student model [9]. Some studies [8, 9] have shown that ensembling predictions can improve the results, so we assemble the information in different training steps through updating the teacher weights θ' as an EMA of the student's weights θ.

2.2. Auxiliary convolutional module

If only the main module is used to learn the consistency regularity, the predicted targets from the teacher model may be unreliable and noisy. One latest study shows that in semi-supervised learning, compared with only using the segmentation result to calculate the consistency loss, adding the consistency loss between the feature maps allows the network to learn diversified information, thereby improving the network performance [16]. Therefore, in order to learn the consistency regularity between the unlabeled image feature maps, we design an auxiliary module consisting of a student network and a teacher network, which is trained by ground truth of the labeled image and the image feature map. Like the main module, it is also a mean teacher framework, and the teacher network is an exponential moving average (EMA) of the student network [9]. The student network and teacher network have the same structure, which both consist of 4 layers of convolution, and does not change the image size.

Specifically, the model is optimized by minimizing the following formula:

$$\mathcal{L}_{aux} = \sum_{i \in D_l} \mathcal{L}_{aux}(f_{aux}(\eta_i; \theta_{aux}), y_i) \tag{4}$$

where the subscript aux represents the student and teacher network in the auxiliary module; η_i is the feature map generated by the labeled image through the main module. We use labeled images to train the auxiliary student network, and the auxiliary teacher network ensembles parameters of the student network through EMA. Then the feature map of the unlabeled image is used as the input of the auxiliary module, and the difference between the feature maps can be obtained by this module, the formula is as follows:

$$\mathcal{L}_{u_2} = \sum_{i \in D_u} \mathcal{L}_{cons}(f_{aux}(\gamma_i; \theta'_{aux}, \eta'_{aux}), f_{aux}(\gamma_i; \theta_{aux}, \eta_{aux})) \tag{5}$$

where $\mathcal{L}_{u_2}$ measures the consistency of the teacher model and student model in the auxiliary module for the same feature map γ_i under different disturbances. Finally, we make $\mathcal{L}_{u_2}$ as part of the unsupervised loss of the main module to jointly guide the training network, the total unsupervised loss can be written as:

$$\mathcal{L}_u = 0.5 \times (\mathcal{L}_{u_1} + \mathcal{L}_{u_2}) \tag{6}$$

because the original loss is as important as the regular loss of the feature map, we take half of the sum of the two losses as the total unsupervised loss.

2.3. Noise in the network level

In the semi-supervised learning method based on the consistency regularity, the disturbance of the data or the network is very important, because these methods [9, 11, 17, 18] encourage the segmentation predictions to be consistent under different perturbations for the same input.

There are many disturbances to unlabeled data in existing methods, including flip, rotate, rescale, noise, and so on [11, 19]. In addition to dropout, the parameter difference between the student and teacher model is also a kind of network disturbance. The student-teacher framework is a self-ensembling model, which can ensemble previously information through *exponential moving average* (EMA).

In order to make full use of disturbances of networks, we design a new method of generating teacher network parameters. The existing data perturbation method adds noise to the input data, and we add random noise to the parameter calculations of the teacher model on the main module and auxiliary module. Specifically, existing methods use Gaussian noise to disturb the data, so we add random Gaussian noise to the student model every time the teacher model is updated. Parameters of the teacher model can be calculated by the following function:

$$\theta' = \alpha\theta'_{t-1} + (1-\alpha)(\theta_t + noise) \tag{7}$$

where θ_t and θ'_t are student's weights and teacher's weights at training step t; α is the EMA decay that controls the updating rate and $noise$ is Gaussian noise added to student network parameters.

3. EXPERIMENTS AND RESULTS

3.1. Dataset and evaluation metrics

We evaluated the proposed method on the small cell lung cancer (SCLC) dataset and CVC-ClinicDB [20].

For SCLC dataset, we performed H&E staining of lung tissue sections from 31 patients using a digital slide scanner (Hamamatsu Nano Zoomer Digital Pathology) with an objective magnification of 20 times. The SCLC was morphologically diagnosed according to the criteria of World Health Organization (WHO) classification of Tumors of the Lung, Pleura, Thymus and Heart (the fourth Edition). The area containing tumor cells was circled by one thoracic pathologist. Since the image is too huge, it is impossible to input the original image into the network during segmentation, so it must be cut into patches one by one. We intercepted each annotation according to the annotation information and resized it to 1024 x 1024 size. Then we got 134 patches. Next, we used data enhancement methods such as mirroring, flipping, rotating and offsetting to expand the dataset. We finally got 650 patches. The dataset was randomly split into 630 cases for training, 20 cases for testing. For training images, only 63(i.e., 10%) were used as labeled and the remaining 567 were used as unlabeled. We normalized each image to zero mean and unit variance as pre-processing.

Table 1: Quantitative comparison between our method and other semi-supervised methods on the SCLC dataset.

Method	Labeled/Unlabeled	DSC	JI	95HD	ASD
UNet [15]	630/0	90.81	83.30	131.68	34.04
UNet [15]	63/0	84.51	74.57	150.11	54.68
MT [9]	63/567	84.82	75.13	160.55	56.56
DAN [6]	63/567	85.18	75.70	150.58	54.50
EM [7]	63/567	85.20	75.52	156.82	56.30
UAMT [12]	63/567	84.88	75.18	157.89	56.61
ICT [22]	63/567	85.19	75.39	154.57	53.87
URPC [23]	63/567	85.36	75.74	156.72	54.45
Ours	63/567	85.29	75.92	145.86	55.27
Ours(noise)	63/567	**85.77**	**76.56**	**144.05**	**53.65**

Table 2: Quantitative comparison between our method and other semi-supervised methods on CVC-ClinicDB.

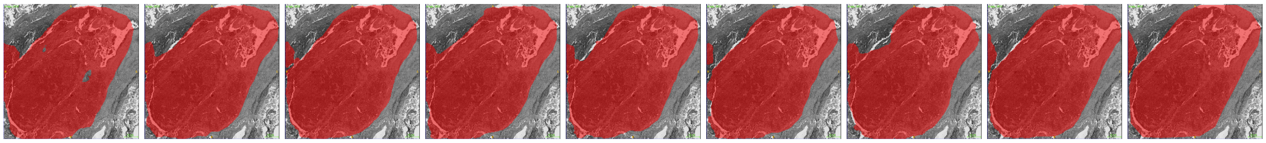
Method	Labeled/Unlabeled	DSC	JI	95HD	ASD
UNet[15]	490/0	84.75	76.59	32.58	11.54
UNet[15]	49/0	55.94	44.68	101.57	39.94
MT [9]	49/441	57.84	45.64	105.55	40.52
DAN [6]	49/441	57.07	46.23	89.53	36.76
EM [7]	49/441	57.06	45.38	100.86	41.98
UAMT [12]	49/441	57.20	45.51	94.52	38.43
ICT [22]	49/441	56.91	45.45	91.74	36.13
URPC [23]	49/441	57.66	47.49	**85.42**	34.05
Ours	49/441	58.50	47.11	98.77	40.52
Ours(noise)	49/441	**58.70**	**47.60**	94.47	**34.01**

We also test the proposed method on the public dataset CVC-ClinicDB. CVC-ClinicDB is an open-access dataset of 612 images with a resolution of 384x288 from 31 colonoscopy sequences. It is used for medical image segmentation, in particular polyp detection in colonoscopy videos. In our experiments, the dataset was randomly split into 490 cases for training, 61 cases for validation and 61 cases for testing [21].

Four evaluation metrics were employed to evaluate the segmentation results, including dice similarity coefficient (DSC), region-based jaccard index (JI), boundary-based 95% hausdorff distance (HD95), average surface distance metric (ASD).

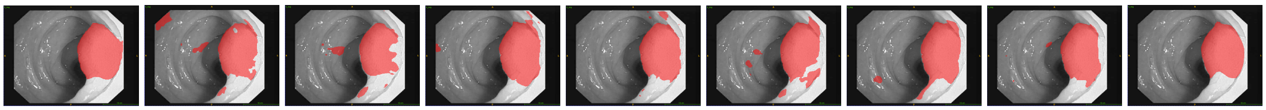
3.2. Implementation details

The framework was implemented in PyTorch, using a TITAN X GPU. UNet [15] is utilized as backbone for both student and teacher model in the main module. We used the SGD optimizer to update the network parameters, the weight decay was set 0.0001 and momentum was set 0.9 and without pre-trained weights. The initial learning rate was set as 0.01 and divided by 10 every 2500 iterations. We totally trained 10000 iterations. The batch size was 4, consisting of 2 annotated images and 2 unannotated images. We used the standard data augmentation techniques on-the-fly to avoid overfitting [24], including randomly flipping and rotating. The auxiliary module uses 3×3 kernels in its convolutional layers and an SGD optimizer. In our experiments, we set the EMA decay rate $a = 0.99$ and unsupervised weight factor λ as a time-dependent Gaussian warming-up function $\lambda(t) = 0.001 * e^{-5(1-\frac{t}{t_{max}})^2}$, where t and t_{max} indicated the current and last training step.



(a) UNet (b) MT (c) DAN (d) EM (e) UAMT (f) ICT (g) URPC (h) Ours (i) GT

Fig. 2: Visual comparison of small cell lung cancer segmentation results, where the red labels indicate predictions.



(a) UNet (b) MT (c) DAN (d) EM (e) UAMT (f) ICT (g) URPC (h) Ours (i) GT

Fig. 3: Visual comparison of CVC-ClinicDB, where the red labels indicate predictions.

3.3. Comparison between our method and other methods

As shown in Table 1, we implemented several state-of-the-art semi-supervised segmentation methods for comparison, including MT [9], DAN [6], EM [7], UAMT [12], ICT [22], URPC [23]. Moreover, UNet [15] using all the annotated images gives the performance upper bound. For a fair comparison, all these methods were implemented by using UNet as the backbone. Compared with fully supervised UNet trained with only 63 labeled data, the performance of all semi-supervised methods have improved significantly. Notably, our method achieves the best performance over the state-of-the-art semi-supervised methods. $Ours$ indicates that we add the auxiliary module to the baseline (mean teacher), and $Ours(noise)$ indicates that we add EMA noise to the main module and auxiliary module. We make 1.0%, 0.6%, 0.5%, 0.9%, 0.6%, 0.4% improvements on dice compared with [9, 6, 7, 12, 22, 23] respectively.

Table 2 shows the performance of our method and other state-of-the-art methods under 10% labeled data on CVC-ClinicDB. It can be found that compared with supervised-only under 10% labeled data settings, all semi-supervised methods improve the performance of segmentation. Notably, our method achieves the best performance over the state-of-the-art semi-supervised methods on three metrics. Specifically, DSC has increased by 0.86%, 1.63%, 1.64%, 1.5%, 1.79%, 1.04% compared with MT, DAN, EM, UAMT, ICT, URPC respectively by fully exploiting 10% labeled data and unlabeled data.

To intuitively demonstrate the difference, the segmentation results are visualized in Fig. 2 and Fig. 3. Compared with other methods, our results have a higher overlap ratio with the ground truth and clearer boundaries on two datasets.

3.4. Analysis and ablation study

In order to explore the effect of EMA noise on the entire network, we have done experiments to add EMA noise to some existing semi-supervised image segmentation methods that use the mean teacher network as the backbone. As shown in Table 3, after adding EMA noise to MT [9], UAMT [12] and our method with no EMA noise, these networks improve by 0.3%, 0.5%, 0.5% dice, respectively. Experiments show that our DNMT framework has the full capability to draw out the rich information from the unlabeled data.

In Addition, we design the ablation experiments on SCLC dataset. We design two methods to use only the EMA noise and the auxiliary module. In the method with auxiliary module, only L_{u1} or L_{u2} is used as unsupervised loss. From the Table 4, we can see the performance of Ours(Noise) and Ours(Aux) are higher than the MT, and the performance of Ours(Aux) has a better performance when using the sum of two unsupervised losses. Finally, we can find that when the EMA noise and the auxiliary module are used at the same time, the network performs best, indicating that the two parts work together to improve the performance of the network.

Table 3: Quantitative analysis of our EMA noise on the small cell lung cancer dataset.

Method	Labeled/Unlabeled	DSC	JI	95HD	ASD
MT [9]	63/567	84.82	75.13	160.55	56.56
MT(noise)	63/567	**85.09**	**75.65**	**154.22**	**56.35**
UAMT [12]	63/567	84.88	75.18	157.89	56.61
UAMT(noise)	63/567	**85.40**	**75.90**	**154.12**	**56.00**
Ours	63/567	85.29	75.92	145.86	55.27
Ours(noise)	63/567	**85.77**	**76.56**	**144.05**	**53.65**

Table 4: Ablation study under 10% labeled data on the small cell lung cancer dataset.

Method	Labeled/Unlabeled	DSC	JI	95HD	ASD
MT [9]	63/567	84.82	75.13	160.55	56.56
Ours(Noise)	63/567	85.09	75.65	154.22	56.35
Ours(Aux)(L_{u1})	63/567	85.47	76.11	151.12	55.71
Ours(Aux)(L_{u2})	63/567	85.49	76.19	152.81	56.22
Ours(Aux)($L_{u1} + L_{u2}$)	63/567	85.53	76.10	144.51	55.26
Ours(Noise+Aux)	63/567	**85.77**	**76.56**	**144.05**	**53.65**

4. CONCLUSION

In this paper, we have proposed a novel semi-supervised segmentation framework DNMT for medical image segmentation. An auxiliary module is designed to make full use of information of the image feature map to improve the model performance. Moreover, we add the noise in EMA to increase the disturbance at the network level. Extensive experiments on two datasets demonstrate the superiority of our method over other state-of-the-art semi-supervised learning methods. Analysis and ablation experiments also show the effectiveness of each part. In the future, we will evaluate our framework on other semi-supervised tasks.

5. REFERENCES

[1] Bram Van Ginneken, Cornelia M Schaefer-Prokop, and Mathias Prokop, "Computer-aided diagnosis: how to move from the laboratory to the clinic," *Radiology*, vol. 261, no. 3, pp. 719–732, 2011.

[2] Geert Litjens, Thijs Kooi, Babak Ehteshami Bejnordi, Arnaud Arindra Adiyoso Setio, Francesco Ciompi, Mohsen Ghafoorian, Jeroen Awm Van Der Laak, Bram Van Ginneken, and Clara I Sánchez, "A survey on deep learning in medical image analysis," *Medical image analysis*, vol. 42, pp. 60–88, 2017.

[3] Fausto Milletari, Nassir Navab, and Seyed-Ahmad Ahmadi, "V-net: Fully convolutional neural networks for volumetric medical image segmentation," in *2016 fourth international conference on 3D vision (3DV)*, 2016, pp. 565–571.

[4] Özgün Çiçek, Ahmed Abdulkadir, Soeren S Lienkamp, Thomas Brox, and Olaf Ronneberger, "3d u-net: learning dense volumetric segmentation from sparse annotation," in *International conference on medical image computing and computer-assisted intervention*, 2016, pp. 424–432.

[5] Wenjia Bai, Ozan Oktay, Matthew Sinclair, Hideaki Suzuki, Martin Rajchl, Giacomo Tarroni, Ben Glocker, Andrew King, Paul M. Matthews, and Daniel Rueckert, "Semi-supervised learning for network-based cardiac mr image segmentation," in *Medical Image Computing and Computer-Assisted Intervention – MICCAI 2017*, 2017, pp. 253–260.

[6] Yizhe Zhang, Lin Yang, Jianxu Chen, Maridel Fredericksen, David P. Hughes, and Danny Z. Chen, "Deep adversarial networks for biomedical image segmentation utilizing unannotated images," in *Medical Image Computing and Computer Assisted Intervention – MICCAI 2017*, 2017, pp. 408–416.

[7] Tuan-Hung Vu, Himalaya Jain, Maxime Bucher, Matthieu Cord, and Patrick Perez, "Advent: Adversarial entropy minimization for domain adaptation in semantic segmentation," in *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, 2019.

[8] Samuli Laine and Timo Aila, "Temporal ensembling for semi-supervised learning," in *The International Conference on Learning Representations (ICLR)*, 2017, pp. 1–13.

[9] Antti Tarvainen and Harri Valpola, "Mean teachers are better role models: Weight-averaged consistency targets improve semi-supervised deep learning results," in *Advances in Neural Information Processing Systems*, 2017.

[10] Zhiqiang Xie, Enmei Tu, Hao Zheng, Yun Gu, and Jie Yang, "Semi-supervised skin lesion segmentation with learning model confidence," in *ICASSP 2021 - 2021 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, 2021, pp. 1135–1139.

[11] Xiaomeng Li, Lequan Yu, Hao Chen, Chi-Wing Fu, Lei Xing, and Pheng-Ann Heng, "Transformation-consistent self-ensembling model for semisupervised medical image segmentation," *IEEE Transactions on Neural Networks and Learning Systems*, pp. 523–534, 2021.

[12] Lequan Yu, Shujun Wang, Xiaomeng Li, Chi-Wing Fu, and Pheng-Ann Heng, "Uncertainty-aware self-ensembling model for semi-supervised 3d left atrium segmentation," in *Medical Image Computing and Computer Assisted Intervention – MICCAI 2019*, 2019, pp. 605–613.

[13] Shuailin Li, Chuyu Zhang, and Xuming He, "Shape-aware semi-supervised 3d semantic segmentation for medical images," in *Medical Image Computing and Computer Assisted Intervention – MICCAI 2020*, 2020, pp. 552–561.

[14] Xiangde Luo, Jieneng Chen, Tao Song, and Guotai Wang, "Semi-supervised medical image segmentation through dual-task consistency," in *Proceedings of the AAAI Conference on Artificial Intelligence*, 2021, pp. 8801–8809.

[15] Olaf Ronneberger, Philipp Fischer, and Thomas Brox, "U-net: Convolutional networks for biomedical image segmentation," in *Medical Image Computing and Computer-Assisted Intervention – MICCAI 2015*, 2015, pp. 234–241.

[16] Shumeng Li, Ziyuan Zhao, Kaixin Xu, Zeng Zeng, and Cuntai Guan, "Hierarchical consistency regularized mean teacher for semi-supervised 3d left atrium segmentation," *CoRR*, vol. abs/2105.10369, 2021.

[17] Mehdi Sajjadi, Mehran Javanmardi, and Tolga Tasdizen, "Regularization with stochastic transformations and perturbations for deep semi-supervised learning," in *Advances in Neural Information Processing Systems*, 2016, pp. 1163–1171.

[18] Jisoo Jeong, Seungeui Lee, Jeesoo Kim, and Nojun Kwak, "Consistency-based semi-supervised learning for object detection," in *Advances in Neural Information Processing Systems*, 2019, pp. 10759–10768.

[19] Xiaomeng Li, Lequan Yu, Hao Chen, Chi-Wing Fu, and Pheng-Ann Heng, "Semi-supervised skin lesion segmentation via transformation consistent self-ensembling model," in *British Machine Vision Conference (BMVC)*, 2018, pp. 1–12.

[20] Jorge Bernal, F Javier Sánchez, Gloria Fernández-Esparrach, Debora Gil, Cristina Rodríguez, and Fernando Vilariño, "Wm-dova maps for accurate polyp highlighting in colonoscopy: Validation vs. saliency maps from physicians," *Computerized medical imaging and graphics*, vol. 43, pp. 99—111, 2015.

[21] D. Jha, M. A. Riegler, D. Johansen, P. Halvorsen, and H. D. Johansen, "Double u-net: A deep convolutional neural network for medical image segmentation," in *2020 IEEE 33rd International Symposium on Computer-Based Medical Systems (CBMS)*, 2020, pp. 558–564.

[22] Vikas Verma, Alex Lamb, Juho Kannala, Yoshua Bengio, and David Lopez-Paz, "Interpolation consistency training for semi-supervised learning," in *Proceedings of the Twenty-Eighth International Joint Conference on Artificial Intelligence, IJCAI-19*, 2019, pp. 3635–3641.

[23] Xiangde Luo, Wenjun Liao, Jieneng Chen, Tao Song, Yinan Chen, Guotai Wang, and Shaoting Zhang, "Semi-supervised segmentation via uncertainty rectified pyramid consistency and its application to gross target volume of nasopharyngeal carcinoma," *CoRR*, vol. abs/2012.07042, 2020.

[24] Lequan Yu, Jie-Zhi Cheng, Qi Dou, Xin Yang, Hao Chen, Jing Qin, and Pheng-Ann Heng, "Automatic 3d cardiovascular mr segmentation with densely-connected volumetric convnets," in *Medical Image Computing and Computer-Assisted Intervention – MICCAI 2017*, 2017, pp. 287–295.