Tripled-Uncertainty Guided Mean Teacher Model for Semi-supervised Medical Image Segmentation

Kaiping Wang[1], Bo Zhan[1], Chen Zu[2], Xi Wu[3], Jiliu Zhou[1,3], Luping Zhou[4], and Yan Wang[1(✉)]

[1] School of Computer Science, Sichuan University, Chengdu, China
[2] Department of Risk Controlling Research, JD.COM, Chengdu, China
[3] School of Computer Science, Chengdu University of Information Technology, Chengdu, China
[4] School of Electrical and Information Engineering, University of Sydney, Sydney, Australia

Abstract. Due to the difficulty in accessing a large amount of labeled data, semi-supervised learning is becoming an attractive solution in medical image segmentation. To make use of unlabeled data, current popular semi-supervised methods (e.g., temporal ensembling, mean teacher) mainly impose data-level and model-level consistency on unlabeled data. In this paper, we argue that in addition to these strategies, we could further utilize auxiliary tasks and consider task-level consistency to better leverage unlabeled data for segmentation. Specifically, we introduce two auxiliary tasks, i.e., a foreground and background reconstruction task for capturing semantic information and a signed distance field (SDF) prediction task for imposing shape constraint, and explore the mutual promotion effect between the two auxiliary and the segmentation tasks based on mean teacher architecture. Moreover, to handle the potential bias of the teacher model caused by annotation scarcity, we develop a tripled-uncertainty guided framework to encourage the three tasks in the teacher model to generate more reliable pseudo labels. When calculating uncertainty, we propose an uncertainty weighted integration (UWI) strategy for yielding the segmentation predictions of the teacher. Extensive experiments on public 2017 ACDC dataset and PROMISE12 dataset have demostrated the effectiveness of our method. Code is available at https://github.com/DeepMedLab/Tri-U-MT.

Keywords: Semi-supervised segmentation · Mean teacher · Multi-task learning · Tripled-uncertainty

1 Introduction

Segmentation is a basic yet essential task in the realm of medical image processing and analysis. To enable clinical efficiency, recent deep learning frameworks [1, 2] have made a quantum leap in automatic segmentation with sufficient annotations. However, such

K. Wang and B. Zhan—The authors contribute equally to this work.

M. de Bruijne et al. (Eds.): MICCAI 2021, LNCS 12902, pp. 450–460, 2021.
https://doi.org/10.1007/978-3-030-87196-3_42

annotations are hard to acquire in real world due to their expensive and time-consuming nature. To alleviate annotation scarcity, a feasible approach is to take semi-supervised learning [3, 4] which leverages both labeled and unlabeled data to train the deep network effectively.

Considerable efforts have been devoted to the semi-supervised segmentation community, which can be broadly categorized into two groups. The first group refers to those methods trying to predict pseudo labels on unlabeled images and mixing them with ground truth labels to provide additional training information [5–7]. However, the segmentation result of such self-training based method is susceptible due to the uneven quality of the predicted pseudo labels. The second group of semi-supervised segmentation methods lies in the consistency regularization, that is, encouraging the segmentation predictions to be consistent under different perturbations for the same input. A typical example is Π-model [8] which minimizes the distance between the results of two forward passes with different regularization strategies. To improve the stability, a temporal ensembling model [8] is further proposed based on Π-model, which aggregates the exponential moving average (EMA) predictions and encourages the consensus between the ensembled predictions and current predictions for unlabeled data. To accelerate the training and enable the online learning, mean teacher [9] further improves the temporal ensembling model by enforcing prediction consistency between the current training model (i.e., the student model) and the corresponding EMA model (i.e., the teacher model) in each training step. Nevertheless, unreliable results from the teacher model may mislead the student model, thus deteriorating the whole training. Researchers therefore incorporated uncertainty map to the mean teacher model, forcing the student to learn high confidence predictions from the teacher model [10, 11]. Our research falls in the second group of semi-supervised approaches.

On the other hand, different from the above semi-supervised segmentation methods which mainly focus on the consistency under the disturbances at data level or model level, there are also research works [12–14] that explore to improve segmentation from another perspective: multi-task learning. These methods jointly train multiple tasks to boost segmentation performance. For instance, Chen et al. [13] applied a reconstruction task to assist the segmentation of medical images. Li et al. [14] proposed a shape-aware semi-supervised model by incorporating a signed distance map generation task to enforce a shape constraint on the segmentation result. By exploring the relationship between the main and the auxiliary tasks, the learned segmentation model could bypass the over-fitting problem and learn more general representations.

In this paper, inspired by the success of semi-supervised learning and multi-task learning, we propose a novel end-to-end semi-supervised mean teacher model guided by tripled-uncertainty maps from three highly related tasks. Concretely, apart from the segmentation task, we bring in two additional auxiliary tasks, i.e., a foreground and background reconstruction task and a signed distance field (SDF) prediction task. The reconstruction task can help the segmentation network capture more semantic information, while the predicted SDF describes the signed distance of a corresponding pixel to its closest boundary point after normalization, thereby constraining the global geometric shape of the segmentation result. Following the spirit of mean teacher architecture, we

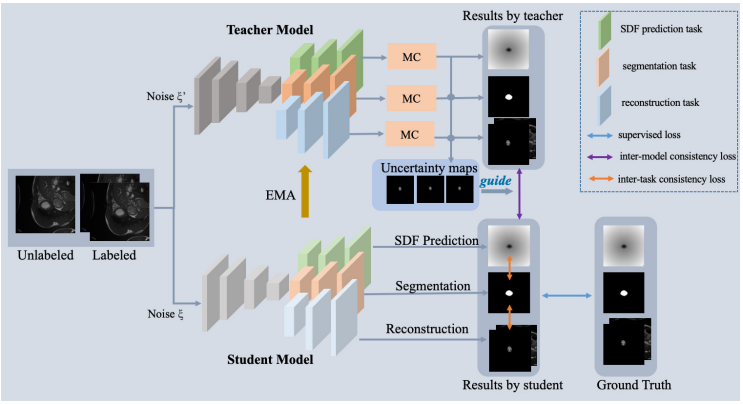

Fig. 1. Overview of our tripled-uncertainty guided mean teacher model.

build a teacher model and a student model, each targeting at all three tasks above. Additionally, to tackle the unreliability and noise in teacher model predictions, we impose uncertainty constraints on all the three tasks, expecting the student model can learn as accurate information as possible.

Our main contributions are three-fold: (1) We inject the spirit of multi-task learning into mean teacher architecture, so that the segmentation task could also benefit from the enhanced semantic and geometric shape information by extracting the correlations among the segmentation task, the reconstruction task, and the SDF prediction task. In this manner, our mean teacher model simultaneously takes account of the data-, model- and task-level consistency to better leverage unlabeled data for segmentation. (2) We impose the uncertainty estimation on all tasks and develop a tripled-uncertainty to guide the student model to learn more reliable predictions from the teacher model. (3) Current approaches tend to generate uncertainty maps by averaging the results from multiple Monte Carlo (MC) samplings, neglecting the discrepancy of different results. In contrast, we propose an uncertainty weighted integration (UWI) strategy to assign different weights for different sampling results, generating a more accurate segmentation prediction.

2 Methodology

An overview of our proposed network is illustrated in Fig. 1, consisting of a teacher model and a student model, following the idea of mean teacher. The student model is the trained model, and it assigns the exponential moving average of its weights to the teacher model at each step of training. On the other hand, the predictions of the teacher model would be viewed as additional supervisions for the student model to learn. These two models share a same encoder-decoder structure, where the encoder is shared among different tasks while the decoders are task-specific. In our problem setting, we are given a training set containing N labeled data and M unlabeled data samples, where $N \ll M$. The labeled set is defined as $\mathcal{D}^l = \left\{\mathbf{X}^i, \mathbf{Y}^i\right\}_{i=1}^{N}$ and the unlabeled set as $\mathcal{D}^u = \left\{\mathbf{X}^i\right\}_{i=N+1}^{N+M}$, where

$\mathbf{X}^i \in \mathbb{R}^{H\times W}$ is the intensity image, $\mathbf{Y}^i \in \{0,1\}^{H\times W}$ is the corresponding segmentation label. For labeled data, we can also obtain the ground truth of SDF $\mathbf{Z}^i \in \mathbb{R}^{H\times W}$ from $\mathbf{Y}^i$ via the SDF function in [17] for the SDF prediction task. Similarly, for the reconstruction task, the ground truths of foreground and background $\mathbf{G}^i \in \mathbb{R}^{2\times H\times W}$ can be obtained by $\mathbf{Y}^i \odot \mathbf{X}^i$ and $\left(1-\mathbf{Y}^i\right)\odot \mathbf{X}^i$ where $\odot$ refers to element-wise multiplication. Given $\mathcal{D}^l$ and $\mathcal{D}^u$, the student model is optimized by minimizing 1) the supervised segmentation loss $\mathcal{L}_{\mathrm{s}}$ on labeled data $\mathcal{D}^l$, 2) the inter-model consistency loss $\mathcal{L}_{\mathrm{cons}}^{\mathrm{model}}$ between the student model and teacher model on both $\mathcal{D}^l$ and $\mathcal{D}^u$, and 3) the inter-task consistency loss $\mathcal{L}_{\mathrm{cons}}^{\mathrm{task}}$ among different tasks on $\mathcal{D}^l$ and $\mathcal{D}^u$. On the other hand, the tripled-uncertainty maps with respect to the three tasks generated by the teacher model guide the student model to learn more reliable predictions from the teacher model. Moreover, to enhance the robustness of our mean teacher model, different perturbations ξ and ξ' are fed into the student and teacher model, respectively. More details will be introduced in subsequent sections.

2.1 Student Model

Our student model employs U-net [2] as backbone with respect to three tasks, i.e., segmentation task, reconstruction task and SDF prediction task. Note that the encoder is shared by the three tasks with the same parameters while the parameters in the task-specific decoders are different to fit the different tasks. In this manner, the encoder is forced to capture features related to the semantic information and geometric shape information, leading to a low disparity between the output segmentation result and the ground truth. Concretely, the encoder consists of three down-sampling blocks with 3×3 convolutional layers while the decoders take similar structures with three up-sampling blocks but differ in the last task head layers. The segmentation head is equipped with softmax activation while the reconstruction head and the SDF head utilize sigmoid activation. Following [13], we drop the skip connection in the reconstruction task. Given an input image $\mathbf{X}^i \in \mathcal{D}^l$, these tasks can generate the segmentation result $\tilde{\mathbf{Y}}_S^i$, the reconstruction result $\tilde{\mathbf{G}}_S^i$, and the SDF result $\tilde{\mathbf{Z}}_S^i$, as follows:

$$\tilde{\mathbf{Y}}_S^i = f_{seg}\left(\mathbf{X}^i;\theta_{seg},\xi\right), \tilde{\mathbf{G}}_S^i = f_{rec}\left(\mathbf{X}^i;\theta_{rec},\xi\right), \tilde{\mathbf{Z}}_S^i = f_{sdf}\left(\mathbf{X}^i;\theta_{sdf},\xi\right), \tag{1}$$

where $f_{seg}, f_{rec}, f_{sdf}$ represent the segmentation network, the reconstruction network and the SDF prediction network with corresponding parameters θ_{seg}, θ_{rec}, θ_{sdf}, and ξ is the noise perturbation of the student model.

2.2 Teacher Model

The network of our teacher model is reproduced from the student model, yet they have different ways for updating parameters. The student model updates its parameters $\boldsymbol{\theta} = \left\{\theta_{seg}, \theta_{rec}, \theta_{sdf}\right\}$ by gradient descent while the teacher model updates its parameters $\boldsymbol{\theta'} = \left\{\theta'_{seg}, \theta'_{rec}, \theta'_{sdf}\right\}$ as the EMA of the student model parameters $\boldsymbol{\theta}$ in different training steps. In particular, at training step t, the parameters of the teacher model, i.e., $\boldsymbol{\theta_t'}$, are updated according to:

$$\boldsymbol{\theta}_{\boldsymbol{t}}^{'} = \tau\boldsymbol{\theta}_{\boldsymbol{t-1}}^{'} + (1-\tau)\boldsymbol{\theta_t}, \tag{2}$$

where τ is the coefficient of EMA decay to control the updating rate.

Moreover, as there is no label on $\mathcal{D}^u$, the results of the teacher model may be biased. To relieve such unreliability, we bring in the uncertainty estimation in the teacher model. Specifically, we perform K times forward passes with Monte Carlo (MC) dropout, thus obtaining K preliminary results of all the tasks with regard to the input $\mathbf{X}^i$, i.e., $\left\{\tilde{\mathbf{Y}}_T^{ij}\right\}_{j=1}^K$, $\left\{\tilde{\mathbf{G}}_T^{ij}\right\}_{j=1}^K$, and $\left\{\tilde{\mathbf{Z}}_T^{ij}\right\}_{j=1}^K$.

Please note that, for the main segmentation task, we design an uncertainty weighted integration (UWI) strategy to assign different weights for different sampling results, instead of averaging the K preliminary results simply. Specifically, we derive the uncertainty maps U_{seg}^{ij} for each preliminary result by calculating the entropy $-\sum_{c\in C}\tilde{\mathbf{Y}}_T^{ijc}\log_C\tilde{\mathbf{Y}}_T^{ijc}$, where the base of the log function, C, is the number of classes to be segmented and set to 2 here. By doing so, the value range of uncertainty maps is between 0 and 1, and a larger value represents a higher degree of uncertainty. Since entropy reflects the uncertainty degree of information, we use 1-entropy to measure the confidence level for each preliminary result, leading to K confidence maps while each pixel corresponds to a vector with length of K. The values of the vector are further normalized to [0, 1] by applying a softmax operation. Afterwards, we can regard the softmax probability map at channel j, i.e., $softmax\left\{1-U_{\text{seg}}^{ij}\right\}_{j=1}^K$, as a weight map $\mathbf{W}_j$ which guides the teacher model implicititly to heed the areas with higher confidence during aggregation. Thus, the aggregated prediction of segmentation $\tilde{\mathbf{Y}}_T^i$ can be formulated as $\tilde{\mathbf{Y}}_T^i=\sum_{j=1}^K\mathbf{W}_j\odot\tilde{\mathbf{Y}}_T^{ij}$.

As for the other two auxiliary tasks, it is noteworthy that they predict the real regression values rather than the probabilistic values as the segmentation task. Accordingly, the entropy is unsuitable for the uncertainty estimation for them. Therefore, we obtain the aggregated results directly by averaging operation, that is, $\tilde{\mathbf{G}}_T^i=\frac{1}{K}\sum_{j=1}^K\tilde{\mathbf{G}}_T^{ij}$, $\tilde{\mathbf{Z}}_T^i=\frac{1}{K}\sum_{j=1}^K\tilde{\mathbf{Z}}_T^{ij}$. For the same reason, we utilize the variance instead of entropy as the uncertainty of the aggregated results of these two auxiliary tasks by following [20]. To sum up, leveraging the aggregated results of three tasks, we can acquire tripled-uncertainty maps of all the tasks by:

$$U_{\text{seg}}=-\sum_{c\in C}\tilde{\mathbf{Y}}_T^{ic}\log_C\tilde{\mathbf{Y}}_T^{ic},\ U_{\text{rec}}=\frac{1}{K}\sum_{j=1}^K\left(\tilde{\mathbf{G}}_T^{ij}-\tilde{\mathbf{G}}_T^{i}\right)^2,\ U_{\text{sdf}}=\frac{1}{K}\sum_{j=1}^K\left(\tilde{\mathbf{Z}}_T^{ij}-\tilde{\mathbf{Z}}_T^{i}\right)^2 \quad (3)$$

With the tripled-uncertainty guidance, the student model can avoid the misleading information from the teacher model and learn more trustworthy knowledge.

2.3 Objective Functions

As aforementioned, the objective function is composed of three aspects: 1) Supervised loss $\mathcal{L}_{\text{s}}$ on labeled data $\mathcal{D}^l$; 2) Inter-model consistency loss $\mathcal{L}_{\text{cons}}^{\text{model}}$ between the student model and teacher model on both $\mathcal{D}^l$ and $\mathcal{D}^u$; 3) Inter-task consistency loss $\mathcal{L}_{\text{cons}}^{\text{task}}$ among different tasks in the student model on $\mathcal{D}^l$ and $\mathcal{D}^u$.

Specifically, $\mathcal{L}_s$ is the weighted sum of the supervised losses on three tasks, i.e., $\mathcal{L}_s^{seg}$, $\mathcal{L}_s^{rec}$, $\mathcal{L}_s^{sdf}$, and can be formulated as:

$$\mathcal{L}_s = \mathcal{L}_s^{seg} + \alpha_1 \mathcal{L}_s^{rec} + \alpha_2 \mathcal{L}_s^{sdf}, \tag{4}$$

where $\mathcal{L}_s^{seg}$ uses Dice loss following [18], $\mathcal{L}_s^{rec}$ and $\mathcal{L}_s^{sdf}$ use mean squared error (MSE) loss, α_1 and α_2 are coefficients for balancing the loss terms.

For the same input from $\mathcal{D}^l$ or $\mathcal{D}^u$, since the teacher model is an ensembling of the student model, the outputs of both models on three tasks should be identical. Therefore, we employ the inter-model consistency loss $\mathcal{L}_{cons}^{model}$ to constrain this condition as follows:

$$\begin{aligned}
\mathcal{L}_{cons}^{model} &= \mathcal{L}_{cons}^{seg} + \mu_1 \mathcal{L}_{cons}^{rec} + \mu_2 \mathcal{L}_{cons}^{sdf}, \\
\mathcal{L}_{cons}^{seg} &= \frac{1}{N+M} \sum_{i=1}^{N+M} exp\left(-U_{seg}\right) \odot \left(\tilde{\mathbf{Y}}_S^i - \tilde{\mathbf{Y}}_T^i\right)^2, \\
\mathcal{L}_{cons}^{rec} &= \frac{1}{N+M} \sum_{i=1}^{N+M} exp(-U_{rec}) \odot \left(\tilde{\mathbf{G}}_S^i - \tilde{\mathbf{G}}_T^i\right)^2, \\
\mathcal{L}_{cons}^{sdf} &= \frac{1}{N+M} \sum_{i=1}^{N+M} exp(-U_{sdf}) \odot \left(\tilde{\mathbf{Z}}_S^i - \tilde{\mathbf{Z}}_T^i\right)^2,
\end{aligned} \tag{5}$$

where the tripled-uncertainty U_{seg}, U_{rec}, U_{sdf} are used as weight maps to encourage the student model learning meaningful information from the teacher model, and μ_1, μ_2 are balancing coefficients.

Similarly, owing to the shared encoder, the results of three tasks are supposed to have consistent semantic information for the same input. Based on this, we devise the inter-task consistency loss $\mathcal{L}_{cons}^{task}$ to narrow the gap between $\tilde{\mathbf{Y}}_S^i$ and $\tilde{\mathbf{G}}_S^i$, $\tilde{\mathbf{Z}}_S^i$. Accordingly, $\mathcal{L}_{cons}^{task}$ is formulated as:

$$\mathcal{L}_{cons}^{task} = \frac{1}{N+M} \sum_{i=1}^{N+M} \left(\left(\tilde{\mathbf{Z}}_S^i - SDF\left(\tilde{\mathbf{Y}}_S^i\right)\right)^2 + \left(\tilde{\mathbf{G}}_S^i - Mask\left(\tilde{\mathbf{Y}}_S^i, \mathbf{X}^i\right)\right)^2\right), \tag{6}$$

where $SDF\left(\tilde{\mathbf{Y}}_S^i\right)$ converts $\tilde{\mathbf{Y}}_S^i$ to the domain of SDF following the function in [17], and $Mask\left(\tilde{\mathbf{Y}}_S^i, \mathbf{X}^i\right)$ is the concatenation of $\tilde{\mathbf{Y}}_S^i \odot \mathbf{X}^i$ and $\left(1 - \tilde{\mathbf{Y}}_S^i\right) \odot \mathbf{X}^i$.

Finally, the total objective function $\mathcal{L}_{total}$ can be summarized as:

$$\mathcal{L}_{total} = \mathcal{L}_s + \lambda_1 \mathcal{L}_{cons}^{model} + \lambda_2 \mathcal{L}_{cons}^{task}, \tag{7}$$

where λ_1 and λ_2 are the ramp-up weighting coefficients for balancing the supervised loss and the two consistency losses.

2.4 Training Details

Our network is implemented by Pytorch and trained with two NVIDIA GeForce 2070SUPER GPUs with total 16 GB memory. We utilize Adam optimizer to train the whole network for 100 epochs with learning rate of 1e−4 and batchsize of 2. To achieve a balance betweem the training efficiency and uncertainty map quality, we perform $K = 8$ times MC dropout in the teacher model. While in testing phase, we turn off the dropout to generate the estimation directly. For updating Eq. (2), we set τ as 0.99 according

to [9]. Based on our trial studies, the hyper-parameters α_1, α_2 in Eq. (4) are set to 1, μ_1, μ_2 in Eq. (5) are set to 0.2 and 1, respectively. As for λ_1 and λ_2 in Eq. (7), following [9], we set them equally as a time-dependent Gaussian warming-up function $\lambda(t) = 0.1 * e^{(-5(1-t/t_{max})^2)}$ where t and t_{max} indicate the current training step and total training steps, respectively. Note that, only the student model is retained for generating segmentation predictions in the test phase.

3 Experiment and Analysis

Table 1. Quantitative comparison results on 2017 ACDC dataset. * means our method is significantly better than the compared method with $p < 0.05$ via paired t-test.

n/m	5/70		10/65		20/55	
	Dice [%]	JI [%]	Dice [%]	JI [%]	Dice [%]	JI [%]
U-net [2]	60.1(24.7)*	47.3(23.4)*	70.9(23.6)*	53.1(28.5)*	90.0(7.7)	82.5(11.3)
Curriculum [3]	67.5(9.9)*	51.8(10.5)*	69.2(14.6)*	50.0(14.5)*	86.6(9.0)*	77.5(13.0)*
Mean Teacher [9]	52.1(18.7)*	37.3(16.2)*	80.0(14.3)*	68.8(17.5)*	88.8(9.3)*	80.9(13.4)*
MASSL [13]	77.4(16.7)*	66.0(18.5)*	86.0(14.9)	77.8(18.5)	90.6(8.8)	84.0(12.6)
Shape-aware [17]	81.4(14.2)*	70.8(16.9)*	85.0(12.2)*	75.6(15.8)*	91.0(7.8)	84.3(11.5)
UA-MT [10]	70.7(14.1)*	56.4(15.9)*	80.6(17.8)*	70.7(21.8)*	88.7(10.5)*	81.2(14.7)*
MI-SSS [19]	81.2(20.9)*	72.4(23.4)*	84.7(15.2)	75.8(18.6)	91.2(5.6)	84.3(8.9)
Proposed	**84.6(13.9)**	**75.6(17.6)**	**87.1(11.5)**	**78.8(15.5)**	**91.3(7.6)**	**84.9(11.6)**

Dataset and Evaluation. We evaluate our method on the public datasets of 2017 ACDC challenge [15] for cardiac segmentation and PROMISE12 [16] for prostate segmentation. The 2017 ACDC dataset contains 100 subjects, where 75 subjects are assigned to training set, 5 to validating set and 20 to testing set. The PROMISE12 dataset has 50 transversal T2-weighted magnetic resonance imaging (MRI) images, from which we randomly selected 35 samples as training set, 5 as validating set and 10 as testing set. In the training set, the partition of the labeled set and the unlabeled set is denoted as n/m, where n and m are the numbers of labeled and unlabeled samples, respectively. To quantitatively assess the performance, we use two standard evaluation metrics, i.e., Dice coefficient (Dice $= \frac{2*|X \cap Y|}{|X|+|Y|}$) and Jaccard Index (JI $= \frac{|X \cap Y|}{|X \cup Y|}$). Higher scores indicate better segmentation performance.

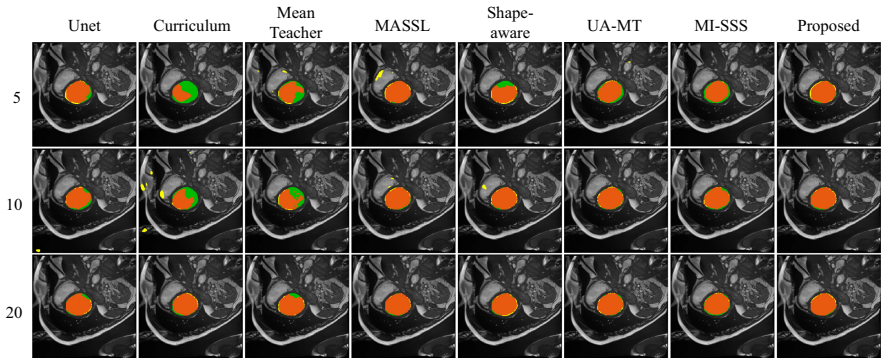

Fig. 2. Visual comparison results on 2017 ACDC dataset. Orange indicates the correct segmented area, green the unidentified and yellow the miss-identified. (Color figure online)

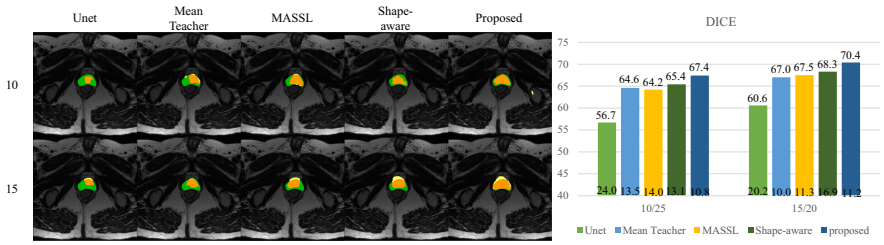

Fig. 3. Visual comparison results on PROMISE12 dataset. Averages and standard deviations are provided above and below the bars, respectively.

Comparison with Other Methods. To demonstrate the superiority of the proposed method in leveraging unlabeled and labeled data, we compare our method with several methods including U-net [2], Curriculum [3], Mean Teacher [9], MASSL [13], Shape-aware [17], UA-MT [10] and MI-SSS [19]. It is worth noting that only U-net is trained in a fully supervised manner while others are semi-supervised. Table 1 is a summary of quantitative results on 2017 ACDC dataset in different n/m settings. As observed, our proposed method outperforms all the compared methods with the highest Dice and JI values in all n/m settings. Specifically, for the fully supervised U-net, our method can leverage the unlabeled data and largely improve Dice and JI from 60.1%, 47.3% to 84.6%, 75.6%, respectively, when $n = 5$. Besides, when n is small, the improvement of our method is statistically significant with $p < 0.05$ via paired t-test. With more labeled data available, all the methods show an upward trend with a narrowing gap, but our method still ranks first. We also give a qualitative comparison result on 2017 ACDC dataset in Fig. 2. It can be seen that the target area is better sketched by our method with more accurate boundaries. On the other hand, our method produces the fewest false positive predictions, thereby generating the most similar segmentation results to ground truth over all of the compared methods. Figure 3 provides comparison results on PROMISE12 dataset. Clearly, our methods achieves the best performance in both qualitative and quantitative measures.

Ablation Study. To investigate the contributions of key components of our method, we further conduct a series of experiments in different model settings on 2017 ACDC dataset. First, to validate the effectiveness of auxiliary tasks, we compare the models of (1) the segmentation task alone (Seg), (2) the segmentation task and the SDF prediction task (Seg+SDF), and (3) all the three tasks (Seg+SDF+Rec). The quantitative results are shown in the upper part in Table 2, from where we can see that the Seg model exhibits the worst performance. With the SDF prediction task and the reconstruction task joining, Dice and JI are improved to varying degrees. Especially when n is small, the improvement by SDF is significant, revealing its large contribution to the utilization of unlabeled data.

Table 2. Ablation study of our method on 2017 ACDC dataset. * means our method is significant better than compared method with $p < 0.05$ via paired t-test.

n/m	5/70		10/65		20/55	
	Dice [%]	JI [%]	Dice [%]	JI [%]	Dice [%]	JI [%]
Seg	60.1(24.7)*	47.3(23.4)*	70.9(23.6)*	53.1(28.5)*	90.0(7.7)	82.5(11.3)
Seg+SDF	81.0(16.7)*	70.8(19.5)*	83.4(17.2)*	74.5(20.2)*	90.2(8.5)	83.1(12.5)
Seg+SDF+Rec	81.2(13.8)*	70.9(17.8)*	84.2(15.1)*	75.4(18.6)*	90.7(7.2)	83.8(10.8)
S+T	52.1(18.7)*	37.3(16.2)*	80.0(14.3)*	68.8(17.5)*	88.8(9.3)*	80.9(13.4)*
S+T+UncA	70.8(22.4)*	58.8(22.6)*	82.8(20.5)*	74.7(23.0)*	90.2(8.2)	83.1(12.0)
S+T+UncW	79.3(23.4)*	70.5(25.3)*	83.2(18.7)*	74.7(21.8)*	90.7(7.8)	83.8(11.6)
Proposed w/o. Tri-U	82.6(13.0)	72.4(16.6)	86.0(12.2)	77.1(16.0)	90.9(8.4)	84.2(12.4)
Proposed	**84.6(13.9)**	**75.6(17.6)**	**87.1(11.5)**	**78.8(15.5)**	**91.3(7.7)**	**84.9(11.6)**

Second, we regard the Seg model as the student model (S) and construct variant models by incorporating (1) the teacher model (T), (2) the uncertainty estimation with averaging (UncA), and (3) the uncertainty estimation with the proposed UWI (UncW). The middle part of Table 2 presents the detailed results. We can find that the S+T model decreases Dice and JI by 8.0% and 10% compared with S only when $n = 5$. This may be explained as that the teacher model is susceptible to noise when labeled data is few, thus degrading the performance. However, by considering the uncertainty estimation, the S+T+UncA model rises Dice and JI remarkably by 18.7% and 21.5% for $n = 5$, and the proposed S+T+UncW model yields higher indicator values, proving their effectiveness.

Third, to verify the guiding role of the tripled-uncertainty, we compare the proposed model and that without the tripled uncertainty, i.e., proposed w/o. tri-U, and display the results in the Table 2. As observed, our complete model gains better performance, demonstrating the promotion effect of the devised tripled-uncertainty.

4 Conclusion

In this paper, we propose a tripled-uncertainty guided semi-supervised model for medical image segmentation. Based on a mean teacher architecture, our model explores the relationship among the segmentation task, the foreground and background reconstruction

task and the SDF prediction task. To eliminate the possible misdirection caused by the noisy unlabeled data, we employ uncertainty estimation on all three tasks in the teacher model. In contrast to the common uncertainty averaging integration strategy, we consider the differences of each sampling and develop a novel uncertainty weighted integration strategy. The experimental results demonstrate the feasibility and superiority of our method.

Acknowledgement. This work is supported by National Natural Science Foundation of China (NFSC 62071314) and Sichuan Science and Technology Program (2021YFG0326, 2020YFG0079).

References

1. Chen, J., Yang, L., Zhang, Y., et al.: Combining fully convolutional and recurrent neural networks for 3D biomedical image segmentation. In: NIPS (2016)
2. 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
3. Kervadec, H., Dolz, J., Granger, É., Ayed, I.B.: Curriculum semi-supervised segmentation. In: Shen, D., et al. (eds.) MICCAI 2019. LNCS, vol. 11765, pp. 568–576. Springer, Cham (2019). https://doi.org/10.1007/978-3-030-32245-8_63
4. Bortsova, G., Dubost, F., Hogeweg, L., Katramados, I., 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
5. Zheng, Z., et al.: Semi-supervised segmentation with self-training based on quality estimation and refinement. In: Liu, M., Yan, P., Lian, C., Cao, X. (eds.) MLMI 2020. LNCS, vol. 12436, pp. 30–39. Springer, Cham (2020). https://doi.org/10.1007/978-3-030-59861-7_4
6. Park, S., Hwang, W., Jung, K.H.: Integrating reinforcement learning to self- training for pulmonary nodule segmentation in chest x-rays. arXiv preprint arXiv:1811.08840 (2018)
7. Zheng, H., et al.: Cartilage segmentation in high-resolution 3D micro-CT images via uncertainty-guided self-training with very sparse annotation. In: Martel, A.L., et al. (eds.) MICCAI 2020. LNCS, vol. 12261, pp. 802–812. Springer, Cham (2020). https://doi.org/10.1007/978-3-030-59710-8_78
8. Laine, S., Aila, T.: Temporal ensembling for semi-supervised learning. arXiv preprint arXiv:1610.02242 (2016)
9. Tarvainen, A., Valpola, H.: Mean teachers are better role models: weight-averaged consistency targets improve semi-supervised deep learning results. In: Proceedings of the 31st International Conference on Neural Information Processing Systems, pp. 1195–1204 (2017)
10. Lequan, Y., Wang, S., Li, X., Chi-Wing, F., Heng, P.-A.: Uncertainty-aware self-ensembling model for semi-supervised 3D left atrium segmentation. In: Shen, D., et al. (eds.) MICCAI 2019. LNCS, vol. 11765, pp. 605–613. Springer, Cham (2019). https://doi.org/10.1007/978-3-030-32245-8_67
11. Wang, Y., et al.: Double-uncertainty weighted method for semi-supervised learning. In: Martel, A.L., et al. (eds.) MICCAI 2020. LNCS, vol. 12261, pp. 542–551. Springer, Cham (2020). https://doi.org/10.1007/978-3-030-59710-8_53

12. Luo, X., Chen, J., Song, T., et al.: Semi-supervised medical image segmentation through dual-task consistency. arXiv preprint arXiv:2009.04448 (2020)
13. Chen, S., Bortsova, G., Juárez, A.-U., Tulder, G., Bruijne, M.: Multi-task attention-based semi-supervised learning for medical image segmentation. In: Shen, D., et al. (eds.) MICCAI 2019. LNCS, vol. 11766, pp. 457–465. Springer, Cham (2019). https://doi.org/10.1007/978-3-030-32248-9_51
14. Li, S., Zhang, C., He, X.: Shape-aware semi-supervised 3d semantic segmentation for medical images. In: Martel, A.L., et al. (eds.) MICCAI 2020. LNCS, vol. 12261, pp. 552–561. Springer, Cham (2020). https://doi.org/10.1007/978-3-030-59710-8_54
15. Bernard, O., Lalande, A., Zotti, C., 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)
16. Litjens, G., Toth, R., van de Ven, W., et al.: Evaluation of prostate segmentation algorithms for MRI: the PROMISE12 challenge. Med. Image Anal. **18**(2), 359–373 (2014)
17. Xue, Y., Tang, H., Qiao, Z., et al.: Shape-aware organ segmentation by predicting signed distance maps. In: Proceedings of the AAAI Conference on Artificial Intelligence, vol. 34, no. 7, pp. 12565–12572 (2020)
18. 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)
19. Peng, J., Pedersoli, M., Desrosiers, C., et al.: Boosting semi-supervised image segmentation with global and local mutual information regularization. arXiv preprint arXiv:2103.04813 (2021)
20. Kendall, A., Gal, Y.: What uncertainties do we need in bayesian deep learning for computer vision? arXiv preprint arXiv:1703.04977 (2017)