An Effective Semi-supervised Approach for Liver CT Image Segmentation

Kai Han, Lu Liu, Yuqing Song, Yi Liu, Chengjian Qiu, Yangyang Tang, Qiaoying Teng, Zhe Liu*

***Abstract*— Despite the substantial progress made by deep networks in the field of medical image segmentation, they generally require sufficient pixel-level annotated data for training. The scale of training data remains to be the main bottleneck to obtain a better deep segmentation model. Semi-supervised learning is an effective approach that alleviates the dependence on labeled data. However, most existing semi-supervised image segmentation methods usually do not generate high-quality pseudo labels to expand training dataset. In this paper, we propose a deep semi-supervised approach for liver CT image segmentation by expanding pseudo-labeling algorithm under the very low annotated-data paradigm. Specifically, the output features of labeled images from the pretrained network combine with corresponding pixel-level annotations to produce class representations according to the mean operation. Then pseudo labels of unlabeled images are generated by calculating the distances between unlabeled feature vectors and each class representation. To further improve the quality of pseudo labels, we adopt a series of operations to optimize pseudo labels. A more accurate segmentation network is obtained by expanding the training dataset and adjusting the contributions between supervised and unsupervised loss. Besides, the novel random patch based on prior locations is introduced for unlabeled images in the training procedure. Extensive experiments show our method has achieved more competitive results compared with other semi-supervised methods when fewer labeled slices of LiTS dataset are available.**

***Index Terms*— semi-supervised learning; medical image segmentation; data-augmentation; liver segmentation**

I. Introduction

Automated medical image segmentation is the crucial component of Computer Aided Diagnosis (CAD). Accurate segmentation results can not only be used for subsequent quantitative evaluation of the region of interest, but also benefit accurate diagnosis, formulate surgical plans and intra-operative guidance [1]. In recent years, various segmentation approaches based on deep learning have made remarkable progress, such as FCN [2], SegNet [3], *etc.*. However, the scale of training data is still the main bottleneck of deep models. Generally, the model trained on the insufficient dataset is prone to over-fitting and fails to obtain better generalization ability. Compared with natural images, it is more difficult for acquiring rich medical pixel-annotated images. For one thing, the process of pixel-annotations is tedious and time-consuming due to the low contrast and resolution of medical images. For another, annotating organs or diseased areas often requires veteran radiologists who have corresponding domain knowledge. Compared with annotated medical images, unlabeled images are easier to obtain. A natural ideal is to utilize unlabeled images to improve the performance of supervised learning.

Semi-supervised learning is a more effective approach that reduces the dependence of labeled data with a large amount of unlabeled data [4]. For semi-supervised image segmentation, there are several popular methods according to the manner of training: Self-training [5], [6], Co-training [7], [8] and Mean Teacher-based methods [9], [10]. Self-training method usually contains one model while Co-training and Mean Teacher-based methods contain two or more models. Bai *et al.* [5] proposed a semi-supervised segmentation approach where the parameters of the network and the segmentation for unlabeled data are alternatively updated. Besides, they adopted Conditional Random Field (CRF) to refine the segmentation for unlabeled data. Fan *et al.* [6] designed a semi-supervised segmentation framework based on a randomly selected propagation strategy for enlarging the training dataset with unlabeled data, which only requires a few labeled images and leverages primarily unlabeled data. Compared with Self-training, Co-training usually utilizes two or more models to produce pseudo labels for each other, which is a typical approach of multi-view training. Zhou *et al.* [7] learned a deep model in a co-training style which mines consensus information from multiple planes like the sagittal, coronal and axial planes. This multi-plane fusion method is used to generate more reliable pseudo labels, and the error generated by pseudo labels is reduced. Xia *et al.* [11] further designed an uncertainty aware multi-view co-training framework where co-training is conducted by exploiting multi-viewpoint consistency of 3D data, and estimated the reliability of each view's prediction with Bayesian deep learning. Mean Teacher-based framework is a more prevalent framework for semi-supervised medical image segmentation. The main difference between Mean Teacher and other models is that exponential moving average (EMA) is used in the training steps to average the weight of the model, and it tends to generate a more accurate model instead of directly using output prediction. These above approaches embrace their own shortcomings. Self-training's core is to ensure the quality of pseudo labels for unlabeled data. In other words, the model would make incorrect reasoning and even encounter the collapse of networks if pseudo labels contain more noises at the beginning of training. Co-training and Mean Teacher-based methods usually have more parameters to be optimized and are at high computing expense. Besides, Mean Teacher and its variants with exponential moving averaging (EMA) on parameter updating, which encounters a parameter-correlation between teacher and student models [12]. Thus, the reliability of used pixels/voxels may not be stable enough.

For liver segmentation, Jin *et al.* [13] utilized an improved fully convolution network to segment liver, but the results of this approach are not accurate and nor stable enough. Liu *et al.* [14] designed a DFS U-Net to segment small liver areas and has achieved a better performance. The H-DenseUnet [15] cascaded 2D dense U-Net and 3D residual dense U-Net and achieved satisfactory accuracy segmentation performance. However, this model has a more complex structure with a large number of parameters to be optimized, which undoubtedly exacerbates the cost of calculation. Recently, Li *et al.* [16] designed a nested attention-aware segmentation network (Attention ++), where attention mechanism between nested convolutional blocks is used,

This work was supported by the National Natural Science Foundation of China (61976106, 61772242, 61572239); China Postdoctoral Science Foundation (2017M611737); Six talent peaks project in Jiangsu Province (DZXX-122).

K.Han, Y.Song, Y.Liu, C.Qiu, Y.Tang and Z.Liu are with the School of Computer Science and Telecommunication, Jiangsu University, Zhenjiang, Jiangsu Province, China (Email: 1000004088@ujs.edu.cn)

Lu Liu is with the School of Computing and Mathematics, University of Leicester, Leicester, UK, LE1 7RH. (E-mail: l.liu@Leicester.ac.uk)

Q.Teng is with the School of Computer Science, Jilin Normal University, Siping, China.

*Corresponding author

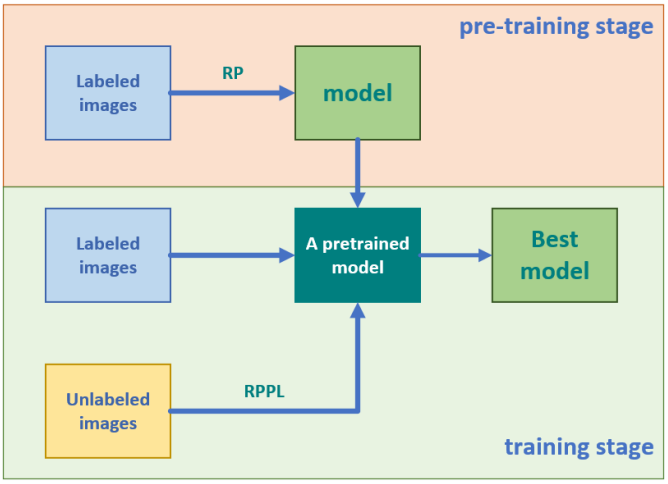

Fig. 1: The training procedure of our method.

so that the features extracted at different levels could be merged with a task-related selection, which has obtained a 97.48% dice score on LiTS dataset. Although these methods have higher accuracy performance for liver segmentation, they are usually susceptible to over-fitting risk and lack of better generalization ability when training images are insufficient.

To this end, we propose a novel semi-supervised liver segmentation approach by expanding pseudo-labeling algorithm, which leverages very few labeled images to guide the generation of pseudo labels for unlabeled images and achieves better performance. Meanwhile, our method of using original U-Net network [17] is more simplified and has fewer parameters to be optimized. In summary, our main contributions are as follows:

(i) We propose a novel semi-supervised segmentation framework based on pseudo-labeling algorithm for liver CT image segmentation;

(ii) We obtain the high-quality pseudo labels by adopting location constraints and morphological operations;

(iii) Our experimental results on LiTS have competitive performance than other typical semi-supervised segmentation methods under very few labeled data paradigm.

II. Method

This section mainly covers notations, data pre-processing, and semi-supervised image segmentation. For data pre-processing, the Random-Patch (RP) method is adopted for labeled images in pre-training stage and a novel Random-Patch method based on Prior Locations (RPPL) is designed to constrain unlabeled images during the training stage. In semi-supervised image segmentation, features and their pixel-level annotations for labeled images are combined to produce class representations. Pseudo labels are generated by computing the distance between feature vectors of unlabeled images and each class representation. Moreover, a series of operations are implemented to improve the quality of pseudo labels. The final total loss is calculated by weighting supervised loss and unsupervised loss. Fig 1 shows the procedure of training. The pipeline of our semi-supervised segmentation framework is shown in Fig 2.

A. Notations and data pre-processing

Assuming that a dataset has two subsets $D = D_L \cup D_U$, including the labeled dataset $D_L = (I_i, M_i)_{i=1}^{N_L}$ and the unlabeled dataset $D_U = (I_j)_{j=1}^{N_U}$, where I and M represent images and corresponding masks, N_L and N_U denote the number of labeled and unlabeled images, respectively.

The slices of human abdomen could be fuzzy when not pre-processed, then recognizing organs and tissues would be difficult. Thus, HU windowing [18] is used to pre-process abdomen images according to the following formula:

$$HU = pixel\_value \times rescale\_slope + rescale\_intercept \quad (1)$$

Where, $rescale\_slope$ and $rescle\_intercept$ are set to 1 and -1024, respectively. To neglect organs and tissues that are not of interest, the original CT slices are windowed to a Hounsfield Unit in the range of -75 to 175 HU. Besides, the contrast for CT images is enhanced through histogram equalization. The comparison between before and after processing is shown in Fig 3.

Due to the scarceness of available training images, it is vital to conduct data-augmentation and constraint. Thus, the Random-Patch method (RP) [19] is adopted for labeled images in pre-training stage and a novel Random-Patch based on Prior Locations (RPPL) for unlabeled images is adopted during the training stage. The processes of the two methods are shown in Fig 4.

In pre-training stage, RP is that the patch at random is chosen from the original image. By doing data-augmentation, a better initial network is obtained before the semi-supervised training stage. In the training stage, observing that the location of liver in CT images has a certain range, we impose prior constraints before random patch selection. Specifically, RPPL first crops unlabeled images and keeps the range of width and height of each patch $(W - W') : W, (H - H') : H$ (*i.e.* The red solid line in right figure of Fig 4) according to the range of liver location. Then the random patches are generated in the limited areas. Compared with RP in pre-training stage, the purpose of RPPL is to keep unlabeled images containing fewer background areas.

B. Semi-supervised image segmentation

For each input labeled image $I_i \in \mathbb{R}^{W \times H}$,

$$F_i = f_\Theta(I_i) \quad (2)$$

Where, f is the segmentation network parameterised by Θ and it is pre-trained on labeled images by RP, $F_i \in \mathbb{R}^{W \times H \times C}$ is the feature map. Then it is fed into 1×1 convolution layer and get probability map p_i after a $sigmoid$ function. Furthermore, the supervised loss is calculated by Binary Cross-Entropy (BCE) loss function between predicted probability p_i and ground-truth M_i. The formulas of the above operations are as follows:

$$p_i = Sigmoid(Conv_{1\times1}(F_i)) \quad (3)$$

$$L_{sup} = -\frac{1}{B_L}(\sum_{i=1}^{B_L} M_i log p_i + (1 - M_i) log(1 - p_i)) \quad (4)$$

Where, B_L is batch size.

1) The generation of pseudo labels: Although unlabeled images cover abundant latent information, they have no annotations available for supervision training. Combining labeled images and corresponding masks produces class representations for guiding unlabeled images. Specifically, The foreground class representation is obtained by multiplying the mask by the network output feature map and then computing mean value. Besides, swapping 0 and 1 in the mask, the background class representation is calculated like the above

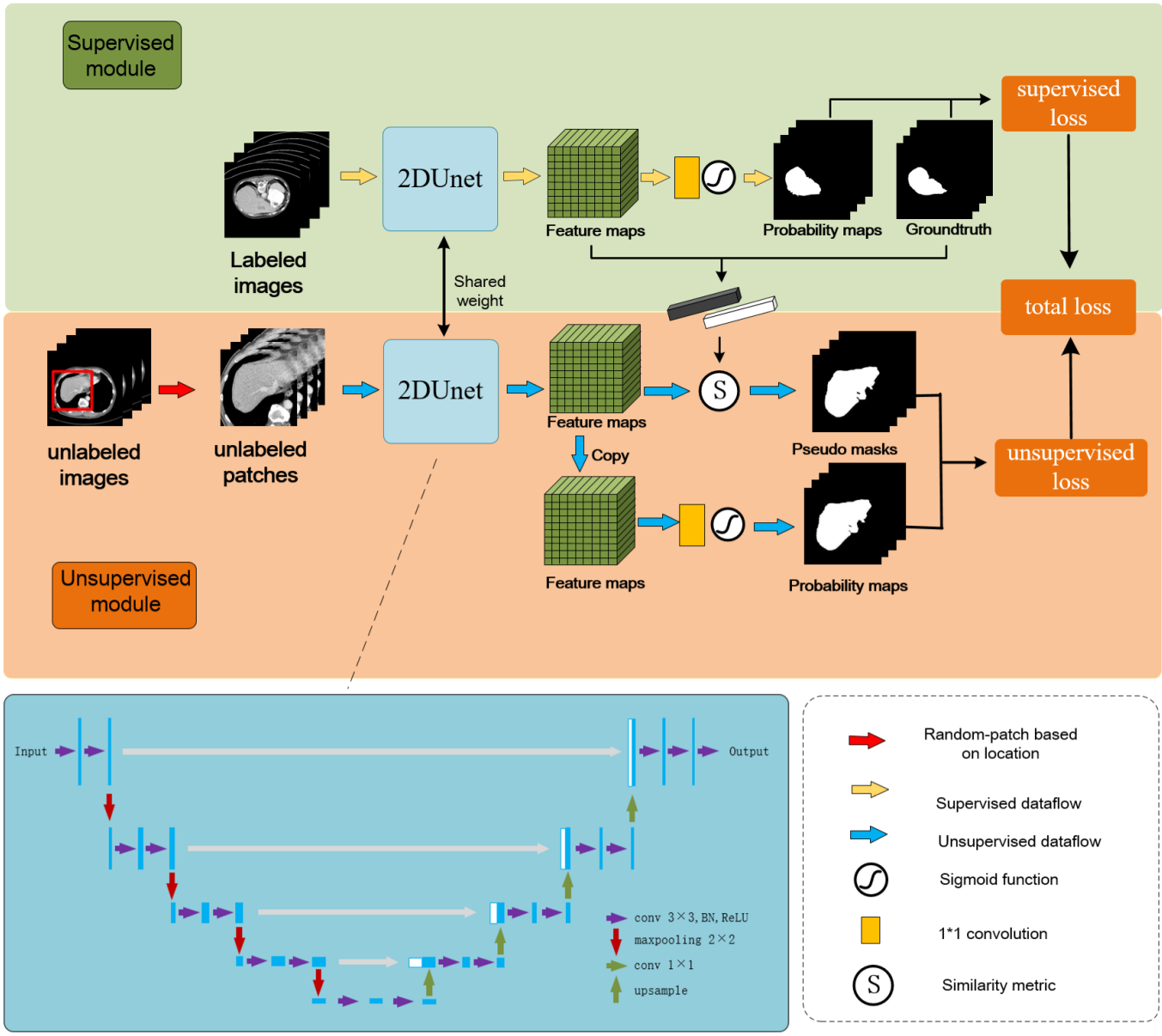

Fig. 2: The framework of proposed method.

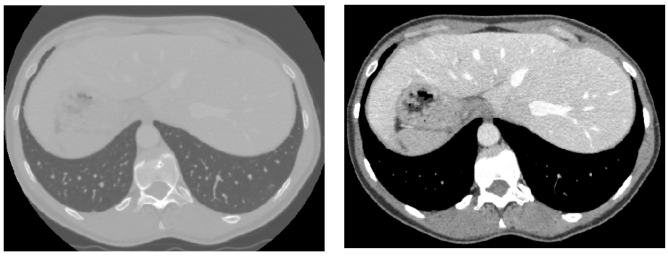

Fig. 3: The raw CT slice (left) and the CT slice after HU constrain and histogram equalization (right).

process. The foreground class representation P_{fg} and background class representation P_{bg} are calculated as follows:

$$P_{fg} = \frac{1}{B_L}\sum_{i=1}^{B_L}\frac{\sum_{x,y} F_i^{(x,y)}\mathbb{I}(M_i^{(x,y)} = fg)}{\sum_{x,y}\mathbb{I}(M_i^{(x,y)} = fg)} \tag{5}$$

$$P_{bg} = \frac{1}{B_L}\sum_{i=1}^{B_L}\frac{\sum_{x,y} F_i^{(x,y)}\mathbb{I}(M_i^{(x,y)} = bg)}{\sum_{x,y}\mathbb{I}(M_i^{(x,y)} = bg)} \tag{6}$$

Where, $P_{fg}, P_{bg} \in \mathbb{R}^{1\times C}$. $\mathbb{I}(.)$ is an *indicator* function, outputting value 1 if the argument is true or 0 otherwise. (x, y) represents the spatial location of feature map F_j.

Similar to labeled images, each unlabeled patch is fed into the network and get F_j, which further produces predicted probability map p_j by convolution layer and sigmoid function. $F_j \in \mathbb{R}^{W'\times H'\times C}$ can be regarded as $W' \times H'$ vectors of dimension $1 \times C$. For each vector in F_j, the distance between foreground or background

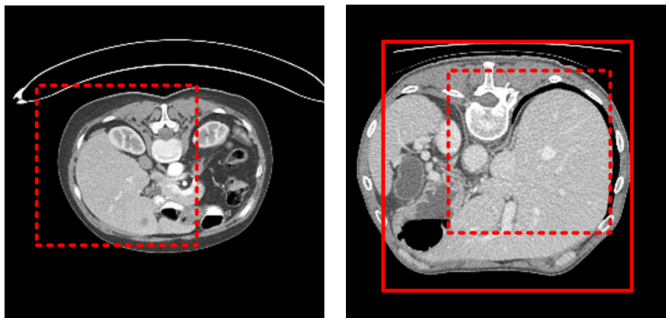

Fig. 4: The processes of the random patch (RP) for labeled images in the pre-training stage (left) and the process of the random patch based on prior locations (RPPL) for unlabeled images in the training stage (right). The red spot line represents the random patch.

class representations and each vector in F_j is calculated by cosine similarity metrics, which is shown in formula (7):

$$Dist[k] = F_j^{(x,y)} \cdot P_k / (\|F_j^{(x,y)}\| \cdot \|P_k\|), \quad k \in \{fg, bg\} \tag{7}$$

Pseudo label P_{mj} is achieved by the $softmax$ function and $argmax$ function:

$$P_{mj} = argmax_k(softmax(Dist[k])), \quad k \in \{fg, bg\} \tag{8}$$

By measuring the distance between the feature vector and foreground/background class representation, the corresponding pixel is classified by selecting small distance value.

2) The improvement of pseudo labels: Since pseudo labels usually contain more noises, which is harmful to the reasoning of the network, there is a high possibility to remove these noises and improve the quality of pseudo labels. Morphological operations rely on the correlation of pixel values rather than their absolute values, thus they are very suitable for binary image optimization. Morphological operations contain erosion and dilation. The corrosion removes small bumps or burrs from objects smaller than structural elements. By selecting structural elements of different sizes, objects of different sizes could be removed from the image. The expansion operation enlarges the image and fill the holes in the objects. The formulas of erosion and dilation operations are shown as follows:

$$(f \oplus b)(s,t) = min\{f(s+x, t+y) - b(x,y) | (x,y), (t+y) \in D_f; (x,y)\} \tag{9}$$

$$(f \ominus b)(s,t) = max\{f(s-x, t-y) - b(x,y) | (s-x), (t-y) \in D_f; (x,y)\} \tag{10}$$

Where, $f(x,y)$ is the input image and $b(x,y)$ is the structure element, respectively. The process of first corrosion and then expansion is open operation, which is used to smooth the boundary of larger objects without changing their areas. The process from first expansion to corrosion is a closed operation, which fills the internal holes of the object connecting the adjacent objects, smooth the boundary without changing its area. Since noise and details may be located above or below the target signal, a single erosion or expansion operation only eliminates the noise and details above or below the target signal, and the smoothed image is always above or below the original image, so it will cause the position shift of the target information. Constructing a morphological opening and closing hybrid operation could achieve the purpose of eliminating the detail noise in the image while maintaining the integrity and position of the target information.

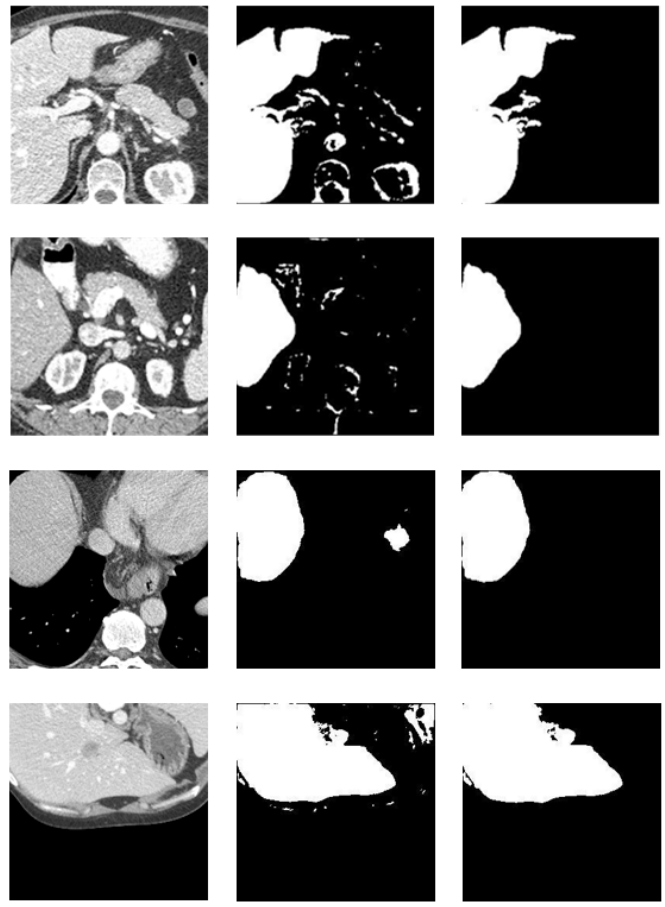

Fig. 5: The results of the improvement of pseudo labels. The first column to last column denote unlabeled images, original pseudo labels and the optimized pseudo labels, respectively.

Considering there are other disturbances around the liver, the number of pixels is computed for each connected area and then the largest connected area is selected as the final pseudo labels. The visual results of pseudo labels improvement are shown in Fig 5.

3) Semi-supervised learning: The unsupervised loss is shown as follows:

$$L_{unsup} = -\frac{1}{B_U}(\sum_{j=1}^{B_U} P_{mj} log p_j + (1 - P_{mj}) log(1 - p_j)) \tag{11}$$

The process of semi-supervised learnings (SSL) includes both supervised learning and unsupervised learning. Final total loss is the weighted sum between supervised loss and unsupervised loss:

$$L_{total} = \lambda L_{sup} + (1 - \lambda) L_{unsup} \tag{12}$$

Note that, parameter $\lambda \in [0, 1]$ that controls the contribution of the supervised and unsupervised term. Fig 6 shows the pseudo code of the entire procedure for semi-supervised segmentation.

III. Experiments

A. Dataset

Our proposed method is evaluated on the training set of LiTS Liver Tumor Segmentation Challenge. The dataset contains 131 contrast-enhanced CT images provided by hospitals around the world. Note that, 3DIRCADb dataset is a subset of LiTS dataset with case

Algorithm 1 Semi-supervised segmentation

Input: labeled dataset D_L, labeled dataset D_U, λ
Output: model parameters θ;

$D_L^{RP} \leftarrow Random\ patch(D_L)$
Get θ pre-trained on D_L^{RP};
t $\leftarrow 0$
repeat
  $t \leftarrow t+1$
  for sample (I, M) from D_L **do**
    Compute L_{sup} by Eq.(4)
    Obtain P_{fg} and P_{bg} by Eq.(5) and (6);
  end for
  $D_U^{RPPL} \leftarrow RPPL(D_U)$
  for sample (I) from D_U^{RPPL} **do**
    Compute $Dist$ by Eq.(7)
    Obtain psudo labels by Eq.(8)
    Optimize psudo labels
    Compute L_{unsup} by Eq.(11);
  end for
  Compute L_{total} by Eq.(12);
until the model converge.

Fig. 6: The pseudo-code of semi-supervised segmentation method.

numbers from 27 to 48. Our model is trained with 111 cases from LiTS after removing 3DIRCADb and is evaluated on 3DIRCADb dataset. To further verify our method in a low-labeled data paradigm, 100 labeled slices and 900 unlabeled slices are randomly selected from different subjects in the training dataset.

B. Implementation Details

All models are trained with Stochastic Gradient Descent (SGD) optimization algorithm. In the pre-training stage, the original CT slices are windowed to a Hounsfield Unit in the range of -75 to 175 HU. Besides, the initial learning rate (lr) of pretrained model is set 1e-2 with a momentum 0.9 and the learning rate is decreasing by $lr \times ((1 - epoch)/max\_epoch)^{0.9}$. While the learning rate in semi-supervised training stage is initialized by the learning rate of best pre-trained model. The patch size is set to 256× 256 for labeled images by using RP. During training, the patch size of unlabeled images is set to 256× 256 and labeled images still keep the size of 512× 512. All experiments are performed on a machine with CPU Intel Core i7-7700K @ 4.2 GHz, GPU NVIDIA GeForce GTX 1080 Ti, and 11 GB of RAM.

C. Evaluation Criterion

Following the evaluation criterion used in the 2017 LiTS challenge, we use four evaluation metrics to measure the segmentation performance for all experiments, including Dice coefficient, Jaccard index, Precision and Recall.

D. Ablation Study

Traditional pseudo-labeling method [20] generates pseudo labels by output predictions with the pre-trained model. Our proposed metrics-based pseudo-labeling method is different from it, which leverages the features of labeled images and ground-truth to produce pseudo labels while [20] does not. To verify the quality of pseudo labels produced by two methods, 500 unlabeled patches are selected

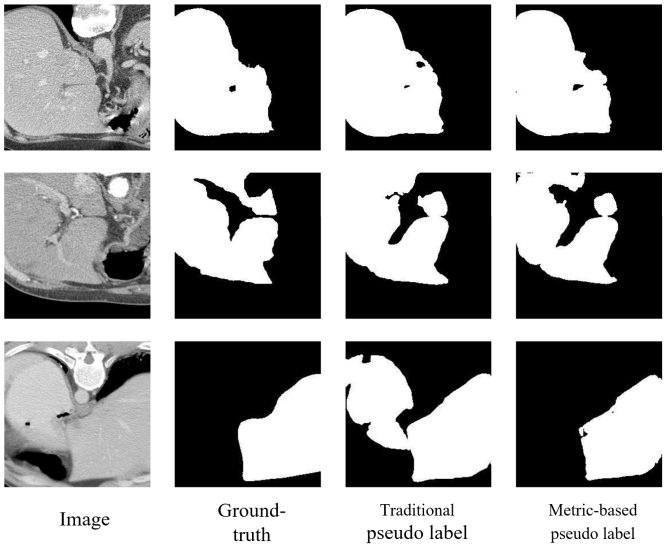

Fig. 7: The comparisons of pseudo labels generated by two types of pseudo-labeling approaches.

TABLE I: The mean dice of two types of pseudo labels.

Method	Dice	Jaccord	Precision	Recall
Pseudo-labeling [20]	78.41%	68.56%	72.14%	79.15%
Metrics-based pseudo-labeling	79.68%	70.81%	75.36%	79.41%

at random after pre-training. The mean dice of the two methods is shown in Table I. Besides, several typical cases are chosen to visual comparison in Fig 7. From Table I and Fig 7, metrics-based pseudo-labeling method can utilize labeled images to guide the generation of higher quality pseudo labels for unlabeled images.

To validate the effectiveness of the proposed semi-supervised segmentation method, relevant ablation experiments are conducted and the results are shown in Table II. The original U-Net with 16 output channels is adopted as the baseline model. From Table II, RP is very helpful to increase the diversity of the number of images, which shows great potential under the scarceness of data. While RPPL limits the range of liver area so that pseudo labels contain more of interest areas, which could provide much useful information for network training. Why RP is not implemented for the labeled images during the training phase? The class representations will be not robust when computing the mean value of feature vectors for patches in a batch size, which affects subsequent measurements, we thus keep labeled images their original size instead of patched size. In addition, the number of labeled images needs to be kept quite large in a batch B_L.

The parameter λ is the core of our semi-supervised segmentation framework. If λ is closed to 1, which means the total loss is dominated by unsupervised loss and the network is forced to learn inaccurate predictions, even resulting in the collapse of the model. If λ is close to 0, unlabeled data has almost no contribution to the entire training procedure. Thus, it is crucial to choose a reasonable value λ so that the model could be guided by supervised learning and at the same time, gain benefits from unsupervised learning. In fewer semi-supervised works, the proportion of the unsupervised loss or semi-supervised loss is equal to supervised loss [21], [22], they usually need two or more parameters-related models and the outputs of different model have fewer differences. In most cases, the unsupervised or semi-supervised weights tend to change dynamically [23], [24] or controlled by a fixed hyper-parameter [25]–[27] due to

TABLE II: The results of the ablation study on the proposed approach. Where RP, SSL and RPPL represent random-patch, semi-supervised learning and random-patch based on prior locations, respectively. Dynamic Function (DF) is $\mu(T) = k * e^{-5(1-T)^2}$.

Method	Labeled data	Unlabeled data	Lambda	Dice score	Jaccard	Precision	Recall
Baseline	100	0	-	76.64%	66.76%	71.00%	78.21%
Baseline+RP	100	0	-	78.04%	68.43%	72.84%	79.45%
Baseline+RP+SSL	100	900	0.95	80.00%	71.14%	76.06%	77.94%
Baseline+RP+RPPL+SSL	100	900	0.8	80.21%	70.15 %	74.65%	79.14%
	100	900	0.85	81.72%	72.51 %	78.65%	83.73%
	100	900	0.9	84.02%	75.80 %	84.49%	82.52%
	100	900	0.95	83.65%	76.57 %	**85.17%**	83.24%
	100	900	DF	**84.85**%	**77.80** %	84.93%	**84.65%**

TABLE III: The results of three same experiments when sampling 1000 images (100 labled images and 900 unlabeled images) and using DF as λ.

Group	Dice	Jaccord	Precision	Recall
1	84.85%	77.80%	84.93%	84.65%
2	83.49%	78.01%	84.56%	83.79%
3	84.77%	77.54%	83.92%	84.26%
Average	84.37%	77.78%	84.47%	84.23%

the large fluctuations of unsupervised outputs. In our experiments, it is detrimental that the network cannot be dominated by supervised part if $\lambda \leq 0.5$. Experimentally, we choose other fixed values when interval is 0.05 and Dynamic Function (DF) as λ. By constantly adjusting the parameters, the model has the highest dice coefficient when lambda is DF. With the continuous learning of the model, the accuracy of the model is gradually improved and the accuracy of the unsupervised part is also improved. The dynamic increase of the unsupervised proportion will also promote the model form better learning. In order to valid the robustness of the experiment, we also carry out three same experiments using same data, which is shown in TABLE III.

E. Comparison to other semi-supervised segmentation methods

We conduct two experimental settings that includes 10% labeled data and 20% labeled data when total number is 1000. Four typical semi-supervised segmentation methods [9], [28]–[30] are selected for comparison. Where, Entropy Mini [28] is a single model and other methods are double models. The results of the comparison are shown in Table IV. Generally, two or more networks (*e.g.* the Self-ensembling model) often have higher accuracy for medical image segmentation than a single network. By observing Table IV, CPS [29] has achieved 81.59% dice score, it is deduced that two models of CPS may generate low-quality pseudo labels and utilize them in supervising each other, which undoubtedly increases the bias. Besides, although Entropy Mini [28] has better dice score than CPS, it has poor performance to specific cases from Fig 8. It is obvious that UA-MT [9] obtains better segmentation performance based on the Mean-Teacher model and utilization of an uncertainty-aware scheme exploiting the uncertainty information. Compared with the above methods, our proposed approach has fewer parameters and the segmentation performance is better than UA-MT. During an epoch, our pseudo labels generated by leveraging information from labeled images supervise unlabeled images and take part in back-propagation. In other words, the unsupervised part of other methods is guided by the network trained on labeled images instead of network and direct information from labels. Thus our method has better performance for extreme cases in Fig 8, which is exactly where we are distinguished.

IV. Discussion

In this paper, we propose an effective semi-supervised segmentation framework for liver CT image segmentation when labeled images are insufficient. Different from previous studies, our method generates high-quality pseudo labels guided from labeled images. We first conduct a random patch on labeled images to obtain a better pretrained model in pre-training stage and propose a novel random patch based on prior locations to constrain unlabeled images during training. By combining the features and masks of labeled images, the class representations are calculated and further used for guiding the generation of pseudo labels. To optimize pseudo labels, a series of operations are adopted to fill holes and remove noises near the liver. A more accurate network is achieved by combining the supervised loss and unsupervised loss. With the contributions provided by these components, our approach achieves a remarkable segmentation performance.

Compared with other semi-supervised segmentation approaches, the difference of our methods is that the pseudo-labels are generated by the labeled image and the network instead of the network output. In addition, a series of operations to optimize pseudo labels, including a good initial network, conducting location constrains on unlabeled images and morphological operations, which all improve the quality of pseudo labels. In addition, our method is suitable to various networks with encoding and decoding structures.

However, our method embraces certain shortcomings. Our network needs to be pre-trained before training and is not one-stage. Moreover, the number of vectors for class representation calculation is kept as large as possible. This is because the robustness of class representations are affected by two factor: one is the network performance of the current epoch and the other is the number of class representation vectors involved in the calculation.

Our future work is to achieve more robust class representations formed by the networ architecture exploring and the number of output channel, to better guide the generation of pseudo labels. At the same time, more strategies [31] should be adopted to adjust our framework into a more flexible single phase. Besides, the lightweight network is also a potential research topic when data is insufficent [32].

V. Conclusion

This paper introduces a novel semi-supervised framework for liver image segmentation under the low-labeled images paradigm, which generates high-quality pseudo labels for unlabeled images by utilizing the guidance from labeled images. To reduce noises and improve the quality of pseudo labels, a series of operations are adopted to optimize pseudo labels. Besides, a novel random patch based on prior locations is proposed for unlabeled images during the training. Our experiments on LiTS dataset substantiate that our method can obtain a competitive performance over other models for liver segmentation.

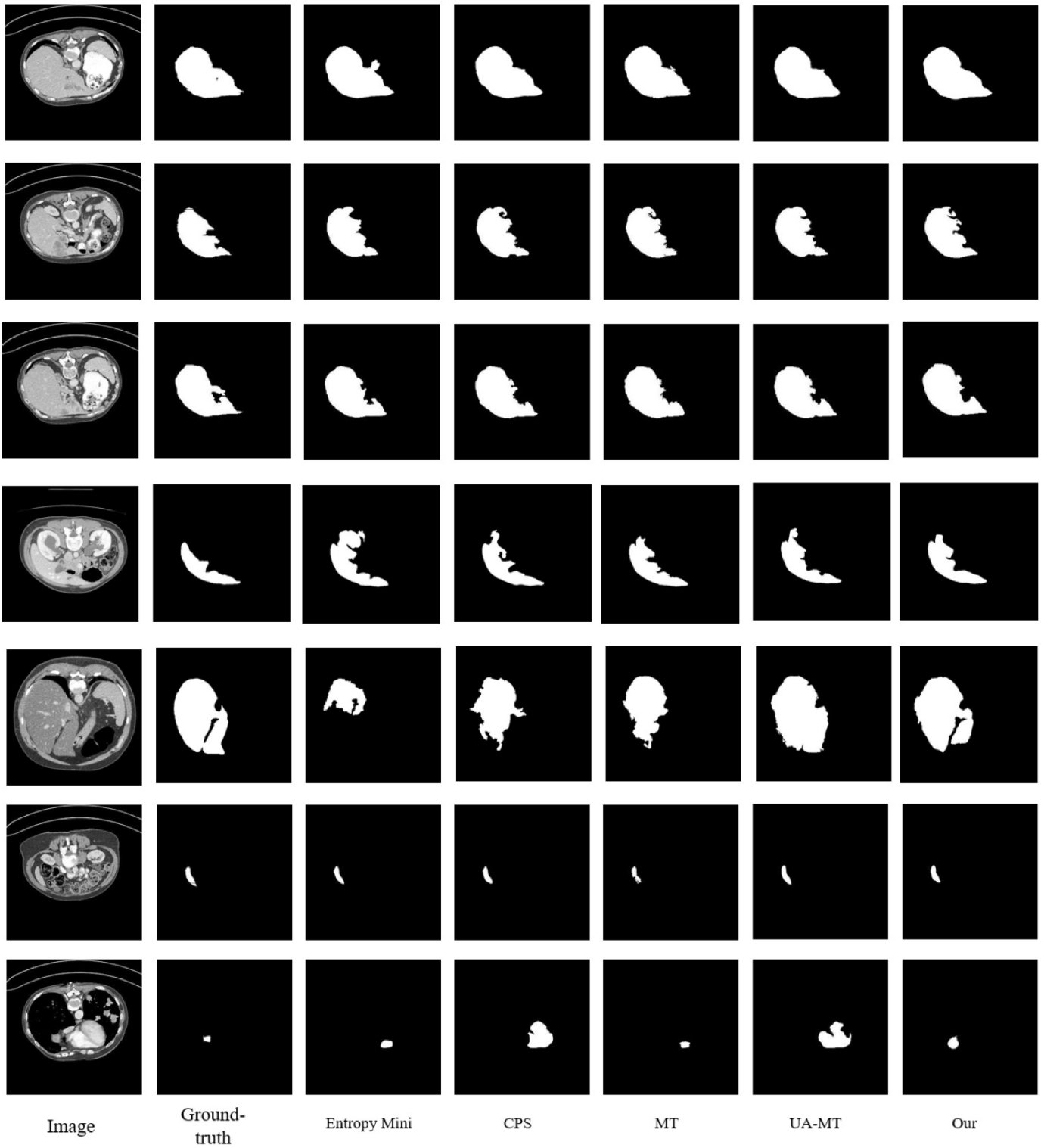

Fig. 8: The visual results of various methods.

REFERENCES

[1] Q. Yu, L. Xie, Y. Wang, Y. Zhou, E. K. Fishman, and A. L. Yuille, "Recurrent saliency transformation network: Incorporating multi-stage visual cues for small organ segmentation," in *Proceedings of the IEEE conference on computer vision and pattern recognition*, 2018, pp. 8280–8289.

[2] J. Long, E. Shelhamer, and T. Darrell, "Fully convolutional networks for semantic segmentation," in *Proceedings of the IEEE conference on computer vision and pattern recognition*, 2015, pp. 3431–3440.

[3] V. Badrinarayanan, A. Kendall, and R. Cipolla, "Segnet: A deep convolutional encoder-decoder architecture for image segmentation," *IEEE transactions on pattern analysis and machine intelligence*, vol. 39, no. 12, pp. 2481–2495, 2017.

[4] X. J. Zhu, "Semi-supervised learning literature survey," 2005.

[5] W. Bai, O. Oktay, M. Sinclair, H. Suzuki, M. Rajchl, G. Tarroni, B. Glocker, A. King, P. M. Matthews, and D. Rueckert, "Semi-supervised learning for network-based cardiac mr image segmentation," in *International Conference on Medical Image Computing and Computer-Assisted*

TABLE IV: The comparison with other semi-supervised segmentation methods.

Method	labeled data	unlabeled data	Dice	Jaccard	Precision	Recall	Parameters
Entropy Mini [28]	100	900	82.98%	74.21%	78.18%	80.34%	1.81M
CPS [29]	100	900	81.59%	73.64%	80.91%	81.46%	3.63M
MT [30]	100	900	83.02%	75.83%	79.01%	79.25%	3.63M
UA-MT [9]	100	900	84.37%	73.12%	80.58%	**82.01%**	3.63M
Our proposed	100	900	**84.85%**	**77.80%**	**81.93%**	81.65%	1.81M
Entropy Mini [28]	200	800	83.56%	78.40%	80.71%	82.49%	1.81M
CPS [29]	200	800	84.71%	80.48%	81.04%	83.82%	3.63M
MT [30]	200	800	85.58%	79.13%	80.94%	81.94%	3.63M
UA-MT [9]	200	800	85.08%	81.90%	82.71%	82.56%	3.63M
Our proposed	200	800	**86.83%**	**82.03%**	**83.51%**	**83.70%**	1.81M

Intervention. Springer, 2017, pp. 253–260.

[6] D.-P. Fan, T. Zhou, G.-P. Ji, Y. Zhou, G. Chen, H. Fu, J. Shen, and L. Shao, "Inf-net: Automatic covid-19 lung infection segmentation from ct images," *IEEE Transactions on Medical Imaging*, vol. 39, no. 8, pp. 2626–2637, 2020.

[7] Y. Zhou, Y. Wang, P. Tang, S. Bai, W. Shen, E. Fishman, and A. Yuille, "Semi-supervised 3d abdominal multi-organ segmentation via deep multi-planar co-training," in *2019 IEEE Winter Conference on Applications of Computer Vision (WACV)*. IEEE, 2019, pp. 121–140.

[8] J. Peng, G. Estrada, M. Pedersoli, and C. Desrosiers, "Deep co-training for semi-supervised image segmentation," *Pattern Recognition*, vol. 107, p. 107269, 2020.

[9] L. Yu, S. Wang, X. Li, C.-W. Fu, and P.-A. Heng, "Uncertainty-aware self-ensembling model for semi-supervised 3d left atrium segmentation," in *International Conference on Medical Image Computing and Computer-Assisted Intervention*. Springer, 2019, pp. 605–613.

[10] Y. Wang, Y. Zhang, J. Tian, C. Zhong, Z. Shi, Y. Zhang, and Z. He, "Double-uncertainty weighted method for semi-supervised learning," in *International Conference on Medical Image Computing and Computer-Assisted Intervention*. Springer, 2020, pp. 542–551.

[11] Y. Xia, F. Liu, D. Yang, J. Cai, L. Yu, Z. Zhu, D. Xu, A. Yuille, and H. Roth, "3d semi-supervised learning with uncertainty-aware multi-view co-training," in *Proceedings of the IEEE/CVF Winter Conference on Applications of Computer Vision*, 2020, pp. 3646–3655.

[12] Z. Ke, D. Wang, Q. Yan, J. Ren, and R. W. Lau, "Dual student: Breaking the limits of the teacher in semi-supervised learning," in *Proceedings of the IEEE/CVF International Conference on Computer Vision*, 2019, pp. 6728–6736.

[13] X. Jin, H. Ye, L. Li, and Q. Xia, "Image segmentation of liver ct based on fully convolutional network," in *2017 10th international symposium on computational intelligence and design (ISCID)*, vol. 1. IEEE, 2017, pp. 210–213.

[14] Z. Liu, K. Han, Z. Wang, J. Zhang, Y. Song, X. Yao, D. Yuan, and V. S. Sheng, "Automatic liver segmentation from abdominal ct volumes using improved convolution neural networks," *Multimedia Systems*, vol. 27, no. 1, pp. 111–124, 2021.

[15] X. Li, H. Chen, X. Qi, Q. Dou, C.-W. Fu, and P.-A. Heng, "H-denseunet: hybrid densely connected unet for liver and tumor segmentation from ct volumes," *IEEE transactions on medical imaging*, vol. 37, no. 12, pp. 2663–2674, 2018.

[16] C. Li, Y. Tan, W. Chen, X. Luo, Y. Gao, X. Jia, and Z. Wang, "Attention unet++: A nested attention-aware u-net for liver ct image segmentation," in *2020 IEEE International Conference on Image Processing (ICIP)*. IEEE, 2020, pp. 345–349.

[17] O. Ronneberger, P. Fischer, and T. Brox, "U-net: Convolutional networks for biomedical image segmentation," in *International Conference on Medical image computing and computer-assisted intervention*. Springer, 2015, pp. 234–241.

[18] P. F. Christ, F. Ettlinger, F. Grün, M. E. A. Elshaera, J. Lipkova, S. Schlecht, F. Ahmaddy, S. Tatavarty, M. Bickel, P. Bilic *et al.*, "Automatic liver and tumor segmentation of ct and mri volumes using cascaded fully convolutional neural networks," *arXiv preprint arXiv:1702.05970*, 2017.

[19] Y. Xu, B. Du, F. Zhang, and L. Zhang, "Hyperspectral image classification via a random patches network," *ISPRS journal of photogrammetry and remote sensing*, vol. 142, pp. 344–357, 2018.

[20] D.-H. Lee *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, no. 2, 2013, p. 896.

[21] H. Yang, C. Shan, R. A. Bouwman, L. Dekker, A. Kolen, and P. H. De With, "Medical instrument segmentation in 3d us by hybrid constrained semi-supervised learning," *IEEE Journal of Biomedical and Health Informatics*, 2021.

[22] Y. Zhang and J. Zhang, "Dual-task mutual learning for semi-supervised medical image segmentation," in *Chinese Conference on Pattern Recognition and Computer Vision (PRCV)*. Springer, 2021, pp. 548–559.

[23] X. Li, L. Yu, H. Chen, C.-W. Fu, L. Xing, and P.-A. Heng, "Transformation-consistent self-ensembling model for semisupervised medical image segmentation," *IEEE Transactions on Neural Networks and Learning Systems*, vol. 32, no. 2, pp. 523–534, 2020.

[24] X. Cao, H. Chen, Y. Li, Y. Peng, S. Wang, and L. Cheng, "Uncertainty aware temporal-ensembling model for semi-supervised abus mass segmentation," *IEEE transactions on medical imaging*, vol. 40, no. 1, pp. 431–443, 2020.

[25] X. Luo, W. Liao, J. Chen, T. Song, Y. Chen, S. Zhang, N. Chen, G. Wang, and S. Zhang, "Efficient semi-supervised gross target volume of nasopharyngeal carcinoma segmentation via uncertainty rectified pyramid consistency," in *International Conference on Medical Image Computing and Computer-Assisted Intervention*. Springer, 2021, pp. 318–329.

[26] C. You, R. Zhao, L. Staib, and J. S. Duncan, "Momentum contrastive voxel-wise representation learning for semi-supervised volumetric medical image segmentation," *arXiv preprint arXiv:2105.07059*, 2021.

[27] Y. Wu, M. Xu, Z. Ge, J. Cai, and L. Zhang, "Semi-supervised left atrium segmentation with mutual consistency training," in *International Conference on Medical Image Computing and Computer-Assisted Intervention*. Springer, 2021, pp. 297–306.

[28] T.-H. Vu, H. Jain, M. Bucher, M. Cord, and P. Pérez, "Advent: Adversarial entropy minimization for domain adaptation in semantic segmentation," in *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, 2019, pp. 2517–2526.

[29] X. Chen, Y. Yuan, G. Zeng, and J. Wang, "Semi-supervised semantic segmentation with cross pseudo supervision," in *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, 2021, pp. 2613–2622.

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

[31] A. Miller, J. Panneerselvam, and L. Liu, "A review of regression and classification techniques for analysis of common and rare variants and gene-environmental factors," *Neurocomputing*, 2021.

[32] X. Zhu, K. Guo, S. Ren, B. Hu, M. Hu, and H. Fang, "Lightweight image super-resolution with expectation-maximization attention mechanism," *IEEE Transactions on Circuits and Systems for Video Technology*, 2021.