CTANet: Confidence-based Threshold Adaption Network for Semi-supervised Segmentation of Uterine Regions from MR Images for HIFU Treatment

Chen Zhang, Guanyu Yang, Faqi Li, Yingang Wen, Yuhan Yao, Huazhong Shu, Antoine Simon, Jean-Louis Dillenseger, Jean-Louis Coatrieux

CTANet: Confidence-based Threshold Adaption Network for Semi-supervised Segmentation of Uterine Regions from MR Images for HIFU Treatment

Chen Zhang[a,b,c], Guanyu Yang[a,b,c], Faqi Li[d], Yingang Wen[e], Yuhan Yao[a,b,c], Huazhong Shu[a,b,c,d,*], Antoine Simon[c,f], Jean-Louis Dillenseger[c,f], Jean-Louis Coatrieux[c,f]

[a] *LIST, Key Laboratory of Computer Network and Information Integration (Southeast University), Ministry of Education, Nanjing 210096, China*
[b] *Jiangsu Provincial Joint International Research Laboratory of Medical Information Processing, Southeast University, Nanjing, 210096, China*
[c] *Centre de Recherche en Information Biomédicale Sino-Français (CRIBs), Rennes, F-35000, France*
[d] *State Key Laboratory of Ultrasound in Medicine and Engineering, Chongqing Medical University, Chongqing, 400016, China*
[e] *National Engineering Research Center of Ultrasound Medicine, Chongqing, 401121, China*
[f] *Univ Rennes, Inserm, LTSI - UMR1099, Rennes, F-35000, France*

Abstract

Objectives: The accurate preoperative segmentation of the uterus and uterine fibroids from magnetic resonance images (MRI) is an essential step for diagnosis and real-time ultrasound guidance during high-intensity focused ultrasound (HIFU) surgery. Conventional supervised methods are effective techniques for image segmentation. Recently, semi-supervised segmentation approaches have been reported in the literature. One popular technique for semi-supervised methods is to use pseudo-labels to artificially annotate unlabeled data. However, many existing pseudo-label generations rely on a fixed threshold used to generate a confidence map, regardless of the proportion of unlabeled and labeled data.

Materials and Methods: To address this issue, we propose a novel semi-supervised framework called Confidence-based Threshold Adaptation Network (CTANet) to improve the quality of pseudo-labels. Specifically, we propose an online pseudo-labels method to automatically adjust the threshold, producing high-confident unlabeled annotations and boosting segmentation accuracy. To further improve the network's generalization to fit the diversity of different patients, we design a novel mixup strategy by regularizing the network on each layer in the decoder part and introducing a consistency regularization loss between the outputs of two sub-networks in CTANet.

Results: We compare our method with several state-of-the-art semi-supervised

*Corresponding authors.
Email address: shu.list@seu.edu.cn (Huazhong Shu)

segmentation methods on the same uterine fibroids dataset containing 297 patients. The performance is evaluated by the Dice similarity coefficient, the precision, and the recall. The results show that our method outperforms other semi-supervised learning methods. Moreover, for the same training set, our method approaches the segmentation performance of a fully supervised U-Net (100% annotated data) but using 4 times less annotated data (25% annotated data, 75% unannotated data).
Conclusion: Experimental results are provided to illustrate the effectiveness of the proposed semi-supervised approach. The proposed method can contribute to multi-class segmentation of uterine regions from MRI for HIFU treatment.

Keywords: HIFU therapy, semi-supervised segmentation,
threshold-adaptation, uterine fibroids

---

1. Introduction

Uterine fibroids are common benign clonal neoplasms of the uterus and are clinically apparent in 20-40% of child-bearing women [1]. They may cause persistent menstrual bleeding and iron deficiency anemia [2]. Surgical therapy has been the primary treatment for uterine fibroids, and hysterectomy eliminates symptoms and reduces the risk of recurrence [3]. However, patients with uterine fibroids-related infertility who strictly refuse surgical treatment or who have very high or unacceptable surgical risks can only try other surgical options.

High-intensity focused ultrasound (HIFU) is a high-precision medical procedure for local heating and ablation of diseased tissue and has been widely used to treat uterine fibroids. Compared to other surgical therapies, HIFU has the advantages of being non-invasive and having a low number of registered complications. HIFU can be considered as a promising treatment option for women who wish to conceive a child [4].

HIFU can be guided either by MRI (Magnetic resonance-guided HIFU - MRgHIFU) or ultrasound (Ultrasound-guided HIFU - USgHIFU). The clinical use of MRgHIFU is limited due to the need for a dedicated MR device to guide the treatment and the duration of the procedure [5]. In this paper, we explore USgHIFU, which provides real-time performance, but the low resolution of the ultrasound itself may make the surgery less accurate. Therefore, the segmentation of the uterine regions from preoperative MR is crucial to determine the exact position of the targeted fibroid and the adjacent organs in order to guide the HIFU procedure. In uterus region segmentation from MR images, the doctors need to manually annotate, slice by slice, the contours of the uterus, uterine fibroids, and organs at risk (OAR) such as the spine. Therefore, automatic segmentation of uterine regions could therefore free physicians from this tedious and time-consuming work.

Recently, deep learning (DL) has made tremendous progress in medical image segmentation, such as U-Net [6] and V-Net [7]. These fully-supervised learning (FSL)-based methods can handle the medical image segmentation tasks.

However, the accuracy and robustness of the DL methods rely on a large number of learning data annotated by experts. Acquiring good and accurate annotations requires laborious work, and inter-expert delineation results vary.

Unlike the above FSL methods, semi-supervised learning (SSL) can take advantage of large numbers of unlabeled data to improve network performance when labeled data is insufficient. One popular SSL approach is to adopt consistency learning, which regularizes the network to be consistent with the predictions of perturbation [8]. Another common way for SSL is pseudo-labeling by producing an artificial label for unlabeled images. Specifically, a pre-trained model is first used on a small number of labeled data. Then, the unlabeled data are fed to the model, and the class with the maximum predicted probability is selected and called pseudo-label. After that, the labeled data are co-trained with the pseudo-labeled data. The above procedures are repeated to make the model more efficient. The effectiveness of this mechanism is similar to that of entropy regularization. This pseudo-labeling method can combine almost all neural network models and training methods. However, it has been shown that the use of Denoising Auto-Encoder and Dropout gave very good results [9].

Because of the lack of annotated data, some data augmentation methods have been considered in SSL. One simple and effective method is Mixup [10], which can improve the generalization and robustness of the model by mixing the data pairs. This strategy can also be interpreted as an empirical risk minimization on data modified with random perturbations [11]. Based on the Mixup, CutMix [12], and Mixmatch [13] were developed to further improve the performance of the SSL.

However, the quality of pseudo-labels will affect the optimization of the model when it is updated. In the existing methods, the generation of pseudo-labels relies on time-consuming artificial offline selection, usually based on experience or after experiments on a small validation set. Then a threshold is set to generate a credible confidence map. This threshold is usually determined on an ad-hoc basis and is not generalizable. We believe that an adaptive threshold strategy needs to be developed for an automatic adaptation to different semi-supervised data distributions. This should improve the generalization and robustness of the semi-supervised methods. In addition, we also plan to extend the regularization in the Mixup to adapt the framework to more complex semi-supervised medical image segmentation tasks. To this end, we propose a novel approach named Confidence-based Threshold Adaptation Network (CTANet). The contributions of our method are as follows:

- CTANet introduces a novel online threshold-adaptation strategy to generate high-confidence maps for pseudo-labels, improving the model performance even under different ratios of annotated to unannotated data numbers.

- CTANet adopts an augmentation method named Hidden Mixup inspired by Mixup to improve the generalization under different patient data distribution.

- CTANet contains two networks in which consistency regularization and dropout are used to address the overfitting issue.
- To the best of our knowledge, CTANet is the first network that investigates the problem of semi-supervised learning for uterine regions segmentation for HIFU surgery.
- Experiments conducted on a uterus dataset indicate that CTANet is better than state-of-the-art (SOTA) semi-supervised methods and has the potential to generalize well to other segmentation tasks.

2. Related work

2.1. Segmentation Methods of Uterus and Uterine fibroids

Some conventional methods for uterus and uterine fibroids segmentation have been reported in the literature. Methods based on level sets [14, 15] usually rely on user feedback and intervention. Fuzzy C-Means [16, 17] and Region-growing [18, 19] based methods suffer from the well-known problem of sensitivity to noise. Also, these methods are not efficient on large-scale datasets and are limited by their computational time and power requirements.

Recently, with the development of deep learning, some CNN-based automatic segmentation methods have been applied to the uterine fibroid segmentation task. Kurata *et al.* [20] used an optimized U-Net to segment the uterus on T2-weighted MR images. Zhang *et al.* [21] proposed a large kernel Encoder-Decoder Network based on a 2D segmentation model named HIFUNet to segment all uterine regions from MR images. Ning *et al.* [22] developed a multistage segmentation network to segment the fibroids. In the first stage, a classification network is used to select the slices with existing fibroids. Then, the processed images are fed into the second stage to remove the unrelated regions in preparation for the final fibroids segmentation in the third stage. Behboodi *et al.* [23] combined 2D U-Net and MobileNet-v2 [24] to make the first fully-automatic segmentation try in ultrasound uterus images. Generally, these methods rely on very large amounts of annotated data to perform segmentation.

2.2. Semi-supervised Learning

The emergence of SSL avoids the dependence on large amounts of labeled data as in traditional FSL neural networks. Therefore, the SSL approach is closer to the real-world application. The main idea is to improve the model performance by training it on the available labeled data and then using unlabeled data as constraints. The algorithms can be divided into 4 categories: graph-based, Generative Adversarial Networks(GAN)-based, self or co-training, and consistency training methods.

Graph-based methods: The advantage of graphs is that they are convex, scalable, and efficient for modeling relationships between different entities [25]. Kipf *et al.* [26] presented an approach to use graph topology and node side information for semi-supervised classification. Wang *et al.* [27] proposed the mixed-order

graph convolutional networks (MOGCN) to improve the performance of semi-supervised node classification by employing a novel ensemble module.

Generative Adversarial Networks (GAN)-based methods: Berthelot *et al.* proposed GAN [13] to learn deep representations and capture statistical distribution by playing two neural networks (a Generative network and an Adversarial network) against each other. GAN-based SSL methods can generate a perfect discriminator by learning imperfect generators on labeled and unlabeled data. Sricharan *et al.* [28] proposed the semi-supervised GAN (SS-GAN) to conditionally generate data-given attributes by using a pair of stacked discriminators. Hung *et al.* [29] applied adversarial learning for semi-supervised semantic segmentation by combining two semi-supervised loss terms to leverage the unlabeled data.

Self-training or co-training: Self-training or co-training is a proxy label method that produces proxy labels on unlabeled data without supervision. Self-training methods use labeled data to pre-train a model and then predict pseudo-labels to unlabeled data. Yalniz *et al.* [30] trained a teacher-student model to exploit the large-scale unlabeled data and gained a 4.8% accuracy improvement compared to ResNet-50. Similarly, Xie *et al.* [31] designed a simple and efficient teacher-student model based on EfficientNet to improve ImageNet classification. Chen *et al.* [32] considered the global shape constraint in anatomical landmarks and proposed a model-agnostic shape-regulated self-training framework.

Co-training was originally proposed to describe a model where a set of unlabeled data is used to augment a smaller set of labeled data [33]. Inspired by this idea, Qiao *et al.* [34] extended co-training to deep co-training for semi-supervised image recognition. Especially, adversarial examples are used in different views to prevent a model from collapsing. Peng *et al.* [35] applied co-training to the image segmentation problem, and experiments conducted on three public medical image datasets showed the robustness and generalization of their method. They also added the adversarial learning technique to enforce diversity in the ensemble models.

Consistency training: Consistency regularization can obtain similar output results for the same input with different data enhancements or networks. Π-Model and temporal ensembling [8] are typical implementations of consistency regularization. However, temporal ensembling has its limitations for large datasets because each target is updated once per epoch. Tarvainen *et al.* [36] overcome the problem by averaging the model weights instead of the predictions. The method is called Mean Teacher, which includes two networks: the teacher network and the student network. The two networks have the same structure but are updated in different ways. The student network updates parameters by back-propagating gradient descent and the teacher network updating parameters by exponential moving average (EMA) of the student network parameters. Compared with the temporal ensembling, Mean Teacher can update the moving average of the network parameters once per backpropagation, which is more efficient. Ouali *et al.* [37] proposed a cross-consistency training (CCT) network, where an invariance of the predictions is enforced over different perturbations applied to the outputs of the encoder. In addition, adopting adversarial learn-

ing can enforce the distributions of segmentation of unannotated images to be similar to those of the annotated images.

2.3. Semi-supervised Learning for Medical Image Segmentation

Semi-supervised learning approaches have been widely applied to medical image segmentation tasks. Nie *et al.* [38] employed an adversarial network named ASDNet to produce high-confidence unannotated data to train the segmentation network. Li *et al.* [39] used transformation consistency learning to do the semi-supervised skin lesion segmentation and got a new record in the ISIC2017 skin lesion segmentation challenge. Bai *et al.* [40] proposed an SSL framework in which the pseudo labels of the unlabeled data were obtained by non-rigid image registration in the cardiac cycle. This method was evaluated on a short-axis cardiac MR image dataset and obtained average Dice values of 0.92, 0.85, and 0.89 for the left ventricular (LV) cavity, LV myocardium, and right ventricular (RV) cavity, respectively. Similarly, Ito *et al.* [41] also used the registration-based SSL method to achieve the brain tissue segmentation and evaluated it on human and marmoset brain image datasets to show the effectiveness of the method.

Compared to existing semi-supervised learning methods, our work differs in the following aspects: (1) It is the first effective work to segment uterine regions using semi-supervised methods. Thus, we address a segmentation challenge due to the ambiguous boundaries of uterine fibroids and the fact that they are highly variable in size and shape. Therefore, we design a novel Mixup strategy in hidden layers to maintain the robustness of the model against perturbations brought by the "noisy" background regions or unreliable pseudo labels. In addition, the feature space of different layers is "smoothed" to improve the model's generalization and extract more multi-scale features. (2) A confidence-based adaptive thresholding module is proposed for the multi-class segmentation problem in the uterine region. The existing segmentation algorithms adopt time-consuming ad-hoc threshold selection (or empirically set it to 0.5) in the pseudo-label generation. This fixed threshold ignores the differences among patients and segmented targets. Our method can automatically and adaptively perform pixel-level weighting, resulting in more attention to regions with high weights. In addition, the module prompts the model to learn features ranging from easy (spine) to challenging (uterine fibroids).

3. Methodology

In this paper, we aim to take advantage of unannotated data by optimizing the generation of pseudo-labels and improving the generalization capability of the model. In this section, we first introduce the main structure of CTANet, then describe the adopted confidence-based threshold adaptation strategy, and finally, we describe four essential components of the framework: Confidence-based threshold adaption module, Hidden Mixup augmentation, consistency regularization and dropout.

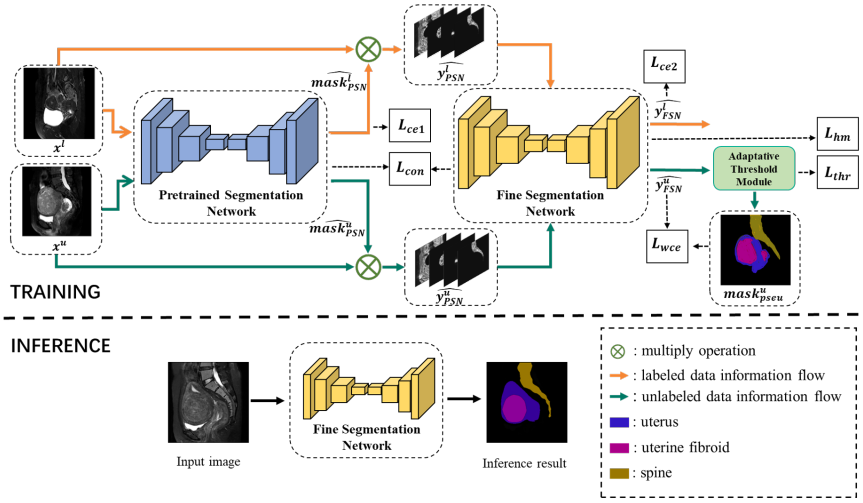

Figure 1: The framework of the proposed CTANet. It consists of a Pretrained Segmentation Network (PSN) and a Fine Segmentation Network (FSN).

3.1. Overview of CTANet

The training pipeline: The proposed CTANet is shown in Fig.(1) which consists of a Pre-trained Segmentation Network (PSN) and a Fine Segmentation Network (FSN). We use the simple and efficient CNN model U-Net as the backbone. These two base networks are both trained from scratch. Note that for the cascade between the two networks, we modify the output of the PSN. The output of PSN consists of the probability values of the four channels (One per class: background, uterus, fibroid, and spine). And each of these probability maps is multiplied using a dot-product by the input raw image so that each channel has only the region of interest for the current category. We now delve into the details.

We first train the PSN on labeled data x^l supervised by the cross-entropy loss L_{ce1} between the predicted outputs of PSN $\widehat{mask^l_{PSN}}$ and the ground-truth annotated by experts $mask^l_{GT}$. In the second stage, FSN is also trained on the labeled data x^l supervised by the cross-entropy loss L_{ce2} between the predicted outputs of FSN $\widehat{mask^l_{FSN}}$ and $mask^l_{GT}$. Similarly, for the unlabeled image x^u, we use $\widehat{mask^u_{PSN}}$ and $\widehat{y^u_{FSN}}$ to define the outputs of two networks. Specifically, $\widehat{mask^l_{PSN}}$ and $\widehat{mask^u_{PSN}}$, which are four-channel probability maps, are respectively associated by a dot product with x^l, x^u to generate $\widehat{y^l_{PSN}}$ and $\widehat{y^u_{PSN}}$. In the second stage, $\widehat{y^u_{FSN}}$ is produced by feeding $\widehat{y^u_{PSN}}$ into the FSN. The pseudo-labels for x^u generated by FSN are used to optimize PSN by the weighted cross-entropy loss L_{wce}. However, pseudo-labels with low confidence must be discarded from the optimization process. For this, usually a confidence threshold T is set. The network is trained only if prediction confidence is over

T. In most of the paper, this threshold is fixed at the start. In our case, the threshold T is designed on the loss function in L_{thr}, which continuously optimizes the estimated global threshold as training proceeds, thus continually improving the quality of the pseudo-label.

The two networks carry out cooperative training and constantly update under four losses: the Hidden Mixup loss L_{hm}, the weighted cross-entropy loss L_{wce}, the threshold loss L_{thr}, and consistency regularization loss L_{con}. Specifically, L_{hm} is used to improve the generalization performance of the model, L_{wce} and L_{thr} are related to the loss of unlabeled data to achieve the automatic online generation of high-confidence pseudo-labels and L_{con} helps the two networks to produce similar reference results. These losses will be detailed later on.

The inference pipeline: The bottom part of Fig.(1) shows the inference process. Simply, a new input image to be segmented is fed into the previously trained FSN to obtain the inference result.

3.2. Confidence-based Threshold Adaptation

In the pseudo-label generation, the previous methods rely on grid search or experience, which are very time-consuming and have limitations in the multi-class medical image segmentation tasks. Inspired by [42], which was the first attempt of an online weighted pseudo-label in unsupervised domain adaptation (UDA), we introduce an adaptive method as an alternative to the manual grid search method to obtain the threshold T. Specifically, for each FSN output pixel, if it is lower than T, we set the weight of this pixel as 0. On the contrary, if the pixel output of FSN is higher than T, the weight of this pixel will be calculated by:

$$\omega = \begin{cases} \frac{\max\left(\widehat{p^u_{FSN}}\right)-T}{1-T} & \text{if } \max\left(\widehat{p^u_{FSN}}\right) \geq T \\ 0 & \text{otherwise} \end{cases} \tag{1}$$

where $\max(\widehat{p^u_{FSN}})$ refers to the maximum confidence value for each pixel $\widehat{p^u_{FSN}}$ in $\widehat{y^u_{FSN}}$ andω represents the pixel-wise weights of the pseudo-labels. Notice that $\widehat{y^u_{FSN}}$ is actually a feature map containing multiple channels, where each channel represents a segmentation category. In this way, the pixels with higher confidence can be used to calculate the loss, while a lower ones will be discarded. By using pixel-by-pixel weighting, the network can pay more attention to pixels with correct predictions in pseudo-labels, and reduce the negative impact of pixels with inaccurate predictions. The loss function of unlabeled data, namely L_u, is defined by:

$$L_u = L_{wce} + L_{thr} \tag{2}$$

where L_{wce} is the weighted cross-entropy loss function, and L_{thr} is the adaptive threshold loss function. They are given by:

$$L_{wce} = -\frac{1}{N}\sum_{i\in I}\omega p^{u}_{pseu,i}\log\left(\widehat{p^{u}_{FSN}}\right) \tag{3}$$

$$L_{thr} = \log^2(1-T) \tag{4}$$

Here N is the number of pixels in one image and $p^{u}_{pseu,i}$ represents each pixel in one pseudo-label generated by the FSN. I is the number of channels. For the initialization of T, in the early training process, we first chose a threshold value that is high enough to quickly get a good result for easy-to-segment targets. As the training process proceeds, the threshold value is gradually reduced so that high weights are learned for the hard-to-segment pixels. Considering the training efficiency, we set the initial value of T to 0.8, which provides acceptable results.

3.3. Hidden Mixup

Mixup [10] aims to improve the network generalization by a linear combination of paired input data and their labels. Recently, [43] extended this regularization strategy to both the input space and latent space to regularize different parts of the network. Considering the semi-supervised multi-category image segmentation task, we should use limited labels to generate more data to avoid over-fitting and achieve generalization over different patients.

Hidden Mixup in FSN achieves this goal by regularizing the output of each layer in the decoder part. The Hidden Mixup loss (Fig.(2)) is computed in the hidden layers of FSN. For the encoder input in FSN, we first generate a multiple-channel attention map $\widehat{y^{l}_{PSN}}$ by multiplying the labeled image x^l with its predicted result of PSN defined as $\widehat{mask^{l}_{PSN}}$, then the unlabeled image gets its corresponding attention map $\widehat{y^{u}_{PSN}}$. After that, we linearly mix the two attention maps as follows:

$$\lambda \sim \mathrm{Beta}(\alpha,\alpha) \tag{5}$$

$$\lambda' = \max(\lambda, 1-\lambda) \tag{6}$$

$$\widehat{y^{mix}} = \lambda'\widehat{y^{l}_{PSN}} + (1-\lambda')\,\widehat{y^{u}_{PSN}} \tag{7}$$

$$z^{mix} = \lambda' z^{l} + (1-\lambda')\,z^{u} \tag{8}$$

where Beta is a Beta distribution with α its positive shape parameter. α is considered as a hyperparameter in this work. In (8), z^l is x^l multiplied by $mask^{l}_{GT}$ and z^u is x^u multiplied by $mask^{u}_{pseu}$. $\widehat{y^{mix}}$ and z^{mix} are respectively

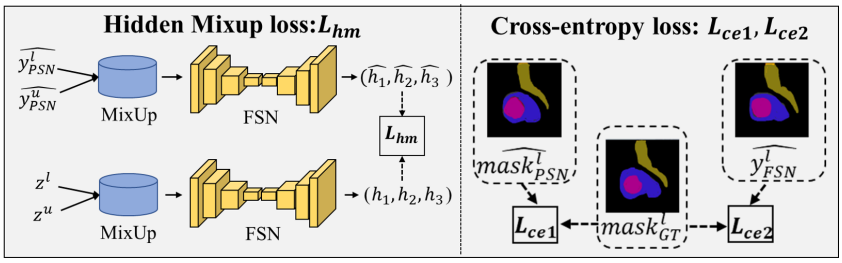

Figure 2: The illustration of the Hidden Mixup loss (L_{hm}) and cross-entropy loss in the CTANet.

sent into the FSN, and their outputs at each layer k of the decoder are marked as two sets $\hat{H} = \left\{\hat{h_k}\right\}$ and $H = \{h_k\}$ where $1 \leq k \leq K$. K is the depth of FSN. Here we set K to 3.
Cross entropy loss function is used for computing Hidden Mixup:

$$L_{hm} = -\sum_{k=1}^{K} h_k \log\left(\widehat{h_k}\right) \tag{9}$$

3.4. Consistency Regularization and Dropout

Due to the existence of pseudo-labels, some noise will inevitably be introduced. Therefore, regularization plays an important role in our task. The purpose of consistency regularization is to ensure that the sample and the extended version of its network prediction have the same conceptual meaning as possible in the problem.

The dropout layer is another technique to prevent our model from overfitting. It randomly drops neurons from the network during training. The consistency regularization loss is defined as:

$$L_{\text{con}} = 0.5^* D_{KL}\left(\widehat{y_{PSN}} \| \widehat{y_{FSN}}\right) + 0.5^* D_{KL}\left(\widehat{y_{FSN}} \| \widehat{y_{PSN}}\right) \tag{10}$$

where $\widehat{y_{PSN}}$ and $\widehat{y_{FSN}}$ represent the prediction outputs of PSN and FSN for both labeled and unlabeled data, respectively.

3.5. Loss function

The training of CTA-Net is divided into three steps: 1) the PSN is trained with a limited proportion of labeled data; 2) then the parameters are shared to the FSN; 3) the whole network is trained with all training data, including labeled and unlabeled data. The loss function is as follows:

$$L = L_l + L_u + L_{hm} + L_{con} \tag{11}$$

where the L_u, L_{hm}, L_{con} were introduced in sections 3.2, 3.3 and 3.4 respectively. L_l is the loss function of the labeled data and is composed of two standard cross-entropy loss functions (Fig.(2)):

$$L_l = 0.5 * (L_{ce1} + L_{ce2}) \tag{12}$$

Here L_{ce1} is computed between the PSN for the output of the labeled image $\widehat{mask^l_{PSN}}$ and its ground-truth $mask^l_{PSN}$, and L_{ce2} is computed using the FSN for the output of the labeled image y^l_{PSN} and $mask^l_{PSN}$. Under the supervision of the labeled loss function, both PSN and FSN can improve the predicted segmented regions.

4. Experimental configurations

4.1. Data Description

The HIFU dataset was collected at the State Key Laboratory of Ultrasound in Medicine and Engineering (Chongqing Medical University, Chongqing, China). Sagittal T2-weighted images were performed using a 3.0-T MRI system (Signa HD, GE Healthcare, Milwaukee, WI, USA). The standardized parameters of the T2WI sequence were as follows: repetition time (TR) 3030 ms, echo/time (TE) 107.5 ms, slice thickness 6 mm, slice gap 1 mm. The median age of the patients was 40.8 years.

The MR dataset is a uterine fibroids dataset containing 297 labeled 3D fat-suppressed T2-weighted MRI scans with the uterus, uterine fibroids, and spine. Each MR volume consists of 25 slices of 304×304 pixels. We split them into 260 scans for training and the remaining 37 scans for testing. The ground-truth was manually annotated and confirmed by two radiologists through a proper annotation process to ensure the consensus agreement of the annotation.

The ethics committee approved the study at Chongqing Medical University. The patients signed an informed consent form before each procedure.

4.2. Experimental Setup and details

The framework is implemented using Pytorch and trained by the Adam optimizer. In the training data, the segmentation models of PSN and FSN are called M1, M2, respectively. Firstly, M1 is pre-trained with labeled data for 50 epochs, and the model parameters are saved. We then import the pre-trained model parameters and train M1 and M2 with both labeled and unlabeled data for 50 epochs. The training set is further divided into training and validation sets in an 8:2 ratio. For example, if 10% of the training data is used as labeled data, we use data from 26 patients for training (21 patients as the training set and 5 patients as the validation set). The data division is randomized and different slices of the same patient do not appear in both the training or validation sets. All the models are trained with an initial learning rate of 0.001, which decays by 0.5 after every 10 epochs. The best model is saved based on the validation accuracy, and then the best weights are saved to infer the test data. The Mixing

parameter α for the datasets is set to 1.0. The batch size is 4. The computation was performed on an RTX 2080Ti GPU.

The data and hyperparameters are fixed during the comparison experiments and ablation experiments.

4.3. Evaluation Criteria

To evaluate the performance of the segmentation, we employed some of the most commonly used metrics such as the Dice similarity coefficient (Dice or DSC), precision (PR), and recall (RE).

4.4. Comparison with Other Deep Learning Methods

We compared our CTANet with four SOTA semi-supervised learning approaches, including ASDNet [38], Latent Mixup [43], and Cross-Consistency Training (CCT) [37], this for 3 different labeled/unlabeled data ratios. Besides, we added two fully-supervised methods: the classical Vanilla U-Net [6] and HIFUNet [21] with the whole set of labeled data as the performance reference. All the experiments were conducted in a fair way with the same training, test data, and network hyperparameters.

For quantitative comparison, Table (1) shows the Dice, PR, and RE indices obtained on the HIFU dataset by the different methods. In order to test the impact of the ratio of labeled/unlabeled data on the results of the methods, we used respectively 10%, 25%, and 40% of the training data as labeled data (26, 65, and 104 patients) and the remainder as unlabeled data. As can be seen in this table, our method is better than other semi-supervised learning methods, and this is for all ratios of labeled/unlabeled data. This trend is also more pronounced for low ratios of labeled/unlabeled data. As expected, the segmentation performance is improved when the number of labeled data increases. However, it should be noted that our method still shows segmentation performance close to that of a fully supervised U-Net (100% of the labeled data) even when the number of labeled datasets is only 65 scans (25% of the labeled data)[1]. If we now look at the segmentation performance on an organ by organ basis, we can see that the spine has the highest segmentation accuracy due to the large contrast difference between the spine and the surrounding organs, which makes segmentation less difficult. As we suspected, the performance is lower for organs with a smaller contract and/or greater variability in shape.

Fig.(3) compares the segmentation results on two different slices (one with one fibroid -bottom- and one with multiple fibroids -top-) by using the different SOTA methods at different labeled/unlabeled ratios to the corresponding ground-truth. On these images, we can make several qualitative observations on the behavior of the different methods:

[1]These same observations were also confirmed in a comparative study of our method performed on multicentric cardiac data (different machines, sequences and operators). The results of this study (see supplementary material) show the robustness and adaptability of our method whatever the data.

Table 1: Quantitative comparison of three evaluation metrics of different segmentation methods on the testing dataset (best results are indicated in bold)

Method	#Scans used		Dice(%)			PR(%)			RE(%)		
	Labeled	Unlabeled	Uterus	Fibroid	Spine	Uterus	Fibroid	Spine	Uterus	Fibroid	Spine
U-Net	260 (100%)	0	75.34	77.58	78.15	76.97	78.39	89.10	74.81	79.23	71.46
HIFUNet	260	0	82.37	83.51	85.01	79.45	84.48	82.51	86.00	83.70	88.69
U-Net	104 (40%)	156 (60%)	73.32	70.73	83.76	68.43	**80.35**	76.44	**80.49**	65.43	**93.39**
CCT	104	156	58.28	42.34	73.97	48.05	65.11	76.60	76.87	33.02	75.16
ASDNet	104	156	71.05	69.76	85.60	71.95	66.09	**86.53**	71.41	81.10	85.44
Latent Mixup	104	156	73.72	74.17	84.38	69.92	69.87	81.77	79.00	83.95	87.99
CTA-Net	104	156	**75.20**	**75.80**	**86.24**	**75.26**	71.89	86.13	76.53	**84.57**	87.19
U-Net	65 (25%)	195 (75%)	67.39	65.72	84.56	68.1	63.24	82.22	68.27	74.56	88.03
CCT	65	195	51.50	37.75	72.94	59.69	67.29	**82.51**	48.33	28.17	68.33
ASDNet	65	195	67.52	69.67	**84.72**	67.44	63.47	79.14	69.3	82.28	91.97
Latent Mixup	65	195	71.41	73.36	82.47	71.02	68.91	73.96	73.48	82.83	**94.07**
CTA-Net	65	195	**74.72**	**73.51**	84.61	**73.19**	**74.94**	78.73	**77.45**	**75.95**	92.10
U-Net	26 (10 %)	234 (90%)	58.94	55.48	80.18	54.88	58.35	**79.56**	66.24	60.84	82.16
CCT	26	234	52.68	36.55	70.29	43.30	62.25	65.49	69.67	27.24	78.46
ASDNet	26	234	64.76	52.48	81.25	59.14	66.85	86.37	**74.07**	52.40	77.81
Latent Mixup	26	234	65.02	64.66	84.42	66.24	61.23	83.10	65.43	**76.31**	86.52
CTA-Net	26	234	**70.54**	**71.58**	**84.50**	**70.25**	**73.18**	79.37	72.31	74.25	**91.13**

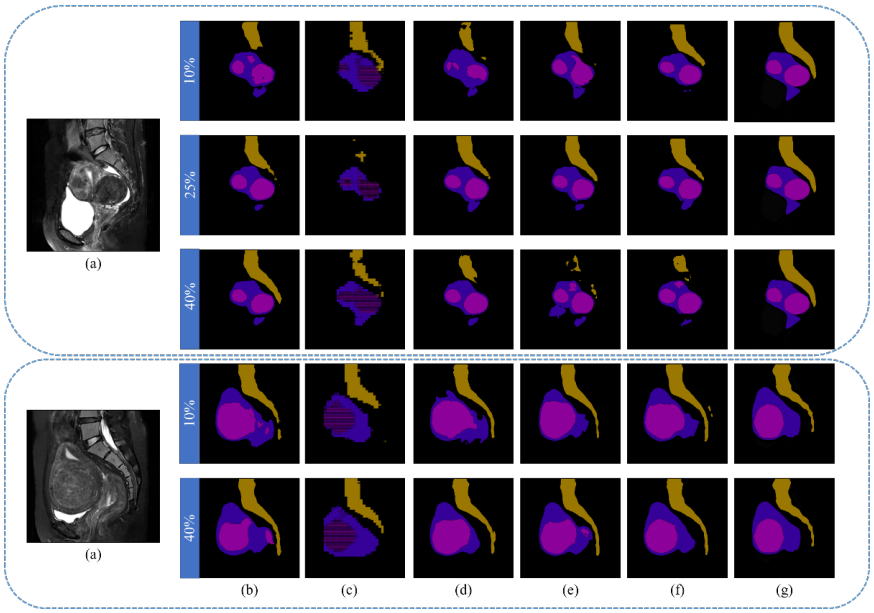

Figure 3: Segmentation results of 2 slices obtained by different SOTA methods with 3 different percentages of labeled data (10%, 25%, and 40%) and the corresponding ground-truth. From left to right are the (a) raw image, (b) results of U-Net, (c) CCT, (d) ASDNet, (e) Latent Mixup, (f) our CTANet and, (g) the ground-truth. Blue represents the uterus, pink the fibroids and yellow the spine.

1) The spine is segmented with higher accuracy than the uterus and the fibroid. This is due to the more fixed size and shape of the spine, and the more obvious difference in contrast with the surrounding tissues on the MRI image.

2) The case of multiple fibroids is much more complex than that of single fibroids. One explanation could be that in cases of multiple fibroids, the contrast and size of the fibroids are sometimes not the same. As shown in the case of multiple-fibroids at the top of Fig.(3), the two fibroids have different intensities, and it is easy to confuse them with the surrounding tissues whose contrast is similar to those of the fibroids.

3) As the ratio of annotated data increases, it does not always improve the segmentation results. For example, on the case of multiple-fibroids(the top of Fig.(3)), we can see that from 10% to 25%, the segmentation result of each method is improved significantly. However, from 25% to 40%, the performance of all the methods decreases except U-Net (b) and CCT(c). However, in the single-fibroid case (bottom of Fig.(3)), from 10% to 40%, the segmentation results of all methods are significantly improved.

4) Some of the methods show relatively poor results. For example, the CCT (c) method shows jagged boundaries. On the other hand, our method shows a segmentation behavior relatively consistent with the ground truth.

Table 2: Dice(%) of the proposed Confidence-Based Threshold Adaptation module on a 25% labeled dataset (best results are indicated in bold)

Threshold	Uterus	Fibroid	Spine	Average
0.1	69.57	69.78	82.61	73.99
0.2	68.29	68.88	83.07	73.41
0.3	71.56	73.17	84.54	76.42
0.4	71.36	**74.42**	83.37	76.38
0.5	70.61	69.63	82.41	74.22
0.6	69.78	70.60	82.87	74.42
0.7	68.60	69.13	81.50	73.08
0.8	70.38	67.90	77.58	71.95
Adaptive	**74.72**	73.51	**84.61**	**77.61**

4.5. Ablation Studies

We also conducted a series of ablation studies to justify the effectiveness of the proposed approach.

First, we analyzed the confidence-based threshold adaptation under 25% training data of the HIFU dataset. We compared our automatic adaptive threshold strategy with different offline fixed threshold settings, ranging from 0.1 to 0.8. The results in Table (2) show that some of the fixed thresholds can give a good segmentation performance for one specific organ. For example, 0.4 is the best threshold for fibroid segmentation, while a threshold of 0.3 shows better performance for uterus and spine segmentation. However, our method achieves the best average performance and the best performance for almost all the target organs.

Fig.(4) shows the adaptation of the threshold value during the training process. It can be seen that the threshold gradually converges from 0.80 to about 0.25, and there is a sharp to slow decrease during the training process. This finding indicates that more training rounds are needed when the network learns difficult segmentation regions.

We also wanted to estimate the impact of our several improvements on the segmentation results. Table (3) presents the ablation study of our CTANet with the several components or variants introduced in Section 3. All these experiments were performed using the same dataset with a 25%/75% labeled/unlabeled ratio. First, we compared the classic Vanilla U-Net (Net1) with our PSN-FSN architecture based on 2 U-Nets but without any other improvements (Net2). The 2 U-Nets slightly improved the results. Next, we added the CTA (Confidence-based Threshold Adaptation) to the architecture (Net3). CTA brought steady improvements in the segmentation of the uterus and fibroids. On this basis, we then compared the original Mixup (Net4) with our HMP (Hidden MixUp, Net5) and found that our HMP improved the segmentation accuracy by more than 1% on each of the three targets. Finally, the full solution (Net6) with RAD (consistency Regularization and Dropout) added to the system gave the best results. This table shows that each component plays an important role in our

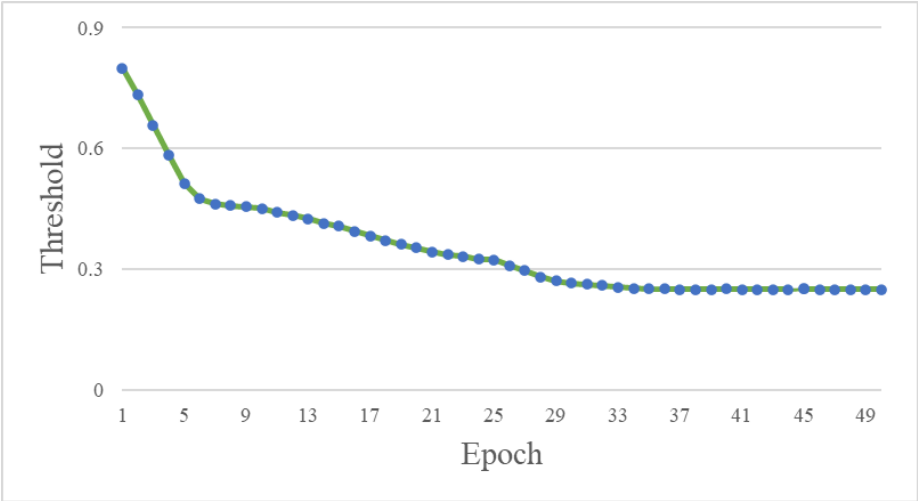

Figure 4: CTANet threshold adaptation during the training process (from 0.8 to 0.25).

semi-supervised scheme. Moreover, each of these components is independent of the other and can be applied to other semi-supervised learning networks. In the last column, we also report the computation time corresponding to each network to process a slice of the test data. We find that the additional computation time brought by each new component is acceptable.

Fig.(**??**) shows the segmentation results for 3 different images obtained after adding several components or variants, *i.e.*, with the networks Net1 to Net6 (Fig.(**??**). (b)-(g)). Visually comparing (b) and (c), we can see that the segmentation accuracy is improved for all three segmentation targets when we add the semi-supervised mechanism. Looking at (d), we see that although the adaptive threshold may have a negative impact on the segmentation of the spine. It allows to solve the "hole" problem that occurs in the segmentation. When comparing (e) and (f), we can see that HMP can further refine the segmentation of the uterus and improve the boundary segmentation problem compared to the traditional Mixup. One explanation could be the data augmentation at each scale of the feature map, especially for the continuous segmentation of the spine or, more obviously, the refinement of the cervical part of the uterus. Finally, using the strategy of incorporating regularization (g), the generalization and robustness of the model are further improved, which allows to enhance of the details of the segmentation targets to reduce the false segmentation caused by the noise in the pseudo-labels, and to make the morphology of the segmentation targets (smoothness of the fibroid, integrity of the uterine shape, and continuity of the spine) more consistent with the real situation.

Table 3: Effectiveness of the proposed techniques on the HIFU dataset using 25% of the labeled data (best results are indicated in bold)

Method	Baseline	Components					Dice(%)			Average	Inference time /slice(ms)
		Semi	CTA	Mixup	HMP	RAD	Uterus	Fibroid	Spine		
Net1	U-Net						67.39	65.72	84.56	72.56	38.5
Net2		✓					69.46	69.71	**85.01**	74.73	42.22
Net3		✓	✓				70.21	71.04	84.24	75.16	58.66
Net4		✓	✓	✓			71.34	72.83	81	75.05	60.36
Net5		✓	✓		✓		73.01	72.86	82.76	76.21	66.47
Net6 (ours)		✓	✓	✓	✓	✓	**74.72**	**73.51**	84.61	**77.62**	66.52

5. Discussion and Conclusion

By setting pixel-level weights (see (1)), a higher-confidence pixel corresponds to higher weights. In this way, the network focuses more attention on regions with high weights, thus improving the quality of the training data. Second, the thresholds are determined using an adaptive approach, which can alleviate the problem of time-consuming manual grid search. The threshold value gradually converges from the initial set value of 0.8 to 0.25. In the early epochs, the network focuses on pixels with higher confidence, corresponding to the easy-to-learn areas in an image. During the learning progression, the threshold gradually decreases, and the network starts to focus on some areas with lower confidence, which correspond to the hard-to-learn areas. The network gradually learns the features of the images by the threshold adaptive method from easy to difficult.

The introduction of the Hidden Mixup improves the generalizability of the model and effectively avoids over-fitting the data. The core of the strategy is to add complexity control to the space that is not covered by the training data. Our strategy performs linear interpolation using the data points generated in the encoding-decoding structure. This data augmentation model allows the neural network to learn a simple linear interpolation function in the "blank region", thus reducing the complexity of the uncovered space. The input of FSN is the product of the original image and the four-channel feature. We consider making the network pay more attention to the region of interest by adding soft attention to each channel, reducing the effect of irrelevant regions on the segmentation results.

By introducing the consistency regularization loss, the intermediate representation of the same data in two networks tends to be consistent, which improves the robustness and generalization of the model.

One possible limitation would be that the model is biased toward the dominant class. The model is trained based on labeled data distribution at the early training stage. Thus, a biased distribution of categories for the pretrained labeled data may directly affect the quality of pseudo-label generation and may impact the training results. Second, we used a simple U-Net for feature extraction, ignoring the difficulties of low tissue contrast around the uterus and fibroids with varying sizes and shapes, which is insufficient for feature extraction.

In conclusion, we have proposed a novel semi-supervised framework named Confidence-based Threshold Adaptation Network (CTANet) to improve the quality of pseudo-labels. The main contribution of our approach is the adaptive thresholds learning to automatically generate high-quality pseudo-labels for semi-supervised learning. This allows us to abandon offline threshold tuning. We validated our method on data used for HIFU fibroid treatment planning. This evaluation demonstrated that our segmentation network performed better than the SOTA semi-supervised learning methods.

The most prominent future work is to improve the quality of pseudo-labels by designing class-wise thresholds to generate unbiased pseudo-labels. In addition, we plan to extend our approach to external datasets from different sites. We will explore how to select and annotate representative data and extract richer

feature representations from limited data annotation.

Acknowledgments

This work was supported by the National Key Research and Development Program of China (2021ZD0113202) and, National Natural Science Foundation under grants (62171125, 31800825, 31640028), and the Natural Science Foundation of Jiangsu Province under grant (BE2019748), and in part by the China Scholarship Council under NO.201906090391.

Appendix A. Supplementary material

The folowing is the Supplementary material related to this article.

Complementary test to evaluate the robustness and adaptability of our method

In order to complete the study conducted in section 4.4 on the comparison with other semi-supervised deep learning methods and to prove the robustness and adaptability of our method, we applied our experimental setting on a dataset proposed in the Multi-Center, Multi-Vendor & Multi-Disease Cardiac Image Segmentation Challenge (M&Ms) [1]. The M&Ms dataset is composed of 375 patients and scanned in clinical centers from different countries using four different MRI platforms of major companies. The images have been segmented by experienced clinicians from the respective institutions, including contours for the left (LV) and right ventricles (RV), as well as for the left ventricular myocardium (MYO). The training set contains 150 annotated images from two different MRI vendors (75 each) and 25 unannotated images from a third company. The testing set compromises 200 patients from three mentioned vendors and one unseen vendor for testing the model generalization. The acquisition size of each case is different, uniformly cropped to 256×256.

We performed the same experimental setting as described in section 4.2 and compared our CTANet to U-Net and Latent Mixup with 3 different labeled/unlabeled data ratios. We compared the methods in terms of Dice similarity coefficient and Hausdorff Distance between the segmented Left Ventricle (LV), Myocardium (MYO) and Right Ventricle (RV) and the annotated ones.

Table S 1 shows the quantitative results on the M&Ms dataset. As for the experiment on the HIFU dataset we used 10%, 25%, and 40% of training data as labeled data. The trends observed in the HIFU study are also found when applying the methods to this cardiac data. Our method shows better results than other methods and achieves the best results when trained with different amounts of labeled data. Our method overcame the difficulty of this dataset, which lay in the fact that the data are collected from different equipment and hospitals, which places higher demands on the robustness and generalizability of the model. Besides that, we can also notice that the scores obtained by the different methods are higher on M&Ms dataset than on the HIFU dataset. This

Table S 1: Quantitative comparisons of semi-supervised segmentation models on the M&M dataset (best results are indicated in bold)

Method	#Scan used		Dice(%)			95HD (voxel)		
	Labeled	unlabeled	LV	MYO	RV	LV	MYO	RV
Latent Mixup	60(40%)	90(60%)	80.95	76.91	79.18	2.76	1.35	**0.91**
Ours	60	90	**86.17**	**78.5**8	**79.84**	**0.35**	**1.0**8	1.17
U-Net	38(25%)	112(75%)	77.04	65.13	69.10	2.58	4.16	2.6
U-Net+semi	38	112	81.49	72.09	74.55	2.06	2.29	2.16
Latent Mixup	38	112	82.79	72.61	74.05	1.28	1.44	1.51
Ours	38	112	**85.57**	**77.29**	**79.20**	**0.98**	**1.27**	**1.10**
Latent Mixup	15(10%)	135(90%)	77.55	64.59	56.19	7.17	7.11	11.5
Ours	15	135	**81.64**	**71.68**	**71.24**	**2.33**	**1.47**	**2.01**

proves that the segmentation of fibroids is more challenging than the segmentation of cardiac elements.

[1] V. M. Campello et al., "Multi-Centre, Multi-Vendor and Multi-Disease Cardiac Segmentation: The M&Ms Challenge," in IEEE Transactions on Medical Imaging, vol. 40, no. 12, pp. 3543-3554, Dec. 2021, doi: 10.1109/TMI.2021.3090082.

References

[1] E. E. Wallach, N. F. Vlahos, Uterine myomas: an overview of development, clinical features, and management, Obstetrics & Gynecology 104 (2) (2004) 393–406. doi:10.1097/01.AOG.0000136079.62513.39.

[2] E. A. Stewart, S. K. Laughlin-Tommaso, W. H. Catherino, S. Lalitkumar, D. Gupta, B. Vollenhoven, Uterine fibroids, Nature Reviews Disease Primers 2 (1) (2016) 1–18. doi:10.1038/nrdp.2016.43.

[3] E. A. Stewart, Uterine fibroids, The Lancet 357 (9252) (2001) 293–298. doi:10.1016/S0140-6736(00)03622-9.

[4] K. J. Anneveldt, H. J. van 't Oever, I. M. Nijholt, J. R. Dijkstra, W. J. Hehenkamp, S. Veersema, J. A. F. Huirne, J. M. Schutte, M. F. Boomsma, Systematic review of reproductive outcomes after High Intensity Focused Ultrasound treatment of uterine fibroids, European Journal of Radiology 141 (2021) 109801. doi:10.1016/j.ejrad.2021.109801.

[5] Y. Wang, Z.-B. Wang, Y.-H. Xu, Efficacy, efficiency, and safety of magnetic resonance-guided high-intensity focused ultrasound for ablation of uterine fibroids: comparison with ultrasound-guided method, Korean Journal of Radiology 19 (4) (2018) 724–732. doi:10.3348/kjr.2018.19.4.724.

[6] O. Ronneberger, P. Fischer, T. Brox, U-Net: Convolutional networks for biomedical image segmentation, in: Medical Image Computing and Computer Assisted Intervention – MICCAI 2015, Vol. 9351 of Lecture

Notes in Computer Science, Springer, 2015, pp. 234–241. doi:10.1007/978-3-319-24574-4_28.

[7] F. Milletari, N. Navab, S.-A. 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. doi:10.1109/3dv.2016.79.

[8] S. Laine, T. Aila, Temporal ensembling for semi-supervised learning, arXiv preprint arXiv:1610.02242doi:10.48550/arxiv.1610.02242.
URL https://arxiv.org/abs/1610.02242

[9] D.-H. Lee, Pseudo-label: The simple and efficient semi-supervised learning method for deep neural networks, in: ICML 2013 Workshop : Challenges in Representation Learning (WREPL), Vol. 3, 2013, p. 896, issue: 2.

[10] H. Zhang, M. Cissé, Y. N. Dauphin, D. Lopez-Paz, mixup: Beyond empirical risk minimization, arXiv preprint arXiv:1710.09412doi:10.48550/arXiv.1710.09412.
URL http://arxiv.org/abs/1710.09412

[11] L. Carratino, M. Cissé, R. Jenatton, J.-P. Vert, On mixup regularization, arXiv preprint arXiv:2006.06049doi:10.48550/arxiv.2006.06049.
URL https://arxiv.org/abs/2006.06049

[12] S. Yun, D. Han, S. J. Oh, S. Chun, J. Choe, Y. Yoo, Cutmix: Regularization strategy to train strong classifiers with localizable features, in: 2019 IEEE/CVF International Conference on Computer Vision (ICCV), 2019, pp. 6023–6032. doi:10.1109/iccv.2019.00612.

[13] D. Berthelot, N. Carlini, I. Goodfellow, N. Papernot, A. Oliver, C. Raffel, Mixmatch: A holistic approach to semi-supervised learning, arXiv preprint arXiv:1905.02249doi:10.48550/arxiv.1905.02249.
URL https://arxiv.org/abs/1905.02249

[14] N. Ben-Zadok, T. Riklin-Raviv, N. Kiryati, Interactive level set segmentation for image-guided therapy, in: 2009 IEEE International Symposium on Biomedical Imaging: From Nano to Macro, 2009, pp. 1079–1082. doi:10.1109/isbi.2009.5193243.

[15] T. F. Chan, L. A. Vese, Active contours without edges, IEEE Transactions on Image Processing 10 (2) (2001) 266–277. doi:10.1109/83.902291.

[16] A. Fallahi, M. Pooyan, H. Ghanaati, M. A. Oghabian, H. Khotanlou, M. Shakiba, A. H. Jalali, K. Firouznia, Uterine segmentation and volume measurement in uterine fibroid patients' MRI using fuzzy C-mean algorithm and morphological operations, Iranian Journal of Radiology 8 (3) (2011) 150–156. doi:10.5812/kmp.iranjradiol.17351065.3142.

[17] L. Ma, R. C. Staunton, A modified fuzzy C-means image segmentation algorithm for use with uneven illumination patterns, Pattern Recognition 40 (11) (2007) 3005–3011. doi:10.1016/j.patcog.2007.02.005.

[18] K. Antila, H. J. Nieminen, R. B. Sequeiros, G. Ehnholm, Automatic segmentation for detecting uterine fibroid regions treated with MR-guided high intensity focused ultrasound (MR-HIFU), Medical Physics 41 (7) (2014) 073502. doi:10.1118/1.4881319.

[19] L. Rundo, C. Militello, S. Vitabile, C. Casarino, G. Russo, M. Midiri, M. C. Gilardi, Combining split-and-merge and multi-seed region growing algorithms for uterine fibroid segmentation in MRgFUS treatments, Medical & Biological Engineering & Computing 54 (7) (2016) 1071–1084. doi:10.1007/s11517-015-1404-6.

[20] Y. Kurata, M. Nishio, A. Kido, K. Fujimoto, M. Yakami, H. Isoda, K. Togashi, Automatic segmentation of the uterus on MRI using a convolutional neural network, Computers in Biology and Medicine 114 (2019) 103438. doi:10.1016/j.compbiomed.2019.103438.

[21] C. Zhang, H. Shu, G. Yang, F. Li, Y. Wen, Q. Zhang, J.-L. Dillenseger, J.-L. Coatrieux, HIFUNet: multi-class segmentation of uterine regions from MR images using global convolutional networks for HIFU surgery planning, IEEE Transactions on Medical Imaging 39 (11) (2020) 3309–3320. doi:10.1109/tmi.2020.2991266.

[22] G. Ning, X. Zhang, Q. Zhang, Z. Wang, H. Liao, Real-time and multimodality image-guided intelligent HIFU therapy for uterine fibroid, Theranostics 10 (10) (2020) 4676–4693. doi:10.7150/thno.42830.

[23] B. Behboodi, H. Rivaz, S. Lalondrelle, E. Harris, Automatic 3D ultrasound segmentation of uterus using deep learning, in: 2021 IEEE International Ultrasonics Symposium (IUS), 2021, pp. 1–4. doi:10.1109/ius52206.2021.9593671.

[24] M. Sandler, A. Howard, M. Zhu, A. Zhmoginov, L.-C. Chen, MobileNetV2: inverted residuals and linear bottlenecks, in: 2018 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), 2018. doi:10.1109/cvpr.2018.00474.

[25] Y. Chong, Y. Ding, Q. Yan, S. Pan, Graph-based semi-supervised learning: A review, Neurocomputing 408 (2020) 216–230. doi:10.1016/j.neucom.2019.12.130.

[26] T. N. Kipf, M. Welling, Semi-supervised classification with graph convolutional networks, arXiv preprint arXiv:1609.02907doi:10.48550/arxiv.1609.02907.
URL https://arxiv.org/abs/1609.02907

[27] J. Wang, J. Liang, J. Cui, J. Liang, Semi-supervised learning with mixed-order graph convolutional networks, Information Sciences 573 (2021) 171–181. doi:10.1016/j.ins.2021.05.057.

[28] K. Sricharan, R. Bala, M. Shreve, H. Ding, K. Saketh, J. Sun, Semi-supervised conditional GANs, arXiv preprint arXiv:1708.05789doi:10.48550/arxiv.1708.05789.
URL https://arxiv.org/abs/1708.05789

[29] W.-C. Hung, Y.-H. Tsai, Y.-T. Liou, Y.-Y. Lin, M.-H. Yang, Adversarial learning for semi-supervised semantic segmentation, arXiv preprint arXiv:1802.07934doi:10.48550/arxiv.1802.07934.
URL https://arxiv.org/abs/1802.07934

[30] I. Z. Yalniz, H. Jégou, K. Chen, M. Paluri, D. Mahajan, Billion-scale semi-supervised learning for image classification, arXiv preprint arXiv:1905.00546doi:10.48550/arxiv.1905.00546.
URL https://arxiv.org/abs/1905.00546

[31] Q. Xie, M.-T. Luong, E. Hovy, Q. V. Le, Self-training with noisy student improves imagenet classification, in: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, 2020, pp. 10687–10698.

[32] R. Chen, Y. Ma, L. Liu, N. Chen, Z. Cui, G. Wei, W. Wang, Semi-supervised anatomical landmark detection via shape-regulated self-training, Neurocomputing 471 (2022) 335–345. doi:10.1016/j.neucom.2021.10.109.

[33] A. Blum, T. Mitchell, Combining labeled and unlabeled data with co-training, in: Proceedings of the eleventh annual conference on Computational learning theory, 1998, pp. 92–100. doi:10.1145/279943.279962.

[34] S. Qiao, W. Shen, Z. Zhang, B. Wang, A. Yuille, Deep co-training for semi-supervised image recognition, in: Proceedings of the European Conference on Computer Vision – ECCV 2018, 2018, pp. 142–159. doi:10.1007/978-3-030-01267-0_9.

[35] J. Peng, G. Estrada, M. Pedersoli, C. Desrosiers, Deep co-training for semi-supervised image segmentation, Pattern Recognition 107 (2020) 107269. doi:10.1016/j.patcog.2020.107269.

[36] A. Tarvainen, H. Valpola, Mean teachers are better role models: Weight-averaged consistency targets improve semi-supervised deep learning results, arXiv preprint arXiv:1703.01780doi:10.48550/arxiv.1703.01780.
URL https://arxiv.org/abs/1703.01780

[37] Y. Ouali, C. Hudelot, M. Tami, Semi-supervised semantic segmentation with cross-consistency training, in: 2020 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), 2020, pp. 12674–12684. doi:10.1109/cvpr42600.2020.01269.

[38] D. Nie, Y. Gao, L. Wang, D. Shen, Asdnet: Attention based semi-supervised deep networks for medical image segmentation, in: Medical Image Computing and Computer Assisted Intervention – MICCAI 2018, Vol. 11073 of Lecture Notes in Computer Science, 2018, pp. 370–378. doi:10.1007/978-3-030-00937-3_43.

[39] X. Li, L. Yu, H. Chen, C.-W. Fu, P.-A. Heng, Semi-supervised skin lesion segmentation via transformation consistent self-ensembling model, arXiv preprint arXiv:1808.03887doi:10.48550/arxiv.1808.03887.
URL https://arxiv.org/abs/1808.03887

[40] W. Bai, O. Oktay, M. Sinclair, H. Suzuki, M. Rajchl, G. Tarroni, B. Glocker, A. King, P. M. Matthews, D. Rueckert, Semi-supervised learning for network-based cardiac MR image segmentation, in: Medical Image Computing and Computer-Assisted Intervention – MICCAI 2017, Vol. 10434 of Lecture Notes in Computer Science, 2017, pp. 253–260. doi:10.1007/978-3-319-66185-8_29.

[41] R. Ito, K. Nakae, J. Hata, H. Okano, S. Ishii, Semi-supervised deep learning of brain tissue segmentation, Neural Networks 116 (2019) 25–34. doi:10.1016/j.neunet.2019.03.014.

[42] F. Pizzati, R. d. Charette, M. Zaccaria, P. Cerri, Domain bridge for unpaired image-to-image translation and unsupervised domain adaptation, in: 2020 IEEE Winter Conference on Applications of Computer Vision (WACV), 2020, pp. 2990–2998. doi:10.1109/wacv45572.2020.9093540.

[43] P. K. Gyawali, S. Ghimire, P. Bajracharya, Z. Li, L. Wang, Semi-supervised medical image classification with global latent mixing, in: Medical Image Computing and Computer Assisted Intervention – MICCAI 2020, Vol. 12261 of Lecture Notes in Computer Science, 2020, pp. 604–613. doi:10.1007/978-3-030-59710-8_59.