ITC 3/54
Information Technology and Control
Vol. 54 / No. 3/ 2025
pp. 958-975
DOI 10.5755/j01.itc.54.3.38178

MEMatch: A Semi-supervised Cardiac MRI Segmentation Method Guided by Entropy and Multi-scale Joint Strong-weak Consistency

Received 2024/07/23 | Accepted after revision 2024/12/04

HOW TO CITE: Ren, F., Bao, S., Yu, C., Yang, Y., Xu, X. (2025). MEMatch: A Semi-supervised Cardiac MRI Segmentation Method Guided by Entropy and Multi-scale Joint Strong-weak Consistency. *Information Technology and Control*, 54(3), 958-975. https://doi.org/10.5755/j01.itc.54.3.38178

MEMatch: A Semi-supervised Cardiac MRI Segmentation Method Guided by Entropy and Multi-scale Joint Strong-weak Consistency

Fuquan Ren*, Shuang Bao, Chunlei Yu, Yuromg Yang, Xing Xu
School of Science, Yanshan University, Qinhuangdao, 066004, China

Corresponding author: renfu_quan@ysu.edu.cn

Cardiac magnetic resonance imaging (cardiac MRI) is a very powerful tool for the diagnosis of cardiovascular disease. Accurate automated segmentation of cardiac MRI can enhance clinical diagnostic efficiency. While supervised cardiac MRI segmentation methods have achieved brilliant achievements, they mostly rely on an enormous quantity of labeled samples, which is extremely challenging and expensive to acquire. To alleviate the challenge of manual annotation, semi-supervised based methods offer an effective solution. Despite some progress in semi-supervised segmentation for cardiac MRI, there remains a gap in clinical application and accuracy needs further improvement. This paper proposes MEMatch, a novel semi-supervised cardiac MRI segmentation method. MEMatch introduces multi-scale joint strong-weak consistency, which applies strong-weak consistency to the prediction results of multiple scales of the network, to more fully utilize the discrepancy between the outputs of different scales of the same network. Additionally, entropy minimization is applied to the average prediction of multiple scales, which enforces the average prediction to generate high-confidence predictions and further reduces the discrepancy among the prediction results from different scales. As demonstrated by the experimental results, utilizing only 3 labeled samples, the proposal improves Dice and Jaccard scores by 1.48%/0.17% and 2.15%/0.25%, respectively, compared to optimal performances of other methods on the ACDC/LA dataset. The experimental results on 2D and 3D segmentation illuminate the effectiveness and superiority of our method in various semi-supervised settings compared to state-of-the-art techniques.

KEYWORDS: Cardiac MRI, Semi-supervised segmentation, Strong-weak consistency, Entropy minimization

1. Introduction

Cardiovascular disease (CVD) remains to be a leading cause of mortality in the global population [14, 34]. In 2021 alone, approximately 20.5 million deaths were attributed to CVD, nearly one-third of all global deaths, according to the World Heart Report 2023 by the World Heart Foundation. Additionally, there has been an increasing trend of CVD affecting younger populations in recent years, attributed to unhealthy lifestyles and obesity. Early diagnosis and treatment of CVD can significantly increase patient survival rates. In the diagnosis of CVD, cardiac MRI [32, 35] have become a main auxiliary tool for assessing patient conditions due to its safety and reliability, and were considered as the golden standard for assessing the structure and function of the heart.

Accurate segmentation and identification of intracardiac structures from cardiac MRI data play a crucial role in guiding the diagnosis of cardiovascular diseases (CVD) and serve as an important preparatory step before cardiac surgery. In clinical practice, it is necessary for experts to manually delineate the contours of the left ventricle, right ventricle, and myocardium. However, due to the structural complexity of cardiac MRI, the manual segmentation process can be time-intensive, laborious, and inefficient. Therefore, there is an urgent need to develop fast, accurate, reproducible, and fully automated cardiac magnetic resonance imaging segmentation methods using computer technology to improve clinical diagnostic efficiency and reduce cardiovascular disease mortality.

In the field of medical image processing, the segmentation of cardiac MRI images has been a prominent subject. Numerous algorithms have emerged over time [36]. However, traditional approaches such as threshold-based methods [22, 41], edge detection and active contour-based methods [30, 44, 28], clustering based methods [15, 7] and their hybrid algorithms [52, 39, 31] often necessitate prior knowledge of heart structure and shape information, along with manual intervention. Despite theoretical advancements made by these methods, they still possess limitations when it comes to practical clinical applications.

Deep learning (DL) methodology has been extensively used for tasks in computer vision and the analysis and processing of cardiac MRI, achieving a succession of breakthroughs over the past several years. Simultaneously, deep learning-based approaches have also received widespread attention and achieved good results in solving cardiac MRI segmentation problems, especially after the proposal of U-Net. U-Net [33] was proposed and applied to the field of medical image segmentation in 2015, which significantly enhances the performance of cardiac MRI segmentation compared to traditional algorithms. Furthermore, numerous enhancements to the U-Net [11, 9, 27] have been studied to further enhance the accuracy of cardiac MRI segmentation.

Although these methods have achieved encouraging results, they all rely on an extensive quantity of pixel/voxel-level expert labeling as training data. However, due to the complex structure of cardiac MRI images, it is inherently time-consuming for specialized physicians to obtain their pixel-level labels. In addition, cardiac MRI images present challenges such as the lack of clear tissue boundaries and variability in the size and location of individual hearts, which further complicates the task of sample labels for segmentation. To reduce the burden of manual labeling, methods such as weakly supervised learning (WSL) [53, 19], semi-supervised learning (SSL) [49], and unsupervised learning [26, 13] have been proposed successively. This paper focuses on the application of SSL to address the issue of cardiac MRI segmentation, aiming to integrate a small number of labeled images with a large number of unlabeled images to achieve the training of network models. Currently, the mainstream semi-supervised cardiac MRI segmentation methods include: pseudo labels [8, 25], consistency regularization [51, 21, 23, 46, 29, 24, 37], entropy minimization [18, 1], and so on. The basic concept behind consistency regularization is that, for the same input, the predictions after applying different forms of perturbation to it should be as similar as possible. Currently, many methods construct consistency by adding multiple decoders to the same encoder and constructing consistency between the predictions of the different decoders and do not focus on the consistency and similarity of the predictions at different scales in a single decoder. For example, MC-Net [47] improves accuracy by constraining mutual consistency between two different decoders with an identical encoder.

In addition, Despite the remarkable achievements of these methods, there is still a certain gap to be bridged before its application in the diagnosis of CVD in clinical practice. Based on the idea that predictions at different scales of the same input from the same network should be as similar as possible, this paper innovatively proposes MEMatch, a semi-supervised cardiac MRI segmentation model. Overall, the contributions of the paper can be summarized in three aspects:

1. This paper proposes MEMatch, a novel semi-supervised cardiac MRI segmentation method guided by multi-scale joint strong-weak consistency and entropy minimization.
2. A multi-scale joint strong-weak consistency method is proposed to more fully utilize the discrepancies between the outputs of different scales within the same network. This method is then combined with entropy minimization based on average prediction to further enhance the model's capabilities.
3. Numerous experiments are conducted on our approach with different ratio labeled samples and make comparisons to several state-of-the-art segmentation techniques on the ACDC and LA datasets.

The remaining sections mainly include the following contents: Section 2 describes the current state of development of the research field and some of the best approaches. Section 3 presents the proposal MEMatch algorithm in detail. The semi-supervised experimental settings, comparative experimental results, and ablation studies of this paper are introduced in Section 4. In Section 5, a summary of the proposed methodology is presented.

2. Related Work

Medical image segmentation is an important assistive technology for various clinical diagnoses such as radiotherapy and computer-aided diagnosis [40]. Cardiac MRI segmentation is a crucial stage in the diagnosis of CVD. As the research on deep learning continues, U-Net and its improved models have become the most popular models as well as in cardiac MRI segmentation. However, it still suffers from underutilization of contextual information due to structural defects. Therefore, a series of variants of U-net were proposed and widely used. For example, Attention U-Net [11] introduces a novel Attention Gate module and applies it to the end of the skip connections of the U-Net network to enable the model to capture more useful information by suppressing the features of irrelevant regions.

However, because of the characteristics of convolution itself, the capability of convolution-based approaches to capture long-distance relationships does not meet the practical requirements. To address this limitation, people have attempted to use Transformer models [42], which are influenced by the natural language domain, to replace or combine with CNN. However, this approach still relies on the paradigm of fully supervised segmentation and remains restricted by the availability and quality of data annotation for dynamic cardiac MRI.

2.1. Deep Semi-Supervised Learning

Deep semi-supervised learning (DSSL), which effectively leverages a large amount of unlabeled data in combination with a small amount of labeled data for learning, thereby reducing manual labeling costs and improving learning accuracy, is garnering increasing attention and research [50]. When the number of labeled samples is limited, DSSL technology can enhance the learning ability by incorporating unlabeled instances. This distinguishes DSSL from supervised learning algorithms, which rely solely on labeled data.

The primary challenge in DSSL lies in devising efficacious and appropriate supervision signals for unlabeled samples. Over the years, numerous approaches have been introduced to address this issue in image segmentation. The mainstream methods can be roughly categorized into the following groups: proxy-label methods [8, 25], consistency regularization [51, 21, 23, 46, 29, 24, 16], generative models [12, 20], graph-based methods [10, 38] and entropy minimization [17, 45]. Besides the main categories mentioned above, there are also some hybrid-based DSSL techniques that combine ideas from the aforementioned approaches to enhance performance in learning [50]. For example, in the MixMatch [6], Berthelot et al. introduce a novel method of multiple predictions with sharpening to improve the quality of pseudo-labels. Furthermore, they innovatively extend the mixup algorithm to semi-supervised learning by incorporating the generated labels. In

subsequent research, ReMixMatch [5] improves the MixMatch method by introducing Distribution Alignment (DA) and Augmentation Anchoring (AA). Distribution Alignment aims to align the marginal distribution of predictions for unlabeled samples with the marginal distribution of true labels, encouraging closer alignment between the two distributions. On the other hand, Augmentation Anchoring involves providing multiple strongly augmented versions $\{x_i\}$ of an input sample x into the model and promoting each output y_i to align closely with the prediction from weaker augmented version of x. As an effective and intuitive medical image segmentation algorithm based on SSL, Fixmatch [37] employs the combined strategy of consistency regularization and pseudo-labeling to enhance the segmentation performance. It first applies weak and strong augmentations to an unlabeled sample and then uses the prediction of the weakly-augmented sample to generate a pseudo-label, that serves as a supervision signal for the prediction of the strongly-augmented sample to compute the consistency loss.

2.2. DSSL Based Cardiac MRI Segmentation

Due to its effectiveness in reducing the dependency on manual segmentation annotations, SSL is increasingly becoming as a viable solution and has been widely applied in medical image segmentation, particularly in addressing the challenge in segmenting cardiac MRI images. A diverse range of algorithms has been proposed and implemented for cardiac MRI segmentationin recent years, including pseudo labels [25, 51], consistency regularization [51, 21, 23, 46, 29, 16], and entropy minimization [18, 1, 47].

Pseudo-labeling methods are typically implemented iteratively, with the network first being trained using labeled samples. Then unlabeled samples are applied to the network after being trained to generate pseudo-labels. Subsequently, the network training is facilitated by employing pseudo-labels along with a small amount of ground truth labels. For example, Chang et al. [8] introduce an effective dynamic computational algorithm for generating pseudo-labels and regularized constraints on unlabeled data using a specific loss function. Ma et al. [25] perform Fourier transform enhancement on the image, then generate pseudo-labels based on the input image, and finally calculate the variance between the prediction results and the pseudo-labels, and use this to optimize the network parameters. However, the pseudo-labeling approach is highly dependent on the suitable threshold, an unsuitable value can cause substantial training bias in the model.

Consistency regularization methods are founded on the straightforward assumption that the model's predictions should remain unchanged in response to perturbations of either the model or the images. They impose consistency constraints on one or more perturbed branches and show excellent ability in semi-supervised cardiac MRI segmentation. For example, SASSNet [21] introduces more flexible geometric representations in the model to implement global shape constraints on the predicted results. SS-Net [46] introduces adversarial noise as a strong perturbation to encourage inter-class separation by shrinking each class distribution, thereby achieving better performance. UA-MT [51] extends the Mean-Teacher framework and introduces an uncertainty-aware method that improves the utilization of valid information in the student model. Cross-ALD [29] proposes a Cross-ALD regularization that is different from other regularization methods to improve the smoothness assumption. Similar to our proposed method, UniMatch [49] also improves the Fixmatch model. It has been demonstrated by extensive experiments that applying a strong perturbation to unlabeled samples can significantly enhance the model's segmentation capabilities, and based on this, it proposes a strong perturbation branch based on feature enhancement. Simultaneously, to make full use of the benefits of strong perturbation, it adds another strong perturbation branch to the network, which greatly enhances the performance of the model. Basak et al. [3] introduce a special interpolation-based regularization algorithm that uses the interpolation of the predictions generated from two unlabeled samples as pseudo-labels to supervise the predictions they interpolate. CAML [16] further improves the model's ability to utilize labeled information by injecting labeled data information into unlabeled data.

Entropy minimization is also a kind of simple, effective, and very widely used semi-supervised method. LLUB [1] proposes a novel algorithm to reconstruct the segmentation mask and estimate the pixel-level uncertainty by this segmentation mask. LG-ER-MT [18] first introduces entropy minimization into the

student network of the Mean Teacher framework, to guide it to produce high-confidence predictions. Second, it also proposes two different types of consistency losses, local and global, to obtain the local and global structural information.

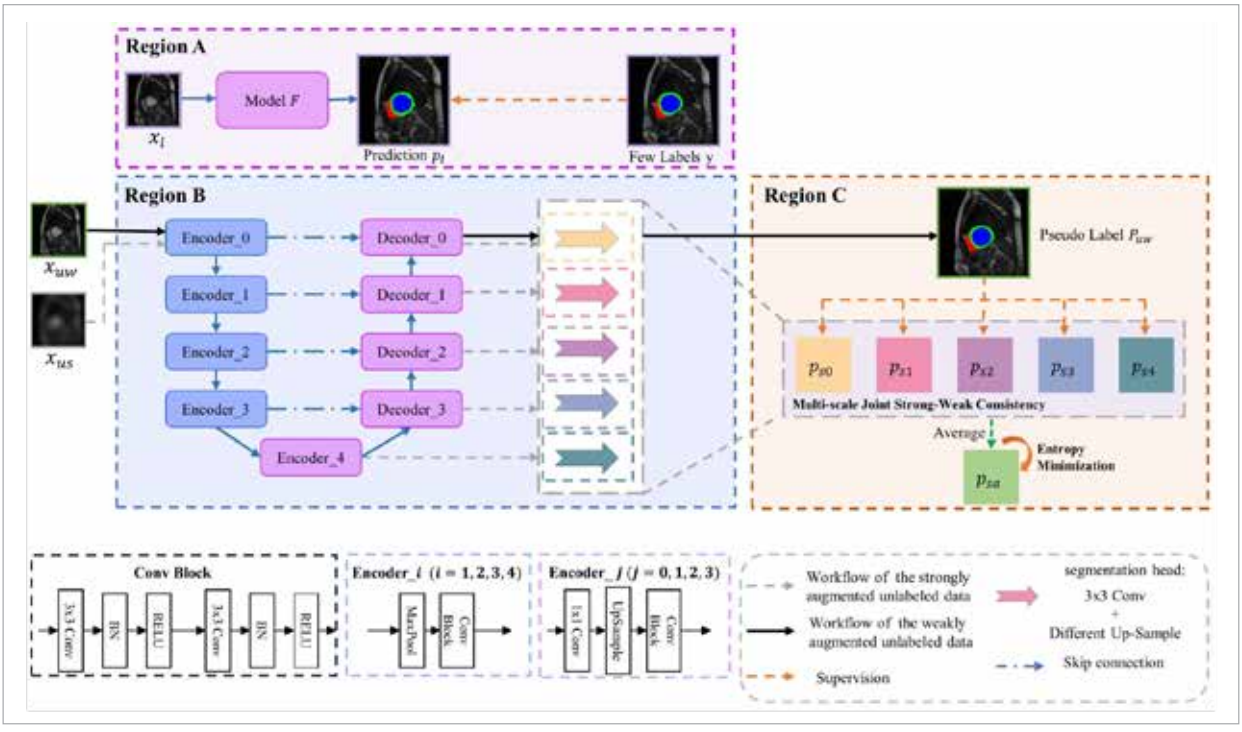
3. Method

In this section, the overall workflow of our MEMatch is first presented. After that, our proposed multi-scale joint strong-weak consistency and entropy minimization based on average prediction are introduced in detail, respectively. Figure 3 displays the structure of MEMatch for cardiac MRI semi-supervised segmentation.

– For labeled data, the process is shown in Region A in Figure 3. The labeled data are fed into the segmentation network, which utilizes the encoder and decoder to generate segmentation predictions. These predictions are then compared with the ground truth labels using cross-entropy loss and Dice loss, resulting in a supervised loss that is a linear combination of the two losses.
– For unlabeled data, the process is shown in Region B in Figure 3. Randomized weak and strong augmentation are applied to unlabeled data, respectively.
  - The segmentation network receives the weakly-augmented unlabeled data and filters out the pseudo-labels based on a confidence threshold.
  - For strongly-augmented unlabeled data, we add a segmentation head with different up-sampling strategies after the encoder output block and each decoder block of the segmentation network to obtain segmentation predictions at different scales. As shown in Region C in the Figure 3, the strongly-augmented unlabeled data is incorporated into the segmentation network, passing through the encoder and decoder, and obtaining multi-scale segmentation

Figure 1

The architecture of our proposed MEMatch. Encoder_i (j=1,2,3,4) and Decoder_j (j=0,1,2,3) are the encoder block and decoder block of the network. x_l denotes the labeled data. x_{uw} and x_{us} represent the weakly-augmented unlabeled data and the strongly-augmented unlabeled data, respectively. p_{si} (i=1,2,3,4) denotes the segmentation predictions obtained after different segmentation heads. p_{sa} denotes the average prediction.



prediction results p_{s0}, p_{s1}, p_{s2}, p_{s3}, and p_{s4} from multiple segmentation heads with different upsampling strategies. Subsequently, the multi-scale segmentation prediction results are used to calculate Dice loss with pseudo-labels, resulting in consistency loss and the so-called multi-scale Joint Strong-Weak Consistency loss. Furthermore, the average prediction result p_{sa} is derived based on the multi-scale segmentation prediction results, leading to entropy minimization loss.

For strongly-augmented unlabeled data, we provide supervisory signals for predictions at multiple scales rather than for a single final output. This ensures that predictions at all scales are closer to the pseudo-labels, enabling the model to learn more diverse features. Based on this idea, we propose multi-scale joint strong-weak consistency, which uses the pseudo-labels filtered from the weakly-augmented unlabeled data to supervise multi-scale segmentation predictions on the strongly-augmented unlabeled data. To further enhance the consistency between the predictions of different scales and make full use of the prediction information of multiple scales, we average the predictions of multiple scales to obtain the average prediction and introduce entropy minimization as a regularization constraint for the average prediction. If the prediction results of multiple scales for a certain pixel deviation is large, the entropy minimization loss of the model will be large, and then the entropy minimization loss will constrain the prediction results of multiple scales to converge to 0 or 1 to minimize the entropy, and at the same time, reduce the uncertainty of the model and improve the consistency of the prediction results of the multiple scales.

Our MEMatch consists of two core modules: multi-scale joint strong-weak consistency and entropy minimization. To introduce our method in more detail, the meaning of some mathematical symbols is described. Our training set consists of the vast majority of unlabeled samples X_u and a few labeled samples X_l, where $X_l = \{(x_l^i, y^i)\}_{i=1}^{N}$ (y^i represents ground-truth annotations), $X_u = \{x_u^i\}_{i=N+1}^{N+M}$ ($N=M$). Let x_u^i undergo weak augmentation and strong augmentation to obtain x_{uw}^i and x_{us}^i, respectively. In addition, we denote the cross-entropy and Dice loss between two probability distributions p and q as $H(p,q)$ and $C(p,q)$.

As a framework based on SSL, our approach is designed to works with unlabeled samples as well as labeled samples. During the supervised training, the segmentation network F is applied to the labeled sample x_l to obtain the segmentation prediction p_l. p_l is supervised by the ground-truth label y, then the supervised loss can be expressed as:

$$L_{seg} = \frac{1}{2}(H(p_l, y) + C(p_l, y)). \quad (1)$$

3.1. Multi-scale Joint Strong-weak Consistency

To more fully utilize the discrepancy between the outputs of different scales of the same network to obtain more consistency information, the multi-scale joint strong-weak consistency is proposed. Specifically, the unlabeled samples are divided into two branches after undergoing weak and strong data augmentation, namely weakly-augmented unlabeled data and strongly-augmented unlabeled data.

For the weakly-augmented unlabeled data branch, the weakly-augmented unlabeled samples pass through the segmentation network F, producing segmentation predictions, and selecting pseudo-labels according to the confidence threshold τ.

For the branch of strongly-augmented unlabeled data, we implement our proposed multi-scale joint strong-weak consistency. Specifically, a segmentation head with different upsampling strategies is added after the encoder output block and each decoder block of the segmentation network. Firstly, the strongly-augmented unlabeled samples pass through the segmentation network, producing segmentation predictions at different scales through the corresponding segmentation heads. Subsequently, taking the pseudo-labels output by the weakly-augmented unlabeled data branch as the true segmentation masks, consistency losses are calculated with each segmentation prediction, respectively.

For unsupervised training, we use

$$P_{uw} = \arg\max(F(x_{uw}))$$

as the pseudo-label for the unlabeled data x_{uw}. As shown in Figure3, the strongly-augmented data x_{us} passes through Decoder_0 and its segmentation head to obtain the segmentation prediction p_{s0}. Then the consistency loss L_{con} can be expressed as:

$$L_{con} = \frac{1}{B_u} \sum \Gamma(\max(F(x_{uw}) \geq \tau)) C(P_{uw}, p_{s0}), \quad (2)$$

where B_u is the batchsize for unlabeled data, Γ denotes the indicator function, τ is the threshold.

Let p_{sj} (j=1,2,3) denote the segmentation predictions at different scales obtained by applying Decoder_j (j=1,2,3) and its segmentation head to x_{us}. Let p_{s4} denote the segmentation prediction obtained by applying Encoder_4 and its segmentation head to x_{us}. Then the multi-scale joint strong-weak consistency loss L_{mc} can be written as:

$$L_{mc} = \sum_{j=1}^{4} \frac{1}{B_u} \sum \Gamma(\max(F(x_{uw}) \geq \tau)) C(P_{uw}, p_{sj}). \quad (3)$$

3.2. Entropy Minimization Based on Average Prediction

Entropy [43] is a very important concept in the field of SSL, which is used as a measure of a model's uncertainty about unlabeled samples. Entropy minimization is an effective method for imposing consistency constraints on unlabeled samples and is extensively applied in the field of SSL-based cardiac MRI segmentation [18, 47]. Entropy minimization can drive the model to generate predictions with a high level of confidence.

To further constrain multi-scale predictions of the same network to converge to similarity, we propose entropy minimization based on average prediction. Specifically, we average the multi-scale segmentation predictions obtained from multiple segmentation heads to get the average prediction, and then apply an entropy minimization constraint on it.

Let $p_{sa} = \frac{\sum_{j=0}^{4} p_{sj}}{5}$ denote the average prediction, then the entropy minimization loss can be written as:

$$L_{em} = -\frac{1}{M} \sum_{i=1}^{M} \sum_{c=1}^{C} \log(p_{sa}^{ic}) p_{sa}^{ic}, \quad (4)$$

where M and C represent the number of pixels and the number of classes for segmentation (including background), respectively, and p_{sa}^{ic} denotes the average prediction probability of class c at pixel i. Our optimization objective is to minimize the L_{em}. Observing Equation 4, it is evident that in order to minimize the L_{em}, it is necessary to drive each p_{sa}^{ic} towards convergence to 0.0 or 1.0. This will result in the average prediction of the multiple scales of our proposed method generating a high-confidence prediction for class c or any other class.

3.3. The Overall Loss Function

The proposed method aims to learn the model parameters by minimizing the total loss function L_{total}, which consists of a linear combination of L_{seg}, L_{con}, L_{mc}, and L_{em}. Summarizing Sections 3.1-3.2, the unsupervised loss of the proposed method can be expressed as:

$$L_{un} = \alpha L_{con} + \beta L_{mc} + \gamma L_{em} \quad (5)$$

where L_{con}, L_{mc}, L_{em} represent the consistency loss, the multi-scale joint strong-weak consistency loss, and the entropy minimization loss, respectively. α, β, γ represent the weight of L_{con}, L_{mc}, L_{em}. To explore the best loss weights, we conduct many ablation studies on the ACDC dataset with different weights of our loss function. If not specified, we set the hyperparameters α=0.5, β=0.25, γ=0.1, and τ=0.95.

Finally, the total loss function of the proposed method can be expressed as follows.

$$L_{total} = L_{seg} + L_{un}. \quad (6)$$

4. Experimental

4.1. Implementation Details

4.1.1. Dataset

We evaluate our proposed method on the publicly available medical image dataset ACDC and LA.

ACDC (Automatic Cardiac Diagnosis Challenge) [4] dataset is a multi-category cardiac 3D MRI image dataset that is widely used for developing and evaluating automatic segmentation and diagnosis methods in the field of cardiac imaging. It contains 200 MRI scans of 100 patients with three categories, left ventricle (LV), right ventricle (RV), and myocardium (Myo). The ACDC dataset is utilized for the purposes of training, validation, and testing by employing scans from 70, 10, and 20 patients, respectively. The 3D MRI volumes are sliced into 2D images during train-

ing, with 1312, 20, and 40 slices utilized for training, validation, and testing respectively, and each image is randomly cropped to a size of 112x122x80 during the training process before input into the network.

LA (2018 Left Atrium Segmentation Challenge) dataset [48] contains a total of 100 left atrium MRI images, of which 80 and 20 images are used for training and testing, respectively. In our experiment, 4, and 8 cases are used as labeled data, respectively. During the training process, 3D patches with dimensions of are randomly cropped to serve as the input for the network.

For the unlabeled samples in the ACDC and LA datasets, we apply strong and weak augmentation on them during training. Specifically, the strong augmentations include rotate, blur, and cutmix, while the weak augmentations include resize, crop, and flip.

4.1.2. Evaluation Metric

Four different metrics are used to evaluate our approach: Dice Similarity Coefficient (Dice), Jaccard Similarity Coefficient (Jaccard), Average Symmetric Distance (ASD), and Hausdorff Distance 95 (HD95). For two object regions, the Jaccard and Dice coefficients calculate the percentage of overlap between them, the Average Surface Distance (ASD) calculates the average distance between their boundaries, and the Hausdorff Distance 95% (HD95) measures the distance to the nearest point between them.

Let X denote the segmentation prediction result of the model and Y denote the segmentation mask. Then the mathematical definitions of the four metrics are expressed as follows:

$$Dice(Y,X) = \frac{2|Y \cap X|}{|X| + |Y|} \tag{7}$$

$$Jaccard = \frac{|Y \cap X|}{|Y \cup X|} \tag{8}$$

$$ASD = \frac{1}{2}\left(\underset{x\in\bar{X}}{mean}\min_{y\in\bar{Y}} d(x,y) + \underset{y\in\bar{Y}}{mean}\min_{x\in\bar{X}} d(x,y)\right) \tag{9}$$

$$HD95 = \max\left(\max_{x\in\bar{X}}\left(\min_{y\in\bar{Y}} d(x,y)\right), \max_{y\in\bar{Y}}\left(\min_{x\in\bar{X}} d(x,y)\right)\right) \times 0.95 \tag{10}$$

where $\bar{X}$, $\bar{Y}$ denote the surfaces of X, Y, respectively. $d(x,y)$ denotes the Euclidean distance between the points x and y.

4.1.3. Experimental Environment and Parameter Settings

When training on the ACDC dataset, the hyperparameters batch size, epochs, and the initial learning rate are set to 8, 300, and 0.01, respectively, while on the LA dataset, the hyperparameters batch size, iterations, and the initial learning rate are set to 4, 9000, and 0.01, respectively. On both ACDC and LA datasets, the proposed network is trained to employ the Stochastic Gradient Descent (SGD) optimizer, incorporating momentum of 0.9 and weight decay of 0.0001. Throughout the training process, the poly learning rate strategy is implemented to facilitate learning rate decay. Specifically, the initial learning rate of 0.01 is multiplied by $(1.0 - \frac{t}{t_{max}})^{\eta}$ with $\eta = 0.9$.

We set the hyperparameters α, β, γ of the loss functions L_{con}, L_{mc}, and L_{em} to 0.5, 0.25, and 0.1, respectively. In particular, for γ, since our entropy minimization constraint is based on average prediction, when epoch is less than 15, the prediction results generated in the early stage of model training are not accurate enough, and to prevent the model from deviating from the right direction because of this entropy minimization constraint, it is set to 0. When the epoch is greater than 15, the model training is stable and set to 0.1.

The U-net and V-net models are utilized as the segmentation sub-networks for the ACDC and LA datasets, respectively. The model is trained to utilize the PyTorch framework (torch 1.10.0 + cu113), making use of a single NVIDIA 3080 GPU with 10 GB of memory for the training process.

4.2. Comparison with State-of-the-Art Methods

The experimental results of comparing different algorithms on the ACDC dataset are presented in Table 1. To demonstrate the effectiveness of the aforementioned SSL-based methods compared to the fully supervised methods, we present the experimental results obtained from fully supervised training using only 5% (3 cases), 10% (7 cases), and 100% (full cases) of the total data as training data with U-Net network. These results are then compared with the performance of the aforementioned semi-supervised methods with corresponding ratio labeled cases.

Table 1 shows the average metrics of three categories (right ventricle, myocardium, left ventricle) obtained by different methods on the ACDC test set. when uti-

lizing only 5% (3 labeled cases) and 10% (7 labeled cases) of labeled data, most of the semi-supervised approaches exhibit a significant superiority over fully supervised methods across all four metrics. For instance, our proposed method demonstrates a substantial improvement over the fully supervised method when employing 5% labeled data, with enhancements of 41.24%, 43.81%, 28.58%, and 11.87% for the Dice, Jaccard, HD95, and ASD evaluation metrics, respectively. These results suggest that the semi-supervised methods have utilized the information contained in the unlabeled samples to varying degrees. By further observing Table 1, it is evident that the proposed method achieves the best performance across the four metrics compared to other semi-supervised methods under different semi-supervised settings. For 3 and 7 labeled cases, our method demonstrates superior performance compared to the second-best method in terms of Dice and Jaccard metrics, with improvements of 1.48% and 2.15%, as well as 0.66% and 1.39%, respectively, which indicates that the two core modules proposed by our method, namely multi-scale joint strong-weak consistency and entropy minimization, augment the model's capacity to apprehend and

Table 1

Comparisons with state-of-the-art semi-supervised segmentation methods on the ACDC dataset. The best results for each metric are **bolded**.

Method	Scans used		Metrics			
	Labeled	Unlabeled	Dice(%)↑	Jaccard(%)↑	95HD(voxel)↓	ASD(voxel)↓
U-Net	3(5%)	0	47.83	37.01	31.16	12.62
U-Net	7(10%)	0	79.41	68.11	9.35	2.70
U-Net	70(ALL)	0	91.44	84.59	4.30	0.99
UA-MT [51] (MICCAI'19)	3(5%)	67(95%)	46.04	35.97	20.08	7.75
SASSNet [21] (MICCAI'20)			57.77	46.14	20.05	6.06
DTC [23] (AAAI'21)			56.90	45.67	23.36	7.39
URPC [24] (MICCAI'21)			55.87	44.64	13.60	3.74
MC-Net [47] (MICCAI'21)			62.85	52.29	7.62	2.33
SS-Net [46] (MICCAI'22)			65.82	55.38	6.67	2.28
Cross-ALD [29] (MICCAI'23)			80.60	69.08	5.96	1.9
Fixmatch [37]			87.29	78.08	2.71	0.81
BCP [2] (CVPR'23)			87.59	78.67	**1.90**	**0.67**
Ours (MEMatch)			**89.07**	**80.82**	2.58	0.75
UA-MT [51] (MICCAI'19)	7(10%)	63(90%)	81.65	70.64	6.88	2.02
SASSNet [21] (MICCAI'20)			84.50	74.34	5.43	1.86
DTC [23] (AAAI'21)			84.29	73.96	12.81	4.01
URPC [24] (MICCAI'21)			83.10	72.41	4.84	1.53
MC-Net [47] (MICCAI'21)			86.44	77.04	5.50	1.84
SS-Net [46] (MICCAI'22)			86.78	77.67	6.07	1.40
Cross-ALD [29] (MICCAI'23)			87.52	78.62	4.81	1.6
Fixmatch [37]			88.62	80.16	3.79	114
BCP [2] (CVPR'23)			88.84	80.62	3.98	1.17
Ours (MEMatch)			**89.80**	**82.01**	**2.89**	**0.81**

harness the information encompassed in the labeled and unlabeled data.

The results of the comparison experiments on the LA dataset are presented in Table 2. It can be found that although our method, MEMatch, exhibits slightly inferior performance to both CAML and UniMatch in terms of the Dice and Jaccard coefficients when utilizing 10% of the labeled data, it achieves the best results in terms of Dice, Jaccard, and 95HD when using only 5% of the labeled data. The comparative results on the LA dataset further substantiate the effectiveness of our proposed MEMatch approach.

To illustrate the effectiveness of the proposed MEMatch approach from multiple perspectives, Figures 2-3 show the qualitative comparison of the different methods on the ACDC dataset with 3 and 7 labeled cases. Meanwhile, Figures 4-5 illustrate the qualitative comparison of the different methods on the LA dataset with 4 and 8 labeled cases. The figure shows that the segmentation results of the proposed MEMatch exhibit a greater degree of similarity to the ground truth labels. Meanwhile, in terms of segmentation edges and segmentation shapes, the segmentation results of the proposed

Table 2

Comparisons with state-of-the-art semi-supervised segmentation methods on the LA dataset. The best results for each metric are **bolded**.

Method	Scans used		Metrics			
	Labeled	Unlabeled	Dice(%)↑	Jaccard(%)↑	95HD(voxel)↓	SD(voxel)↓
V-Net	4(5%)	0	52.55	39.60	47.05	9.87
V-Net	8(10%)	0	82.74	71.72	13.35	3.26
V-Net	80(ALL)	0	91.47	84.36	5.48	1.51
UA-MT [51] (MICCAI'19)	4(5%)	76(95%)	82.26	70.98	13.71	3.82
SASSNet [21] (MICCAI'20)			81.60	69.63	16.16	3.58
DTC [23] (AAAI'21)			81.25	69.33	14.90	3.99
URPC [24] (MICCAI'21)			82.48	71.35	14.65	3.65
MC-Net [47] (MICCAI'21)			83.59	72.36	14.07	2.70
SS-Net [46] (MICCAI'22)			86.33	76.15	9.97	2.31
Fixmatch [37]			85.63	75.12	10.61	**2.14**
UniMatch [49] (CVPR'23)			86.08	75.83	12.04	2.85
CAML [16] (MICCAI'23)			87.34	77.65	9.76	2.49
Ours (MEMatch)			**87.51**	**77.93**	**8.74**	2.31
UA-MT [51] (MICCAI'19)	8(10%)	72(90%)	81.65	87.79	78.39	8.68
SASSNet [21] (MICCAI'20)			87.54	78.05	9.84	2.59
DTC [23] (AAAI'21)			87.51	78.17	8.23	2.36
URPC [24] (MICCAI'21)			86.92	77.03	11.13	2.28
MC-Net [47] (MICCAI'21)			87.62	78.25	10.03	**1.82**
SS-Net [46] (MICCAI'22)			88.55	79.62	7.49	1.90
Fixmatch [37]			86.00	75.62	10.60	2.09
UniMatch [49] (CVPR'23)			89.09	80.47	12.50	3.59
CAML [16] (MICCAI'23)			**89.62**	**81.28**	**8.76**	2.02
Ours (MEMatch)			88.81	80.00	9.34	2.45

method in this paper are smoother and finer compared to other methods.

Figure 6-7 illustrate the changes of train loss and dice metrics with epoch on the ACDC dataset. Analyzing the two figures, it is evident that each loss function of the proposed method keeps decreasing and finally stabilizes with the growth of epochs, while the dice metrics of each segmentation category are increasing

Figure 2

Qualitative comparison of different approaches on ACDC dataset with 3 labeled cases. From left to right they represent: a) Ground truth, b) Ours, c) UA-MT [51], d) URPC [24], e)MC-Net [47], f) SS-Net [46], and g) Fixmatch [37].

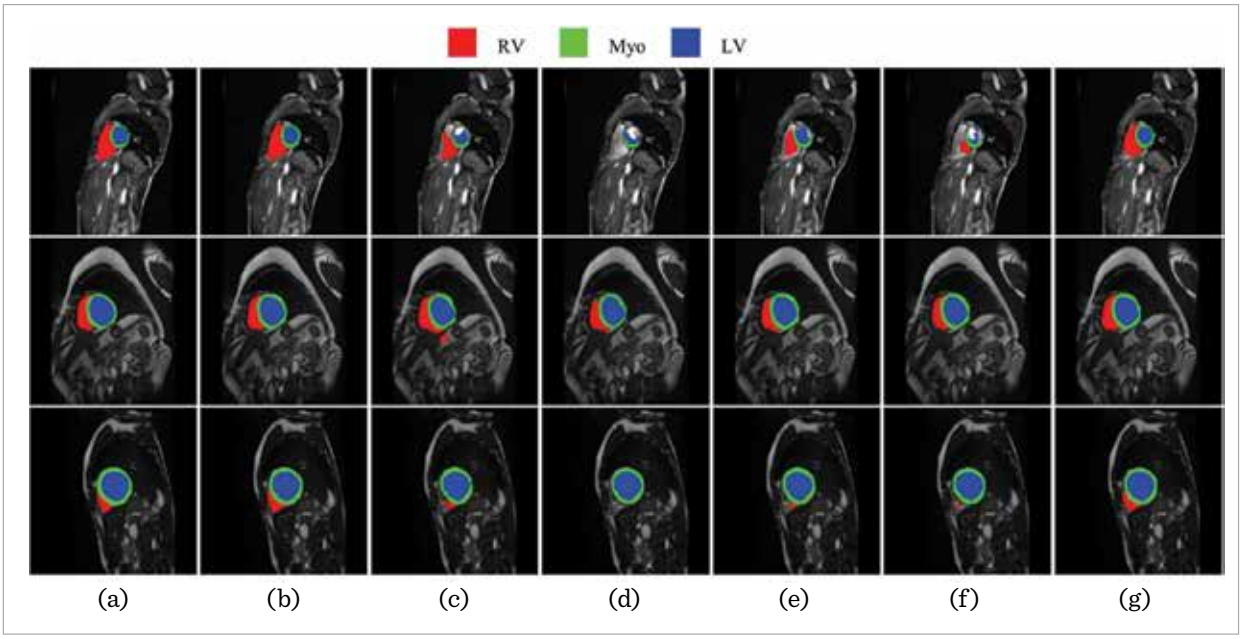


Figure 3

Qualitative comparison of different approaches on ACDC dataset with 7 labeled cases. From left to right they represent: a) Ground truth, b) Ours, c) UA-MT [51], d) URPC [24], e)MC-Net [47], f) SS-Net [46], and g) Fixmatch [37].

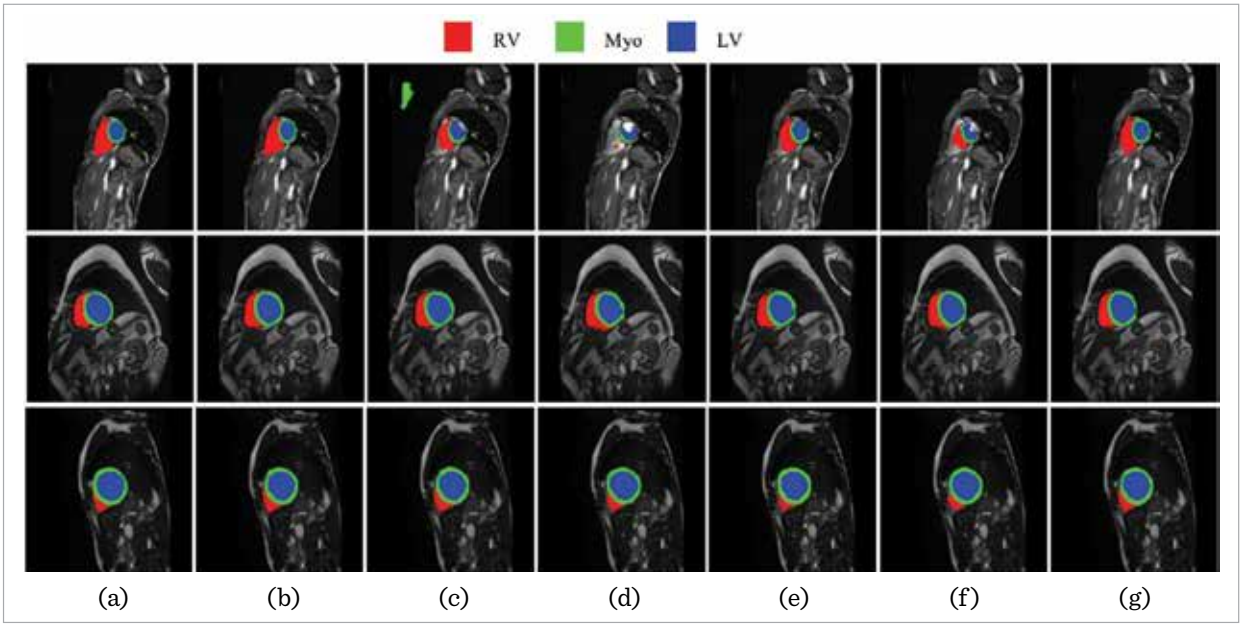


Figure 4

Qualitative comparison of different approaches on LA dataset with 4 labeled cases. From left to right they represent: a) Ground truth, b) Ours, c) DTC [23], d) MC-Net [47], e) SS-Net [46], f) Fixmatch [37], and g) CAML [16].

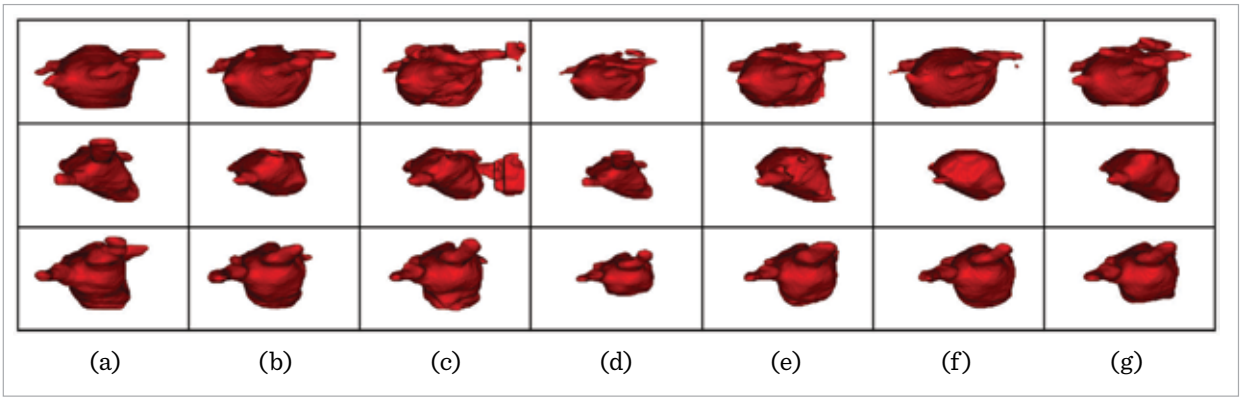


Figure 5

Qualitative comparison of different approaches on LA dataset with 8 labeled cases. From left to right they represent: a) Ground truth, b) Ours, c) DTC [23], d) MC-Net [47], e) SS-Net [46], f) Fixmatch [37], and g) CAML [16].

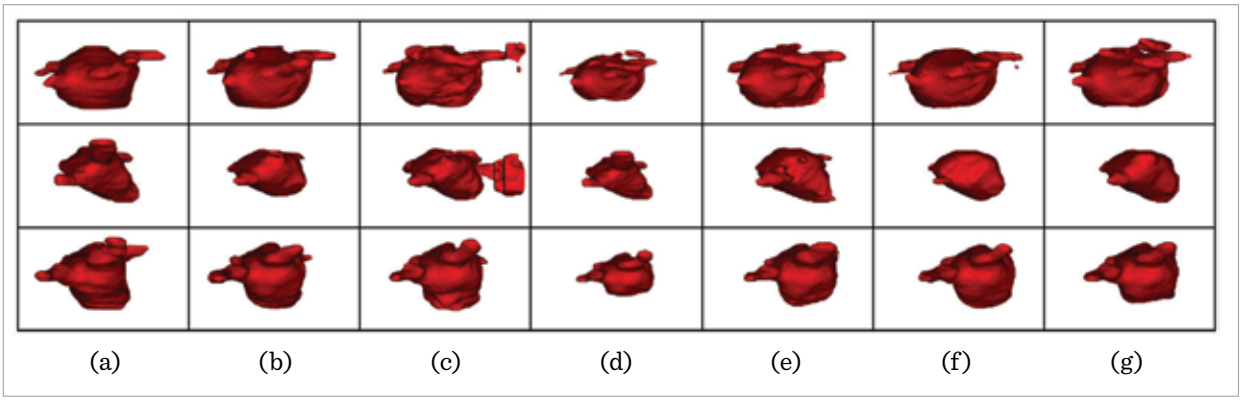


Figure 6

Training loss when using 3 labeled cases. L_{con}, L_{mc}, L_{em} are not multiplied by the weight values. L_{con}, L_{mc}, L_{em} are the abbreviations for consistency loss, multi-scale joint strong-weak consistency loss, and entropy minimization loss on unlabeled data, respectively.

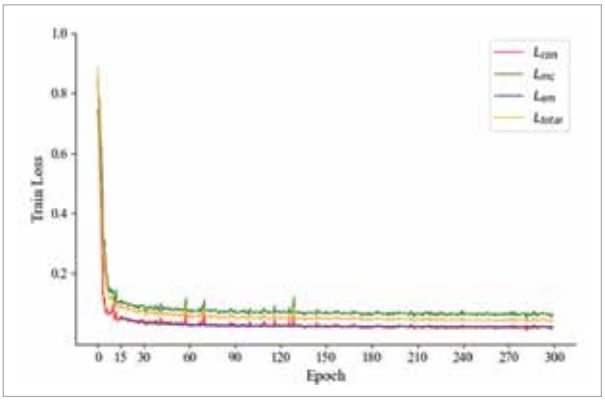
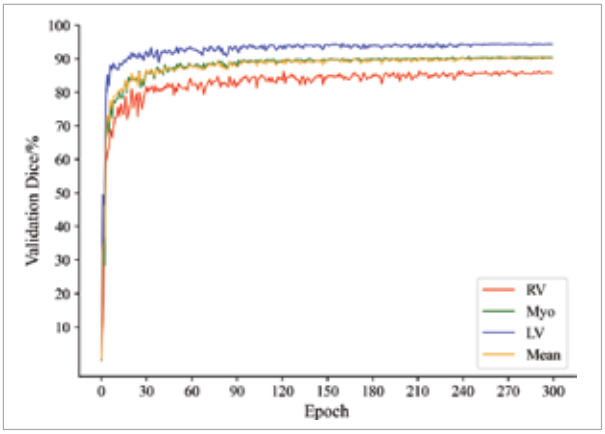


Figure 7

Dice metrics on the validation set when using 3 labeled cases. RV, Myo, and LV denote three different categories, and Mean denotes the average Dice metric for the three categories.



and finally stabilizing. In addition, the standard deviation of the Dice/Jaccard coefficients is 0.035/0.056 for the samples in the test set of the ACDC dataset. This observation confirms the convergence and stability of our proposed approach.

4.3. Ablation Study

4.3.1. Ablation Study on Different Components

To assess the contribution of the different loss functions of our proposed method to the model performance, we conduct ablation studies using the ACDC dataset with 5% labeled data (3 labeled cases). The components under study include L_{seg}, L_{con}, L_{mc}, and L_{em}. The evaluation metrics used are Dice and Jaccard. We present the segmentation metrics results for three categories - RV (right ventricle), Myo (myocardium), and LV (left ventricle) with different combinations of loss functions to facilitate a more detailed comparison. Table 3 displays the results of the ablation studies, showcasing the efficacy of each component proposed in this paper by gradually incorporating the loss functions of each component multiplied by their corresponding weights into the total loss function.

As shown in the first row of Table 3, when utilizing only the linear combination of L_{seg} and L_{con} as the loss function, our method achieves 87.29% and 78.08% on the Dice and Jaccard metrics, respectively. According to the second line of the table, after adding the L_{mc} with weight β to the total loss function, we observe that both metrics of the three categories have significant improvements, with an average increase of 1.35%, and 2.11%. This suggests that our multi-scale joint strong-weak consistency L_{mc} provides consistency on different scales, and captures more valuable information from the unlabeled data. In the third row of the table, it is evident that complete MEMatch with all components achieves optimal performance on both evaluation metrics. This indicates that our L_{em} reduces discrepancies between outputs from different segmentation heads. The above-mentioned findings emphasize the importance of each component in our proposed approach.

4.3.2. Value of the Weights of Loss β and γ

We perform ablation experiment on the weights β for L_{mc} and γ for L_{em} to investigate their impact on the model performance. The experiment is conducted on the ACDC dataset with 5% labeled data (3 labeled cases). The results of the ablation studies for the hyper-parameters β and γ are shown in Table 4.

For the weight β, we set it to 0.05, 0.10, 0.25 (default), and 0.50, respectively, to investigate the effect of various weights of L_{mc} on the segmentation performance of the model. According to the results presented in Table 4, both the Dice and Jaccard metrics generally increase as β increases from 0.05 to 0.25, reaching their highest values when β was set at 0.25, thus confirming the effectiveness of our proposed L_{mc}.

Table 4

Ablation studies on the Hyperparameter β and γ with 5% labeled data (3 labeled cases).

β	0.05	0.10	0.25(default)	0.50
Dice(%)	88.22	88.29	**89.07**	88.44
Jaccard(%)	79.52	79.66	**80.82**	79.89
γ	**0.05**	**0.10 (default)**	**0.25**	**0.50**
Dice(%)	88.68	**89.07**	88.57	88.61

Table 3

Ablation studies of different loss combinations, when using 5% labeled data (3 labeled cases) on the ACDC. L_{seg} represents the supervision loss on the labeled data. L_{con}, L_{mc}, L_{em} are the abbreviations for consistency loss, multi-scale joint strong-weak consistency loss, and entropy minimization loss on unlabeled data, respectively.

Loss				RV		Myo		LV		Mean	
L_{seg}	L_{con}	L_{mc}	L_{em}	Dice(%)	Jaccard(%)	Dice(%)	Jaccard(%)	Dice(%)	Jaccard(%)	Dice(%)	Jaccard(%)
✓	✓			86.88	77.42	84.41	73.37	90.57	83.46	87.29	78.08
✓	✓	✓		87.84	78.95	86.33	76.21	91.76	85.42	88.64	80.19
✓	✓	✓	✓	**88.16**	**79.41**	**86.56**	**76.55**	**92.50**	**86.48**	**89.07**	**80.82**

Regarding the weight γ, we also set it to 0.05, 0.10 (default), 0.25, 0.50 respectively. Observing Table 4, we can see that both Dice and Jaccard achieve scores of 89.07% and 80.82%, respectively, when γ=0.10, which are identified as the optimal scores obtained across the four different ablation values.

4.3.3. Value of the Confidence Threshold τ

Figure 2 illustrates the results of the ablation study of our approach utilizing the Dice and Jaccard metrics to evaluate the impact of the hyperparametric confidence threshold τ on the ACDC dataset. The ablation values for the confidence threshold τ are set to 0.90, 0.95 (default), 0.98, and 0.99, respectively. From Figure 2, it is evident that the highest Jaccard score of 80.82% and Dice index of 89.07% are achieved at a confidence threshold of 0.95 under semi-supervised settings with only 5% labeled data (3 labeled cases).

4.4. Effectiveness of the Proposed Method on Lower Labeled Cases

For evaluating the performance of the MEMatch on lower labeled data, we utilize only 1 labeled case from the ACDC and LA datasets as our labeled data. Table 5 displays the results of the ablation experiments. For ACDC, our method achieved performance scores of 85.70%, 76.06%, 6.17, and 1.55 on the four metrics, respectively. For LA dataset, our method achieved performance scores of 81.31%, 69.10%, 17.25, and 3.52 on the four metrics, respectively. Upon analyzing the experimental results and comparing them with those in Tables 1-2. Upon analysis of the tables, it is evident that our method on the ACDC dataset achieves segmentation results using only 1 labeled case that is even superior to those obtained by most methods using 3 labeled cases. Meanwhile, on the LA dataset, our method achieved comparable performance to DTC [23] using only 1 labeled case, whereas DTC utilized 4 labeled cases to reach its metrics. These results illustrate the efficacy of the proposal approach in handling lower amounts of labeled samples.

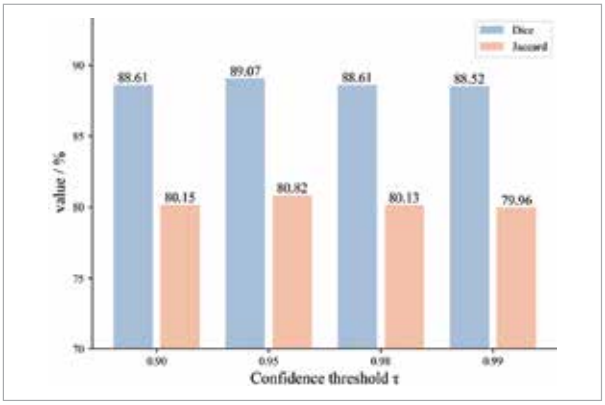
Table 5

Experimental results on 1 labeled case

Dataset	Dice	Jaccard	95HD	ASD
ACDC	85.70	76.06	6.17	1.55
LA	81.31	69.10	17.25	3.52

Figure 8

Ablation studies on the threshold τ with 5% labeled data (3 labeled cases).



5. Conclusion

In this paper, the MEMatch framework, utilized for the semi-supervised segmentation of cardiac MRI images, is presented. The main idea of MEMatch is to leverage the multi-scale predictions generated by strongly-augmented unlabeled images and the pseudo-labels generated by weakly-augmented unlabeled images to construct multi-scale consistency. To further constrain multi-scale predictions of the same network to converge to similarity, we apply entropy minimization to the average prediction over multiple scales. The comparison results on the ACDC and LA datasets illustrate superior performance and reliability of the proposed MEMatch. This improvement has led to more precise cardiac diagnoses and more personalized treatment plans, resulting in significantly improved patient outcomes in the field of cardiac care. However, our method does not specifically add different network perturbations for different segmentation heads but simply uses different up-sampling strategies to enhance the variability of segmentation predictions at different scales. In subsequent work, we can further extend this variability by adding different attention mechanisms to different segmentation heads, or by using different types of convolution in different segmentation heads, or by constructing a loss of mutual consistency between the segmentation results of the individual segmentation heads to obtain stronger consistency constraints. In addition, this paper will further improve the network structure in future work to reduce the consumption of computational resources for model training.

Acknowledgments

This paper is supported in part by the National Natural Science Foundation of China (grant No. 61807029), the Natural Science Foundation of Hebei Province (grant No. F2019203427).

Data Availability Statement

The dataset used in this article is from the Automated Cardiac Diagnosis Challenge [4] at https://www.creatis.insa-lyon.fr/Challenge/acdc/ (accessed on 11 January 2024)

and the 2018 Atria Segmentation Challenge [48] at

https://www.cardiacatlas.org/atriaseg2018-challenge/ (accessed on 11 January 2024).

Conflicts of Interest

The authors declare no conflicts of interest.

References

1. Adiga Vasudeva, S., Dolz, J., Lombaert, H. Leveraging Labeling Representations in Uncertainty-Based Semi-Supervised Segmentation. In Wang, L., Dou, Q., Fletcher, P. T., Speidel, S., Li, S. (Eds.), Medical Image Computing and Computer Assisted Intervention - MICCAI 2022, Springer, Cham, 2022, 265-275. https://doi.org/10.1007/978-3-031-16452-1_26
2. Bai, Y., Chen, D., Li, Q., Shen, W., Wang, Y. Bidirectional Copy-Paste for Semi-Supervised Medical Image Segmentation. In 2023 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), 2023, 11514-11524. https://doi.org/10.1109/CVPR52729.2023.01108
3. Basak, H., Bhattacharya, R., Hussain, R., Chatterjee, A. An Exceedingly Simple Consistency Regularization Method for Semi-Supervised Medical Image Segmentation. In 2022 IEEE 19th International Symposium on Biomedical Imaging (ISBI), 2022, 1-4. https://doi.org/10.1109/ISBI52829.2022.9761602
4. Bernard, O., Lalande, A., Zotti, C., Cervenansky, F., Yang, X., Heng, P.-A., Cetin, I., Lekadir, K., Camara, O., Gonzalez Ballester, M. A., Sanroma, G., Napel, S., Petersen, S., Tziritas, G., Grinias, E., Khened, M., Kollerathu, V. A., Krishnamurthi, G., Rohé, M.-M., Pennec, X., Sermesant, M., Isensee, F., Jäger, P., Maier-Hein, K. H., Full, P. M., Wolf, I., Engelhardt, S., Baumgartner, C. F., Koch, L. M., Wolterink, J. M., Išgum, I., Jang, Y., Hong, Y., Patravali, J., Jain, S., Humbert, O., Jodoin, P.-M. Deep Learning Techniques for Automatic MRI Cardiac Multi-Structures Segmentation and Diagnosis: Is the Problem Solved? IEEE Transactions on Medical Imaging, 2018, 37(11), 2514-2525. https://doi.org/10.1109/TMI.2018.2837502
5. Berthelot, D., Carlini, N., Cubuk, E. D., Kurakin, A., Sohn, K., Zhang, H., Raffel, C. ReMixMatch: Semi-Supervised Learning with Distribution Alignment and Augmentation Anchoring. arXiv preprint arXiv:1911.09785 (2019). doi:10.48550/arXiv.1911.09785.
6. Berthelot, D., Carlini, N., Goodfellow, I., Papernot, N., Oliver, A., Raffel, C. A. MixMatch: A Holistic Approach to Semi-Supervised Learning. In Wallach, H., Larochelle, H., Beygelzimer, A., d'Alché-Buc, F., Fox, E., Garnett, R. (Eds.), Advances in Neural Information Processing Systems, Vol. 32, Curran Associates, Inc., 2019, 5049-5059. doi:10.48550/arXiv.1905.02249.
7. Bhan, A., Goyal, A., Dutta, M. K., Riha, K., Omran, Y. Image-Based Pixel Clustering and Connected Component Labeling in Left Ventricle Segmentation of Cardiac MR Images. 2015 7th International Congress on Ultra Modern Telecommunications and Control Systems and Workshops (ICUMT), 2015, 339-342. https://doi.org/10.1109/ICUMT.2015.7382454
8. Chang, Q., Yan, Z., Lou, Y., Axel, L., Metaxas, D. N. Soft-Label Guided Semi-Supervised Learning for Bi-Ventricle Segmentation in Cardiac Cine MRI. 2020 IEEE 17th International Symposium on Biomedical Imaging (ISBI), 2020, 1752-1755. https://doi.org/10.1109/ISBI45749.2020.9098546
9. Chen, J., Lu, Y., Yu, Q., Luo, X., Adeli, E., Wang, Y., Lu, L., Yuille, A. L., Zhou, Y. TransUNet: Transformers Make Strong Encoders for Medical Image Segmentation. CoRR abs/2102.04306 (2021). arXiv:2102.04306. doi:10.48550/arXiv.2102.04306.
10. Chong, Y., Ding, Y., Yan, Q., Pan, S. Graph-Based Semi-Supervised Learning: A Review. Neurocomputing, 2020, 408, 216-230. https://doi.org/10.1016/j.neucom.2019.12.130
11. Das, N., Das, S. Attention-UNet Architectures with Pretrained Backbones for Multi-Class Cardiac MR Image Segmentation. Current Problems in Cardiology, 2024, 49(1, Part C), 102129. https://doi.org/10.1016/j.cpcardiol.2023.102129

12. Decourt, C., Duong, L. Semi-Supervised Generative Adversarial Networks for the Segmentation of the Left Ventricle in Pediatric MRI. Computers in Biology and Medicine, 2020, 123, 103884. https://doi.org/10.1016/j.compbiomed.2020.103884

13. Dike, H. U., Zhou, Y., Deveerasetty, K. K., Wu, Q. Unsupervised Learning Based on Artificial Neural Network: A Review. 2018 IEEE International Conference on Cyborg and Bionic Systems (CBS), 2018, 322-327. https://doi.org/10.1109/CBS.2018.8612259

14. Gaidai, O., Cao, Y., Loginov, S. Global Cardiovascular Diseases Death Rate Prediction. Current Problems in Cardiology, 2023, 48(5), 101622. https://doi.org/10.1016/j.cpcardiol.2023.101622

15. Galic, S., Loncaric, S. Cardiac Image Segmentation Using Spatiotemporal Clustering. Medical Imaging: Image Processing, 2001, 1199-1206. https://doi.org/10.1117/12.430996

16. Gao, S., Zhang, Z., Ma, J., Li, Z., Zhang, S. Correlation-Aware Mutual Learning for Semi-Supervised Medical Image Segmentation. In Greenspan, H., Madabhushi, A., Mousavi, P., Salcudean, S., Duncan, J., Syeda-Mahmood, T., Taylor, R. (Eds.), Medical Image Computing and Computer Assisted Intervention - MICCAI 2023, Springer, Cham, 2023, 98-108. https://doi.org/10.1007/978-3-031-43907-0_10

17. Grandvalet, Y., Bengio, Y. Semi-Supervised Learning by Entropy Minimization. In Saul, L., Weiss, Y., Bottou, L. (Eds.), Advances in Neural Information Processing Systems, MIT Press, 2004, 17, 529-536.

18. Hang, W., Feng, W., Liang, S., Yu, L., Wang, Q., Choi, K.-S., Qin, J. Local and Global Structure-Aware Entropy Regularized Mean Teacher Model for 3D Left Atrium Segmentation. In Martel, A. L., Abolmaesumi, P., Stoyanov, D., Mateus, D., Zuluaga, M. A., Zhou, S. K., Racoceanu, D., Joskowicz, L. (Eds.), Medical Image Computing and Computer Assisted Intervention - MICCAI 2020, Springer, Cham, 2020, 562-571. https://doi.org/10.1007/978-3-030-59710-8_55

19. Lei, W., Su, Q., Jiang, T., Gu, R., Wang, N., Liu, X., Wang, G., Zhang, X., Zhang, S. One-Shot Weakly-Supervised Segmentation in 3D Medical Images. IEEE Transactions on Medical Imaging, 2024, 43(1), 175-189. https://doi.org/10.1109/TMI.2023.3294975

20. Li, C., Xu, K., Zhu, J., Liu, J., Zhang, B. Triple Generative Adversarial Networks. IEEE Transactions on Pattern Analysis and Machine Intelligence, 2022, 44(12), 9629-9640. https://doi.org/10.1109/TPAMI.2021.3127558

21. Li, S., Zhang, C., He, X. Shape-Aware Semi-Supervised 3D Semantic Segmentation for Medical Images. In Medical Image Computing and Computer-Assisted Intervention - MICCAI 2020: 23rd International Conference, Lima, Peru, October 4-8, 2020, Proceedings, Part I, Springer, 2020, 552-561. https://doi.org/10.1007/978-3-030-59710-8_54

22. Liu, H., Hu, H., Xu, X., Song, E. Automatic Left Ventricle Segmentation in Cardiac MRI Using Topological Stable-State Thresholding and Region Restricted Dynamic Programming. Academic Radiology, 2012, 19(6), 723-731. https://doi.org/10.1016/j.acra.2012.02.011

23. Luo, X., Chen, J., Song, T., Wang, G. Semi-Supervised Medical Image Segmentation Through Dual-Task Consistency. Proceedings of the AAAI Conference on Artificial Intelligence, 2021, 35(10), 8801-8809. https://doi.org/10.1609/aaai.v35i10.17066

24. Luo, X., Liao, W., Chen, J., Song, T., Chen, Y., Zhang, S., Chen, N., Wang, G., Zhang, S. Efficient Semi-Supervised Gross Target Volume of Nasopharyngeal Carcinoma Segmentation via Uncertainty Rectified Pyramid Consistency. In de Bruijne, M., Cattin, P. C., Cotin, S., Padoy, N., Speidel, S., Zheng, Y., Essert, C. (Eds.), Medical Image Computing and Computer Assisted Intervention - MICCAI 2021, Springer, Cham, 2021, 318-329. https://doi.org/10.1007/978-3-030-87196-3_30

25. Ma, W., Yao, H., Lin, Y., Guo, J., Li, X. Semi-Supervised Domain Generalization for Cardiac Magnetic Resonance Image Segmentation with High Quality Pseudo Labels. In Camara, O., Puyol-Antón, E., Qin, C., Sermesant, M., Suinesiaputra, A., Wang, S., Young, A. (Eds.), Statistical Atlases and Computational Models of the Heart. Regular and CMRxMotion Challenge Papers, Springer, Cham, 2022, 383-391. https://doi.org/10.1007/978-3-031-23443-9_35

26. Meena, K. S., Suriya, S. A Survey on Supervised and Unsupervised Learning Techniques. In Kumar, L. A., Jayashree, L. S., Manimegalai, R. (Eds.), Proceedings of International Conference on Artificial Intelligence, Smart Grid and Smart City Applications, Springer, Cham, 2020, 627-644. https://doi.org/10.1007/978-3-030-24051-6_58

27. Milletari, F., Navab, N., Ahmadi, S.-A. V-Net: Fully Convolutional Neural Networks for Volumetric Medical Image Segmentation. 2016 Fourth International Conference on 3D Vision (3DV), 2016, 565-571. https://doi.org/10.1109/3DV.2016.79

28. Miranda Teixeira, G., Ramalho Pommeranzembaum, I., de Oliveira, B. L., Lobosco, M., Weber dos Santos,

R. Automatic Segmentation of Cardiac MRI Using Snakes and Genetic Algorithms. In Bubak, M., van Albada, G. D., Dongarra, J., Sloot, P. M. A. (Eds.), Computational Science - ICCS 2008, Springer, Berlin, Heidelberg, 2008, 168-177. https://doi.org/10.1007/978-3-540-69389-5_20

29. Nguyen-Duc, T., Le, T., Bammer, R., Zhao, H., Cai, J., Phung, D. Cross-Adversarial Local Distribution Regularization for Semi-Supervised Medical Image Segmentation. In Greenspan, H., Madabhushi, A., Mousavi, P., Salcudean, S., Duncan, J., Syeda-Mahmood, T., Taylor, R. (Eds.), Medical Image Computing and Computer Assisted Intervention - MICCAI 2023, Springer, Cham, 2023, 183-194. https://doi.org/10.1007/978-3-031-43907-0_18
30. Pluempitiwiriyawej, C., Moura, J., Wu, Y.-J. L., Ho, C. STACS: New Active Contour Scheme for Cardiac MR Image Segmentation. IEEE Transactions on Medical Imaging, 2005, 24(5), 593-603. https://doi.org/10.1109/TMI.2005.843740
31. Priya, E., Srinivasan, S., Nageswararao, A. V. Inhomogeneity Correction and Hybrid-Based Segmentation in Cardiac MRI. International Journal of Biomedical Engineering and Technology, 2018, 28(4), 349-365. https://doi.org/10.1504/IJBET.2018.095983
32. Rajiah, P. S., François, C. J., Leiner, T. Cardiac MRI: State of the Art. Radiology, 2023, 307(3), e223008. https://doi.org/10.1148/radiol.223008
33. Ronneberger, O., Fischer, P., Brox, T. U-Net: Convolutional Networks for Biomedical Image Segmentation. In Medical Image Computing and Computer-Assisted Intervention - MICCAI 2015: 18th International Conference, Munich, Germany, October 5-9, 2015, Proceedings, Part III, Springer, 2015, 234-241. https://doi.org/10.1007/978-3-319-24574-4_28
34. Sadeghi, Z., Alizadehsani, R., Cifci, M. A., Kausar, S., Rehman, R., Mahanta, P., Bora, P. K., Almasri, A., Alkhawaldeh, R. S., Hussain, S., Alatas, B., Sheibi, A., Moosaei, H., Hladik, M., Nahavandi, S., Pardalos, P. M. A Brief Review of Explainable Artificial Intelligence in Healthcare. arXiv preprint arXiv:2304.01543 (2023). https://doi.org/10.2139/ssrn.4600029
35. Sharifrazi, D., Alizadehsani, R., Hoseini Izadi, N., Roshanzamir, M., Shoeibi, A., Khozeimeh, F., Alizadeh Sani, F., Alizadeh Sani, Z., Hussain, S., Harlapur, C., Gorriz, J. M., Khosravi, A., Nahavandi, S., Sarrafzadegan, N., Shariful Islam, S. M. Hypertrophic Cardiomyopathy Diagnosis Based on Cardiovascular Magnetic Resonance Using Deep Learning Techniques. Research Square, 2021. https://doi.org/10.2139/ssrn.3855445
36. Shewaye, T. N. Cardiac MR Image Segmentation Techniques: An Overview. CoRR abs/1502.04252 (2015). arXiv:1502.04252. doi:10.48550/arXiv.1502.04252.
37. Sohn, K., Berthelot, D., Carlini, N., Zhang, Z., Zhang, H., Raffel, C. A., Cubuk, E. D., Kurakin, A., Li, C.-L. FixMatch: Simplifying Semi-Supervised Learning with Consistency and Confidence. Advances in Neural Information Processing Systems, 33 (2020), 596-608. doi:10.48550/arXiv.2001.07685.
38. Song, Z., Yang, X., Xu, Z., King, I. Graph-Based Semi-Supervised Learning: A Comprehensive Review. IEEE Transactions on Neural Networks and Learning Systems, 2023, 34(11), 8174-8194. https://doi.org/10.1109/TNNLS.2022.3155478
39. Thampi, S. M., Bandyopadhyay, S., Krishnan, S., Li, K. C., Mosin, S., Ma, M. Automatic Left Ventricle Segmentation in Cardiac MRI Images Using a Membership Clustering and Heuristic Region-Based Pixel Classification Approach. In Intelligent Systems Technologies and Applications, Springer International Publishing, 2016, 615-623. https://doi.org/10.1007/978-3-319-28658-7_52
40. Van Ginneken, B., Schaefer-Prokop, C. M., Prokop, M. Computer-Aided Diagnosis: How to Move from the Laboratory to the Clinic. Radiology, 2011, 261(3), 719-732. https://doi.org/10.1148/radiol.11091710
41. Varga-Szemes, A., Muscogiuri, G., Schoepf, U. J., Wichmann, J. L., Suranyi, P., De Cecco, C. N., Cannao, P. M., Renker, M., Mangold, S., Fox, M. A. Clinical Feasibility of a Myocardial Signal Intensity Threshold-Based Semi-Automated Cardiac Magnetic Resonance Segmentation Method. European Radiology, 2016, 26(5), 1503-1511. https://doi.org/10.1007/s00330-015-3952-4
42. Vaswani, A., Shazeer, N., Parmar, N., Uszkoreit, J., Jones, L., Gomez, A. N., Kaiser, Ł., Polosukhin, I. Attention Is All You Need. In Guyon, I., Luxburg, U. V., Bengio, S., Wallach, H., Fergus, R., Vishwanathan, S., Garnett, R. (Eds.), Advances in Neural Information Processing Systems, 2017, 30, 5998-6008. https://doi.org/10.5040/9781350101272.00000005
43. Vu, T.-H., Jain, H., Bucher, M., Cord, M., Pérez, P. ADVENT: Adversarial Entropy Minimization for Domain Adaptation in Semantic Segmentation. In 2019 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), 2019, 2512-2521. https://doi.org/10.1109/CVPR.2019.00262
44. Wei Zhu, S. H. K., Biros, G. A Geodesic-Active-Contour-Based Variational Model for Short-Axis Cardiac MR Image Segmentation. International Journal of

Computer Mathematics, 2013, 90(1), 124-139. https://doi.org/10.1080/00207160.2012.695355

45. Wu, J., Fan, H., Zhang, X., Lin, S., Li, Z. Semi-Supervised Semantic Segmentation via Entropy Minimization. In 2021 IEEE International Conference on Multimedia and Expo (ICME), 2021, 1-6. https://doi.org/10.1109/ICME51207.2021.9428304
46. Wu, Y., Wu, Z., Wu, Q., Ge, Z., Cai, J. Exploring Smoothness and Class-Separation for Semi-Supervised Medical Image Segmentation. In International Conference on Medical Image Computing and Computer-Assisted Intervention, Springer, 2022, 34-43. https://doi.org/10.1007/978-3-031-16443-9_4
47. Wu, Y., Xu, M., Ge, Z., Cai, J., Zhang, L. Semi-Supervised Left Atrium Segmentation with Mutual Consistency Training. In de Bruijne, M., Cattin, P. C., Cotin, S., Padoy, N., Speidel, S., Zheng, Y., Essert, C. (Eds.), Medical Image Computing and Computer Assisted Intervention - MICCAI 2021, Springer, Cham, 2021, 297-306. https://doi.org/10.1007/978-3-030-87196-3_28
48. Xiong, Z., Xia, Q., Hu, Z., Huang, N., Bian, C., Zheng, Y., Vesal, S., Ravikumar, N., Maier, A., Yang, X., Heng, P.-A., Ni, D., Li, C., Tong, Q., Si, W., Puybareau, E., Khoudli, Y., Géraud, T., Chen, C., Bai, W., Rueckert, D., Xu, L., Zhuang, X., Luo, X., Jia, S., Sermesant, M., Liu, Y., Wang, K., Borra, D., Masci, A., Corsi, C., de Vente, C., Veta, M., Karim, R., Preetha, C. J., Engelhardt, S., Qiao, M., Wang, Y., Tao, Q., Nuñez-Garcia, M., Camara, O., Savioli, N., Lamata, P., Zhao, J. A Global Benchmark of Algorithms for Segmenting the Left Atrium from Late Gadolinium-Enhanced Cardiac Magnetic Resonance Imaging. Medical Image Analysis, 2021, 67, 101832. https://doi.org/10.1016/j.media.2020.101832
49. Yang, L., Qi, L., Feng, L., Zhang, W., Shi, Y. Revisiting Weak-to-Strong Consistency in Semi-Supervised Semantic Segmentation. Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, 2023, 7236-7246. https://doi.org/10.1109/CVPR52729.2023.00699
50. Yang, X., Song, Z., King, I., Xu, Z. A Survey on Deep Semi-Supervised Learning. IEEE Transactions on Knowledge and Data Engineering, 2023, 35(9), 8934-8954. https://doi.org/10.1109/TKDE.2022.3220219
51. Yu, L., Wang, S., Li, X., Fu, C.-W., Heng, P.-A. Uncertainty-Aware Self-Ensembling Model for Semi-Supervised 3D Left Atrium Segmentation. In Shen, D., Liu, T., Peters, T. M., Staib, L. H., Essert, C., Zhou, S., Yap, P.-T., Khan, A. (Eds.), Medical Image Computing and Computer Assisted Intervention - MICCAI 2019, Springer, Cham, 2019, 605-613. https://doi.org/10.1007/978-3-030-32245-8_67
52. Zhao, F., Fan, J. Image Segmentation by Multi-Level Thresholding Based on C-Means Clustering Algorithms and Fuzzy Entropy. In Fuzzy Information and Engineering: Proceedings of the Second International Conference of Fuzzy Information and Engineering (ICFIE), Springer, 2007, 113-121. https://doi.org/10.1007/978-3-540-71441-5_13
53. Zhou, Z.-H. A Brief Introduction to Weakly Supervised Learning. National Science Review, 2018, 5(1), 44-53. https://doi.org/10.1093/nsr/nwx106