Received: 15 July 2024 | Revised: 23 October 2024 | Accepted: 7 November 2024

DOI: 10.1002/mp.17547

RESEARCH ARTICLE

MEDICAL PHYSICS

Semi-supervised medical image segmentation network based on mutual learning

Junmei Sun[1] | **Tianyang Wang**[1] | **Meixi Wang**[1] | **Xiumei Li**[1] | **Yingying Xu**[2]

[1]School of Information Science and Technology, Hangzhou Normal University, Hangzhou, China

[2]Zhejiang Lab, Hangzhou, China

Correspondence
Xiumei Li, School of Information Science and Technology, Hangzhou Normal University, Hangzhou, China.
Email: lixiumei@hznu.edu.cn

Funding information
Science and Technology Plan Project of Hangzhou City, China, Grant/Award Number: 20201203B124

Abstract

Background: Semi-supervised learning provides an effective means to address the challenge of insufficient labeled data in medical image segmentation tasks. However, when a semi-supervised segmentation model is overfitted and exhibits cognitive bias, its performance will deteriorate. Errors stemming from cognitive bias can quickly amplify and become difficult to correct during the training process of neural networks, resulting in the continuous accumulation of erroneous knowledge.

Purpose: To address the issue of error accumulation, a novel learning strategy is required to enhance the accuracy of medical image segmentation.

Methods: This paper proposes a semi-supervised medical image segmentation network based on mutual learning (MLNet) to alleviate the issue of continuous accumulation of erroneous knowledge. The MLNet adopts a teacher-student network as the backbone framework, training student and teacher networks on labeled data and mutually updating network parameter weights, enabling the two models to learn from each other. Additionally, an image partial exchange algorithm (IPE) as an appropriate perturbation addition method is proposed to reduce the introduction of erroneous information and the disruption to the contextual information of the image.

Results: In the 10% labeled experiment on the ACDC dataset, our Dice coefficient reached 89.48%, a 9.28% improvement over the baseline model. In the 10% labeled experiment on the BraTS2019 dataset, the proposed method still performs exceptionally well, achieving 84.56%, surpassing other comparative methods.

Conclusions: Compared with other methods, experimental results demonstrate that our approach achieves optimal performance across all metrics, proving its effectiveness and reliability.

KEYWORDS
mean-teacher, medical image segmentation, mutual learning, semi-supervised learning

1 | INTRODUCTION

Medical images, such as computed tomography (CT) and magnetic resonance imaging (MRI), play a crucial role in disease diagnosis by physicians. Medical image segmentation is a key step in building computer-aided diagnostic systems. With precise segmentation results, the morphological properties of organs can be quantitatively analyzed, providing reliable diagnostic evidence for clinical physicians.

With the development of deep learning technology, its extensive applications in the field of medical image segmentation have achieved significant results.[1] However, the success of deep learning often relies on large-scale labeled data. Unlike natural images, the labeling of medical images requires manual work by experts, which is both time-consuming and labor-intensive. To address this challenge, a semi-supervised learning approach is usually adopted, utilizing a small amount of labeled data and a large amount of unlabeled

data for model training to reduce reliance on labeled data.

The Mean-Teacher (MT) method is a classic approach for semi-supervised medical image segmentation based on consistency learning.[2] The core concept of consistency learning involves utilizing a consistency loss function to encourage the model to produce similar outputs for the same sample when it is subjected to different perturbations. By encouraging this consistency, the model learns to recognize the underlying distribution and features of the unlabeled data. As a result, the model's robustness and generalization abilities are significantly improved. In the MT method, typically two models are constructed: a teacher model and a student model. Small perturbations, such as Gaussian noise, are added to the input or feature layers of the student model. Subsequently, a consistency constraint is enforced between the predictions of the student and teacher models. The student model's parameters are updated using backpropagation and gradient descent, while the teacher model's parameters are updated using the Exponential Moving Average (EMA). The EMA is a technique that updates the teacher model's parameters by combining the current student model's parameters and the historical teacher model's parameters using an exponential decay. Many methods have been inspired by the MT method and achieved good performance in semi-supervised learning.[3–5] However, the MT method also has some shortcomings: (1) As iterations increase, the teacher model accumulates all student model knowledge, including errors from cognitive bias. Without supervised training, the teacher model cannot correct these errors, hindering the student's further optimization. (2) Choosing the appropriate perturbation method is challenging: weak perturbations may be ineffective, while strong ones can introduce errors, reducing performance.

To address the above issues in the MT method, this paper proposes a semi-supervised medical image segmentation network based on mutual learning (MLNet). MLNet consists of three modules: the image partial exchange algorithm (IPE), mutual learning strategy (MLS), and cross pseudo supervision module (CPSM).[6] The main contributions of this paper are threefold:

1. A MLS is proposed, enabling the teacher model to learn knowledge from labeled data similar to the student model, thereby alleviating the issue of accumulation of errors in the teacher model. Furthermore, the two networks share their learned knowledge by mutually updating parameters, which helps improve the model's performance.
2. To alleviate the issues of perturbations affecting model performance, an IPE algorithm is proposed. IPE is an appropriate perturbation addition method designed to minimize erroneous information and disruption to the contextual information of the image.
3. Through extensive experimentation, the proposed MLNet outperforms state-of-the-art semi-supervised methods on both the ACDC and BraTS2019 datasets. With only 10% labeled data, our method achieves a Dice similarity coefficient of 89.48% on the ACDC dataset and 84.56% on the BraTS2019 dataset.

2 | METHODS

For general semi-supervised image processing methods, an image X is defined as $X \in R^{H\times W}$, where the goal of semi-supervised image segmentation is to predict a label $Y \in \{0, 1, \ldots, K-1\}^{H\times W}$ within the image X, where K represents the number of classes. The training set is generally divided into two parts: labeled data $\{X^l, Y^l\}$ and unlabeled data $\{X^u\}$.

The overall structure of the proposed semi-supervised medical image segmentation MLNet is shown in Figure 1. The identities of the two networks $f(\theta)$ and $f(\theta')$, which have identical structures but different initialization parameters, are not fixed as in the MT method and can be interchanged. Both networks are trained on labeled data, updating network parameters through the supervision signal provided by the labels. Additionally, these two networks can learn from each other, utilizing EMA to update parameters mutually. Furthermore, to explicitly reinforce class information, pseudo-labels are provided for these two networks for cross pseudo supervision. MLNet leverages the different knowledge learned from both the CPSM and a consistency learning method, significantly improving performance in scenarios with varying levels of supervision through the exchange of knowledge between $f(\theta)$ and $f(\theta')$. The MLNet consists of two training steps. In the first step, as shown in Figure 1a, $f(\theta)$ is treated as the student model, and $f(\theta')$ as the teacher model. Labeled and unlabeled data are selected and obtained augmented data using IPE. The labeled data is used to supervise $f(\theta)$, while the unlabeled data and augmented data are used to calculate consistency loss between predictions. In the second step, as shown in Figure 1b, MLNet treats $f(\theta')$ as the student model and $f(\theta)$ as the teacher model. Labeled data is input into $f(\theta')$ and supervised by its label. Unlabeled data is input into both models to generate pseudo-labels, and CPSM is applied.

2.1 | Mutual learning strategy

Deep Mutual Learning (DML) proposed the idea of mutual learning and applied it in the training process of multiple networks.[7] This means that not only can teachers impart knowledge to students, but students can also share their insights, experiences, and perspectives with teachers. In this learning environment, the

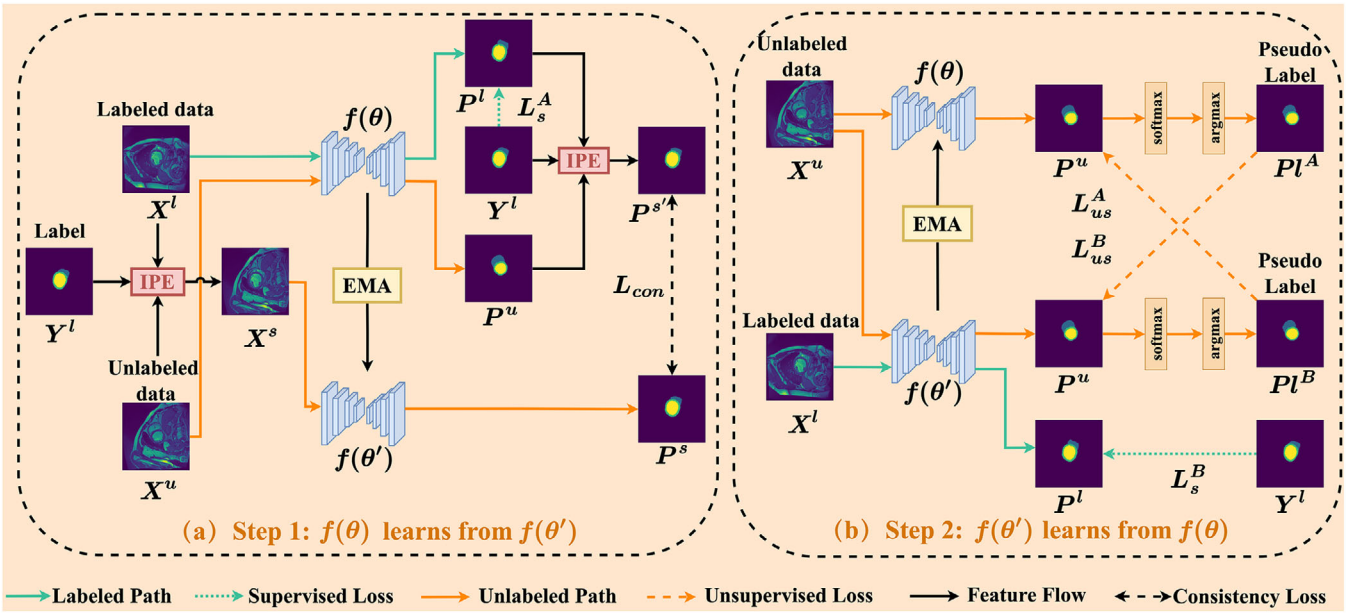

FIGURE 1 The overall framework of the proposed MLNet. The training process employs a MLS divided into two steps: (a) $f(\theta)$ learns from $f(\theta')$, and (b) $f(\theta')$ learns from $f(\theta)$. In (a), MLNet obtains augmented data through the IPE algorithm and uses labeled data to train $f(\theta)$, while using unlabeled data and augmented data for consistency learning. In (b), MLNet uses labeled data to train $f(\theta')$ and applies the CPSM with unlabeled data. CPSM, cross pseudo supervision module; IPE, image partial exchange algorithm; MLNet, network based on mutual learning; MLS, mutual learning strategy.

relationship between teachers and students is more like that of partners, jointly exploring the depth and breadth of knowledge. DML enhances the performance of the entire model by enabling multiple neural networks to learn from each other during training. However, mutual learning solely through mutual supervision is insufficient, as the model may still suffer from the problem of error accumulation. Inspired by DML, this paper applies the idea of mutual learning to MT Method, proposing a MLS. The MLS further enhances the mutual learning capability by allowing two networks to mutually update parameters, thereby alleviating the issue of the teacher network in the MT method continuously accumulating erroneous knowledge. The flowchart of this strategy is shown in Figure 1.

This paper employs a MLS using two networks with identical structures but different initialization parameters. Both segmentation networks are set to trainable mode and can undergo supervised training on labeled data. The MLS consists of two steps. In the first step as shown in Figure 1a, labeled data X^l is input into the $f(\theta)$, while unlabeled data X^u and augmentation data X^s are input into the $f(\theta)$ and $f(\theta')$ respectively. Then, the supervised loss L_s^A between label Y^l and prediction P^l is computed. The consistency loss L_{con} between prediction $P^{s'}$ and P^s is also computed. Finally, the parameters of the teacher network are updated through EMA. The process of the first step is shown from Equations (1) to (4), where P^l represents the $f(\theta)$'s output logits for labeled data, P^u represents the $f(\theta)$' s output logits for unlabeled data, P^s represents the $f(\theta')$'s output logits for augmented data, and $P^{s'}$ is the output obtained from P^l, P^u, and Y^l through the IPE algorithm. The formula for EMA is shown in Equation (5):

$$P^l = f\left(X^l, \theta\right) \tag{1}$$

$$P^u = f\left(X^u, \theta\right) \tag{2}$$

$$P^s = f\left(X^s, \theta'\right) \tag{3}$$

$$P^{s'} = IPE\left(P^l, P^u, Y^l\right) \tag{4}$$

$$\theta_t' = \alpha\theta_{t-1}' + (1-\alpha)\,\theta_t \tag{5}$$

We set the smoothing coefficient α to 0.99. Here, θ_t' represents the parameters of $f(\theta')$ at the t-th iteration, while θ_t represents the parameters of $f(\theta)$ at the t-th iteration.

In the second step, as shown in Figure 1b, both labeled data X^l and unlabeled data X^u are input into the $f(\theta')$. Unlabeled data X^u is input into the $f(\theta)$. The supervised loss L_s^B between label Y^l and prediction P^l is computed. Pseudo-labels Pl^A and Pl^B are generated for the unlabeled data through the student network and the teacher network. Subsequently, the teacher network and the student network are supervised with the pseudo-labels Pl^A and Pl^B to compute the cross pseudo supervision loss of L_{us}^A and L_{us}^B. Finally, the parameters of the student network are updated via EMA as shown in Equation (6).

ALGORITHM 1 IPE algorithm

Input: Two images that need to be exchanged, A^l, A^u, and label Y^l

Output: Exchanged image A^s

Top-left corner point $(x_1, y_1) \leftarrow (0, 0)$
Bottom-right corner point $(x_2, y_2) \leftarrow (0, 0)$
Initialize an image: $Mask \leftarrow 0$
Iterate pixel in Y^l by line from top-left to bottom-right:
    $(x_1, y_1) \leftarrow$ The position of the first non-zero pixel
Iterate pixel in Y^l by line from bottom-right to top-left
    $(x_2, y_2) \leftarrow$ The position of the first non-zero pixel
The midpoint $(x_m, y_m) \leftarrow ((x_1 + x_2)/2, (y_1 + y_2)/2)$
The radius $r \leftarrow \sqrt{(x_1 - x_2)^2 + (y_1 - y_2)^2}$
for x_i in Y^l.shape[0], y_j in Y^l.shape[1] **do**
    if $(x_i - x_m)^2 + (y_j - y_m)^2 \leq r^2$ **then**
        (x_i, y_j).value in $Mask \leftarrow 1$
    end if
end do
$A^s = A^l * Mask + A^u * (1 - Mask)$
Return A^s

The process of generating pseudo-labels is shown in Equations (7) and (8):

$$\theta_t = \alpha\theta_{t-1} + (1 - \alpha)\,\theta_t' \tag{6}$$

$$PI^A = argmax\,(softmax\,(f\,(X^u, \theta))) \tag{7}$$

$$PI^B = argmax\left(softmax\left(f\left(X^u, \theta'\right)\right)\right) \tag{8}$$

Through the MLS, the teacher network and the student network can be trained using labeled data, acquiring self-correcting abilities, thereby addressing the issue of error accumulation in the teacher model. This mutual learning process helps improve the model's generalization ability, facilitating further optimization of the model.

2.2 | Image partial exchange algorithm

To avoid the potential disruption of contextual information and reduce the introduction of erroneous information, this paper proposes the IPE algorithm. The IPE algorithm is designed to operate not only between the labeled and unlabeled data but also between the prediction results generated by the student model. The entire process of the IPE algorithm is shown in Algorithm 1 and Figure 2. For brevity, we consider two images, A^l and A^u, to be exchanged, both of which are randomly selected, along with the label Y^l corresponding to A^l. The IPE needs to determine the position of its exchanged part within the entire label Y^l. The specific method is to find the first nonzero pixel in the label starting from the top-left and bottom-right. To ensure that the exchanging step can proceed normally, the size of the partial image to be exchanged must remain consistent. Therefore, after obtaining the positional coordinates, the midpoint coordinates and the distance between the two points are calculated. This results in a circular region on a *Mask* where all pixel values are initially set to 0, with the midpoint as the center and the distance between the two points as the diameter. The pixel values within this circular region are then set to 1. Finally, the IPE is completed by obtaining A^s using the Equation (9):

$$A^s = A^l * Mask + A^u * (1 - Mask) \tag{9}$$

where $*$ denotes the element-wise multiplication.

Data augmentation with context has been proven helpful in semantic segmentation.[8] Since the shape of tissues or lesions is rarely rectangular and is instead closer to circular, IPE uses a circular mask. Unlike,[8] the circular mask in IPE is used for image partial exchange rather than context-aware augmentation, and it preserves more contextual information by smoothly integrating the exchanged regions, thereby reducing the introduction of erroneous information. This approach preserves the original structure of the input image during cropping to better align with the shape of the segmentation target boundaries, thereby retaining more contextual information. Using the IPE algorithm can minimize the introduction of erroneous information while ensuring the effectiveness of the consistency loss.

2.3 | Cross pseudo supervision module

To enhance knowledge interaction, MLNet incorporates a CPSM that complements the knowledge not captured by consistency learning. This module generates diverse pseudo-labels using different models and employs cross pseudo supervision loss for mutual supervision. This approach allows the models to learn richer features and better capture the complexity and underlying structure of the data, while also facilitating information sharing. By combining consistency learning with the CPSM, MLNet can fully leverage the diversity and redundancy of the data, thereby enhancing its generalization capabilities. Consistency learning ensures that the model produces similar outputs for the same sample under different perturbations. Meanwhile, the CPSM guides the model to capture latent patterns and details from multiple dimensions through model collaboration, thereby improving overall learning performance.

2.4 | Loss function

In the proposed semi-supervised medical image segmentation network, there are three components of loss:

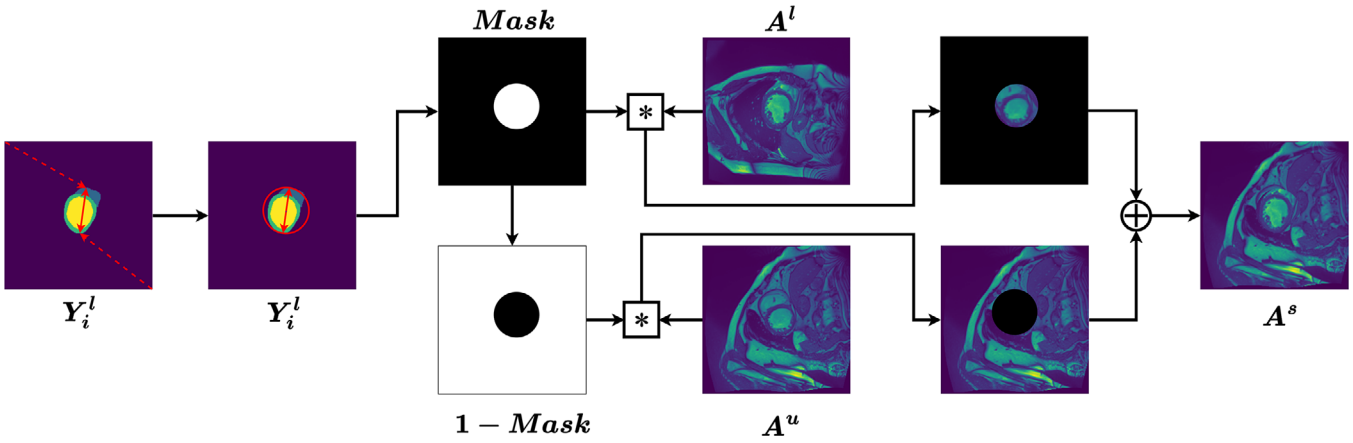

FIGURE 2 Flowchart of the IPE algorithm. First, identify the top-left and bottom-right points of the target segmentation area on Y^l. Generate a *Mask* based on these two points, then element-wise multiply A^l and A^u by the *Mask* and its inverse, respectively. Finally, add the two parts to obtain the augmented image A^s. IPE, image partial exchange algorithm.

supervised loss, consistency loss, and cross pseudo supervision loss. The supervised loss is represented in Equations (10) and (11):

$$L_s^A = \frac{1}{2N} \sum_{X^l, Y^l} L_{ce}\left(f\left(X^l, \theta\right), Y^l\right) + L_{dice}\left(f\left(X^l, \theta\right), Y^l\right) \tag{10}$$

$$L_s^B = \frac{1}{2N} \sum_{X^l, Y^l} L_{ce}\left(f\left(X^l, \theta'\right), Y^l\right) + L_{dice}\left(f\left(X^l, \theta'\right), Y^l\right) \tag{11}$$

N represents the number of labeled data, L_s^A is the supervised loss of $f(\theta)$, L_s^B is the supervised loss of $f(\theta')$, L_{ce} is the cross-entropy loss, and L_{dice} is the dice loss. The definitions of L_{ce} and L_{dice} are shown in Equations (12) and (13) respectively:

$$L_{ce}(P, Y) = -\frac{1}{N} \sum_{i=1}^{N} \sum_{k=1}^{K} Y_i^k \log\left(P_i^k\right) \tag{12}$$

$$L_{Dice}(P, Y) = 1 - \frac{2\sum_{i=1}^{N}(P_i \times Y_i)}{\sum_{i=1}^{N} P_i + \sum_{i=1}^{N} Y_i} \tag{13}$$

K represents the total number of categories, N is the total number of pixels in the image, and P denotes the predicted values while Y denotes the labels.

The consistency loss ensures that the predictions of the student model on unlabeled data remain consistent with the predictions of the teacher model. The consistency loss in this paper is shown in Equation (14):

$$L_{con} = d_{MSE}\left(P^s, P^{s'}\right) \tag{14}$$

d_{MSE} represents the mean squared error loss function to measure the average squared difference between the predictions of two models on unlabeled data, as shown in Equation (15):

$$d_{MSE}(y_i, \hat{y}_i) = \frac{1}{N} \sum_{i=1}^{N} (y_i - \hat{y}_i)^2 \tag{15}$$

N represents the number of samples, $\hat{y}_i$ represents the predicted value of the model for the i-th sample, and y_i represents the ground truth of the i-th sample.

Finally, there is the cross pseudo supervision loss, which supervises the predictions of one network using pseudo-labels generated by another network, as shown in Equations (16) and (17):

$$L_{us}^A = \frac{1}{M} \sum_{X^u} L_{ce}\left(f\left(X^u, \theta'\right), PI^A\right) \tag{16}$$

$$L_{us}^B = \frac{1}{M} \sum_{X^u} L_{ce}\left(f\left(X^u, \theta\right), PI^B\right) \tag{17}$$

M represents the number of unlabeled data.

Therefore, the final loss function is a linear combination of the aforementioned loss functions, as shown in Equations (18) and (19):

$$L_{overall}^A = L_s^A + \lambda L_{con} \tag{18}$$

$$L_{overall}^B = L_s^B + \lambda \times \left(L_{us}^A + L_{us}^B\right) \tag{19}$$

$L_{overall}^A$ represents the total loss function in the first step of the MLS, and $L_{overall}^B$ represents the total loss function in the second step of the MLS. λ is a weight factor, usually defined by a time-related Gaussian preheating

function: $\lambda\ (t) = 0.1 \times e^{(-5\times(1-\frac{t_i}{t_{total}})^2)}$. t_i represents the current iteration training times, and t_{total} represents the total iteration training times.[9]

3 | EXPERIMENTS

3.1 | Experimental details

This paper implements the proposed MLNet using the PyTorch framework.[10] The training and testing of the network model are accelerated using the NVIDIA 2080 Ti GPU. The poly learning rate strategy is adopted in this paper, where the initial learning rate is set to 1e-2. On the ACDC dataset, we trained the model for 30 000 epochs on the training set with a batch size of 24, where 12 are labeled data and 12 are unlabeled data. On the BraTS2019 dataset, we trained the model for 30 000 epochs on a training set with a batch size of 4, consisting of 2 labeled, and 2 unlabeled data. For a fair comparison, we use the same training settings as MLNet to train other currently classical and advanced semi-supervised methods for comparison.

3.2 | Datasets

Automated Cardiac Diagnosis Challenge (ACDC) public dataset.[11] The ACDC dataset is a multi-object segmentation dataset consisting of four classes (background, left ventricle, right ventricle, and myocardium), with scan data from 100 patients segmented into 70, 10, and 20 cases for training, validation, and testing, respectively. All 3D scans are converted into 2D slices, including 1312 slices in the training set, 200 slices in the validation set, and 400 slices in the test set, respectively. The semi-supervised experiments in this paper explored the model's performance with 10% labeled and 5% labeled data scenarios on the ACDC dataset. In the 10% labeled experiment, 7 cases out of the 70 training sets are considered labeled data, while the remaining 63 cases are treated as unlabeled data without using their labels. The situation for the 5% labeled experiment is the same.

Multimodal Brain Tumor Segmentation Challenge 2019 (BraTS2019).[12] The BraTS2019 dataset is a widely used medical imaging dataset for the task of brain tumor segmentation. It consists of MRI data collected from multiple institutions and various scanning devices, primarily intended for evaluating and comparing the performance of different algorithms in the automatic segmentation of brain tumors. The BraTS2019 dataset comprises a total of 335 labeled data. We have divided the BraTS2019 dataset into 250 cases for the training set, 25 cases for the validation set, and 60 cases for the test set, including 37 500 slices in the training set, 3750 slices in the validation set, and 9000 slices in the test set.

3.3 | Evaluation metrics

3.3.1 | Dice coefficient

The Dice coefficient is a commonly used metric for evaluating the similarity between predicted results and ground truth in image segmentation tasks, which calculates the ratio of the intersection area between the predicted results and the ground truth to their total area, as shown in Equation (20):

$$Dice = \frac{2 \times TP}{2 \times TP + FP + FN} \tag{20}$$

3.3.2 | 95% Hausdorff distance (HD95)

The HD95 is commonly used to measure the similarity between predicted segmentation results and ground truth annotations. It assesses the accuracy of segmentation algorithms and the matching degree of boundaries. The formula for HD95 is shown in Equation (21):

$$\mathrm{HD95}\,(A, B) = \max\{\mathrm{h}\,(\mathrm{A}, \mathrm{B})\,, \mathrm{h}\,(\mathrm{B}, \mathrm{A})\} \tag{21}$$

where $h\,(A, B) = \max_{a\in A}\{\min_{b\in B}\|a - b\|\}$ and $h\,(B, A) = \max_{b\in B}\{\min_{a\in A}\|b - a\|\}$.

3.3.3 | Average surface distance (ASD)

The ASD is commonly used to measure the proximity between two surfaces to assess the disparity or deviation between segmentation results and ground truth labels. ASD quantifies the average distance from each point on the segmented surface to the nearest point on the true structure surface, as shown in Equation (22):

$$ASD\,(A, B) = \frac{1}{A + B}\left(\sum_{a\in A}\min_{b\in B}\|a - b\| + \sum_{b\in B}\min_{a\in A}\|b - a\|\right) \tag{22}$$

3.4 | Comparison with state-of-the-art methods

Many recently proposed semi-supervised medical image segmentation methods can mainly be classified into two categories: consistency regularization methods and pseudo-labeling methods. Additionally, there have been some other methods emerging, such as adversarial network methods and contrastive learning

TABLE 1 Segmentation results of various methods under the scenario of 10% labeled data in the ACDC dataset.

Method	Scan used		Metrics		
	Labeled	Unlabeled	Dice↑	HD95↓	ASD↓
MT (NIPS'2017)[2]	7 (10%)	63 (90%)	80.20	10.65	3.08
CCT (CVPR'2020)[3]			82.85	7.66	2.10
CPS (CVPR'2021)[6]			85.28	6.02	1.88
MCF (CVPR'2018)[13]			85.78	5.45	1.80
UA-MT (MICCAI'2019)[14]			81.26	7.55	2.93
ICT (Neural Networks)[15]			83.13	9.98	2.79
URPC (MICCAI'2021)[16]			81.87	4.10	1.09
DCT (ECCV'2018)[17]			80.88	10.46	3.83
R-Drop (NIPS'2021)[18]			83.20	6.29	1.94
SCTNET (PMLR'2022)[19]			86.62	8.90	2.31
BCP (CVPR'2023)[20]			88.69	3.86	1.11
ADVENT (CVPR'2019)[21]			82.13	11.56	3.36
DAN (MICCAI'2017)[22]			79.45	13.78	4.12
MLNet (Ours)			**89.48**	**2.86**	**0.80**

Note: Bold indicates the best result.
Abbreviations: ACDC, automated cardiac diagnosis challenge; MLNet, network based on mutual learning; MT, mean-teacher.

methods. The proposed MLNet is compared with mainstream methods. The consistency regularization methods include MT,[2] cross-consistency targets (CCT),[3] CPS,[6] mutual correction framework (MCF),[13] uncertainty aware mean teacher (UA-MT),[14] interpolation consistency training (ICT),[15] uncertainty rectified pyramid consistency (URPC),[16] deep co-training (DCT),[17] and regularized dropout (R-Drop).[18] The pseudo-labeling methods include semi-supervised cross teaching network (SCTNET)[19] and bidirectional copy-paste (BCP),[20] and the adversarial methods include adversarial entropy (ADVENT)[21] and deep adversarial network (DAN).[22] CCT establishes consistency not only between labeled and unlabeled data but also between feature maps at different resolutions. MCF introduces two distinct subnetworks to explore and leverage the differences between them, thereby correcting the model's cognitive bias. UA-MT presents a model that integrates uncertainty estimation with self-ensembling techniques to enhance the accuracy of 3D left atrium segmentation in a semi-supervised learning context. ICT interpolates unlabeled data to generate pseudo-labels and then requires the model to ensure consistency between the predictions of these pseudo-labels and the original labeled data. URPC introduces a novel uncertainty rectification module, enabling the framework to progressively learn from meaningful and reliable consensus regions across different scales. CPS utilizes pseudo-supervision signals and cross-sample consistency to enhance the performance of semi-supervised semantic segmentation. BCP proposes a bidirectional copy-paste method that alleviates the issue of empirical distribution mismatch between labeled and unlabeled data by performing copy-paste operations between them. All experiments are conducted under the same conditions. Table 1 presents the performance of different methods on the ACDC dataset under the scenario of 10% labeled data, evaluated using three metrics.

The MLS effectively mitigates the issue of the teacher model continuously accumulating erroneous knowledge. Furthermore, the proposed IPE algorithm not only reduces the introduction of erroneous information but also addresses the issue of perturbations disrupting contextual information in the image. CPSM helps networks to learn better and understand class information, enabling mutual supervision between networks. Finally, the obtained Dice coefficient result is 89.48%, surpassing all the aforementioned methods.

Table 2 presents the results obtained under the same experimental conditions on the ACDC dataset with 5% labeled data. It can be observed that compared to the experiments with 10% labeled data, some methods show a significant decrease in performance metrics when the number of labeled data decreases. In contrast, the proposed method still performs excellently even with half the amount of labeled data, outperforming other comparative methods.

Figure 3 illustrates the visual comparisons of various methods on the ACDC dataset using 10% labeled data and 5% labeled data. It can be seen that for the four MRI images, other methods begin to exhibit noticeable segmentation errors. Therefore, compared with other methods, the segmentation results of the proposed method are more reliable, with fewer missing segmentation areas.

To validate the effectiveness of MLNet, we compare it with several classical methods for brain tumor

TABLE 2 Segmentation results of various methods under the scenario of 5% labeled data in the ACDC dataset.

Method	Scan used		Metrics		
	Labeled	Unlabeled	Dice↑	HD95↓	ASD↓
MT (NIPS'2017)[2]	3 (5%)	67 (95%)	49.73	47.12	18.75
CCT (CVPR'2020)[3]			56.21	16.58	7.20
CPS (CVPR'2021)[6]			66.93	8.54	1.82
MCF (CVPR'2018)[13]			75.64	5.09	1.41
UA-MT (MICCAI'2019)[14]			58.06	15.10	6.13
ICT (Neural Networks)[15]			60.42	12.05	3.71
URPC (MICCAI'2021)[16]			54.42	17.80	8.23
DCT (ECCV'2018)[17]			54.96	33.65	13.51
R-Drop (NIPS'2021)[18]			61.98	16.12	5.26
SCTNET (PMLR'2022)[19]			75.60	5.03	1.29
BCP (CVPR'2023)[20]			87.74	2.94	1.03
ADVENT (CVPR'2019)[21]			58.46	20.61	6.60
DAN (MICCAI'2017)[22]			56.42	31.04	10.17
MLNet (Ours)			**88.15**	**2.05**	**0.93**

Note: Bold indicates the best result.
Abbreviations: ACDC, automated cardiac diagnosis challenge; MLNet, network based on mutual learning; MT, mean-teacher.

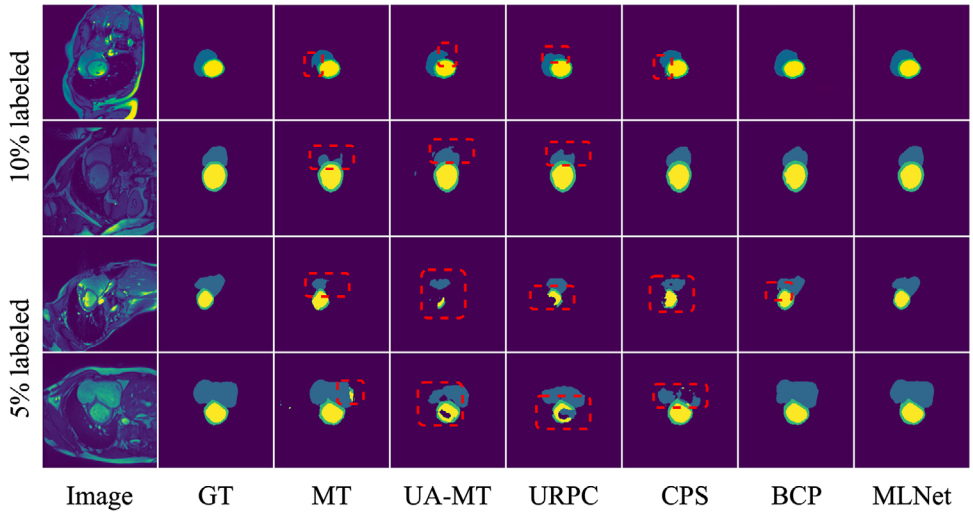

FIGURE 3 Multi-object segmentation visualization comparisons of the ACDC dataset with 10% labeled data and 5% labeled data, where the blue, green, and yellow areas represent the segmentation results of the right ventricle, myocardium, and left ventricle, respectively. ACDC, automated cardiac diagnosis challenge.

segmentation on the BraTS2019 dataset, as shown in Tables 3 and 4. The semi-supervised methods are trained using 10% labeled data (25 labeled cases and 225 unlabeled cases) and 5% labeled data (12 labeled cases and 238 unlabeled cases). The experimental results demonstrate that MLNet achieves superior performance in the brain tumor segmentation task compared to other methods. Compared to MT, MLNet achieves a significant improvement in the Dice evaluation metric, increasing by 2.63% points with 10% labeled data and 5.41% points with 5% labeled data. The visual brain tumor segmentation results are shown in Figure 4.

4 | DISCUSSION

4.1 | Ablation studies

To verify the effectiveness of CPSM, the IPE algorithm (IPE), and the MLS, ablation studies are conducted. Experiments are conducted on the ACDC dataset using 10% labeled data and 90% unlabeled data, with the MT method as the baseline. The experimental results are detailed in Table 5. It can be seen that by incorporating CPSM into the baseline, the Dice score increased by 2.16% points, as it avoids prediction errors that a

TABLE 3 Segmentation results of various methods under the scenario of 10% labeled data in the BraTS2019 dataset.

Method	Scan used		Metrics		
	Labeled	Unlabeled	Dice↑	HD95↓	ASD↓
MT (NIPS'2017)[2]	25 (10%)	225 (90%)	81.93	10.04	2.82
CPS (CVPR'2021)[6]			82.57	10.81	2.97
MCF (CVPR'2018)[13]			83.33	11.72	2.99
UA-MT (MICCAI'2019)[14]			82.55	10.43	2.88
ICT (Neural Networks)[15]			81.77	15.51	4.29
URPC (MICCAI'2021)[16]			82.59	13.88	3.72
R-Drop (NIPS'2021)[18]			80.80	10.07	2.23
BCP (CVPR'2023)[20]			83.16	9.23	3.09
ADVENT (CVPR'2019)[21]			82.29	10.73	2.79
DAN (MICCAI'2017)[22]			81.68	11.01	2.80
MLNet (Ours)			**84.56**	**8.62**	**2.17**

Note: Bold indicates the best result.
Abbreviations: MLNet, network based on mutual learning; MT, mean-teacher.

TABLE 4 Segmentation results of various methods under the scenario of 5% labeled data in the BraTS2019 dataset.

Method	Scan used		Metrics		
	Labeled	Unlabeled	Dice↑	HD95↓	ASD↓
MT (NIPS'2017)[2]	12 (5%)	238 (95%)	78.06	14.65	3.78
CPS (CVPR'2021)[6]			78.80	14.97	4.16
MCF (CVPR'2018)[13]			81.61	10.53	3.98
UA-MT (MICCAI'2019)[14]			79.31	11.65	3.12
ICT (Neural Networks)[15]			78.66	22.84	7.43
URPC (MICCAI'2021)[16]			78.74	16.42	4.73
R-Drop (NIPS'2021)[18]			77.93	13.88	2.89
BCP (CVPR'2023)[20]			81.36	11.28	2.95
ADVENT (CVPR'2019)[21]			78.02	15.67	3.93
DAN (MICCAI'2017)[22]			76.97	15.28	4.08
MLNet (Ours)			**82.47**	**10.63**	**2.72**

Note: Bold indicates the best result.
Abbreviations: MLNet, network based on mutual learning; MT, mean-teacher.

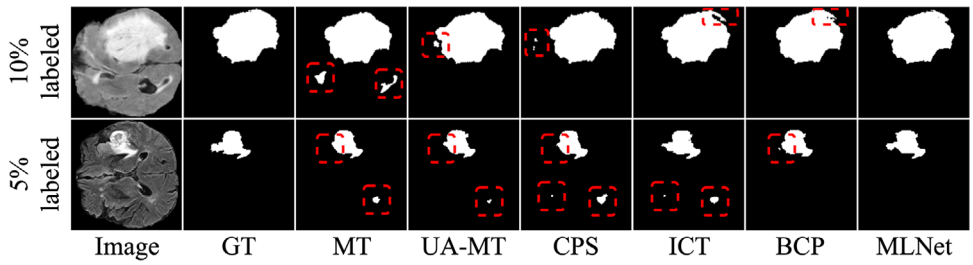

FIGURE 4 Visualization results of the BraTS2019 dataset with 10% labeled data and 5% labeled data.

single model might cause. When the baseline is combined with the IPE algorithm, it resolves the issues of contextual information disruption and the introduction of erroneous information, increasing the Dice score by 3.38% points compared to the baseline. Combining the baseline with the MLS can significantly improve the semi-supervised segmentation results, increasing the Dice score by 4.13% points. This indicates that this

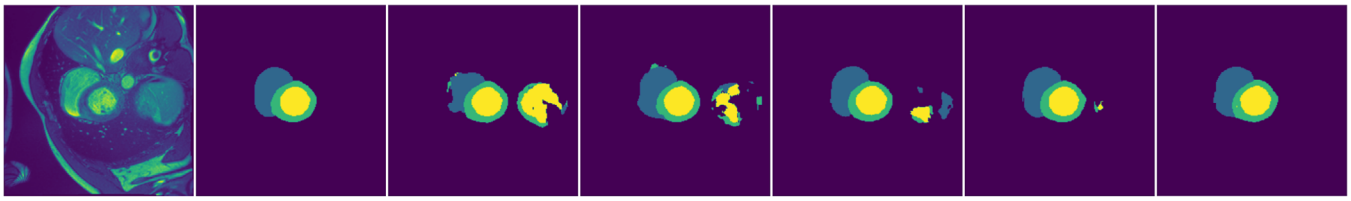

FIGURE 5 Visual results of the ablation experiments on the ACDC dataset with 10% labeled data. ACDC, automated cardiac diagnosis challenge.

TABLE 5 Module ablation experiments on ACDC dataset with 10% labeled data.

Module				Metrics		
Baseline	CPSM	IPE	MLS	Dice↑	HD95↓	ASD↓
√				80.20	10.65	3.08
√	√			82.36	8.03	2.63
√		√		83.58	7.03	2.63
√			√	84.33	6.10	2.06
√	√	√	√	**89.48**	**2.86**	**0.80**

Note: Bold indicates the best result.
Abbreviations: ACDC, automated cardiac diagnosis challenge; CPSM, cross pseudo supervision module; IPE, image partial exchange algorithm; MLS, mutual learning strategy.

strategy can effectively alleviate the issues of accumulated erroneous information in MT and enhance model performance. Overall, the integration of CPSM, the IPE algorithm, and the MLS into the baseline significantly improves the semi-supervised segmentation accuracy, demonstrating their effectiveness in enhancing model performance.

To better visualize the advantages of each module, Figure 5 shows the visualization of the output predictions from our model. With the CPSM, prediction errors caused by a single model are avoided, resulting in better coverage of the overall outline. However, the segmented areas are still relatively fragmented, with noticeable over-segmentation. By applying the IPE algorithm, more contextual information during training can be preserved, therefore effectively reducing over-segmentation and providing more stable boundary handling. By employing the MLS, more complete target regions and fewer over-segmented areas can be achieved due to knowledge sharing and continuous error correction between networks.

4.2 | Error accumulation analysis

To further demonstrate that the MLS can help alleviate error accumulation, we conduct experiments using the proposed MLNet and MLNet without MLS. During the experiments, we output visualizations of the model every 6000 epochs and record the changes in segmentation error regions, as shown in Figure 6. It can be observed that as the training progresses, MLNet (w/o MLS) fails to correct the accumulated errors, leading to increasingly severe inaccuracies. In contrast, MLNet continuously corrects the erroneous regions, getting closer to the Ground Truth. This indicates that MLNet can mitigate the problem of error accumulation.

4.3 | Effectiveness of the IPE algorithm compared to other data augmentation methods

To verify the effectiveness of the IPE method, it was compared with several other data augmentation methods. CutMix generates a new image where one region comes from one image and the remaining regions come from another image, with the proportions of the two images in the new image controlled by a mask, and the new labels are linear combinations of the proportions of each image in the new image.[23] The mask for CutMix is randomly generated rectangles. FMix blends two images by randomly selecting another image for each training sample and mixing their pixel values in proportion, with the mixing process controlled by a random mask.[24] The random mask is created by generating a random low-frequency image in Fourier space and applying a threshold to produce the mask, and it is not related to the selected samples. Fourier transform-based data augmentation enhances image samples by exchanging the central regions of the amplitude spectra of two images in Fourier space, where the central region is controlled by a mask.[25] The mask for Fourier contains the central region of the amplitude spectrum that contains low-frequency information. The experimental results are shown in Table 6. The experimental results indicate that: (1) Methods using data augmentation as perturbation are more effective than those traditionally adding Gaussian noise. (2) The CutMix method, which swaps using rectangular boxes, disrupts the contextual information at the boundaries, which is disadvantageous for cardiac segmentation on the ACDC dataset and lacks flexibility due to its singular swap position. (3) The FMix method, using randomly generated mask images for swapping, exhibits high flexibility but poor stability due to

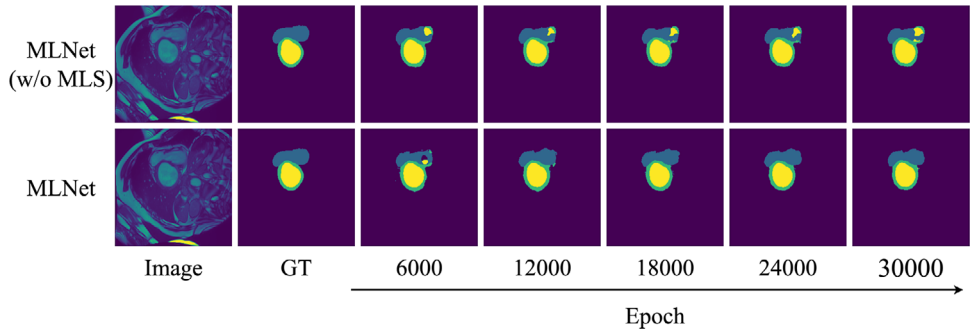

FIGURE 6 The changes during the training process of MLNet and MLNet (w/o MLS). The numbers represent epochs count. It can be observed that MLNet (w/o MLS) is unable to correct errors during the training process, while MLNet continuously corrects the learned erroneous knowledge through mutual learning. MLNet, network based on mutual learning; MLS, mutual learning strategy.

TABLE 6 Comparison of data augmentation methods in the ACDC dataset with 10% labeled data.

Method	Scan used		Metrics		
	Labeled	Unlabeled	Dice↑	HD95↓	ASD↓
MT[2]	7 (10%)	63 (90%)	80.20	10.65	3.08
CutMix[23]			81.56	10.20	3.02
FMix[24]			81.92	9.53	2.87
Fourier[25]			80.34	10.91	3.17
IPE			**83.58**	**7.03**	**2.63**

Note: Bold indicates the best result.
Abbreviations: ACDC, automated cardiac diagnosis challenge; IPE, image partial exchange algorithm.

excessive randomness. (4) The Fourier transform-based data augmentation method, employing amplitude spectrum swapping, may introduce erroneous signals.

Attention heatmap is a visualization technique used to display the level of attention that a model gives to different regions or features when processing input data.[26] To more clearly demonstrate the effect of adding perturbations, we process the two images to be exchanged using attention heatmaps, to help us clearly see the important parts of the images and the preservation of important content after exchanging with different data augmentation methods. The attention heatmaps of different data augmentation methods are shown in Figure 7. The results in Figure 7 are obtained from Equation (23):

$$Result = Image\ A * Mask + Image\ B * (1 - Mask) \tag{23}$$

It can be seen that the perturbation generated by the IPE algorithm is minimal, with its result being closest to image A. In summary, the IPE algorithm proposed in this paper performs the best.

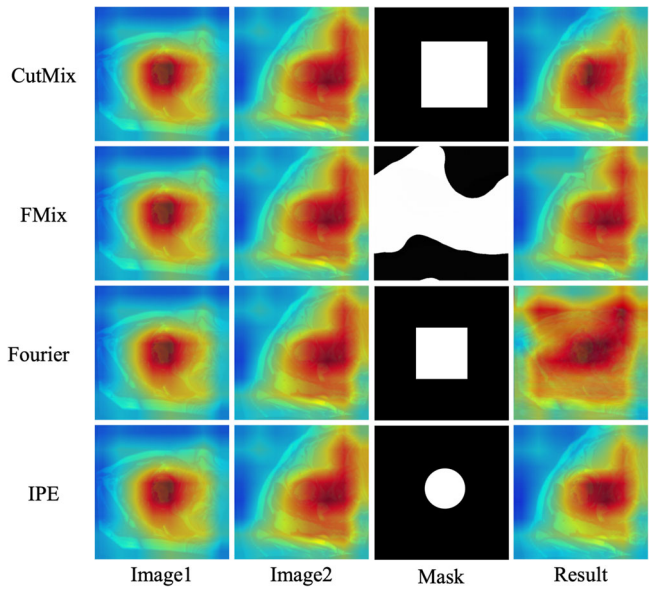

FIGURE 7 Comparing the attention heatmaps of different data augmentation methods in the ACDC dataset. The first and second columns are the attention heatmap representations of the two images that need enhancement, the third column represents the masks used by different methods to control the exchange of regions, and the last column shows the attention heatmap representations of resulting augmented images. The mask for CutMix is randomly generated rectangles, while the mask for Fmix is obtained by thresholding the low-frequency image sampled in the Fourier space. The mask for Fourier contains the central region of the amplitude spectrum that contains low-frequency information. ACDC, automated cardiac diagnosis challenge.

5 | CONCLUSION

In this paper, a semi-supervised medical image segmentation MLNet is proposed to alleviate the issues of error accumulation in the MT method. Employing the IPE algorithm, not only addresses the issue of disruptions to contextual information in images but also mitigates the introduction of erroneous information. Additionally, through the MLS between the student model and the teacher model, the problem of error accumulation in the teacher model is effectively addressed, enabling the teacher model to learn during the teaching process, thus achieving mutual learning. The results on the ACDC and BraTS2019 datasets demonstrate the superiority

of the proposed MLNet. With only 10% labeled data, our Dice coefficients reach 89.48% and 84.56%, respectively, outperforming state-of-the-art semi-supervised segmentation models. In future work, further exploration of leveraging unlabeled data knowledge and generating high-quality pseudo-labels will be our main research focus.

ACKNOWLEDGMENTS

This work was supported in part by grants from the Science and Technology Plan Project of Hangzhou City, China (No. 20201203B124).

CONFLICT OF INTEREST STATEMENT

The authors declare no conflicts of interest.

DATA AVAILABILITY STATEMENT

The data that support this study are openly available in the ACDC dataset at https://www.creatis.insa-lyon.fr/Challenge/acdc/ and the BraTS2019 dataset at https://www.med.upenn.edu/cbica/brats-2019/.

REFERENCES

1. Shen W, Xu W, Sun Z, et al. Automatic segmentation of the femur and tibia bones from X-ray images based on pure dilated residual U-Net. *Inverse Probl Imaging*. 2020;15(6):1333-1346
2. Tarvainen A, Valpola H. Mean teachers are better role models: weight-averaged consistency targets improve semi-supervised deep learning results. *Adv Neural Inf Process, NIPS*. 2017;30:1195-1204.
3. Ouali Y, Hudelot C, Tami M. Semi-supervised semantic segmentation with cross-consistency training. In: *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, CVPR*. IEEE; 2020:12674-12684.
4. Luo X, Chen J, Song T, et al. Semi-supervised medical image segmentation through dual-task consistency. In: *Proceedings of the Association for the Advancement of Artificial Intelligence*, *AAAI*. AAAI; 2021;35:8801-8809.
5. Li S, Zhang C, He X. Shape-aware semi-supervised 3D semantic segmentation for sedical images. In: *Conference on Medical Image Computing and Computer-Assisted Intervention, MICCAI*. Springer; 2020:552-561.
6. Chen X, Yuan Y, Zeng G, Wang J. Semi-supervised semantic segmentation with cross pseudo supervision. In: *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition CVPR*. IEEE; 2021:2613-2622.
7. Zhang Y, Xiang T, Hospedales TM, Lu H. Deep mutual learning. In: *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, CVPR*. IEEE; 2018:4320-4328.
8. He Q, Duan Y, Yang Z, et al. Context-aware augmentation for liver lesion segmentation: shape uniformity, expansion limit and fusion strategy. *Quant Imaging Med Surg*. 2023;13(8):5043-5047
9. Laine S, Aila T. Temporal ensembling for semi-supervised learning. In: *Proceedings of the International Conference on Learning Representations, ICLR*. ICLR; 2017:1-13.
10. Paszke A, Gross S, Chintala S, et al. Automatic differentiation in PyTorch. *Adv Neural Inf Process Syst Workshop, NIPS-W*. 2017.
11. Bernard O, Lalande A, Zotti C, et al. Deep learning techniques for automatic MRI cardiac multi-structures segmentation and diagnosis: is the problem solved?. *IEEE Trans Med Imaging*. 2018;37(11):2514-2525
12. Menze BH, Jakab A, Bauer S, et al. The multimodal brain tumor image segmentation benchmark (BRATS). *IEEE Trans Med Imaging*. 2014;34(10):1993-2024.
13. Wang Y, Xiao B, Bi X, Li B, Gao, X. MCF: mutual correction framework for semi-supervised medical image segmentation. In: *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, CVPR*. IEEE; 2023:15651-15660.
14. 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: *Conference on Medical Image Computing and Computer-Assisted Intervention, MICCAI*. Springer; 2019:605-613.
15. Verma V, Kawaguchi K, Lamb A, et al. Interpolation consistency training for semi-supervised learning. *Neural Netw*. 2022;145:90-106
16. Luo X, Liao W, Chen J, et al. Efficient semi-supervised gross target volume of nasopharyngeal carcinoma segmentation via uncertainty rectified pyramid consistency. In: *Conference on Medical Image Computing and Computer-Assisted Intervention, MICCAI*. ACM;2021:318-329.
17. Qiao S, Shen W, Zhang Z, Wang B, Yuille A. Deep co-training for semi-supervised image recognition. In: *Proceedings of the European Conference on Computer Vision, ECCV*. Springer; 2018:135-152.
18. Wu L, Li J, Wang Y, et al. R-Drop: regularized dropout for neural networks. *Adv Neural Inf Process Syst, NIPS*. 2021;34:10890-10905.
19. Luo X, Hu M, Song T, et al. Semi-supervised medical image segmentation via cross teaching between CNN and transformer. In: *Proceedings of Machine Learning Research, PMLR*. 2022:820-833.
20. Bai Y, Chen D, Li Q, Shen W, Wang Y. Bidirectional copy-paste for semi-supervised medical image segmentation. In: *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, CVPR*. IEEE; 2023:11514-11524.
21. Vu TH, Jain H, Bucher M, Cord M, Pérez P. ADVENT: adversarial entropy minimization for domain adaptation in semantic segmentation. In: *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, CVPR*. IEEE; 2019:2517-2526.
22. Zhang Y, Yang L, Chen J, Fredericksen M, Hughes DP, Chen DZ. Deep adversarial networks for biomedical image segmentation utilizing unannotated images. In: *Conference on Medical Image Computing and Computer-Assisted Intervention, MICCAI*. IEEE; 2017:408-416.
23. Yun S, Han D, Oh SJ, Chun S, Yoo Y, Choe J. CutMix: regularization strategy to train strong classifiers with localizable features. In: *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, CVPR*. IEEE; 2019:6023-6032.
24. Harris E, Marcu A, Painter M, Niranjan M, Prugel-Bennett A, Hare JS. FMix: enhancing mixed sample data augmentation. arXiv preprint arXiv:2002.12047. 2020.
25. Yao H, Hu X, Li X. Enhancing pseudo label quality for semi-supervised domain-generalized medical image segmentation. *Proceedings of the Association for the Advancement of Artificial Intelligence, AAAI*. AAAI; 2022;36(3):3099-3107.
26. Selvaraju RR, Cogswell M, Das A, Vedantam R, Parikh D, Batra D. Grad-CAM: visual explanations from deep networks via gradient-based localization. In: *Proceedings of the IEEE International Conference on Computer Vision, ICCV*. IEEE; 2017:618-626.

How to cite this article: Sun J, Wang T, Wang M, Li X, Xu Y. Semi-supervised medical image segmentation network based on mutual learning. *Med Phys*. 2024;1-12. https://doi.org/10.1002/mp.17547