International Journal of Research in Engineering and Science (IJRES)
ISSN (Online): 2320-9364, ISSN (Print): 2320-9356
www.ijres.org Volume 12 Issue 12 ǁ December 2024 ǁ PP. 212-227

Semi-supervised Medical Image Segmentation with Hybrid Cropping in a Multi-Model Co-learning Framework

Ziyuan Zhang
*[*1]School of Computer Science and Technology, Tongji University, Shanghai, China*
Corresponding Author: Ziyuan Zhang

Abstract
Training deep convolutional neural networks typically requires a large amount of labeled data. However, in medical image segmentation tasks, annotating data is both expensive and time-consuming. In this paper, we propose a novel semi-supervised multi-model co-learning framework for segmenting the left atrium from 3D MR images. Our framework encourages models to generate segmentation results from multiple perspectives for the same input image, thereby leveraging the strengths of multiple models. The framework consists of two student models and one teacher model. The weights of the teacher model are obtained using a weighted Exponential Moving Average (EMA) of the two student models, and inter-model loss constraints are applied during training to prevent overfitting to input noise. To further enhance segmentation performance, we design a new medical image augmentation method inspired by CutMix. Specifically, we employ a hybrid augmentation approach that mutually crops and mixes regions from labeled and unlabeled images. This allows the model to better unify the data distribution of labeled and unlabeled samples, leading to the generation of higher-quality pseudo-labels. Experimental results demonstrate that our method achieves outstanding performance in the 3D left atrium segmentation task, highlighting its effectiveness.

Keywords: *Mutual learning, Semi-supervised medical image segmentation.*

Date of Submission: 05-12-2024 Date of acceptance: 17-12-2024

I. INTRODUCTION

Automatic segmentation of the left atrium (LA) in magnetic resonance (MR) images is of great significance for advancing the treatment of atrial fibrillation. Deep learning has greatly advanced LA segmentation using large amounts of labeled data[1]. However, in the field of medical imaging, obtaining reliable annotations from 3D medical images slice-by-slice by experienced experts is both expensive and time-consuming. Since unlabeled data is often abundant, we focus on semi-supervised methods that leverage a limited amount of labeled data and a large volume of unlabeled data for LA segmentation.

The combination of consistency and pseudo-labeling provides an innovative approach to joint learning between dual students and a teacher model. Consistency learning between the teacher and students enhances the network's accuracy in medical image segmentation. On the other hand, cross-learning between the two students using pseudo-supervision — where students leverage each other's pseudo-labels for supervised learning — encourages the models to better understand and capture specific structures and features in medical images. Throughout the learning process, multiple constraints are applied, effectively improving segmentation performance. This integrated learning strategy enables the model to better adapt to various medical imaging scenarios, achieving more precise and robust segmentation results, even for complex structures.

In recent years, semi-supervised learning methods based on mutual learning have emerged as a research hotspot in medical image segmentation. Due to the high cost and time-consuming nature of annotating medical imaging data, traditional supervised learning methods often face significant limitations. In contrast, mutual learning improves the utilization of unlabeled data through collaboration among multiple models, thereby enhancing segmentation performance. Mutual learning is an innovative training strategy designed to boost learning efficiency through interactions between different models[2]. In semi-supervised learning, mutual learning allows models to share information, enabling them to jointly learn from unlabeled data and overcome the challenges that a single model might encounter when dealing with complex tasks.

Mutual learning by enabling collaboration among multiple models, can overcome the difficulties that a single model might encounter when dealing with complex tasks, thus effectively improving the utilization of unlabeled data and enhancing segmentation performance. Currently, in the field of semi-supervised segmentation, mutual learning is often implemented by combining consistency regularization and pseudo-

labeling, where cross-supervision is applied between subnetworks. However, mutual learning faces several challenges due to the need to account for the collaborative interactions among multiple models. For example, if the models do not sufficiently diverge during training, they may prematurely converge, causing the models to fail to learn in different directions and degrade into a single-model training scenario. Furthermore, in semi-supervised learning, labeled data is usually much smaller than unlabeled data. The labeled data can be seen as a small sampling of the overall data distribution within the same medical image segmentation dataset, and it may not accurately represent the full data distribution. As a result, the model's learning performance can be affected by data distribution biases.

In this paper, we propose a multi-model mutual learning framework that captures image information from multiple perspectives. This framework ensures sufficient diversity among models, effectively mitigating the issue of self-degradation. Building on this, to address the distributional discrepancy between labeled and unlabeled data, we employ a hybrid augmentation method that mutually crops and mixes regions from labeled and unlabeled images. This approach enables the model to better align the data distributions of labeled and unlabeled samples, thereby generating higher-quality pseudo-labels. Experimental results demonstrate that our method achieves excellent performance on the 3D left atrium segmentation task, fully validating its effectiveness.

II. METHOD

2.1 Overview

We introduce uncertainty relationships to construct model diversity. As shown in Figure 1, the construction of uncertainty relationships involves both structural differences in the subnetworks and differences in the training data. Specifically, regarding structural differences, the mutual learning framework consists of a teacher model and two student models. The teacher model is a self-ensemble of the student models' averages, and each model employs a different backbone structure. The purpose of the teacher model is to regulate the learning process of the student networks, preventing them from fitting incorrect information. The teacher's parameters are updated using a weighted Exponential Moving Average (EMA), as:

$$\theta_t = (1-\alpha-\beta)\theta_{t-1} + \alpha\theta_{s1} + \beta\theta_{s2} \tag{1}$$

where α and β represent the weighted update coefficients for Student Subnetwork 1 and Student Subnetwork 2, respectively. θ_{s1} and θ_{s2} represent the parameters of Student Subnetwork 1 and Student Subnetwork 2, respectively.

For a semi-supervised mutual learning network, the total loss consists of the labeled loss and the unlabeled loss, as:

$$L = L_{labeled} + \lambda L_{unlabeled} \tag{2}$$

where λ is the loss weight parameter.

For labeled data, the supervised loss is represented as the sum of the semantic segmentation losses from the two student subnetworks. In our experiments, we use the Dice loss as the semantic segmentation loss for medical image segmentation. The labeled loss is formulated as shown in Equation (3), where N represents the total number of labeled data, Y_i denotes the Ground Truth.

$$L_{labeled} = \frac{1}{N}\sum_{i=1}^{N}\left\{L_{dice}\left(f\left(X_i;\theta_{s_1}\right),Y_i\right) + L_{dice}\left(f\left(X_i;\theta_{s_2}\right),Y_i\right)\right\} \tag{3}$$

For unlabeled data, we use two types of losses for computation. We borrow the CPS method used in MC-Net[3], where the pseudo-labels from the two subnetworks are cross-supervised to compute the loss. This approach promotes mutual learning between the two student subnetworks and strengthens their consistency. Let M denote the number of unlabeled data. The CPS loss for Student Subnetwork 1 and Student Subnetwork 2 is given by Equations (4) and (5), respectively:

$$L_{cps_1} = \frac{1}{M}\sum_{j=1}^{M} L_{dice}\left(f\left(X_j;\theta_{s_1}\right),\widehat{Y_j^{s_2}}\right) \tag{4}$$

$$L_{cps_2} = \frac{1}{M}\sum_{j=1}^{M} L_{dice}\left(f\left(X_j;\theta_{s_2}\right),\widehat{Y_j^{s_1}}\right) \tag{5}$$

From Equations (4) and (5), the total CPS loss is represented as follows:

$$L_{cps} = L_{cps_1} + L_{cps_2} \tag{6}$$

As previously mentioned, to prevent the two students from training in the wrong direction, we introduce a teacher model to guide the optimization of the student models. The parameters of the teacher model are obtained using a weighted EMA method, as shown in Equation (1). Therefore, the teacher model's parameters can be considered as a self-ensemble of the two student models' parameters. During training, the teacher model should have better representational power compared to the student models.

Thus, we can employ a teacher-student loss to constrain the student models and reduce the risk of overfitting to noise. The design of the teacher-student loss is given by Equations (7) to (9), where $\widehat{Y_j^t}$ represents the pseudo-label generated by the teacher model.

$$L_{teacher_1} = \frac{1}{M}\sum_{j=1}^{M} L_{dice}\left(f\left(X_j;\theta_{s_1}\right),\widehat{Y_j^t}\right) \tag{7}$$

$$L_{teacher_2} = \frac{1}{M}\sum_{j=1}^{M} L_{dice}\left(f\left(X_j;\theta_{s_2}\right),\widehat{Y_j^t}\right) \tag{8}$$

$$L_{teacher} = L_{teacher_1} + L_{teacher_2} \tag{9}$$

In summary, the unsupervised loss of the mutual learning model is expressed as follows:

$$L_{unlabeled} = L_{cps} + L_{teacher} \tag{10}$$

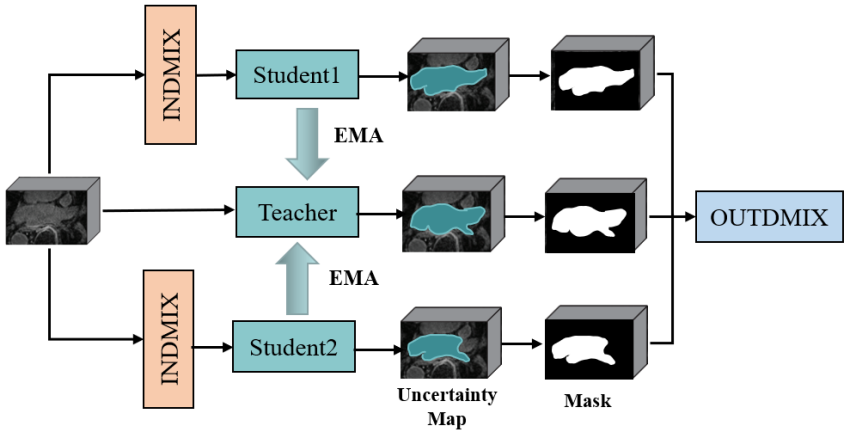

Figure 1: Overview of our mutual learning method

2.2 INDMIX module

The operation of our INDMIX module is as follows: We randomly select two unlabeled images and two labeled images from the training set. Then, we randomly crop a region, and the foreground from the labeled cropped part is stitched onto the background of the unlabeled image. Simultaneously, the foreground of the unlabeled region is stitched onto the background of the labeled image, thereby generating two mixed images. In Figure 2, the areas inside the blue and red boxes represent the cropped regions. These two images are then input into the Student network to generate predicted segmentation masks, which are subsequently passed to the OUTDMIX module for further processing.

To conduct copy-paste between a pair of images, we first generate a zero-centered mask:$Mask \in \{0,1\}^{W\times H\times L}$, where 0 indicating the voxel comes from the foreground and 1 from the background image. Then we can get the mixed data as follows step 1 and step 2 given by Equations (11) and (12), and $\odot$ means element-wise multiplication.:

$$Mixed_1 = X_{labeled}\odot Mask + X_{unlabeled}\odot(1 - Mask) \tag{11}$$

$$Mixed_2 = X_{unlabeled}\odot Mask + X_{labeled}\odot(1 - Mask) \tag{12}$$

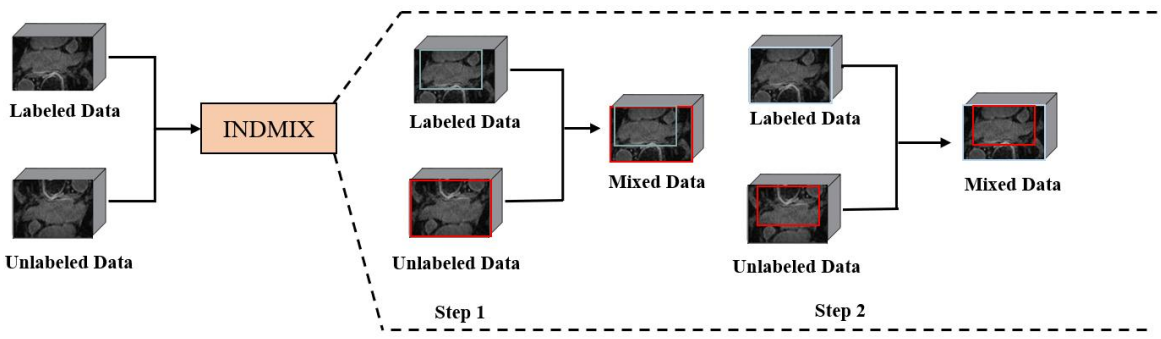

Figure 2: Overview of INDMIX

2.3 OUTDMIX module

The operation of the OUTDMIX module is as follows: First, we input the mixed images generated by the INDMIX module into the two Student networks, respectively, to generate two segmentation masks for the mixed images. Then, we input the cropped regions into the Teacher model to obtain the segmentation masks for the cropped regions. Since the Teacher model's parameters are obtained through weighted EMA, it can be considered as a self-ensemble of the two student models' parameters, and it possesses better discriminative power. We mix the segmentation results from the Teacher model with the Ground Truth for the cropped regions to create a "Ground Truth" for the mixed region. Finally, we compute the MSE loss between the Teacher-generated "Ground Truth" and the segmentation masks produced by the Student networks, which gives us the loss for the proposed hybrid algorithm.

The outputs of the Teacher network are as follows:

$$Y_{mix1} = \widehat{Y}_1 \odot Mask + Y_{labeled} \odot (1 - Mask) \tag{13}$$

$$Y_{mix2} = \widehat{Y}_2 \odot Mask + Y_{labeled} \odot (1 - Mask) \tag{14}$$

Then we can get the output of the Student networks:

$$Y_1 = f(Mixed_1; \theta_{s_1}) \tag{15}$$

$$Y_2 = f(Mixed_2; \theta_{s_2}) \tag{16}$$

So we can get the loss of our module use the mse loss:

$$L_{mix} = MSE(Y_{mix1}, Y_1) + MSE(Y_{mix2}, Y_2) \tag{17}$$

The total loss of our proposed multi model mutual learning framework is:

$$L_{total} = L_{labeled} + \lambda L_{unlabeled} + \mu L_{mix} \tag{18}$$

where λ and μ are the weight parameters of the loss function.

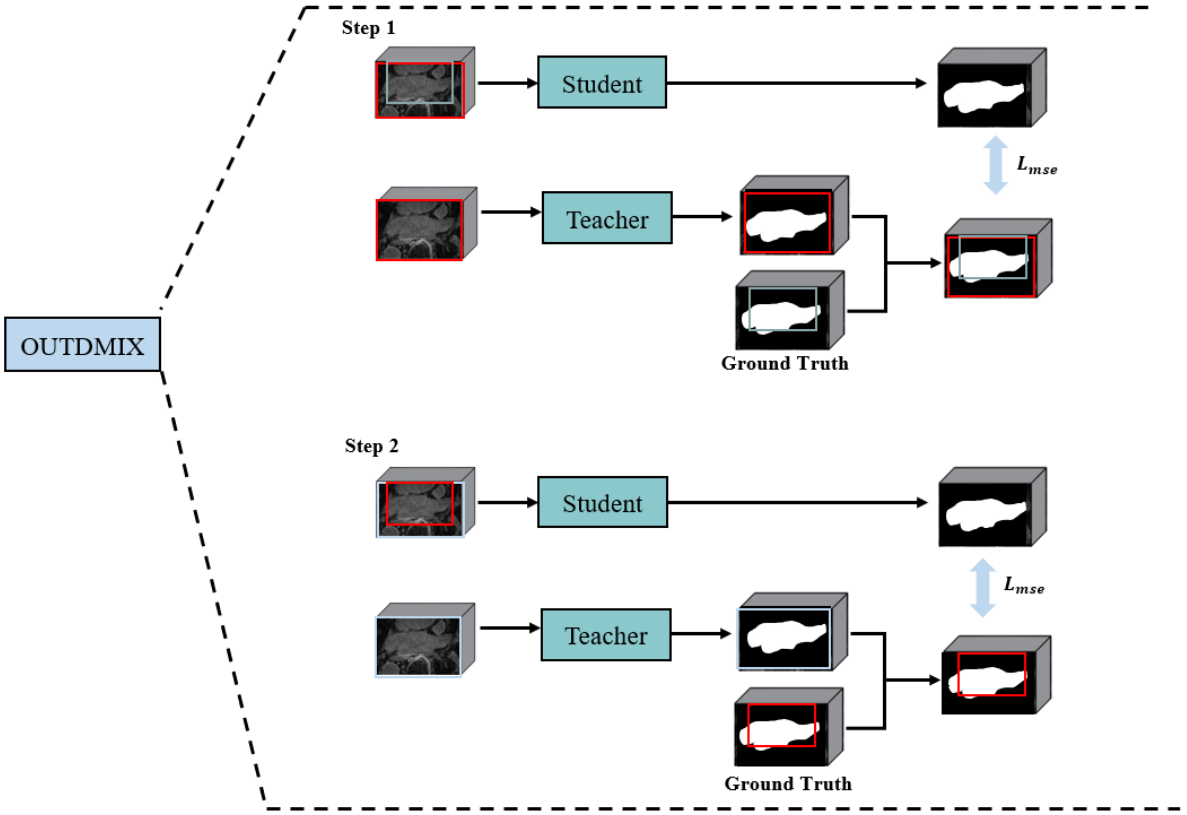

Figure 3: Overview of OUTDMIX

III. EXPERIMENTS

We evaluated our method on the Atrial Segmentation Challenge dataset. It provides 100 3D gadolinium-enhanced MR imaging scans (GE-MRIs) and LA segmentation mask for training and validation. These scans have an isotropic resolution of 0:625 × 0:625 × 0:625mm^3.We split the 100 scans into 80 scans for training and 20 scans for evaluation. All the scans were cropped centering at the heart region for better comparison of the segmentation performance of different methods, and normalized as zero mean and unit variance.

The framework was implemented in PyTorch, using two NVIDIA RTX 3090 GPU. We used the SGD optimizer to update the network parameters(weight decay=0:0001, momentum=0.9). The initial learning rate was set as 0.01 and divided by 10 every 2500 iterations. We totally trained 6000 iterations as the network has converged. The batch size was 4, consisting of 2 annotated images and 2 unannotated images. We randomly cropped $112 \times 112 \times 80$ sub-volumes as the network input and the final segmentation results were obtained using a sliding window strategy. We used the standard data augmentation techniques on-the-fly to avoid overfitting following, including randomly flipping, and rotating with 90, 180 and 270 degrees along the axial plane. The parameters λ and μ both set as 1 , the cropping parameter set as 0.5.

Table 1 Comparison between our method and various methods on LA dataset

Method	Scans used	Metrics			
	Labeled Data	Dice	Jaccard	95HD	ASD
UA-MT[4]	4(5%)	82.53	70.62	13.64	3.79
SASSNet[5]		81.24	69.51	16.12	3.64
DTC[6]		81.23	69.12	14.34	3.89
URPC[7]		82.36	71.32	14.49	3.61
MC-Net		83.44	72.96	14.14	2.74
SS-Net[8]		86.29	76.23	10.06	2.42
Ours		**87.96**	**77.64**	**8.36**	**2.16**
UA-MT[4]	8(10%)	87.62	78.19	8.63	2.15
SASSNet[5]		87.36	78.15	9.72	2.61
DTC[6]		87.48	78.21	8.17	2.34
URPC[7]		86.91	76.96	11.06	2.39
MC-Net		87.59	78.13	10.06	1.86
SS-Net[8]		88.16	79.63	7.52	1.93
Ours		**89.24**	**80.52**	**6.91**	**1.77**

We compare our framework on LA dataset with various competitors: UA-MT, SASSNet ,DTC, URPC, MC-Net and SS-Net. Semi-supervised experiments of different labeled ratios (i.e. 5% and 10%) are carried out. As shown in Table 1, our method achieves the best performance on all four evaluation metrics, outperforming other competitors by a big margin. We can see that in the case of 5% labeled data, our model improved by 1.67, 1.41, 1.7, and 0.26 in four metrics compared to the optimal cases of other algorithms. This indicates that our algorithm can achieve excellent segmentation performance even with a small number of labels, because our Hybrid Cropping algorithm can effectively learn the overall distribution of the segmented dataset. In the case of 10% labeled data, we still achieved the best performance, achieving improvements of 1.08, 0.89, 0.61, and 0.09 compared to other algorithms. This reflects the excellent learning ability of our multi model learning method when the amount of labeled data increases. In summary, our algorithm has demonstrated its excellent segmentation performance on the LA dataset.

REFERENCES

[1]. Xiong, Z., Fedorov, V. V., Fu, X., Cheng, E., Macleod, R., & Zhao, J. (2018). Fully automatic left atrium segmentation from late gadolinium enhanced magnetic resonance imaging using a dual fully convolutional neural network. IEEE transactions on medical imaging, 38(2), 515-524.

[2]. Zhang, Y., Xiang, T., Hospedales, T. M., & Lu, H. (2018). Deep mutual learning. In Proceedings of the IEEE conference on computer vision and pattern recognition (pp. 4320-4328).

[3]. Wu, Y., Ge, Z., Zhang, D., Xu, M., Zhang, L., Xia, Y., & Cai, J. (2022). Mutual consistency learning for semi-supervised medical image segmentation. Medical Image Analysis, 81, 102530.

[4]. Yu, L., Wang, S., Li, X., Fu, C. W., & Heng, P. A. (2019). Uncertainty-aware self-ensembling model for semi-supervised 3D left atrium segmentation. In Medical image computing and computer assisted intervention–MICCAI 2019: 22nd international conference, Shenzhen, China, October 13–17, 2019, proceedings, part II 22 (pp. 605-613). Springer International Publishing.

[5]. Li, S., Zhang, C., & He, X. (2020). 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 23 (pp. 552-561). Springer International Publishing.

[6]. Luo, X., Chen, J., Song, T., & Wang, G. (2021, May). Semi-supervised medical image segmentation through dual-task consistency. In Proceedings of the AAAI conference on artificial intelligence (Vol. 35, No. 10, pp. 8801-8809).

[7]. Luo, X., Liao, W., Chen, J., Song, T., Chen, Y., Zhang, S., ... & Zhang, S. (2021). Efficient semi-supervised gross target volume of nasopharyngeal carcinoma segmentation via uncertainty rectified pyramid consistency. In Medical Image Computing and Computer Assisted Intervention–MICCAI 2021: 24th International Conference, Strasbourg, France, September 27–October 1, 2021, Proceedings, Part II 24 (pp. 318-329). Springer International Publishing.

[8]. Wu, Y., Wu, Z., Wu, Q., Ge, Z., & Cai, J. (2022, September). Exploring smoothness and class-separation for semi-supervised medical image segmentation. In International conference on medical image computing and computer-assisted intervention (pp. 34-43). Cham: Springer Nature Switzerland.