Beyond General Edge Utilization: Edge Attention Mean Teacher for Semi-Supervised Medical Image Segmentation

Kaiwei Sun[1], Luhan Wang[1] ✉, and Jin Wang[1]

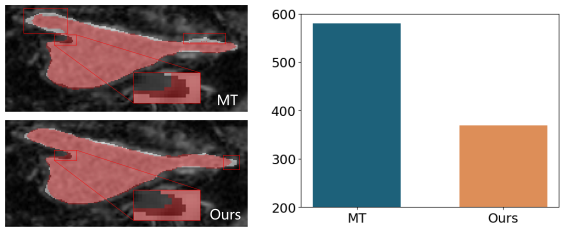
Key Laboratory of Data Engineering and Visual Computing, Chongqing University of Posts and Telecommunications, CHINA sunkw@cqupt.edu.cn, s230201109@stu.cqupt.edu.cn, wangjin@cqupt.edu.cn

Abstract. Deep learning has achieved substantial success in the field of semi-supervised medical image segmentation. Current researches mainly concentrate on enhancing pseudo-label generation process and refining consistency regularization architectures. However, the edge information, which is essential for medical image segmentation but scarce in semi-supervised scenario, is often overlooked. To address this problem, we present an edge attention mean teacher (EAMT) method that goes beyond general edge extraction to better leverage the edge information for improved segmentation performance. Particularly, based on a novel definition of edge, we propose a new edge extraction method to boost the edge extraction capability of model. Furthermore, we elaborately design an edge-aware loss function that uses the extracted edges as additional supervision for labeled data and as masks for unlabeled data. The EAMT method is characterized by its capability to extract and leverage robust edge information to promote the learning process for both labeled and unlabeled data. We evaluate the segmentation performance of the proposed EAMT method on two public 3D datasets (LA and Pancreas-CT). Experimental results demonstrate that EAMT achieves superior segmentation performance compared to several state-of-the-art methods in semi-supervised medical image segmentation.

Keywords: Semi-supervised learning · Medical image segmentation · Edge attention · Mean teacher.

1 Introduction

Medical image segmentation, such as computed tomography (CT) and magnetic resonance imaging (MRI), is essential for many clinical applications [8, 23, 36]. Recent years have witnessed the remarkable success of deep learning in fully supervised scenario. These supervised methods heavily rely on a large number of annotated medical images to guarantee their segmentation performance. However, well annotation of medical images is time-consuming. Semi-supervised learning aims at learning robust models with only a little labeled data along with a large amount of unlabeled data, making it more data-efficient.



(a) visualization of edge regions (b) number of bias pixels

Fig. 1. Edge information is crucial to medical image segmentation. In figure(a), red pixels mean the predictions while white pixels are the ground-truth. The predictions on bias areas marked with red rectangles are key to differentiate the segmentation performance. Figure(b) shows the number of the bias pixels predicted by MT method and our proposed EAMT method.

The core issue in semi-supervised learning is how to effectively utilize unlabeled data. Currently, two mainstream techniques, consistency regularization and pseudo-labeling, are widely recognized for their excellent performance. Most semi-supervised methods incorporate dual networks to get additional features from unlabeled data [8]. Mean Teacher (MT) [18] serves as a foundational benchmark for consistency regularization, and its variants [19, 25, 33] have a significant impact on medical image segmentation. Another approach is pseudo-labeling, which focuses on enhancing the quality of pseudo-labels [24] . Despite the promising outcomes of these methods, they usually overlook the edge information, which is crucial for accurate segmentation of medical images, as depicted in Fig. 1(a). However, the edge regions, which usually encompass the ground-truth and the background, pose a significant challenge for model to accurately segment. In computer vision, conventional wisdom has it that the shallow layers of deep convolutional neural network (CNN) architectures are rich in edge information. Moreover, due to the robust feature extraction abilities of attention modules, numerous studies have integrated attention blocks into their works to bolster the extraction of edge information [20, 35]. Owing to the accessibility of a large number of labeled medical images, these supervised methods have successfully harnessed edge information. However, edge information has not yet been effectively adapted to the semi-supervised learning domain.

In the context of semi-supervised medical image segmentation, maximizing the use of edge information can be achieved by prompting the model to capture extensive edge cues from labeled data, which can then be utilized to inform the segmentation of unlabeled data. We accomplish this by introducing a novel definition of edge and proposing the EAMT method. Specifically, the EAMT method is built on a typical MT architecture and employs the V-Net [14] as backbone. We incorporate an edge attention module within the top two layers of the encoder. Furthermore, we introduce an edge extraction method to get edge from labeled data, which is further utilized as additional supervision for labeled data and as masks for unlabeled data. We also develop a novel edge-aware loss

function that leverages edge information to enhance learning from labeled data and imposes edge consistency to promote learning from unlabeled data. As shown in Fig. 1(b), our proposed EAMT method obtains much less predicted bias pixels than MT. We further evaluate the performance of the proposed EAMT method on two benchmark 3D datasets in semi-supervised medical image segmentation: LA dataset and Pancreas-CT dataset.

In summary, our work makes the following contributions:

- Based on the MT architecture, we propose the EAMT method that aims at leveraging edge information to enhance segmentation performance.
- We present a novel definition of edge, which goes beyond the general edge extraction to gain more robust edge information.
- Based on the new definition of edge, we develop an edge extraction module that leverages attention mechanism to extract rich edge information.
- We design an edge-aware loss function that utilizes edge information to promote learning from unlabeled data, and uses the edge information as additional supervision to strengthen learning from labeled data.
- We validate our proposed EAMT method on two public benchmark datasets, and the experimental results show that EAMT outperforms several state-of-the-art methods, demonstrating the effectiveness of our method.

2 Related Work

2.1 Semi-Supervised Medical Image Segmentation

Consistency regularization based methods are prevalent in semi-supervised medical image segmentation. These methods build upon the smoothness assumption that minor perturbations should not significantly alter the outputs of model. The pioneering work MT [18] adds noise to teacher network and utilizes an exponential moving average (EMA) to update the parameters of teacher network, which is formulated as:

$$\theta'_t = \lambda\theta'_{t-1} + (1-\lambda)\theta_t, \tag{1}$$

where θ'_t denotes the parameters of the teacher network at time step t, θ_t denotes the parameters of the student network, and λ is the smoothing hyperparameter that regulates the degree of smoothing between the new and old parameters. Another common approach is pseudo-labeling, where a segmentation network first generates predictions for the unlabeled data, and these predictions are subsequently used as pseudo-labels to guide the supervised learning process [31].

Whether employing consistency regularization or adopting pseudo-labeling, mainstream methods in semi-supervised medical image segmentation adopt a structure with dual networks to better leverage unlabeled data. For instance, in the work of ABD [5], an adaptive bidirectional displacement mechanism is utilized to mitigate the limitations that mixed perturbations impose on two subnets. The MCF framework [24] allows two networks to learn from each other by a mutual correction mechanism. The BCP [1] uses two networks and a bidirectional copy-paste technique to learn common features from unlabeled data. The

UAMT [33] explores uncertainty information, which improves the performance of general MT. Luo et al. [13] use two different types of subnets to obtain more robust information. The SAMT-PCL [3] constructs one encoder with dual decoders to obtain predictions and uncertainty maps from different perspectives. Despite their success, those methods usually overlook the valuable edge information. Our proposed EAMT places greater emphasis on edge information and employs a robust attention module to enhance the model's capability of edge feature extraction.

2.2 Edge-Related Works

Edge information, which is widely recognized as a critical feature in computer vision, is extensively utilized in supervised medical image segmentation. In the ET-Net [35], an edge guidance module is incorporated to extract edge details from the encoder, and this extracted edge information is subsequently utilized in a weight aggregation module. In the DCAN [2], a CNN is employed to handle substantial changes in appearance and produce detailed probabilistic maps with high accuracy. The EANet [20] designs an edge attention module to enhance the edge information extraction ability of model. Cheng et al. [4] utilized directional feature maps to tackle the blurred margins problem. In their work, Yang et al. [32] developed an improved active contour model, which is capable of extracting robust edge information from images. Despite the advancements, edge extraction strategies prevalent in supervised scenario have not been extended to the semi-supervised medical image segmentation domain.

2.3 Attentions in Computer Vision

Attention mechanisms have demonstrated remarkable efficacy in Natural Language Processing (NLP) tasks and have been increasingly integrated into Computer Vision (CV) tasks. The applications of attention in CV can be categorized into three primary categories: spatial attention, channel attention, and hybrid attention. In CNNs, each layer generates a feature map. Spatial attention focuses on learning a weight for each pixel across all channels within the feature map. The non-local neural networks [22] calculate the response at a specific location by taking a weighted sum of the features. The SMSA module [17] captures spatial information from each feature channel, thereby improving the network's capacity to discern fine details. The channel attention approach involves learning distinct weights for each channel. In the milestone work SENet [9], inter-channel relationships are modeled to adjust the feature responses on a per-channel basis. The ECA-Net [21] improves the structure of SENet and gets lower model complexity. In BA-Net [34], a bridge attention module is proposed, which amplifies channel attention through the integration of feature information from various convolutional layers. The hybrid attention, which represents a synergistic application of different attentions, has been explored in numerous studies [6, 17, 26]. Attention mechanisms have exhibited extraordinary feature extraction capabilities; however, simply stacking these attentions is far from being effective. Our

proposed edge attention module strategically employs channel attention solely in the shallow layers of the encoder.

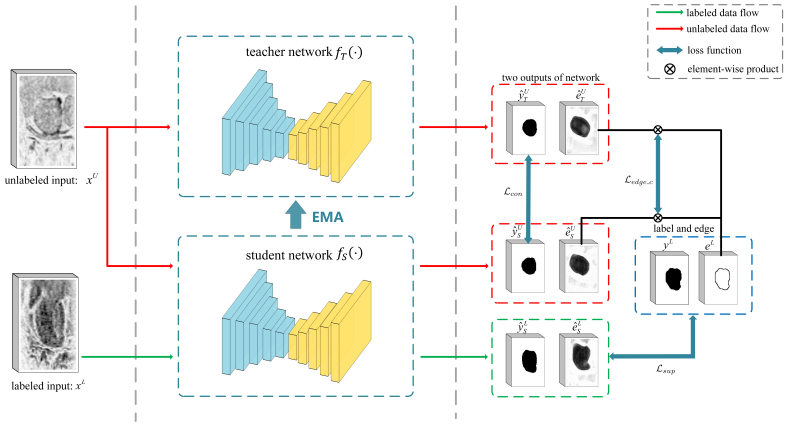


Fig. 2. Architecture of the proposed EAMT method.

3 Methodology

3.1 Overview of EAMT Method

In the semi-supervised medical image segmentation scenario, we assume that the training data consists of a labeled dataset $\mathcal{D}_L = (x_i^L, y_i^L)_{i=1}^N$, and an unlabeled dataset $\mathcal{D}_U = (x_i^U)_{i=N+1}^{N+M}$. The number of labeled images is much less than that of unlabeled images, i.e. $N \ll M$. Here, $x_i \in \mathbb{R}^{H\times W\times D}$ represents the medical image, $y_i \in \{0,1\}^{H\times W\times D}$ is the corresponding ground-truth. In our proposed EAMT method, we integrate the extracted edge $e_i \in \{0,1\}^{H\times W\times D}$ into labeled dataset, and reformulate the labeled dataset as $\mathcal{D}_L = (x_i^L, y_i^L, e_i^L)_{i=1}^N$. The proposed EAMT method is built upon the MT architecture, with a well designed edge attention module equipped within both teacher network $f_T(\cdot)$ and student network $f_S(\cdot)$. Moreover, a novel edge-aware loss function is also integrated into the EAMT. The architecture of EAMT is shown in Fig. 2.

During the training phase, both labeled and unlabeled data are fed into the student network. While only the unlabeled data is fed into the teacher network. The outputs of two networks include segmentation prediction $\hat{y}$ and edge prediction $\hat{e}$:

$$\hat{y}_S^L, \hat{e}_S^L, \hat{y}_S^U, \hat{e}_S^U = f_S((x^L, x^U)), \tag{2}$$

$$\hat{y}_T^U, \hat{e}_T^U = f_T(x^U + \epsilon), \tag{3}$$

where the subscripts S and T denote the student network and teacher network, respectively, ϵ represents the random noise (perturbations).

The loss function of EAMT comprises both supervised loss and unsupervised loss. And the loss function is utilized to update the parameters of student network while parameters of teacher network are updated by EMA (see Eqn.1).

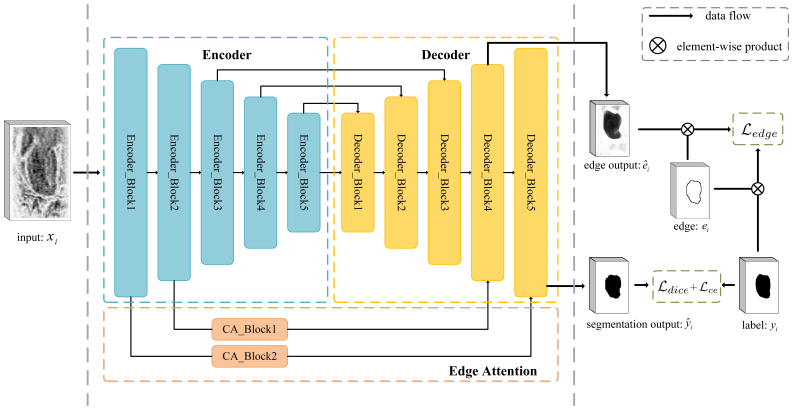


Fig. 3. Improved network structure.

3.2 Edge Extraction Module

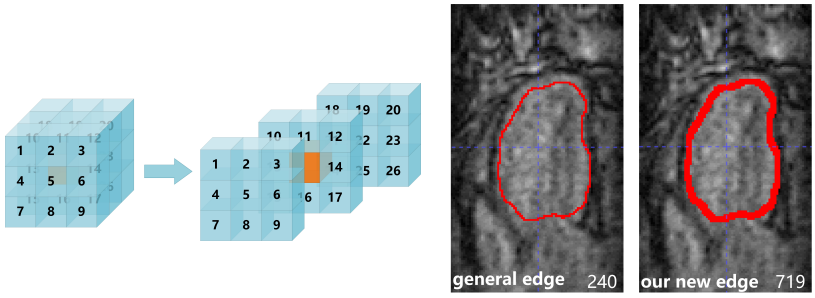
Robust edge information extraction holds a core position in our proposed EAMT. The design of edge extraction module comprises two parts: a novel definition of edge and an edge attention block.

Definition of edge. The general edge is defined as the outermost one pixel along the boundary, containing only a few pixels. Consequently, the edge extracted by conventional methods tends to be sparse and is very sensitive to noise. To achieve more robust edge extraction, we present a novel edge definition. Specifically, instead of viewing the outermost one pixel as edge, we define the outermost boundary as b_i and consider its surrounding 26 pixels as edge. The new definition of edge is explained in Fig.4(a) and formulated as follows:

$$b_i(p) = \big(1 \in \mathcal{N}(p)\big) \wedge \big(0 \in \mathcal{N}(p)\big) \wedge \big(x_i^L(p) == 1\big), \tag{4}$$

$$e_i^L(p) = b_i(p) \vee \big(\exists_{p' \in \mathcal{N}(p)} b_i(p')\big), \tag{5}$$

where p and p' represent a pixel, $x_i^L(\cdot)$ denotes the value of the pixel in the image, $e_i^L(\cdot)$ denotes the value of the pixel in the edge, b_i signifies the general edge, and $b_i(\cdot)$ indicates the value of the pixel, $\mathcal{N}(\cdot)$ refers to the 26 neighbouring pixels of a given pixel. By considering more neighbouring pixels along the boundary, the newly defined edge can encompass more pixels along the boundary, improving the robustness of edge. The difference between general edge and edge extracted by our method is shown in Fig.4(b).



(a) explanation of new edge definition (b) visualization of extracted edge

Fig. 4. A novel definition of edge. In figure (a), each element in the cube represents a pixel. The value of this central orange pixel in the edge map is determined by all 27 pixels within a larger cube. The left side of figure (b) shows the general edge extracted by conventional method, the outermost edge contains only 240 pixels over 10000 pixels, while the right side of figure (b) depicts the edge extracted by our method, which contains 719 pixels (more than 5%).

Edge attention. Inspired by the success of attention mechanism, we propose to displace the standard residual connections between the first two layers of encoder and last two layers of decoder with an edge attention module that comprises two channel attention (CA) blocks, which consists of three layers: an average pooling layer, a linear layer with ReLU, and a linear layer with Sigmoid. The average pooling layer ensures that every pixel of a channel contributes to the weight of this channel, and then two linear layers are used to calculate the weights of channels, which are also viewed as attention scores. The calculation of channel attention block is formulated as:

$$\mathbf{Y} = \mathbf{X} \times \sigma\left(\text{Linear}_2(\text{ReLU}(\text{Linear}_1(\text{AvgPool}(\mathbf{X}))))\right), \tag{6}$$

where $\mathbf{X}$ is the input of channel attention block, $\mathbf{Y}$ is the output of channel attention block, $\sigma(\cdot)$ represents the sigmoid function, $\text{Linear}(\cdot)$ denotes a linear transformation, and $\text{AvgPool}(\cdot)$ means average pooling.

3.3 Edge-aware Loss Function

The edge extraction module can extract robust edge information. Subsequently, another core issue is how to leverage the edge information to promote the learning. To this end, we elaborately design an edge-aware loss function, which is a combination of supervised loss $\mathcal{L}_{sup}$ and unsupervised loss $\mathcal{L}_{unsup}$:

$$\mathcal{L}_{overall} = \mathcal{L}_{sup} + \alpha\mathcal{L}_{unsup}, \tag{7}$$

where α is a balancing factor that controls the weight of inherent consistency.

For labeled data, we use the combination of Dice loss [14] and cross entropy loss to supervise the model training, as done in many medical image segmentation

researches [1, 24, 33]. Beyond that, we utilize edge information extracted from the labeled data as additional supervision. Thus, the loss function on labeled data is formulated as follows:

$$\mathcal{L}_{sup}(\hat{y}_S^L, y^L, \hat{e}_S^L, e^L) = 0.5\left(\mathcal{L}_{dice}(\hat{y}_S^L, y^L) + \mathcal{L}_{ce}(\hat{y}_S^L, y^L)\right) + \beta\mathcal{L}_{edge}, \tag{8}$$

$$\mathcal{L}_{edge} = \frac{\mathrm{SUM}\left(\left((\hat{e}_S^L - y^L)\cdot e^L\right)^2\right)}{\sum_{p=1}^{P} e^L(p)}, \tag{9}$$

where $\hat{y}_S^L$ is the predicted segmentation, y^L is the ground-truth, $\hat{e}_S^L$ is the edge prediction, e^L is the edge, $\mathcal{L}_{dice}(\hat{y}_S^L, y^L)$ denotes the Dice loss, $\mathcal{L}_{ce}(\hat{y}_S^L, y^L)$ represents the cross entropy loss, β is a balancing factor, $\mathcal{L}_{edge}$ is edge loss, function $\mathrm{SUM}(\cdot)$ is the sum of the pixel-level values in a feature map, P represents the total number of pixels in the image, and p signifies an individual pixel within that image, the calculation of $e^L(p)$ is shown in Eqn.5. From Fig.3, we can see that the edge output is extracted from the penultimate decoder, which obtains edge information from the edge attention module. Therefore, the loss $\mathcal{L}_{edge}$ helps to update the parameters of edge attention module. Additionally, to avoid redundant computations, the loss $\mathcal{L}_{edge}$ only takes into account the edge regions that are crucial for medical image segmentation.

For unlabeled data, we calculate the consistency loss between student network and teacher network. Specifically, the consistency loss consists of segmentation consistency loss $\mathcal{L}_{con}$ and edge consistency loss $\mathcal{L}_{edge\_c}$. The loss function $\mathcal{L}_{unsup}$ on unlabeled data is formulated as:

$$\begin{aligned}\mathcal{L}_{unsup}(\hat{y}_S^U, \hat{y}_T^U, \hat{e}_S^U, \hat{e}_T^U, e^L) = \mathcal{L}_{con}(\hat{y}_S^U, \hat{y}_T^U)\\ +\gamma\mathcal{L}_{edge\_c}(\hat{e}_S^U, \hat{e}_T^U, e^L),\end{aligned} \tag{10}$$

$$\mathcal{L}_{con}(\hat{y}_S^U, \hat{y}_T^U) = \left(\hat{y}_S^U - \hat{y}_T^U\right)^2, \tag{11}$$

$$\mathcal{L}_{edge\_c}(\hat{e}_S^U, \hat{e}_T^U, e^L) = \frac{\mathrm{SUM}\left(\left((\hat{e}_S^U - \hat{e}_T^U)\cdot e^L\right)^2\right)}{\sum_{p=1}^{P} e^L(p)}, \tag{12}$$

where γ is also a balancing factor, $\hat{y}$ represents the segmentation prediction, the subscripts S and T denote student network and teacher network, respectively, and the superscript U means unlabeled data. From Eqn.12, we can see that the edge e^L extracted from the labeled data are utilized as masks for unlabeled data. By imposing edge consistency between student and teacher networks, the learning from unlabeled data can leverage the edge information from labeled data to achieve better segmentation. By utilizing the edge extracted from labeled data as masks, a synergistic learning framework between labeled and unlabeled data is created, where edge information from labeled data can promote learning from unlabeled data and unlabeled data can provide more details about edge regions to help robust edge extraction on labeled data.

4 Experiments

4.1 Datasets

Two public datasets are used in our experiments, including the LA dataset [30], which contains 100 3D MRI images, and the Pancreas-CT dataset [15], which comprises 82 CT scans. For fair comparison, we preprocess the two datasets following the previous works [1, 16]. We also compare the segmentation performance of our proposed EAMT method with that of several state-of-the-art methods for semi-supervised medical image segmentation.

4.2 Implementation and Experimental Setting

Implementation Configurations. Our proposed EAMT is implemented using PyTorch and trained on an NVIDIA 4070 GPU. We employ the 3D V-Net architecture as the backbone, which is a de facto choice for many medical image segmentation tasks. We utilize the SGD optimizer with a weight decay of 0.0001 and a momentum factor of 0.9. The initial learning rate is set to 0.01, and we adopt a polynomial decay strategy to adjust the learning rate at each iteration.

For the LA dataset, we set the maximum number of iterations to 15k. We set the batch size to 8, comprising 4 labeled and 4 unlabeled samples. For the Pancreas-CT dataset, we set the maximum iteration to 10k, and a batch size of 4, with 2 labeled and 2 unlabeled samples. The parameter α in Eqn. 7 is set following the work in [18]. The parameter β in Eqn. 8 is setting to 0.1 for Pancreas-CT and 0.5 for LA dataset, besides, every 1.5k iterations, we will multiply this parameter for the LA dataset by 0.5. And γ in Eqn. 10 is set to 1.0 for both two datasets.

Following previous works in semi-supervised medical image segmentation [1, 5, 10, 33], our experiments were conducted with two typical semi-supervised settings, i.e. training with 10% labeled data and training with 20% labeled data. Four metrics were adopted to evaluate the segmentation performance: Dice similarity coefficient (Dice), Jaccard similarity coefficient (Jaccard), 95% Hausdorff Distance (95HD), and Average Surface Distance (ASD).

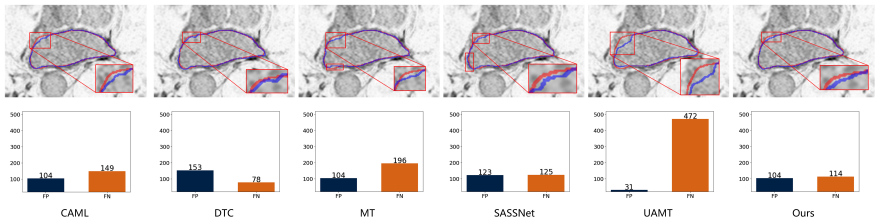


Fig. 5. Visualization of segmentations on LA dataset. The blue line represents predicted segmentation and the red line means the ground-truth. The bias areas are highlighted by red rectangles. We also count the FP and FN pixels in each slice.

Table 1. Segmentation Results on LA dataset

Method	Volumes Used		Dice(%)↑	Jaccard(%)↑	95HD(voxel)↓	ASD(voxel)↓
	Labeled	Unlabeled				
V-Net	8(10%)	0	82.74	71.72	3.26	13.35
V-Net	16(20%)	0	86.03	76.06	3.51	14.26
V-Net	80(100%)	0	91.65	83.82	1.60	5.28
MT	8(10%)	72	83.50	72.72	2.67	12.74
UAMT	8(10%)	72	84.25	73.48	3.36	13.84
DTC	8(10%)	72	87.42	**78.06**	**2.40**	**8.37**
CAML	8(10%)	72	**87.54**	77.95	2.57	10.76
SASSNet	8(10%)	72	86.79	76.90	4.10	14.56
EAMT(ours)	8(10%)	72	**89.93**	**81.86**	**1.74**	**7.02**
UMCT	16(20%)	64	89.36	81.01	2.60	7.25
MT	16(20%)	64	88.22	79.20	2.73	10.75
UAMT	16(20%)	64	88.59	79.67	2.14	8.51
DTC	16(20%)	64	89.42	80.98	2.10	7.32
UPC	16(20%)	64	89.65	81.36	2.15	6.71
MC-Net	16(20%)	64	90.34	82.48	1.77	**6.00**
CAML	16(20%)	64	**90.71**	**83.07**	**1.59**	6.08
SASSNet	16(20%)	64	89.17	80.69	2.86	8.57
EAMT(ours)	16(20%)	64	**90.95**	**83.51**	**1.71**	**6.61**

↑: the higher the better; ↓: the lower the better; best two results are marked in bold.

4.3 Experimental Results

Segmentation Results on LA Dataset. We compared the segmentation performance of our proposed EAMT method with that of the benchmark semi-supervised learning method MT [18] and several state-of-the-art methods, including UMCT [28], UAMT [33], DTC [12], UPC [11], MC-Net [27], CAML [7] and SASSNet [10]. The segmentation results of our proposed EAMT and the comparing methods are presented in Table 1, where the best two results in terms of four evaluation metrics are marked in bold. Overall, in 10% labeled data setting, EAMT achieves the best segmentation performance in terms of four evaluation metrics. The proposed EAMT improves the Dice from 82% to almost 90% with only 10% labeled data, and yields nearly identical Jaccard scores compared with fully supervised learning methods when using 20% labeled data. In both 10% and 20% settings EAMT outperforms several state-of-the-art semi-supervised methods, which demonstrates the superiority of our method. Moreover, EAMT significantly outperforms MT which is the base architecture of EAMT, indicating its effectiveness in leveraging edge information. In Fig. 5, we have visualized the segmentation results on LA dataset. Compared with other edge-cutting methods, the proposed EAMT method can yield more accurate segmentation, especially for complex region areas.

Table 2. Segmentation results on Pancreas-CT dataset

Method	Volumes Used		Dice(%)↑	Jaccard(%)↑	95HD(voxel)↓	ASD(voxel)↓
	Labeled	Unlabeled				
V-Net	6	0	58.41	46.81	18.43	50.03
V-Net	12	0	71.63	56.81	8.67	19.54
V-Net	62	0	82.46	69.65	1.42	6.76
UMCT	6	56	67.74	53.59	7.41	16.34
MT	6	56	65.13	51.98	7.03	23.06
UAMT	6	56	66.44	51.02	**5.19**	20.42
DTC	6	56	67.58	52.79	6.16	**15.57**
FUSSNet	6	56	**68.32**	**54.01**	5.85	17.46
EAMT(ours)	6	56	**70.20**	**55.51**	**2.57**	**14.17**
UMCT	12	50	76.42	62.98	5.40	14.34
MT	12	50	75.82	62.03	3.37	13.09
UAMT	12	50	78.26	62.72	**3.09**	10.43
DTC	12	50	77.19	63.75	4.25	**9.36**
FUSSNet	12	50	**79.25**	**63.71**	3.47	9.52
EAMT(ours)	12	50	**79.60**	**66.57**	**2.18**	**8.22**

Segmentation Results on Pancreas-CT Dataset. We compare the proposed EAMT method with several state-of-the-art and milestone works, including UMCT [28], MT, UAMT [33], DTC [12], and FUSSNet [29]. The segmentation results in terms of four evaluation metrics are reported in Table 2. Overall, the proposed EAMT method yields the best segmentation results over four evaluation metrics in both semi-supervised settings, which indicates the effectiveness of our proposed EAMT method in semi-supervised medical image segmentation. The visualization of segmentation results is shown in Fig.6. It can be observed that as the number of training iterations increases, the model's segmentation performance on edges improves progressively.

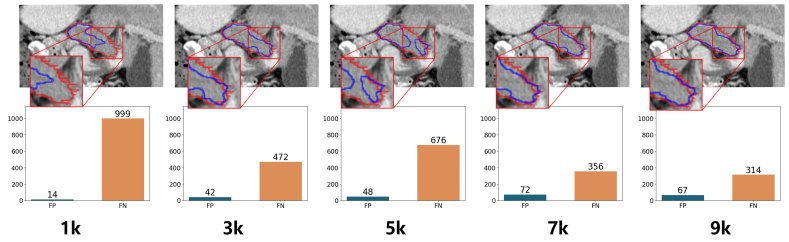


Fig. 6. Visualization of segmentations on Pancreas-CT dataset with different iterations. The blue line represents predicted segmentation and the red line means the ground-truth. The bias areas are highlighted by red rectangles. We also count the FP and FN pixels in each slice.

4.4 Ablation Study

To verify the effectiveness of our EAMT, we conducted several ablation experiments. The experimental settings were kept consistent: two datasets (20% labeled and 80% unlabeled).

Table 3. Segmentation results with different attention block numbers

Block Numbers	dataset	Dice(%)↑	Jaccard(%)↑
0	LA dataset	89.63	81.37
1	LA dataset	90.16	82.28
2	LA dataset	**90.95**	**83.51**
3	LA dataset	90.47	82.74
4	LA dataset	90.71	83.12
0	Pancreas-CT	77.41	64.13
1	Pancreas-CT	78.32	65.10
2	Pancreas-CT	**79.60**	**66.57**
3	Pancreas-CT	77.63	64.60
4	Pancreas-CT	78.30	64.95

Table 4. Segmentation results with different types of attention block

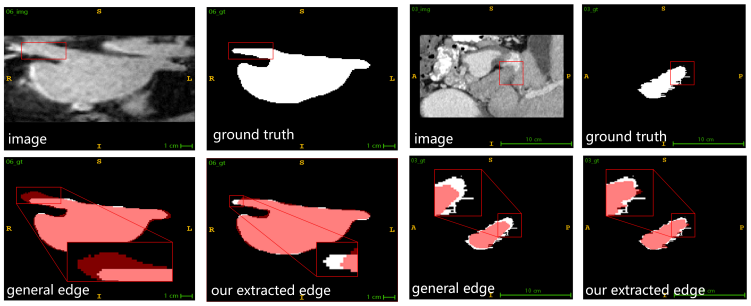
attention types	dataset	parameters↓	dice(%)↑
Spatial Attention	LA dataset	1372	89.05
Channel Attention	LA dataset	**374**	**90.95**
Spatial Attention	Pancreas-CT	1372	77.51
Channel Attention	Pancreas-CT	**374**	**79.60**

The Effect of Edge Attention. We conducted experiments on the edge attention module in two dimensions: the number of attention blocks and the types of attention modules. Firstly, we have conducted with different numbers, the results are presented in Table 3. We can see that setting the number to 2 yields the best performance, and an increased number of attention blocks does not directly lead to better performance. This can be attributed to the fact that a single layer attention block is insufficient for capturing robust edge information, while 3 and 4 layers tend to capture more abstract features rather than concrete edge details. Furthermore, we have investigated the impact of different types. The segmentation results are reported in Table 4. We can observe that the channel attention block not only has fewer parameters but also outperforms the spatial attention block, making it a more suitable choice for our application.

Edge Extraction. In order to verify the robustness of our edge extraction method, we conducted experiments on two datasets. From Table 5, we can observe that our designed edge extraction method obtains superior segmentation performance than the general edge extraction method. Furthermore, from Fig.7, we can see that our proposed EAMT method produces more accurate segmentation over the complex edge regions, which also demonstrates the robustness of our edge extraction method.

Table 5. Segmentation results with different edge extraction methods

method	dataset	Dice(%)↑	Jaccard(%)↑
general edge	LA dataset	89.32	80.93
our edge	LA dataset	**90.95**	**83.51**
general edge	Pancreas-CT	70.67	56.85
our edge	Pancreas-CT	**79.60**	**66.57**



(a) segmentation on LA dataset (b) segmentation on Pa dataset

Fig. 7. Visualization of segmentations using general edge and our proposed new edge.

Edge-aware Loss Function. In order to verify the importance of edge-aware loss functions, we conducted four ablation experiments. The results are presented in Table 6. We can observe that both types of edge-aware loss functions perform effectively, yielding improved segmentation performance compared to the baseline model, which does not utilize edge-aware loss functions. Moreover, the combination of the two edge-aware functions results in an augmented effect.

Table 6. Segmentation results with different loss functions

used loss functions		dataset	Dice(%)↑	Jaccard(%)↑
$\mathcal{L}_{edge}$	$\mathcal{L}_{edge\_c}$			
×	×	LA dataset	88.84	80.47
✓	×	LA dataset	89.87	81.80
×	✓	LA dataset	89.70	81.67
✓	✓	LA dataset	**90.95**	**83.51**
×	×	Pancreas-CT	77.63	64.42
✓	×	Pancreas-CT	78.68	65.64
×	✓	Pancreas-CT	78.28	64.92
✓	✓	Pancreas-CT	**79.60**	**66.57**

5 Conclusion

The edge attention mean teacher (EAMT) method presented in this study aims to enhance the performance of semi-supervised medical image segmentation by effectively harnessing edge information. Our approach introduces several innovative methods and modules to achieve this goal. We define a novel edge segmentation task that can be addressed by a plug-and-play edge attention module. Notably, we introduce a new edge extraction method and an edge-aware loss function, which allow us to utilize the edge extracted from labeled data for both supervising the learning process on labeled data and guiding the learning process on unlabeled data. The experimental results on the LA dataset and Pancreas-CT dataset substantiate the effectiveness of our EAMT method in leveraging edge information. This work underlines the significance of edge information in the field of semi-supervised medical image segmentation. And future work will focus on optimizing these methods.

Acknowledgments. The authors have no competing interests to declare that are relevant to the content of this article.

References

1. 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). pp. 11514–11524 (2023)
2. Chen, H., Qi, X., Yu, L., Dou, Q., Qin, J., Heng, P.A.: Dcan: Deep contour-aware networks for object instance segmentation from histology images. Medical Image Analysis **36**, 135–146 (2017)
3. Chen, Y., Chen, F., Huang, C.: Combining contrastive learning and shape awareness for semi-supervised medical image segmentation. Expert Systems with Applications **242**, 122567 (2024)
4. Cheng, F., Chen, C., Wang, Y., Shi, H., Cao, Y., Tu, D., Zhang, C., Xu, Y.: Learning directional feature maps for cardiac mri segmentation. In: Medical Image

Computing and Computer Assisted Intervention – MICCAI 2020. pp. 108–117 (2020)
5. Chi, H., Pang, J., Zhang, B., Liu, W.: Adaptive bidirectional displacement for semi-supervised medical image segmentation. In: 2024 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR). pp. 4070–4080 (2024)
6. Fu, J., Liu, J., Tian, H., Li, Y., Bao, Y., Fang, Z., Lu, H.: Dual attention network for scene segmentation. In: 2019 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR). pp. 3141–3149 (2019)
7. Gao, S., Zhang, Z., Ma, J., Li, Z., Zhang, S.: Correlation-aware mutual learning for semi-supervised medical image segmentation. In: Medical Image Computing and Computer Assisted Intervention - MICCAI. vol. 14220, pp. 98–108 (2023)
8. Han, K., Sheng, V.S., Song, Y., Liu, Y., Qiu, C., Ma, S., Liu, Z.: Deep semi-supervised learning for medical image segmentation: A review. Expert Systems with Applications **245**, 123052 (2024)
9. Hu, J., Shen, L., Albanie, S., Sun, G., Wu, E.: Squeeze-and-excitation networks. IEEE Transactions on Pattern Analysis and Machine Intelligence **42**, 2011–2023 (2020)
10. 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. vol. 12261, pp. 552–561 (2020)
11. Lu, L., Yin, M., Fu, L., Yang, F.: Uncertainty-aware pseudo-label and consistency for semi-supervised medical image segmentation. Biomedical Signal Processing and Control **79**, 104203 (2023)
12. Luo, X., Chen, J., Song, T., Wang, G.: Semi-supervised medical image segmentation through dual-task consistency. In: Proceedings of the AAAI Conference on Artificial Intelligence. vol. 35, pp. 8801–8809 (2021)
13. Luo, X., Hu, M., Song, T., Wang, G., Zhang, S.: Semi-supervised medical image segmentation via cross teaching between cnn and transformer. In: International Conference on Medical Imaging with Deep Learning. vol. 172, pp. 820–833 (2022)
14. Milletarì, F., Navab, N., Ahmadi, S.A.: V-net: Fully convolutional neural networks for volumetric medical image segmentation. In: 2016 Fourth International Conference on 3D Vision (3DV). pp. 565–571 (2016)
15. Roth, H.R., Lu, L., Farag, A., Shin, H.C., Liu, J., Turkbey, E.B., Summers, R.M.: Deeporgan: Multi-level deep convolutional networks for automated pancreas segmentation. In: Medical Image Computing and Computer-Assisted Intervention – MICCAI 2015. vol. 9349, pp. 556–564 (2015)
16. Shi, Y., Zhang, J., Ling, T., Lu, J., Zheng, Y., Yu, Q., Qi, L., Gao, Y.: Inconsistency-aware uncertainty estimation for semi-supervised medical image segmentation. IEEE Transactions on Medical Imaging **41**, 608–620 (2022)
17. Si, Y., Xu, H., Zhu, X., Zhang, W., Dong, Y., Chen, Y., Li, H.: Scsa: Exploring the synergistic effects between spatial and channel attention. arXiv preprint arXiv:2407.05128 (2024)
18. Tarvainen, A., Valpola, H.: Mean teachers are better role models: Weight-averaged consistency targets improve semi-supervised deep learning results. In: Proceedings of the 31st International Conference on Neural Information Processing Systems. p. 1195–1204 (2017)
19. Wang, K., Zhan, B., Zu, C., Wu, X., Zhou, J., Zhou, L., Wang, Y.: Semi-supervised medical image segmentation via a tripled-uncertainty guided mean teacher model with contrastive learning. Medical Image Analysis **79**, 102447 (2022)

20. Wang, K., Zhang, X., Zhang, X., Lu, Y., Huang, S., Yang, D.: Eanet: Iterative edge attention network for medical image segmentation. Pattern Recognition **127**, 108636 (2022)
21. Wang, Q., Wu, B., Zhu, P., Li, P., Zuo, W., Hu, Q.: Eca-net: Efficient channel attention for deep convolutional neural networks. In: 2020 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR). pp. 11531–11539 (2020)
22. Wang, X., Girshick, R., Gupta, A., He, K.: Non-local neural networks. In: 2018 IEEE Conference on Computer Vision and Pattern Recognition (CVPR). pp. 7794–7803 (2018)
23. Wang, Y., Zhou, Y., Shen, W., Park, S., Fishman, E.K., Yuille, A.L.: Abdominal multi-organ segmentation with organ-attention networks and statistical fusion. Medical Image Analysis **55**, 88–102 (2019)
24. Wang, Y., Xiao, B., Bi, X., Li, W., Gao, X.: Mcf: Mutual correction framework for semi-supervised medical image segmentation. In: 2023 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR). pp. 15651–15660 (2023)
25. Wang, Z., Zheng, J.Q., Voiculescu, I.: An uncertainty-aware transformer for mri cardiac semantic segmentation via mean teachers. In: Medical Image Understanding and Analysis. vol. 13413, pp. 494–507 (2022)
26. Woo, S., Park, J., Lee, J.Y., Kweon, I.S.: Cbam: Convolutional block attention module. In: Computer Vision – ECCV 2018. pp. 3–19 (2018)
27. Wu, S., Li, J., Liu, C., Yu, Z., Wong, H.S.: Mutual learning of complementary networks via residual correction for improving semi-supervised classification. In: 2019 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR). pp. 6493–6502 (2019)
28. Xia, Y., Liu, F., Yang, D., Cai, J., Yu, L., Zhu, Z., Xu, D., Yuille, A., Roth, H.: 3d semi-supervised learning with uncertainty-aware multi-view co-training. In: 2020 IEEE Winter Conference on Applications of Computer Vision (WACV). pp. 3635–3644 (2020)
29. Xiang, J., Qiu, P., Yang, Y.: Fussnet: Fusing two sources of uncertainty for semi-supervised medical image segmentation. In: Medical Image Computing and Computer Assisted Intervention – MICCAI 2022. vol. 13438, pp. 481–491 (2022)
30. Xiong, Z., Xia, Q., Hu, Z., Huang, N., Zhao, J.: A global benchmark of algorithms for segmenting the left atrium from late gadolinium-enhanced cardiac magnetic resonance imaging. Medical Image Analysis **67**, 101832 (2021)
31. Yang, X., Song, Z., King, I., Xu, Z.: A survey on deep semi-supervised learning. IEEE Transactions on Knowledge and Data Engineering **35**, 8934–8954 (2023)
32. Yang, Y., Hou, X., Ren, H.: Efficient active contour model for medical image segmentation and correction based on edge and region information. Expert Systems with Applications **194**, 116436 (2022)
33. 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: Medical Image Computing and Computer Assisted Intervention – MICCAI 2019. vol. 11765, pp. 605–613 (2019)
34. Zhang, R., Zou, R., Zhao, Y., Zhang, Z., Chen, J., Cao, Y., Hu, C., Song, H.: Banet: Bridge attention in deep neural networks. arXiv preprint arXiv:2410.07860 (2024)
35. Zhang, Z., Fu, H., Dai, H., Shen, J., Pang, Y., Shao, L.: Et-net: A generic edge-attention guidance network for medical image segmentation. In: Medical Image Computing and Computer Assisted Intervention – MICCAI 2019. pp. 442–450 (2019)

36. Zhao, X., Qi, Z., Wang, S., Wang, Q., Wu, X., Mao, Y., Zhang, L.: Rcps: Rectified contrastive pseudo supervision for semi-supervised medical image segmentation. IEEE Journal of Biomedical and Health Informatics **28**, 251–261 (2024)