Neurocomputing 548 (2023) 126411

Contents lists available at ScienceDirect

Neurocomputing

journal homepage: www.elsevier.com/locate/neucom

Mutually aided uncertainty incorporated dual consistency regularization with pseudo label for semi-supervised medical image segmentation

Shanfu Lu [b], Zijian Zhang [a,*], Ziye Yan [b], Yiran Wang [c], Tingting Cheng [a], Rongrong Zhou [a], Guang Yang [d]

[a] Department of Oncology, National Clinical Research Center for Geriatric Disorder, Xiangya Hospital, Central South University, Changsha 410008, China
[b] Perception Vision Medical Technologies Co. Ltd., Guangzhou 510530, China
[c] Department of Statistics, University of British Columbi., 3182 Earth Sciences Building, 2207 Main Mall, Vancouver, BC, Canada
[d] National Heart and Lung Institute, Imperial College London, Cardiovascular Research Centre, Royal Brompton Hospital, London and School of Biomedical Engineering Imaging Sciences, King's College London, London, United Kingdom

ARTICLE INFO

Article history:
Received 6 March 2023
Revised 25 April 2023
Accepted 29 May 2023
Available online 2 June 2023
Communicated by Zidong Wang

Keywords:
Medical image segmentation
Semi-supervised learning
Consistency-Regularization
Pseudo label
Cycle loss

ABSTRACT

Semi-supervised learning has contributed plenty to promoting computer vision tasks. Especially concerning medical images, semi-supervised image segmentation can significantly reduce the labor and time cost of labeling images. Among the existing semi-supervised methods, pseudo-labelling and consistency regularization prevail; however, the current related methods still need to achieve satisfactory results due to the poor quality of the pseudo-labels generated and needing more certainty awareness the models. To address this problem, we propose a novel method that combines pseudo-labelling with dual consistency regularization based on a high capability of uncertainty awareness. This method leverages a cycle-loss regularized to lead to a more accurate uncertainty estimate. Followed by the uncertainty estimation, the certain region with its pseudo-label is further trained in a supervised manner. In contrast, the uncertain region is used to promote the dual consistency between the student and teacher networks. The developed approach was tested on three public datasets and showed that: 1) The proposed method achieves excellent performance improvement by leveraging unlabeled data; 2) Compared with several state-of-the-art (SOTA) semi-supervised segmentation methods, ours achieved better or comparable performance.

1. Introduction

Medical image computer-aided diagnosis (CAD) technology has achieved significant success in the medical field by using advanced deep-learning algorithms to support clinical decision-making [1–3]. Medical image segmentation as an essential part of medical image CAD has made significant progress. Recently, with the rising of computing capabilities and novel deep learning based methods, the quality of image segmentation has been improved significantly [4–8]. However, a majority of these models utilize the data in a supervised manner which is both labour and time-intensive. As it is more convenient to obtain unlabelled data, the idea of semi-supervised learning is proposed [9–11].

Semi-supervised methods performs supervised training on the labelled data, and exploit the unlabelled data to enrich the model. Among current semi-supervised learning methods, consistency regularization and pseudo-labelling have achieved promising results. Pseudo-label based methods [12] initially train the model on the labelled data and generate pseudo-labels on the unlabelled data based on the predicted probability map. For example, [12] generate pseudo-labels for the unlabelled data by picking up the class which had the maximum predicted probability every weights update. By minimizing the entropy for unlabeled data, the overlap of class probability distribution could be reduced which consequently favored a low-density separation between classes. In addition, regarding the consistency regularization methods, data augmentation is applied to unlabeled data and optimizes the model to make consistent predictions regardless of the augmentations applied. The Pi model [13] generated variations of the unlabelled data by applying random augmentations and network dropout methods. The mean teacher network consisted of a student and a teacher network [14]. Specifically, the teacher network's parameters were obtained by applying an exponential moving average on the student network, and consistency regularization was performed on the two networks' outputs. [15] proposed a net-

* Corresponding author.
E-mail addresses: lushanfu@pvmedtech.com (S. Lu), wanzzj@csu.edu.cn (Z. Zhang).

https://doi.org/10.1016/j.neucom.2023.126411
0925-2312/

work that enforces consistency regularization between two tasks rather than two predicted samples. It builds the consistency between a global level set function regression task and a pixel-wise classification task to consider geometric constraints.

Despite that these methods greatly made the semi-supervised training on medical images more effective and the improved segmentation accuracy proved the rationale behind these methods, there still some limitations: 1) Considering that the pseudo-labels are generated on a pixel-level, some predicted pixel might not be accurate due to its high uncertainty, so a pseudo-label for the entire sample might be misleading. 2) Most consistency regularization methods promote consistency on the uncertain region, so this method could be inferior if the uncertainty calculation is unreliable. Presently, the uncertainty calculation depends on the result of two models with mostly identical settings, and the certain area is the pixels with the same predictions after a softmax transformation, whereas the pixels with different outcomes represent the uncertain region. While this method splits the two regions successfully, it loses efficacy at the beginning stage, because the uncertain region could occupy the majority of the image and it could be scatteredly distributed given that the model performs poorly initially. While this issue propagates, it does not construct a positive cooperation between the consistency loss or the pseudo labelling part of the model. Consequently, we can compute the uncertainty from a different but more efficient perspective, eliminating scatteredly distributed and large-scale uncertain regions.

To this end, we propose a dual consistency regularization network with pseudo labels for semi-supervised medical image segmentation. In particular, we combine pseudo-labelling with consistency regularization to make better use of the certain and uncertain regions of the unlabelled samples. In addition, we further investigate the uncertainty calculation for unlabelled images from a new perspective. Specifically instead of having the model outputs two predictions and directly calculate the uncertainty from them, we propose a cycle loss to enforce a consistent and rather accurate prediction between the two outcomes while assigning a different weight to the background and foreground elements.

Our model is structured on the basis of the mean teacher network [14] with some modifications to the student network. As illustrated in Fig. 1, the uncertainty calculation is incorporated into the student network indicated by an encoder-aux decoders structure. The main segmentation head is the encode-decoder part of the student network, and the teacher network has the same structure as the main segmentation head in the student network but its parameters are obtained as the exponential moving average of the student network's parameters.

Our main contribution can be summarised as follows:

- A semi-supervised structure for medical image segmentation that combines pseudo-labeling with consistency regularization.
- A novel uncertainty calculation method was obtained by enforcing a cycle loss between the two model predictions.
- A asymmetry mean-teacher model architecture was proposed to compute prediction consistency and reduce computation compared to the symmetric structure.
- Comprehensive experiments studies have been done on three public datasets for organ and tumor segmentation, demonstrating the effectiveness of the proposed semi-supervised.

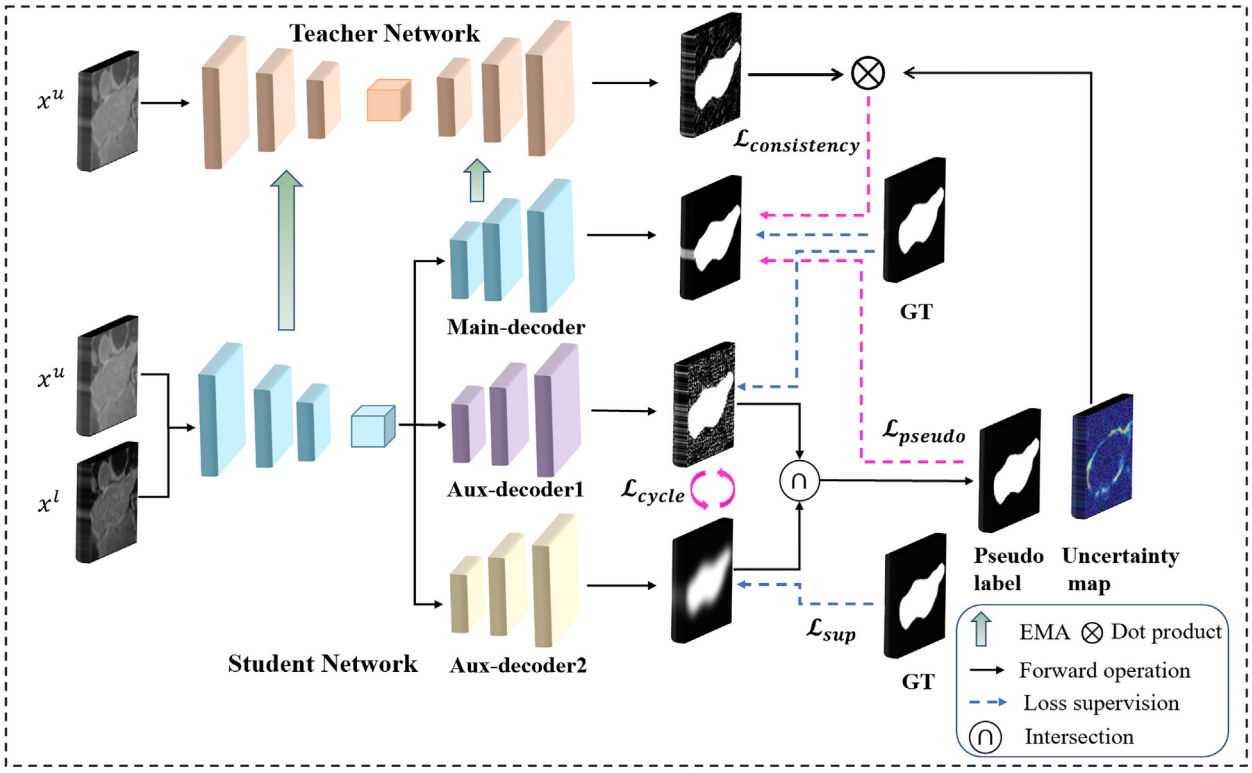

Fig. 1. The overview of our proposed framework for semi-supervised medical image segmentation. The framework adopts a variant of the mean-teacher architecture, where the student branch uses multiple decoders. Two auxiliary decoders are used for the uncertainty calculation and pseudo-label generation, and the teacher branch is used for the consistency calculation. The teacher branch and pseudo labels are used to construct the dual Consistency Loss so that the student branch can learn more reliable information from the unlabeled data for semi-supervised learning. The model's input is the labeled image $x^l \in D_L$ and the unlabeled image $x^u \in D_U$, and the output is the segmentation results of x^l and x^u, as well as the uncertainty estimation map.

2. Related work

2.1. Medical image segmentation in the supervised setting

Recently, with the rapid development of deep learning methods, the convolutional neural networks have been applied to various applications including but not limited to medical image segmentation and achieved SOTA. As one of the most popular models, U-Net [6] has earned its reputation in the field of supervised medical image segmentation. This encoder-decoder structured model performs efficient segmentation using a limited number of labelled images and its unique architecture combines the location information obtained from the down sampling path and the contextual information obtained from the up sampling path to predict a fair segmentation map. To make this architecture more robust, and training efficient, a few variations of the u-net model were also proposed such as the U-Net++ [16], VNet [17], Attention-UNet [18], SD-Unet [19], 3D-Unet [20], etc. Combined with more sophisticated and effective loss function, fusion methods, and attention mechanisms, supervised deep learning has proved its significance in medical image segmentations.

2.2. Semi-supervised learning

Many deep-learning-based semi-supervised learning models have been developed to further take advantage of the labelled data while maintaining or improving the accuracy that supervised learning models can achieve. Current models can be broadly divided into the following groups depending on the method applied: Consistency regularization, Pseudo-labelling, entropy minimization, and generative models [21]. Consistency regularization methods are based on the idea that if a realistic perturbation is added to the image, a good and consistent model should output similar segmentation maps. In general, the perturbation can be added to either the input images or the model parameters. Besides, the model is trained to give consistent predictions on the unlabelled images and their perturbed versions. Pseudo-labelling works around labelling the unlabelled images. These methods train the model on the labelled data to provide pseudo-labels for the unlabelled data. Generally, self-training [22,23] or multi-view training [24] is applied in such methods. The confidence (or certainty) of a prediction can be estimated by the entropy of the probability map. As a result, the entropy minimization approach minimizes the entropy of the predictions in order to best maximize the confidence of the predictions. The generative method of the semi-supervised learning can be interpreted as an extension of unsupervised learning. Unsupervised learning uses clustering to determine the distribution of the given unlabelled data; however, this estimates distribution might be untrustworthy given that only a sample of the real-world data is used. Therefore, the labelled data can be used to help the model to correct the decision boundaries of the distribution generated by the unlabelled data. Variational autoencoders [25,26] and generative adversarial networks [27] are the two main generative models used for semi-supervised models.

2.3. Uncertainty awareness

To better analyze the confidence of the predictions on the unlabelled images, it is crucial to have an uncertainty calculation method. Data augmentation is one of the most common methods of computing uncertainty [28]. For example, a small perturbation can be added to the images or features space, and the ideal case would be that this noise would not affect the final result. By comparing the regions with and without a consensus, uncertain and certain regions can simply be obtained. While this method makes changes to the images, another perspective is to modify the network while keeping the input image the same. An example is to use Monte Carlo dropout at test time, this way the model randomly deactivates some parameters to enforce the model and makes different predictions. [29]. developed a conservative and radical segmentation, specifically putting different costs on mis-segmenting background pixels and object class pixels. Then the outcomes of the two segmentation models are compared for certain and uncertain regions separation. While these methods perform an effective calculation of the uncertainty, it depends highly on the number of unlabelled data and its quality. In other words, the uncertain region might be scatteredly distributed or irregularity. And by proceeding with these inaccurate certain and uncertain regions, the model could lead in the wrong prediction.

2.4. Pseudo-labelling

The pseudo-labelling method is trained firstly on the labelled dataset, and then generates pseudo-labels for the unlabelled images, then these pseudo-labels are combined with the labelled images for further training. However, depending on the quality and quantity of the unlabelled images, it might not be ideal to simply produce pseudo labels and use them for further training as this may bring errors to the training process. The more common approach is to divide the predictions of the unlabelled images into certain and uncertain regions, [29] apply pseudo labelling on the rather certain region. [30] proposed a cycled pseudo label to transform the two probability outputs into soft-pseudo labels, then leverage the uncertainty information to learn more critical features by encouraging mutual consistency.

2.5. Consistency regularization

The pi-model opens up the door for consistency regularization. Using data augmentation and monte Carlo dropout, the model takes inputs that are slightly perturbed and outputs results that are different [13]. By regularizing the consistency between the model predictions while minimizing the cross-entropy loss between the labelled images and their labels, the model converges. To capture the model information from previous epochs, the temporal ensembling method is proposed on the basis of the pi-model. Furthermore, the mean teacher model once again extends the temporal ensembling method by updating the teacher network once per training step, and the consistency loss between the student's and teacher's prediction is minimized [14]. [31] developed a dual-task segmentation network which trains the model to enhance the consistency between two tasks: the global-level level set function regression task and a pixel-wise classification task. This method focuses on the consistency between two tasks other than two output datasets.

In summary, the existing methods mainly have the following shortcomings: 1) The calculation of uncertainty perception is complex and requires multiple forward propagations. 2) Weak perception of uncertainty. 3) Failure to effectively combine consistency regularities with pseudo-label techniques.

3. Proposed method

This section introduces a semi-supervised medical image segmentation framework based on pseudo-labelling and dual consistency regularization. The main contributions are as follows. An encoder-decoder architecture incorporates consistency regularization and pseudo-labelling as the main semi-supervised learning techniques. Moreover, we propose a cycle loss mechanism to

compute the pixel-level uncertainty of the predicted un-labelled images hence improving the quality of the pseudo-labels.

We first introduce the default symbol representation to describe the method proposed in this paper accurately. Suppose we train data set D, consisting of two parts D_L and D_U. $\theta_{enc}, \theta_{dec}$ are parameters of the encoder and decoder of the leading segmentation branch of the student's model. $\theta_{aux\_1}, \theta_{aux\_2}$ which are parameters of auxiliary-decoder1 (Aux-decoder1) and auxiliary-decoder2 (Aux-decoder2). θ'_{enc} and θ'_{dec} are encoder and decoder of the teacher's model, are shared from student's model θ_{enc} and θ_{dec} by EMA. $\mathcal{F}_{s\_seg}$ and $\mathcal{F}_{t\_seg}$ represent the output prediction of the student model main segmentation branch and teacher model segmentation branch to generate a probabilistic segmentation map. $\mathcal{F}_{aux\_1}$ and $\mathcal{F}_{aux\_2}$ represent the output prediction by two decoder branches of different learning abilities to generate segmentation, respectively.

3.1. Model architeture

Fig. 1 shows that the overall network architecture follows the mean teacher structure. The network comprises a teacher and a student network, but unlike the traditional mean teacher network, our student network consists of two additional auxiliary decoder modules. The primary branch of the student network is the encoder-decoder module which is the main segmentation head, and Aux-decoder1 and Aux-decoder2 are responsible for computing the uncertainty, which then is used for pseudo-label generating. The teacher module is constructed with the same encoder-decoder architecture as the student module. The segmentation model takes 3D medical images as input, and the segmentation result is the pixel-level label of the input images. Moreover, regarding the labeled data, the supervised segmentation loss is defined as follows:

$$\mathcal{L}_{sup} = \beta_1 \mathcal{L}_{dice}(\mathcal{F}_{s\_seg}, y) + \beta_2 \mathcal{L}_{ce}(\mathcal{F}_{s\_seg}, y) + \beta_3 \mathcal{L}_{ce}(\mathcal{F}_{aux\_1}, y) + \beta_4 \mathcal{L}_{ce}(\mathcal{F}_{aux\_2}, y) \quad (1)$$

$\mathcal{L}_{dice}(\mathcal{F}_{s\_seg}, y), \mathcal{L}_{ce}(\mathcal{F}_{s\_seg}, y)$ are used to calculate the dice loss and the cross entropy loss between the ground truth and the prediction made by the main segmentation branch of the student module. $\mathcal{L}_{ce}(\mathcal{F}_{aux\_1}, y)$ in Eq. 3, $\mathcal{L}_{ce}(\mathcal{F}_{aux\_2}, y)$ in Eq. 4 computes the cross-entropy of the predictions made by the two minor aux-decoders for the labelled images. $\beta_1, \beta_2, \beta_3, \beta_4$ are balance factors, all set to 1.

As mentioned earlier, Monte Carlo Dropout [11] is an ordinary method that measures uncertainty, but it requires multiple forward propagations to compute the uncertainty. More specifically, applies Monte Carlo dropout to measure the uncertainty in the following way:

$$\mu_i^c = \frac{1}{N}\sum_{n=1}^{N} p_i^{t^c} \quad and \quad \mu_i = -\sum_{c=1}^{C} \mu_i^c \log \mu_i^c \quad (2)$$

Where $p_i^{t^c}$ is the i-th pixel's probability of c-class in the n-th time prediction, $\mu_i^c$is the mean of N prediction of i-th pixel, C is the category counts, μ_i is the uncertainty estimation, $\mu_i \in \mathbb{R}^{W\times D\times H}$ N is usually set to be 8 according to [11], and while this method is effective, but needs multiple forward propagations. Different from the previous methods, we referenced and improved the CoraNet architecture [32]. In the student's network, two encoder-aux-decoders were used for uncertainty estimation for the unlabelled images. Specifically, the two auxiliary decoders share the decoder of the student network. The decoders Aux-decoder1 and Aux-decoder2 have the same structure, and they each make predictions on the background and target objects for the same input image. In other words, a binary pixel-level classification is performed. To guarantee that each of the two networks focuses on a different target, we give different weights to the background and target objects when training the model. In particular, these two auxiliary decoders are assigned different penalty weights for misclassifying the background and target objects, and the loss is defined as a binary cross-entropy loss as the following:

$$\mathcal{L}_{aux-decoder1} = \sum_{i=1}^{n} k_1 y^{(i)} \log \hat{y}^{(i)} + (1 - y^{(i)}) \log (1 - \hat{y}^{(i)}) \quad (3)$$

$$\mathcal{L}_{aux-decoder2} = \sum_{i=1}^{n} y^{(i)} \log \hat{y}^{(i)} + k_2 (1 - y^{(i)}) \log (1 - \hat{y}^{(i)}) \quad (4)$$

In which $y^{(i)}$ is the ground truth, $\hat{y}^{(i)}$ is the predicted label, k_1 and k_2 are the hyper-parameters which can be adjusted to change the model's feature learning ability on different backgrounds and target objects. The prediction of a 3D input sample $X \in \mathbb{R}^{W\times D\times H}$ made by the 2 aux-decoders based on different feature learning capacities can be written as $\mathcal{F}_{aux_1} \in \mathbb{R}^{H\times W\times D\times class}, \mathcal{F}_{aux_2} \in \mathbb{R}^{H\times W\times D\times class}$. Afterwards, the uncertainty of this input sample can be computed based on these two predictions. The uncertain region is represented as the pixels that are predicted to be different, whereas the certain region is the pixels that are predicted to be the same after a SoftMax transformation. They are considered to be certain because even these two networks are set to put different weights on learning the background and target object, they still make the same predictions on those pixels. Notation wise, they can be written as follows:

$$U^{certainty} = \mathcal{F}_{aux\_1} \bigcap \mathcal{F}_{aux\_2} \quad (5)$$

$$U^{uncertainty} = 1 - U^{certainty} \quad (6)$$

in which $\bigcap$ represents the region of the two predictions with consensus, namely the certain region. Next, we generate pseudo-labels based on the certain region of the unlabelled image. To guarantee the accuracy of the pseudo-labels, only the certain region is used for pseudo-labelling.

3.2. Cycle loss

Based on the current model architecture, to further improve the quality of the uncertainty awareness mechanism and the pseudo-labels, we propose the cycle loss calculation in order to regularize the two aux-decoders. First, we apply function $\mathcal{F}$ to transform the outputs made by the two aux-decoders $\mathcal{F}_{aux\_1}, \mathcal{F}_{aux\_2}$ to their corresponding pseudo-labels $PL_{aux\_1}, PL_{aux\_2} \in [0 or 1]^{H\times W\times D}$. Then we use PL_{aux_1}, PL_{aux_2} to mutually supervise, hence to calculate the cycled loss. In this way, when there is a limited number of labelled data, by encouraging the two aux-decoders with different feature learning capabilities to mutually learn while best maintaining the consistency and low entropy, the uncertainty and pseudo-labels obtained can have their quality improved substantially. In this way, the unlabeled images are utilized with a higher efficacy. Therefore, our proposed cycle loss is interpreted as the cross-entropy loss of the two aux-decoders which is formulated as:

$$\mathcal{L}_{cycle} = \mathcal{L}_{ce}(\mathcal{F}_{aux_1}, PL_{aux\_2}) + \mathcal{L}_{ce}(\mathcal{F}_{aux_2}, PL_{aux\_1}) \quad (7)$$

3.3. Dual consistency regularization

For semi-supervised medical image segmentation, unsupervised loss functions are used to generate supervised information. Consequently, prediction capability of the model and generalization performance can be enhanced by performing unsupervised learning with the unlabelled data. We employ the uncertainty-aware-mean teacher model to be the main approach of our semi-supervised segmentation network. Regarding consistency regular-

ization learning, the student model aims to minimize the Euclidean distance between the teacher's model's output and its own output. For training with pseudo-labels, they are the predictions made by the teacher's network after being transformed from the probability map to pseudo-labels and the student's network then utilizes them to perform further supervised training. In our work, uncertainty-aware and pseudo-labelling are completed using two aux-decoders, specifically by adjusting the penalty weights on misclassifying the background and target objects, thus promoting consistency on the certain region. The generation of pseudo-labels depends on the output results of the aux-decoders from the last iteration, and the computation of the loss of pseudo-labels only depends on the certain region generated by the model which is expressed as:

$$\mathcal{L}_{pseudo} = \frac{\sum_i \mathbb{1}(\mu^{\text{certainty}})\left(\mathcal{L}_{\text{dice}}\left(\mathcal{F}_{\text{s\_seg}}, \mathbb{y}^{\text{pseudo}}\right)\right)}{\sum_i \mathbb{1}(\mu^{\text{certainty}})} + \frac{\sum_i \mathbb{1}(\mu^{\text{certainty}})\left(\mathcal{L}_{\text{ce}}\left(\mathcal{F}_{\text{s\_seg}}, \mathbb{y}^{\text{pseudo}}\right)\right)}{\sum_i \mathbb{1}(\mu^{\text{certainty}})} \tag{8}$$

where y^{pseudo} is the pseudo-label for the unlabelled image, $\mathcal{F}_{s\_seg}$ is the probability map output made by the student network indicator function $\mathbb{1}$ represents 0 or 1, $\in R^{W\times H\times D}$ to select certain regions, certain region $\mu^{certainty} = 1 - \mu^{uncertainty}$ in Eq. 5 and Eq. 6.

The consistency loss computes the difference between the prediction made by the teacher's network and the student's network regarding the uncertain region for the unlabelled images. It is expressed as:

$$\mathcal{L}_{consistency} = \frac{\sum_i \mathbb{1}(\mu^{\text{uncertainty}})\left\|\mathcal{F}_{\text{t\_seg}} - \mathcal{F}_{\text{s\_seg}}\right\|^2}{\sum_i \mathbb{1}(\mu^{\text{uncertainty}})} \tag{9}$$

where $\mu^{certainty}$ represents the uncertain region, $\mathbb{1}(\mu^{\text{uncertainty}})$ is the indicator function used to select the uncertain regions.

3.4. Overall training procedure

The main goal of this training framework is to minimize the supervised loss $\mathcal{L}_{sup}$ in Eq. 1, pseudo-label loss $\mathcal{L}_{pseudo}$ in Eq. 8, consistency loss $\mathcal{L}_{consistency}$ in Eq. 9, as well as cycled loss $\mathcal{L}_{cycled}$ in Eq. 7. The student network first performs supervised learning on labelled data, then utilizes uncertainty estimation and consistency regularization to further learn from the unlabelled data. Therefore, the final loss is defined as:

$$\begin{aligned}\min\ & \mathcal{L}_{sup}(\theta_{enc};\theta_{dec};\theta_{aux\_1};\theta_{aux\_2};D_L)\\ &+ \lambda_1\mathcal{L}_{pseudo}(\theta_{enc};\theta_{dec};\theta_{aux\_1};\theta_{aux\_2};D_U)\\ &+ \lambda_2\mathcal{L}_{consistency}\left(\theta_{enc};\theta_{dec};\theta'_{enc};\theta'_{dec};D_U\right)\\ &+ \lambda_3\mathcal{L}_{cycle}(\theta_{aux\_1};\theta_{aux\_2};D_U)\end{aligned} \tag{10}$$

where $\lambda_1, \lambda_2, \lambda_3$ are the gradient weight parameters,which controls the balance between supervised loss and unsupervised loss, hence reduce the perturbation between supervised loss and unsupervised loss during the early training phase. The training objective of our proposed dual consistency regularization with pseudo label for medical image segmentation framework can be formulated as Algorithm 1.

Algorithm 1: Dual Constraint between Consistency Regularization and Pseudo Label for Semi-Supervised Medical Image Segmentation

Input: $x_i \in D_L + D_U, y_i \in D_L$
Output: Student model's parameter $\theta_{enc}, \theta_{dec}$, θ_{aux_1} for aux-decoder1 and θ_{aux_2} for aux-decoder2. Teacher model's parameter θ'_{enc} and θ'_{dec} , are shared from student model's parameter θ_{enc} and θ_{dec} by EMA.

```
while stop criterion not met (epochs) do
    Sample batch b_l = (x_i, y_i) ∈ D_L
    Generate output segmentation maps F_s_seg, F_aux_1, F_aux_2
    Calculate supervised segmentation loss L_sup as Eq. 1.
    Update the student model's weights θ_enc, θ_dec, θ_aux_1, θ_aux_2
    with L_sup
end while
whilestop criterion not met do
    Sample batch b_l = (x_i, y_i) ∈ D_L, and b = b_l + b_u where
    b_u = x_i ∈ D_U
    Generate output segmentation maps F_s_seg, F_aux_1, F_aux_2
    Calculate uncertainty region and pseudo label generate
    by F_aux_1 and F_aux_2
    Calculate supervised segmentation loss L_sup as Eq. 1.
    Calculate cycle loss L_cycled as Eq. 7.
    Calculate pseudo predict loss L_pseudo as Eq. 8.
    Calculate consistency regularization loss L_regularization as
    Eq. 9.
    Update the student model's weights
    θ_enc, θ_dec, θ_aux_1, θ_aux_2 with Eq. 10.
    Update the teacher model's weights θ'_enc, θ'_dec with
    exponential moving average (EMA) of the student model's
    weights as θ'_enc_t = αθ'_enc_t−1 + (1 − α)θ_enc_t, θ'_dec_t =
    αθ'_dec_t−1 + (1 − α)θ_dec_t
end while
return θ_enc, θ_dec
```

4. Experiments results

We evaluate the performance of the proposed semi-supervised medical image semantic segmentation network on three datasets, Pancreas [33], LA [34] and BraTs2019 [35]. In this section, we first present the data and the implementation details. Then, we report the results on common metrics (DSC, Jaccard, 95HD and ASD) and compare them with other methods. Finally, we study the effectiveness of each part of our proposed approach.

4.1. Dataset and experimental setup

Now, we proceed to introduce each of the three dataset and experimental designs as follows:

Pancreas [33]: This dataset is the most authoritative and popular CT scans on the pancreas and all of the images are labelled by experienced doctors. In our experiment, we collected 82 abdominal enhanced scans and their size ranges from 512x512x182 to 512x512x466 voxels. In the pre-processing stage, the image contrast is adjusted, and the CT window is changed to [-100, 240]. We eliminate the background that is unrelated to the segmentation target, then crop each slice to 192x240 pixels. Afterwards the data is split into the training and testing set which consists

of 62 and 20 samples respectively and we treat 20% of the training data as labelled data, the rest of training data as unlabelled data.

LA [34]: this dataset comes from the 2018 atrium segmentation competition, which consists of 100 3D gadolinium-enhanced MR images. The resolution of the images is 0.625x0.625x0.625 mm^3. Same as above, we split the data into training and testing sets in an 80%-20% ratio. And for the training set, we split it into labelled and unlabelled portions in two different ratios: (10%, 90%) and (20%, 80%) to see how much the results vary.

BraTs 2019 [35]: this dataset comes from different organizations and it includes the pre-operation MRI brains scans of gliomas that come from 355 patients. Every patient has MRI scans of 4 modalities including T1, T1Gd, T2, and T2-FLAIR with neuroradiolist-examined labels. All the scans are resampled to the same resolution of 1x1x1mm^3 with intensity normalized to zero mean and unit variance. In our experiment, following the method introduced in [32], we use 250 for training, 25 for validation and 60 for testing. In the training set, 2 experiments are performed for comparison: experiment 1 uses 10% of the training data as labelled data, and experiment 2 uses 20% as labelled data, the rest as unlabeled data.

4.2. Implementing details and metrics

All the experiments are developed on python and the PyTorch framework with an Nvidia p6000 GPU with 24 GB memory. In our work, to make a comparison with the current works, we use the VNet as the major network and the mean teacher model with modifications to the student network. The model training process consists of two stages, first, we train the model on the labelled images for 60 epochs, then combine it with the unlabelled data to further train for 200 epochs. During the co-training process, pseudo-labels are generated every 5 epochs. The entire model is optimized using the Adam optimizer, and we set the learning rate to be 0.001, betas to be (0.5, 0.999). The batch size is 8 which includes 4 labelled images and 4 unlabelled images. Same as the settings in CoraNet, $k_1, k_2 = 5$, $\lambda_1 = \lambda_3 = 0.5, \lambda_2(t) = 0.05 * e^{-5(1-t/t_{max})}$. During the training stage, we applied standard augmentation including flipping the images by 90, 180, and 270 degrees along the axial plane. For the LA dataset, we randomly cropped the images to 112x112x80 sub-volumes. For the pancreas and BraTs 2019 datasets, they have been randomly cropped to 96x96x96 sub-volumes as the network input. During the testing

Table 1
The comparison on 3D CT pancreas segmentation employed 20% labeled data.

Method	Scans used		Metrics			
	labeled	unlabeled	DSC(%)	Jaccard(%)	95HD	ASD
VNet([17])	12	0	70.63	56.72	22.54	6.29
VNet([17])	62	0	81.78	69.65	5.13	1.34
DAN([36])	12(20%)	50	76.74	63.29	11.13	2.97
ADVNET([37])	12(20%)	50	75.31	61.73	11.72	3.88
UA-MT([11])	12(20%)	50	77.26	63.82	11.90	3.06
SASSNet([38])	12(20%)	50	77.66	64.08	10.93	3.05
DTC([31])	12(20%)	50	78.27	64.75	8.36	2.25
CoraNet([29])	12(20%)	50	79.67	66.69	**7.59**	**1.89**
Ours	12(20%)	50	**81.31**	**69.50**	8.73	2.63

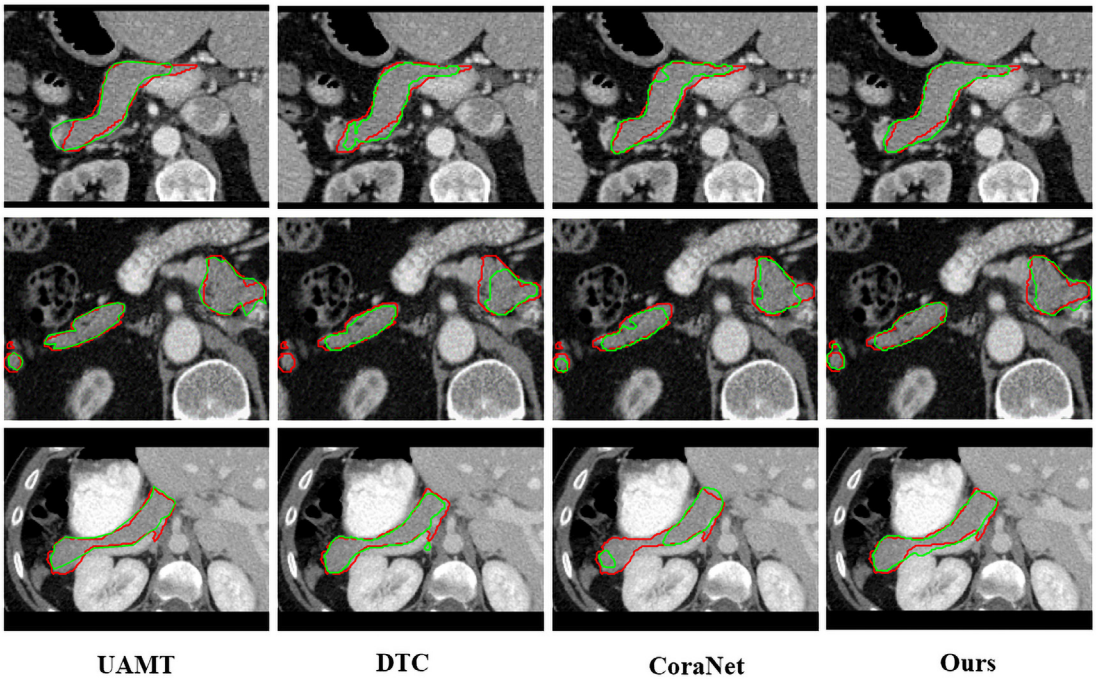

Fig. 2. Visual comparison of the pancreas results of our proposed framework with partially other state-of-the-art model using 20% labeled data. The red curve represents the ground-truth and the green represents the model prediction.

Table 2
Comparison of different methods on the LA dataset with different ratio labeled Data.

Method	Scans used		Metrics			
	labeled	unlabeled	DSC(%)	Jaccard(%)	95HD	ASD
VNet([17])	8(10%)	0	79.99	68.12	21.11	5.48
VNet([17])	16(20%)	0	86.03	76.06	14.26	3.51
VNet([17])	80(all)	0	91.14	83.82	5.75	1.52
DAP([39])	8(10%)	72	81.80	71.23	15.81	3.80
UA-MT([11])	8(10%)	72	84.25	73.48	13.84	3.36
SASSNet([38])	8(10%)	72	87.32	77.72	**9.62**	2.55
LG-ER-MT([40])	8(10%)	72	85.54	75.12	13.29	3.77
DUWM([41])	8(10%)	72	85.91	75.75	12.67	3.31
DTC([31])	8(10%)	72	86.57	76.55	14.47	3.74
CoraNet([29])	8(10%)	72	86.11	76.02	10.49	**2.46**
Ours	8(10%)	72	**87.75**	**78.50**	12.30	2.99
DAP([39])	16(20%)	64	87.89	78.72	9.29	2.74
UA-MT([11])	16(20%)	64	88.88	80.21	7.32	2.26
SASSNet ([38])	16(20%)	64	89.54	81.24	8.24	2.20
LG-ER-MT([40])	16(20%)	64	89.62	81.31	7.16	2.06
DUWM([41])	16(20%)	64	89.65	81.35	7.04	2.03
DTC([31])	16(20%)	64	89.42	80.98	7.32	2.10
CoraNet([29])	16(20%)	64	89.94	81.89	8.38	2.29
Ours	16(20%)	64	**90.87**	**83.35**	**5.79**	**1.71**

phase, we apply the sliding windows technique to obtain the final results. In order to fairly evaluate the results, we have used four complementary metrics: dice similarity coefficient (DSC), and Jaccard Index which is two region-based method to evaluate the mismatching between the regions. In addition, we also used the Average distance (ASD) and 95% Hausdorff Distance (95HD) to evaluate the difference between the boundaries of the predictions and the ground truths. To ensure a fair comparison principle, no post-processing methods were utilized.

4.3. Experiment results on pancreas dataset

We compare our proposed method with six SOTA semi-supervised segmentation models including DAN [36], ADVNet [37], UA-MT [11], SASSNet [38], DTC [31], and CoraNet [29]. To ensure a fair comparison, we used the same VNet structure as the backbone of our network exactly as in the other networks. Table 1 shows multiple metrics for different semi-supervised segmentation models on the pancreas dataset. In comparison, we also trained our model with 20% and 100% labelled images only to form the lowest and highest boundaries of the model performances. As a result, with effective incorporation of the unlabeled images, all the methods can significantly improve the segmentation accuracy, but it is worth noting that our model achieved better results on the DSC and Jaccard metric., and competitive results on the ASD and 95HD metrics. Fig. 2 visually compares the segmentation results by the baseline UAMT, DTC, CoraNet, and the proposed model when 20% of the data is labeled. It can be observed that compared to other methods, our proposed method has better accuracy and segmentation integrality.

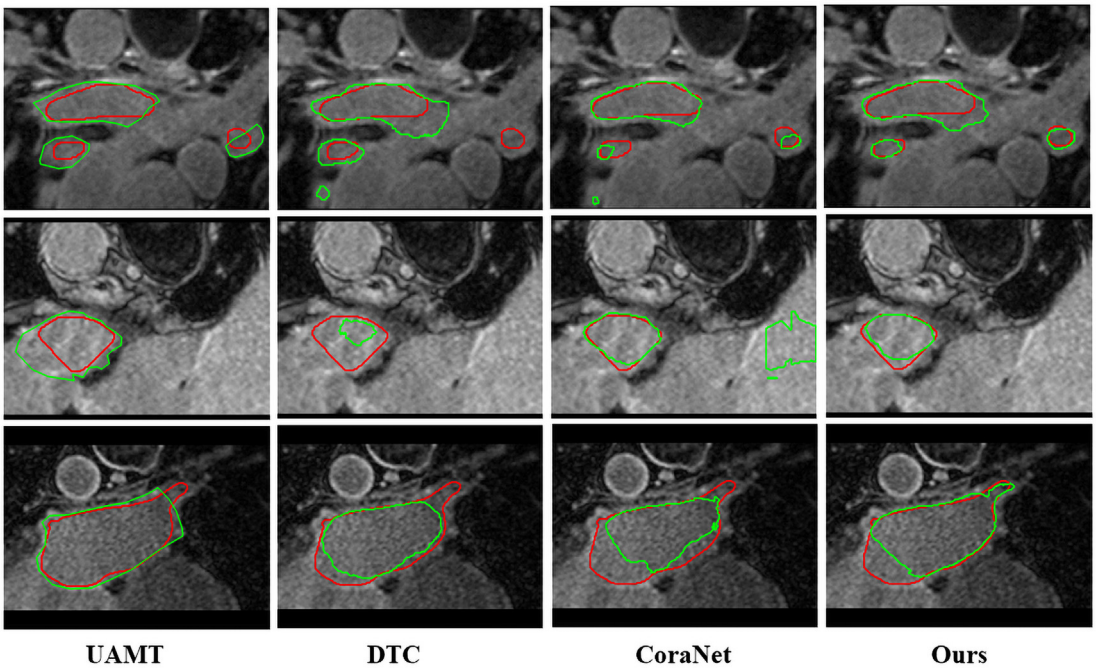

Fig. 3. Visual comparison of the LA results of our proposed framework with partially other state-of-the-art model using 20% labeled data. The red curve represents the ground-truth and the green represents the model prediction.

4.4. Experiment results on LA dataset

For the LA dataset, we also compared our method with the current SOTA semi-supervised method: DAP [39], UA-MT [11], SASS-Net [38], LG-ER-MT [40], DUWM [41], DTC [31], and CoraNet [29]. We validate the performance of 10% and 20% labeled data, respectively, and the quantitative results are shown in Table 2. From this table, it is evident that all semi-supervised methods have achieved more significant performance improvement under different proportions of labeled data than using only part of the labeled data for supervised training. Meanwhile, when the proportion of labeled data is 10% and 20%, the proposed method achieves the best dice value, 87.75 and 90.87, respectively, and other indicators also achieve a competitive result. In addition, the advantage of our model is more evident under the 10% labeled data by comparing the dice values of segmentation results with different semi-supervised models. Fig. 3 visually compares the segmentation results by the baseline UAMT, DTC, CoraNet, and the proposed model when 20% of the data is labeled. The results show that the proposed method can better utilize unlabeled data and improve the segmentation performance of the network under the condition of limited data labeling.

4.5. Experiment results on BraTs 2019 Dataset

To further validate the proposed method's effectiveness, in addition to the segmentation of some organs, we also performed brain tumor segmentation experiments. Compared with organ segmentation, tumor segmentation is a more significant challenge due to tumors' irregular and heterogeneous shape. We compared with MT [14], UA-MT [11], Entropy Mini [37], URPC [15], and DTC [31] on the BraTS 2019 dataset. In order to better explore the performance of the proposed model, we experimented on 10% (25 labeled) and 20% (50 labeled) labeled data; the results are shown in Table 3. From this table, it is evident that all semi-supervised methods have achieved more significant performance improvement under different proportions of labeled data than using only part of the labeled data for supervised training with DSC = 73.00

Table 3
Comparison of different methods on the BraTS2019 dataset with different ratio labeled Data

Method	Scans used		Metrics			
	labeled	unlabeled	DSC(%)	Jaccard(%)	95HD	ASD
VNet([17])	25	0	73.00	59.96	42.64	2.96
MT([14])	25	225	81.21	70.83	14.72	2.45
UA-MT([11])	25	225	80.85	70.32	14.61	2.57
Entropy Mini([37])	25	225	81.74	71.42	13.71	**2.37**
URPC([15])	25	225	81.80	71.63	**11.50**	2.48
DTC([31])	25	225	81.96	71.84	12.08	2.43
Ours	25	225	**83.15**	**72.90**	11.68	2.74
VNet([17])	50	0	76.14	64.15	36.01	2.70
MT([14])	50	200	82.38	72.32	14.54	2.21
UA-MT([11])	50	200	81.57	71.42	13.98	2.49
Entropy Mini([37])	50	200	82.37	72.28	15.83	2.30
URPC([15])	50	200	82.80	72.72	12.48	2.72
DTC ([31])	50	200	82.78	72.47	13.43	2.20
Ours	50	200	**84.20**	**74.78**	**7.90**	**1.82**

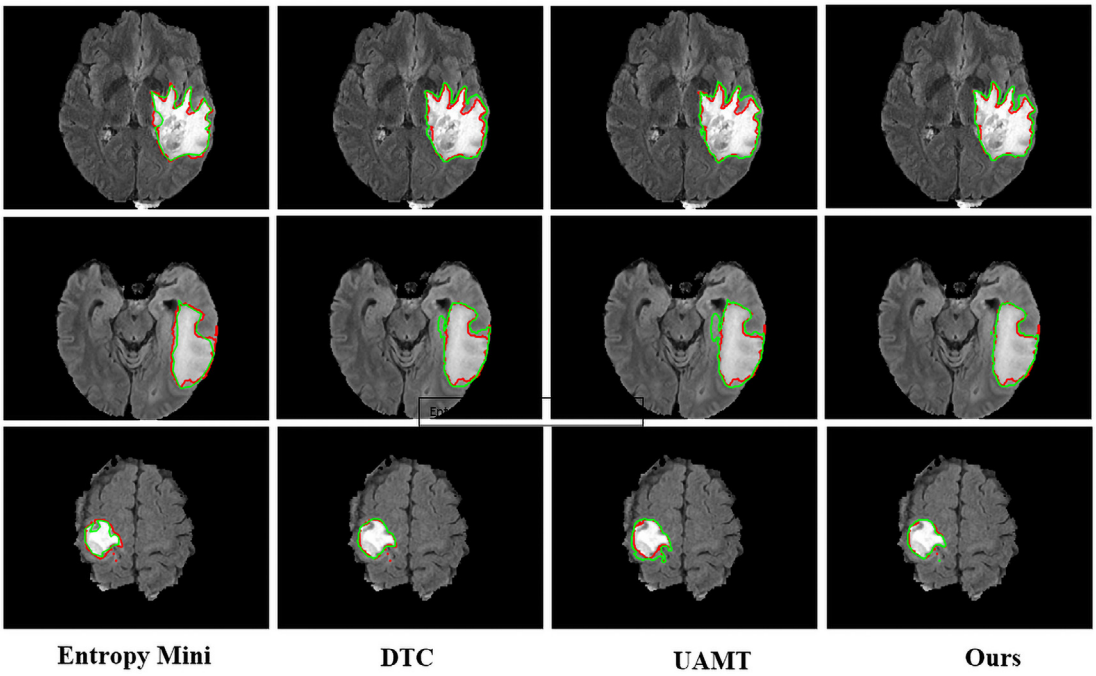

Fig. 4. Visual comparison of the BraTS2019 results of our proposed framework with partially other state-of-the-art model using 20% labeled data. The red curve represents the ground-truth and the green represents the model prediction.

and 76.14, which indicates the effectiveness of semi-supervised learning. In addition, compared with other semi-supervised methods, our method achieved the best performance on the main metrics, such as dice and Jaccard, and also achieved competitive performance on other metrics. Fig. 4 visually compares the segmentation results by the baseline DTC, UA-MT, Entropy Mini and the proposed model when 20% of the data is labeled. The red curve is the ground truth(gt), and the green is the model prediction. Intuitively, our results are more similar to gt.

4.6. Ablation studies

All ablation experiments were conducted with the pancreas dataset, which used 12 labeled and 50 unlabeled samples to further investigate the effect of different settings on segmentation.

4.6.1. Different Components Analysis

Table 4 shows the impact of each component in our method. The module ablation setting as follows:

- $\mathcal{L}_{sup}$: Employing the dice loss and cross-entropy loss for labeled data Eq. (1) for training.
- $\mathcal{L}_{consistency}$: Employing $\mathcal{L}_{sup}$ for labeled data and the consistency loss for unlabeled data Eq. (9) to joint training.
- $\mathcal{L}_{pseudo}$: Employing $\mathcal{L}_{sup}$ for labeled data and the pseudo label loss for unlabeled data Eq. (8) to joint training.
- $\mathcal{L}_{sup} + \mathcal{L}_{consistency} + \mathcal{L}_{pseudo}$: Employing $\mathcal{L}_{sup}$ for labeled data and the pseudo label loss,consistency loss for unlabeled data Eq. (8), Eq(9) to joint training.
- $\mathcal{L}_{sup} + \mathcal{L}_{consistency} + \mathcal{L}_{pseudo} + \mathcal{L}_{cycle}$: Employing $\mathcal{L}_{sup}$ for labeled data and the pseudo label loss,consistency loss,cycle loss for unlabeled data Eq. (10) to joint training.

For simplicity, we use the sign ✓ to indicate that a specific component was added during model training, as shown in Table 4. Compared with the fully supervised training method, the model performance using consistency regularity or pseudo-label has significantly improved (DSC≈8%-9%). Semi-supervised training using only pseudo-labels or consistency regularities achieves similar results. The joint optimization of consistency regularization and pseudo-label further improve the model's performance, indicating that the combination of the two has a positive gain on the model's performance. In addition, after adding cycle loss, the model achieves the best performance(dice = 81.51), which indicates that the proposed cycle loss can improve the segmentation performance of the model by using more information on unlabeled data.

4.6.2. Consistency Constraint Objective Analysis

Table 5 reports the impact of different consistency constraint objectives of the proposed method. Experiments are conducted on three objective functions: mean square error ($\mathcal{L}_{MSE}$), cross-entropy ($\mathcal{L}_{CE}$), and KL divergence ($\mathcal{L}_{KL}$). The segmentation results using the cross-entropy loss are slightly lower than the mean square error loss. KL divergence is better than the mean square error in dice, 95HD, and other metrics. One possible explanation is that it makes more sense to use the difference between the two distributions for constraints in the uncertainty region than to predict the same result.

5. Discussion

Accurate and automatic segmentation of medical images plays a vital role in improving the efficiency of doctors' diagnoses and treatment. However, it is also limited by requiring a large number of fine annotations when training a deep model with very high performance. Semi-supervised learning can learn from limited labeled data and a large amount of unlabeled data, which has shown great potential. Therefore, we proposed mutually aided uncertainty incorporated dual consistency regularization with the pseudo label for semi-supervised medical image segmentation in this paper.

Estimating uncertainty regions or generating high-quality pseudo-labels is the primary goal of semi-supervised learning [11,29,30]. There are some problems in the current work, such as the poor effect of uncertainty estimation, the complexity of the process, and the low quality of pseudo-label generation. In this paper, we also follow the basic paradigm of semi-supervised learning, and the main goal is uncertainty estimation and pseudo-label generation. Unlike the previous method of only calculating the uncertainty region or the pseudo-label, we linked the uncertainty calculation and the pseudo-label generation to optimize the model jointly. In addition, we also use the cyclic consistency constraint to enhance the perception ability of uncertainty. From Table 4, by introducing the cycle consistency loss, the unlabeled data information can be fully utilized, and the segmentation ability of the model can be improved. Compared to CoraNet [29], our approach is more straightforward and uses an asymmetric design based on the architecture of mean-teacher. The performance of the model is improved while the network parameters reduce. Despite the

Table 4
Ablation analysis of each module loss on the pancreas dataset.

$\mathcal{L}_{sup}$	$\mathcal{L}_{consistency}$	$\mathcal{L}_{pseudo}$	$\mathcal{L}_{cycle}$	Metrics			
				DSC(%)	Jaccard(%)	95HD	ASD
✓				70.63	56.72	22.54	6.29
✓	✓			79.42	66.62	8.85	2.96
✓		✓		78.75	65.66	10.12	2.77
✓	✓	✓		79.72	66.75	8.91	2.91
✓	✓	✓	✓	**81.31**	**69.50**	**8.73**	**2.63**

Table 5
Different consistent constraint objective analysis of our proposed framework on pancreas dataset.

$\mathcal{L}_{MSE}$	$\mathcal{L}_{CE}$	$\mathcal{L}_{KL}$	Metrics			
			DSC(%)	Jaccard(%)	95HD	ASD
✓			81.31	**69.50**	8.73	2.63
	✓		81.12	69.34	7.51	2.71
		✓	**81.53**	69.28	**6.82**	**2.54**

promising results of the proposed method on several datasets, we only focus on relatively small datasets for binary classifications. In future work, we aim to perform multi-classification tasks on more diversified and challenging datasets. Furthermore, consider including the discrepancy between the distribution of labeled and unlabeled images in our optimization targets and select the representative unlabeled images as the learning targets to further improve and validate the performance of semi-supervised medical image segmentation models in real clinical applications.

6. Conclusion

This paper presents a semi-supervised learning framework for medical image segmentation based on consistency regularization and pseudo-labeling. We designed two auxiliary decoders to calculate the uncertainty and generate pseudo-labels by adjusting their learning strategies. A cycle loss mechanism was proposed to improve the capability of uncertainty awareness and the quality of the pseudo-labels. The model was optimized based on the prediction of the uncertain regions and pseudo-labels so that more information could be obtained from the unlabelled images. The proposed method has been evaluated on three popular public datasets and achieved better performance than other semi-supervised methods.

CRediT authorship contribution statement

Shanfu Lu: Software, Formal analysis, Writing - original draft. **Zijian Zhang:** Conceptualization, Funding acquisition, Writing - original draft. **Ziye Yan:** Project administration. **Yiran Wang:** Investigation, Validation. **Tingting Cheng:** Resources, Funding acquisition. **Rongrong Zhou:** Writing - review & editing. **Guang Yang:** Writing - review & editing, Supervision.

Data availability

Data will be made available on request.

Declaration of Competing Interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Acknowledgment

The work was supported in part by the Science Foundation of Hunan Province (Grant No.: 2023JJ5014 and 2022JJ70072) and the Royal Society (IEC/NSFC/211235). This work was supported in part by the High Performance Computing Center of Central South University.

References

[1] P. Wu, Z. Wang, B. Zheng, H. Li, F.E. Alsaadi, N. Zeng, Aggn: Attention-based glioma grading network with multi-scale feature extraction and multi-modal information fusion, Computers in Biology and Medicine 152 (2023) .
[2] H. Li, N. Zeng, P. Wu, K. Clawson, Cov-net: A computer-aided diagnosis method for recognizing covid-19 from chest x-ray images via machine vision, Expert Systems with Applications 207 (2022) .
[3] H. Li, P. Wu, Z. Wang, J. Mao, F.E. Alsaadi, N. Zeng, A generalized framework of feature learning enhanced convolutional neural network for pathology-image-oriented cancer diagnosis, Computers in Biology and Medicine 151 (2022) .
[4] Z. Gu, J. Cheng, H. Fu, K. Zhou, H. Hao, Y. Zhao, T. Zhang, S. Gao, J. Liu, Ce-net: Context encoder network for 2d medical image segmentation, IEEE transactions on medical imaging 38 (2019) 2281–2292.
[5] X. Mao, Y. Zhao, B. Chen, Y. Ma, Z. Gu, S. Gu, J. Yang, J. Cheng, J. Liu, Deep learning with skip connection attention for choroid layer segmentation in oct images, in: 2020 42nd Annual International Conference of the IEEE Engineering in Medicine & Biology Society (EMBC), IEEE, 2020, pp. 1641–1645.
[6] O. Ronneberger, P. Fischer, T. Brox, U-net: Convolutional networks for biomedical image segmentation, in: International Conference on Medical image computing and computer-assisted intervention, Springer, 2015, pp. 234–241.
[7] S. Wang, M. Zhou, Z. Liu, Z. Liu, D. Gu, Y. Zang, D. Dong, O. Gevaert, J. Tian, Central focused convolutional neural networks: Developing a data-driven model for lung nodule segmentation, Medical image analysis 40 (2017) 172–183.
[8] T. Chen, Z. Zijian, X. Yang, S. Lu, D. Qian, X. Wang, H. Zhu, Automatic delineation of organ at risk in cervical cancer radiotherapy based on ensemble learning, Journal of Central South University, Medical sciences 47 (2022) 1058–1064.
[9] S. Chen, G. Bortsova, A. García-Uceda Juárez, G. v. Tulder, M. d. Bruijne, Multi-task attention-based semi-supervised learning for medical image segmentation, in: International Conference on Medical Image Computing and Computer-Assisted Intervention, Springer, 2019, pp. 457–465.
[10] K. Ta, S.S. Ahn, J.C. Stendahl, A.J. Sinusas, J.S. Duncan, A semi-supervised joint network for simultaneous left ventricular motion tracking and segmentation in 4d echocardiography, in: International Conference on Medical Image Computing and Computer-Assisted Intervention, Springer, 2020, pp. 468–477.
[11] L. Yu, S. Wang, X. Li, C.-W. Fu, P.-A. Heng, Uncertainty-aware self-ensembling model for semi-supervised 3d left atrium segmentation, in: International Conference on Medical Image Computing and Computer-Assisted Intervention, Springer, 2019, pp. 605–613.
[12] D.-H. Lee, et al., Pseudo-label: The simple and efficient semi-supervised learning method for deep neural networks, in: Workshop on challenges in representation learning, ICML, volume 3, 2013, p. 896.
[13] S. Laine, T. Aila, Temporal ensembling for semi-supervised learning, arXiv preprint arXiv:1610.02242 (2016).
[14] A. Tarvainen, H. Valpola, Mean teachers are better role models: Weight-averaged consistency targets improve semi-supervised deep learning results, Advances in neural information processing systems 30 (2017).
[15] X. Luo, W. Liao, J. Chen, T. Song, Y. Chen, S. Zhang, N. Chen, G. Wang, S. Zhang, Efficient semi-supervised gross target volume of nasopharyngeal carcinoma segmentation via uncertainty rectified pyramid consistency, in: International Conference on Medical Image Computing and Computer-Assisted Intervention, Springer, 2021, pp. 318–329.
[16] Z. Zhou, M.M. Rahman Siddiquee, N. Tajbakhsh, J. Liang, Unet++: A nested u-net architecture for medical image segmentation, in: Deep learning in medical image analysis and multimodal learning for clinical decision support, Springer, 2018, pp. 3–11.
[17] F. Milletari, N. Navab, S.-A. Ahmadi, V-net: Fully convolutional neural networks for volumetric medical image segmentation, in: 2016 fourth international conference on 3D vision (3DV), IEEE, 2016, pp. 565–571.
[18] O. Oktay, J. Schlemper, L.L. Folgoc, M. Lee, M. Heinrich, K. Misawa, K. Mori, S. McDonagh, N.Y. Hammerla, B. Kainz, et al., Attention u-net: Learning where to look for the pancreas, arXiv preprint arXiv:1804.03999 (2018).
[19] P.K. Gadosey, Y. Li, E.A. Agyekum, T. Zhang, Z. Liu, P.T. Yamak, F. Essaf, Sd-unet: Stripping down u-net for segmentation of biomedical images on platforms with low computational budgets, Diagnostics 10 (2020) 110.
[20] Ö. Çiçek, A. Abdulkadir, S.S. Lienkamp, T. Brox, O. Ronneberger, 3d u-net: learning dense volumetric segmentation from sparse annotation, in: International conference on medical image computing and computer-assisted intervention, Springer, 2016, pp. 424–432.
[21] Y. Ouali, C. Hudelot, M. Tami, An overview of deep semi-supervised learning, arXiv preprint arXiv:2006.05278 (2020).
[22] D. Yarowsky, Unsupervised word sense disambiguation rivaling supervised methods, in: 33rd annual meeting of the association for computational linguistics, 1995, pp. 189–196.
[23] H. Scudder, Probability of error of some adaptive pattern-recognition machines, IEEE Transactions on Information Theory 11 (1965) 363–371.
[24] A. Kumar, H. Daumé, A co-training approach for multi-view spectral clustering, in: Proceedings of the 28th international conference on machine learning (ICML-11), Citeseer, 2011, pp. 393–400.
[25] D.P. Kingma, M. Welling, Auto-encoding variational bayes, arXiv preprint arXiv:1312.6114 (2013).
[26] C. Doersch, Tutorial on variational autoencoders, arXiv preprint arXiv:1606.05908 (2016).
[27] I. Goodfellow, J. Pouget-Abadie, M. Mirza, B. Xu, D. Warde-Farley, S. Ozair, A. Courville, Y. Bengio, Generative adversarial nets, Advances in neural information processing systems 27 (2014).
[28] Y. Ouali, C. Hudelot, M. Tami, Semi-supervised semantic segmentation with cross-consistency training, in: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, 2020, pp. 12674–12684.
[29] Y. Shi, J. Zhang, T. Ling, J. Lu, Y. Zheng, Q. Yu, L. Qi, Y. Gao, Inconsistency-aware uncertainty estimation for semi-supervised medical image segmentation, IEEE Transactions on Medical Imaging 41 (2021) 608–620.
[30] Y. Wu, M. Xu, Z. Ge, J. Cai, L. Zhang, Semi-supervised left atrium segmentation with mutual consistency training, in: International Conference on Medical Image Computing and Computer-Assisted Intervention, Springer, 2021, pp. 297–306.
[31] X. Luo, J. Chen, T. Song, G. Wang, Semi-supervised medical image segmentation through dual-task consistency, in: Proceedings of the AAAI Conference on Artificial Intelligence, volume 35, 2021, pp. 8801–8809.

[32] Y. Zhang, Q. Liao, R. Jiao, J. Zhang, Uncertainty-guided mutual consistency learning for semi-supervised medical image segmentation, arXiv preprint arXiv:2112.02508 (2021).
[33] H. Roth, A. Farag, E.B. Turkbey, L. Lu, J. Liu, R.M. Summers, Data from pancreas-ct, 2016. URL: https://wiki.cancerimagingarchive.net/x/ellXAQ. DOI: 10.7937/K9/TCIA.2016.TNB1KQBU.
[34] Z. Xiong, Q. Xia, Z. Hu, N. Huang, C. Bian, Y. Zheng, S. Vesal, N. Ravikumar, A. Maier, X. Yang, et al., A global benchmark of algorithms for segmenting the left atrium from late gadolinium-enhanced cardiac magnetic resonance imaging, Medical Image Analysis 67 (2021) .
[35] S.S. Bakas, Brats miccai brain tumor dataset, 2020. URL: https://ieee-dataport.org/competitions/brats-miccai-brain-tumor-dataset. 10.21227/HDTD-5J88.
[36] Y. Zhang, L. Yang, J. Chen, M. Fredericksen, D.P. Hughes, D.Z. Chen, Deep adversarial networks for biomedical image segmentation utilizing unannotated images, in: International conference on medical image computing and computer-assisted intervention, Springer, 2017, pp. 408–416.
[37] T.-H. Vu, H. Jain, M. Bucher, M. Cord, P. Pérez, Advent: Adversarial entropy minimization for domain adaptation in semantic segmentation, in: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, 2019, pp. 2517–2526.
[38] S. Li, C. Zhang, X. He, Shape-aware semi-supervised 3d semantic segmentation for medical images, in: International Conference on Medical Image Computing and Computer-Assisted Intervention, Springer, 2020, pp. 552–561.
[39] H. Zheng, L. Lin, H. Hu, Q. Zhang, Q. Chen, Y. Iwamoto, X. Han, Y.-W. Chen, R. Tong, J. Wu, Semi-supervised segmentation of liver using adversarial learning with deep atlas prior, in: International Conference on Medical Image Computing and Computer-Assisted Intervention, Springer, 2019, pp. 148–156.
[40] W. Hang, W. Feng, S. Liang, L. Yu, Q. Wang, K.-S. Choi, J. Qin, Local and global structure-aware entropy regularized mean teacher model for 3d left atrium segmentation, in: International Conference on Medical Image Computing and Computer-Assisted Intervention, Springer, 2020, pp. 562–571.
[41] Y. Wang, Y. Zhang, J. Tian, C. Zhong, Z. Shi, Y. Zhang, Z. He, Double-uncertainty weighted method for semi-supervised learning, in: International Conference on Medical Image Computing and Computer-Assisted Intervention, Springer, 2020, pp. 542–551.

Shanfu Lu received the M.S. degree from the Zhejiang normal university in 2018 with a major in Software Engineering. He is an algorithm engineer of clinical research branch, Perception Vision Medical Technologies Co Ltd. His research interests include medical AI, medical image processing, Data mining and computer vision.

Zijian Zhang received the M.S. degree from the University of South China in 2009 with a major in Nuclear Science and Technology. He is a Senior medical physicist of the Department of Onocology. National Clinical Research Center for Geriatric Disorders, Xiangya Hospital, Central South University. His research interests include medical AI, medical image processing, Data mining and computer vision.

Ziye Yan received the B.S. and M.S. degrees both from Harbin Institute of Technology in 2002 and 2005 respectively, and the Ph.D. degree from Beijing Institute of Technology in 2011. He is the professor-level engineer and the director of clinical research branch, Perception Vision Medical Technologies Co Ltd. His research interests include medical AI, medical image processing and information security.

Yiran Wang received the B.S. degree from the University of British Columbia in 2023 with a major in Mathematics and minor in Statistics. He is currently pursuing a career in actuarial science with research interests in computer vision and discrete optimization.

Tingting Cheng received the M.D. degrees from Xiangya Hospital, Central South University in 2020. She is an assistant professor from the Department of Onocology, National Clinical Research Center for Geriatric Disorders, Xiangya Hospital, Central South University. Her research interests focus on Multiomics and artificial intelligence.

Rongrong Zhou received the M.D. degrees from Xiangya Hospital, Central South University in 2010. She is a professor from the Department of Onocology, National Clinical Research Center for Geriatric Disorders, Xiangya Hospital, Central South University. Her research interests focus on clinical treatment of lung cancer and artificial intelligence.

Guang Yang (Senior Member, IEEE) received the M.Sc. degree in vision imaging and virtual environments from the Department of Computer Science, University College London, London, U.K., in 2006, and the Ph.D. degree in medical image analysis from the Centre for Medical Image Computing (CMIC), Department of Computer Science and Medical Physics, University College London, in 2012.,He is currently an Honorary Lecturer with the Neuroscience Research Centre, Cardiovascular and Cell Sciences Institute, St. George's, University of London, London. He is also an Image Processing Physicist and an Honorary Senior Research Fellow with the Cardiovascular Research Centre, Royal Brompton Hospital, Chelsea, London. He is also with the National Heart and Lung Institute, Imperial College London, London.,Dr. Yang is also a member of the International Society for Magnetic Resonance in Medicine (ISMRM) and the Society of Photo-Optical Instrumentation Engineers (SPIE). (Based on a document published on August 17, 2020.)