Computers in Biology and Medicine 143 (2022) 105252

Contents lists available at ScienceDirect

Computers in Biology and Medicine

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

Shape and boundary-aware multi-branch model for semi-supervised medical image segmentation

Xiaowei Liu [a], Yikun Hu [a], Jianguo Chen [b,*], Keqin Li [a,c]

[a] College of Computer Science and Electronic Engineering, Hunan University, Changsha, 410082, China
[b] Institute for Infocomm Research, Agency for Science, Technology and Research, 138 632, Singapore
[c] Department of Computer Science, State University of New York, New Paltz, NY, 12 561, USA

ARTICLE INFO

Keywords:
Boundary aware
Medical images segmentation
Multi-branch consistence
Semi-supervised learning

ABSTRACT

Supervised learning-based medical image segmentation solutions usually require sufficient labeled training data. Insufficient available labeled training data often leads to the limitations of model performances, such as over-fitting, low accuracy, and poor generalization ability. However, this dilemma may worsen in the field of medical image analysis. Medical image annotation is usually labor-intensive and professional work. In this work, we propose a novel shape and boundary-aware deep learning model for medical image segmentation based on semi-supervised learning. The model makes good use of labeled data and also enables unlabeled data to be well applied by using task consistency loss. Firstly, we adopt V-Net for Pixel-wise Segmentation Map (PSM) prediction and Signed Distance Map (SDM) regression. In addition, we multiply multi-scale features, extracted by Pyramid Pooling Module (PPM) from input X, with $2 - |SDM|$ to enhance the features around the boundary of the segmented target, and then feed them into the Feature Fusion Module (FFM) for fine segmentation. Besides boundary loss, the high-level semantics implied in SDM facilitate the accurate segmentation of boundary regions. Finally, we get the ultimate result by fusing coarse and boundary-enhanced features. Last but not least, to mine unlabeled training data, we impose consistency constraints on the three core outputs of the model, namely PSM1, SDM, and PSM3. Through extensive experiments over three representative but challenging medical image datasets (LA2018, BraTS2019, and ISIC2018) and comparisons with the existing representative methods, we validate the practicability and superiority of our model.

1. Introduction

Medical imaging can help doctors make a rapid diagnosis and clinical interventions based on the visual manifestation of organs, tissues, and lesions in medical images such as Computed Tomography (CT), X-ray, Ultrasound, and Magnetic Resonance Imaging (MRI) [1,2]. Visual segmentation of organs, tissues, and lesions is one of the basic technologies for automatic and intelligent analysis of medical images. In most cases, deep neural networks are trained in a supervised fashion, which requires sufficient labeled training data. However, manual labeling is an experience-oriented and time-consuming task, while unlabeled samples in the hospital are abundant. If we can make good use of these unlabeled data, intelligent analysis of medical images will make good progress. Therefore, even if new technologies or directions emerge in the future, Semi-Supervised Learning (SSL) is still one of the most promising fields of machine learning, especially in the application of medical image analysis.

In recent years, many medical image segmentation methods have emerged to cope with the limitations of available labeled training datasets. For instance, to reduce the workload of manual labeling, researchers have developed interactive segmentation techniques to select specific organs or lesions with less manual intervention [3–7]. This semi-automatic segmentation technology can effectively improve the efficiency of manual image annotation. Weakly supervised learning adopts image-level or object-level annotations with bounding boxes, instead of pixel-wise annotations that are more difficult to obtain [8,9]. The SSL approaches directly mine essential information from limited labeled data and a large amount of unlabeled data. Unlike fully supervised learning, the SSL methods can obtain better segmentation results by using training data with limited annotations. Our goal in this work is to improve the accuracy of fully automatic pixel-by-pixel segmentation when only limited labeled data are available, so weak supervision and

[*] Corresponding author.
E-mail addresses: liuxiaowei@hnu.edu.cn (X. Liu), yikunhu@hnu.edu.cn (Y. Hu), chen_jianguo@i2r.a-star.edu.sg (J. Chen), lik@newpaltz.edu (K. Li).

https://doi.org/10.1016/j.compbiomed.2022.105252
Received 29 September 2021; Received in revised form 4 January 2022; Accepted 20 January 2022
Available online 26 January 2022

interactive methods are not our choices.

In this work, we mainly concentrate on semi-supervised medical image segmentation methods, which are more effective and suitable for current actual needs. Only a handful of labeled data are available, and the rest are massive amounts of unlabeled data. In the early years, semi-supervised learning of medical image segmentation relied on the self-training and data augmentation strategies [10–13]. In the past two or three years, many SSL approaches have emerged that utilize unlabeled data by performing consistent regularization [14–16]. For example, in Ref. [14], Shuailin Li, Chuyu Zhang, and Xuming He developed a multi-task model to predict a Pixel-wise Segmentation Map (PSM) and the corresponding Signed Distance Map (SDM) according to the same input. They used an adversarial loss to enhance the consistency between the predicted SDMs of labeled and unlabeled samples, thereby ensuring more effective capture of shape-aware features. In Ref. [17], Yu et al. employed a mean teacher model [18] to mine consistency information from unlabeled samples. In Ref. [15], Li et al. introduced a transformation consistency to enhance the regularization effect of medical image segmentation models. For the same input, the output of the student and teacher models should be the same. Most of the existing methods promote the application of semi-supervised learning in medical image segmentation.

Inspired by the consistency of parallel dual-task [16] and boundary-aware [19], we propose a shape and boundary-aware multi-branch model for semi-supervised medical image segmentation. The core idea of the proposed method is to construct the consistency between two sequence-related tasks, namely the regularized SDM regression task and the pixel-level segmentation task. The SDM obtained from the previous branch captures redundant information such as the boundary and shape of the segmented object, and has substantial guiding significance for the subsequent pixel-level segmentation. The main contributions of this work are summarized as follows:

- We propose a shape and boundary-aware multi-branch model for medical image segmentation. First, this model utilizes a V-shaped model to predict coarse PSM and regress SDM, containing rich information about shape and boundary. What's more, through SDM and FFM, this model further extracts boundary-enhanced features. Finally, under the guidance of coarse PSM and boundary-enhanced features, we get the final segmentation of the targets.
- We exploit the task consistency among the three critical outputs on unlabeled samples. At the same time, through adversarial training, a large amount of unlabeled data is further utilized for model training.
- We conduct extensive experiments using 5-fold cross-validation on the BraTS2019, Atrial Segmentation 2018, and ISIC2018 datasets to evaluate the effectiveness of our model. Experimental results show that our model outperforms the state-of-the-art methods in terms of multiple evaluation metrics.

The rest of this paper is arranged as follows. The basic knowledge of semi-supervised medical image segmentation is reviewed in Section 2. Section 3 introduces the proposed shape and boundary-aware multi-branch model. Section 4 discusses the comparison experiments and results. Section 5 gives a general summary of the whole work.

2. Related work

We review the fundamental technologies involved in this work, such as semi-supervised medical image segmentation, consistency regularization, and signed distance map.

2.1. Semi-supervised medical image segmentation

The widespread popularity of Deep Learning (DL) has greatly promoted the rapid development of computing power, the availability of massive labeled samples, and the transition from manual feature extraction to automatic feature extraction. Thanks to the technological progress, the performance of semantic segmentation has been rapidly improved.

In recent years, some popular DL-bases semi-supervised methods have been proposed, such as self-training [10], co-training [20], adversarial learning [21,22], consistency regularization [18,23–26], and data augmentation-based methods [12]. The main idea of self-training is to train a pre-trained network on unlabeled data by using estimated labels (pseudo labels) in an iterative way. For instance, in Ref. [10], Bai et al. developed an iterative semi-supervised strategy for cardiac MRI image segmentation. The initial weights of the model come from training the network only on labeled samples. Here exit the next two steps: step 1 is to use the pre-trained model to predict segmentation map on unlabeled samples, and step 2 is to use the ground truths of labeled samples and estimated segmentation of unlabeled samples to update the network parameters. By iteratively executing the two steps, the model would be optimized to obtain better segmentation results.

Nowadays, combined with adversarial learning, the segmentation of medical images in a semi-supervised manner is also a popular research topic. In Ref. [21], Zhang et al. proposed a new deep adversarial network for biomedical image segmentation. The model contains two sub-networks: a generator for performing segmentation and a discriminator for evaluating the quality of segmentation. The generator encourages good segmentation results, and the discriminator decides whether the segmentation results come from unlabeled input or labeled input. In Ref. [14], Li et al. introduced an adversarial loss of semi-supervised learning to calculate the error between the predicted SDMs of labeled data and unlabeled data, where the SDMs contain rich geometric shape and boundary information. In Ref. [20], Peng et al. proposed a segmentation method based on the idea of model integration. Like the collaborative training method, the proposed method uses labeled data subsets for training and unlabeled data subsets for information exchange.

One way to solve the scarcity of labeled samples is to apply random geometry, color and intensity transformation, and interpolation strategies for data augmentation. Recently, many scholars choose Generative Adversarial Networks (GANs) to product samples as a supplement to model training [12,27]. For instance, in Ref. [12], Chaitanya et al. reported a creative data augmentation approach for learning using limited labeled training samples, where two conditional generative models are responsible for modeling shape and intensity characteristics, respectively. The holistic method is another way to alleviate the predicament of insufficient training samples for available markers, and is dedicated to integrating various SSL paradigms. For example, in Ref. [28], Berthelot et al. mixed consistency regularization, entropy minimization, and MixUp [29] together, and achieved significantly better performance.

2.2. Consistency regularization

In the field of computer vision, consistency regularization solutions, such as data-level consistency and task-level consistency, play a prominent role in self-supervised learning, unsupervised learning, and semi-supervised learning [16]. Data-level consistency encourages the assumption that decision boundaries maybe located in low-density areas where the predictions of the same input should be the same before and after the interference. Task-level consistency strives to ensure that the representations of similar tasks are equivalent. These representations can be uniquely converted to each other to reflect the consistency of tasks.

Most consistency regularization is at the data level. In Ref. [30], Antti Tarvainen and Harri Valpola implemented a Mean Teacher (MT) model to improve performances of SSL by using data-level consistency. In the MT model, they trained two sub-models to promote each other. Specifically, this consistency was embodied in that taking one sample as the input of the two sub-models, the outputs of the teacher and student sub-models should be consistent, even if different disturbances were attached to the two sub-networks [15]. Significantly, the student model updated the parameters by Exponential Moving Average (EMA) method, while the parameters of the student model depend on loss function optimization. In Ref. [15], Li et al. devised a semi-supervised method based on MT model and transformation consistency. The model encourages consistent predictions for two sub-networks with the same input under different data disturbances, including flipping, rotation, re-scaling, adding noise, etc. In Ref. [24], Ouali et al. described a Cross Consistency Training (CCT) model for semi-supervised semantic segmentation. The model consists of an encoder and multiple decoders, where the primary decoder uses labeled samples, and multiple secondary decoders use unlabeled samples for training. For all samples, the output of the primary decoder is consistent with the output of multiple secondary decoders using different perturbations.

Recently, Luo et al. proposed a task-level consistency method for semi-supervised medical image segmentation, and obtained different representation diversity of segmentation results, including signed distance maps and pixel-level segmentation maps [16]. In Ref. [31], Zamir et al. illustrated that cross-task consistent learning can make more accurate predictions and better generalize outliers, which is due to the invariance of inference paths on any task map. In Ref. [32], Navarro et al. adopted the idea of multi-task learning and constructed two additional sub-tasks, such as distance map regression and contour map detection. Unfortunately, they trained the model only in a fully-supervised manner. Inspired by the studies in Refs. [19,32–34], we propose an SSL model for medical image segmentation and use unlabeled data for consistency regularization and adversarial training to obtain a more robust model.

2.3. Signed distance map

By calculating the distance from each pixel marked 1 in the binary segmentation map to the nearest boundary pixel, we can get the distance map [35] of the binary segmentation mask, which gives rich and robust information about the boundary, size, shape, and position of the segmented objects. For a binary segmentation mask, the Signed Distance Map (SDM) is usually formulated as:

$$\varphi(x) = \begin{cases} 0, \ \ x \in \partial\Omega; \\ -\inf\limits_{y \in \partial\Omega} \|x - y\|_2, \ \ x \in \Omega, \Omega \neq \varnothing; \\ +\inf\limits_{y \in \partial\Omega} \|x - y\|_2, \ \ x \notin \Omega, \Omega \neq \varnothing; \\ 1, \Omega = \varnothing, \end{cases} \tag{1}$$

where $\Omega = \{x_i | y_i = 1, i \in \mathcal{S}\}$ is the pixel set of the foreground, x_i is the i-th point/pixel, y_i is the corresponding label, the index i traverses the entire input image or the corresponding segmentation mask, and $\mathcal{S}$ is the index set. At the same time, we use the symbol $\partial\Omega$ to represent the boundary pixel set. The middle two terms in Eq. (1) also need to be normalized to [-1,1] by using the mini-max normalization. Specifically, the absolute value of SDM indicates the distance from a pixel/voxel to the nearest pixel/voxel on the contour, and the symbol indicates the internal $(-)$ or external $(+)$ boundaries of the segmentation object. SDM is also a general definition of level set functions, so we name it $\varphi(x)$. Zero means that the point is exactly on the boundary. It is worth noting that when there is no foreground $(\Omega = \varnothing)$, each element in SDM should be 1. In other contexts, SDM should normalize to $[-1, 1]$ by using the mini-max normalization method.

3. Proposed approach

In this section, we propose a novel network for medical image segmentation. The model can learn image features from labeled and unlabeled data by using the corresponding losses.

3.1. Network architecture

As shown in Fig. 1, the overall framework includes a V-shaped backbone network for dual tasks, and a Pyramid Pooling Module (PPM) for extracting multi-scale features, and a Feature Fusion Module (FFM) for refining Pixel-level Segmentation Map (PSM). The dual tasks are completed by two branches, where branch A is for coarse PSM generation, and branch B is for Signed Distance Map (SDM) regression. First, branch A is in charge of PSM1 generation, and the intermediate product, $F1$ also guide the refinement of the final result PSM3. Second, branch B is responsible for the regression of SDM containing information such as boundaries and shapes. Third, the FFM module mines information from the boundary enhanced features, i.e., $(2 - |SDM|)*PPM(x)$, and outputs the boundary refined feature, $F2$. Then, PSM2 is obtained. Finally, $F1$ and $F2$ pass through a **Concatenation**, a **Convolution** and a **Sigmoid** function in turn, then we can acquire the ultimate PSM3. Essentially, it is a coarse-to-fine strategy.

3.1.1. Backbone network

The backbone network of the proposed model includes two components: the encoder and the decoder. The encoder uses multiple down-sampling layers to extract high-level features, while the decoder upsamples the features to recovery the size. As shown in Fig. 1, in the down-sampling stage, the input images pass through a series of convolutional layers to get high-level feature maps. Then, in the upsampling stage, bilinear interpolation or deconvolution is used to restore the feature scale step by step. Finally, there is a skip connection that combines the symmetric layer features of the encoder and decoder. In our experimental section, we will adopt a typical V-Net [36] structure in our model.

3.1.2. Coarse pixel-wise segmentation branch

As shown in Fig. 1, branch A is a coarse pixel-wise segmentation branch, the upward branch at the end of the backbone. In branch A, a convolutional operation is performed on the extracted features in the decoding stage of the backbone for adjusting the channels to 1 ($F1$), and then run a sigmoid to get a coarse PSM, called $PSM1$. In particular, the intermediate $F1$ will be conducive to the generation of final PSM, i.e., $PSM3$.

3.1.3. Pyramid Pooling Module

At the bottom of our proposed model as Fig. 1, Pyramid Pooling Module (PPM) is used to extract multi-scale features of the input image x. The pipeline of PPM is demonstrated as Fig. 2. This module takes a 2D image or 3D patch as input and consists of four parallel pooling branches, which pools inputs into the initial size of 1/2, 1/4, 1/8, and 1/16, respectively. After that, multi-scale feature maps can be obtained through **Convolution**, **Batch Normalization**, and **ReLu**, respectively. Finally, these features are upsampled to the original scale and spliced together to form a multi-scale feature pool.

3.1.4. SDM regression branch

Due to the area difference between the boundary and all other regions, it is difficult to directly obtain the accurate contour of the segmented object even using a cross-entropy loss. Therefore, as an alternative, we choose the SDM regression method. As shown in Fig. 1, branch B is to regress SDM, which implies redundant information, such as coarse contour, shape, and segmentation target size. Based on the PPM and predicted SDM, we can obtain the boundary enhanced features by $(2 - |SDM|) \times PPM(X)$, which strengthens the contour of the segmented object and does not alienate the regions far from the boundary. Simultaneously, $(1 - |SDM|) \times PPM(x)$ ignores features far from the boundary. In branch B, there are only two operations, including a **convolution** to adjust the channel to 1 and a **tanh** to regress the SDM.

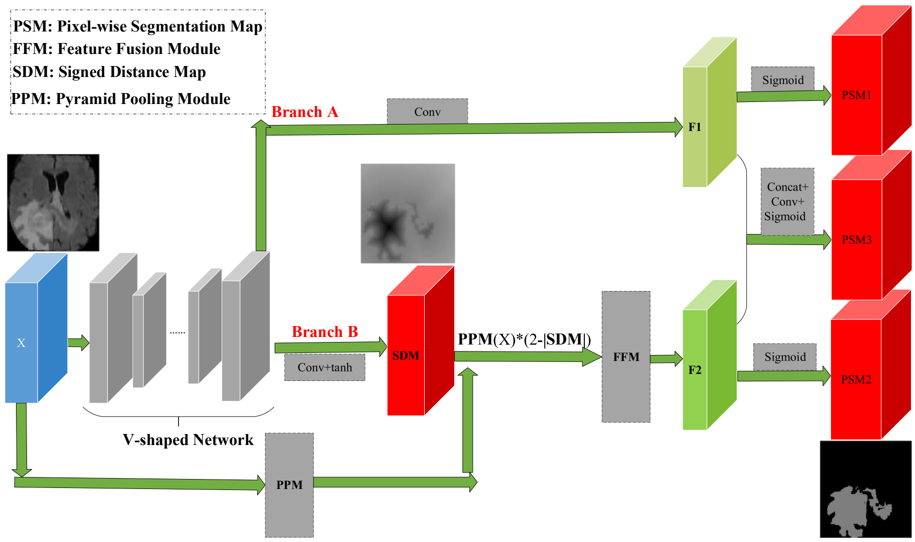

Fig. 1. Overview framework of the proposed shape and boundary-aware deep learning model for medical image segmentation based on semi-supervised learning. The model consists of a U-shaped or V-shaped backbone network for feature extraction, coarse pixel-wise segmentation branch (branch A) for coarse Pixel Segmentation Map (PSM) generation, Signed Distance Map (SDM) regression branch (branch B), Pyramid Pooling Module (PPM) for extracting multi-scale features and Feature Fusion Module (FFM) for PSM refinement.

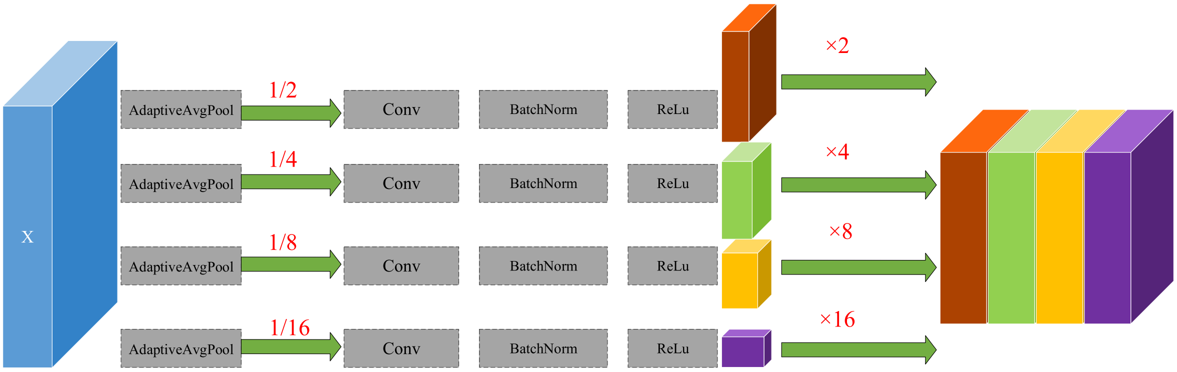

Fig. 2. Pyramid Pooling Module (PPM). This module includes several parallel pooling branches with different pooling parameters responsible for multi-scale feature extraction.

3.1.5. *Feature Fusion Module*

In our model, FFM is a channel spatial group module [37], which combines a spatial and a channel attention module. As shown in Fig. 3, FFM receives boundary-enhanced feature maps $X \in \mathbb{R}^{B\times C\times H\times W\times D}$, where B, C, H, W, D are the batch size, channel number, height, width and depth, respectively. Before grouping, we firstly employ “channel shuffle” to make different groups of information flow freely along the channel, which is inspired by ShuffleNet v2 [38]. Secondly, FFM divides X into g groups along the channel dimension, i.e., $X = [X_1, \ldots, X_g]$, $X_k \in \mathbb{R}^{B\times C/g\times H\times W\times D}$. Thirdly, the parallel attention module generates the

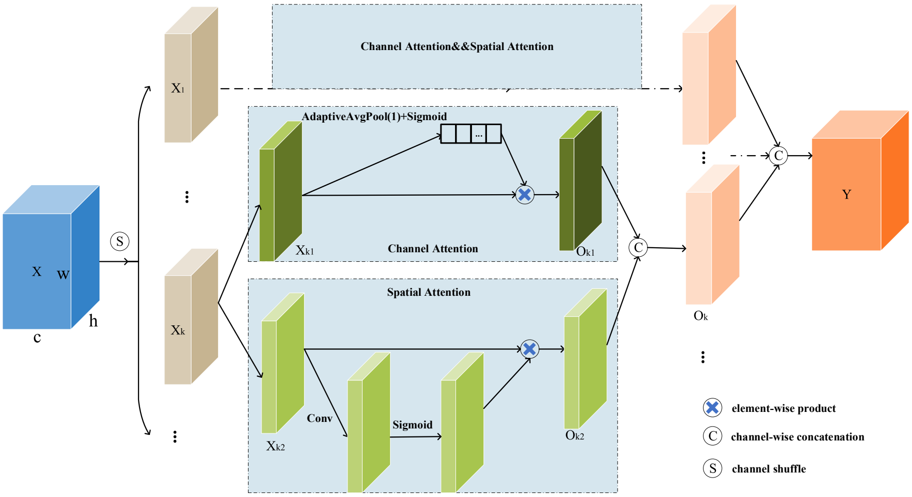

Fig. 3. Feature Fusion Module (FFM). This pipeline has four steps. First, FFM shuffles all feature maps along channels randomly. Second, it evenly splits shuffled features into g sub-groups. Third, it fuses those sub-features through spatial and channel attention blocks in parallel and finally aggregates these feature maps.

corresponding importance coefficients for sub-features as each group and fuses the sub-features themselves from both the channel and spatial views. Specifically, at the front of each attention block, the sub-features are equally divided into two partitions of equal channel sizes, i.e., $X_k = [X_{k1}, X_{k2}], X_{k1}, X_{k2} \in \mathbb{R}^{B\times C/2g\times H\times W\times D}, k = 1, 2, \ldots, g, k1 = 0, k2 = 1$. As shown in Fig. 3, one partition X_{k1} is sent into a channel attention block, while the other partition X_{k2} is fed to a spatial attention block. So, this is a simplified version of the dual attention mechanism. Channel Attention Block (CAB) utilizes a **AdaptiveAvgPool(1)** to pool X_{k1} to $\mathbb{R}^{B\times C/2g\times 1\times 1\times 1}$ and a **Sigmoid** to generate a channel attention map, which implies the importance of different channels. Spatial Attention Block (SAB) adopts a **Conv** to adjust X_{k2} from $\mathbb{R}^{B\times C/2g\times H\times W\times D}$ to $\mathbb{R}^{B\times 1\times H\times W\times D}$ and a **Sigmoid** to produce a spatial attention map, which implies the importance of different position. Then, the channel feature fusion result O_{k1} and the spatial feature fusion result O_{k2} are spliced together to form O_k, as the same size as X_k. And finally, all O_k are also stitched together to gain Y, as the same size as X.

3.2. Loss functions

Next, we will introduce each loss function used at all outputs in detail below.

3.2.1. SDM loss

Inspired by the work in Ref. [34], we use L_1 loss function in our model, which is the L_1 difference between the predicted value and the real SDM. The L_1 loss function can be written as:

$$\mathcal{L}_{L_1} = \|SDM_{pred} - SDM_{gt}\|_1, \tag{2}$$

where SDM_{pred} denotes the predicted SDM from branch B, and SDM_{gt} is from the binary mask of GT. Based on GT, we can acquire the SDM by using morphological methods. For multi-object segmentation, we can obtain the L_1 loss by adding all the L_1 losses corresponding to different binary segmentation maps. Although the L_1 loss is robust to outliers, it may causes the training process unstable. To overcome the disadvantage of the L_1 loss function, we further combine it with a product loss, as defined below:

$$\mathcal{L}_{\text{product}} = -\frac{1}{N}\sum\nolimits_{i=1}^{N}\frac{y_i p_i}{(y_i p_i + p_i{}^2 + y_i{}^2 + \varepsilon)}, \tag{3}$$

where N is the number of all pixels/voxels, y_i and p_i are the element of real SDM and predicted SDM, respectively, and ε is a small constant to prevent division by zero. Here, we set it to $1e-5$. We can find that Eq. (3) has the following properties:

$$\begin{cases} -\frac{1}{3} \le L_{product} < 0, & \text{if } y_i p_i > 0; \\ 0 < L_{product} \le 1, & \text{if } y_i p_i < 0; \\ L_{product} = 0, & \text{if } y_i = 0 \text{ or } p_i = 0, \end{cases} \tag{4}$$

where if y_i and p_i have the same sign, the smaller the gap between y_i and p_i, the smaller the loss. Especially if $y_i = p_i \neq 0$, the loss value will be $-1/3$. If y_i and p_i have different signs, the loss will be a positive value, which means the penalty is heavy right now.

Based on the above properties, we can conclude that the $\mathcal{L}_{product}$ enhances the perception of boundaries by emphasizing the correctness of the signs. Therefore, the final SDM loss is defined as:

$$\mathcal{L}_{\text{SDM}} = \mathcal{L}_{L_1} + \mathcal{L}_{product}. \tag{5}$$

SDM loss pays more attention to the elements close to zero, that is, the boundary represented by SDM. Therefore, SDM loss in our model can improve boundary perception.

3.2.2. PSM loss

There are many ways to calculate the Pixel-wise Segmentation Map (PSM) loss. Here, we hire Dice and Cross-Entroy losses, as defined below:

$$\mathcal{L}_{\text{PSM}} = \mathcal{L}_{\text{Dice}} + \mathcal{L}_{\text{CE}}, \tag{6}$$

where

$$\mathcal{L}_{\text{dice}} = C - \sum\nolimits_{i=1}^{C}\frac{2\sum y_{\text{i}} p_{\text{i}}}{\sum y_{\text{i}}^2 + \sum p_{\text{i}}^2 + \varepsilon}, \tag{7}$$

and

$$\mathcal{L}_{\text{CE}} = -\sum\nolimits_{i=1}^{\text{C}} y_i \log(p_i) \tag{8}$$

where C is the number of categories, p_i is the predicted value, and y_i is the corresponding ground truth. Since the model has three PSM outputs, the final PSM loss is summarized as:

$$\mathcal{L}_{\text{TotalPSM}} = \sum\nolimits_{i=1}^{3}\mathcal{L}_{PSMi}. \tag{9}$$

3.2.3. Boundary loss

To recognize the boundary more accurately, the model needs to pay more attention to the area near the contour of the segmented target. We hire the boundary loss [39] for that, defined as follows:

$$\mathcal{L}_B(\theta) = \int_\Omega \varphi_G(q) p_\theta(q) dq, \tag{10}$$

where $p_\theta(q)$ is the predicted probability map of the input q, θ indicates model parameters and φ_G stands for a real SDM. From Eq. (10), the boundary loss can be seen as a weighted summation of all elements of predicted probability maps (i.e., PSM), and φ_G acts as this weight.

The uncertainty of the segmentation task often focuses on the boundary, so it is unwise to pay too much attention to it. In Eq. (16), the boundary loss only plays a small role, and the weight β is set to 0.1. So, we apply this loss to PSM2 only.

3.2.4. Consistency loss

We use consistency loss to guarantee the consistency of SDM, PSM1, and PSM3 in Fig. 1. To reduce computational consumption and training instability, we don't take PSM2 into account. And our experiments confirm this programme. In Fig. 4, we show this consistency among the three outputs, i.e., SDM, PSM1, and PSM3.

First, to guarantee consistency between SDM and PSM1, i.e., C_{1-2} and C_{2-1}, we hire L2 loss, i.e, Eq. (11) as follow:

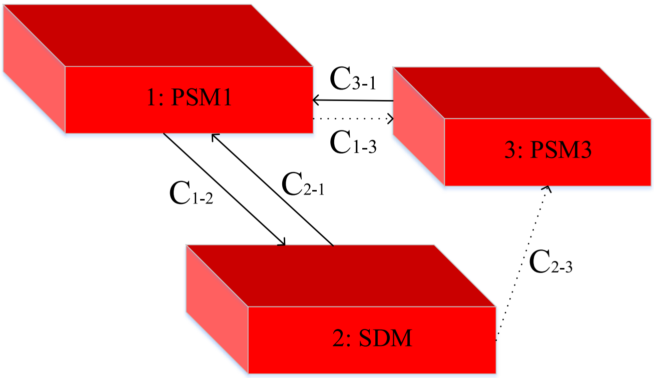

Fig. 4. Cycle consistency of the three outputs. We take advantage of the network structure itself and design consistency losses to guarantee the consistency of the three outputs. The direction of arrows implies the promotion of the former to the latter. The dotted line means that the network structure makes the ascension by itself, and the solid line indicates that it relies on the loss function.

$$\mathcal{L}_{c_1} = \|F(SDM) - PSM1\|_2, \tag{11}$$

where

$$F(x) = \frac{1}{1+e^{-k\cdot x}} = \sigma(k\cdot x) \ \ , \tag{12}$$

where k is a constant coefficient, which affects the slope of the curve. The larger the value of $|k|$, the steeper the slope, closer to a step function. The sign of k determines whether the curve is uphill or downhill. Inspired by the work in Ref. [34], we set k to -1500, which matches the normalized SDM defined in our work. As we can see, Eq. (12) convert the predicted SDM to PSM. Moreover, due to the differentiability of Eq. (12), bi-directional promotion, that is, C_{1-2} and C_{2-1}, can be realized by Eq. (11). Second, we use Eq. 13

$$\mathcal{L}_{c_2} = \|PSM3 - PSM1\|_2. \tag{13}$$

to relize C_{3-1}. That is, for the learning of unsupervised samples, PSM3 provides the learning objectives of PSM1. According to the design of network architecture, SDM and PSM1 guide the generation of PSM3, so it is no need to design additional loss functions to ensure the consistency of C_{1-3} and C_{2-3}.

Finally, the total consistency loss (Eq. (14)) is the sum of these two sub-consistency losses, as defined below:

$$\mathcal{L}_{cl} = \mathcal{L}_{c_1} + \mathcal{L}_{c_2}. \tag{14}$$

3.2.5. *Adversarial loss function*

Inspired by the work in Ref. [14], we introduce an adversarial loss to regularize model training further. To this end, we use a discriminator to distinguish whether the final predicted PSM3 comes from the labeled or unlabeled input. We assume that the quality of PSM3 in the supervised learning method is high, while the other parts in the unsupervised learning method are of low quality. As the training progresses, the discriminator will not discriminate whether the input data corresponding to the predicted PSM3 is from a subset of labeled or unlabeled data. From there, the predicted results of PSM3 based on the unlabeled subset are as good as predictions based on the labeled subset. It is worth noting that the adversarial loss increases more perturbation, making the model easier to approach the optimal solution of this problem.

Specifically, the adopted discriminator takes PSM3 and the corresponding image as input and predicts its probability from the labeled data subset. Given the discriminator D and segmentation network G, an adversarial loss on a batch of training data can be formulated as:

$$\begin{aligned}\min_{\theta}\max_{\zeta} \mathcal{L}_{adv}(\theta,\zeta) = &\frac{1}{N}\sum\nolimits_{n=1}^{N} \log D(\mathbf{X}_n, G(\mathbf{X}_n|\theta)_{PSM}|\zeta) \\ &+\frac{1}{N}\sum\nolimits_{m=N+1}^{2N} \log\big(1 - D(\mathbf{X}_m, G(\mathbf{X}_m|\theta)_{PSM}|\zeta)\big),\end{aligned} \tag{15}$$

where 2 N is the batch size, consisting of N labeled images and N unlabeled images. X_n and X_m are the labeled and unlabeled inputs, respectively. $D(\cdot|\theta)$ and $G(\cdot|\zeta)$ represent the discriminator and segmentation networks, where θ and ζ are the weight parameters of the networks. Since the adversarial loss imposes on PSM3, the output of $G(\mathbf{X}|\theta)_{PSM}$ is namely PSM3.

On the one hand, when $D(\cdot|\zeta)$ is fixed, sufficient training of $G(\cdot|\theta)$ will make $D(\cdot|\zeta)$ unable to determine whether the input is from labeled or unlabeled data. In Eq. (15), these two terms are minimized at that time. On the other hand, given a fixed generator (a segmentation network) $G(\cdot|\theta)$, the goal is to sufficiently train the discriminator and output 0 for the unlabeled input and 1 for the labeled input. In Eq. (15), these two terms are maximized at the same time.

3.2.6. *Hybrid loss function*

We linearly combine all of the above losses to form a hybrid loss function:

$$\mathcal{L}_{final} = (1-\gamma)(\mathcal{L}_{TotalPSM} + \alpha\mathcal{L}_{SDM}) + \gamma(\mathcal{L}_{cl} + \mathcal{L}_{adv} + \beta\mathcal{L}_{boundary}), \tag{16}$$

where α, β, and γ balance these different sub-losses. In comparison experiments, we will set α to 0.3 and β to 0.1, which is determined by the grid search and through past experience values. $\gamma(t)$ is defined as a function with the number of training steps as the independent variable, as shown in Eq. (17) below:

$$\gamma(t) = a\cdot\exp\left(-5\left(1-\frac{t}{t_{max}}\right)^2\right), \tag{17}$$

where t denotes the current training step, t_{max} represents the maximum training step, and a is the maximum weight, a constant coefficient greater than zero. Therefore, as the training progresses, the proportion of $\mathcal{L}_{cl} + \mathcal{L}_{adv} + \beta\mathcal{L}_{boundary}$ will increase with the increase of t, and will eventually be equal to a. In comparison experiments, we will set a to 1. Therefore, $\mathcal{L}_{final}$ can exploit both the labeled subset and the unlabeled subset to optimize all modules in a semi-supervised manner.

4. Experiments

4.1. *Dataset*

To evaluate our proposed model, we conduct comparison experiments on three different medical image datasets to complete the tasks of 3D brain tumor segmentation, 3D left ventricular segmentation and 2D skin lesion segmentation.

BraTS 2019[1]**:** BraTS 2019 focuses on the segmentation of brain tumors (gliomas) using preoperative MRI scans from multiple hospitals. Each scan has four modalities, namely T1, T1ce, T2, and Flair, while each tumor region is labeled as core, whole, and enhancement. For simplicity, we evaluate the segmentation effect of the whole tumor region only using a flair-modal scan instead of segmenting three subregions. In the first two rows of Fig. 5, we list a brain tumor case and the corresponding label from three perspectives (axial, sagittal, and coronal). The white part in the first row represents the whole tumor area, and the red region in the second row corresponds to the white part in the first row. We divide 335 original training samples into a training set (335 × 90% = 302) and a test set (335 × 10% = 33). The training subset is for 5-fold cross-validation, and the test subset is for visualization by fusing the generated five models. Note that each partition is hierarchical according to the ratio of HGG to LGG, which leads to stratified k-fold cross-validation.

Left Atrial 2018 (LA2018)[2]: This work involves a total of 154 3D MRIs with a resolution of 0.625 × 0.625 × 0.625 mm^3 from patients with atrial fibrillation. Since the test data is not labeled and the challenge is over, only training data (100 samples) is available. We divided them into two groups, one group holds 90% × 100 = 90 samples for five-fold cross-validation, and the other one holds 10%*100 = 10 samples for visualization, which is obtained by fusing five models (from 5-fold cross-validation). In Fig. 5, the middle two rows show an atrial sample (line 3) and the corresponding label (line 4).

ISIC 2018[3]**:** The ISIC 2018 challenge [40] contains three independent tasks: lesion segmentation, lesion attribute detection, and disease classification. Here, we only focus on the first task, namely lesion segmentation. In Fig. 5, the last row demonstrates a sample and the corresponding label. Only the training set containing 2594 images is for 5-fold cross-validation, and the validation set containing 100 images is for testing. All images are normalized based on the mean and variance of

[1] https://www.med.upenn.edu/cbica/brats-2019.

[2] http://atriaseg2018.cardiacatlas.org.

[3] https://challenge2018.isic-archive.com.

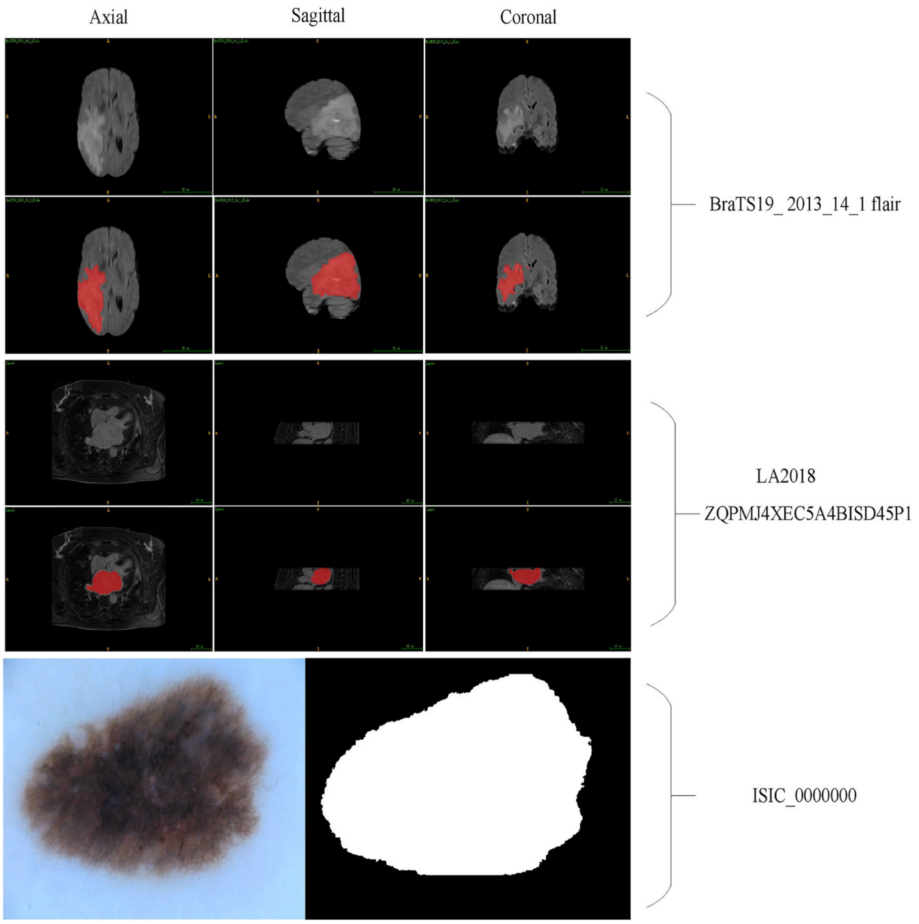

Fig. 5. Images and their segmentation results of three samples. The first two rows are from BraTS2019, which shows 2D slices from three views (axial, sagittal, and coronal) and the corresponding labels. The middle two rows are from LA2018, and the last row is from ISIC2018.

the data population and resized to 512 × 512.

Last but not least, we only report the performances of cross-validation on all datasets, which means that we exhibit the mean and standard deviation of each model on all training data. For visualization, we input each testing sample into all five models (obtained by 5-fold cross-validation) at the same time and calculate the average value of the five probability maps for the final results. In summary, the testing data is only for visual presentation.

4.2. *Implementation details and evaluation metrics*

In this work, we adopt the V-Net [36] as a backbone in all comparative experiments. We implement our model in PyTorch, and list all hardware and software environments in Table 1. Unlike other models, our model consists of an SDM regression branch B and a PSM branch A, where these two branches are attached to the end of the V-Net, as shown in Fig. 1. After the two branches, an FFM is used to refine the segmentation results. All models involved in the comparison are trained by an SGD optimizer for 20K iterations, with an initial learning rate 0.01 decayed by 10% every 5K iterations.

For 3D datasets, i.e., BraTS 2019 and LA2018, we obtain 3D patches by cropping each scan and normalizing these patches by using the Min-Max Normalization. The patch size of training and prediction is [96, 96, 96]. We set the labeled and unlabeled batch size to 3 and prediction stride to [64, 64, 64] along three axes. Before being fed into the network, we perform random rotation ([−20, 20]) and flipping operations on the fly. For the 2D dataset, i.e., ISIC 2018, we resize all images into 512 × 512 first and then normalize them by adopting Z-score Normalization. Similarly, we perform random rotation (±90°, ±180°*and* ± 270°) and flipping operations on the fly. The labeled and unlabeled batch size are set to 3 both.

Table 1
Software and hardware experimental environment configurations.

Software	OS	CUDA	Pytorch	V-Net
	Ubuntu18.4	11.1	1.7.1	[36]
Hardware	CPU	Memery	GPU	Video Memery
	Xeon(R)	62G	2 × Tesla V100	2 × 32G

We hire several highly correlated metrics for quantitative analysis, including Accuracy, Dice, Jaccard index, Average Surface Distance (ASD), and 95% Hausdorff Distance (95HD).

4.3. *Quantitative comparison*

We verify the effectiveness of the proposed model in comparison experiments from quantitative and qualitative perspectives. From a quantitative point of view, we design two groups of comparative experiments. One group verifies the positive effects of the proposed semi-supervised strategy, and the other group compares the proposed model with other methods. From a qualitative point of view, we discuss the segmentation results of the comparative methods, as shown in Fig. 6.

4.3.1. *Effectiveness of proposed model*

To justify the effectiveness of the proposed semi-supervised learning model, we conduct comparison experiments on LA 2018 by comparing a baseline and two variants of the proposed model. They are a V-Net model, a fully supervised version of our model, and a semi-supervised

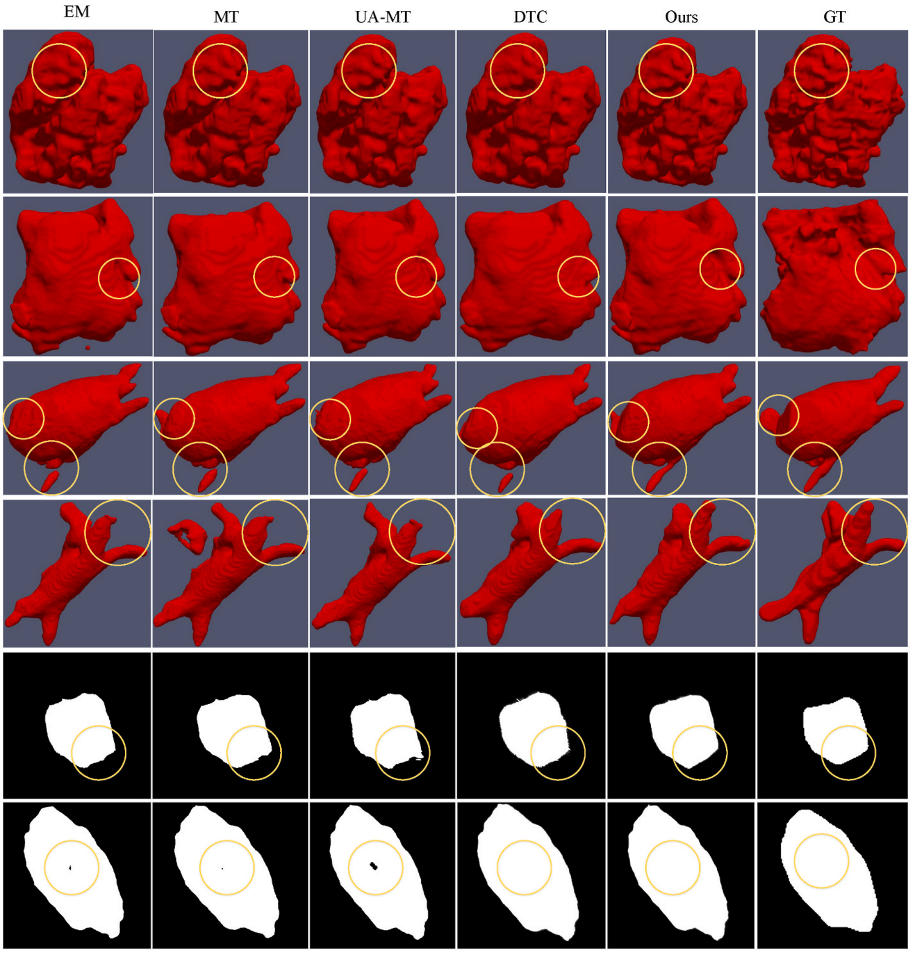

Fig. 6. Visualization of comparative methods. Each row refers to a case, and each column refers to a different method.

variant of our model, respectively. We report the performance of these three models in Table 2, where there are three groups of comparison experiments. The variable is the different proportion of labeled samples and unlabeled samples. In the last row, we apply all labeled data to train V-Net and get the best performance. The results confirm that the performance of the semi-supervised versions is always better than the corresponding fully-supervised variants. As the labeled data increases, their performance gap is narrowing. In particular, when the number of labeled samples is 20, the performance of our semi-supervised method is better than that of full supervision in Accuracy, Dice, and Jaccard, and it is slightly inferior in the two distance-based criteria (ASD and 95HD). It shows that when there is considerable labeled data, the role of semi-supervised learning is not necessarily positive totally. In short, thanks to the effective use of unlabeled data, the proposed semi-supervised model is better than the fully-supervised models when there is less labeled data.

4.3.2. *Comparison with semi-supervised methods*

We compare our model with four state-of-the-art semi-supervised methods, including the Mean Teacher (MT) model [30], Entropy Minimization (EM) model [41], Uncertainty-Aware Mean Teacher (UA-MT) model [17], and Dual-Task Consistency (DTC) model [16]. All comparative experiments were conducted on the training sub-sets from three data sets. We report the mean and standard deviation of every measure. All data in the table comes from 5-fold cross-validation.

4.3.2.1. Per0066ormance comparisons on BraTS2019. Firstly, we evaluate our proposed model on the BraTS2019. In each fold, about 240

Table 2
Segmentation performance of the comparative models on the LA2018 dataset using 5-fold cross validation.

Method	Scans used		Metrics				
	Labeled	Unlabeled	Accuracy	Dice	Jaccard	ASD [voxel]	95HD [voxel]
V-Net	5	0	0.962 4 ± 0.023 9	0.759 6 ± 0.176 7	0.643 3 ± 0.184 3	5.237 1 ± 6.608 5	22.034 5 ± 16.336 6
Ours	5	0	0.964 6 ± 0.024 5	0.762 2 ± 0.213 2	0.651 5 ± 0.210 6	3.600 4 ± 4.733 0	19.364 7 ± 21.286 5
Ours	5	67	**0.967 7 ± 0.019 0**	**0.806 6 ± 0.130 6**	**0.692 0 ± 0.149 1**	**3.445 8 ± 1.927 3**	**14.253 0 ± 9.210 0**
V-Net	10	0	0.971 4 ± 0.014 6	0.830 1 ± 0.116 3	0.722 2 ± 0.130 1	2.917 3 ± 2.251 0	18.556 6 ± 14.747 4
Ours	10	0	0.972 3 ± 0.020 0	0.835 8 ± 0.172 2	0.738 2 ± 0.173 0	2.431 0 ± 1.625 0	13.545 8 ± 19.248 9
Ours	10	62	**0.976 3 ± 0.014 7**	**0.867 1 ± 0.695 2**	**0.772 0 ± 0.097 9**	**2.426 6 ± 1.626 6**	**10.025 2 ± 6.668 0**
V-Net	20	0	0.977 2 ± 0.010 1	0.870 1 ± 0.057 5	0.775 8 ± 0.078 5	2.195 7 ± 0.881 2	14.317 7 ± 11.235 3
Ours	20	0	0.978 9 ± 0.012 2	0.874 2 ± 0.101 9	0.786 8 ± 0.114 5	**1.930 0 ± 0.897 4**	**8.410 2 ± 9.934 1**
Ours	20	52	**0.979 8 ± 0.007 9**	**0.884 9 ± 0.042 6**	**0.795 9 ± 0.063 3**	2.560 3 ± 1.723 7	9.773 1 ± 7.441 8
V-Net	72	0	0.981 4 ± 0.009 2	0.891 7 ± 0.070 7	0.810 1 ± 0.088 0	1.839 9 ± 1.141 5	6.554 0 ± 4.176 8

samples are prepared for model training and 62 samples for validation. The first three rows of Table 3 lists the performances of V-Net under fully supervised settings (with 34, 68, and 240 labeled samples) as the reference. Apart from these, there are two groups of semi-supervised models, where the first group uses 34 labeled samples and the second increases to 68. Compared with fully supervised V-Net trained with only 34 labeled samples, all semi-supervised approaches improve the segmentation performance significantly with the blessing of the remaining unlabeled samples. From rows 4–13, UA-MT is slightly better than MT on most metrics, and the uncertainty map plays a good role here. DTC model has good advantages in Accuracy, but it is not satisfactory in other indicators. When there are 34 labeled samples and 206 unlabeled samples, our model ranks first on Dice, Jaccard, and 95HD, and second on Accuracy. When using 68 labeled samples and 172 unlabeled samples, our model ranks first on Accuracy, Dice, Jaccard, and 95HD, but fourth on ASD.

4.3.2.2. Performance comparisons on LA2018. We further verify our model on the LA2018 dataset. A quantitative comparison is shown in Table 4. We can see that no matter how much labeled data, our approach achieves the best performance than other approaches on almost all evaluation metrics. Surprisingly, UA-MT performs poorly, while the MT model performs very well. DTC is almost the worst when the number of labeled samples is 20.

4.3.2.3. Performance comparisons on ISIC2018. Finally, we validate our method on a 2D medical image set, ISIC 2018, and report the results in Table 5. Based on this data set, UA-MT is better than MT in Accuracy, Dice, and Jaccard but worse in ASD and 95HD. DTC performs very well when the number of available labeled samples increases to 580. When 290 labeled samples are available, our method is best than all the other models on all metrics. When 580 labeled data are available, it is still best on Accuracy, Dice, Jaccard, and 95HD, but not very poor on ASD. Therefore, our framework achieves better performance than other frameworks, almost on all evaluation indicators.

4.3.3. *Visualization comparison*

The quantitative comparison and analysis in the previous sections have verified the effectiveness of our model on three commonly used data sets. We utilize test subsets for visualization. Fig. 6 shows a total of six visualization results, of which two results are selected for display in each dataset. We acquire each result by taking the average value of the probability maps of the five models obtained from 5-fold cross-validation. Specifically, the first two cases (row 1–2) are from BraTS2019, the middel two cases (row 3–4) are from LA2018, and the last two cases (row 5–6) are from ISIC2018. For the details of the circle part, our method is closer to the ground truth. For example, in the last row, the results of columns 1–3 are significantly worse than DTC's and our results. Our results are closer to GT because of the thin and long characteristics of the segmented region. Compared with other methods, our results possess a higher overlapping degree with GT, fewer false positives, and more details visually. All of them further reveal the effectiveness, generalization, and robustness of our proposed approach.

4.4. *Ablation study*

From Eq. (16), we can know that the loss function is composed of five parts: $\mathcal{L}_{SDM}$, $\mathcal{L}_{TotalPSM}$, $\mathcal{L}_{boundary}$, $\mathcal{L}_{cl}$, and $\mathcal{L}_{adv}$. To better demonstrate the importance of each loss part, we conduct ablation experiments on these parts. We respectively use 5 and 20 labeled samples to train our atrial segmentation model. We compare four different combinations of the loss combinations: (1) $\mathcal{L}_{TotalPSM} + \alpha\mathcal{L}_{SDM}$; (2) $(1-\gamma(t))(\mathcal{L}_{TotalPSM} + \alpha\mathcal{L}_{SDM}) + \gamma(t)\beta\mathcal{L}_{boundary}$; 3) $(1-\gamma(t))(\mathcal{L}_{TotalPSM} + \alpha\mathcal{L}_{SDM}) + \gamma(t)(\beta\mathcal{L}_{boundary} + \mathcal{L}_{cl})$; and (4) $(1-\gamma(t))(\mathcal{L}_{TotalPSM} + \alpha\mathcal{L}_{SDM}) + \gamma(t)(\beta\mathcal{L}_{boundary} + \mathcal{L}_{cl} + \mathcal{L}_{adv})$.

Here, all hyper-parameter configurations of our model are the same as previous comparative experiments on LA2018, i.e, $\alpha = 0.3$, $\beta = 0.1$. For $\mathcal{L}_{TotalPSM}$, $\mathcal{L}_{boundary}$ and $\mathcal{L}_{SDM}$, only labeled are used to train the model. $\mathcal{L}_{cl}$ uses samples without ground truth, while $\mathcal{L}_{adv}$ requires labeled samples too.

Table 6 reports the performance of the four variants with different losses. In Case 1, these two-loss combinations apply to an SDM output and three PSMs. In the first and second cases, only labeled data is used. In Case 2, The growing weight of boundary loss enables the network to pay more attention to the boundary and decrease 95HD. In this way, the boundary loss improves the ability to recognize the boundary area, especially the boundary line. In Case 3, the segmentation performance of the model is improved with the help of consistency loss, which is mainly due to the use of unlabeled data. Especially, when the number of labeled samples is 20, semi-supervised learning still leads to the decline of 95HD. In Case 4, the adversarial loss enhances the robustness of the model and slightly increases the Dice values. Therefore, we can conclude that every part of the loss function in our model is essential and effective for model performance.

5. Conclusion

In this paper, we proposed a semi-supervised method for medical image segmentation based on both labeled samples and unlabeled samples. The proposed method consists of a backbone network, an SDM regression branch, a Pyramid Pooling Module (PPM), and a Feature Fusion Module (FFM). First, we adopted a dual-task learning strategy to obtain a PSM and the corresponding SDM. Second, with the help of PPM and FFM, the model is conducive to accurately acquiring boundary regions by mining valuable information from $(2 - |SDM| \times PPM(X))$, where X denotes the input. Meanwhile, we utilized boundary loss to enhance that in the boundary area. Third, the model fuses the coarse

Table 3
Quantitative comparison on the BraTS2019 dataset using 5-fold validation.

Method	Scans used		Metrics				
	Labeled	Unlabeled	Accuracy	Dice	Jaccard	ASD [voxel]	95HD [voxel]
V-Net	34	0	0.986 6 ± 0.057 8	0.815 7 ± 0.163 6	0.713 8 ± 0.184 4	2.163 1 ± 2.052 5	14.480 ± 18.538 8
V-Net	68	0	0.992 7 ± 0.006 4	0.852 3 ± 0.129 0	0.759 9 ± 0.156 5	2.152 7 ± 4.939 8	10.228 ± 16.879 3
V-Net	240	0	0.994 1 ± 0.004 6	0.883 6 ± 0.094 0	0.801 6 ± 0.122 2	1.810 0 ± 4.951 4	7.963 1 ± 14.209 3
MT	34	206	0.985 5 ± 0.080 8	0.836 6 ± 0.153 4	0.742 1 ± 0.177 8	2.012 7 ± 2.042 2	9.671 3 ± 14.138 3
EM	34	206	0.985 6 ± 0.080 8	0.837 0 ± 0.152 4	0.742 4 ± 0.177 2	1.999 7 ± 1.876 4	10.369 9 ± 15.349 7
UA-MT	34	206	0.989 0 ± 0.057 4	0.839 4 ± 0.146 1	0.744 4 ± 0.171 7	**1.891 5 ± 1.730 8**	10.791 2 ± 15.797 3
DTC	34	206	**0.991 6 ± 0.008 9**	0.834 3 ± 0.164 3	0.741 3 ± 0.185 8	2.917 2 ± 9.666 1	9.319 5 ± 14.633 6
Ours	34	206	0.989 2 ± 0.057 5	**0.847 2 ± 0.141 2**	**0.755 4 ± 0.171 1**	2.265 8 ± 4.163 5	**8.536 3 ± 12.216 8**
MT	68	172	0.989 6 ± 0.057 4	0.859 6 ± 0.121 0	0.769 2 ± 0.151 6	1.712 3 ± 1.328 5	9.785 3 ± 15.995 3
EM	68	172	0.989 9 ± 0.057 3	0.862 4 ± 0.117 9	0.773 1 ± 0.148 2	1.706 0 ± 1.363 4	8.815 0 ± 12.973 2
UA-MT	68	172	0.989 7 ± 0.057 3	0.860 8 ± 0.114 5	0.770 1 ± 0.145 7	**1.664 2 ± 1.127 5**	10.246 0 ± 15.745 7
DTC	68	172	0.992 6 ± 0.008 1	0.862 1 ± 0.155 5	0.772 9 ± 0.180 2	4.072 7 ± 7.014 1	9.555 8 ± 12.230 4
Ours	68	172	**0.993 1 ± 0.006 6**	**0.863 3 ± 0.120 1**	**0.774 8 ± 0.148 7**	2.203 3 ± 4.688 2	**8.070 3 ± 12.387 3**

Table 4
Quantitative comparison on the LA2018 dataset using 5-fold cross validation.

Method	Scans used		Metrics				
	Labeled	Unlabeled	Accuracy	Dice	Jaccard	ASD [voxel]	95HD [voxel]
V-Net	10	0	0.971 4 ± 0.014 6	0.830 1 ± 0.116 3	0.722 2 ± 0.130 1	2.917 3 ± 2.251 0	18.556 6 ± 14.747 4
V-Net	20	0	0.977 2 ± 0.010 1	0.870 1 ± 0.057 5	0.775 8 ± 0.078 5	2.195 7 ± 0.881 2	14.317 7 ± 11.235 3
V-Net	72	0	0.981 4 ± 0.009 2	0.891 7 ± 0.070 7	0.810 1 ± 0.088 0	1.839 9 ± 1.141 5	6.554 0 ± 4.176 8
MT	10	62	0.974 5 ± 0.013 7	0.858 2 ± 0.070 7	0.757 5 ± 0.095 0	**2.149 0 ± 0.794 8**	16.821 0 ± 13.814 3
EM	10	62	0.971 6 ± 0.014 8	0.845 6 ± 0.070 9	0.738 5 ± 0.097 8	2.150 4 ± 0.870 4	18.617 9 ± 13.302 9
UA-MT	10	62	0.970 9 ± 0.015 4	0.840 0 ± 0.082 7	0.731 8 ± 0.108 4	2.300 0 ± 1.085 5	17.746 1 ± 13.142 2
DTC	10	62	0.973 9 ± 0.015 4	0.854 8 ± 0.072 7	0.752 6 ± 0.098 2	3.853 4 ± 3.180 7	14.699 4 ± 11.657 1
Ours	10	62	**0.974 7 ± 0.014 3**	**0.867 1 ± 0.095 2**	**0.772 0 ± 0.097 9**	2.426 6 ± 1.626 6	**10.025 2 ± 6.668 0**
MT	20	52	0.978 5 ± 0.010 2	0.881 8 ± 0.047 0	0.791 4 ± 0.069 0	1.881 8 ± 0.746 3	12.771 3 ± 11.588 4
EM	20	52	0.977 8 ± 0.010 0	0.877 3 ± 0.049	0.784 7 ± 0.072 8	1.929 5 ± 0.748 9	14.350 7 ± 12.595 8
UA-MT	20	52	0.977 4 ± 0.009 3	0.876 0 ± 0.044 0	0.782 0 ± 0.067 0	**1.840 9 ± 0.561 9**	14.068 1 ± 12.086 2
DTC	20	52	0.976 3 ± 0.014 2	0.870 8 ± 0.061 2	0.775 7 ± 0.083 0	3.644 0 ± 3.131 3	13.164 3 ± 11.486 6
Ours	20	52	**0.979 8 ± 0.007 9**	**0.884 9 ± 0.042 6**	**0.795 9 ± 0.063 3**	2.560 3 ± 1.723 7	**9.773 1 ± 7.441 8**

Table 5
Quantitative comparison on the ISIC2018 dataset using 5-fold cross validation.

Method	Scans used		Metrics				
	Labeled	Unlabeled	Accuracy	Dice	Jaccard	ASD [pixel]	95HD [pixel]
V-Net	290	0	0.909 9 ± 0.141 9	0.760 8 ± 0.253 5	0.668 3 ± 0.270 0	34.399 7 ± 38.628 9	78.903 0 ± 76.666 5
V-Net	580	0	0.924 0 ± 0.127 2	0.801 2 ± 0.227 8	0.714 7 ± 0.248 1	26.473 3 ± 33.359 3	61.994 4 ± 68.585 1
V-Net	2075	0	0.933 5 ± 0.124 0	0.836 6 ± 0.189 0	0.753 0 ± 0.212 7	21.440 6 ± 26.482 4	53.690 2 ± 60.333 8
MT	290	1785	0.912 8 ± 0.146 7	0.776 9 ± 0.253 2	0.690 0 ± 0.269 3	31.309 9 ± 39.327 5	71.853 7 ± 77.637 9
EM	290	1785	0.910 0 ± 0.156 3	0.772 2 ± 0.255 7	0.684 3 ± 0.270 1	29.015 8 ± 35.927 0	67.522 9 ± 71.654 6
UA-MT	290	1785	0.913 2 ± 0.144 0	0.784 8 ± 0.241 7	0.696 6 ± 0.259 2	31.000 5 ± 38.198 0	72.261 1 ± 74.705 5
DTC	290	1785	0.903 9 ± 0.150 0	0.771 0 ± 0.253 1	0.682 8 ± 0.273 2	38.632 8 ± 47.284 3	82.625 8 ± 84.708 5
Ours	290	1785	**0.920 7 ± 0.132 3**	**0.792 3 ± 0.238 7**	**0.706 2 ± 0.258 0**	**28.894 0 ± 37.279 9**	**67.248 5 ± 74.968 6**
MT	580	1495	0.929 0 ± 0.131 1	0.816 3 ± 0.219 3	0.733 2 ± 0.239 3	**22.898 3 ± 30.427 0**	55.233 8 ± 64.468 1
EM	580	1495	0.927 7 ± 0.131 0	0.813 1 ± 0.217 7	0.728 1 ± 0.239 3	23.780 2 ± 30.582 6	57.864 3 ± 65.485 1
UA-MT	580	1495	0.929 8 ± 0.120 4	0.816 3 ± 0.214 5	0.731 9 ± 0.236 8	24.969 5 ± 31.909 7	59.364 1 ± 65.990 1
DTC	580	1495	0.931 2 ± 0.114 9	0.821 2 ± 0.210 0	0.737 7 ± 0.233 7	27.124 8 ± 35.613 2	62.429 8 ± 70.662 0
Ours	580	1495	**0.931 8 ± 0.125 8**	**0.821 8 ± 0.215 9**	**0.740 1 ± 0.236 7**	23.187 6 ± 31.218 5	**54.788 1 ± 64.094 6**

Table 6
Ablation Study of combinations of different losses on LA2018.

Different loss combination	Scans used		Metrics	
	Labeled	Unlabeled	Dice	95HD
$\mathcal{L}_{TotalPSM} + \alpha\mathcal{L}_{SDM}$	5	0	0.761 5	20.364 7
$(1-\gamma(t))(\mathcal{L}_{TotalPSM} + \alpha\mathcal{L}_{SDM}) + \gamma(t)\beta\mathcal{L}_{boundary}$	5	0	0.763 2	18.661 2
$(1-\gamma(t))(\mathcal{L}_{TotalPSM} + \alpha\mathcal{L}_{SDM}) + \gamma(t)(\beta\mathcal{L}_{boundary} + \mathcal{L}_{cl})$	5	67	0.798 9	14.352 7
$(1-\gamma(t))(\mathcal{L}_{TotalPSM} + \alpha\mathcal{L}_{SDM}) + \gamma(t)(\beta\mathcal{L}_{boundary} + \mathcal{L}_{cl} + \mathcal{L}_{adv})$	5	67	0.806 6	14.253 0
$\mathcal{L}_{TotalPSM} + \alpha\mathcal{L}_{SDM}$	20	0	0.874 5	9.314 7
$(1-\gamma(t))(\mathcal{L}_{TotalPSM} + \alpha\mathcal{L}_{SDM}) + \gamma(t)\beta\mathcal{L}_{boundary}$	20	0	0.874 2	8.410 2
$(1-\gamma(t))(\mathcal{L}_{TotalPSM} + \alpha\mathcal{L}_{SDM}) + \gamma(t)(\beta\mathcal{L}_{boundary} + \mathcal{L}_{cl})$	20	52	0.882 5	9.744 2
$(1-\gamma(t))(\mathcal{L}_{TotalPSM} + \alpha\mathcal{L}_{SDM}) + \gamma(t)(\beta\mathcal{L}_{boundary} + \mathcal{L}_{cl} + \mathcal{L}_{adv})$	20	52	0.884 9	9.773 1

PSM and boundary-enhanced features to obtain the final result. Forth, we use consistency loss and adversarial loss to mine the knowledge of unlabeled samples, where these two loss functions can be imposed on the unlabeled and labeled samples. Finally, extensive experiments on BraTS 2019, LA 2018, and ISIC 2018 are given to validate our model by comparing the proposed model with multiple recent typical approaches. Experimental results show that our method achieves good results in most evaluation criteria. It can highlight the boundaries and shapes of the target organs and lesion regions. In conclusion, our method can fully use a large pool of unlabeled data and comparatively few labeled data and has broad prospects in the field of automatic medical image segmentation.

In our future work, we will explore different deep learning models of medical image segmentation and apply to different organs and tissues. We will use semantic context unraveling strategy to roughly segment multiple organs and tissues. We also use the transformer methods to capture the overall dependence between organs or tissues.

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

Thank the reviewers for their hard work and put forward many constructive suggestions to improve the quality of this manuscript. Thank Hunan Malanshan Computing Media Research Institute for providing Tesla V100. This work was supported in part by the National Key R&D Program of China under Grant 2018YFB1003401, the National Natural Science Foundation of China under Grant 62 002 110, and the International Postdoctoral Exchange Fellowship Program (Grant No. 20180024).

References

[1] Bin Pu, Ningbo Zhu, Kenli Li, Shengli Li, Fetal cardiac cycle detection in multi-resource echocardiograms using hybrid classification framework, Future Generat. Comput. Syst. 115 (2021) 825–836.

[2] Laifa Ma, Xiao Zheng, Kenli Li, Shengli Li, Jianlin Li, Xiaoping Yi, Game theoretic interpretability for learning based preoperative gliomas grading, Future Generat. Comput. Syst. 112 (2020) 1–10.

[3] Jianguo Chen, Kenli Li, Zhaolei Zhang, Keqin Li, S Yu Philip, A survey on applications of artificial intelligence in fighting against covid-19, ACM Comput. Surv. 54 (8) (2021) 1–32.

[4] Guotai Wang, Wenqi Li, Maria A. Zuluaga, Rosalind Pratt, Premal A. Patel, Michael Aertsen, Doel Tom, L David Anna, Deprest Jan, Sébastien Ourselin, et al., Interactive medical image segmentation using deep learning with image-specific fine tuning, IEEE Trans. Med. Imag. 37 (7) (2018) 1562–1573.

[5] Zian Wang, David Acuna, Huan Ling, Amlan Kar, Sanja Fidler, Object instance annotation with deep extreme level set evolution, in: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, CVPR), 2019, pp. 7500–7508.

[6] Helena Williams, João Pedrosa, Laura Cattani, Susanne Housmans, Tom Vercauteren, Deprest Jan, D'hooge Jan, Interactive segmentation via deep learning and b-spline explicit active surfaces, in: International Conference on Medical Image Computing and Computer-Assisted Intervention, Springer, 2021, pp. 315–325.

[7] Hao Yuying, Yi Liu, Zewu Wu, Lin Han, Yizhou Chen, Guowei Chen, Lutao Chu, Shiyu Tang, Zhiliang Yu, Zeyu Chen, et al., Edgeflow: achieving practical interactive segmentation with edge-guided flow, in: Proceedings of the IEEE/CVF International Conference on Computer Vision, 2021, pp. 1551–1560.

[8] Neeraj Kumar, Ruchika Verma, Sanuj Sharma, Surabhi Bhargava, Abhishek Vahadane, Sethi Amit, A dataset and a technique for generalized nuclear segmentation for computational pathology, IEEE Trans. Med. Imag. 36 (7) (2017) 1550–1560.

[9] Zhi-Hua Zhou, A brief introduction to weakly supervised learning, Natl. Sci. Rev. 5 (1) (2018) 44–53.

[10] Wenjia Bai, Ozan Oktay, Matthew Sinclair, Hideaki Suzuki, Martin Rajchl, Giacomo Tarroni, Ben Glocker, Andrew King, Paul M. Matthews, Daniel Rueckert, Semi-supervised learning for network-based cardiac mr image segmentation, in: International Conference on Medical Image Computing and Computer-Assisted Intervention, Springer, 2017, pp. 253–260.

[11] Chaitanya Krishna, Neerav Karani, Christian F. Baumgartner, Anton Becker, Olivio Donati, Ender Konukoglu, Semi-supervised and task-driven data augmentation, in: International Conference on Information Processing in Medical Imaging, Springer, 2019, pp. 29–41.

[12] Chaitanya Krishna, Neerav Karani, Christian F. Baumgartner, Ertunc Erdil, Anton Becker, Olivio Donati, Ender Konukoglu, Semi-supervised task-driven data augmentation for medical image segmentation, Med. Image Anal. 68 (2021) 101934.

[13] Jianguo Chen, Kenli Li, Kashif Bilal, Keqin Li, S Yu Philip, et al., A bi-layered parallel training architecture for large-scale convolutional neural networks, IEEE Trans. Parallel Distr. Syst. 30 (5) (2018) 965–976.

[14] Shuailin Li, Chuyu Zhang, Xuming 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.

[15] Xiaomeng Li, Lequan Yu, Hao Chen, Chi-Wing Fu, Lei Xing, Pheng-Ann Heng, Transformation-consistent self-ensembling model for semi-supervised medical image segmentation, IEEE Transact. Neural Networks Learn. Syst. (2020).

[16] Xiangde Luo, Jieneng Chen, Tao Song, Guotai Wang, Semi-supervised Medical Image Segmentation through Dual-Task Consistency, AAAI Conference on Artificial Intelligence, 2021.

[17] Lequan Yu, Shujun Wang, Xiaomeng Li, Chi-Wing Fu, Pheng-Ann 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.

[18] Viktor Olsson, Wilhelm Tranheden, Juliano Pinto, Lennart Svensson, Classmix: segmentation-based data augmentation for semi-supervised learning, in: Proceedings of the IEEE/CVF Winter Conference on Applications of Computer Vision, 2021, pp. 1369–1378.

[19] Mengyang Feng, Huchuan Lu, Errui Ding, Attentive feedback network for boundary-aware salient object detection, in: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, CVPR), 2019, pp. 1623–1632.

[20] Jizong Peng, Guillermo Estrada, Marco Pedersoli, Christian Desrosiers, Deep co-training for semi-supervised image segmentation, Pattern Recogn. 107 (2020) 107269.

[21] Yizhe Zhang, Lin Yang, Jianxu Chen, Maridel Fredericksen, David P. Hughes, Danny 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.

[22] Nasim Souly, Concetto Spampinato, Mubarak Shah, Semi-supervised semantic segmentation using generative adversarial network, in: Proceedings of the IEEE International Conference on Computer Vision, ICCV), 2017, pp. 5688–5696.

[23] Gerda Bortsova, Florian Dubost, Laurens Hogeweg, Ioannis Katramados, Marleen de Bruijne, Semi-supervised medical image segmentation via learning consistency under transformations, in: International Conference on Medical Image Computing and Computer-Assisted Intervention, Springer, 2019, pp. 810–818.

[24] Yassine Ouali, Céline Hudelot, Myriam Tami, Semi-supervised semantic segmentation with cross-consistency training, in: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, CVPR), 2020, pp. 12674–12684.

[25] Xuyang Cao, Houjin Chen, Yanfeng Li, Yahui Peng, Shu Wang, Lin Cheng, Uncertainty aware temporal-ensembling model for semi-supervised abus mass segmentation, IEEE Trans. Med. Imag. 40 (1) (2020) 431–443.

[26] Srinivas Parthasarathy, Carlos Busso, Semi-supervised speech emotion recognition with ladder networks, IEEE/ACM Trans. Audio, Speech, Lang. Process. 28 (2020) 2697–2709.

[27] Ian Goodfellow, Jean Pouget-Abadie, Mehdi Mirza, Bing Xu, David Warde-Farley, Sherjil Ozair, Aaron Courville, Yoshua Bengio, Generative adversarial nets, Adv. Neural Inf. Process. Syst. 27 (2014).

[28] David Berthelot, Nicholas Carlini, Ian Goodfellow, Nicolas Papernot, Avital Oliver, Colin Raffel, Mixmatch: A Holistic Approach to Semi-supervised Learning, 2019 arXiv preprint arXiv:1905.02249.

[29] Hongyi Zhang, Moustapha Cisse, Yann N. Dauphin, David Lopez-Paz, Mixup: beyond Empirical Risk Minimization, 2017 arXiv preprint arXiv:1710.09412.

[30] Antti Tarvainen, Harri Valpola, Mean Teachers Are Better Role Models: Weight-Averaged Consistency Targets Improve Semi-supervised Deep Learning Results, 2017 arXiv preprint arXiv:1703.01780.

[31] Amir R. Zamir, Sax Alexander, Nikhil Cheerla, Rohan Suri, Zhangjie Cao, Jitendra Malik, Leonidas J. Guibas, Robust learning through cross-task consistency, in: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, CVPR, 2020, pp. 11197–11206.

[32] Fernando Navarro, Suprosanna Shit, Ivan Ezhov, Johannes Paetzold, Andrei Gafita, Jan C. Peeken, Stephanie E. Combs, Bjoern H. Menze, Shape-aware complementary-task learning for multi-organ segmentation, in: International Conference on Machine Learning in Medical Imaging, Springer, 2019, pp. 620–627.

[33] Xiaowei Liu, Lei Yang, Jianguo Chen, Siyang Yu, Keqin Li, Region-to-boundary deep learning model with multi-scale feature fusion for medical image segmentation, Biomed. Signal Process Control 71 (2022) 103165.

[34] Xue Yuan, Hui Tang, Zhi Qiao, Guanzhong Gong, Yong Yin, Zhen Qian, Chao Huang, Wei Fan, Xiaolei Huang, Shape-aware organ segmentation by predicting signed distance maps, in: Proceedings of the AAAI Conference on Artificial Intelligence (AAAI) 34, 2020, pp. 12565–12572.

[35] Gunilla Borgefors, Distance transformations in digital images, Comput. Vis. Graph Image Process 34 (3) (1986) 344–371.

[36] Fausto Milletari, Nassir Navab, Seyed-Ahmad 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.

[37] Qing-Long Zhang, Yu-Bin Yang, Sa-net: shuffle attention for deep convolutional neural networks, in: ICASSP 2021-2021 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP), IEEE, 2021, pp. 2235–2239.

[38] Ningning Ma, Xiangyu Zhang, Hai-Tao Zheng, Jian Sun, Shufflenet v2: practical guidelines for efficient cnn architecture design, in: Proceedings of the European Conference on Computer Vision, ECCV, 2018, pp. 116–131.

[39] Hoel Kervadec, Jihene Bouchtiba, Christian Desrosiers, Eric Granger, Jose Dolz, Ismail Ben Ayed, Boundary loss for highly unbalanced segmentation, in: International Conference on Medical Imaging with Deep Learning, PMLR, 2019, pp. 285–296.

[40] Codella Noel, Veronica Rotemberg, Philipp Tschandl, M. Emre Celebi, Dusza Stephen, David Gutman, Brian Helba, Aadi Kalloo, Konstantinos Liopyris, Michael Marchetti, et al., Skin Lesion Analysis toward Melanoma Detection 2018: A Challenge Hosted by the International Skin Imaging Collaboration (Isic), 2019 arXiv preprint arXiv:1902.03368.

[41] Tuan-Hung Vu, Himalaya Jain, Maxime Bucher, Matthieu Cord, Patrick Pérez, Advent: adversarial entropy minimization for domain adaptation in semantic segmentation, in: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), 2019, pp. 2517–2526.