

# SSCL: A Spatial-Spectral and Commonality Learning Network for Semi-Supervised Medical Image Segmentation

Yujie Liu      Zhonghao Du      Xuanting Li      Zongmin Li      Jiayue Fan  
Chaozhi Yang  
Qingdao Institute of Software, College of Computer Science and Technology,  
China University of Petroleum (East China)  
QingDao, China

z22070066@s.upc.edu.cn

## Abstract

The semi-supervised method has greatly promoted the development of medical image segmentation, as it alleviates the pressure of obtaining a large number of medical image annotations. In this article, we propose a new spatial-spectral and commonality learning network (SSCL) that better utilizes unlabeled data for semi-supervised medical image segmentation. The motivation of the SSCL model is to observe that, existing methods only focus on the features of the target area during segmentation, ignoring the features which can represent the total image, we call them common features. Because there is also feature information in the common features that can better assist the network in segmentation. In addition, due to the low contrast and high noise characteristics of medical images, only allowing the model to learn features in the spatial domain is not sufficient for the network to learn enough information. Due to insufficient feature information, the network will make more erroneous predictions when segmenting the edges of the target area than when segmenting the central area. Therefore, our proposed SSCL model consists of two new designs to address the above issues. First, we propose a reliable commonality learning module to learn the common features to help the network improve the segmentation performance. Second, we design a spectral convolution module to learn spectral feature information. Experimental results on three medical image datasets show that our framework outperforms previous state-of-the-art methods.

*Keywords: Semi-supervised learning, Medical image segmentation, Spectral feature, Common feature, Commonality learning*

## 1. Introduction

In the medical field, accurately segmenting internal structures from medical images plays an important role in many clinical applications [19], such as disease diagnosis and quantitative analysis. Due to the power of deep neural networks, many researchers have tried to use them to solve the task of medical image segmentation, and segmentation models based on a large amount of labeled data have achieved great results[7, 37, 40, 24, 10].

However, since medical image labeling often requires specialized knowledge and clinical experience, the labeling work is only suitable to be done by doctors or professionals, which makes it very difficult and expensive to acquire large amounts of labeled data. But obtaining a large amount of unlabeled data is much easier than obtaining labeled data, so how to use a small portion of labeled data and a large amount of unlabeled data to train the network to be able to solve the task of medical image segmentation becomes very important, this is called semi-supervised. Semi-supervised medical image segmentation methods have arisen under such a need and become an important research direction in computerized medical vision tasks.

Semi-supervised learning can be subdivided into two categories of methods: pseudo-labeling and consistency learning. Pseudo-labeling [13] methods are based on previously labeled data, where unlabeled data is simply pseudo-labeled and used along with labeled data to train the network; this method aims to expand the training dataset and thus improve the performance of the model. Consistency learning [27, 4, 38] methods encourage agreement in predictions between different perturbations by introducing perturbations to the model structure or input images, an approach designed to further train the network to improve its performance. However, previous methods [4, 38, 32, 31, 9, 25, 14, 16, 17, 30, 34] further improve the accuracy of segmentation results, there are still two problems with current semi-supervised medical image segmen-

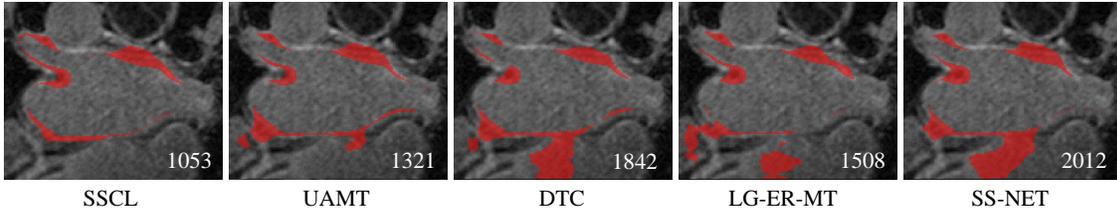


Figure 1. Visualization of the model’s erroneous predictions and the number of incorrectly predicted pixels. The red masks represent the model’s wrong predictions, and the white numbers represent the number of wrong pixels.

tation methods: (1) The network only focuses on the features of the target area, neglecting the features that can represent the total image, we call them common features. In previous work, labels or pseudo labels were usually used to supervise the network learning of the features of the target area in the image, which made the network only focus on extracting the features of the target area and ignoring other feature information in the image. However when using pseudo labels, due to inaccurate supervised signals, it is difficult for the network to learn accurate target area features. There is some information in common features that can help the network better segment the target area. Therefore, learning only the features of the target area limits the network’s ability to segment the target area; (2) Lack of sufficient feature information leads to more errors in edge segmentation. Due to the poor imaging quality of medical images, the edge information of medical images is very difficult to distinguish, they may contain very limited information in the spatial domain, so the networks only learn features in the spatial domain and cannot obtain sufficient feature information, leading to more erroneous predictions when segmenting the target area. As shown in Figure 1, displayed the model’s incorrect predictions and the number of pixels with incorrect predictions. It can be seen that the erroneous pixels are mainly located in the edge area of the target, which is caused by insufficient information obtained by the network from the image. Therefore, learning image features only in the spatial domain can not provide sufficient feature information for the segmentation network.

To alleviate the above existing problems, we propose a spatial-spectral and commonality learning network (SSCL) for semi-supervised medical image segmentation. Firstly, we propose a commonality learning module to learn common features, to solve the problem of networks only focusing on the features of the target region while ignoring common features. Specifically, outside of the segmentation network, we trained a reconstruction network as a commonality learning module. The commonality learning module does not require additional labeling to learn common features, because the label of the reconstruction network is the input image, and the common features learned by the reconstruction network are provided as compensation information to the segmentation network to help it better segment

the target area. Furthermore, in the field of semi-supervised medical image segmentation, we have for the first time introduced the extraction and processing of spectral domain information, to solve the problem of insufficient feature information leading to more edge segmentation errors. Previous work [8] has shown that there are spectral features that existing methods may miss in medical images, and semi-supervised learning requires more information than fully supervised learning. Therefore, we combine spectral and spatial features to provide more information to the network. Specifically, we have added a new spectral encoder to the network. During the training process, the input image will be fed to the spatial encoder and spectral encoder to learn spatial and spectral features respectively. In this way, the network can learn more feature information.

In summary, the main contributions of this paper are as follows:

- We propose a commonality learning module to learn common features, which is used to help the segmentation network better grasp the features of the target area and train the module without adding additional labels, ensuring the accuracy of learning features. (Section 3.2)
- We introduce spectral features that extract and process information from the spectral domain to provide more feature information for the network, to solve the problem of only learning features in the spatial domain without sufficient feature information. (Section 3.3)
- We validate the proposed SSCL on three public benchmark datasets, and the experimental results show that the proposed SSCL outperforms other state-of-the-art methods in all three benchmarks, indicating that SSCL is a superior architecture for semi-supervised medical image segmentation. The ablation experiments further demonstrate the effectiveness of each proposed module. (Section 4)

## 2. Related Work

### 2.1. Medical Image Segmentation

Segmenting target organs or lesions from medical images plays a vital role in many clinical applications, as accurate segmentation information can help doctors better assess a patient’s condition. The development of deep learning has greatly improved the accuracy and stability of med-

ical image segmentation results. In the field of medical image segmentation, U-Net [21, 6] and its encoder-decoder framework have been widely used due to their simplicity and accuracy. The main advantage of U-Net is its skip connection design. To improve performance, various methods have been proposed, such as redesigning skip connections [41], introducing residual/dense convolution blocks in the network [1, 15], or optimizing feature maps by introducing attention mechanisms [20]. nnU-Net [12] proposed by Isensee et al. enables the segmentation network to adapt to training strategies and network frameworks automatically. In recent years, with the remarkable achievements of transformer in traditional computer vision tasks, transformer have also received widespread attention in the field of medical segmentation, and many transformer-based segmentation methods have been proposed. Chen et al. [3] improved the segmentation performance by replacing the encoder in U-Net with a transformer. Cao et al. [2] constructed a pure transformer segmentation network, Swin-Unet, by using transformer blocks to replace the convolution blocks in the U-Net framework. VNet [18] was proposed by Milletari et al. to address the challenges in medical image segmentation, especially when processing 3D volume data. Although these methods have achieved great results in the medical segmentation field, they are all based on supervised training mode, and obtaining medical image annotations is very complex and costly, which limits these methods to a small amount of labeled data.

## 2.2. Semi-supervised learning

Semi-supervised learning trains a network by utilizing a combination of labeled and unlabeled data, allowing the network's performance to approach that of fully supervised learning. The theoretical basis of semi-supervised learning comes from three fundamental assumptions [28]: (1) smoothness assumption: for two adjacent input samples  $x_1, x_2 \in X$  in the input space, the corresponding labels  $y_1$  and  $y_2$  should be the same, and vice versa. (2) low-density assumption: samples within a single class tend to form a cluster, so the decision boundary of the classifier should pass through low-density areas in the input space, and not high-density regions. (3) manifold assumption: samples located on the same low-dimensional manifold should belong to the same class, reflecting the local smoothness of the decision boundary.

As mentioned in the introduction section, semi-supervised learning methods can be divided into two categories: pseudo-labeling and consistency learning. The pseudo-labeling method attempts to generate pseudo-labels similar to ground truth labels to expand the training set for network training. Lee et al. [13] proposed using the predictions of a fully supervised network as pseudo-labels for unlabeled data, but this method introduces a lot of noise

during the training process. To address this issue, Sohn et al. [26] reduced the number of erroneous predictions by setting a threshold for the predicted values and only retaining high-confidence annotations.

The core idea of consistency learning is that for an input, even if perturbed, the network can still produce an output that is consistent with the original. One typical representative is the mean teacher proposed by Tarvainen et al. [27], which applies two different augmentations to the same input to increase perturbation and uses the output of the teacher model to supervise the student model. After that, many works have extended the mean teacher framework differently. Because the teacher network in the mean teacher framework updates parameters through exponential moving average (EMA) [27], the performance of the teacher network is limited by the student network. Therefore, Chen et al. [4] proposed a cross-pseudo-supervision method where two networks with different parameter initialization supervise each other and independently update their parameters.

## 2.3. Semi-supervised Medical Image Segmentation

The difficulty in obtaining labeled medical image data and the success of semi-supervised learning have driven the development of semi-supervised medical image segmentation research. Variants of the mean teacher framework have been widely used in semi-supervised medical image segmentation. Yu et al. proposed a mean teacher framework UA-MT [38] that uses uncertainty information to guide the student network to learn more reliable targets from the teacher network and enhances network performance by introducing transformation perturbations. Wang et al. proposed a dual uncertainty-weighted method Double-UA [32] based on this uncertainty-aware teacher-student model, which simultaneously considers the uncertainty of predictions and features and refines the predictions of the teacher model during training. Afterward, Wang et al. proposed a triple uncertainty-guided mean teacher framework Triple-UA [31] by defining two auxiliary tasks on the mean teacher network to help the model learn different features and make better predictions. Hang et al. used the global-local structure-aware entropy minimization method LG-ER-MT [9] based on the mean teacher network. CoraNet [25] proposed a model that can generate both deterministic and uncertain regions, where the student network assigns different weights to the regions given by the teacher network. In addition, some other methods have also achieved good results. SASSNet [14] introduced a shape-aware semi-supervised segmentation strategy that integrates more flexible geometric representations into the network to improve performance. DTC [16] proposed a dual-task consistency framework by building task-level regularization training, which encourages consistent predictions of

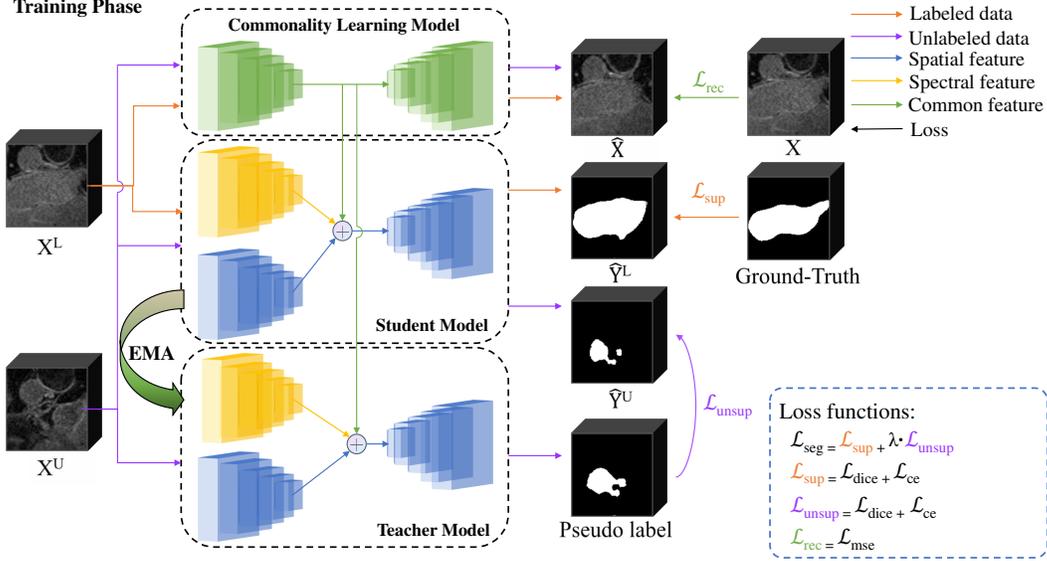


Figure 2. The process of our proposed SSCL framework.

the same input under different tasks. Luo et al. proposed a pyramid multi-scale architecture (URPC) [17], which encourages consistent regularization of predictions of unlabeled inputs at multiple scales. Wang et al. proposed a generative Bayesian deep learning model (GBDL) [30] to learn the joint distribution of data and labels. Wu et al. proposed a new competitive winning method (ComWin) [33] to improve the quality of pseudo labels. Xu et al. proposed a novel ambiguity-consensus mean-teacher (AC-MT) [23] model. Shen et al. proposed a multi-network collaborative training UCMT [36] with high-confidence pseudo labels. Although previous methods have improved the accuracy of segmentation results, they only learn the features in the spatial domain and cannot provide sufficient feature information for the segmentation network; And only learn the features of the target area, ignoring the information in the common features may also play a promoting role in network segmentation. This article proposes solutions to these problems.

### 3. Method

In this section, we first introduce our proposed SSCL segmentation network. After this, we describe the modules in the SSCL framework, and the loss functions are described. Finally, the testing stage is illustrated.

#### 3.1. Overview of the Framework

In the task of semi-supervised medical image segmentation, we assume a complete dataset that contains two types of data: labeled data and unlabeled data. The number of labeled data is  $N$ , and the number of unlabeled data is  $M$ , where  $M \gg N$ . For the sake of simplicity, we define the

two types of data as two subsets of the dataset, denoted as  $D = \{D_L, D_U\}$ . The labeled subset is represented as  $D_L = (X_i^L, Y_i^L)_{i=1}^N$ , and the unlabeled subset is represented as  $D_U = (X_i^U)_{i=1}^M$ . Here,  $X \in \mathbb{R}^{W \times H \times D}$  represents the input image during the training phase, and  $Y$  represents the corresponding labels (only for labeled data). The goal of semi-supervised medical image segmentation is to predict the label mapping  $\hat{Y} \in \{0, 1, \dots, C\}^{W \times H \times D}$  for each voxel  $k \in X$ , where 0 represents the background class, and the rest correspond to target classes.

The overall training process of the proposed SSCL is illustrated in Figure 2. SSCL is based on the Mean Teacher framework, which consists of a teacher network  $T(F_T(\cdot))$ , a student network  $S(F_S(\cdot))$ , and commonality learning module  $R(F_R(\cdot))$ , where the commonality learning module is a reconstruction network. A batch of input data  $X$  contains an equal number of labeled data ( $X_L, Y_L$ ) and unlabeled data  $X_U$ . The input data  $X$  is first passed through the reconstruction network to obtain the common features  $F_c$  and the reconstruction prediction  $\hat{X}$  for the two types of data.

$$\hat{X}, F_c = F_R(X) \quad (1)$$

Note that the common features  $F_c$  are the output of the encoder in the reconstruction network. Then, the common features  $F_c$ , along with the input data  $X$ , are separately fed into the teacher and student networks to obtain the predicted values:

$$\hat{Y}_T = F_T(X^U, F_c) \quad (2)$$

$$\hat{Y}_S = F_S(X, F_c) \quad (3)$$

The overall network output includes the segmentation predictions  $\hat{Y}_T$  and  $\hat{Y}_S$  from the teacher and student networks,

respectively, as well as the reconstruction predictions  $\hat{X}$  for the two types of data. The student network is pseudo-supervised using the predictions from the teacher network. The loss function  $L_{seg}$  consists of both supervised and unsupervised losses, which are combined by cross-entropy loss and dice loss [18], this is a popular approach in medical image segmentation. The input for the reconstruction network, as well as the supervised labels, is the input image  $X$ . Therefore, this network does not differentiate between the two types of data. The loss function  $L_{rec}$  is composed of the MSE loss. The reconstruction network and the student network independently update their parameters based on the calculated losses, while the teacher network updates its parameters using exponential moving average (EMA) [27].

### 3.2. Commonality learning module

The purpose of the commonality learning module is to learn the common features of labeled and unlabeled data and help the segmentation network better capture the features of the target regions. In semi-supervised segmentation networks, the lack of reliable supervision signals prevents the segmentation network from accurately learning the features of the target regions in unlabeled data. The proposed commonality learning module, as shown in Figure 2, consists of a VNet [18]. The input and supervised labels for this module are both the input image, and the output is the reconstructed input image. The entire network is trained by calculating the loss between the reconstructed and input images. Therefore, we use the common features learned by the commonality learning module as compensatory information provided to the segmentation network.

When the input image  $X$  enters the commonality learning module, the network extracts the common features  $F_c$ . Here, we take the features outputted by the last layer of the encoder in the network as the common features  $F_c$  provided to the segmentation network. The common features  $F_c$  are added to the final features  $F_f$  (which will be discussed in the next section) outputted by the last layer of the encoder in the segmentation network. The combined features are then passed to the decoder for segmentation, resulting in the final segmentation result  $\hat{Y}$ :

$$\hat{Y} = D(F_c + F_f) \quad (4)$$

Since the commonality learning module is trained in a fully supervised manner, it can ensure the accuracy of the learned common features. By providing common features as supplementary information to the segmentation network, the segmentation network can improve the segmentation performance of the target area with the help of common features.

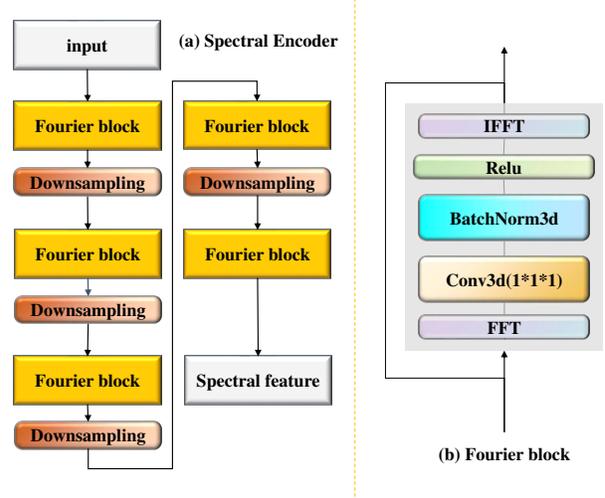


Figure 3. The Spectral Encoder (a) and Fourier block (b).

### 3.3. Spectral Feature extraction

As shown in Figure 2, the student and teacher networks for image segmentation consist of two encoder branches,  $E_{spa}$  and  $E_{spe}$ , and a decoder  $D$ .  $E_{spa}$  is a spatial encoder (the blue part in Figure 2) used to extract spatial features  $F_{spa}$  from the image, which has the same structure as the encoder in the VNet and is composed of five convolution blocks. On the other hand,  $E_{spe}$  is a spectral encoder (the yellow part in Figure 2) used to extract spectral features  $F_{spe}$  from the image. Its structure, as shown in Figure 3(a), consists of five Fast Fourier Convolution (FFC) [5] blocks, which have the same overall structure as the spatial encoder but with FFC blocks instead of convolution blocks.  $D$  is a VNet decoder that outputs the segmentation results. Since medical images have limited inherent information, using only a spatial encoder to learn image features may not be sufficient. Therefore, we introduce a spectral encoder to extract and process spectral features from the spectral domain, providing additional feature information to the segmentation network and helping improve segmentation accuracy.

The input of the spectral encoder  $E_{spe}$  is the same as that of the spatial encoder  $E_{spa}$ . When the input image enters the spectral encoder  $E_{spe}$ , its processing procedure is similar to the spatial encoder  $E_{spa}$ . FFC blocks are used to convolve and learn features, followed by downsampling operations. The features  $F_{spe}$  outputted by the spectral encoder  $E_{spe}$  are added to the features  $F_{spa}$  to get the final features  $F_f$ , complementing the network with spectral feature information.

$$F_f = F_{spa} + F_{spe} \quad (5)$$

The final features  $F_f$ , combined with the common features  $F_c$  obtained from the commonality learning module, are then passed to the decoder  $D$  to obtain the final results.

The FFC block in the spectral encoder, as shown in Figure 3(b), first performs a 3D Fast Fourier Transform (FFT) on the input to obtain the real and imaginary parts,  $a + bi \in \mathbb{C}$ . The concatenated real and imaginary parts were then passed through a convolutional layer with a kernel size of 1 to learn spectral features. After that, the output of the convolutional layer goes through an activation layer and batch normalization layer. The output is then split into two parts, the real and imaginary parts, finally, a 3D inverse Fast Fourier Transform (IFFT) is performed to transform the spectral features back to the spatial domain, aligning the spectral features with the corresponding spatial domain features.

### 3.4. Loss Function

In the SSCL framework, two networks need to be updated through loss computation: the student network and the reconstruction network. Therefore, the loss function includes two parts: the segmentation loss  $L_{seg}$  for training the student network and the reconstruction loss  $L_{rec}$  for training the reconstruction network.

The segmentation loss  $L_{seg}$  for training the student network consists of two components: the supervised loss and the unsupervised loss. It can be represented as follows:

$$L_{seg} = L_{sup} + \lambda L_{unsup} \quad (6)$$

$L_{sup}$  is the supervised loss,  $L_{unsup}$  is the unsupervised loss, and  $\lambda$  is a balancing parameter with a default value of 0.5. Both supervised and unsupervised losses are composed of a combination of dice loss [18] and cross-entropy loss. The difference is that the supervised loss is supervised by ground truth labels, while the unsupervised loss is pseudo-supervised using pseudo-labels generated by the teacher network. The representation is as follows:

$$L_{sup} = DICE(\hat{Y}^L, Y^L) + CE(\hat{Y}^L, Y^L) \quad (7)$$

$$L_{unsup} = DICE(\hat{Y}^U, Y^P) + CE(\hat{Y}^U, Y^P) \quad (8)$$

$Y^P$  represents pseudo-labels, which are determined by applying a common threshold of 0.5 to the predicted values  $\hat{Y}_T$  of the teacher network.

The reconstruction loss  $L_{rec}$  for the reconstruction network is relatively straightforward. It is trained by calculating the MSE loss between the reconstructed image outputted by the network and the input image. It can be represented as follows:

$$L_{rec} = MSE(\hat{X}, X) \quad (9)$$

$\hat{X}$  represents the reconstructed image, and  $X$  represents the input image.

The teacher network updates its parameters  $W_T^{k+1}$  at the (k+1)th iteration using an exponential moving average

(EMA). The update of the teacher network parameters is represented as follows:

$$W_T^{k+1} = \alpha W_T^k + (1 - \alpha) W_S^k \quad (10)$$

Here,  $\alpha$  is the weight parameter, default is 0.999, and  $W_S^k$  represents the parameters of the student network at the k-th iteration.

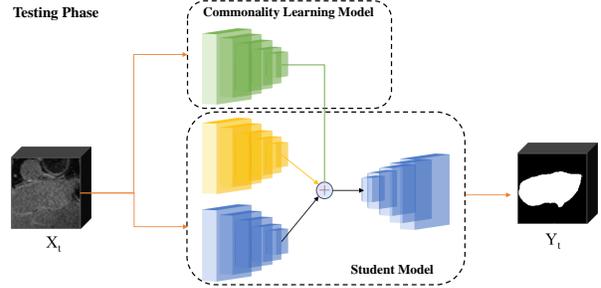


Figure 4. The process of testing phase. Note that only the student model and the commonality learning model are needed in the testing phase.

### 3.5. Testing Phase

The testing phase of SSCL is shown in Figure 4. During testing, we use two networks: the trained student network and the reconstruction network. The reconstruction network is kept during testing because the common features learned by the reconstruction network can be added to the features of the segmentation network to improve the segmentation performance. For a test image  $X_t$ , it is sent to both networks separately. The reconstruction network provides the common features  $F_c$  of the test image  $X_t$  to the student network, and the final prediction  $Y_t$  can be obtained from the student network:

$$Y_t = F_S(X_t, F_c) \quad (11)$$

## 4. Experiments

### 4.1. Dataset and Implementation Details

**Left Atrial Dataset (LA) [35]:** The LA dataset is the benchmark dataset for the 2018 Atrial Segmentation Challenge, containing 100 3D gadolinium-enhanced magnetic resonance imaging (GE-MRI) scans with labels. Following the settings used in DTC [16] and UA-MT [38], we use 80 data for training and 20 data for testing. In our experiments, we conducted two typical semi-supervised settings, using 10% and 20% labeled data for training.

**NIH-pancreas dataset [22]:** The Pancreas CT dataset is publicly available from the National Institutes of Health Clinical Center and includes 82 abdominal CT images manually annotated by experienced physicians. The size of

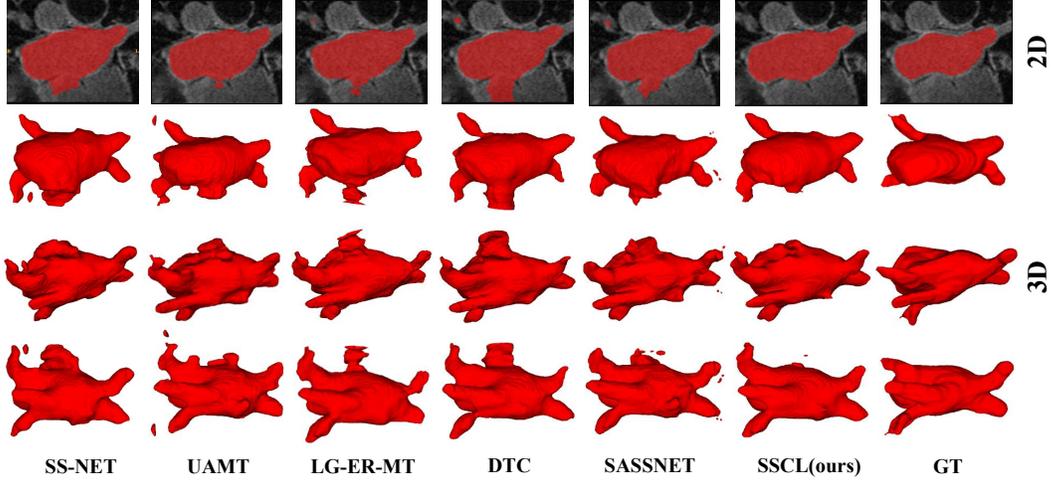


Figure 5. Visualizations of several semi-supervised segmentation methods obtained by SS-NET [34], UAMT [38], LG-ER-MT [9], DTC [16], SASSNET [14], our SSCL model with 20% labeled data and ground truth on LA dataset.

Method	Volumes used		Metrics			
	Labeled	Unlabeled	Dice $\uparrow$	Jaccard $\uparrow$	95HD $\downarrow$	ASD $\downarrow$
VNet	8(10%)	0	0.786	0.670	21.20	6.07
VNet	16(20%)	0	0.870	0.773	11.85	3.22
VNet	80(100%)	0	0.914	0.844	5.48	1.51
UA-MT [38] (MICCAI 2019)			0.843	0.735	13.83	3.36
SASSNet [14] (MICCAI 2020)			0.873	0.777	9.62	2.55
Double-UA [32] (MICCAI 2020)			0.859	0.758	12.67	3.31
LG-ER-MT [9] (MICCAI 2020)			0.855	0.751	13.29	3.77
Tripled-UA [31] (MICCAI 2021)			0.868	0.768	10.42	2.98
URPC [17] (MICCAI 2021)			0.850	0.744	15.37	3.96
DTC [16] (AAAI 2021)	8(10%)	72(90%)	0.875	0.782	8.23	2.36
SS-Net [34] (MICCAI 2022)			0.886	0.796	7.49	1.90
CoraNet [25] (TMI 2022)			0.866	0.781	12.11	2.40
GBDL [30] (CVPR 2022)			0.884	0.792	<b>5.89</b>	<b>1.60</b>
UCMT [23] (IJCAI 2023)			0.881	0.791	9.14	3.06
AC-MT [36] (MedIA 2023)			0.891	0.805	11.05	2.19
SSCL (ours)			<b>0.893</b>	<b>0.807</b>	7.40	2.03
UA-MT [38] (MICCAI 2019)			0.889	0.802	7.32	2.26
SASSNet [14] (MICCAI 2020)			0.895	0.812	8.24	2.20
Double-UA [32] (MICCAI 2020)			0.897	0.814	7.04	2.03
LG-ER-MT [9] (MICCAI 2020)			0.896	0.813	7.16	2.06
Tripled-UA [31] (MICCAI 2021)			0.893	0.810	7.42	2.21
URPC [17] (MICCAI 2021)			0.887	0.799	12.73	3.66
DTC [16] (AAAI 2021)	16(20%)	64(80%)	0.894	0.810	7.32	2.10
SS-Net [34] (MICCAI 2022)			0.889	0.802	8.02	2.36
CoraNet [25] (TMI 2022)			0.887	0.811	7.55	2.45
GBDL [30] (CVPR 2022)			0.894	0.822	<b>4.03</b>	<b>1.48</b>
UCMT [23] (IJCAI 2023)			0.904	0.825	6.31	1.70
AC-MT [36] (MedIA 2023)			0.903	0.824	6.21	1.76
SSCL (ours)			<b>0.905</b>	<b>0.828</b>	6.39	1.94

Table 1. Comparison with state-of-the-art semi-supervised segmentation methods on the LA dataset.

CT scans ranges from  $512 \times 512 \times 181$  to  $512 \times 512 \times 466$  voxels, with interlayer spacing ranging from 1.5 to 2.5mm. During preprocessing, we used a soft tissue CT window range of  $[-125, 275]$ HU and resampled all images to isotropic resolution of  $1.0 \times 1.0 \times 1.0$ mm. In the experiments, we used 62 samples for training and the remaining 20 samples for testing, following the setting in CoraNet [25].

**Kits19 dataset [11]:** It is a kidney tumor segmentation dataset containing 210 labeled 3D computed tomography (CT) scans for training and validation. Following previous work [30], we use 160 data for training and 50 data for testing.

**Implementation details:** We implemented our model in PyTorch and conducted all experiments on an NVIDIA 3090TI GPU with a fixed random seed. We use the SGD optimizer with an initial learning rate of 0.01, weight decay of 0.0001, and momentum of 0.9 to train the model. VNet is used as the backbone and sets the batch size to 4, including 2 labeled data and 2 unlabeled data. The iteration numbers for pre-training and formal training were set to 4k and 30k, respectively. To avoid overfitting, rotation, and flip operations are implemented to augment the data, following previous work. Since 3D data training requires a lot of computation, we cropped all training data into small patches during training. The training patch sizes for LA, Pancreas-CT, and Kits19 were  $112 \times 112 \times 80$ ,  $96 \times 96 \times 96$ , and  $128 \times 128 \times 64$ , respectively. During the testing phase, we use the slide window strategy to obtain the final results, with a step size of  $18 \times 18 \times 4$  for LA and Kits19, and a step size of  $16 \times 16 \times 16$  for Pancreas-CT.

## 4.2. Evaluation Metrics

We use four metrics to evaluate the performance of the model, including Dice, Jaccard, average surface distance (ASD), and 95% Hausdorff distance (95HD). The Dice and Jaccard coefficients are primarily used to calculate the percentage of overlap between two object regions. The ASD measures the average distance between the boundaries of the two object regions, while the 95HD measures the distance between the closest points of the two object regions.

## 4.3. Results

**LA dataset:** We first evaluated our proposed method on the left atrial segmentation task. The compared methods included UA-MT [38], SASSNet [14], Double-UA [32], Tripled-UA [31], CoraNet [25], URPC [17], DTC [16], SS-Net [34], LG-ER-MT [9], GBDL [30], AC-MT [36], and UCMT [23]. We conducted semi-supervised experiments under different labeling ratios (10% and 20%). In addition, VNet is used as the baseline with 10% and 20% labeled data. As shown in Table 1, all methods benefited from the unlabeled data, and our method outperformed the

state-of-the-art methods in terms of Dice and Jaccard coefficients in both cases, for example, compared with UA-MT and GBDL, the Dice coefficients of SSCL increased by 5% and 0.9%, respectively, on 10% of labeled data. It can also be seen that our segmentation results are significantly better than the compared methods from Figure 5, whether in 2D or 3D results, our SSCL prediction is closest to the Ground Truth. Figure 1 also shows that our SSCL reduces more false predictions at the target edge compared to UAMT, DTC, LG-ER-MT, and SS-NET. Demonstrating the superiority of SSCL in semi-supervised medical image segmentation. On the two metrics of 95HD and ASD, it can be seen that our SSCL is better than most methods, but GBDL achieved the best results on both indicators in both cases. We consider that this is possibly due to GBDL extracting voxel data as slices for segmentation instead of directly segmenting the voxel data, as the heart is an irregularly shaped organ, therefore, slice segmentation may achieve better results on these two metrics than directly segmenting voxels.

**NIH-Pancreas dataset:** We further evaluated our proposed method on the pancreas dataset with 20% labeled data. Since the pancreas is located deep in the abdomen and has large variations in size, location, and shape, and the pancreas CT has a more complex background than left atrial MRI, pancreas segmentation is more challenging than left atrial segmentation. However, our proposed SSCL still showed good performance. We compared our method with DAN [39], ADVNET [29], VNet [18], UA-MT [38], SASSNet [14], DTC [16], and CoraNet [25], ComWin [33], as shown in Table 2. Compared with these methods, SSCL performed significantly better in terms of Dice, Jaccard, and 95HD, especially, on Dice, it improved by 0.9% compared to CoraNet, and on Jaccard, it improved by 0.8% compared to ComWin, confirming that our proposed SSCL can better utilize unlabeled data and has stronger information extraction ability.

**Kits19 dataset:** The experimental results on the kidney segmentation dataset are shown in Table 3, where we experimented with 10% labeled data. We compared our method with UA-MT [38], SASSNet [14], Double-UA [32], Tripled-UA [31], CoraNet [25], and GBDL [30]. From the table, we can see that our method SSCL exhibited the best performance in all four evaluation metrics, outperforming all compared methods and further validating the effectiveness of SSCL in medical image segmentation. It is worth noting that our SSCL not only outperforms GBDL in Dice and Jaccard but also outperforms GBDL in 95HD and ASD metrics in this dataset, which was not achieved in the LA dataset. We think this may be because the kidney has a more regular shape compared to the heart, so the advantage of slicing for segmentation on these two metrics is lost.

Method 2-7	Volumes used		Metrics			
	Labeled	Unlabeled	Dice $\uparrow$	Jaccard $\uparrow$	95HD $\downarrow$	ASD $\downarrow$
VNet	12(20%)	0	0.706	0.567	22.54	6.29
VNet	62(100%)	0	0.818	0.697	5.13	1.34
VNet [18](3DV 2016)			0.700	0.556	14.27	1.64
DAN [39](MICCAI 2017)			0.767	0.633	11.13	2.97
ADVNET [29](CVPR 2019)			0.753	0.617	11.72	3.88
UA-MT [38](MICCAI 2019)			0.773	0.638	11.90	3.06
SASSNet [14](MICCAI 2020)	12(20%)	50(80%)	0.777	0.641	10.93	3.05
DTC [16](AAAI 2021)			0.783	0.648	8.36	2.25
CoraNet [25](TMI 2022)			0.797	0.667	7.59	1.89
ComWin [33](TMI 2023)			0.796	0.670	6.95	<b>1.34</b>
SSCL (ours)			<b>0.806</b>	<b>0.678</b>	<b>5.83</b>	1.63

Table 2. Comparison with state-of-the-art semi-supervised segmentation methods on the NIH-Pancreas dataset.

Method 2-7	Volumes used		Metrics			
	Labeled	Unlabeled	Dice $\uparrow$	Jaccard $\uparrow$	95HD $\downarrow$	ASD $\downarrow$
UA-MT [38](MICCAI 2019)			0.883	0.802	9.46	2.89
SASSNet [14](MICCAI 2020)			0.891	0.822	7.54	2.41
Double-UA [32](MICCAI 2020)			0.895	0.828	7.42	2.16
Tripled-UA [31](MICCAI 2021)	16(10%)	144(90%)	0.887	0.815	7.55	2.12
CoraNet [25](TMI 2022)			0.898	0.820	7.23	1.89
GBDL [30](CVPR 2022)			0.911	0.840	6.38	1.51
SSCL (ours)			<b>0.913</b>	<b>0.843</b>	<b>5.08</b>	<b>1.30</b>

Table 3. Comparison with state-of-the-art semi-supervised segmentation methods on the Kits19 dataset.

#### 4.4. Ablation Studies

**Effectiveness of Each Component:** To better understand and evaluate the components of our proposed SSCL method, we conducted ablation experiments on the LA dataset with 10% labeled data, and the results are shown in Table 4. We used VNet as the backbone and trained it using only 10% labeled data, with its performance as the baseline, and trained it using 100% labeled data, with its performance as the upper limit. We evaluated the effectiveness of each module by gradually adding the proposed modules to the baseline model and observing the changes in segmentation performance. REC refers to the commonality learning module, and FFC refers to the spectral feature learning module. From Table 4, we can see that after adding the two modules separately to the baseline method, all metrics improved significantly, with the Dice coefficient increasing by more than 3% for each module. Moreover, the performance obtained by combining the two modules was further improved based on the separate addition of the two modules, and the improvement was very significant. The experimental results indicate that the addition of these two modules contributes to the segmentation performance, and when the two modules are combined, the performance gain is maximized.

**Weight  $\lambda$  in Loss Function:** To balance the weights of supervised and unsupervised losses, we weighted the un-

supervised loss and set the default value of  $\lambda$  to 0.5. We conducted experiments on the LA dataset with 10% labeled data by changing  $\lambda = \{0.5, 1, 1.5, 2, 2.5, 3\}$  to observe the effects of different weights on the network performance and the results are shown in Table 5. From the table, we can see that the best performance was achieved when  $\lambda = 0.5$ , and the performance of the model did not differ much when  $\lambda \leq 2$ . However, when  $\lambda > 2$ , the model performance decreased significantly.

**Weight  $\alpha$  in EMA:** In the teacher-student network framework, the parameters of the teacher network are updated using the exponential moving average (EMA) based on the parameters of the student model. And the weight was set to 0.99 default. However as shown in Table 6, we found that changing the weight from 0.99 to 0.999 improved the performance of the model. This may be because as parameter updates became slower, the robustness of the teacher model was enhanced, allowing the student network to learn more stable information from the teacher network.

**Effectiveness of Threshold:** Figure 6 shows the performance of the SSCL model trained with different pseudo-label thresholds on the LA dataset. The results indicate that as the pseudo-label threshold increases, the overall performance of the model on all four evaluation indicators shows a downward trend. This may be because as the pseudo-label

Method	Volumes used		Metrics			
	Labeled	Unlabeled	Dice $\uparrow$	Jaccard $\uparrow$	95HD $\downarrow$	ASD $\downarrow$
2-7						
Baseline	8(10%)	0	0.831	0.717	21.46	6.55
Upper-bound	80(100%)	0	0.911	0.836	10.09	2.67
Baseline + REC			0.867	0.768	14.86	4.50
Baseline + FFC	8(10%)	72(90%)	0.878	0.784	13.48	3.62
SSCL			<b>0.893</b>	<b>0.807</b>	<b>7.40</b>	<b>2.03</b>

Table 4. Ablation studies of different parts in SSCL on LA dataset.

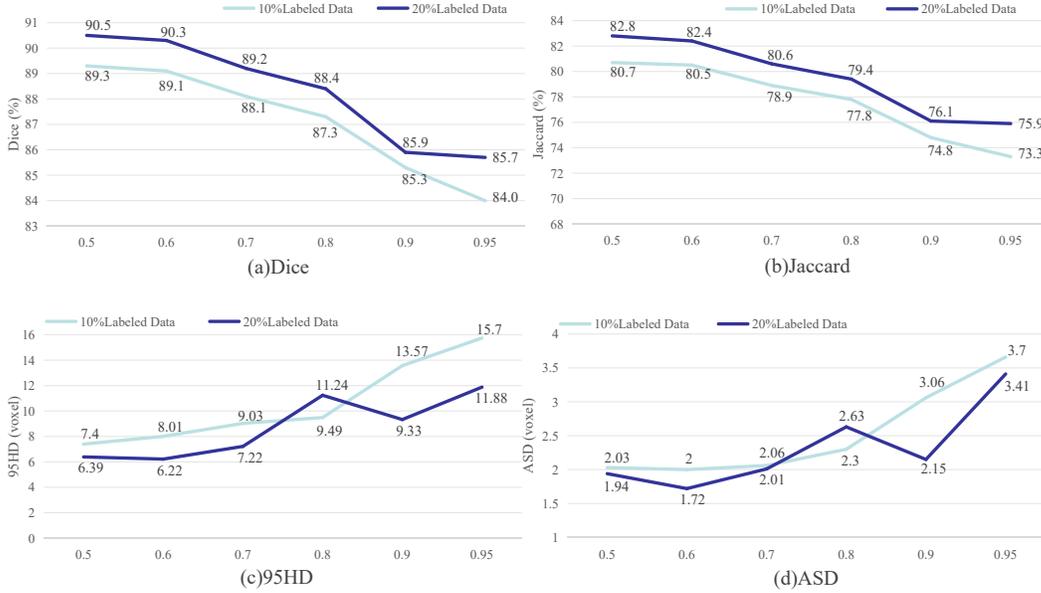


Figure 6. The results of SSCL with different thresholds on the LA datasets.

$\lambda$	Metrics			
	Dice $\uparrow$	Jaccard $\uparrow$	95HD $\downarrow$	ASD $\downarrow$
2-5				
$\lambda = 0.5$	<b>0.893</b>	<b>0.807</b>	<b>7.40</b>	<b>2.03</b>
$\lambda = 1.0$	0.890	0.803	8.28	2.56
$\lambda = 1.5$	0.888	0.800	9.96	3.12
$\lambda = 2.0$	0.887	0.797	11.82	3.29
$\lambda = 2.5$	0.869	0.771	12.02	3.87
$\lambda = 3.0$	0.711	0.560	26.88	9.43

Table 5. Ablation studies of weights  $\lambda$  in the loss function with 10% labeled data.

threshold increases while filtering out erroneous information, effective information is also greatly filtered. Therefore, we ultimately used a pseudo-label threshold of 0.5 to generate pseudo-labels on all datasets.

## 5. Discussion

Compared to other methods, SSCL performs better for two main reasons. Firstly, it provides common features that are missed during segmentation for the segmentation net-

$\lambda$	Metrics			
	Dice $\uparrow$	Jaccard $\uparrow$	95HD $\downarrow$	ASD $\downarrow$
2-5				
$\alpha = 0.99$	0.891	0.805	7.89	2.04
$\alpha = 0.999$	<b>0.893</b>	<b>0.807</b>	<b>7.40</b>	<b>2.03</b>
$\alpha = 0.99$	0.900	0.820	6.99	2.44
$\alpha = 0.999$	<b>0.905</b>	<b>0.828</b>	<b>6.39</b>	<b>1.94</b>

Table 6. Ablation studies of weights  $\alpha$  in EMA with 10% and 20% labeled data.

work. Other methods have ignored the common features of the image, while in our method, we successfully compensated the lost common features back to the segmentation network. The experimental results confirm that helping the segmentation network compensate for common features can better improve the segmentation effect. The second reason is to add a spectral encoder to the segmentation network to learn the spectral features of the image. The visualization of segmentation results also shows that due to the increase in feature information, our SSCL produces fewer erroneous predictions for edge segmentation compared to

other methods. Due to these two reasons, our SSCL ultimately achieved very good segmentation results.

Although the proposed SSCL method has achieved success in semi-supervised medical image segmentation, the fusion strategy of the three features in the proposed model is a simple addition processing, and the three networks trained in the model result in high training time costs. Future work can attempt to find better feature fusion strategies to more effectively utilize feature information, and design a better network structure to reduce the training cost of SSCL.

## 6. Conclusion

In this paper, we propose a semi-supervised medical image segmentation framework called SSCL, which reconstructs input images to obtain common features and introduces spectral feature information to help the network better learn the features of the target region in the image and improve segmentation accuracy. The proposed commonality learning module in SSCL ensures the accuracy of common features without introducing additional labels in a fully supervised manner. The spectral feature extraction module helps the network obtain more feature information by learning the spectral features of the input image. We conducted experiments on three common benchmark datasets with two different modalities, including the left atrial dataset of MR scans, the pancreas dataset, and the kidney segmentation dataset of CT scans, the segmentation results outperformed the previous state-of-the-art methods, demonstrating the effectiveness, robustness, and generalization of SSCL, as well as its potential in semi-supervised medical image segmentation.

## Acknowledgement

This work is partly supported by National key r&d program(Grant no.2019YFF0301800),National Natural Science Foundation of China (Grant no.61379106), the Shandong Provincial Natural Science Foundation (Grant nos.ZR2013FM036,ZR2015FM011).

## References

- [1] M. Z. Alom, C. Yakopcic, M. Hasan, T. M. Taha, and V. K. Asari. Recurrent residual u-net for medical image segmentation. *Journal of Medical Imaging.*, 6:014006, 2019. 3
- [2] H. Cao, Y. Wang, J. Chen, D. Jiang, X. Zhang, Q. Tian, and M. Wang. Swin-unet: Unet-like pure transformer for medical image segmentation. In *European conference on computer vision*, pages 205–218, 2022. 3
- [3] J. Chen, Y. Lu, Q. Yu, X. Luo, E. Adeli, Y. Wang, L. Lu, A. L. Yuille, and Y. Zhou. Transunet: Transformers make strong encoders for medical image segmentation. *arXiv preprint arXiv:2102.04306*, 2021. 3
- [4] X. Chen, Y. Yuan, G. Zeng, and J. Wang. Semi-supervised semantic segmentation with cross pseudo supervision. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 2613–2622, 2021. 1, 3
- [5] L. Chi, B. Jiang, and Y. Mu. Fast fourier convolution. *Advances in Neural Information Processing Systems.*, 33:4479–4488, 2020. 5
- [6] Ö. Çiçek, A. Abdulkadir, S. S. Lienkamp, T. Brox, e.-S. Ronneberger, Olaf\*, L. Joskowicz, M. R. Sabuncu, G. Unal, and W. Wells. 3d u-net: Learning dense volumetric segmentation from sparse annotation. In *Medical Image Computing and Computer-Assisted Intervention – MICCAI 2016*, pages 424–432, 2016. 3
- [7] Q. Dou, Q. Liu, P. A. Heng, and B. Glocker. Unpaired multi-modal segmentation via knowledge distillation. *IEEE Transactions on Medical Imaging.*, 39:2415–2425, 2020. 1
- [8] A. Farshad, Y. Yeganeh, P. Gehlbach, and N. Navab. Y-net: A spatio-spectral network for retinal oct segmentation. In *International Conference on Medical Image Computing and Computer-Assisted Intervention*, 2022. 2
- [9] W. Hang, W. Feng, S. Liang, L. Yu, Q. Wang, K.-S. Choi, and J. Qin. Local and global structure-aware entropy regularized mean teacher model for 3d left atrium segmentation. In *Medical Image Computing and Computer Assisted Intervention–MICCAI 2020*, pages 562–571, 2020. 1, 3, 7, 8
- [10] Q. He, Q. Yang, H. Su, and Y. Wang. Multi-task learning for segmentation and classification of breast tumors from ultrasound images. *Computers in Biology and Medicine.*, 173:108319, 2024. 1
- [11] N. Heller, F. Isensee, K. H. Maier-Hein, X. Hou, C. Xie, F. Li, Y. Nan, G. Mu, Z. Lin, M. Han, and et al. The state of the art in kidney and kidney tumor segmentation in contrast-enhanced ct imaging: Results of the kits19 challenge. *Medical Image Analysis.*, 67:101821, 2021. 8
- [12] F. Isensee, P. F. Jaeger, S. A. Kohl, J. Petersen, and K. H. Maier-Hein. nnu-net: a self-configuring method for deep learning-based biomedical image segmentation. *Nature methods.*, 18:203–211, 2021. 3
- [13] D.-H. Lee. Pseudo-label : The simple and efficient semi-supervised learning method for deep neural networks. *ICML 2013 Workshop : Challenges in Representation Learning (WREPL).*, 2013. 1, 3
- [14] S. Li, C. Zhang, and X. He. Shape-aware semi-supervised 3d semantic segmentation for medical images. In *Medical Image Computing and Computer Assisted Intervention–MICCAI 2020*, pages 552–561, 2020. 1, 3, 7, 8, 9
- [15] X. Li, H. Chen, X. Qi, Q. Dou, C.-W. Fu, and P.-A. Heng. H-denseunet: hybrid densely connected unet for liver and tumor segmentation from ct volumes. *IEEE transactions on medical imaging.*, 37:2663–2674, 2018. 3
- [16] X. Luo, J. Chen, T. Song, and G. Wang. Semi-supervised medical image segmentation through dual-task consistency. In *Proceedings of the AAAI conference on artificial intelligence*, volume 35, pages 8801–8809, 2021. 1, 3, 6, 7, 8, 9
- [17] X. Luo, W. Liao, J. Chen, T. Song, Y. Chen, S. Zhang, N. Chen, G. Wang, and S. Zhang. Efficient semi-supervised gross target volume of nasopharyngeal carcinoma segmenta-

- tion via uncertainty rectified pyramid consistency. In *Medical Image Computing and Computer Assisted Intervention – MICCAI 2021*, pages 318–329, 2021. 1, 4, 7, 8
- [18] F. Milletari, N. Navab, and S.-A. Ahmadi. V-net: Fully convolutional neural networks for volumetric medical image segmentation. In *2016 fourth international conference on 3D vision (3DV)*, pages 565–571, 2016. 3, 5, 6, 8, 9
- [19] M. E. J. Newman and M. Girvan. Abdominal multi-organ segmentation with organ-attention networks and statistical fusion. *Medical Image Analysis.*, 55:88–102, 2019. 1
- [20] O. Oktay, J. Schlemper, L. Folgoc, M. Lee, M. Heinrich, K. Misawa, K. Mori, S. McDonagh, N. Hammerla, B. Kainz, B. Glocker, and D. Rueckert. Attention u-net: Learning where to look for the pancreas. In *Medical Imaging with Deep Learning*, 2018. 3
- [21] O. Ronneberger, P. Fischer, e. N. Brox, Thomas”, J. Hornegger, W. M. Wells, and A. F. Frangi. U-net: Convolutional networks for biomedical image segmentation. In *Medical Image Computing and Computer-Assisted Intervention – MICCAI 2015*, pages 234–241, 2015. 3
- [22] H. R. Roth, L. Lu, A. Farag, H.-C. Shin, J. Liu, E. B. Turkbey, and R. M. Summers. Deeporgan: Multi-level deep convolutional networks for automated pancreas segmentation. In *Medical Image Computing and Computer-Assisted Intervention–MICCAI 2015*, pages 556–564, 2015. 6
- [23] Z. Shen, P. Cao, H. Yang, X. Liu, J. Yang, and O. R. Ziaiane. Co-training with high-confidence pseudo labels for semi-supervised medical image segmentation. In *Proceedings of the Thirty-Second International Joint Conference on Artificial Intelligence, ”IJCAI-23”*, pages 4199–4207, 2023. 4, 7, 8
- [24] T. Shi, H. Jiang, and B. Zheng. C2ma-net: Cross-modal cross-attention network for acute ischemic stroke lesion segmentation based on ct perfusion scans. *IEEE Transactions on Biomedical Engineering.*, 69(1):108–118, 2022. 1
- [25] Y. Shi, J. Zhang, T. Ling, J. Lu, Y. Zheng, Q. Yu, L. Qi, and Y. Gao. Inconsistency-aware uncertainty estimation for semi-supervised medical image segmentation. *IEEE transactions on medical imaging.*, 3:41, 2022. 1, 3, 7, 8, 9
- [26] K. Sohn, D. Berthelot, N. Carlini, Z. Zhang, H. Zhang, C. A. Raffel, E. D. Cubuk, A. Kurakin, and C.-L. Li. Fixmatch: Simplifying semi-supervised learning with consistency and confidence. *Advances in neural information processing systems.*, 33:596–608, 2020. 3
- [27] A. Tarvainen and 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. 1, 3, 5
- [28] J. E. Van Engelen and H. H. Hoos. A survey on semi-supervised learning. *Machine learning.*, 109:373–440, 2019. 3
- [29] T.-H. Vu, H. Jain, M. Bucher, M. Cord, and P. P’éréz. Advent: Adversarial entropy minimization for domain adaptation in semantic segmentation. In *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*, pages 2517–2526, 2019. 8, 9
- [30] J. Wang and T. Lukasiewicz. Rethinking bayesian deep learning methods for semi-supervised volumetric medical image segmentation. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 182–190, 2022. 1, 4, 7, 8, 9
- [31] K. Wang, B. Zhan, C. Zu, X. Wu, J. Zhou, L. Zhou, and Y. Wang. Triple-uncertainty guided mean teacher model for semi-supervised medical image segmentation. In *Medical Image Computing and Computer Assisted Intervention–MICCAI 2021*, pages 450–460, 2021. 1, 3, 7, 8, 9
- [32] Y. Wang, Y. Zhang, J. Tian, C. Zhong, Z. Shi, Y. Zhang, and Z. He. Double-uncertainty weighted method for semi-supervised learning. In *Medical Image Computing and Computer Assisted Intervention–MICCAI 2020*, pages 542–551, 2020. 1, 3, 7, 8, 9
- [33] H. Wu, X. Li, Y. Lin, and K.-T. Cheng. Compete to win: Enhancing pseudo labels for barely-supervised medical image segmentation. *IEEE Transactions on Medical Imaging.*, 42:3244–3255, 2023. 4, 8, 9
- [34] Y. Wu, Z. Wu, Q. Wu, Z. Ge, and J. Cai. Exploring smoothness and class-separation for semi-supervised medical image segmentation. In *International Conference on Medical Image Computing and Computer-Assisted Intervention*, pages 34–43, 2022. 1, 7, 8
- [35] Z. Xiong, Q. Xia, Z. Hu, N. Huang, C. Bian, Y. Zheng, S. Vesal, N. Ravikumar, A. Maier, X. Yang, and others. A global benchmark of algorithms for segmenting the left atrium from late gadolinium-enhanced cardiac magnetic resonance imaging. *Medical image analysis.*, 67:101832, 2021. 6
- [36] Z. Xu, Y. Wang, D. Lu, X. Luo, J. Yan, Y. Zheng, and R. K.-y. Tong. Ambiguity-selective consistency regularization for mean-teacher semi-supervised medical image segmentation. *Medical Image Analysis.*, 88:102880, 2023. 4, 7, 8
- [37] X. Yi, E. Walia, and P. Babyn. Generative adversarial network in medical imaging: A review. *Medical Image Analysis.*, 58:101552, 2019. 1
- [38] L. Yu, S. Wang, X. Li, C.-W. Fu, and P.-A. Heng. Uncertainty-aware self-ensembling model for semi-supervised 3d left atrium segmentation. In *Medical Image Computing and Computer Assisted Intervention–MICCAI 2019*, pages 605–613, 2019. 1, 3, 6, 7, 8, 9
- [39] Y. Zhang, L. Yang, J. Chen, M. Fredericksen, D. P. Hughes, and D. Z. Chen. Deep adversarial networks for biomedical image segmentation utilizing unannotated images. In *Medical Image Computing and Computer Assisted Intervention–MICCAI 2017*, pages 408–416, 2017. 8, 9
- [40] T. Zhao, K. Cao, J. Yao, I. Noguez, L. Lu, L. Huang, J. Xiao, Z. Yin, and L. Zhang. 3d graph anatomy geometry-integrated network for pancreatic mass segmentation, diagnosis, and quantitative patient management. In *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*, pages 13743–13752, 2021. 1
- [41] Z. Zhou, M. M. R. Siddiquee, N. Tajbakhsh, and J. Liang. Unet++: Redesigning skip connections to exploit multiscale features in image segmentation. *IEEE transactions on medical imaging.*, 39:1856–1867, 2019. 3