Mathematical Modeling and Algorithm Application
ISSN: 3006-0842 | Vol. 2, No. 1, 2024

Dual Task Semi-supervised Pancreatic Segmentation Based on Prior Information and Multiple Regularization

Songcai Yan, Xinjun Hu, Qinyuan Xue

School of Mechanical Engineering, Sichuan University of Science and Engineering, 64300, Sichuan, China

Abstract: Due to the relatively small size and complex internal structure of the pancreas, the segmentation is often inaccurate during image processing. A more effective automatic segmentation method is proposed to solve this problem. A multi-task deep neural network architecture based on V-Net architecture is proposed. By capturing the relationship between the prior positions of the pancreas, the target of the pancreas can be constrained at the regional level. In addition, this study uses dual task training methods to simultaneously perform segmentation tasks and regression tasks, and generate high-quality pseudo-label graphs to better utilize the valid information in a large number of unlabeled data. At the same time, this study also introduces the idea of consistency regularization, which uses the consistency regularization of prior information and the consistency regularization of noise disturbance and network disturbance to optimize the segmentation network between double decoders and double tasks, so as to further improve the segmentation effect and generalization ability of the model. Experiments show that compared with the benchmark method, the Dice coefficient of the method in this study is improved by 5.40% (for 10% labeled data) and 3.46% (for 20% labeled data) respectively, which proves the efficiency of the method in processing unlabeled medical images described in this chapter.

Keywords: Convolutional neural network; Medical image segmentation; Multivariate regularization; Semi-supervised learning.

1. Introduction

Pancreatic cancer is a highly lethal malignancy known as the "king of cancer," with a 5-year survival rate of less than 10 percent, according to the American Cancer Society[1]. In the process of diagnosis and treatment of pancreatic cancer, accurate automatic pancreatic segmentation is very important for quantitative evaluation and computer-aided diagnosis, which helps clinicians to quickly detect and diagnose pancreatic cancer, reduce the workload of physicians, and improve diagnostic efficiency. Traditional image segmentation is a method based on pixel color, brightness and other features, by dividing the image into different areas to achieve image segmentation. Common algorithms used in this method include threshold segmentation [2,3], edge detection [4,5], region growth [6,7], etc. Traditional image segmentation is sensitive to noise, and is not effective for complex images or images with multiple objects. In the process of diagnosis and treatment of pancreatic cancer, accurate automatic pancreatic segmentation is very important for quantitative evaluation and computer-aided diagnosis, which helps clinicians to quickly detect and diagnose pancreatic cancer, reduce the workload of physicians, and improve diagnostic efficiency. Traditional image segmentation is a method based on pixel color, brightness and other features, by dividing the image into different areas to achieve image segmentation. The common algorithms of this method include threshold segmentation, edge detection, region growth and so on. Traditional image segmentation is sensitive to noise, and is not effective for complex images or images with multiple objects.

In recent years, deep learning has become one of the hot research topics in medical image segmentation, which mainly uses Convolutional Neural Networks (CNNS) to extract and segment image features, and has high accuracy and robustness. In order to better improve the performance of pancreatic segmentation, a large number of researchers continue to explore and try methods, and have achieved certain results, but there are still some problems to be solved. Li et al [8] first used convolutional neural networks to extract features, and then used convolutional long short-term memory networks to obtain slice context information for pancreas segmentation. However, slicing context information cannot be shared across sequences and parallelism, and there is a problem of information loss in forward propagation. Zhou [9] et al. proposed a fixed-point model method, which first predicted the pancreatic region boundary, reduced the input resolution, and then implemented a fine segmentation strategy. Although this strategy can improve the segmentation accuracy in theory, its implementation process is relatively complicated, and due to the two predictions from coarse to fine, the segmentation results may be more volatile. Schlemper et al[10] proposed an attentional gate module embedded in Unet to eliminate the need for localization, but this mechanism failed to pay attention to the pancreas's clustering features on image space. By using two decoders with different depths in Unet, Bi Xiuli et al[11] solved the problem that a single decoder lost pancreatic location and detail information when decoding deeper encoding features. This method can improve the accuracy of pancreas segmentation, but can not fully capture the spatial structure and context of the pancreas. However, due to the instability of the adversarial network, the training and test results will also fluctuate greatly.

All of the above methods have their own characteristics, but also have their own limitations, mainly in the lack of a combination of the characteristics of the pancreas itself to think deeply about the segmentation network. Due to the

small proportion of pancreas in abdominal CT images and the irregular shape and boundary, the existing pancreas segmentation network can extract strong semantic information to identify pancreas and spleen through subsampling in the coding stage, which will lead to the loss of feature location and detail information of pancreas, and seriously affect the accuracy of boundary pixels in segmentation results. Therefore, for the problem of pancreas segmentation, this study proposes a more advanced end-to-end deep learning technology and a multi-task deep neural network structure based on V-Net architecture. By capturing the relationship between the prior positions of the pancreas, the target of the pancreas can be constrained at the regional level. In addition, this study uses dual task training methods to simultaneously perform segmentation tasks and regression tasks, and generate high-quality pseudo-label graphs to better utilize the valid information in a large number of unlabeled data. At the same time, this study also introduces the idea of consistency regularization, which uses the consistency regularization of prior information and the consistency regularization of noise disturbance and network disturbance to optimize the segmentation network between double decoders and double tasks, so as to further improve the segmentation effect and generalization ability of the model.

2. Data Sets and Preprocessing

2.1. Sources and characteristics of NIH data sets

In this study, a publicly available dataset provided by the National Institutes of Health was used. The dataset contained the results of 82 abdominal enhanced 3D CT scans of 80 subjects, including 53 men and 27 women. The volume of CT image is 512×512×D, where D∈[181,466], indicating the number of slice samples along the long axis of the human body, and the thickness ranges from 0.5mm to 1.0mm. Each sample was manually labeled and examined and corrected by an experienced radiologist. This dataset is publicly available and is widely used to evaluate the performance of pancreatic CT segmentation frameworks.

2.2. Data set preprocessing

In order to better present the features of the pancreas, this study adjusted the window width value of the downloaded NIH 3D image data to 45 and window position value to 325, and processed the 3D image data according to the Z-axis (cross section), and carried out ROI clipping on the images of each slice, and the clipping size resolution was [240,240]. The sections that did not contain the pancreas were also removed. In order to enhance the generalization ability of the network, this study carried out data enhancement processing on the cropped pictures. The pictures of the training set and the verification set were randomly rotated from -25° to +25° and randomly flipped horizontally according to the threshold condition of 0.5. All images used for training and testing are single-channel grayscale images with a bit depth of 8. Finally, when the data set is loaded, the data is normalized. The final amount of data used for training and testing was 14,110 images.

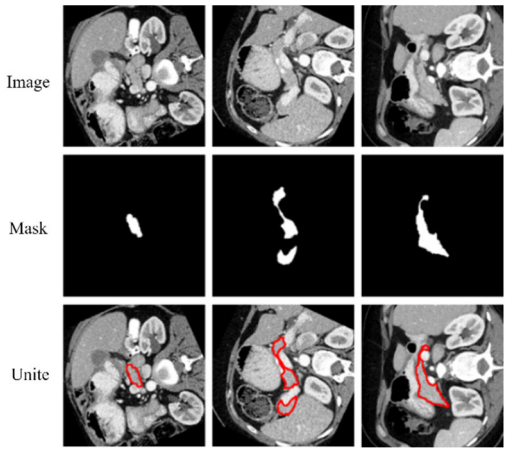

Figure 1. Samples and labels for training and testing purposes

3. Textual Method

3.1. Model design

PIMR-Net (Prior Information and Multivariate Regularization for Dual-Task Segmentation) is designed in this chapter based on prior information and multivariate regularization for dual-task segmentation The overall structure of Consistency Network (PIMR-Net) is shown in Figure 2. The V-Net[12,13] network structure consists of encoder and decoder. The deep feature map is fused with the shallow feature map by upsampling operation, so as to compensate for the feature loss caused by pooling operation in the encoder and accelerate the learning of the feature of target segmentation. This chapter uses two V-net network models, model 1 V-Net and model 2 VNet, which contain decoders A and B, respectively, which are independent of each other. In this paper, random noise perturbations δ and δ 'are added to the output of the encoder and passed to the decoder A of model 1 and to the decoder B of model 2, respectively. The two decoders produce prediction results at the same time after each iteration. In the training process of the model, the ultimate goal is to improve the generalization and noise resistance of the model by reducing the difference between the results of the two parts.

In the V-Net[12,13]structure, composed of encoders and decoders, the network fuses the deep and shallow feature maps through up-sampling operations to compensate for the feature information that may be lost due to pooling operations in the encoder, and speeds up the learning process of target segmentation features. In this study, two V-net models are constructed, namely model 1 V-Net and model 2 V-Net, each of which contains an independent decoder A and decoder B. In the encoder output of these two models, random noise perturbations δ and δ 'are added and then input into encoder A of model 1 and encoder B of model 2, respectively. During each iteration, the encoder output is then passed through the decoder to produce the final prediction result. During the training process, the goal is to improve the generalization performance and noise resistance of the model by reducing the difference in the prediction results of the two parts.

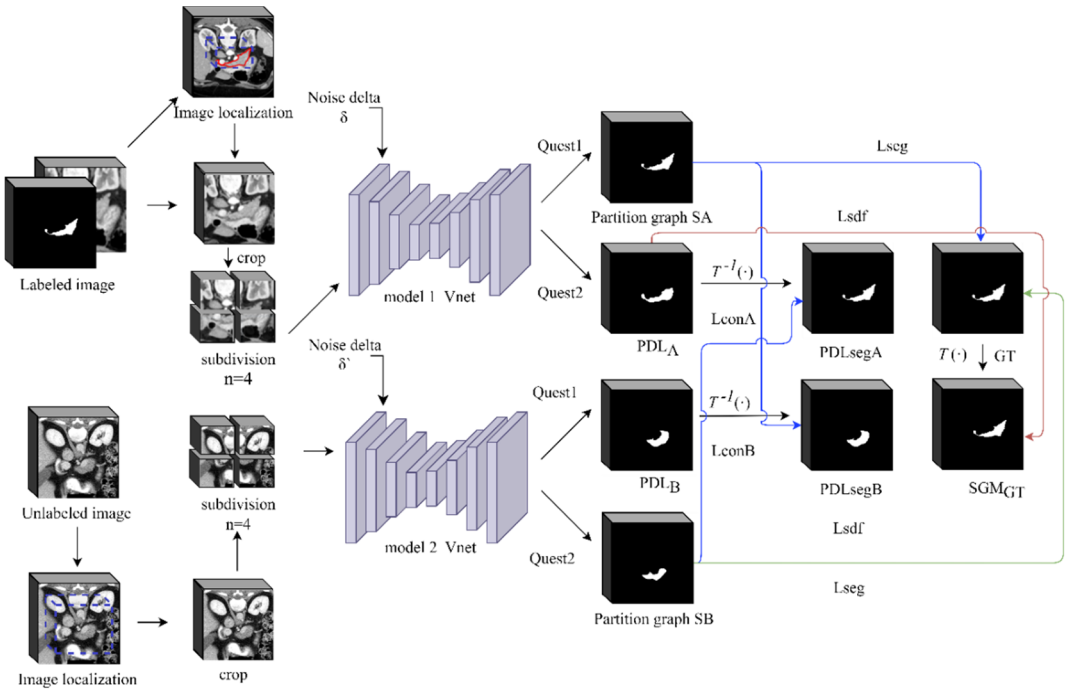

Figure 2. Overall structure of the model

Since the feature distribution of medical image samples after preprocessing is relatively similar, this study considers adding geometric prior information to constrain the model to further improve its learning ability of segmented samples. In this study, a shape sensing regression layer was added at the end of each decoder, and in the training phase, each decoder performed two tasks simultaneously: the segmentation task and the regression task. Therefore, in each iteration, the network outputs A total of four parts: the segmentation graph SA output by model 1 decoder A, the pseudo-label PDLA, the segmentation graph SB output by model 2 decoder B, and the pseudo-label graph PDLB. The ability of the model to learn effective features of the segmented target can be enhanced by implementing the dual task consistency self-constraint and cross-constraint between decoders for each decoding part.

3.2. Prior information pancreatic target localization

This paper introduces a set of symbols to define semi-supervised segmentation tasks. The data set contains two subsets $D = D_L \bigcup D_U$ and $D_L \bigcap D_U = \phi$. A set of symbols is introduced to define semi-supervised segmentation tasks. The data set contains two subsets and sums. Where $D_L = \{(X_i, Y_i)\}_{i=1}^{N_L}$ is the labeled data set and $D_U = \{X_i\}_{i=1}^{N_U}$ is the unlabeled data set. X and Y represent images and labels, respectively. The goal of semi-supervised learning is to improve the performance of the training model using unlabeled data set D_U on labeled data set D_L .

Considering that there is a serious imbalance between foreground and background classes in pancreatic segmentation, it is necessary to extract the region of interest and remove the interference of irrelevant information, as shown in Figure 3. The specific steps are as follows:

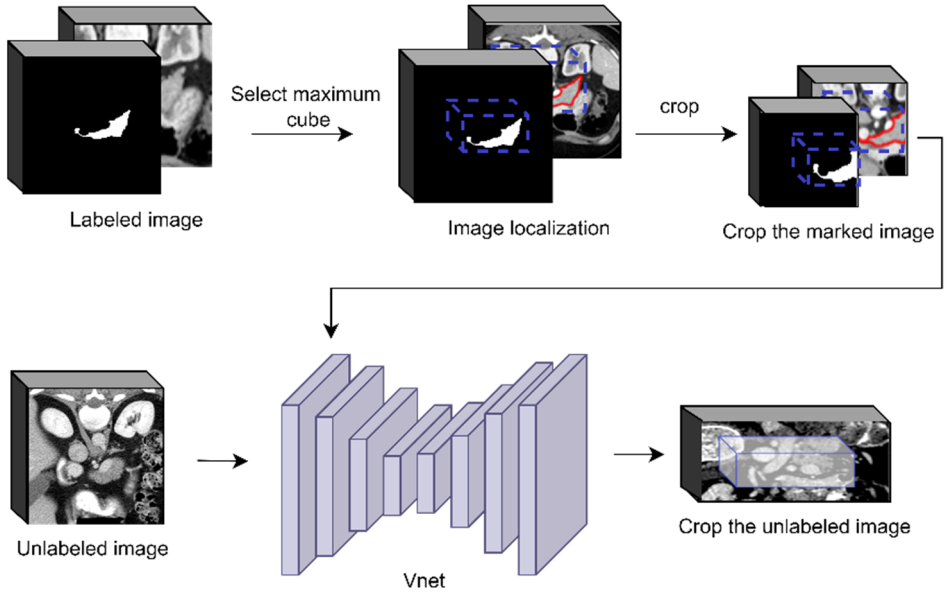

Figure 3. Orientation of prior information

1. The labeled images were preprocessed in this paper. First, the size of the cube containing the pancreatic region in each CT volume was calculated, and then the largest cube was selected so that it could contain all the pancreatic region in the data set. Then, the marked image set is cropped according to the largest cube, and the following data set $X, Y \in \mathbb{R}^{V' \times H' \times W' \times 1}$ is obtained, Where V', H' and W' are the number, length and width of slices of marked data respectively. The model is then pre-trained using this data set.

2. Due to individual differences, the axial plane span fluctuates greatly. By observing that the position of pancreas in 3D CT images has a certain range, the distribution information of labeled image data can be used to impose prior constraints on unlabeled images. Using the pre-trained model in 1) to predict the unlabeled data set D_U, roughly locate the span of the axial plane of the pancreas, and then sample according to the cube sizes H' and W'. The processed unlabeled data is then extended bidirectionally to V' and X^U is obtained.

3. Considering that the pancreas has certain morphological characteristics in the human body, a related hypothesis is proposed: organs in the same position have similar characteristics in different patients. Therefore, the labeled and unlabeled images are divided into n block subregions of the same size, and constraints are imposed on the corresponding positions of the subregions. The size of the partition subregion is V', $\frac{H'}{n}$, and $\frac{W'}{n}$, as shown in Figure 2 (n=4).

3.3. Consistent regularization based on prior information

Training architecture of dual task consistency fusion into multiple decoders. As the overall framework above is shown in Figure 2, in each decoder, there are two parallel execution tasks. These two tasks include generating pixel-level segmentation probability plots and geometric shape plots in traditional regression tasks represented by level set functions [14], respectively. In this section, the level set function is used as the expression method of 3D object contour by using prior knowledge. This function can map the space point X on the input 3D image to one dimensional space, and then generate pseudo-label graph PDL corresponding to the input image. Since the segmentation results can be represented by both pixel-level label maps and global geometric contours, the two prediction tasks need to be synchronized. In order to make more efficient use of unlabeled data, this study incorporates a dual-task consistency strategy into the semi-supervised learning framework and applies this strategy between two decoders. In this paper, we minimize the cross-consistency loss between two decoders and improve the segmentation performance of the model by using geometric prior knowledge.

In order to ensure the consistency between decoder A of model 1 and decoder B of model 2 on geometric constraints, this study uses transformation function $T^{-1}(\cdot)$ to convert pseudo-label graphs PDL_A and PDL_B generated by regression task, and obtains corresponding pixels-level segmentation probability graphs PDL_{segA} and PDL_{segB}. This process is accomplished by smoothing the Heaviside function. In the method of this study, the conversion layer T(·) used to convert the real label GT into the standard segmentation graph SGM_{GT} is defined as follows:

$$T(x) = \begin{cases} -\inf\limits_{y \in \partial D} \| x - y \|_2 & x \in D_i \\ 0 & x \in \partial D \\ +\inf\limits_{y \in \partial D} \| x - y_2 \| & x \in D_o \end{cases} \quad (1)$$

In the formula discussed, $\|x - y\|$ represents the Euclidean distance connecting two different voxel points x and y, and inf represents determining the smallest possible upper bound for that distance value. The variable ∂D represents the zero isometric line, that is, the state of no distance, which is used to describe the outer edge contour of the segmented target. Di refers to the area inside the split target, and Do refers to the area outside the split target. Therefore, in most cases, when the pixel is inside the segmentation target, the value of the conversion will become negative, and conversely, if it is outside, it will take a positive number. In view of the diversity of input image samples in perspective and segmentation target volume, the sigmoid activation function is further used to normalize T(·), so that the value of each input sample is limited to the interval of [-1, 1] [13]. By applying the transformation formula T(·), the training sample GT with labeled data can be converted into the corresponding standard segmentation graph SGM_{GT}. In addition, since the transform itself does not have differentiation, it is impossible to convert pseudo-label graphs PDL_A and PDL_B into segmentation probability graphs directly by using its inverse operation. For this reason, it is necessary to adopt a smooth approximation method to carry out the inverse transformation, the specific process is as follows:

$$T^{-1}(z) = \frac{1}{1 + e^{-k \cdot z}}, z \in SGM \quad (2)$$

In the formula, z represents the value of voxel x in the pseudo-label graph PDL. In order to achieve smooth approximate change effect, the transformation factor k should be selected as large as possible. Through the above transformation and reverse transformation formula, the consistency optimization operation can be carried out on the dual-task output in the model.

3.4. Consistent regularization based on noise perturbation and network perturbation

In the semi-supervised segmentation network framework mentioned above, we design a two-task prediction scheme based on two V-Net structures. In the scheme introduced in this chapter, the input for training the model is data-enhanced 3D CT sample $X \in \mathbb{R}^{H \times W \times D}$. The sample set consists of N labeled samples and M unlabeled samples, and the number of labeled samples is far less than that of unlabeled samples, accounting for no more than 20% of all samples. In this paper, it is defined that the labeled data fields are

$D^l = \{X_n, Y_n, Z_n\}_{n=1}^{N}$ and $D^u = \{X_m\}_{m=N+1}^{N+M}$, and the unlabeled data fields are, and the shared encoder performs N random forward propagation together with decoder A of mode 1 and decoder B of model 2 with added noise perturbation. As shown in Figure 4, when marked sample $X_n \in \mathbb{R}^{H\times W\times D}$ is input, $Y_n \in \{0,1\}^{H\times W\times D}$ is the true segment label of the sample, and $Z_n \in \mathbb{R}^{H\times W\times D}$ is the standard segment graph $SGMG$T converted from the real segment label (GT) Y_n of the sample through the level set function T(·).

During the training process, decoder A of model 1 and decoder B of model 2 backpropagate independently of each other by learning to add training samples with different perturbations. In addition, since the two decoders in the network model also introduce Dropout layers for perturbation at the model level, and the optimization process of the parameters of the neural network model is also random, both decoders can optimize the parameters of the shared encoder during backpropagation, thus improving the encoder's ability in feature extraction. And enhance the ability of the whole model to perturbation and generalization. In this way, the data level and model level perturbation are combined.

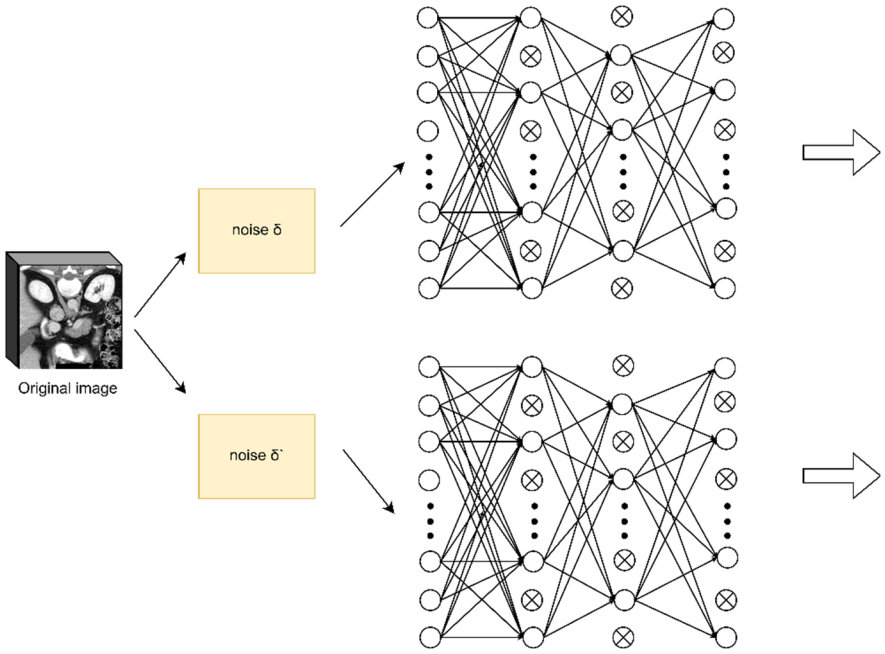

Figure 4. Perturbation operation in model training

4. Experiment and Discussion

4.1. Experimental setup

In this experiment, 13th Gen Intel(R) CoreTM i7-13700KF×24 processor and NVIDIA GeForce RTX 4070 Ti single independent graphics card were used for model training. The deep learning framework used by all network models in this chapter is PyTorch. Using Python as the programming language. In this study, 3D V-Net was used as the backbone. The noise perturbation of the decoder in the model δ is set to [-0.1, 0.1]. The two V-Net backbones share the same parameters. For more environment configurations, see Table 1.

Table 1. Environment configuration description

Name	Hardware model and software version
Operating system	Linux
CPU	13th Gen Intel(R) Core™ i7-13700KF×24
GPU	NVIDIA GeForce RTX 4070 Ti
Internal memory	16GB
Video memory	11GB
CUDA	CUDA 12.2
cuDNN	cuDNN 8.9
Deep learning framework	PyTorch 1.12.0；
Development language	Python 3.7.16
Compiler and environment configuration	Pycharm 2023; Anaconda3
Image processing library	OpenCV 4.7.0; SimpleITK 2.0.2

4.2. Evaluation index and loss function

For image segmentation tasks, the following indicators can be selected as evaluation indicators, the following letters TP represents true cases, FP represents false positive cases, TN represents true negative cases, FN represents false negative cases. The above four terms can be used to calculate the four commonly used medical image segmentation evaluation indicators in pancreatic image analysis, which are Dice coefficient, Intersection Over Union, intersection over union.

IoU, Average surface distance (ASD), and Hausdorff Distance (HD). Its calculation method is shown in formula (3)~(6) :

$$Dice = \frac{2*TP}{2*TP+FP+FN} \tag{3}$$

$$IoU = \frac{TP}{TP+FP+FN} \tag{4}$$

$$ASD = \frac{1}{S(A)+S(B)}\left(\sum_{s_A \in S(A)} d(s_A, S(B)) + \sum_{s_B \in S(B)} d(s_B, S(A))\right) \tag{5}$$

Where S(A) represents the surface voxels in set A, and d(v,S(A)) represents the shortest distance from any voxel to S(A).

$$H^{f_F f_R}(A,B) = \max\{h^{f_F}(A,B), h^{f_R}(B,A)\} \tag{6}$$

In the formula, $f_F, f_R \in [0,1]$ The forward and backward scores, respectively, control forward distance and backward distance.

4.3. Loss function

In order to maintain consistent optimization between data models and tasks, this paper adopts a two-part loss function. One of them is the loss function Lseg and Lsdf used in supervised training on labeled samples. The other is the unsupervised training loss functions LconA and LconB applied to all training samples containing unlabeled data. Wherein, for the labeled training sample $D^l = \{X_n, Y_n, Z_n\}_{n=1}^N$, the voxel (x, y)∈(X, Y) in the 3D sample is defined. Two pixel-level prediction probability graphs output by model 1 decoder A and model 2 decoder B during the training process generate segmentation graph SA and segmentation graph SB by sigmoid function. Dice loss function is used to calculate the loss between the segmentation graph and the real label, which is calculated as follows:

$$L_{\text{seg}}(x,y) = \frac{1}{2}\left(\sum_{x_i, y_i \in D^l}\left(1 - \frac{2\times \sum_{x_j \in x_i, y_j \in y_i} f_{A1}(x_i)\times y_i}{2\times \sum_{x_j \in x_i, y_j \in y_i} f_{A1}(x_j) + \sum_{y_j \in y_i} y_j}\right)\right. \\ \left.\sum_{x_i, y_i \in D^l}\left(1 - \frac{2\times \sum_{x_j \in x_i, y_j \in y_i} f_{B1}(x_i)\times y_i}{2\times \sum_{x_j \in x_i, y_j \in y_i} f_{B1}(x_j) + \sum_{y_j \in y_i} y_j}\right)\right) \tag{7}$$

In the formula, The variables $\Sigma_{x_j \in x_i, y_j \in y_i}$ and $\Sigma_{x_i, y_i \in D^l}$ refer to the voxels in the 3D image sample and the labeled samples in the training dataset, respectively. $f_{A1}(\cdot)$ and $f_{B1}(\cdot)$ represent the segmentation graph SA and the segmentation graph SB generated by the decoder. At the same time, the decoder also generates pseudo-label graphs PDL_A and PDL_B, which will be compared with the standard label SGM_{GT} obtained by the inverse transformation function $T(\cdot)$ to calculate the loss, which is calculated as follows:

$$L_{\text{sdf}}(x,y) = \sum_{x_i, y_i \in D^l} \|f_{A2}(x_i) - T(y_i)\|^2 + \sum_{x_i, y_i \in D^l} \|f_{B2}(x_i) - T(y_i)\|^2 \tag{8}$$

In the formula， $f_{A2}(\cdot)$ and $f_{B2}(\cdot)$ Represents the PDL_A and PDL_B output of the decoder. Thus, the total loss function of the supervised part is defined as:

$$L_{\text{sup}} = L_{\text{seg}} + \beta L_{\text{sdf}} \tag{9}$$

Where β represents a parameter weighing two loss terms. In order to efficiently utilize a large number of unlabeled sample data, this study adopts a strategy that combines the consistency of the task with the consistency of the model. This strategy uses a consistency loss function to mutually constrain the output of two independent decoders on different tasks, which is applicable to all training data. In the optimization process, this study aims at the probability graph SA generated by the segmentation task of decoder A in model 1 and the pseudo-label graph PDL_B generated by the regression task of decoder B in model 2. As well as the consistency optimization between the probability graph SB generated by the segmentation task of decoder B in model 2 and the pseudo-label graph PDL_A generated by the regression task of decoder A in model 1, the specific calculation method is as follows:

$$L_{\text{conA}}(x) = \sum_{x_i \in D} \| f_{A1}(x_i) - T^{-1}(f_{B2}(x_i)) \|^2 \tag{10}$$

$$L_{\text{conB}}(x) = \sum_{x_i \in D} \| f_{B1}(x_i) - T^{-1}(f_{A2}(x_i)) \|^2 \tag{11}$$

$$L_{\text{cross}} = L_{\text{conA}} + L_{\text{conB}} \tag{12}$$

In the formula, $\sum_{x_i \in D}\cdot$ Represents all samples in the training set. It can be concluded from the above that the overall loss function of the semi-supervised image segmentation model designed in this chapter is:

$$L = L_{\text{sup}} + \alpha L_{\text{cross}} \tag{13}$$

In the formula， α is a hyperparameter that balances the weight between supervised and unsupervised training losses to stabilize the overall training process. In this chapter, the annealing strategy of the time-dependent Gaussian warm-up function [15] is adopted to gradually increase the value of α.

4.4. Experimental result

Methods in this chapter are divided as follows: (1) Baseline (BL): three-dimensional class represents semi-supervised medical image methods [16]; (2) PLPI: pancreatic target localization method based on prior information; (3) CRPI: consistency regularization of prior information. (3) CRNN: consistency regularization based on noise perturbation and network perturbation. In order to verify the effectiveness of the prior information pancreatic target positioning method PLPI, the prior information consistent regularization CRPI, the noise disturbance consistent regularization CRNN method

and the network disturbance consistent regularization CRNN method, as well as the superposition of these methods, this paper compares the same segmentation case and presents the segmentation results in a centralized manner. The segmentation results of the centralized method are shown in Figure 5.

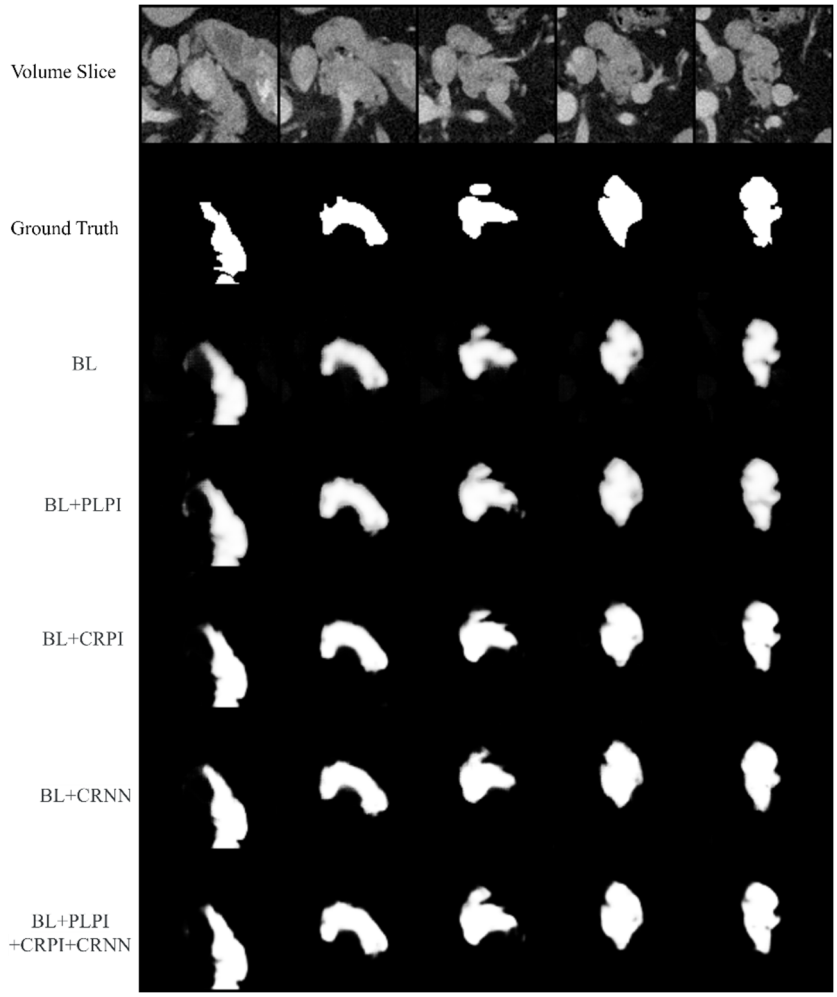

Figure 5. The results were divided by different methods

Note: BL represents a three-dimensional class representing a semi-supervised medical image method [16]. BL+PLPI represents the fusion of prior information pancreatic target location and BL. BL+CRPI represents the fusion of the consistent regularization of prior information and BL. BL+CRNN represents the consistency regularization of noise perturbations and network perturbations and the fusion of BL. BL+PLPI+CRPI+CRNN represents the set of the above methods, namely PIMR-Net, the best model in this paper.

The comparative experimental results of different methods are shown in Table 2. The experiments were carried out with 6 labeled data and 12 labeled data respectively.

Table 2. Ablation experiments using different labeling numbers and different methods

Setting		measures				evaluation			
labelled	unlabelled	BL	PLPI	CRPI	CRNN	Dice (%)	IoU(%)	ASD(mm)	95HD(mm)
6	56	√				71.81	57.36	5.60	20.27
6	56	√	√			74.74	60.80	2.91	11.98
6	56	√		√		72.81	58.27	5.34	16.41
6	56	√			√	75.97	62.74	3.09	13.07
6	56	√	√	√	√	77.21	63.51	2.94	12.10
12	50	√				79.26	66.41	4.88	14.58
12	50	√	√			81.23	68.86	2.10	9.17
12	50	√		√		79.89	67.56	1.22	6.37
12	50	√			√	80.53	67.75	1.19	5.76
12	50	√	√	√	√	82.72	70.86	1.15	4.86

As can be seen from Table 2, after the prior position is added to the method of three-dimensional class representation of semi-supervised medical images, each evaluation index can obtain greater benefits. In particular, the Dice index increased from 71.81% to 74.74% with 6 marked numbers, an increase of 2.93%, and from 79.26% to 81.23% with 12 marked numbers, an increase of 1.97%. It is proved that the unlabeled data can be used more effectively for pancreatic

target localization based on prior information, and adding the decoder with prior information pancreatic target localization can make the model have better overall segmentation performance. Experimental results of BL and BL+CRPI in Table 2 show that Dice coefficient increases by 1.00% and IoU index increases by 0.91% on 6 labeled data sets. In 12 datasets with annotated data, the Dice coefficient increased by 0.63%, while the IoU also increased by 1.15%. This indicates that the consistent regularization method using prior information can improve the effect of pancreatic segmentation to a certain extent. The results of BL and BL+CRNN in Table 2 show that the ASD index is decreased by 2.51mm and the 95HD index is decreased by 7.20mm on the six labeled datasets. In 12 datasets with labeled data, the ASD index is decreased by 3.69mm, and the 95HD index is also decreased by 8.82mm, indicating the consistency regularization of noise disturbance and network disturbance, which makes the model more generalized. When the above three methods are used together in this paper, that is, the best model PIMR-Net proposed in this paper increases the Dice coefficient by 5,4% compared with BL on 6 labeled data sets, and the ASD index decreases by 2.66mm. On the 12 labeled datasets, Iou increased by 4.45% and 95HD decreased by 9.72 mm relative to BL. Through the above ablation comparison experiments, it can be shown that no matter what method is effective for improving the segmentation of pancreas, especially in PIMR-Net proposed in this paper, the generalization performance of the model and the effect of the semi-supervised segmentation model of pancreas can be improved to a greater extent.

The visualization of 3D segmentation results between the methods in this chapter and several advanced segmentation methods in recent years is shown in Figure 6. It can be visually seen from the figure that under the conditions of 6 labeled samples (10% labeled data) and 12 labeled samples (20% labeled data), the segmentation effect of PIMR-Net proposed in this chapter is more overlapping with the real labels labeled by experts in the same test sample in the test set which is fuzzy and difficult to segment compared with other samples. A more complete pancreas shape and more details are predicted, especially at the boundary of the segmentation target, which further proves the validity of the dual-task semi-supervised pancreas network based on prior information and multivariate regularization proposed in this chapter.

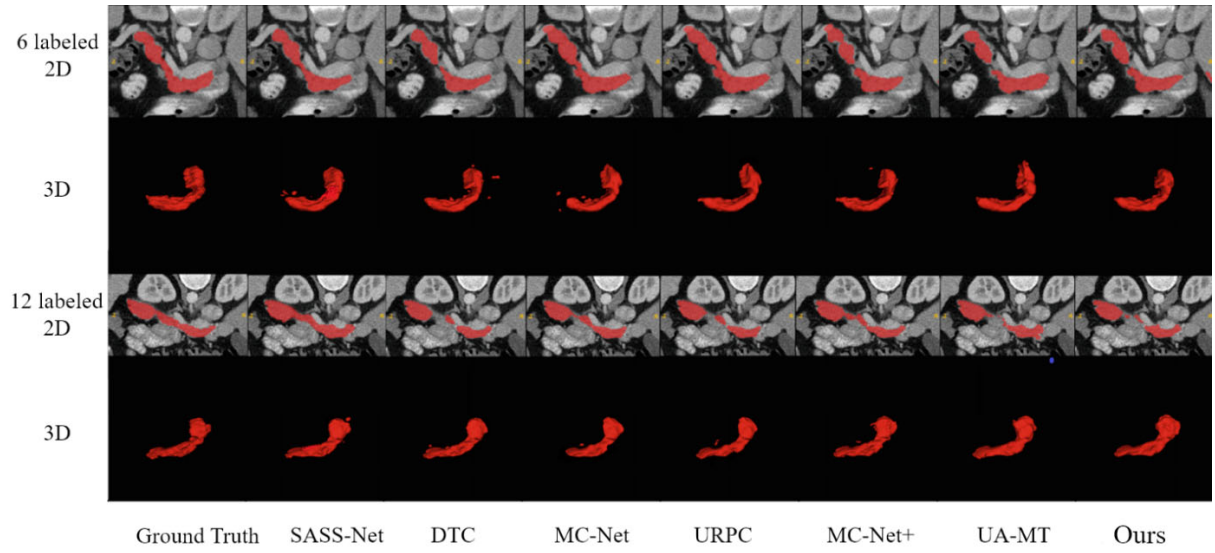

Figure 6. 3D visualization of segmentation results by different segmentation methods

5. Conclusion

In this paper, we propose a multi-task deep neural network architecture based on V-Net architecture. By capturing the relationship between the prior positions of the pancreas, the target of the pancreas can be constrained at the regional level. In addition, this study uses dual task training methods to simultaneously perform segmentation tasks and regression tasks, and generate high-quality pseudo-label graphs to better utilize the valid information in a large number of unlabeled data. At the same time, this study also introduces the idea of consistency regularization, which uses the consistency regularization of prior information and the consistency regularization of noise disturbance and network disturbance to optimize the segmentation network between double decoders and double tasks, so as to further improve the segmentation effect and geeralization ability of the model. Experiments show that compared with the benchmark method, the Dice coefficient of the method in this study is improved by 5.40% (for 10% labeled data) and 3.46% (for 20% labeled data) respectively, which proves the efficiency of the method in processing unlabeled medical images described in this chapter. Compared with other algorithms, the algorithm proposed in this chapter performs better and can perform automatic segmentation of pancreatic targets in CT images.

References

[1] Siegel Rebecca L, Miller Kimberly D, Fuchs Hannah E, et al. Cancer statistics, 2021[J]. Ca Cancer J Clin, 2021, 71(1): 7-33.

[2] Anitha J, Pandian S Immanuel Alex, Agnes S Akila. An efficient multilevel color image thresholding based on modified whale optimization algorithm[J]. Expert Systems with Applications, 2021, 178: 115003.

[3] Lu Jianhong, LIU Haipeng, Wang Meng. Multi-threshold image segmentation algorithm based on improved Gull algorithm [J]. Journal of Optoelectronics Laser, 2022, 33: 932-939.

[4] Fakeeha Jafari, Aamer Nadeem, Uz Zaman Qamar. Evaluation of Metamorphic Testing for Edge Detection in MRI Brain Diagnostics [J]. Applied Sciences, 2022, 12(17).

[5] Tang Shoufeng, Zhai Shaoqi, Tong Guangming, et al. Edge detection based on improved Canny operator and morphological fusion [J]. Computer Engineering and Design, 2023, 44: 224-231.

[6] Siyoum Biratu Erena, Friedhelm Schwenker, Girma Debelee Taye, et al. Enhanced Region Growing for Brain Tumor MR Image Segmentation [J]. Journal of Imaging, 2021, 7(2).

[7] Liu Ying, LI Jun, Yang Haima, et al. Cervical cancer image segmentation based on region growth and level set algorithm [J]. Chinese Journal of Electronic Measurement and Instrumentation, 2020, 34: 146-152.

[8] Li Jun, Lin Xiaozhu, Che Hui, et al. Probability map guided bi-directional recurrent UNet for pancreas segmentation[J]. arXiv preprint arXiv:.00923, 2019.

[9] Zhou Yuyin, Xie Lingxi, Shen Wei, et al. A fixed-point model for pancreas segmentation in abdominal CT scans [C]//International conference on medical image computing and computer-assisted intervention, 2017: 693-701.

[10] Schlemper Jo, Oktay Ozan, Schaap Michiel, et al. Attention gated networks: Learning to leverage salient regions in medical images[J]. Medical image analysis, 2019, 53: 197-207.

[11] Bi Xiu-li, LU Meng, XIAO Bin, et al. Pancreas segmentation based on double-decoded U-convolutional neural networks [J]. Journal of Software, 2022, 33(5): 1947-1958.

[12] Milletari Fausto, Navab Nassir, Ahmadi Seyed-Ahmad. V-net: Fully convolutional neural networks for volumetric medical image segmentation[C]//2016 fourth international conference on 3D vision (3DV), 2016: 565-571.

[13] Su Yixin, Xiao Zhiyong. Semi-supervised dual-task cross-consistency constraint network for 3D cardiac MRI segmentation [J]. Chinese Journal of Image and Graphics, 2023, 28: 1198-1211.

[14] Luo Xiangde, Chen Jieneng, Song Tao, et al. Semi-supervised medical image segmentation through dual-task consistency [C]//Proceedings of the AAAI conference on artificial intelligence, 2021: 8801-8809.

[15] Tarvainen Antti, Valpola Harri. Mean teachers are better role models: Weight-averaged consistency targets improve semi-supervised deep learning results[J]. Advances in neural information processing systems, 2017, 30.

[16] Han Kai, Liu Lu, Song Yuqing, et al. An effective semi-supervised approach for liver CT image segmentation[J]. IEEE Journal of Biomedical, 2022, 26(8): 3999-4007.