Semi-supervised cardiac MRI image of the left ventricle segmentation algorithm based on contrastive learning*

ZHU Enrong[1], ZHAO Haochen[2], and HU Xiaofei[1]**

1. *School of Geography and Bioinformatics, Nanjing University of Posts and Telecommunications, Nanjing 210003, China*
2. *School of Communication and Information Engineering, Nanjing University of Posts and Telecommunications, Nanjing 210003, China*

(Received 20 January 2022; Revised 7 May 2022)

A semi-supervised convolutional neural network segmentation method of medical images based on contrastive learning is proposed. The cardiac magnetic resonance imaging (MRI) images to be segmented are preprocessed to obtain positive and negative samples by labels. The U-Net shrinks network is applied to extract features of the positive samples, negative samples, and input samples. In addition, an unbalanced contrastive loss function is proposed, which is weighted with the binary cross-entropy loss function to obtain the total loss function. The model is pre-trained with labeled samples, and unlabeled images are predicted by the pre-trained model to generate pseudo-labels. A pseudo-label post-processing algorithm for removing disconnected regions and hole filling of pseudo-labels is proposed to guide the training process of semi-supervised networks. The results on the Sunnybrook dataset show that the segmentation results of this model are better, with a higher dice coefficient, accuracy, and recall rate.

Document code: A **Article ID:** 1673-1905(2022)09-0547-6
DOI https://doi.org/10.1007/s11801-022-2010-0

In the future, a large part of intelligent consultation and Internet medical treatment will assist doctors in imaging analysis, and image segmentation will be the basis and key to all such image analysis, understanding, and identification problems. It is of great medical value to achieve precise segmentation of specific tissues in the human body. However, due to the movement between the textures, it may cause a certain degree of noise to the image. The flow of blood may also cause artifacts, affecting the grayscale distribution of the image, making segmentation difficult. Contrastive learning was proposed by VHADSELL et al[1] in 2006. In this method, images similar to the target were distinguished by calculating the Euclidean distance, pulling similar samples close and pushing different samples away, to achieve the purpose of classification. In 2020, the momentum comparison algorithm for unsupervised visual representation knowledge[2] effectively improved the usability of contrastive learning. In 2020, CHEN et al[3] proposed to use a neural network projection head to encode the feature map, calculate the contrastive loss function, and improve the classification results. Medical images are often accompanied by many labeled images, which require a lot of human resources and material resources. Semi-supervised learning can achieve great segmentation results of few annotations while ensuring certain accuracy. In the semi-supervised learning method proposed by LEE et al[4], self-trained pseudo-labels were used to supervise the network to achieve image classification. In 2020, KHOSLA et al[5] extended the self-supervised batch comparison method to full supervision, allowing for efficient use of label information. Points belonging to the same class were clustered together in the embedding space, while clusters of samples from different classes were separated. In the same year, CHAITANYA et al[6] provided a strategy for global and local contrast learning for image segmentation, pre-training neural networks with unlabeled data, and then fine-tuning downstream tasks with limited annotations. In 2021, ZHENG et al[7] built an uncertainty-aware self-enhancement model and performed supervised learning on the student model. For unlabeled data, the segmentation map was predicted by the teacher model as the learning target of the student model. At the same time, the uncertainty of the learning target was evaluated, and the consistency loss function was used to improve the performance of the student model.

In this paper, a semi-supervised medical image segmentation network based on contrastive learning is constructed and the left ventricle of cardiac magnetic resonance imaging (MRI) images is segmented. The algorithm flow chart is shown in Fig.1.

Since the training is performed on a small amount of data, and the U-Net network has a better segmentation effect on a small amount of data, the U-Net network is used as the backbone.

* This work has been supported by the Natural Science Foundation of Jiangsu (No.BK20171443).
** E-mail: huxf@njupt.edu.cn

The semi-supervised medical image segmentation network is shown in Fig.2. The model consists of a U-Net backbone, a contrastive learning module, and a pseudo-label post-processing module. The image and its positive and negative samples pass through the U-Net's contraction path to extract features. The distance between the negative feature map and the original image feature map is calculated by the imbalanced contrastive loss function (ICLF). The segmented image is restored through the expansion path, and its binary crossover loss function is calculated. The initial model parameters are obtained after the iteration round [8]. The pseudo-label of the unlabeled image is obtained by the initial segmentation model parameters, and then the pseudo-label with higher accuracy is provided by the pseudo-label post-processing algorithm. Then the network is trained with pseudo-labeled images and the labeled images to get the final network model parameters.

The essence of contrastive learning is to shorten the distance between the original samples and their positive samples in the feature space while pushing away the original samples from the negative samples[9].

Unbalance means that the number of pixels between the negative images of the background region and the positive samples of the part to be segmented varies greatly. In an image, the gap between the positive and negative samples is obtained when extracting features and mapping them to a feature space. A part of the area to be segmented occupies a small area of the entire image. On the whole, it is an unbalanced distribution of positive and negative samples. Accordingly, an ICLF L_{ICLF} is proposed as follows

$$L_{\mathrm{ICLF}}=\frac{1}{2(D_{\mathrm{w}})^2},\tag{1}$$

where

$$D_{\mathrm{w}}(\boldsymbol{X}_1,\boldsymbol{X}_2)=\left\|\boldsymbol{G}_{\mathrm{w}}(\boldsymbol{X}_1),\boldsymbol{G}_{\mathrm{w}}(\boldsymbol{X}_2)\right\|_2,\tag{2}$$

where $\boldsymbol{X}_1$ represents the negative sample matrix, $\boldsymbol{G}_{\mathrm{w}}(\boldsymbol{X}_1)$ represents the feature matrix of the negative sample after feature extraction, $\boldsymbol{X}_2$ represents the original image matrix, $\boldsymbol{G}_{\mathrm{w}}(\boldsymbol{X}_2)$ represents the original image after feature extraction, and $D_{\mathrm{w}}(\boldsymbol{X}_1, \boldsymbol{X}_2)$ defines the distance between the negative sample features and the original image features.

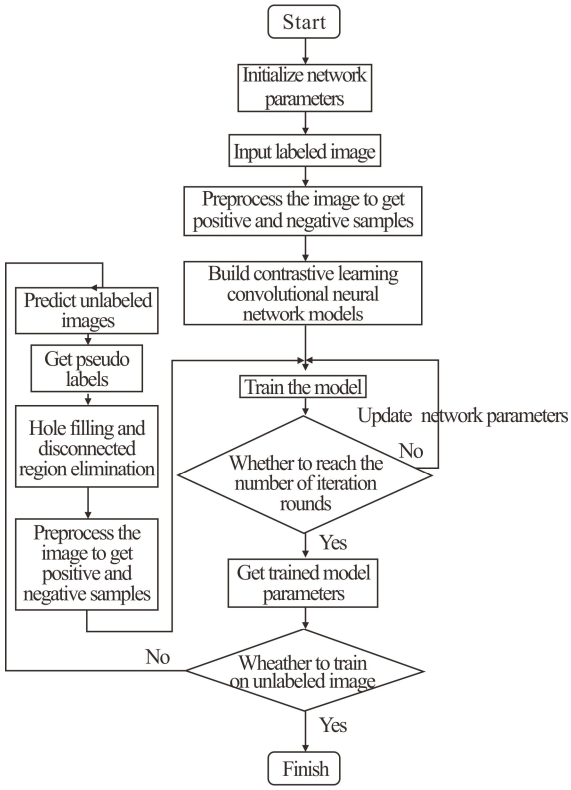

Fig.1 Flow chart of contrastive learning semi-supervised training model

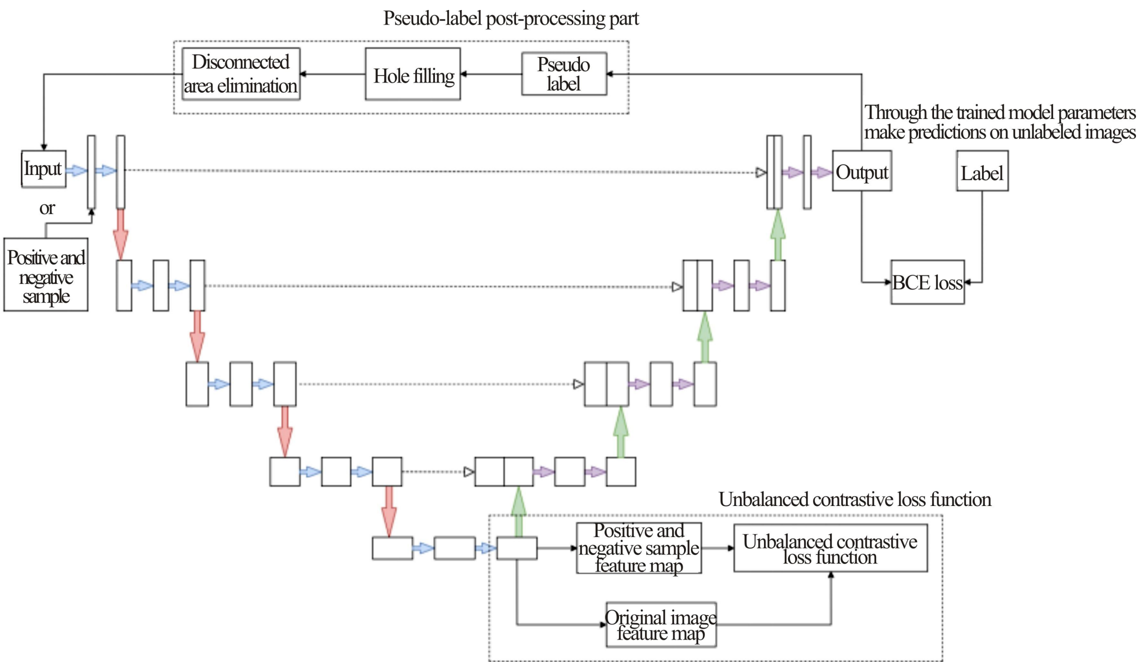

Fig.2 Network model of contrastive learning semi-supervised segmentation

If the calculated distance is small, the features between the region to be segmented and the negative samples are similar. If the distance is large, it means that the difference between the features of the image to be segmented and the features of the negative samples is large. It shows that the difference between the area to be divided and the background can be distinguished. As shown in Eq.(1), if the distance is larger, the ICLF value will become smaller. If the distance is smaller, the loss will become larger, which will form a penalty term and guide the overall loss function to decrease.

The overall loss L_{total} is composed of L_{ICLF} and the binary crossover loss function L_{BCE}, shown as follows

$$L_{total}=L_{BCE}+\lambda L_{ICLF}, \tag{3}$$

$$L_{BCE}=-[y_i \log x_i+(1-y_i)\log(1-x_i)], \tag{4}$$

where x_i represents the probability of the predicted data, y_i is the label of the data, and λ represents the weight parameter, which is 0.2.

Pseudo-labels are obtained by predicting unlabeled images with a pre-trained contrastive learning model. By contrastive learning the segmentation network model, the segmentation accuracy and the certainty of the obtained pseudo-labels can be improved. At the same time, pseudo-labels with higher confidence can be obtained by the pseudo-label post-processing algorithm.

The pseudo-label post-processing algorithm flow is shown in Fig.3, which consists of two parts, disconnected regions elimination and hole filling.

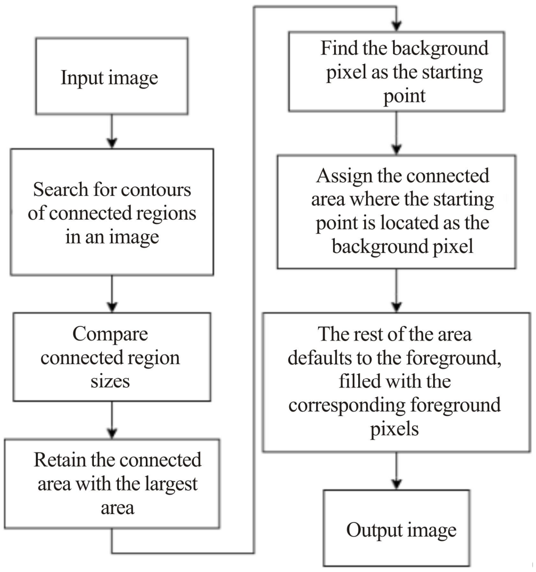

Fig.3 Flow chart of pseudo-label post-processing algorithm

When observing the grayscale distribution of cardiac MRI image, it is found that the gray values of the left ventricle area (foreground area) have a similar distribution interval, while the gray scale distribution outside the left ventricle (background area) is heterogeneous. As shown in Fig.4, where the blue line represents the grayscale distribution of the negative samples (background area), the red line represents the grayscale distribution of the positive samples (the foreground area), and the distribution of the two in the low gray areas has more overlaps. Disconnected regions elimination is to retain the largest connected area in the image.

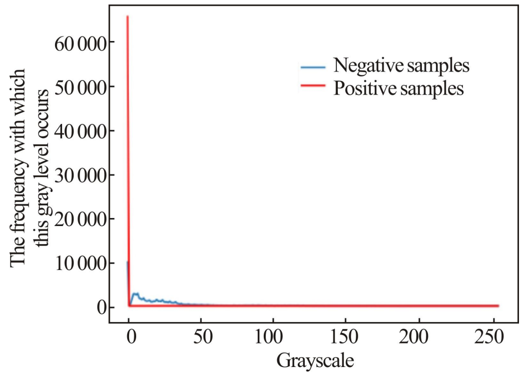

Fig.4 Comparison of grayscale histograms for positive and negative samples

The method of grayscale symbiosis matrix feature quantity is used to feature the positive and negative images in the labeled training set. The lines plotted in Fig.5 and Fig.6 show the distribution of contrast, dissimilarity, homogeneity, correlation and ASM of 161 negative and 161 positive samples. For negative samples, the contrast values are large, and the textures are complex. In contrast, the positive samples are generally smaller, indicating that the negative samples will bring certain interference to the segmentation. For both negative and positive samples, the homogeneity values are relatively high and the distribution is relatively uniform, which will bring about misidentification. For negative samples, the correlation distribution is relatively smooth, the correlation value between samples is large and the internal correlation of a single image is large, but for positive samples, the correlation value between images is small but the internal correlation of most images is large. For negative samples, ASM fluctuations are large, and the texture thickness between images varies greatly, but for positive images, it is generally at a higher level, which may cause to appear small holes when segmenting. Therefore, it is necessary to process these images through post-processing methods to prevent the appearance of small holes in the pseudo-label or the appearance of false target areas due to similar grayscale and texture distribution.

U-Net semi-supervised network and contrastive learning semi-supervised network model are built on Google Cloud Drive, PyTorch framework, written in Python3.8.

The experimental training set, validation set, and test set data are all from the Sunnybrook dataset. The dataset consists of MRI images of cardiac slices from 45 patients. The size of the images is 256×256 pixels. There are 805 cardiac MRI images annotated by experts. The purpose of the experiment is to segment the left ventricle in the MRI image.

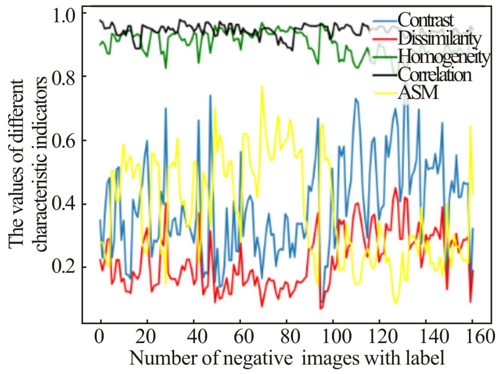

Fig.5 Texture feature distribution plot of negative samples

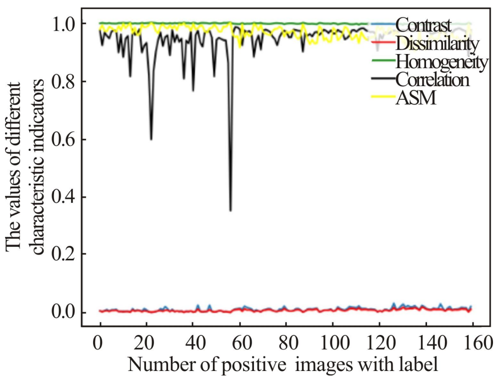

Fig.6 Texture feature distribution plot of positive samples

Fig.7 and Fig.8 illustrate the post-processing effect of disconnected region removal and hole filling. The unlabeled images are predicted by the pre-training model, and then the pseudo-labels are obtained by the pseudo-label post-processing algorithm. It can be seen from the results of Fig.7(d) and Fig.8(d) that adding the pseudo-label post-processing algorithm can effectively improve the credibility and accuracy of the pseudo-label, which has a significant impact on the subsequent network training and segmentation.

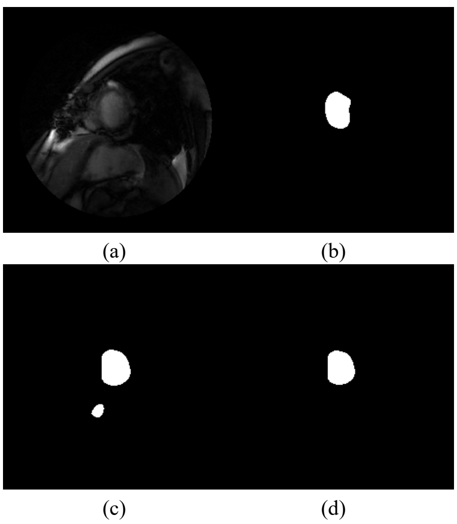

Fig.7 Post-processing effect of disconnected region removal: (a) Original image; (b) Label; (c) Segmentation result of contrastive learning model; (d) Post-processing result

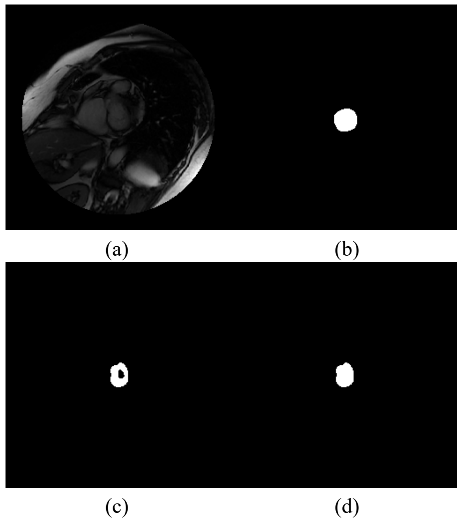

Fig.8 Post-processing effect of hole filling: (a) Original image; (b) Label; (c) Segmentation result of contrastive learning model; (d) Post-processing result

Compared with the U-Net network, supervised U-Net network based on ICFL has a better performance on image segmentation, and can more accurately capture the area to be segmented. Tab.1 shows the dice coefficient, precision, recall, and the number of images whose dice coefficient is lower than 85% with a sample size of 483 training images and 161 testing images. It can be seen that the contrastive learning for the segmentation of the cardiac compared with U-Net has a 4.28% improvement in the dice coefficient, a 3.44% improvement in the precision, and a 2.12% improvement in the recall rate. The effectiveness of contrastive learning in image segmentation is verified.

Tab.1 Comparison of experimental results of U-Net fully supervised network whether using ICLF

	Dice	Precision	Recall	Number of dice coefficients below 85%
U-Net	86.57%	85.83%	93.15%	17
Contrastive learning	90.85%	89.27%	95.27%	13

Furthermore, the data set is divided into 483 training images and 161 testing images, of which the training set is further divided into 161 labeled data and 322 unlabeled data.

Both U-Net semi-supervised learning and contrastive learning semi-supervised learning use Kaiming initialization[10]. The batch training size is set to 8, and 1 000 iterations are performed. The Adam gradient descent method is used, the learning rate is set to 0.000 1, and the weight decay is 10^{-8}.

Tab.2 compares the results of U-Net semi-supervised backbone network, comparison method[10], contrastive learning + U-Net semi-supervised network, and post-processing + contrastive learning + U-Net semi-supervised network in Fig.2 for image segmentation.

Compared with the semi-supervised model, the contrastive learning semi-supervised model can improve the dice coefficient by 3.42%, the precision by 3.81%, and the recall rate by 1.3%. In addition, compared with Ref.[11], it increases by 0.65% on the dice coefficient and 1.83% on the recall rate.

After adding the pseudo-label post-processing algorithm in the experiment, the contrastive learning semi-supervised model can increase the dice coefficient by 1.45%, the precision by 2.21%, and the recall rate by 0.16%. After the addition of the post-processing algorithm, compared with Ref.[11], the dice coefficient is increased by 2.1% and the recall rate by 1.99%.

Tab.2 Experimental results of semi-supervised networks with and without contrastive learning and pseudo-label post-processing algorithm

	Dice	Precision	Recall	Number of dice coefficients below 85%
U-Netsemi-supervised	83.63%	81.54%	89.73%	43
Contrastive method[11]	86.40%	88.70%	89.20%	26
Our method (without post-processing)	87.05%	85.35%	91.03%	28
Our method (with post-processing)	88.50%	87.56%	91.19%	25

U-Net semi-supervised segmentation model, contrastive learning semi-supervised model, contrastive learning semi-supervised model with post-processing segmentation results, and corresponding objective indicators are shown in Fig.9 and Fig.10.

It is not difficult to find from Fig.9 and Fig.10 that the subjective quality and objective indicators of the segmented images are improved after the contrastive learning and pseudo-label post-processing algorithm, and they are closer to the labels.

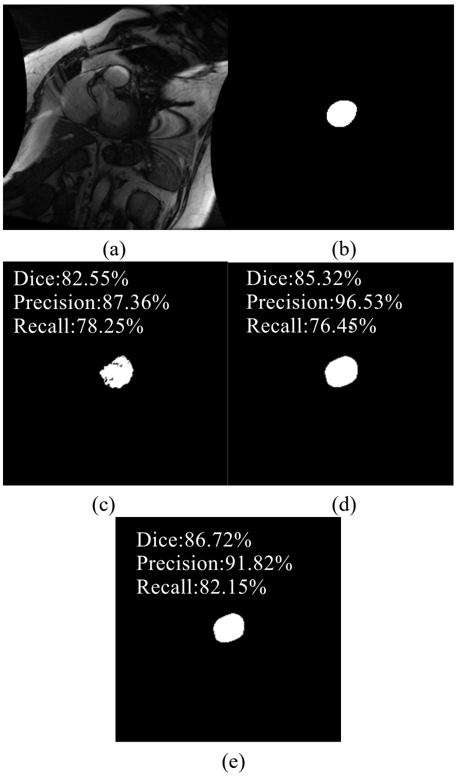

Fig.9 Group 1: (a) Original image to be segmented; (b) Label; (c) U-Net semi-supervised segmentation result; (d) Contrastive learning semi-supervised segmentation result; (e) Contrastive learning semi-supervised segmentation result with post-processing

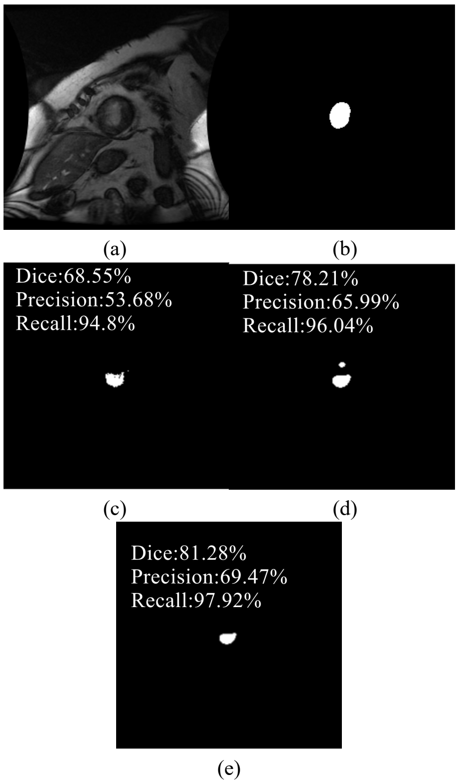

Fig.10 Group 2: (a) Original image to be segmented; (b) Label; (c) U-Net semi-supervised segmentation result; (d) Contrastive learning semi-supervised segmentation result; (e) Contrastive learning semi-supervised segmentation result with post-processing

In this paper, we propose a semi-supervised medical image segmentation network based on contrastive learning. To achieve a better semi-supervised effect, we optimize the segmentation results of the U-Net network by ICLF, use the contrastive learning network to generate pseudo-labels, and present a hole filling and disconnected regions elimination processing algorithm to form supervision for unlabeled images which can achieve the purpose of improving the confidence level of pseudo-labels.

Contrastive learning achieves a better prediction effect by pushing away the distance between the image features containing the region to be segmented and the image features of negative samples. Combined with post-processing methods of hole filling and disconnected region elimination, pseudo-labels are improved, and better segmentation results are achieved.

From the perspective of future development, we can start from the direction of image and image correlation. For MRI medical images, they have a continuous-time character of each volume. If the adjacent frames information of the MRI images is utilized, the location of the left ventricle of the heart of the current frame can be roughly predicted, which is equivalent to adding a strong positional attention mechanism. It is more helpful for accurate segmentation.

Statements and Declarations

The authors declare that there are no conflicts of interest related to this article.

References

[1] VHADSELL R, CHOPRA S, LECUN Y. Dimensionality reduction by learning an invariant mapping[C]//2006 IEEE Computer Society Conference on Computer Vision and Pattern Recognition (CVPR'06), June 19-21, 2006, New York, NY, USA. New York：IEEE, 2006, 2：1735-1742.

[2] HE K, FAN H, WU Y, et al. Momentum contrast for unsupervised visual representation learning[C]//Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, June 13-19, 2020, Seattle, WA, USA. New York：IEEE, 2020：9729-9738.

[3] CHEN T, KORNBLITH S, NOROUZI M, et al. A simple framework for contrastive learning of visual representations[C]//International Conference on Machine Learning, July 12-18, 2020, Vienna, Austria. San Diego：ICML, 2020：1597-1607.

[4] LEE D H. Pseudo-label：the simple and efficient semi-supervised learning method for deep neural networks[C]//Workshop on Challenges in Representation Learning, June 16-21, 2013, Atlanta, GA. San Diego：ICML, 2013：896.

[5] KHOSLA P, TETERWAK P, WANG C, et al. Supervised contrastive learning[J]. Advances in neural information processing systems, 2020, 33：18661-18673.

[6] CHAITANYA K, ERDIL E, KARANI N, et al. Contrastive learning of global and local features for medical image segmentation with limited annotations[J]. Advances in neural information processing systems, 2020, 33：12546-12558.

[7] ZHENG X, FU C, XIE H, et al. Uncertainty-aware deep co-training for semi-supervised medical image segmentation[EB/OL]. (2021-11-23) [2022-04-26]. https：//arxiv.org/abs/2111.11629v1.

[8] CHAKRABORTY S, GOSTHIPATY A R, PAUL S. G-SimCLR：self-supervised contrastive learning with guided projection via pseudo labelling[C]//2020 International Conference on Data Mining Workshops (ICDMW), November 17-20, 2020, Sorrento, Italy. New York：IEEE, 2020：912-916.

[9] DIPPEL J, VOGLER S, HÖHNE J. Towards fine-grained visual representations by combining contrastive learning with image reconstruction and attention-weighted pooling[EB/OL]. (2021-04-09) [2022-04-26]. https：//arxiv.org/abs/2104.04323.

[10] HE K, ZHANG X, REN S, et al. Delving deep into rectifiers：surpassing human-level performance on Imagenet classification[C]//Proceedings of the IEEE International Conference on Computer Vision, June 7-12, 2015, Boston, MA, USA. New York：IEEE, 2015：1026-1034.

[11] ZHAO X, FANG C, FAN D J, et al. Cross-level contrastive learning and consistency constraint for semi-supervised medical image segmentation[EB/OL]. (2021-02-13) [2022-04-26]. https：//arxiv.org/abs/2202.04074.