TSU-net: Two-stage multi-scale cascade and multi-field fusion U-net for right ventricular segmentation

Xiuquan Du[a], Xiaofei Xu[b], Heng Liu[c] and Shuo Li[d]

[a]Key Laboratory of Intelligent Computing and Signal Processing, Ministry of Education, Anhui University, Hefei, Anhui, China
[b]School of Computer Science and Technology, Anhui University, Hefei, Anhui, China
[c]Department of Gastroenterology, The First Affiliated Hospital of Anhui Medical University, Hefei, Anhui, China
[d]Department of Medical Imaging, Western University, London, ON, Canada

ARTICLE INFO

Keywords:
Cardiac MRI
Right Ventricle Segmentation
Automated Segmentation Method
Right Ventricle Segmentation Challenge

ABSTRACT

Accurate segmentation of the right ventricle from cardiac magnetic resonance images (MRI) is a critical step in cardiac function analysis and disease diagnosis. It is still an open problem due to some difficulties, such as a large variety of object sizes and ill-defined borders. In this paper, we present a TSU-net network that grips deeper features and captures targets of different sizes with multi-scale cascade and multi-field fusion in the right ventricle. TSU-net mainly contains two major components: Dilated-Convolution Block (DB) and Multi-Layer-Pool Block (MB). DB extracts and aggregates multi-scale features for the right ventricle. MB mainly relies on multiple effective field-of-views to detect objects at different sizes and fill boundary features. Different from previous networks, we used DB and MB to replace the convolution layer in the encoding layer, thus, we can gather multi-scale information of right ventricle, detect different size targets and fill boundary information in each encoding layer. In addition, in the decoding layer, we used DB to replace the convolution layer, so that we can aggregate the multi-scale features of the right ventricle in each decoding layer. Furthermore, the two-stage U-net structure is used to further improve the utilization of DB and MB through a two-layer encoding/decoding layer. Our method is validated on the RVSC, a public right ventricular data set. The results demonstrated that TSU-net achieved an average Dice coefficient of 0.86 on endocardium and 0.90 on the epicardium, thereby outperforming other models. It effectively assists doctors to diagnose the disease and promotes the development of medical images. In addition, we also provide an intuitive explanation of our network, which fully explain MB and TSU-net's ability to detect targets of different sizes and fill in boundary features.

1. Introduction

The World Health Organization reports that cardiovascular diseases (CVDs) with a high incidence have been the leading cause of death worldwide [1]. The evaluation technique for cardiac ventricle function plays an essential role in the diagnosis and treatment of CVDs. As is known to all, cardiac magnetic resonance images (MRI) are considered the most accurate method to estimate clinical indicators, such as ventricular volume, ejection fraction, myocardial mass, etc. [2, 3]. As a prerequisite to obtaining clinical cardiac indicators, cardiac left/right ventricle cavity segmentation is necessary. We give a few examples that show the contours of the left/right ventricles from Fig. 1. Compared with the right ventricular contour, the left ventricular contour target is large and regular, which is relatively easy and well-studied in [4-6]. Furthermore, the right ventricle is more challenging due to its irregular gravity, blurry boundaries, crescent shape, and larger object size [7]. Besides, the right ventricle is manually segmented by clinical experts that are time-consuming, tedious, and very sensitive to intra-expert and inter-expert variability [8, 9]. Therefore, an accurate and automatic right ventricle segmentation algorithm is urgently needed in the clinic compared with left ventricle segmentation.

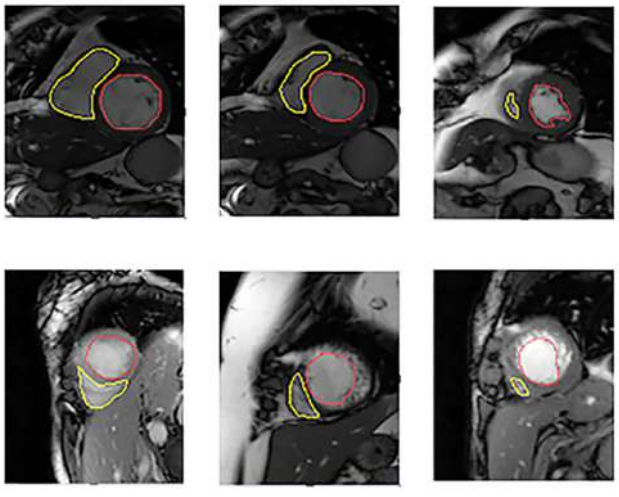

Figure 1: Contours of the left ventricle (red line) and right ventricle (yellow line).

To promote the development of automated right ventricular segmentation technology, Medical Image Computing and Computer-Aided Intervention (MICCAI) launched the Right Ventricular Segmentation Challenge (RVSC) in 2012 [7]. There were many traditional medical image segmen-

*This work was supported in part by the Provincial Natural Science Research Program of Higher Education Institutions of Anhui province under Grant KJ2020A0035.
ORCID(s):

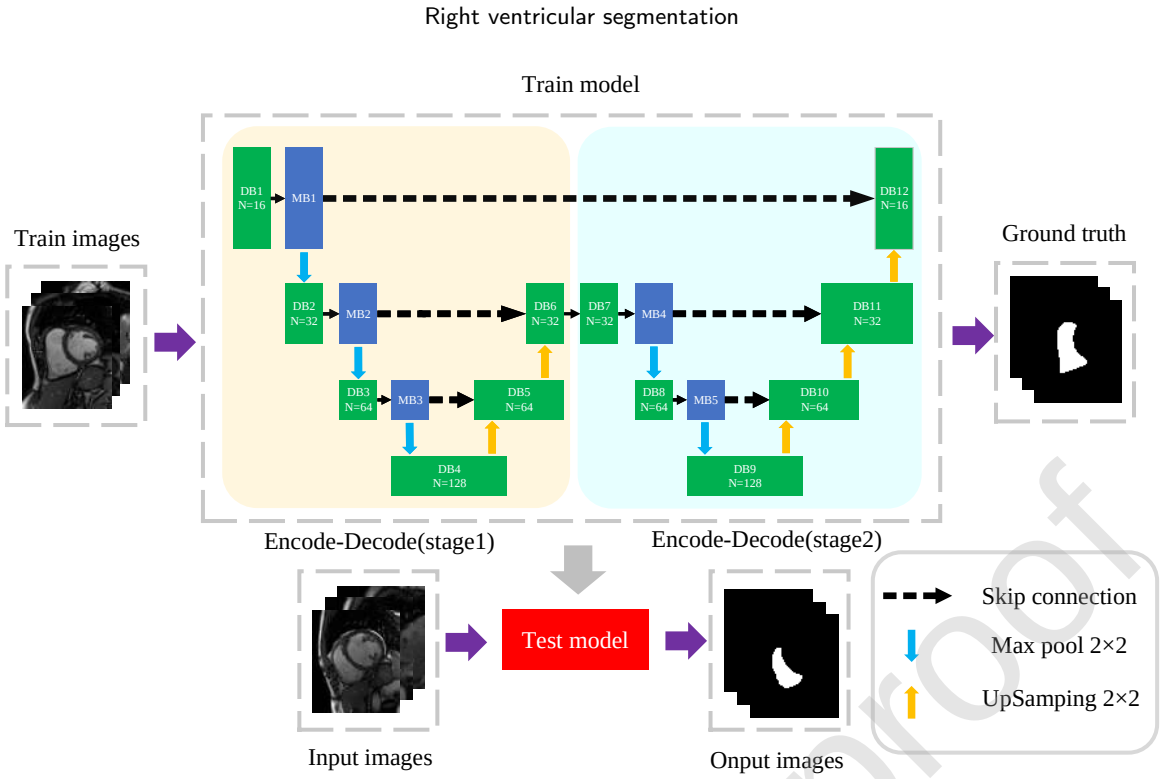

Figure 2: Contours of the left ventricle (red line) and right ventricle (yellow line).

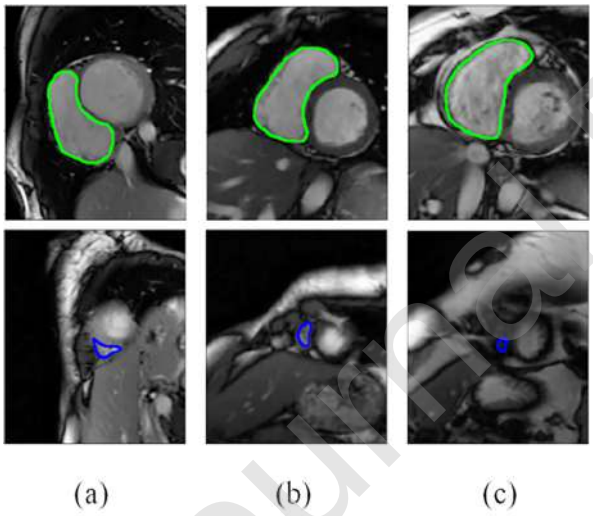

Figure 3: end-diastolic (green) and end-systolic (blue) of Right ventricular.

tation methods in the previous studies, such as the atlas-based, the prior model-based, the image-based, and the pixel classification [10-17]. The atlas-based method CMIC [12] won the first place in RVSC, and another atlas-based method SBIA won the third place [13]. However, the limitation of the atlas-based process depends on the quality of registration and huge computing cost. In order to utilized the prior knowledge of right ventricular structure, a prior based model was proposed to improve the performance of right ventricular segmentation [14]. Nevertheless, it requires strong expertise, so its robustness and accuracy were very low. Relatively speaking, the image-based model had the advantages of high time efficiency and high segmentation accuracy. The image-based method won the second place in RVSC [15]. Unlike the above, neural network and pixel classification methods such as random forest were based on a training database [16, 17]. Nevertheless, the underlying data set needs to be large enough. Despite the good meaning of traditional methods, they were still far from the expert level [18]. This is because the right ventricle boundary is fuzzy and the size of the target varies greatly, while the traditional methods are shallow structures and do not have the ability to extract complex features. Therefore, we urgently need a novel method that can extract complex features and detect different size targets simultaneously.

In recent years, convolutional neural networks (CNNs) in deep learning architecture, have been achieved state of the art performance in a breadth of visual recognition tasks. Compared with the traditional methods, the deep learning method automatically learns features and overcomes manual features' limitations. In other words, the deep learning model has a deeper structure, which can grasp more complex features and obtain better performance. In this respect, the most significant deep learning network was based on the end-to-end convolutional neural network (FCN) [19] and U-net [20]. Their medical image segmentation performance, especially heart segmentation, has been greatly improved [19, 21]. In addition, many researchers have done a lot of research on the ventricular data set. In the research of antagonistic learning, C.J. Yu et al. [22] found that the joint

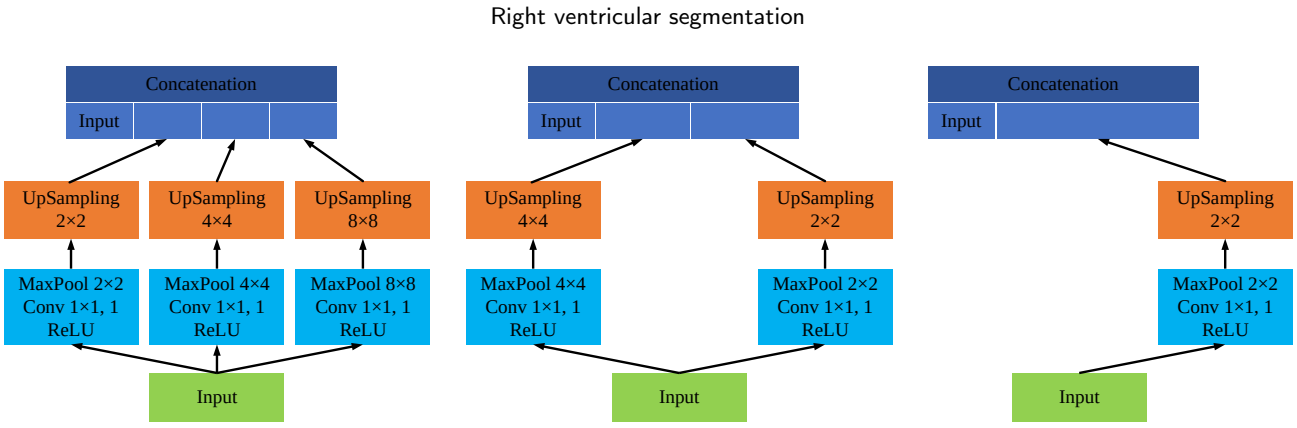

Figure 4: Multi-Layer-Pool Block consists of 1-3 different pool layers, 1-3 different UpSmaping layers, and one concatenation.

distribution of reverse mapping network and multi task network is matching, and further proved that reverse mapping network can effectively restrict the learning of multi-task network. Therefore, their multi-task learning method based on reverse mapping has achieved good results in MRI and CT multi-type cardiac index estimation. R.J.G et al. [23] applied the research of heterogeneous task learning to the segmentation task of pixel level classification of left ventricle and the direct quantization task of image level regression, introduced segmentation information through interaction, jointly improved the spatial attention map, guided the quantization to focus on the relevant regions of left ventricle, transferred quantitative feedback, and made global constraints on segmentation. Based on this idea, the proposed K-Net has achieved good results in the segmentation and quantification of left ventricle. H.K et al. [24] studied the scale space theory to improve the left ventricular invariance by considering the diffusion process of geometry and appearance. Based on this study, the proposed scheme of scale space combined with data enhancement induced by thermal equation has high accuracy in left ventricle. T.Liu et al. [25] studied the influence of spatial information on heart data sets. Their proposed network learning can learn spatial information well and combine with 3D information context. Therefore, they have achieved good results on right ventricular data sets. C.Xu et al. [26] committed to solve the influence of contrast agent injection on cardiac segmentation. They are also the first to propose a segmentation method without contrast agent. These studies on cardiac have achieved good results in their respective fields, but there is still a research gap, and the small target problem is not considered.

Compared with other cardiac data sets, the method based on RVSC right ventricular data set still has some defects, which need to be solved and improved. For example, C. Szegedy et al. [27] proposed an inception network, which used a 22-layer network to deepen the network's depth and increase the level of segmentation of the network. However, it will dramatically increase the number of parameters and computational costs, leading to an over-fitting problem on a small dataset like the benchmark RVSC database. In view of such difficulties, F. Yu et al. [28] proposed a Dilated convolution, which obtained large-scale features without increasing too many parameters and computing costs. In practice, features in different scales can be obtained by dilated convolution in different dilation rates. This method makes good use of multi-scale information, but ignores the problem that the size of right ventricular target changes greatly, which affects the efficiency of network segmentation. J.Li et al. [29] proposed a Dilated-Inception Net (DIN), which combined the advantages of the Dilated (atrous) structure [28] and Inception structure [27], each layer of convolution in the U-net network was replaced by an embedded Dilated-Inception Block (DIB). DIN automatically aggregated the multi-scale features of the right ventricle in all layers of the network. However, for RVSC, a small RVSC database, reducing the use of convolution may affect deep feature extraction. Moreover, this method is slightly insufficient for small target acquisition. For obtaining deeper right ventricular characteristics. For example, Z. Liu et al. [30] proposed a Residual U-net, which was added four layers of the residual block in each layer of encoding layer to capture complex features of the right ventricle [31]. In addition, J. Bullock proposed an XNet [32], which used the two-stage U-net structure to deepen the network architecture without changing U-net's internal structure and has made great breakthroughs in small data sets. To sum up, the latest deep learning method on the RVSC data set has a strong ability to extract complex features. Unfortunately, they missed a crucial point: large differences in the size of right ventricular targets, which may affect the performance of newwork. The lack of small target capture ability of the network will affect the overall segmentation performance because of the poor segmentation performance of individual small targets. Moreover, compared with the large target segmentation, small target segmentation is more challenging and severe.

Therefore, in this paper, we aim to improve the ability of the network to detect targets of different sizes and to fill the boundaries of targets. Motivated by DIN [29] and XNet [32], we propose a TSU-net network to improve the right ventricular segmentation by extracting and aggregating multi-scale features from Dilated-Convolution Block (DB). Meanwhile, to compensate for the ability of the network to detect targets of different sizes, we propose a Multi-Layer Pool

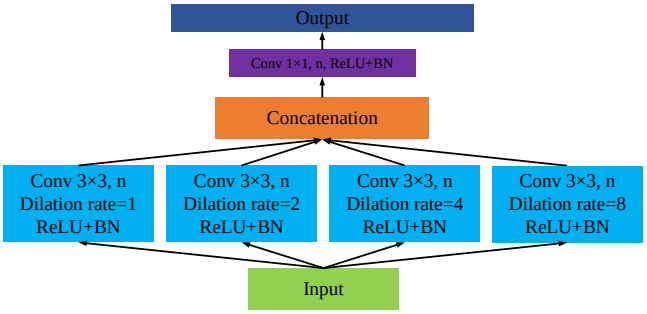

Figure 5: Dilated-Convolution Block (DB) consists of four dilated Convolutional layers, one concatenation layer, and one 1×1 Convolutional layer.

Block (MB), which mainly relies on multiple effective field of views to detect objects at different sizes and fill boundary features. Moreover, we embed MB into the encoding layer of TSU-net, which detects different size targets and fill boundary information in each encoding layer. Compared with the previous work, our TSU-net considers the problem of large changes in the size of the right ventricular target and uses MB to detect different small targets and supplement the boundary information of the target to achieve more accurate segmentation. Thus, through this operation, we can not only detect the size of the target but also extract the target features with different sizes. In addition, TSU-net uses a two-stage U-net network structure. It adds the number of MB and DB,which not only enhances the ability of the network to detect targets and extract features but also enhances the ability of the network to dilute shallow features.

The main contributions of this work are summarized as follows:

(1) Our proposed MB can detect targets of different sizes and fill boundary features.

(2) We integrate DB block and MB block and replace the convolutional network of encoding layer with DB and MB, and the convolutional network of decoding layer with DB. Therefore, the proposed model can automatically aggregate the multi-scale features of the right ventricle in all layers of the network and detect different size targets in the encoding layer.

(3) Our proposed TSU-net's performance on the RVSC data set is superior to that of the latest deep learning network.

2. Method

The proposed TSU-net consists of two encoder-decoder frameworks in Fig. 2. Different from other network models in the past, we used DB and MB to replace the convolution layer in the encoding layer, thus, we can gather multi-scale information of right ventricle, detect different size targets and fill boundary information in each layer of the encoding layer. Accordingly, in the decoding layer, we used DB to replace the convolution layer, so that we can aggregate the multi-scale features of the right ventricle in each layer of the decoding layer. In addition, we used the idea of XNet [32] two-stage U-net network, which uses a two-stage encode-decode structure (stage1 and stage2) to deepen the network's ability to capture target features and reduce the forgetting ability of the model to previous features.

2.1. Multi-Layer-Pool Block

One of the difficulties in medical image segmentation is the large change in the image's target size. As shown in Fig. 3, the images of three different patients ((a), (b) and (c) in Fig. 3) end-diastolic and end-systolic are extracted. The target size of the right ventricular end-diastolic (green) and end-systolic (blue) of the same patient changes greatly,which will cause the model to lose some target features and affect the extraction of target features. Therefore,we propose MB relies on multiple effective receptive fields to detect objects of different sizes and supplement the missing feature information. The size of the receptive field roughly determines how much context information we can use. The general max-pooling operation just employs a single pooling kernel, such as 2 x 2, it will weaken the ability of different size target detection. Therefore, as illustrated in Fig. 4,the proposed MB1 encodes global context information with three different-size receptive fields: 2×2, 4×4, and 8×8. The three-level outputs contain the feature maps of various sizes.After that, the feature graphs of different sizes are connected with the input to supplement the feature information of different target sizes. In short, the function of MB is to make up the lost target feature information by multi receiving domain information fusion.

As we all know, the encoding layer needs to dilute the target features while retaining more target features, which coincides with our proposed MB function. Therefore, we embed MB into the encoding layer, and reduce the number of our receiving domains as the encoding layer undergoes 2 $\times$ 2 pooling. MB is like arranging a quality inspector in each coding layer to check and make up for each passing item. We show the steps of algorithm 1 in Table. 1, and show the steps of Multi-Layer-Pool Block in the form of pseudo code.

2.2. Dilated-Convolution Block

Dilated convolution can significantly extend the receiving domain without loss of resolution and coverage [28].Through the amplification and convolution of different expansion rates, different sizes of acceptance regions can be obtained to extract multi-scale features. Dilated-Convolution Block (DB) combines multiple dilated convolutional layers within different dilation rates that multi-scale features are extracted and aggregated.

As shown in Fig. 5, four convolution layers with 3×3 cores have different dilation rates (1, 2, 4, and 8) applied to features of different scales (3×3, 5×5, 9×9, and 17×17) in DB. Then four dilated convolution layers are output combined by cascade layers and aggregated through the standard 1×1 convolution layer. The typical 1×1 convolution layer is widely used to combine different characteristics

Table 1
Multi-Layer-Pool Block algorithm steps.

Algorithm1:Multi-Layer-PoolBlock
Input: *Three-dimensional feature maps* W, *kernel size* $K_i(i = 0, 1, 2, 4)$, *stride* S , *padding* P
Output: *Three-dimensional feature maps* Y
1. $W_1 = \frac{W-K_1}{S} + 1$
2. $W_1 = \frac{W_1-K_4+2P}{S} + 1$
3. $W_1 = K_1 * W_1$
4. $W_2 = \frac{W-K_2}{S} + 1$
5. $W_2 = \frac{W_2-K_4+2P}{S} + 1$
6. $W_2 = K_2 * W_2$
7. $W_3 = \frac{W-K_3}{S} + 1$
8. $W_3 = \frac{W_3-K_4+2P}{S} + 1$
9. $W_3 = K_3 * W_3$
10. $Y = convat[W, W_1, W_2, W_3]$
Return Y

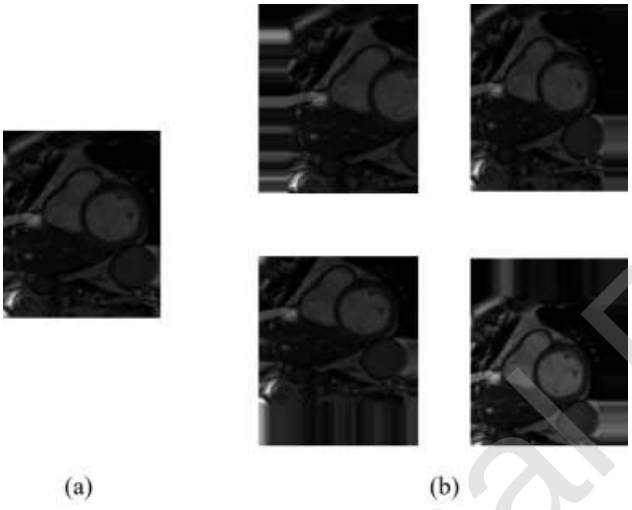

Figure 6: The process of data augmentation. Fig. 6 (a) is the original image, and the four images in Fig. 6 (b) are the results of data augmentation.

[27]. In Fig. 5, each convolution layer represents the number of convolution kernels. Each convolution layer's active unit is the unit of linear correction (ReLU) [35]. We show the steps of algorithm 2 in Table 2, and show the steps of Dilated-Convolution Block in the form of pseudo code.

2.3. Loss Function

Our framework is an end-to-end deep learning system. As shown in Fig. 2, we needed to train the proposed method to predict each pixel as foreground or background, which is a pixel-level classification problem. The input signal can be trained to predict the relationship between the results and the actual situation by minimizing the cross-entropy error. The loss function is defined as:

$$L = -\sum_{i\in\theta}\sum c y_{ci} log(p_{ci}) + (1 - y_{ci}) log(1 - p_{ci}) \quad (1)$$

Table 2
Dilated-Convolution Block algorithm steps.

Algorithm 2:Dilated-Convolution Block
Input: *Three-dimensional feature maps* W, *kernel size* $K_i(i = 1, 2)$, *dilation rate* $D_i(i = 1, 2, 3, 4)$ *stride* S , *padding* $P_i(i = 1, 2, 3, 4, 5)$
Output: *Three-dimensional feature maps* Y
1. $W_1 = \frac{W+2*P_1-D_1*(K_1)-1}{S} + 1$
2. $W_2 = \frac{W+2*P_2-D_2*(K_2)-1}{S} + 1$
3. $W_3 = \frac{W+2*P_3-D_3*(K_3)-1}{S} + 1$
4. $W_4 = \frac{W+2*P_4-D_4*(K_4)-1}{S} + 1$
5. $W_5 = convat[W_1, W_2, W_3, W_4]$
6. $Y = \frac{W_5-K_2+2*P_5}{S} + 1$
Return Y

Where p_{ci} denotes the predicted probability of c-th class for pixel$i$in the predicted result $p, y_{ci} \in \{0, 1\}$is the corresponding ground truth,i.e.,y_{ci}=1 if pixel i belongs to the c-th class,otherwise 0.c=1 the class of right ventricle cavity while c=2 denotes the background. θ denotes the space of the predicted result p and the ground truth y. The predicted result p is output of TSU-net. The ground truth y is the segmentation result generated from the segmentation contour provided in the RVSC training database [36]. By minimizing the loss on a training database, the parameters of TSU-net can be optimized. Then the trained TSU-net can be applied for automated right ventricle segmentation.

3. Experiment

In this part, we analyze the MB, TSU-net, and the conjecture of relieving BN and small bitch-size. The pretreatment process and the setting of training parameters are described in detail. In addition, TSU-net is compared with the most advanced methods.

3.1. Experimental configuration

In U-net, four max-pooling layers and four up-sampling layers are used to segment 512 × 512 microscopic images [21]. The cardiac MRI image size is 256 × 216; accordingly, we used three max-pooling layers and three up-sampling layers in the left and right U-net structures. And the Convolutional layers in the U-net were replaced with the proposed DBs and MBs. Therefore, TSU-net contains 12 DBs, 5 MBs, 5 pooling layers, 5 upper sampling layers, and 1 output layer. In Fig. 2, n represents the number of cores in each core. The pool window size of each maximum pooling layer is 2 × 2. The window size for each up-sampling is 2 × 2. In the concatenation layers, skip connections are applied to combine the output features of lower and upper layers in TSU-net for segmentation. According to previous studies [20, 34], jump connections helps speed up the training process and improve a deep segment network's performance. In the output layer of TSU-net, a 1 × 1 Convolutional layer with a softmax activation function is applied for predicting the segmentation results. Our network parameter is Adam Optimizer [40] with a

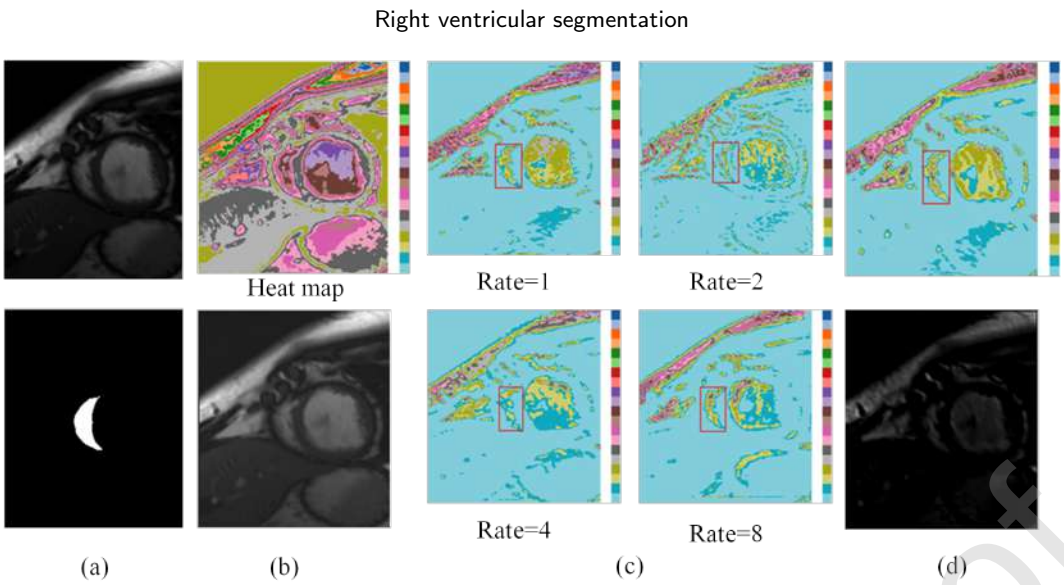

Figure 7: Comparison of feature maps transferred by the DB module. (a) Original image and ground truth, (b) the feature map before inserting DB modules,(c) Characteristic graphs obtained by dilation convolution with different dilation rates, and (d) the feature map before inserting DB modules.

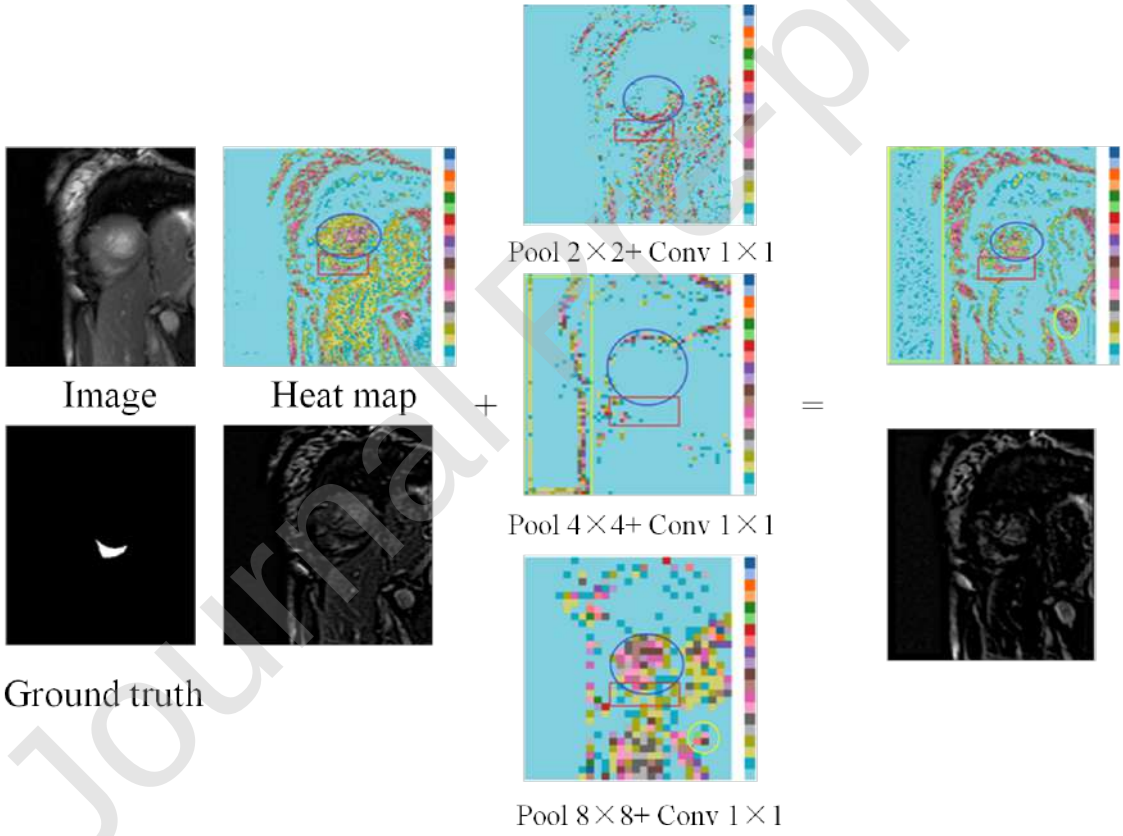

Figure 8: Comparison of feature maps transferred by the MB module. (a) Original image and ground truth, (b) the feature map before inserting MB modules, (c) Characteristic graphs obtained from different receiving domains, and (d) the feature map before inserting MB modules.

learning rate of 3e-4, a drop-out rate of 0.3, and a mini-batch size of 5 was applied.

In the experiments, the testing hardware and software conditions are listed as follows: Desktop, Intel i7 3.6 GHz CPU, 32G DDR4 RAM, Nvidia GeForce GTX 1080 Ti, Keras with TensorFlow backend.

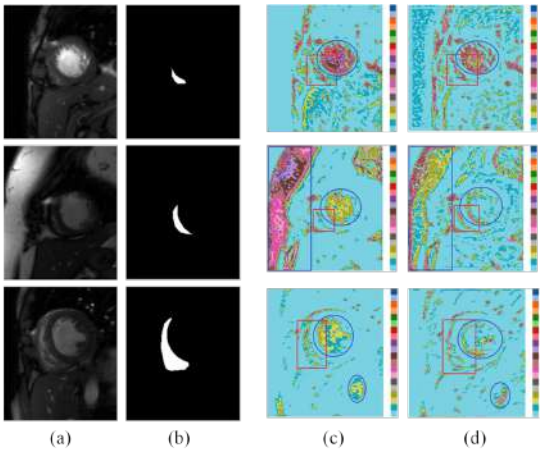

Figure 9: Comparison of feature maps transferred by the MB module. (a) Original image, (b) ground truth, (c) the feature map before inserting MB modules, and (d) the feature map after inserting MB modules.

3.2. Dataset

3.2.1. RVSC database

The RVSC database included 48 patients (36 males and 12 females) with an age of 52.1 ± 18.1 years. MR images of the right ventricle in about 16 subjects were manually segmented by clinical experts and then treated as ground truth. In the RVSC database, there is a training set containing MR images, and the contour of 16 patients is segmented. The images of 16 patients in the test set (test1 group) and the data of 16 patients in an additional test set (test group 2) are established. The public data set address is http://pagesperso.litislab.fr /cpetitjean /mr-images-and-contour-data/.

In the RVSC training database, 243 cardiac MR images were obtained from 16 patients. To normalize, each MRI image was subtracted from the mean value, divided by the standard deviation. Then the average value of each normalized image is zero unit difference,and the normalization formula is as follows:

$$X_n = \frac{X_y - X_{min}}{X_{max} - X_{min}} \tag{2}$$

Where X_y represents the original pixel value,X_{max} represents the maximum value of a pixel and X_{min} denotes the minimum value of a pixel.

During the training procedure, data augmentation techniques (rotation, vertical / horizontal flipping, height / width shift, zoom, etc) were applied to each MR image in the training set. The data augmentation techniques can help to mitigate over-fitting and achieve more accurate and robust segmentation performance of deep networks [19, 21]. As shown in Fig. 6, Fig. 6(a) is the original image, and the four images in Fig. 6(b) are the results of data augmentation technologies (rotation, vertical / horizontal flipping, height / width shift, zoom, etc.)

3.2.2. MMWHS database

MMWHS 2017 challenge data set contains 20 CT and 20 Mr labeled data for training and 40 unlabeled data for testing [43]. Get the slice in the axial view. The in-plane resolution is about 0.434×434mm2, average thickness 0.596mm. Each training data has seven cardiac substructures, including left ventricle (LV), right ventricle (RV), left atrium (LA), right atrium (RA), left ventricular myocardium (myo), ascending aorta (AA) and pulmonary artery (PA). Because the challenge has stopped, we selected 20 CT images with tags in the training set as our experimental data set. From these 20 CT training sets, the first 16 are selected as training sets, and the last four are selected as test sets.

3.3. Evaluation Metrics

In order to evaluate cardiac right segmentation performance, Hausdorff Distance (HD) and dice metric (DM) was applied. The DM is a metric of area overlap between the predicted segmentation result and ground truth, defined as:

$$DM(U,V) = 2\frac{U \cap V}{U \cup V} \tag{3}$$

U denotes the predicted segmentation result, V is ground truth, $U \cap V$ denotes the intersection of U and V. $U \cup V$ denotes the union of U and V. DM varies from 0 (total mismatch) to 1(perfect match). The HD measures the distance of segmentation prediction and where U denotes the predicted segmentation result, V denotes ground truth. The HD is defined as:

$$\begin{aligned} HD(U,V) = max(&max_{u\in U}(min_{v\in V}(u,v)),\\ &max_{v\in V}(min_{u\in U}d(u,v))) \end{aligned} \tag{4}$$

Where U denotes the predicted segmentation result, V denotes ground truth, $d(.,.)$ indicates Euclidean distance. A smaller value of HD denotes high proximity between the segmentation prediction and ground truth.

Accuracy(ACC) and Error rate are often used as an indicator of binary classification problemsThe formula is as follows:

$$ACC = \frac{TP+TN}{TP+FP+TN+FN} \tag{5}$$

$$ER = 1 - ACC \tag{6}$$

Among them, true positive (TP) represents that the label is a positive sample and predicts to be a positive sample, false positive (FP) represents that the label is a negative sample and predicts to be a positive sample, true negative (TN) represents that the label is a negative sample and predicts to be a negative sample, and false negative (FN) represents that the label is a positive sample and predicted to be a negative sample.

3.4. Quantitative Analysis

3.4.1. Dilated-Convolution Block (DB) experimental analysis

DB mainly utilizes multi-scale fusion to extract and aggregate the multi-scale characteristics of the right ventricle.In short, DB is a feature fusion extracted by dilation convolution with different expansion rates to obtain more accurate

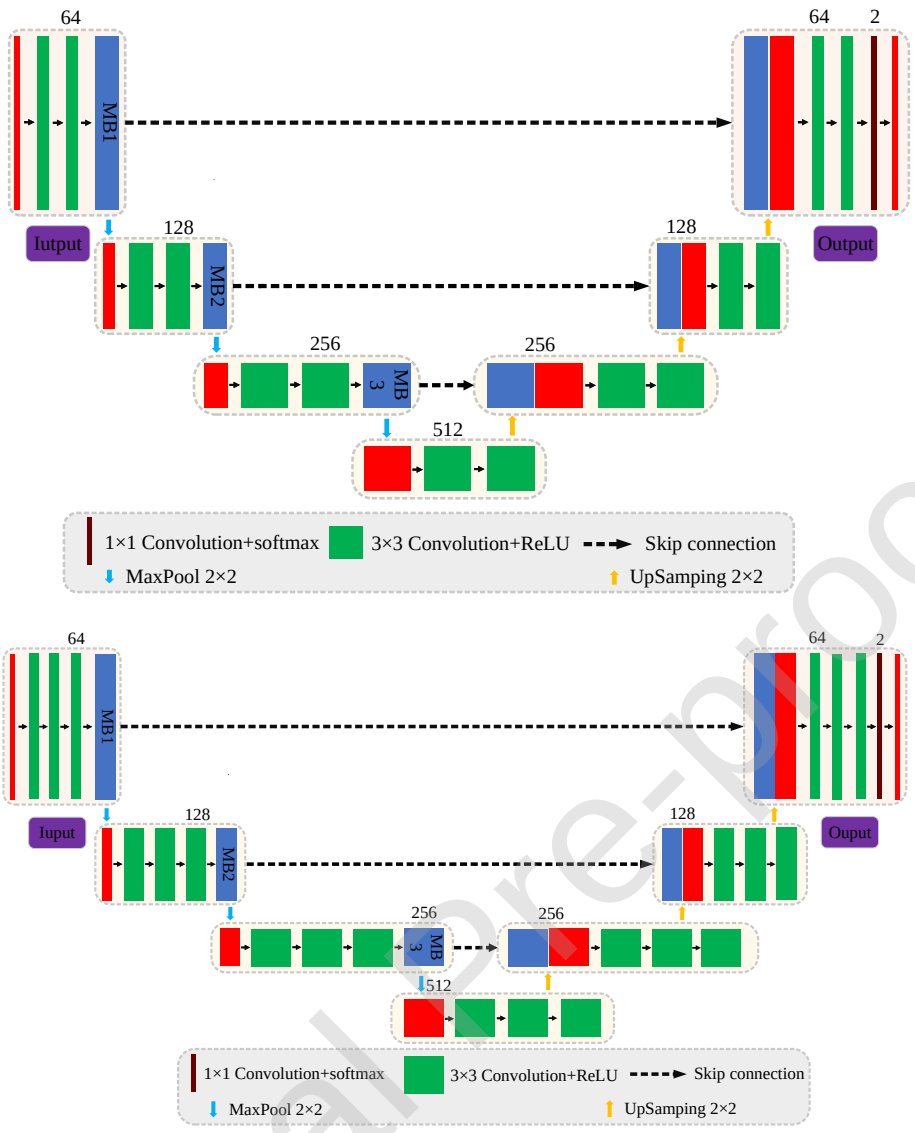

Figure 10: Application of MB in the basic framework. The top is MB / U-net, and the bottom is MB/R3U-net

target feature information. We show the process of extracting target features by DB from the Fig. 7. The expansion convolution with different expansion rates will dilute the information of irrelevant features and retain the target feature information as much as possible (red box). Before passing through DB (b), the image contains a large amount of irrelevant feature information, while after passing through DB (d), DB tries its best to retain the target feature while greatly diluting the irrelevant feature information. Therefore, these visual results indicate that DB can successfully capture the target feature information and dilute the shallow feature information. However, after comparing the target information of (b) and (d), we can clearly see the information loss of target features. This shows that only passing through the DB is not enough to capture the complete target features. Therefore, we need to capture as much as possible the missing target information.

Table 3
Experimental results of DM indexes for right ventricular segmentation.

Method	Test1		Test2	
	Endo	Epi	Endo	Epi
U-net	0.74	0.80	0.80	0.84
MB/U-net	0.76	0.81	0.83	0.85
R3U-net	0.75	0.80	0.81	0.85
MB/R3U-net	0.76	0.81	0.83	0.86

3.4.2. *Multi-Layer Pool Block (MB) experimental analysis*

MB mainly relies on multiple effective fields of view to obtain different target information after DB and supplement the missing target information of DB. In Fig. 8, we show the process of MB supplementing target information. MB

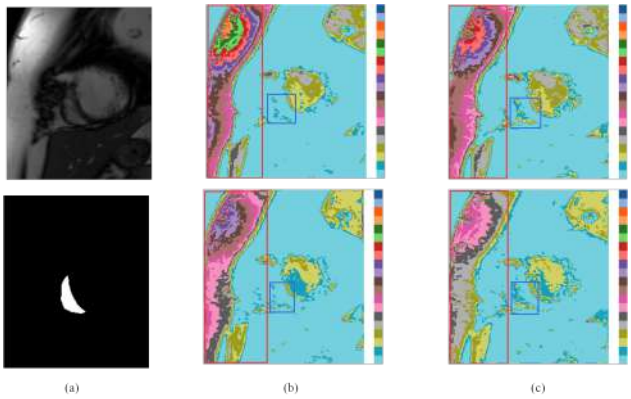

Figure 11: Comparison of feature maps transferred by the MB module. (a) Original image and groundtruth, (b) the feature map before inserting MB modules (upper layer is U-net and the lower layer is R3U-net), (c) the feature map before inserting MB modules, and (d) the feature map after inserting MB modules.

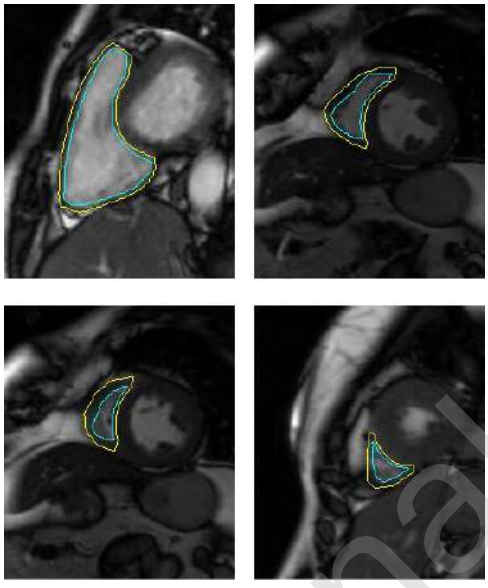

Figure 12: The yellow line is the epicardium and the green line is the endocardium.

Table 4
Experimental results of the MMWHS 2017.

Method	AA	LA	LV	MYO	PA	RA	RV
U-net	0.88	0.82	0.83	0.76	0.74	0.74	0.82
MB/U-net	0.91	0.83	0.81	0.75	0.75	0.78	0.81
R3U-net	0.90	0.83	0.82	0.73	0.75	0.76	0.80
MBR3U-net	0.91	0.85	0.83	0.75	0.76	0.79	0.82

mainly uses several different receiving domains (max pool 2×2, 4×4, and 8×8) to capture different target information (Fig. 8 (c)). It can further dilute irrelevant feature information (blue box) and add target information (red box). However, it will also add some irrelevant target features (yellow box), so that we do not use MB to supplement target information in the decoding layer. In short, MB uses different receiving domains to capture targets, which improves the weight proportion of target features (red box) and reduces the weight proportion of irrelevant target features (blue box). In addition, we show the feature change process of more images which are difficult to extract the target features after MB from the Fig. 9. After MB, the weight proportion of shallow features (blue frame) decreases, and the weight proportion of target features (red frame) increases. These visual results indicate that MB can supplement target feature information and dilute shallow features.

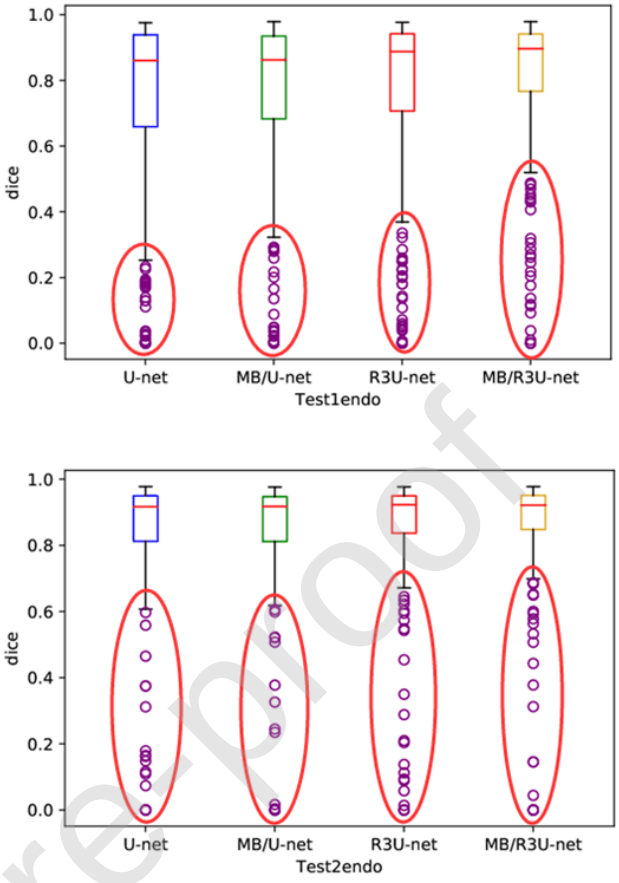

Figure 13: Endocardial results of U-net, MB/U-net, R3U-net and MB/R3U-net in Test1 and Test2.

Not only that, we attempted to embed MB into other network modules to further illustrate its ability to capture and fill in target feature information. U-net network is a common basic network, and has good segmentation performance on RVSC dataset [21]. Therefore, we chose U-net and its similar network R3U-net as the basic network architecture to verify the role of MB. As shown in Fig. 10, each layer of U-net structure contains two layers of 3×3 convolutions, while each layer of R3U-net contains three layers of 3×3 convolutions. We only considered the impact of MB on the basic framework; therefore, we only embedded the MB structure in the encoding layer. In addition, we used the same RVSC dataset as TSU-net and specify the network parameters as Adam optimizer [34] with a learning rate of 1e−3, a drop-

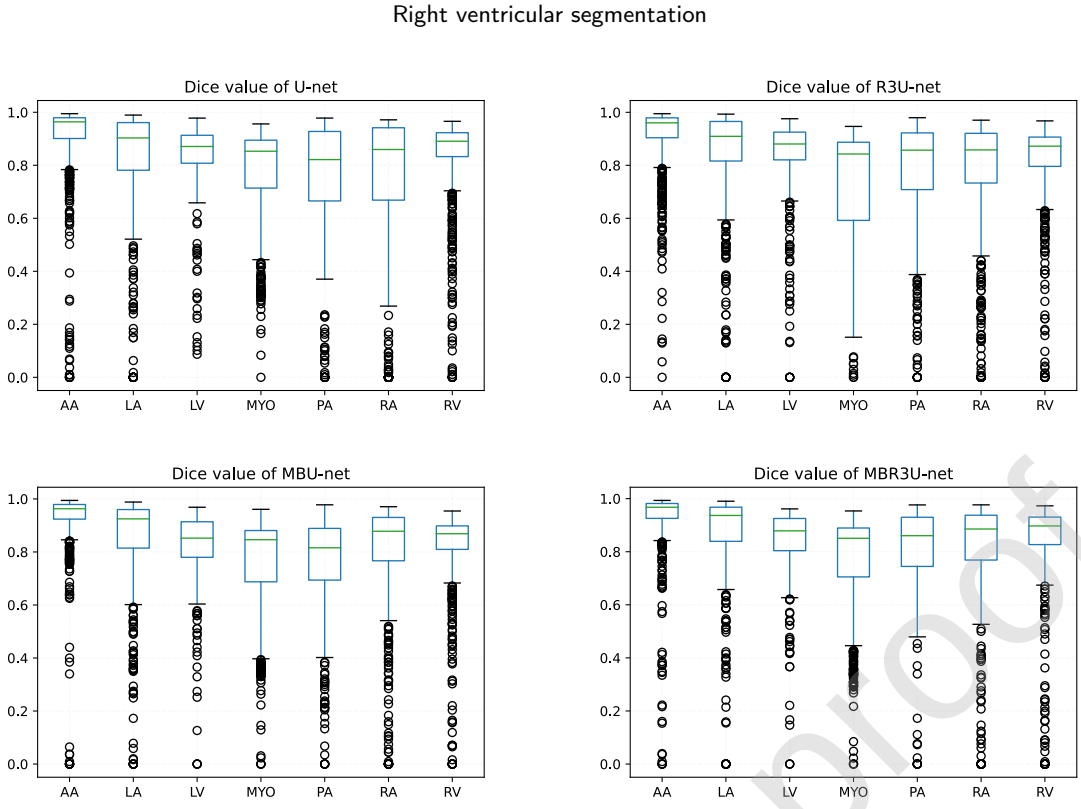

Figure 14: Results of U-net, MB/U-net, R3U-net and MB/R3U-net in MMWHS 2017.

out rate of 0.1, and a mini-batch size of 5 was applied. We think that what the basic network lacks is the ability to detect targets of different sizes and to supplement target features. Therefore, we believe that embedding MBs into the coding layer of the basic network will effectively improve the capability of the basic network.

As shown in Fig. 11, we show the changes in U-net (upper layer) and R3U-net (lower layer) before and after MB. U-net and R3U-net networks have two significant disadvantages: 1) the ability of network to dilute shallow features is weak (red box in Fig. 11(b)); 2) target capture ability is weak (blue box in Fig. 11 (b)). MB can dilute the shallow features and supplement the performance of target features, which can effectively alleviate these two defects. As shown in Fig. 11, after MB, the shallow features in the feature map are diluted to a certain extent (red box in Fig. 11 (c)) and target features are supplemented to some extent (blue frame in Fig. 11 (c)). Therefore, these views just further reflect MB's ability to detect and supplement target features. In addition, we show the experimental structure of the comparative experiment in Table 1. After adding MB, the experimental structure of U-net and R3U-net has been greatly improved. Moreover, as shown in the Fig. 12 , the target size of endocardium is smaller than that of epicardium. Therefore, after adding MB, the DM index growth of endocardium is larger than that of epicardium, which can show that MB can improve the capture ability of small targets. In addition, as shown in Fig. 13, we show the whole experimental structure intuitively in the way of a box diagram. Because some difficult points are effectively improved (red box in Fig. 13), the experimental results in Table 3 are greatly improved. This also reflects that the basic network cannot effectively solve the problem of large changes in target size. After adding MB, the ability of network to capture different size targets is enhanced, so that the network performance is effectively improved.

In addition, in order to verify the generalization ability of the model, we also conducted experiments on mmwhs 2017 data set. As shown in Table. 4, the basic model can effectively improve the network performance after adding MB module. Futhermore, after adding MB module, the model has made a breakthrough for RA with many small targets. Besides, as shown in Fig. 14, after adding MB module to the basic model, the number of singular points is mostly reduced ($DM < 0.6$). This not only shows that MB module does help model capture small targets, but also proves that MB module can help basic model capture small targets of different data sets and different segmentation targets.

3.4.3. Batch Normalization and small bitch-size

With the deepening of the network layer, the influence of parameters on distribution is uncertain. As a result, each layer's input distribution and the same layer in different iterations changes, which makes the network need to adapt to the new distribution, forcing us to reduce the learning rate and reduce the impact [37, 38]. However, the BN process uses the mean and variance of samples in batch to simulate all data's mean and variance; thus, when the batch size is

TSU-net				
MB1	MB2	MB3	MB4	MB5
Max pool 2,4,8	Max pool 2,4	Max pool 2	Max pool 2,4	Max pool 2
Concat	Concat	Concat	Concat	Concat
Conv1×1+BN	Conv1×1+BN	Conv1×1+BN	Conv1×1	Conv1×1
DB1	DB2,DB6,DB7	DB3,DB5,DB8	DB4,DB9	DB10
Conv 3×3,16	Conv 3×3,32	Conv 3×3,64	Conv 3×3,128	Conv 3×3,64
Rate=1,2,4,8	Rate=1,2,4,8	Rate=1,2,4,8	Rate=1,2,4,8	Rate=1,2,4,8
Relu+BN	Relu+BN	Relu+BN	Relu+BN	Relu+BN
Concat	Concat	Concat	Concat	Concat
Conv1×1+BN	Conv1×1+BN	Conv1×1+BN	Conv1×1+BN	Conv1×1
DB11	DB12			
Conv 3×3,32	Conv 3×3,16			
Rate=1,2,4,8	Rate=1,2,4,8			
Relu+BN	Relu+BN			
Concat	Concat			
Conv1×1	Conv1×1			

Figure 15: The basic structure of TSU-net.MB4 and MB5 reduce the use of BN, while db10, DB11 and DB12 reduce the use of BN.

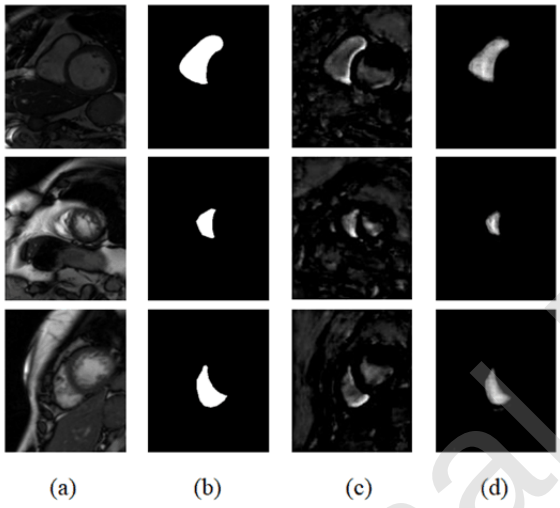

Figure 16: (a) Original image, (b) ground truth, (c) stage1 encode-decode output and (d) stage2 encode-decode output.

small, the simulation results are pathetic [38]. There are two main ways in this research. One is to use BN together with a little bit of size regardless of its influence. One is to give up BN when using a small bitch size. However, these two ways have not been reasonable to alleviate the contradiction. In order to be compatible with the advantages of BN and small bitch-size for our dataset, we boldly reduce the use of BN in the network. We believe that reducing BN use in the connection layer at the end of the network can alleviate the negative impact between BN and small bitch-size. In order to show more clearly what we mean by reducing usage, we show the main structure of TSU-net in Fig. 15. We clearly see that in MB4, MB5, DB11, and DB12, we reduce BN use. We believe that reducing BN use at the end of the network can alleviate the impact of BN and small bit size.

In order to better show the effectiveness of reducing the use of BN proposed by us, we set up three different networks

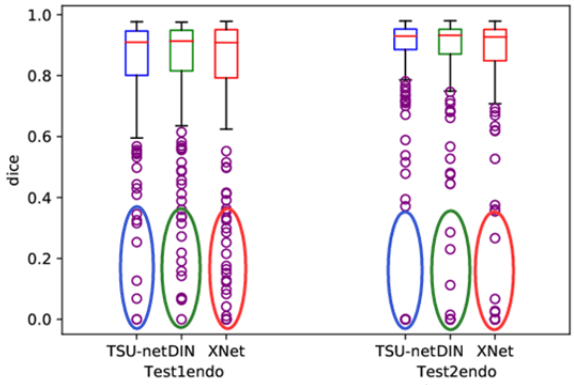

Figure 17: (a) Original image, (b) ground truth, (c) stage1 encode-decode output and (d) stage2 encode-decode output.

without changing the parameters to verify our conjecture. To better show the effectiveness of reducing BN use, we set up three systems with all BN, some BN, and no BN at all without changing the parameters and applying them to bitch-size of different sizes.

In Table. 5, we can see the positive impact of BN and small bitch size equal to 5 on our network structure and data set. When we did not use BN in our network, our results will still show a cliff-like decline even when compared with using BN. We can find that RVSC is especially suitable for small bitch-size. This is why we are not willing to abandon either BN or small bitch-size. Fortunately, we had founded a way to utilize BN and little bitch-size at the same time. In Table. 5, we can clearly see that the results have been greatly improved when using the method we set up. In Table. 6 and Table. 7, we can also see that our experimental results outperform the best practical results at present.

3.4.4. TSU-net experimental results

The proposed MB effectively detects the target features and supplements the target information. However, it is a fly in the ointment that MB will still supplement a small number of shallow features to affect the experimental results. In addition, compared with the two-stage encode-decode architecture, the ability of one-stage encode-decode architecture to dilute shallow features is limited, and it is easy to survive a large number of shallow features. Moreover, two-stage encode-decode not only has a better ability to dilute shallow features but also can improve the memory ability of the network for target features [27]. Therefore, the proposed TSU-net structure used two-stage encode and decode (Stage1 and Stage2 in Fig. 2)) to further dilute the shallow features and improve the experimental performance of the network. As shown in Fig. 16), we show the results of one-stage (Fig. 16(c))) and two-stage (Fig. 16(d))). As we expected, compared with the two-stage encode-decode network structure, only one-stage encode -decode will retain a large number of shallow features and reduce the experimental results.

In the Test1 and Test2 groups, the DM values and HD

Table 5
DM indexes selected for different number of BN as well as different size of bite size values.

Method	bitch-size=2		bitch-size=5		bitch-size=10	
	Test1 Endo	Test2 Endo	Test1 Endo	Test2 Endo	Test1 Endo	Test2 Endo
All have BN	0.823	0.852	0.837	0.871	0.830	0.851
NO BN	0.810	0.810	0.820	0.853	0.812	0.840
TSU-net	0.836	0.861	0.844	0.881	0.834	0.856

Table 6
DM(±standard deviation) of different methods on test1 set and test2 set

Method	Test1		Test2		Average	
	Endo	Epi	Endo	Epi	Endo	Epi
CMIC [41]	0.78(0.23)	0.82(0.19)	0.73(0.27)	0.77(0.24)	0.76(-)	0.80(-)
NTUST [15]	0.57(0.23)	0.62(0.35)	0.61(0.34)	0.64(0.35)	0.59(-)	0.63(-)
SBIA [13]	0.55(0.32)	0.58(0.29)	0.61(0.29)	0.68(0.25)	0.58(-)	0.63(-)
Jord.[42]	0.83(0.16)	0.86(0.11)	0.83(0.18)	0.86(0.14)	0.83(-)	0.86(-)
FCN[19]	-	-	-	-	0.84(0.21)	0.86(0.20)
U-net[21]	0.74(0.32)	0.80(0.27)	0.80(0.28)	0.84(0.25)	0.77(0.28)	0.82(0.26)
Dilated CNN[28]	0.75(0.30)	0.80(0.26)	0.80(0.27)	0.85(0.23)	0.78(0.28)	0.82(0.25)
Inception CNN[22]	0.80(0.26)	0.86(0.19)	0.85(0.22)	0.89(0.17)	0.82(0.24)	0.87(0.18)
Residual U-net [30]	-	-	-	-	0.84(0.16)	0.89(0.15)
XNet[32]	0.80(0.27)	0.85(0.23)	0.85(0.20)	0.89(0.16)	0.82(0.23)	0.87(0.19)
Dilated-Inception Net[35]	0.83(0.22)	0.89(0.11)	0.86(0.19)	0.90(0.16)	0.85(0.20)	0.89(0.14)
TSU-net	0.84(0.22)	0.90(0.10)	0.88(0.18)	0.91(0.17)	0.86(0.19)	0.90(0.14)

of TSU-net implementation are listed in Table 6 and Table. 7. In addition, many of the most advanced methods are available for comparison. Among these methods, CMIC [41], SBIA [13], and NTUST [15] are the best three methods in the right ventricular segmentation challenge [36]. At present, the most advanced right ventricular segmentation methods are DIN [29], Residual U-net [30], Jord [42]. Dilated CNN[28], Inception CNN[27], FCN [19], U-net [21] and XNet [32] are new method for biomedical image segmentation. As shown in Fig. 17, we present the experimental results intuitively by box diagram of TSU- net, DIN [28] and XNet [32]. As we analyzed earlier, the number of singular points in TSU-net (blue box in Fig. 17) is significantly less than that in DIN (green box in Fig. 17 and XNet (red box in Fig. 17). This also reflects the necessity of solving the problem of large difference of right ventricular target. Therefore, as shown in Table 6 and Table. 7, our experimental results are better than the latest methods. In Table 6, the enhancement of TSU-net network to endocardium is greater than that to epicardium, which indicates that TSU-net network has stronger ability to capture small targets (the size of endocardium target is smaller than that of epicardium target) compared with other networks. In addition, as shown in Fig. 18, we take out some experimental results of different sizes of targets, which can further prove that the segmentation performance of our network for different size targets is better than the latest methods.

In addition, we generalize our network with other methods that achieve excellent results on RVSC datasets. As shown in Table 8, we show the DM metrics of these methods on the

Table 7
HD(±standard deviation) of different methods on test1 set and test2 set

Method	Test1		Test2		Average	
	Endo	Epi	Endo	Epi	Endo	Epi
CMIC [35]	10.51(9.17)	10.94(8.32)	12.50(10.95)	12.70(10.44)	11.49(-)	11.80(-)
NTUST [15]	28.44 (23.57)	26.71(22.90)	22.20(21.74)	22.14(25.38)	25.38(-)	24.47 (-)
SBIA [13]	23.16(19.86)	22.53(18.06)	15.08(8.91)	15.17(8.88)	19.20(-)	18.92(-)
Jord.[36]	9.05(6.98)	9.60(7.01)	8.73(7.62)	9.00(7.46)	8.89(-)	9.31(-)
FCN[19]	-	-	-	-	8.86(11.27)	9.33(10.79)
U-net[20]	15.91(21.58)	15.65(20.67)	12.78(19.83)	11.78(18.00)	14.38(20.03)	13.75(19.13)
Dilated CNN[23]	17.05(24.31)	15.12(20.88)	10.57(15.93)	11.23(17.02)	13.87(20.87)	13.21(19.17)
Inception CNN[22]	10.43(15.51)	9.84(13.29)	8.44(14.73)	7.36(10.36)	9.46(15.15)	8.62(12.00)
Dilated-Inception Net[30]	7.71(7.60)	7.46(7.41)	5.93(6.01)	6.49(7.96)	6.84(6.93)	6.99(7.69)
Residual U-net [25]	-	-	-	-	8.05(11.14)	7.14(8.49)
TSU-net	8.06(7.16)	6.95(6.49)	5.12(5.01)	5.65(6.42)	6.56(6.21)	6.35(6.45)

Table 8
DM of different methods on MMWHS 2017 test

Method	AA	LA	LV	MYO	PA	RA	RV
U-net[21]	0.88	0.82	0.83	0.76	0.74	0.74	0.82
XNet[27]	0.90	0.86	0.86	0.78	0.76	0.78	0.85
DIN[30]	0.88	0.84	0.82	0.73	0.72	0.75	0.81
TSU-net	0.92	0.86	0.85	0.80	0.79	0.83	0.86

Table 9
Accuracy and Error rate of different methods on RVSC test

Method	Accuracy	Error rate
U-net	0.98	0.02
TSU-net	0.99	0.01

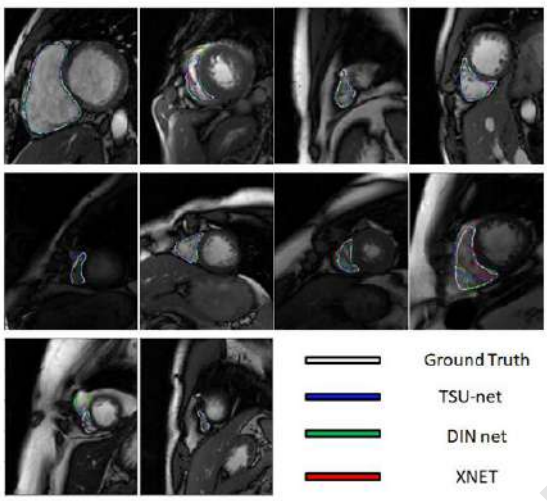

Figure 18: Endocardial results of TSU-net, DIN and XNet in Test1 and Test2.

MMWHS dataset. As we expected, DIN can improve the network performance based on lifting capture RVSC right ventricular characteristics, but this makes the network have strong limitations. Therefore, in the MMWHS data set, the performance of DIN network is greatly degraded. XNet is based on improving the application ability of the network for small data, so XNet still achieves good performance on small data such as MMWHS. Our method can not only improve the application ability of the network for small data, but also capture small targets well. Therefore, we have achieved good results on MMWHS dataset. As shown in Fig. 19, the number of singular points (DM < 0.6) of our method in each target is significantly less than that of other methods, which just shows that our method improves the performance of the network by improving the ability of the network to capture small targets. In addition, compared with other networks, our network has achieved very good performance for RA and PA. this is because RA and PA contain a large number of small targets, and other networks are difficult to effectively use the information of small targets. Our network just makes up for this defect and achieves such superior performance.

As shown in Fig. 20, because the proportion of black background (pixel equal to 0) is large enough and easy to be classified successfully, while the white target area (pixel equal to 1) is small enough and difficult to be classified successfully. Therefore, on the RVSC data set, the accuracy and error rate can easily achieve good results, and a small deviation of accuracy and error rate can reflect that the segmentation performance of white target region is very different. As shown in table R5, the accuracy of our model on the test set is 0.99 and the error rate is as low as 0.01, which is better than that of U-net network model (0.98) and error rate (0.02). This shows that our network performance is obviously better than U-net network model. In addition, our network can predict 99% successful pixels and only 1% error rate, which indicates that our network has excellent performance. This shows that our network performance is obviously better than the U-net network model. Moreover, our network can predict 99% of the successful pixels, which shows that our network has excellent performance.

3.4.5. Clinical Performance

In practice, it is very important to know whether the proposed right ventricular segmentation method is suitable for clinical application. Therefore, we compared the clinical manifestations of the proposed TSU-net with those of clinical experts. Specifically, we analyzed four clinical cardiac indices predicted by din and clinical experts.

In clinical routine, end diastolic endocardial volume (EDV), systolic endocardial volume (ESV), ejection fraction (EF) and ventricular mass (VM) are four clinical cardiac indicators, which are widely used to quantify and analyze global and local cardiac function [7]. EDV and ESV were calculated in ml, that is, the sum of all right ventricular areas multiplied by the 8.4 mm spacer value in rvsc database [7]. EF and VM are defined by EDV and ESV as:

$$EF = \frac{EDV - ESV}{EDV} \tag{7}$$

$$VM = P * EDV_{epi} - EDV_{endo} \tag{8}$$

where denotes epicardial volume at end-diastole, is endocardial volume at end-diastole, is 1.05 g/cm3 for each patient in RVSC database [7].

Based on the data of 32 patients with test1 and test2, the cardiac indexes (EF and VM) obtained by din and clinical experts were analyzed by linear regression and Bland Altman diagram [44,45]. According to the recommended din and clinical experts, linear regression method was used to fit the cardiac index of 32 patients. Bland Altman diagram was used to analyze the differences between the clinical manifestations of the model and clinical experts.

As shown in the top of Fig. 21,the linear regression line of EF is

$$y = 0.94x + 0.02 \tag{9}$$

and the the linear regression line of VM is

$$y = 1.02x + 2.21 \tag{10}$$

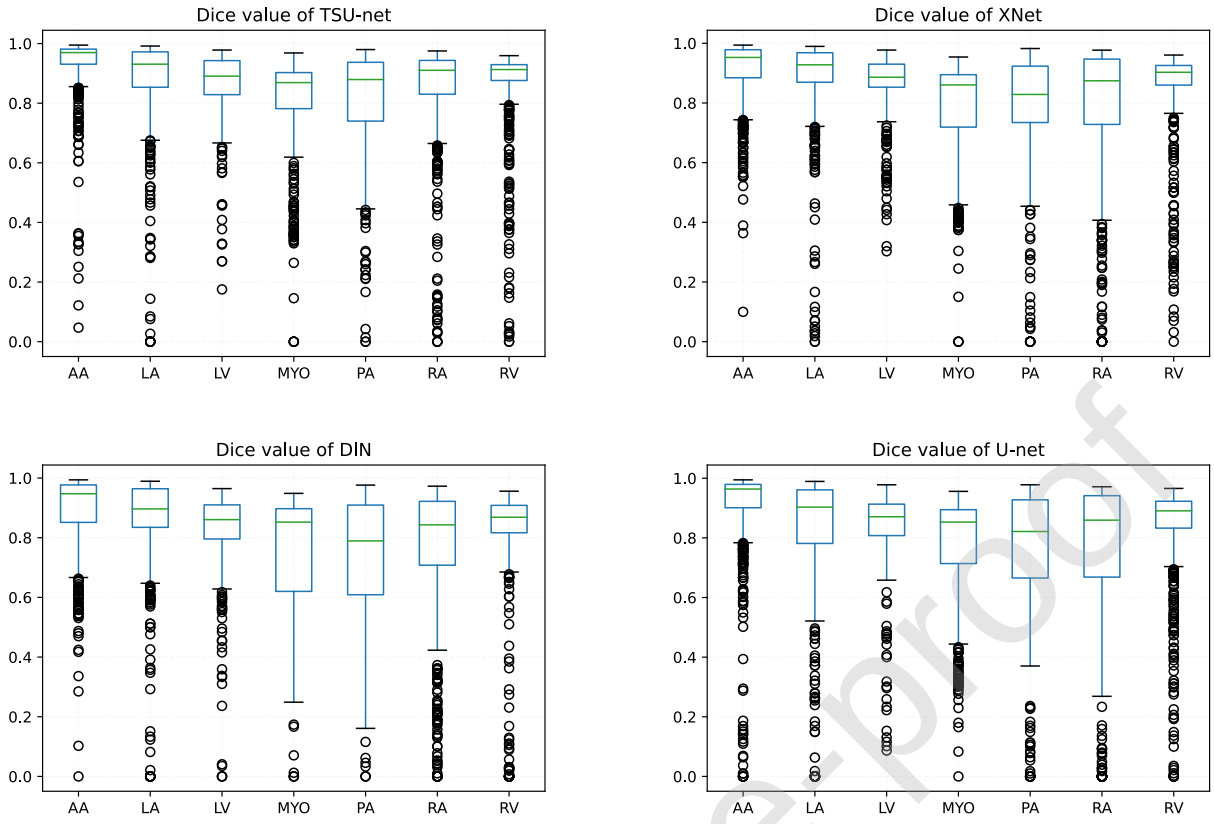

Figure 19: Results of TSU-net, DIN, XNet and U-net in MMWHS 2017.

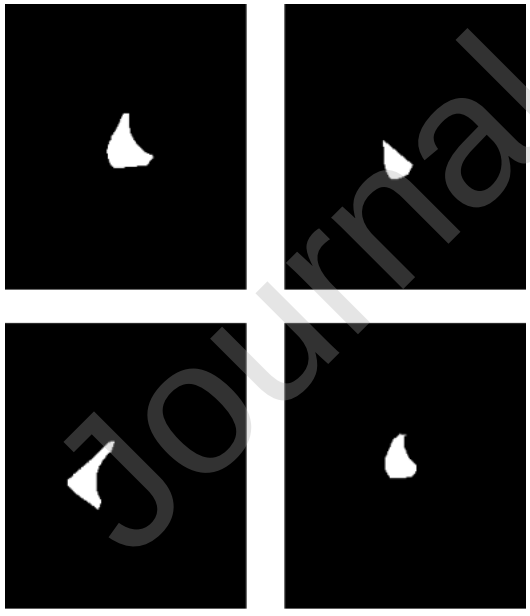

Figure 20: The size of most targets in the right ventricular dataset

,This shows that the EF and VM of TSU-net and clinical experts are very close. As shown in the low of Fig. 21, the differences between EF and VM obtained by TSU-net(automatic method) and clinical experts (manual method) are shown in the Bland Altman graph. Bland Altman diagram is widely used to evaluate the reproducibility within and between observers to determine whether the new observer's prediction is applicable. Bland Altman diagram showed that EF and VM of 32 patients (32 cases) were within 95% of the coincidence range, which could be used for clinical diagnosis $\pm$ 2). In addition, Bland Altman diagram shows that TSU-net can be used to evaluate cardiac index in many patients. Therefore, the proposed TSU net has a further clinical application prospect in cardiac diagnosis.

4. Discussion

Different from the latest methods, we consider the problem of large changes in target size in the RVSC dataset. In Fig. 22, we show the small target segmentation result of Test1 Endo, which is the most difficult to segment. Our network improves the segmentation ability of these very difficult small targets and reduces the occurrence of singular points in the segmentation results. Because of this, the experimental results of our network are better than other experimental results. However, as shown in Fig. 17, we greatly improve the acquisition performance of small targets, but slightly reduce the acquisition performance of large targets. Therefore, in the following work, we need to further improve our network,

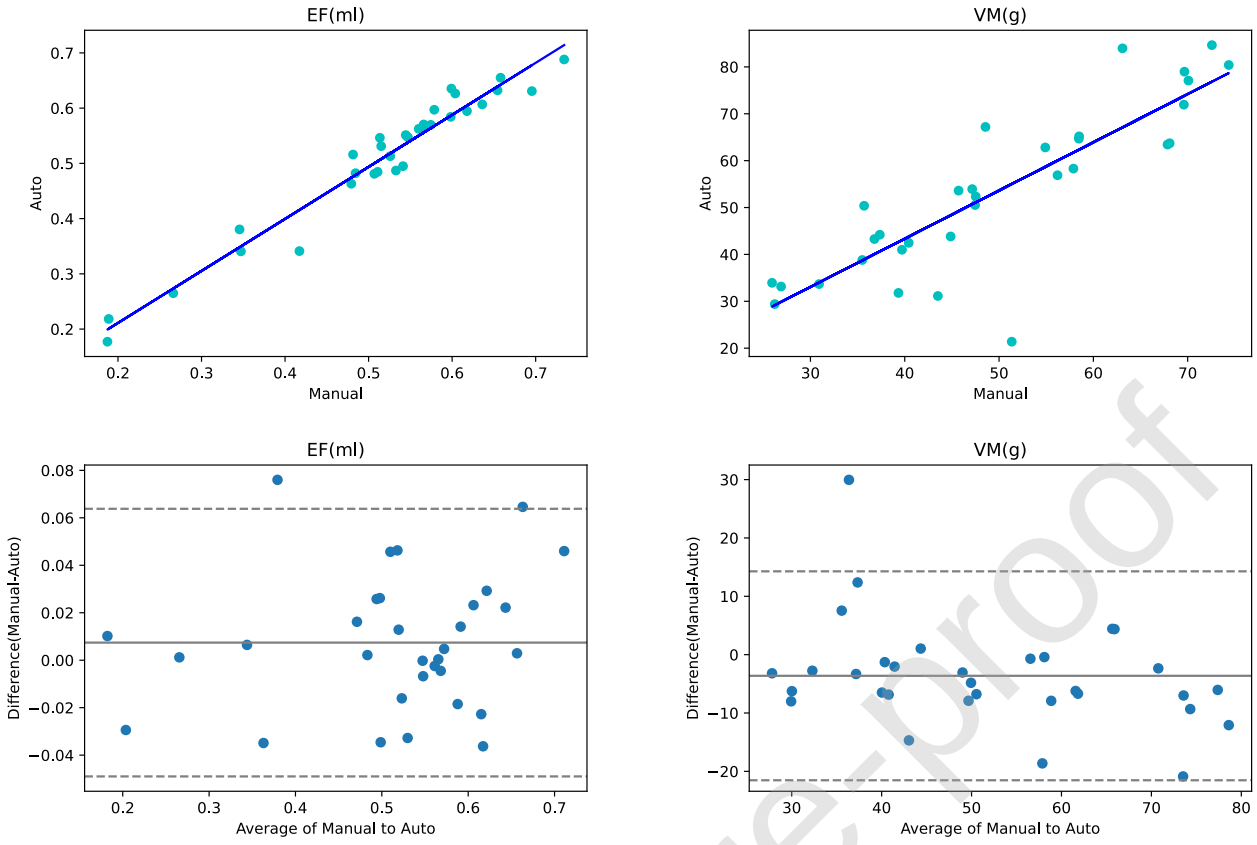

Figure 21: Analysis of different cardiac indices of 32 patients based on TSU-net (auto approach) and clinical experts (manual approach). Top row: Linear regression (blue lines are regression lines). Bottom row: Bland-Altman plots (x axis = average indices obtained by auto and manual approaches, y axis = differences between two approaches, black lines denotes mean difference and denotes standard deviation of differences, dashed lines indicate the 95% limits of agreements (± 2)).

Figure 22: On the left is the result of TSU-net segmentation, and on the right is the ground truth.

to ensure that the network can improve the performance of small targets and the performance of large targets.

5. Conclusion

MRI data of automatic right ventricular segmentation is of great significance for cardiac pathology diagnosis. However, the difference in the size of the right ventricular target I have always been a difficulty in the right ventricular segmentation. Therefore, this paper proposes a Multi-Layer-Pool Block (MB), which is embedded into the coding layer, and uses multiple receiving domains to detect targets of different sizes. We used the multi-scale convolutional neural network to capture the deep features, and propose the combination of TSU-net network and MB. In the present RVSC dataset, we have obtained good results. However, we still need to search for more MRI right ventricular data sets next time to find a better the right ventricular segmentation network.

However, we are not limited to the results achieved at present. In the future work, we will improve our network, improve the ability of network to capture small targets, and alleviate the negative impact between large targets and small targets. Of course, the lack of right ventricular data sets is also a problem we are facing. We need to find more right ventricular data sets to better verify the performance of our network. At the same time, we also hope to use the features of other right ventricular datasets to assist our model segmentation.

References

[1] Virani, S. S. , Alonso, A. , Benjamin, E. J. , Bittencourt, M. S. , Tsao, C. W. . (2020). Heart disease and stroke statistics—2020 update: a report from the american heart association. *Circulation*, 141(9), CIR0000000000000757-.

[2] Attili, A. K. , Schuster, A. , Nagel, E. , Reiber, J. H. C. , Geest, R. J. V. D. . (2010). Quantification in cardiac mri: advances in image acquisition and processing. *The International Journal of Cardiovascular Imaging*.

[3] Petitjean, C. , Dacher, J. N. . (2011). A review of segmentation methods in short axis cardiac mr images. *Medical Image Analysis*, 15(2), 169-184.

[4] Amer, A. , Ye, X. , Zolgharni, M. , Janan, F . . (2020). ResDUnet: Residual Dilated UNet for Left Ventricle Segmentation from Echocardiographic Images. *2020 42nd Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC) in conjunction with the 43rd Annual Conference of the Canadian Medical and Biological Engineering Society.* IEEE.

[5] Du, X. , Tang, R. , Yin, S. , Zhang, Y. , Li, S. . (2018). Direct segmentation-based full quantification for left ventricle via deep multi-task regression learning network. *IEEE Journal of Biomedical Health Informatics*, 1-1.

[6] Dong, Z. , Du, X. , Liu, Y. . (2020). Automatic segmentation of left ventricle using parallel end-end deep convolutional neural networks framework. *Knowledge-Based Systems*, 204, 106210.

[7] C.Petitjean, W.j.Bai, Right ventricle segmentation from cardiac mri: a collation study. *Medical Image Analysis*, 19(1), 187-202.

[8] Caudron, Fares, J. , Lefebvre, V. , Vivier, P. H. , Petitjean, C. , Dacher, J. N. . (2012). Cardiac mri assessment of right ventricular function in acquired heart disease. *Academic Radiology*, 19(8), 991-1002.

[9] Laurent, Bonnemains. (2012). Assessment of right ventricle volumes and function by cardiac mri: quantification of the regional and global interobserver variability. *Magnetic Resonance in Medicine*, 67(6), 1740-1746.

[10] Abinahed, J. , Jolly, M. P. , Yang, G. Z. . (2006). Robust active shape models: a robust, generic and simple automatic segmentation tool.

[11] Bai, W., Shi, W., O'Regan, D., P. , et al. . (2013). A probabilistic patch-based label fusion model for multi-atlas segmentation with registration refinement: application to cardiac mr images. *IEEE Transactions on Medical Imaging*, 32(7), 1302-1315.

[12] Zuluaga, M. A. , Cardoso, M. J. , Modat, M. , Sébastien Ourselin. (2013). Multi-atlas Propagation Whole Heart Segmentation from MRI and CTA Using a Local Normalised Correlation Coefficient Criterion. *International Conference on Functional Imaging Modeling of the Heart*. Springer-Verlag.

[13] Ou, Y. , Sotiras, A. , Paragios, N. , Davatzikos, C. . (2011). DRAMMS: Deformable registration via attribute matching and mutual-saliency weighting. (Vol.15, pp.622-639). es.

[14] Ruan, P. N. D. . (2013). Graph cut segmentation with a statistical shape model in cardiac {mri· Computer Vision and Image Understanding.

[15] C. Wang. . (2012) A simple and fully automatic right ventricle segmentation method for 4-dimensional cardiac MR images.

[16] Mahapatra, D. , Buhmann, J. M. . (2013). Automatic cardiac RV segmentation using semantic information with graph cuts. *IEEE International Symposium on Biomedical Imaging*. IEEE.

[17] Stalidis, G. , Maglaveras, N. , Efstratiadis, S. N. , Dimitriadis, A. S. , Pappas, C. . (2002). Model-based processing scheme for quantitative 4-d cardiac mri analysis. *IEEE Transactions on Information Technology in Biomedicine*, 6(1), 59-72.

[18] Litjens, G. , Kooi, T. , Bejnordi, B. E. , Setio, A. A. A. , Clara I. Sánchez. . (2017). A survey on deep learning in medical image analysis. *Medical Image Analysis*, 42(9), 60-88.

[19] PV Tran. . (2016) A Fully Convolutional Neural Network for Cardiac Segmentation in Short-Axis MRI. arXiv:1604.00494v3.

[20] O Ronneberger, P Fischer, T Brox. . (2015) U-Net: Convolutional Networks for Biomedical Image Segmentation. DOI10.1007/978-3-319-24574-4-28.

[21] iek, zgün, Abdulkadir, A. , Lienkamp, S. S. , Brox, T. , Ronneberger, O. . (2016). 3d u-net: learning dense volumetric segmentation from sparse annotation. DOI: 10.1007/978-3-319-46723-8-49.

[22] C.Yu ,Z.Gao, W.Zhang, G.Yang, S.Li. Multitask learning for estimating multitype cardiac indices in mri and ct based on adversarial reverse mapping, *IEEE Transactions on Neural Networks and Learning Systems*,vol.1,no.14,pp.99,2020.

[23] R.Ge, G.Yang,Y.Chen,L.Luo,S.Li, K-net: integrate left ventricle segmentation and direct quantification of paired echo sequence, *IEEE Transactions on Medical Imaging*, vol.1,no.1,PP.99, 1-1.2019.

[24] H. H.Kim, B. W.Hong, Segmentation neural network incorporating scale-space in the application of cardiac MRI, *Journal of Medical Imaging and Health Informatics*, 2020.

[25] T.Liu, Y.Tian, S.Zhao, X.Huang, Q.Wang, Residual convolutional neural network for cardiac image segmentation and heart disease diagnosis, *IEEE Access*, vol.1,no.1, PP.99, 2020.

[26] C.Xu, L.Xu, P.Ohorodnyk,M.Roth,S.Li, Contrast agent-free synthesis and segmentation of ischemic heart disease images using progressive sequential causal gans, *Medical Image Analysis*, 2020.

[27] C Szegedy, W Liu, Y Jia, P Sermanet, S Reed, D Anguelov, D Erhan, V Vanhoucke, A Rabinovich. . (2014) Going Deeper with Convolutions. DOI: 10.1109/CVPR.2015.7298594.

[28] Yu, F. , Koltun, V. . (2016). Multi-scale context aggregation by dilated convolutions.

[29] Li, J. , Yu, Z. L. , Gu, Z. , Liu, H. , Li, Y. . (2019). Dilated-inception net: multi-scale feature aggregation for cardiac right ventricle segmentation. *IEEE Transactions on Biomedical Engineering*, 66(12), 3499-3508.

[30] Liu, Z. , Feng, Y. , Yang, X. . (2019). Right Ventricle Segmentation of Cine MRI Using Residual U-net Convolutinal Networks. *20th International Conference on Parallel and Distributed Computing, Applications and Technologies (PDCAT)*

[31] S Zagoruyko, N Komodakis. . (2016) Wide Residual Networks, arXiv:1605.07146.

[32] J Bullock, C Cuesta-Lazaro, A Quera-Bofarull. . (2018) XNet: A convolutional neural network (CNN) implementation for medical X-Ray image segmentation suitable for small datasets. DOI:10.1117/12.2512451.

[33] He, K. , Zhang, X. , Ren, S. , Sun, J. . (2014). Spatial pyramid pooling in deep convolutional networks for visual recognition. *IEEE Transactions on Pattern Analysis Machine Intelligence*, 37(9), 1904-16.

[34] Drozdzal, M. , Vorontsov, E. , Chartrand, G. , Kadoury, S. , Pal, C. . (2016). The Importance of Skip Connections in Biomedical Image Segmentation. *International Workshop on Large-scale Annotation of Biomedical Data Expert Label Synthesis International Workshop on Deep Learning in Medical Image Analysis.* Springer International Publishing.

[35] A Krizhevsky, I Sutskever, G Hinton. ImageNet Classification with Deep Convolutional Neural Networks, 2012. DOI:10.1145/3065386.

[36] C. Petitjean, M. A. Zuluaga, W. Bai, J. N. Dacher, Right ventricle segmentation from cardiac MRI: a collation study, Med Image Anal, vol. 19, no. 1, pp. 187-202, Jan, 2015.

[37] S Ioffe, C Szegedy. . (2015) Batch Normalization: Accelerating Deep Network Training by Reducing Internal Covariate Shift..

[38] Santurkar, S. , Tsipras, D. , Ilyas, A. , Madry, A. .(2018) How does batch normalization help optimization?.

[39] Yong, H. , Huang, J. , Meng, D. , Hua, X. , Zhang, L. .(2020). Momentum batch normalization for deep learning with small batch size.

[40] Kingma, D. , Ba, J..(2014). Adam: a method for stochastic optimization.

[41] Zuluaga, M. A. , Cardoso, M. J. , Modat, M. , Sébastien Ourselin,. .(2013). Multi-atlas Propagation Whole Heart Segmentation from MRI and CTA Using a Local Normalised Correlation Coefficient Criterion. *International Conference on Functional Imaging Modeling of the Heart*. Springer-Verlag.

[42] Ringenberg, J. , Deo, M. , Devabhaktuni, V. , Berenfeld, O. , Gold, J. . (2014). Fast, accurate, and fully automatic segmentation of the right ventricle in short-axis cardiac mri. *Computerized Medical Imaging*

Graphics, 38(3).

[43] B, X. Z. A, C, L. L. , D, C. P, E, D. T , E, M. U. Evaluation of algorithms for multi-modality whole heart segmentation: an open-access grand challenge. *Medical Image Analysis*, 58. 2017.

[44] Jérôme, Caudron, Jeannette, Fares, Valentin, Lefebvre, et al. Cardiac MRI assessment of right ventricular function in acquired heart disease: factors of variability, *Academic Radiology*,vol.19,no.8,p.991-1002, 2012.

[45] J. M. Bland, D.G. Altman, Methods for assessing agreement between two methods of clinical measurement, *Lancet*, vol. 327, no. 8476, pp. 307-310,1986.