Applied Intelligence (2022) 52:5225–5240
https://doi.org/10.1007/s10489-021-02720-9

MMNet: A multi-scale deep learning network for the left ventricular segmentation of cardiac MRI images

Ziyue Wang[1] · Yanjun Peng[1,2] · Dapeng Li[1] · Yanfei Guo[1] · Bin Zhang[1]

Accepted: 26 July 2021 / Published online: 6 August 2021

Abstract

With the development of deep learning network models, the automatic segmentation of medical images is becoming increasingly popular. Left ventricular cavity segmentation is an important step in the diagnosis of cardiac disease, but post-processing segmentation is a time-consuming and challenging task. That is why a fully automated segmentation method can assist specialists in increasing their efficiency. Inspired by the power of deep neural networks, a multi-scale multi-skip connection network (MMNet) model is proposed to fully automate the left ventricular segmentation of cardiac magnetic resonance imaging (MRI) images; this model is simple and efficient and has high segmentation accuracy without pre-detecting left ventricular localization. MMNet redesigns the classic encoder and decoder to take advantage of multi-scale feature information, effectively solving the problem of difficult segmentation due to blurred left ventricular edge information and the low accuracy of end-systolic segmentation of the cardiac area. In the model encoding stage, a multi-scale feature fusion module applying dilated convolution is proposed to obtain richer semantic information from different perceptual fields. The decoding stage reconstructs the full-size skip connection structure to make full use of the feature information obtained from different layers for contextual semantic information fusion. At the same time, a pre-activation module is used before each weighting layer to prevent overfitting phenomena from arising. The experimental results demonstrate that the proposed model has better segmentation performance than advanced benchmark models. Ablation experiments show that the proposed modules are effective at improving segmentation results. Therefore, MMNet is a promising approach for the left ventricular fully automated segmentation.

Keywords Left ventricular segmentation · Deep learning networks · Cardiac MRI images · Multi-scale feature fusion

1 Introduction

Today, approximately 17.5 million people worldwide die of heart disease each year, accounting for 30% of all deaths. Cardiovascular disease has become a serious threat to human health and one of the top three factors of human mortality. Cardiac imaging techniques are routinely used to evaluate and diagnose heart disease. Computer-based image analysis methods are also widely used in the segmentation and alignment of the heart to extract its anatomy and systolic function. The analysis of cardiac function using imaging instruments has provided new insights into cardiac pathophysiology and unprecedented characterizations of myocardial tissue. Novel acquisition strategies combined with advanced reconstruction techniques enable high-resolution, true 3D dynamic acquisition [1]. Thus, computer analysis has been shown to be effective in reducing mortality and morbidity from cardiovascular disease, and it can help clinicians interpret medical conditions objectively.

The left ventricle plays an important role in the blood circulation and cardiac cycle, which is responsible for the transport of blood throughout the body. Most cardiovascular diseases affect the physiological shape of the left ventricle of the heart, and examining the left ventricle is an important prerequisite for determining whether the heart is diseased. Accurate segmentation of the left ventricle is related to clinical indicators such as ventricular volume, ejection fraction, left ventricular mass, wall thickening, and wall

✉ Yanjun Peng
pengyanjuncn@163.com

[1] College of Computer Science and Engineering, Shandong University of Science and Technology, Qingdao 266590 Shandong, China

[2] Shandong Province Key Laboratory of Wisdom Mining Information Technology, Shandong University of Science and Technology, Qingdao 266590 Shandong, China

motion abnormality. It is also a condition for the quantitative analysis of the heart as a whole and the local. Left ventricular segmentation is therefore a key step in the diagnosis and treatment of cardiovascular disease [2]. The accuracy of left ventricular segmentation also determines the accuracy of the available information about the overall structure of the heart. In clinical practice, left ventricular (LV) segmentation is usually performed by a specialist who outlines the myocardial border. This is a time-consuming and tedious task. Much effort has gone into automating this process. However, it remains a challenging task to obtain a reliable and accurate image of the left ventricle of the heart fully, automatically and quickly [3].

In recent years, MRI has become a common clinical test for the heart. Compared to those of ultrasound, MRI results are more accurate in that they can accurately quantify ejection fractions and examine myocardial and cardiac functions. Moreover, MRI can also evaluate the patient's myocardial perfusion and the specific circumstances of delayed myocardial strengthening. However, there are still factors that affect left ventricular segmentation on MRI images, such as the following [1, 4–7]:

1. Most current segmentation models require left ventricle pre-positioning and redundant learning parameters and have low segmentation efficiency.
2. The different stages of the cardiac cycle, with different proportions of the left ventricle, require the detection of the target organ's characteristics at different scales.
3. Due to papillary muscle interference, the end-systolic segmentation accuracy is generally low.
4. Blurred left ventricular borders and the overlapping of edge information with the background pixel intensity distribution directly affect the extraction and reconstruction of edge information.
5. Differences in the shapes of the LV contours in different sections and stages make the model more difficult to learn.

Deep learning methods have made impressive achievements in the field of computer vision, and convolutional neural networks are also widely used in medical image segmentation. In this paper, a multi-scale multi-skip connection network (MMNet) model is proposed to solve the problem of accurate left ventricular segmentation in cardiac MRI images. Our main contributions are as follows:

1. Our proposed model, without pre-training and target organ detection, is simple and efficient with high segmentation accuracy.
2. The MMNet model improves feature extraction through multi-scale feature fusion and effectively addresses the problem of low accuracy with respect to end-systolic segmentation of the cardiac system.
3. We propose a multi-scale feature fusion module using dilated convolution to extract multi-scale features from cardiac MRI images. It can effectively extract edge information and reconstruct the left ventricular border contour.
4. We reconstruct a full-size skip connection structure to make full use of the feature information in different layers, thereby enhancing the attentional learning of the left ventricular shape.

2 Related works

As the demand for more accurate left ventricular segmentation in clinical practice has increased, researchers have worked on several segmentation methods.

In the shape priori-based graph cut method, Mahapatra et al. [8] integrated a priori shape information into the graph cutting framework. For each dataset, the priori shape information is combined with the distance function of each pixel relative to the priori shape and the orientation angle histogram, and the final segmentation result is a combination of intensity and shape information. Due to poor edge information and the large shape variations between the different parts of each patient, the method then requires the inclusion of shape information for each dataset, which is inefficient. Auger et al. [9] used a guided point model approach to implement 3D cine DENSE cardiovascular magnetic resonance for the semi-automated segmentation of the left ventricle. An algorithm was presented to robustly propagate and model left ventricular epicardial and endocardial surfaces using displacement information encoded in the phase images of DENSE data. Based on the spatio-temporal continuity of the left ventricle, Wang et al. [10] used an iteratively-reduced threshold-based region growth method for the fully automated segmentation of the left ventricle. Generally, most of the traditional segmentation methods described above have poor robustness and are not fine-grained [11].

Among the machine learning-based methods, Luo et al. [12] combined the hierarchical extreme learning machine (H-ELM) with localization methods to achieve a more compact and meaningful feature representation for performing left ventricle segmentation. This method requires the pre-positioning of the left ventricle and is not highly convergent. Tsang et al. [13] used a new machine learning approach to outline the endocardial and epicardial boundaries of the left ventricle (LV) and to quantify these boundaries. In summary, most machine learning methods tend to be underfitted, they are computationally difficult, and the overall segmentation effect is mediocre [14].

In recent years, medical image processing methods based on deep learning have been favoured due to their

excellent feature representation capabilities. An FCN [15], the pioneer of convolutional neural networks in image segmentation, replaces the fully connected layer with a convolutional layer, thus preserving segmentation location information and making full use of deep feature mapping. However, there are still gaps in accuracy and stability with respect to the requirements of medical image segmentation. Jindong et al. [16] used a 3D FCN-based multi-path structure approach for the segmentation of multimodal brain tumour images. It effectively extracts feature information from multiple MRI images. In 2015, Ronneberger et al. [17] proposed a U-shaped network (UNet) structure applying encoders and decoders to segmentation of cells and the liver, fusing deep and superficial features by skip connections. UNet greatly improved upon previous segmentation methods and made a breakthrough in the field of medical image segmentation. Many subsequent researchers have also achieved improvements based on UNet for other organ and tissue segments. At the MICCAI 2017 Automated Cardiac Diagnosis Challenge (ACDC 2017), Isensee et al. [18] used an improved combination of 2D and 3D UNets to handle segmentation, achieving first place in the competition. Zotti et al. [19] proposed a grid-net CNN to extract global and local information, with global features distinguishing the heart from surrounding organs and local features ensuring accurate segmentation. Khened et al. [20] proposed a novel, highly parameter- and memory-efficient FCN-based architecture with upsampling paths by combining long skip and short-cut connections to overcome the feature graph explosion problem in FCN-like architectures. Yang et al. [21] proposed a dilated block adversarial network (DBAN), which contains a segmenter and a discriminator. An expansion block (EB) captures the multi-scale features of aggregated cardiac MRI images, and the discriminator guides the segmenter to modify the segmentation probability map. Cui et al. [22] used an attention mechanism structure with an input image pyramid and deep supervised output layers (AID), which was able to attend to the sizes and shapes of various heart structures. The attention mechanism emphasizes the desired features in the original image and suppresses irrelevant regions, effectively improving the accuracy of heart segmentation.

Although deep learning methods can learn features autonomously when segmenting the left ventricle without considering the physiological structure of the heart, they require a large amount of labelled data annotated by specialist physicians. Small datasets suffer from undertraining, which can lead to blurred edge segmentation of the left ventricle. In contrast, the resulting model is usually undertrained due to the limited training set, which results in blurred left ventricular edge segmentation. To address these shortcomings, the model proposed in this paper is able to obtain high segmentation accuracy on a limited data set, and the method is efficient, fault tolerant and robust.

3 Materials and methods

In this section, the data source used for LV segmentation is first introduced, and then the architecture of the convolutional neural network model used for this dataset is detailed. Finally, the parameter settings are described.

3.1 Data sources

We use the Automated Cardiac Diagnostic Challenge database from the MICCAI challenge in 2017 (ACDC 2017) [23], the database of the MICCAI 2009 LV segmentation challenge (MICCAI 2009) [24] and the database of the MICCAI 2018 left ventricle full quantification challenge (MICCAI 2018) [35].

The ACDC 2017 database [23] captured cardiac MRI images from a total of 150 subjects in both the end-diastolic and end-systolic phases using two different MRI scanners over a six-year period at the University Hospital of Dijon. The left ventricular short-axis slices have thickness intervals between approximately 5 mm and 8 mm and a spatial resolution of $1.37 - 1.68\ mm^2/pixel$. The database also includes additional information about the test subjects (ages, weights, heights, and diastolic-systolic phase instants). They were divided equally into five subgroups according to the cause of the disease: normal subjects; previous myocardial infarction (ejection fraction of the left ventricle lower than 40% and several myocardial segments with abnormal contraction); dilated cardiomyopathy (diastolic left ventricular volume $>$ 100 mL/m^2 and an ejection fraction of the left ventricle lower than 40%); hypertrophic cardiomyopathy (left ventricular cardiac mass higher than 110 g/m^2, several myocardial segments with thickness higher than 15 mm in diastole and a normal ejection fraction); and abnormal right ventricles (volume of the right ventricular cavity higher than 110 mL/m^2 or ejection fraction of the right ventricle lower than 40%).

This article uses the entire database from MICCAI challenge 2017 [23] with manual annotation labels as our first dataset (a total of 1922 slices from 100 subjects), and Fig. 1 shows the original images in the dataset. As described in the Introduction, the left ventricular cavity varies in shape and size at different stages of the cardiac cycle. There are inherent noise and artefacts in cardiac MRI images.

The database of MICCAI 2009 [24] was collected at Sunnybrook Health Sciences Centre, Toronto, Canada from 45 cardiac short-axis (SAX) datasets. There are four pathological subgroups: 12 ischaemic heart failures, 4 nonischaemic heart failures, 4 left ventricular hyperfullness

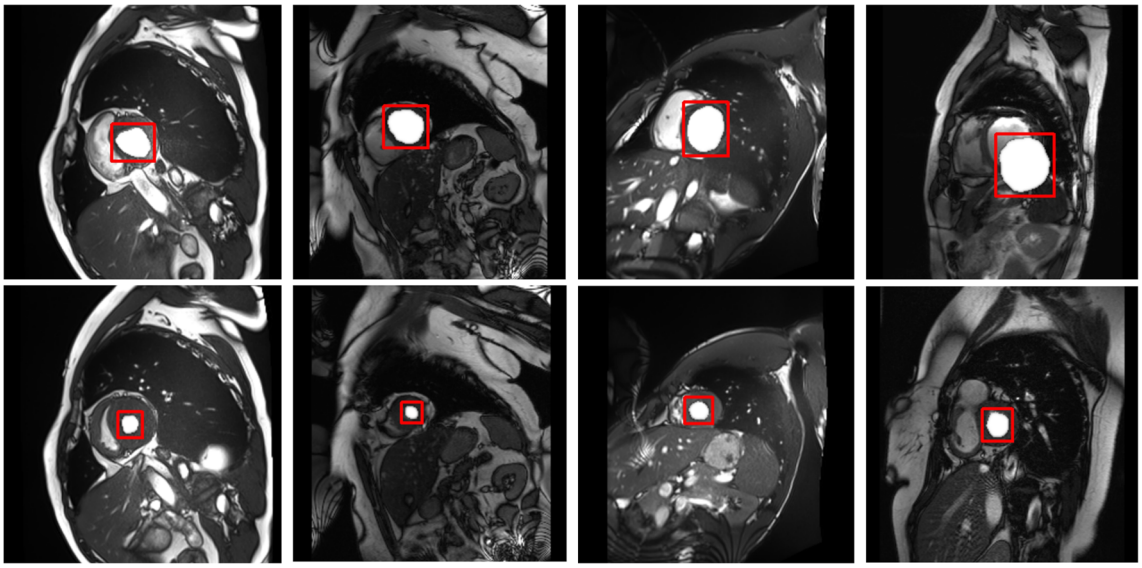

Fig. 1 Examples of cardiac MRI image slices from the ACDC 2017 dataset. The first row shows end-diastolic images of the hearts of the four test subjects, and the second row shows end-systolic images of the hearts. The manually-annotated left ventricular label is superimposed onto the original image, with the left ventricular cavity location selected by the red rectangular box

cases and 3 normal cases. We use 1084 cardiac MRI slices with manual annotation as our second dataset, and the endocardium of the left ventricle is used as the ground truth.

The database of MICCAI 2018 [35] utilized a training dataset of 145 clinical subjects with processed SAX MR sequences for model learning and validation. For each subject, the entire cardiac cycle consists of 20 frames and ground truth values for all LV indices were provided for each frame. We use 2900 cardiac MRI slices with manual annotation as our third dataset, and the endocardium of the left ventricle is used as the ground truth.

3.2 Network architecture

We use a multi-scale multi-skip connection net (MMNet) based on the 2D UNet to train with the MRI images. Figure 2 depicts the overall structure of the network, which consists of multi-scale inception blocks (MIBs), full-scale skip connections (FSCs), and a pre-activation module.

First, we extract multi-scale features from the input cardiac MRI images layer by layer using four MIBs to fully learn the LV position information and border contour information. Then, the features are compressed by dense

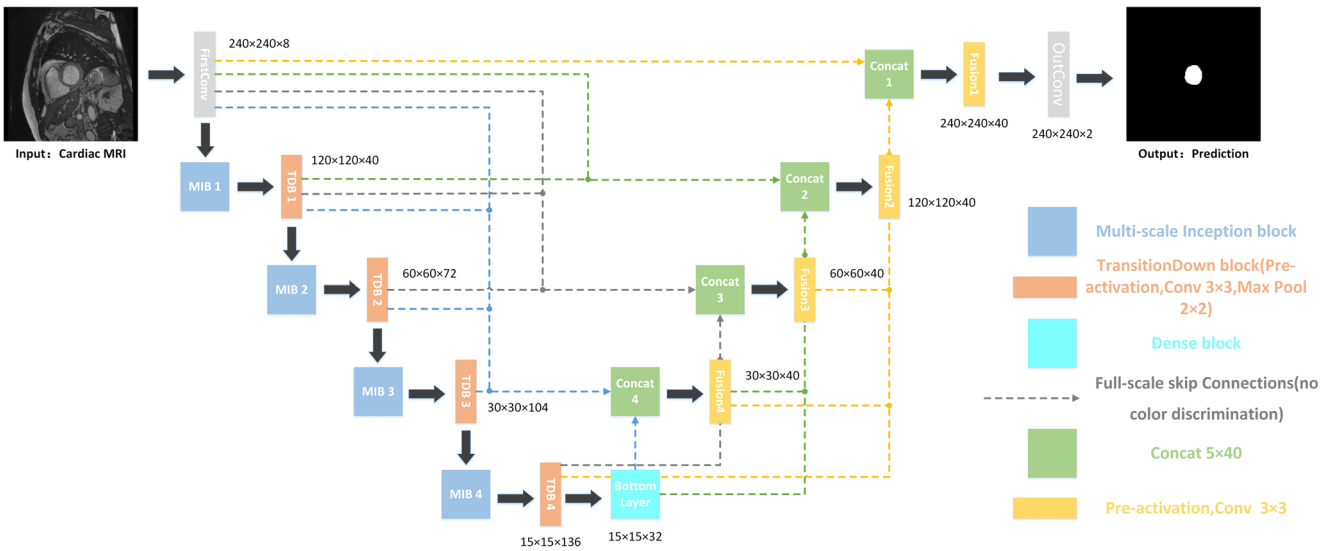

Fig. 2 Network architecture of MMNet. Batch normalization and pooling layers are omitted. The network outputs are two-channel predictions of the LV cavity and background

blocks to reduce redundancy. Finally, the LV contour is reconstructed by multi-layer feature fusion using full-scale skip connections. Pre-activation sub-modules are inserted in each module to propagate feature information backwards and forwards.

In the order of encoding and decoding, this subsection presents the expanded convolutional inception module for multi-scale feature extraction, the pre-activation module, the bottleneck layer, and the full-size skip connection structure for contextual information fusion.

3.2.1 Multi-scale inception block

As the sizes of the left ventricle and the proportions of the whole slice vary across detectors at different stages of the cardiac cycle (Fig. 1), we need to detect features of the target organ at different scales; therefore, multi-scale inference is commonly used in network models. Multi-scale feature fusion is performed at the encoding stage based on the range of different-sized receptive fields to obtain a high-resolution, semantically rich feature map.

In the traditional segmentation model, the convolutional layer performs feature extraction and then reduces the image size by pooling and downsampling to increase the receptive field, where the pooling size and the convolutional kernel size determine the size of the receptive field. If the model is shallow, it is difficult to model fuzzy boundaries. Deepening the model by stacking small convolutional kernels instead of large convolutional kernels [25] can be extremely cumbersome and computationally expensive, and it often results in overfitting problems. PSPNet [26] uses pooling operations of different sizes to control the perceptual field. However, the absence of learnable parameters in the pooling layer often results in the loss of internal data and the inability to reconstruct small pixel-level information. This shortcoming is serious for segmentation tasks that ultimately require the original-size prediction forecast map to be obtained. Inspired by the parallel multi-branch network structure of GoogLeNet [27], we propose a multi-scale inception block (MIB) that applies dilated convolution to perform multi-scale feature fusion during the encoding phase.

Dilated convolution allows for control of the field of perception by changing only the expansion rate while keeping the size of the convolution kernel constant and without increasing the number of parameters or computational redundancy [28]. The dilated convolution input is set as $F : Z^2 \rightarrow R$, make $\Omega_r = [-r, r]^2 \cap Z^2$, $f : \Omega_r \rightarrow R$ discrete filters are made, and we define the convolution operation with an expansion rate d as:

$$F *_d f(p) = \sum_{s+dt=p} F(s) f(t) \tag{1}$$

The size of the feature map output by the 2D dilated convolution operation is expressed as:

$$H_{out}/W_{out} = \left\lfloor \frac{H_{in}/W_{in} + 2 \times p - d \times (k-1) - 1}{s} + 1 \right\rfloor \tag{2}$$

H_{out}/W_{out} denote the height and length of the output feature map, respectively; H_{in}/W_{in} denotes the height and length of the input feature map, respectively; p indicates the padding; s indicates the step size of the convolution operation. Our inception structure uses four 3×3 convolution kernels with four expansion rates (1, 2, 4, 8). To ensure that the size of the output feature map remains the same after the convolution operation, the padding needs to be set to a value equal to the size of the expansion rate, at which point the perceptual field size of the whole convolution is

$$n = [k + (k-1) \times (d-1)]^2 . \tag{3}$$

Figure 3 shows the ranges of the receptive fields for three of the dilated convolutions. In Fig. 3, the inception block takes the input feature map and passes it through four parallel sets of dilated convolution channels to extract features at different scales. Afterwards, the four output feature maps are channel-connected, immediately followed by feature fusion using a set of 1×1 convolution kernels, where each is preceded by a pre-activation module. During the coding process, we use a combination of four such inception modules and a maximum pooling module as a four-sub-layer downsampling layer, with the inception modules only increasing the number of feature map channels and the maximum pooling module reducing the feature map size by half.

During the encoding process, the MIB implicitly learns the location information of the left ventricle using large expansivity convolution, while small expansivity convolution performs the extraction of left ventricular edge information. The amount of LV edge information extracted from the feature maps decreases as the resolution decreases, and multi-scale feature fusion allows the edge information to be preserved. This is important for solving the problems of blurred boundaries and edge information overlapping with background pixels. At the same time, the MIB can learn multi-scale features with fewer parameters, which is well suited to small datasets such as ACDC 2017 [23] and MICCAI 2009 [24].

3.2.2 Pre-activation module

In deep networks, the distribution of data at each layer varies, which can make it difficult for the network training process to converge. Batch normalization [29] is widely used before the activation function (ReLU) normalizes the

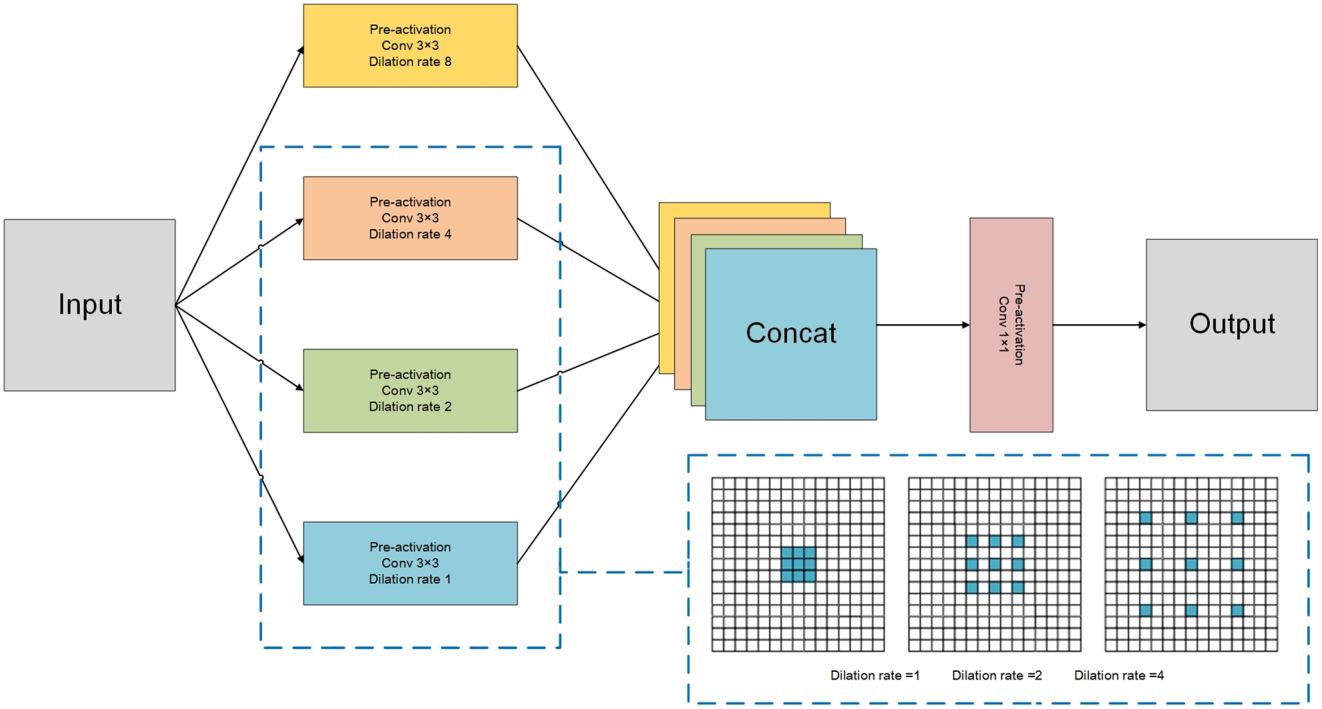

Fig. 3 The Multiscale Inception Module (MIB) uses four 3 × 3 kernel expansion convolutions with expansion rates of 1, 2, 4, and 8 (perceptual fields of 3 × 3, 5 × 5, 9 × 9, and 17 × 17 respectively), three of which are selected by the blue dashed boxes

input to the activation function, addressing the effects of shifting and increasing data. Kaiming et al. [30] rearranged the order of the weight layers and activation functions (e.g., ReLU and/or BN) in the residual cells and demonstrated that the pre-activation model can effectively mitigate the overfitting phenomenon.

The use of the BN+ReLU pre-activation module before each convolution operation in our inception structure ensures that all the heavy layers are normalized. Compared to the results of post-activation, the multiscale features of the left ventricle extracted by parallel dilated convolution operations using the pre-activation module are preserved. The output feature maps of all channels are then reused with the pre-activation module. With BN reducing feature complexity and the ReLU function increasing the non-linearity between the four convolutional layers, inter-parameter dependencies are reduced. For average pooling, the first activation may filter out negative values from the original convolutional output, resulting in some loss of information. However, we use the maximum pooling module in all our networks, and the addition of a pre-activation module before each maximum pooling module does not result in the loss of feature information. In this mode, the overall network training convergence speed is improved, and the feature information can be propagated not only forward but also backwards. The whole network effectively controls gradient explosion, gradient disappearance and overfitting.

3.2.3 Bottleneck layer

The bottleneck layer contains fewer nodes than the last layer of the downsampling stage, allowing for a reduced dimensional representation of the features while ensuring that the size of the feature map remains unchanged. Inspired by DenseNets [31], we use a 4-layer dense block as the bottleneck layer, where each layer in the dense block uses a dropout operation with a dropout rate of 0.5 to prevent overfitting of the network. The number of channels in the feature map is reduced from 136 to 32 after this bottleneck layer, thereby reducing the number of network parameters and helping to combine the features rationally to eliminate redundancy. Ultimately, a compressed feature representation of the network is produced.

3.2.4 Full-scale skip connections

UNet's fine segmentation of medical images benefits from a skip connection between the encoder and decoder. It fuses the fine-grained features of the encoder with the coarse-grained semantic features of the decoder, allowing the optimizer to handle simpler learning tasks when the feature mapping semantics of both are similar and thus improving segmentation efficiency. DenseNet [31] uses short skip connections in the encoding phase, with stable gradients and blocky convergence speed during training but with a large memory footprint. UNet++ [32] designs a dense skip

connection that uses features from all layers of the network to automatically learn the importance levels of features at different depths, reducing the semantic gap between the encoder and decoder. However, UNet++ does not take full advantage of the multi-scale features of the different layers.

In many segmentation studies, different levels of feature maps carry different semantic information. Low-level large-scale feature maps focus on spatial information and highlight target organ boundaries; high-level semantic feature maps locate target organ locations. During the process of downsampling, to increase the robustness of the input image to perturbation and to perform upsampling to restore the image size, the fine-grained details of the left ventricular boundary may be diluted. Therefore, we design a full-scale skip connection (FSC) to improve segmentation accuracy and speed up model training.

In Fig. 4, as an example of a full-scale skip connection in the second-level sampling phase, X_{De}^{2} uses the feature maps in each layer for feature fusion with different semantics. The large feature map X_{En}^{1} in the first layer of the network contains low-level semantic information and needs to be reduced by twice the feature map size with the maximum pooling module to ensure that the channels have the same resolution when they are connected. Whereas the encoded feature maps from the parallel layer in the original UNet are directly connected to the decoded feature map channels, our network uses a pre-activation module and a convolution module before forming the connection. As the encoded feature map X_{En}^{2} is from the second layer of the network in this case, the semantic gap between the encoder and decoder is reduced. Below the parallel layer, skip connections use decoded feature maps. As shown in the figure, X_{De}^{3}, X_{De}^{4} and X_{De}^{5} are expanded twice, four times, and eight times, respectively, using deconvolution respectively to ensure the same resolution. The five sets of feature maps (40 maps each) are connected channel-wise (200 maps in total) and placed in the pre-activation module. X_{De}^{2} is then obtained by convolution operations using 40 3 × 3 convolution kernels. This approach attempts to perform feature fusion, unify the number of channels, and reduce redundant information.

Extending to all layers, i denotes the ith layer in the network along the downsampling direction, and the ith layer decoder (four layers in total) is calculated as follows:

$$x_{De}^{i} = G\left[\underbrace{D\left(X_{En}^{1}\right),\cdots,D\left(X_{En}^{i-1}\right)}_{scales:i-1}, C\left(X_{En}^{i}\right), \underbrace{U\left(X_{De}^{i+1}\right),\cdots,U\left(X_{De}^{5}\right)}_{scales:(i+1)^{th}\sim 5^{th}}\right], i = 1,\cdots,4 \quad (4)$$

where the function $G(\cdot)$ denotes the feature fusion mechanism, which consists of a pre-activation operation and a convolution calculation using 40 3 × 3 convolution

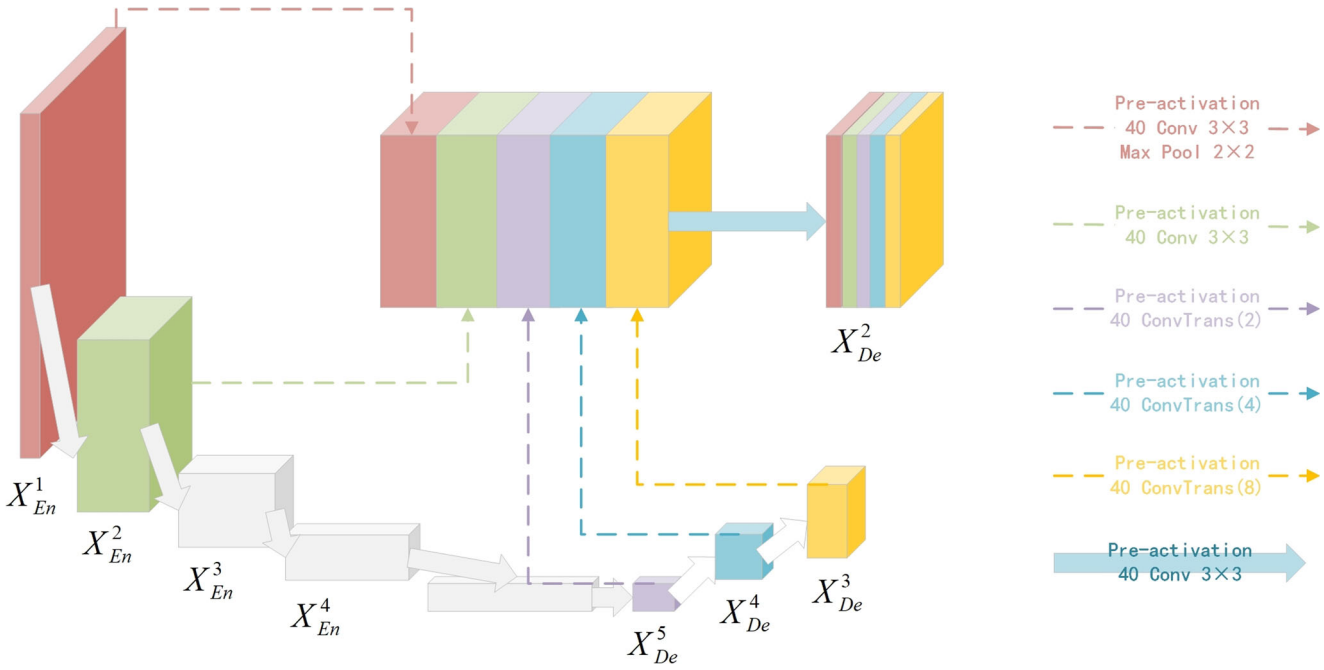

Fig. 4 This figure illustrates the construction of a full-scale aggregated feature map for the 2nd layer decoder X_{De}^{2}. To make full use of the feature information in each layer, both TDB4 and BottomLayer are available as X_{De}^{5} in the 5th layer, and this layer uses BottomLayer as the skip connection object in this example

kernels; [·] indicates a channel connection; the function $D(\cdot)$ indicates a downsampling operation, reducing of the feature map size by utilizing a maximum pooling module (pre-activation module + 40 3 × 3 convolution kernels + maximum pooling layer); the function $C(\cdot)$ represents the convolution operation of a parallel layer (preactivation module + 40 3 × 3 convolution kernels); the function $U(\cdot)$ represents the upsampling operation (pre-activation module + 40 deconvolution cores)used to increase the size of the feature map.

During the segmentation map reconstruction process, the feature information on the LV shape is not all concentrated in the last layer of the feature map, and it may be distributed across different scales of the feature map. By means of gradient propagation, full-scale skip connections can combine information from all layers and learn more about the LV shape features to improve segmentation accuracy.

3.3 Parameter settings

Due to the large gaps (usually 8 mm) between sections along the Z-axis of the input heart MRI image and the large differences between the features of the left ventricle in different sections, the 3D convolution operation is not applicable. Therefore, our network handles each 2D slice individually. For the input data, all 2D sections are centre-clipped to make the patch size 240 × 240, which ensures the category balance between the left ventricle and the background pixels of the sections.

The ground truth of the original ACDC 2017 [23] dataset contains four types of labels, namely, background, right ventricle, myocardium, and left ventricle. We change the label files to make both the right ventricle and myocardium part of the background. The model performs the task of pixel dichotomy. During the process of model training, we use the dichotomous loss function, which is similar to the cross-entropy function:

$$\begin{aligned} loss\,(x, class) &= -log\left(\frac{exp\,(x\,[class])}{\Sigma_j exp\,(x\,[j])}\right) \\ &= -x\,[class] + log\left(\sum_j exp\,(x\,[j])\right) \end{aligned} \tag{5}$$

x represents the output data, and $class$ represents the label data. The formula consists of two steps: a log softmax function and negative log-likelihood loss computation.

4 Experiments

4.1 Network training settings

We used the PyTorch deep learning network framework as our model training framework, applying Kaiming initialization to initialize the weights of the convolutional layers. We also used the Adam [33] optimizer. In this method, the initial learning rate of 0.001 was decayed by 0.95 per epoch, and the weight decay (L2 regularization) was set to $1 \cdot e^{-4}$. The experiments were conducted on an NVIDIA RTX 2080 Ti GPU.

In the first dataset (ACDC 2017 [23]), we divided the cardiac MRI images of 100 testers equally into ten groups, each with an equal number of cases from five cardiac diagnoses. Eight of them were randomly selected as the training set, one was selected as the validation set, and one was selected as the test set. The training set had 90 patients, the validation set had 10 patients, and the test set had 10 patients. For the second dataset (MICCAI 2009 [24]), we randomly selected 805 slices as the training set and 279 slices as the test set. For the third dataset (MICCAI 2018 [35]), we randomly selected 2030 slices as the training set and 870 slices as the test set. The reported results in the paper were all measured on the test set.

The three cardiac MRI image datasets used in the paper were small relative to the segmentation task of natural image processing, which may have resulted in undertraining. We used data enhancement to address the limitations of the small dataset. The original images were subjected to random processing, including vertical flip, mirror flip, rotation, and Gaussian noise. This increased the amount of data available for training and compensated for the small dataset.

We trained and tested the two datasets separately. The batch size was set to 32. We obtained the segmentation results for the left ventricles of all patients in the two test sets after 300 training epochs.

4.2 Evaluation indicators

The most commonly used metric for evaluating the effectiveness of medical image segmentation method is the similarity index called the Dice coefficient (DC), which is a measure of the overlap between the foreground pixels and the ground truth foreground pixel region of the segmented image. It is calculated as follows:

$$Dice\,(R, G) = 2 \times \frac{R \cap G}{R + G} = 2 \times \frac{TP}{TP + FP + TP + FN} \tag{6}$$

R indicates the real predicted results, and G indicates the ground truth. When applied to Boolean data, true positive (TP), false positive (FP), and false negative (FN) can be used to calculate the DC. In this paper, $Dice_{ED}$ denotes the Dice coefficient of the left ventricle at end-diastolic(ED), $Dice_{ES}$ denotes the Dice coefficient of the left ventricle at end-systolic(ES), and $Dice_{Total}$ denotes the Dice coefficient of the left ventricle as a whole.

Another metric, the Jaccard index, indicates the degree of dissimilarity between the foreground pixels and the ground truth foreground pixel region of the segmented image. It is calculated as follows:

$$Jaccard\,(R, G) = \frac{R \cap G}{R + G - R \cap G} = \frac{TP}{TP + FP + FN} \tag{7}$$

R indicates the real predicted results, and G indicates the ground truth. When applied to Boolean data, true positive (TP), false positive (FP), and false negative (FN) were used to calculate the Jaccard index. In this paper, $Jaccard_{ED}$ denotes the Jaccard index of the left ventricle at end-diastolic(ED), $Jaccard_{ES}$ denotes the Jaccard index of the left ventricle at end-systolic(ES), and $Jaccard_{Total}$ denotes the Jaccard index of the left ventricle as a whole.

The Hausdorff distance represents the maximum value of the shortest distance from a point in a point set to another point set, and the two point sets in the segmentation task are the foreground pixels and the ground truth foreground pixel region. This index is a measure of shape similarity and is calculated as follows:

$$\begin{aligned} Hausdorff(X, Y) = max(&sup_{x \in X} inf_{y \in Y} d\,(x, y)\,, \\ &sup_{y \in Y} inf_{x \in X} d(x, y)) \end{aligned} \tag{8}$$

X indicates the real predicted results, and Y indicates the ground truth. sup and inf denote the upper and lower definite boundaries, respectively. In this paper, Hd denotes the Hausdorff distance of the left ventricle as a whole, with units in mm.

We used the Dice coefficient and Jaccard index to measure left ventricular segmentation in both the end-diastolic and end-systolic phases separately for all patients in the test set.

4.3 Experimental results

4.3.1 Training loss comparison

Figures 5 and 6 show the variation of the training loss functions among MMNet, FCN, and the base UNet model over the first 30 rounds of the two dataset experiments. Under the same training environment, MMNet can reach the threshold loss fastest, and the cross-entropy loss can converge to the smallest value (the training losses are similar for both datasets; that of the FCN is 0.0062, that of UNet is 0.0023, and that of MMNet is 0.0018). As a result, our model learns more efficiently on the training set and has a better ability to learn more adequate feature information.

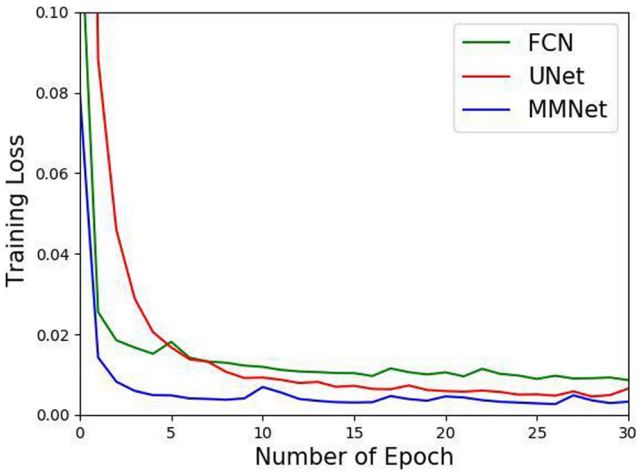

Fig. 5 ACDC 2017 training loss

4.3.2 Model evaluation on the test dataset

We tested on the set of ACDC 2017 dataset [23], and the segmentation results are shown in Fig. 8. In this figure, we compared the prediction maps of MMNet with those of the UNet model. Regarding the assessment of the Dice coefficients for the five pathological subgroups, results of over 90% were achieved for both end-diastolic (ED) and end-systolic (ES) hearts, with mean Dice coefficients of 96.2% for ED and 93.9% for ES. The Jaccard indices at end-diastole (ED) and end-systole (ES) were 0.926 and 0.868, respectively, and the overall mean Hausdorff distance was 7.0. We also tested on the MICCAI 2009 dataset [24], and the segmentation results are shown in Fig. 9. The mean Dice coefficient for the four pathological subgroups was 98.0%, the mean Jaccard index for the four pathological subgroups was 96.4%, and the mean Hausdorff distance for the four pathological subgroups was 5.2. On the MICCAI 2018 dataset [35], the tested Dice factor was 96.8%, the

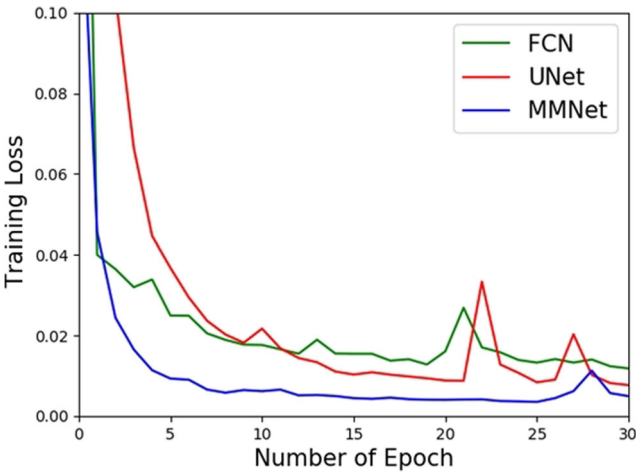

Fig. 6 MICCAI 2009 training loss

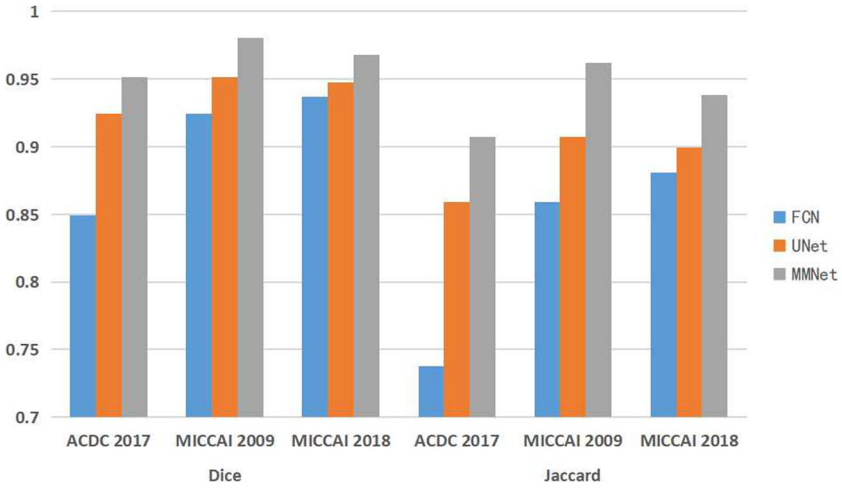

Fig. 7 This image shows the Dice coefficient and Jaccard coefficient statistics for FCN, UNet and MMNet on the three datasets

Jaccard index was 93.7% and the Hausdorff distance was 7.5. The results show that our model has high accuracy and produces stable segmentation results.

As shown in Fig. 7, we have also calculated the test results for FCN, UNet and MMNet on three datasets. The dice coefficients for MMNet tested on all three datasets exceeded 0.95 and were higher than the other two models across the board. In the ACDC 2017 dataset, which has the most comprehensive set of cases and the highest segmentation difficulty, the improvement of MMNet over the other models is particularly clear. The comparison of the Jaccard coefficients is more striking than that of the dice coefficients.

As seen in Fig. 8, two of the UNet model prediction maps showed isolated false positive organs; UNet judged other bright regions in the abdominal cavity to be left ventricular cavities, whereas neither of the MMNet maps showed isolated segmentation errors. For the HCM test subjects, there were distinct areas of varying brightness in the left ventricular cavity due to interference from the papillary muscles at the end of systole. The segmentation results showed that the shape of the ventricular rim matched the ground truth and that there were no isolated segmentation errors within the cavity. The NOR detector segmentation results had low contrast and blurred information on the endocardial and epicardial edges. There was a shape difference between the UNet segmentation output and the ground truth; however, our segmentation method resulted in smoother edges. Noise and motion artefacts were present in the RV detector slices, and there were significant differences between end-diastolic and end-systolic cardiac ventricular contours in the illustrations, which MMNet still allowed for accurate segmentation. The segmentation effect map demonstrates that the method successfully distinguished between ventricular and background regions, even in complex intraventricular chambers.

As shown in Fig. 9, the segmentation of the left ventricle in the dataset was easier, and both MMNet and UNet were able to achieve excellent segmentation results. Benefiting from the MIB block, the edge contour segmentation effect of MMNet was better than that of UNet, and the edge of the left ventricle was smoother. UNet's segmentation map

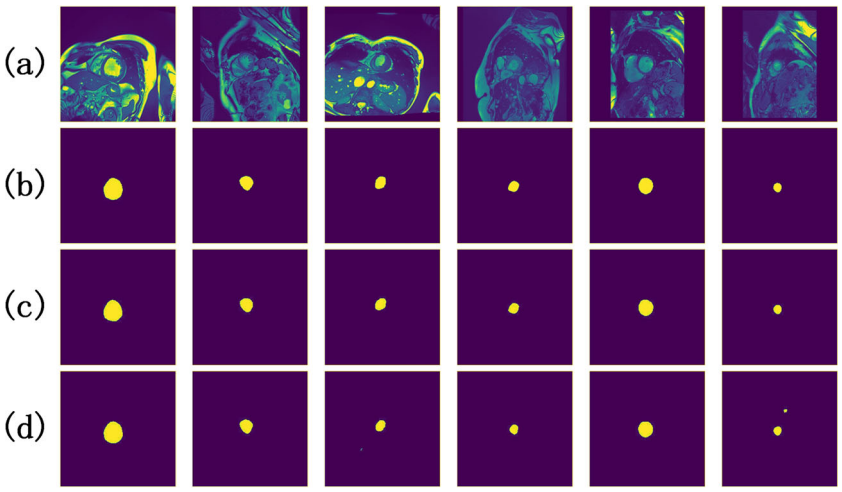

Fig. 8 This figure shows the results of LV segmentation for different patients in the ACDC 2017 dataset. The pathological subgroups from left to right are DCM, HCM, MINF, NOR, RV, and RV. (a) is the original cardiac MRI, (b) is the ground truth (GT), (c) is the prediction map obtained using our model and (d) is the prediction map obtained with the UNet model

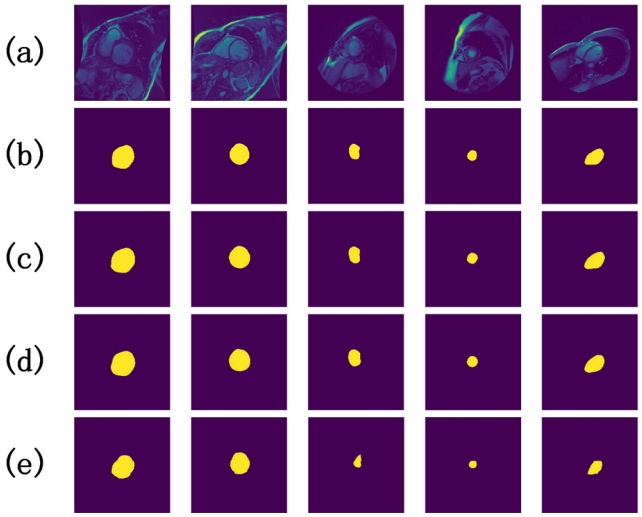

Fig. 9 This figure shows the results of LV segmentation for different patients in the MICCAI 2009 dataset. (a) is the original cardiac MRI, (b) is the ground truth (GT), (c) is the prediction map obtained using our model, (d) is the prediction map obtained with the UNet model and (e) is the prediction map produced by the FCN model

differed somewhat from the ground truth in shape, while MMNet fit the shape of the ground truth more closely. The FCN results were poorly segmented and differed most from the ground truth shape.

4.3.3 Comparison with other approaches on the ACDC 2017 dataset

We also compared our results with those of other state-of-the-art models, including those of the FCN [15], UNet [17], and the online ACDC 2017 leader board [23]. The results are shown in Table 1. MMNet significantly outperformed all other methods in terms of the total Dice coefficient on the test dataset, and the segmentation effect was particularly good for the more difficult segmentation of the end-systole of the heart. The first ranked method, that of Isensee et al. which uses a fusion of a 2D UNet and a 3D UNet, had a slightly higher end-diastolic Dice coefficient than that of our method, but MMNet uses only a single CNN and achieved the highest total Dice coefficient; therefore, our method ranked first.

The 2D and 3D UNet fusion method used by Isensee et al. [18] had the best segmentation performance with respect to the more easily segmented end-diastolic phase of the heart, where pre-processing and post-processing are too complex. The corresponding resource consumption was relatively high, and the efficiency was reduced. MMNet only changes the structure of the model and does not alter the slices too much in terms of pre-processing. The model parameters generated through iterative training can be directly used to automatically segment cardiac MRI slices, thereby achieving higher segmentation accuracy with higher efficiency. Cui et al. [22] introduced an attention mechanism that enhances or weakens the value of each predicted pixel based on its similarity to other pixels in the input image and setting a weight value for each pixel. The attention mechanism relatively increases the number of learning parameters and may not be applicable to every network. The overall Dice coefficient of the left ventricular division was 0.939. UNet++ [32] ensures that the left ventricular features of the encoder and decoder are fused and matched. The Dice coefficient at end-systole was improved, with an overall coefficient of 0.938. However, UNet++ does not take full advantage of the multiscale features at each layer and could be pruned to speed up network convergence. Painchaud et al. [34] used a post-processing method to convert an invalid heart shape to something close to the correct shape. However, this causes blurring in the edge segmentation results, with an overall Dice coefficient of 0.936. Recently developed methods still have difficulty accurately segmenting the heart at the end of systole, and our proposed model is able to segment this phase with an accuracy that is 2.5% higher than that of the average model.

Table 1 LV test results on the ACDC 2017 dataset

Methods	$Dice_{ED}$	$Dice_{ES}$	$Dice_{Total}$	$Jaccard_{ED}$	$Jaccard_{ES}$	$Jaccard_{Total}$	Hd
FCN [15]	0.903	0.795	0.849	0.823	0.660	0.738	12.9
UNet [17]	0.947	0.901	0.924	0.901	0.820	0.859	9.8
Isensee et al. [18]	**0.965**	0.933	0.949	–	–	–	7.1
Zotti et al. [19]	0.964	0.912	0.938	–	–	–	7.3
Khened et al. [20]	0.962	0.911	0.937	–	–	–	8.6
DBAN [21]	0.960	0.905	0.933	–	–	–	7.4
Cui et al. [22]	0.959	0.918	0.939	–	–	–	7.2
UNet++ [32]	0.955	0.921	0.938	0.914	0.854	0.883	8.3
Painchaud et al. [34]	0.961	0.911	0.936	–	–	–	7.2
MMNet	0.962	**0.939**	**0.951**	**0.926**	**0.885**	**0.907**	**7.0**

Table 2 Model versions for the ablation experiment

Methods	Descriptions
UNet	The original four-layer UNet
UNet+MIB	UNet with Multi-scale Inception Block
UNet+FSC	UNet with Full-scale Skip Connections
MMNet+PA	MMNet with Post-activation
MMNet	The model proposed in this paper

4.3.4 Ablation experiments

To verify the effectiveness of the modules in our approach, we designed several versions of our network, as shown in Table 2. This section compares the impact of each module on the overall results and analyses the role of each module.

In Fig. 10, the myocardial wall thickness shown in the slice is thin and uneven, and the edge contour features of the left ventricle were difficult to extract. With the use of a multi-scale inception block (MIB), the model was able to extract the edge information effectively, and the reconstructed LV contours were significantly better than those of the UNet model without the MIB. In Fig. 11, the left ventricular cavity shown in the slice is irregularly shaped, with the shape of the left ventricle varying very little from one slice to another. This requires a much finer fit with the contours and makes it more difficult to learn the correct

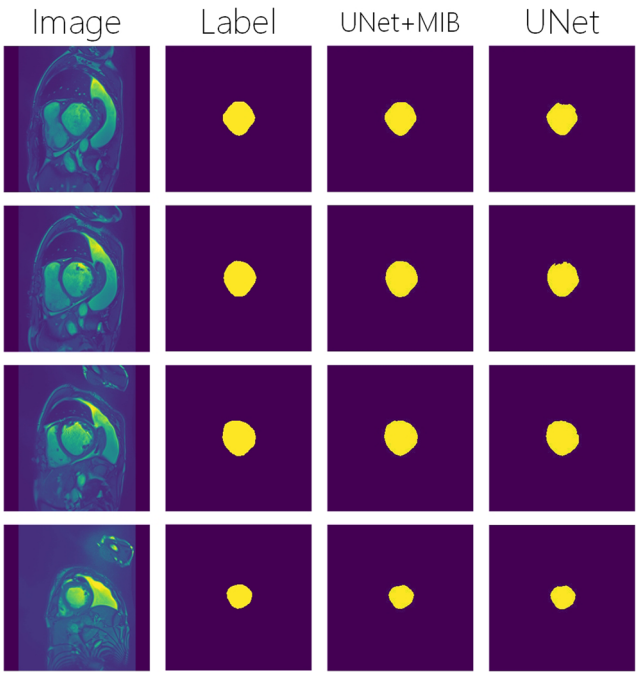

Fig. 10 This figure shows the effect of the MIB on enhancing the model's left ventricular edge segmentation results. From left to right, the images are the original cardiac MRI image, the ground truth, the UNet+MIB model segmentation result, and the UNet segmentation result

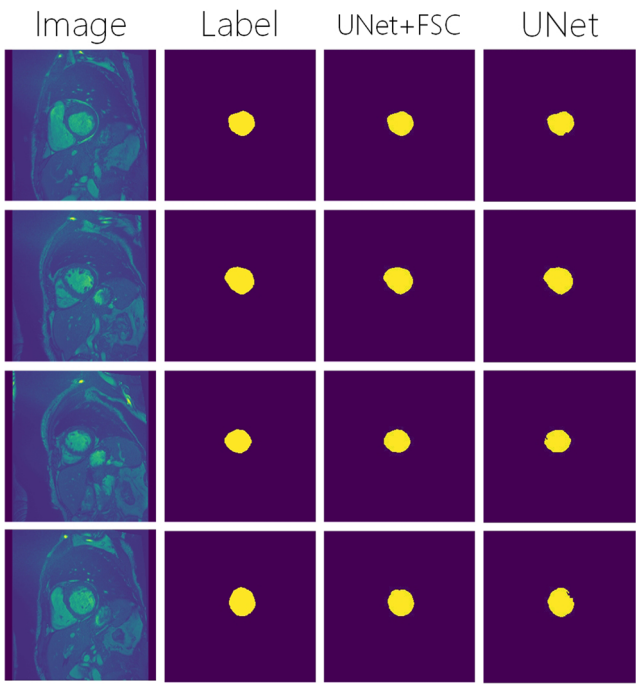

Fig. 11 This figure shows the effect of the FSC-enhanced model in terms of fitting the shape of the left ventricle. From left to right, the images are the original cardiac MRI image, the ground truth, the UNet+FSC model segmentation result, and the UNet segmentation result

shape. With the use of a full-scale skip connection (FSC), on the one hand, the model was able to incorporate more shape information; on the other hand, the shape fit with the ground truth was improved. The reconstructed LV edges are smoother than the UNet model segmentation results.

Figure 12 provided statistics on the lift of the MIB and FSC blocks compared to UNet on different data sets. Statistics shown that the multi-scale inception block (MIB) and the full-scale skip connection (FSC) are effective in improving the segmentation ability of the model on three different datasets. Overall FSC had a slightly higher improvement in dice and jaccard coefficients than MIB. Both modules performed particularly well at ACDC 2017 and MICCAI 2009.

As shown in Tables 3 and 4, the multi-scale inception block (MIB) improved the total Dice coefficient of the original UNet by 2.0% and that of the full-scale skip connection (FSC) structure by 2.1%. On both the ACDC 2017 and MICCAI 2009 datasets, the blocks were able to reduce the Hausdorff distance by 2.0 mm to 2.5 mm. As shown in Table 5, on the MICCAI 2018 dataset, the segmentation of the left ventricle was easier. However, each block segmentation effect was improved relative to that of UNet, with an overall improvement of approximately 0.3 in terms of the Dice coefficient and a 0.7 reduction in the Hausdorff distance. The experimental results demonstrate

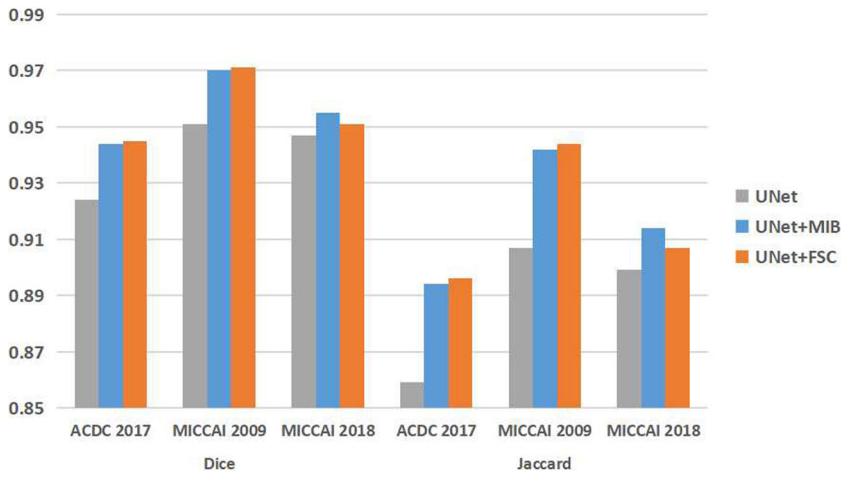

Fig. 12 This image shows the Dice coefficient and Jaccard coefficient statistics for UNet, UNet+MIB and UNet+FSC on the three datasets

Table 3 Results of ablation experiments on the ACDC 2017 dataset

Methods	$Dice_{ED}$	$Dice_{ES}$	$Dice_{Total}$	$Jaccard_{Total}$	Hd
UNet	0.947	0.901	0.924	0.859	9.8
UNet+MIB	0.958	0.930	0.944	0.894	7.4
UNet+FSC	0.957	0.932	0.945	0.896	7.8
MMNet+PA	0.960	0.934	0.947	0.899	7.1
MMNet	0.962	0.939	0.951	0.907	7.0

Table 4 Results of ablation experiments on the MICCAI 2009 dataset

Methods	$Dice_{Total}$	$Jaccard_{Total}$	Hd
FCN [15]	0.924	0.859	8.0
UNet [17]	0.951	0.907	7.8
UNet+MIB	0.970	0.942	5.5
UNet+FSC	0.971	0.944	5.3
MMNet+PA	0.975	0.951	5.1
MMNet	0.980	0.962	5.2

Table 5 Results of ablation experiments with on MICCAI 2018 dataset

Methods	$Dice_{Total}$	$Jaccard_{Total}$	Hd
FCN [15]	0.937	0.881	11.1
UNet [17]	0.947	0.899	8.2
UNet+MIB	0.955	0.914	8.0
UNet+FSC	0.951	0.907	8.3
MMNet+PA	0.965	0.932	7.8
MMNet	0.968	0.938	7.5

that both our proposed FSC and MIB improve upon the segmentation effect of the UNet model. The inception module of the downsampling stage fused the multi-scale features of the obtained feature map to extract a wider range of left ventricle features and improve the segmentation accuracy. The full-scale skip connection maximized the use of the full-scale feature map and improved the segmentation efficiency. The pre-activation module ensured the stability of the segmentation effect. The use of the pre-activation module instead of a post-activation module improved the overall segmentation effect of our model by 0.3%, and this also validates the effectiveness of the pre-activation module.

5 Conclusion

In this paper, we presented a new deep learning model for left ventricular segmentation. Thanks to the multi-scale feature extraction ability of the inception module and the utilized full-scale skip connections, our MMNet model was able to fully learn the left ventricular features of patients with four common cardiovascular diseases and normal cases and automatically segment the left ventricular chambers in cardiac MRI images based on the learned feature information. Even with a limited training data set, the Dice coefficient for end-diastolic LV segmentation on the ACDC 2017 test set reached 96.2% and 93.9% for the more difficult end-systolic segmentation, the Dice coefficient on the MICCAI 2009 test set reached 98.0%, and the Dice coefficient on the MICCAI 2018 test set reached 96.8%. The overall Jaccard indices on the three datasets reached 0.897, 0.964 and 0.937. The overall Hausdorff distances between the three datasets reached 7.0 mm, 5.2 mm and 7.5 mm. We verified through ablation experiments that both of our proposed modules yielded improved segmentation accuracy. The method also outperformed current state-of-the-art approaches in comparative experiments. Thus, MMNet provides an efficient and accurate solution for assisting physicians in the diagnosis of heart disease through cardiac MRI imaging.

Author Contributions Ziyue Wang proposed the method and conducted the experiments, analysed the data and wrote the manuscript. Yanjun Peng supervised the project and participated in manuscript revisions. Dapeng Li and Yanfei Guo provided critical reviews that helped improve the manuscript.

Funding This work was supported in part by the National Natural Science Foundation of China [Grant No. 61976126], Shandong Nature Science Foundation of China [Grant No. ZR2017MF054, ZR2019MF003, ZR2020MF044].

Availability of data and materials Data related to the current study are available from the corresponding author on reasonable request.

Code Availability Some of the codes generated or used during the study are is available from the corresponding author by request.

Declarations

Conflict of Interests The authors declare that they have no conflicts of interest.

References

1. Salerno M, Sharif B, Arheden H, Kumar A, Axel L, Li D, Neubauer S (2017) Recent advances in cardiovascular magnetic resonance: techniques and applications. Circulation: Cardiovascular Imaging 10(6):e003951
2. Xue W, Brahm G, Pandey S, Leung S, Li S (2018) Full left ventricle quantification via deep multitask relationships learning. Medical Image Analysis 43:54–65
3. Bernard O, Lalande A, Zotti C, Cervenansky F, Yang X, Heng PA, Jodoin PM (2018) Deep learning techniques for automatic MRI cardiac multi-structures segmentation and diagnosis: is the problem solved? IEEE Transactions On Medical Imaging 37(11):2514–2525
4. Duan J, Bello G, Schlemper J, Bai W, Dawes TJ, Biffi C, Rueckert D (2019) Automatic 3D bi-ventricular segmentation of cardiac images by a shape-refined multi-task deep learning approach. IEEE Transactions on Medical Imaging 38(9):2151–2164
5. Budai A, Suhai FI, Csorba K, Toth A, Szabo L, Vago H, Merkely B (2020) Fully automatic segmentation of right and left ventricle on short-axis cardiac MRI images. Comput Med Imaging Graph 85:101786
6. Leclerc S, Smistad E, ØStvik A, Cervenansky F, Espinosa F, Espeland T, Bernard O (2020) LU-Net: A Multistage Attention Network to Improve the Robustness of Segmentation of Left Ventricular Structures in 2-D Echocardiography. IEEE Transactions on Ultrasonics, Ferroelectrics, and Frequency Control 67(12):2519–2530
7. Penso M, Moccia S, Scafuri S, Muscogiuri G, Pontone G, Pepi M, Caiani EG (2021) Automated left and right ventricular chamber segmentation in cardiac magnetic resonance images using dense fully convolutional neural network. Comput Methods Prog Biomed 204:106059
8. Mahapatra D (2013) Cardiac image segmentation from cine cardiac MRI using graph cuts and shape priors. Journal of Digital Imaging 26(4):721–730
9. Auger DA, Zhong X, Epstein FH, Meintjes EM, Spottiswoode BS (2014) Semi-automated left ventricular segmentation based on a guide point model approach for 3D cine DENSE cardiovascular magnetic resonance. J Cardiovasc Magn Reson 16:1–12
10. Wang L, Pei M, Codella NC, Kochar M, Weinsaft JW, Li J, Wang Y (2015) Left ventricle: fully automated segmentation based on spatiotemporal continuity and myocardium information in cine cardiac magnetic resonance imaging (LV-FAST). BioMed Research International
11. Liu X, Zhu X, Li M, Wang L, Zhu E, Liu T, Gao W (2019) Multiple kernel k k-means with incomplete kernels. IEEE Transactions on Pattern Analysis and Machine Intelligence 42(5):1191–1204
12. Luo Y, Yang B, Xu L, Hao L, Liu J, Yao Y, van de Vosse F (2018) Segmentation of the left ventricle in cardiac MRI using a hierarchical extreme learning machine model. International Journal of Machine Learning and Cybernetics 9(10):1741–1751
13. Tsang W, Wang B, Ghaziani ZN, Sun J, Chan R, Rakowski H (2020) Machine learning for left ventricular segmentation and scar quantification in hypertrophic cardiomyopathy patients. Can J Cardiol 36(10):S81–S82

14. Yu X, Ye X, Gao Q (2020) Infrared handprint image restoration algorithm based on apoptotic mechanism. IEEE Access 8:47334–47343
15. Long J, Shelhamer E, Darrell T (2015) Fully convolutional networks for semantic segmentation. In: Proceedings of the IEEE conference on computer vision and pattern recognition, pp 3431–3440
16. Sun J, Peng Y, Guo Y, Li D (2021) Segmentation of the multimodal brain tumor image used the multi-pathway architecture method based on 3D FCN. Neurocomputing 423:34–45
17. Ronneberger O, Fischer P, Brox T (2015) U-net: Convolutional networks for biomedical image segmentation. In: International conference on medical image computing and computer-assisted intervention. Springer, Cham, pp 234–241
18. Isensee F, Jaeger PF, Full PM, Wolf I, Engelhardt S, Maier-Hein KH (2017) Automatic cardiac disease assessment on cine-MRI via time-series segmentation and domain specific features. In: International workshop on statistical atlases and computational models of the heart. Springer, Cham, pp 120–129
19. Zotti C, Luo Z, Lalande A, Jodoin PM (2018) Convolutional neural network with shape prior applied to cardiac MRI segmentation. IEEE Journal of Biomedical and Health Informatics 23(3):1119–1128
20. Khened M, Kollerathu VA, Krishnamurthi G (2019) Fully convolutional multi-scale residual DenseNets for cardiac segmentation and automated cardiac diagnosis using ensemble of classifiers. Medical Image Analysis 51:21–45
21. Yang X, Zhang Y, Lo B, Wu D, Liao H, Zhang Y (2020), DBAN: Adversarial Network with Multi-Scale Features for Cardiac MRI Segmentation. IEEE Journal of Biomedical and Health Informatics
22. Cui H, Yuwen C, Jiang L, Xia Y, Zhang Y (2021) Multiscale attention guided U-Net architecture for cardiac segmentation in short-axis MRI images. Comput Methods Prog Biomed 106142
23. Bernard O, Lalande A, Zotti C, Cervenansky F, Yang X, Heng PA, Sanroma G (2018) Deep learning techniques for automatic MRI cardiac multi-structures segmentation and diagnosis: is the problem solved? IEEE Transactions on Medical Imaging 37(11):2514–2525
24. Radau P, Lu Y, Connelly K, Paul G, Dick AJWG, Wright G (2009) Evaluation framework for algorithms segmenting short axis cardiac MRI. The MIDAS Journal-Cardiac MR Left Ventricle Segmentation Challenge 49
25. Simonyan K, Zisserman A (2014) Very deep convolutional networks for large-scale image recognition. arXiv:1409.1556
26. Zhao H, Shi J, Qi X, Wang X, Jia J (2017) Pyramid scene parsing network. In: Proceedings of the IEEE conference on computer vision and pattern recognition, pp 2881–2890
27. Szegedy C, Liu W, Jia Y, Sermanet P, Reed S, Anguelov D, Rabinovich A (2015) Going deeper with convolutions. In: Proceedings of the IEEE conference on computer vision and pattern recognition, pp 1–9
28. Yu F, Koltun V (2015) Multi-scale context aggregation by dilated convolutions. arXiv:1511.07122
29. Ioffe S, Szegedy C (2015) Batch normalization: Accelerating deep network training by reducing internal covariate shift. arXiv:1502.03167
30. He K, Zhang X, Ren S, Sun J (2016) Identity mappings in deep residual networks. In: European conference on computer vision. Springer, Cham, pp 630–645
31. Jégou S, Drozdzal M, Vazquez D, Romero A, Bengio Y (2017) The one hundred layers tiramisu: Fully convolutional densenets for semantic segmentation. In: Proceedings of the IEEE conference on computer vision and pattern recognition workshops, pp 11–19
32. Zhou Z, Siddiquee MMR, Tajbakhsh N, Liang J (2018) Unet++: A nested u-net architecture for medical image segmentation. In: Deep learning in medical image analysis and multimodal learning for clinical decision support. Springer, Cham, pp 3–11
33. Kingma DP, Ba J (2014) Adam: A method for stochastic optimization. arXiv:1412.6980
34. Painchaud N, Skandarani Y, Judge T, Bernard O, Lalande A, Jodoin PM (2020) Cardiac segmentation with strong anatomical guarantees. IEEE Trans Med Imaging 39(11):3703–3713
35. Pop M, Sermesant M, Zhao J, Li S, McLeod K, Young A, ..., Mansi T. (eds) (2019) Statistical Atlases and Computational Models of the Heart. Atrial Segmentation and LV Quantification Challenges, vol 11395. Springer, Berlin

Ziyue Wang received the bachelor's degree in Shandong Jianzhu University institute of science, Jinan, China, in 2019. He is currently studying for MSc at Shandong University of Science and Technology. His research interests include Computer vision and image processing.

Yanjun Peng received Dr. degree in march, 2004. He joined the Department of Computer Science, Shandong University of science and technology, Qingdao, China, in 1996, where he was promoted to professor in 2010. His main research interests include medicine visualization, virtual reality, and image processing.

Dapeng Li received bachelor degree and master degree from Shandong Normal University, Jinan, China, in 2011 and 2014, respectively. He joined the department of information engineering, Shandong Normal University of Lishan college in 2014. He is currently studying for a Ph.D. at Shandong University of Science and Technology. His main research interests include deep learning and medical image process.

Yanfei Guo received the bachelor's degree in Shandong University of Science and Technology of Taishan institute of technology, Taian, China, in 2015, and the master's degree in software engineering from the Beijing Key Laboratory of Information Service Engineering, Beijing Union University, Beijing, China. She is currently studying for Ph.D. at Shandong University of Science and Technology. Her research interests include Computer vision and image processing.