

Research Article

Multiresolution Mutual Assistance Network for Cardiac Magnetic Resonance Images Segmentation

Shaolong Chen , Changzhen Qiu , Weiping Yang , and Zhiyong Zhang 

School of Electronics and Communication Engineering, Sun Yat-sen University, Shenzhen 518107, China

Correspondence should be addressed to Zhiyong Zhang; zhangsysu2020@163.com

Received 23 June 2022; Accepted 21 October 2022; Published 31 October 2022

Academic Editor: V. E. Sathishkumar

Copyright © 2022 Shaolong Chen et al. This is an open access article distributed under the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

The automatic segmentation of cardiac magnetic resonance (MR) images is the basis for the diagnosis of cardiac-related diseases. However, the segmentation of cardiac MR images is a challenging task due to the inhomogeneity of MR images intensity distribution and the unclear boundaries between adjacent tissues. In this paper, we propose a novel multiresolution mutual assistance network (MMA-Net) for cardiac MR images segmentation. It is mainly composed of multibranch input module, multiresolution mutual assistance module, and multilabel deep supervision. First, the multibranch input module helps the network to extract local and global features more pertinently. Then, the multiresolution mutual assistance module implements multiresolution feature interaction and progressively improves semantic features to more completely express the information of the tissue. Finally, the multilabel deep supervision is proposed to generate the final segmentation map. We compare with state-of-the-art medical image segmentation methods on the medical image computing and computer-assisted intervention (MICCAI) automated cardiac diagnosis challenge datasets and the MICCAI atrial segmentation challenge datasets. The mean dice scores of our method in the left atrium, right ventricle, myocardium, and left ventricle are 0.919, 0.920, 0.881, and 0.960, respectively. The analysis of evaluation indicators and segmentation results shows that our method achieves the best performance in cardiac magnetic resonance images segmentation.

1. Introduction

Cardiovascular disease is one of the world's leading causes of death, and it kills more people each year from cardiovascular disease than from any other disease [1, 2]. In recent years, the number of patients with cardiovascular disease has increased sharply. The prevention and treatment of cardiovascular disease should attract public attention. With the development of modern medicine, in order to reduce the mortality rate and misdiagnosis rate of cardiovascular diseases, medical imaging technologies such as magnetic resonance imaging (MRI), computerized tomography (CT), and ultrasound (US) are widely used in the diagnosis and treatment of cardiovascular diseases. Cardiac MRI is currently recognized as the gold standard for evaluating the cardiac function, and MRI has the advantages of less harm to the human body and clear imaging [3–6]. The automatic segmentation of cardiac magnetic resonance (MR) images is the

basis for the diagnosis of cardiac-related diseases. In general, the anatomy of the cardiac MR image includes the left ventricle, right ventricle, epicardium, endocardium, and myocardium. At present, the main segmentation method in clinical use is manual segmentation by doctors, which can obtain accurate results but is very time-consuming. The limitations of manual segmentation have motivated researchers to continue developing automatic segmentation methods for cardiac segmentation [7].

The current cardiac MR images segmentation methods can be mainly divided into traditional methods [8–10] and deep learning-based methods [11–15]. Traditional methods mainly include graph searching based on intensity [8], region growing [9], and active appearance models [10]. However, most of these traditional methods have problems such as complex design, poor versatility, and low segmentation accuracy. In recent years, deep learning has achieved great success in medical image processing [16–19]. Some

researchers introduced deep learning into medical image segmentation. The deep learning-based method gradually replaces the traditional medical image segmentation method due to its good versatility, high segmentation accuracy, and high efficiency [20–23]. The proposal of UNet [24] is a milestone in medical image segmentation. Based on the U-shaped structure and skip connections, UNet fuses low-resolution information and high-resolution information and has been widely used for cardiac MR images segmentation. Li et al. [25] proposed a new multiscale feature attentive UNet for cardiac MR images segmentation and achieved excellent performance. Sharan et al. [26] combined feature pyramid network and UNet architecture to study the automatic segmentation of left ventricle, myocardium, and right ventricle. Sharan et al. [27] proposed a stack attention-based convolutional neural network approach for fully automatic segmentation from short-axis cardiac MR images. Cui et al. [28] added the direction field module, channel self-attention module, and selective kernel module to the UNet framework to improve the segmentation performance, and the segmentation experiments on cardiac MR images demonstrated the effectiveness of the improvements. Wang et al. [29] proposed an auto-weighted supervision framework to solve the problem of scar and edema segmentation in multisequence cardiac MR images although the existing cardiac MR images segmentation methods have achieved good results. However, the segmentation of cardiac MR images is still a challenging task due to the inhomogeneity of MR images intensity distribution and the unclear boundaries between adjacent tissues.

Recently, Fu et al. [30] proposed that a multiscale input layer constructs an image pyramid to achieve multiple level receptive field sizes for optic disc and optic cup segmentation. It is proved that multiscale input can improve the segmentation performance. Shi et al. [31] proposed a multiinput fusion network model based on multiscale input and feature fusion, which automatically extracts and fuses the features of different input scales to realize the detection of cardiac MR images. Chen et al. [32] proposed a T-based multiresolution input network, which achieved good performance in the field of medical image segmentation. Currently, the application of multiresolution input in medical image segmentation is less studied. There is still a lot of room for improvement in the existing methods. Firstly, the existing multiresolution input network only considers the fusion of multiresolution features at the encoder side but does not consider the fusion of multiresolution features at the decoder side. Second, the shallow features extracted from high-resolution images contain a lot of irrelevant background information, and existing methods do not consider how to suppress this irrelevant background information by utilizing deep features extracted from low-resolution images.

In this paper, we propose a novel multiresolution mutual assistance network (MMA-Net) for cardiac MR images segmentation. It is mainly composed of multibranch input module, multiresolution mutual assistance module, and multilabel deep supervision. First, the multibranch input

module is responsible for feature extraction of input images with different resolutions. Each resolution input image has a separate feature extraction branch. The high-resolution input image branch is responsible for learning the local information of the image without worrying about the loss of global information because the extraction of global information is completed by the low-resolution input image branch. Similarly, the low-resolution input image branch is responsible for learning the global information of the image without worrying about the loss of local information. Second, the multiresolution mutual assistance module implements multiresolution feature interaction and progressively improves semantic features to more completely express the information of the tissue. Finally, the multilabel deep supervision is proposed to generate the final segmentation map. In addition, we designed the attention gate that utilizes global features extracted from low-resolution input images to suppress irrelevant background information from local features extracted from high-resolution input images. We compared with state-of-the-art medical image segmentation methods on the medical image computing and computer-assisted intervention (MICCAI) automated cardiac diagnosis challenge datasets (ACDC) [33] and the MICCAI atrial segmentation challenge datasets (ASC) [34]. The mean dice score of our method in the left atrium, right ventricle, myocardium, and left ventricle are 0.919, 0.920, 0.881, and 0.960, respectively. The analysis of evaluation indicators and segmentation results shows that our method achieves the best performance in cardiac magnetic resonance images segmentation.

The main contribution of this work can be summarized as follows:

- (1) A novel multiresolution mutual assistance network (MMA-Net) for cardiac MR images segmentation is proposed. It implements multiresolution feature interaction and progressively improves semantic features to more completely express the information of the tissue.
- (2) We designed the attention gate that utilizes global features extracted from low-resolution input images to suppress irrelevant background information from local features extracted from high-resolution input images.
- (3) A multilabel deep supervision is proposed, which can well handle the problem of inconsistent prediction results and labels caused by up sampling of small-scale feature layers in deep supervision.
- (4) Our method outperforms the existing six excellent medical image segmentation methods.

2. Method

The proposed multiresolution mutual assistance network (MMA-Net) is shown in Figure 1. It is mainly composed of multibranch input module, multiresolution mutual assistance module, and multilabel deep supervision. As shown in Figure 1, first, 2D medical images with resolutions of

224×224 and 112×112 are input to the multibranch input module to extract features, respectively. Second, these extracted features are then input to the multiresolution mutual assistance module for information interaction and progressively improves semantic features to more completely express the information of the tissue. Finally, the multilabel deep supervision to guide the learning of the network and the prediction result of $M_{D,1}$ are used as the final result.

2.1. Multibranch Input Module. The multiresolution input has been shown to be effective in improving segmentation quality [30]. The current multiresolution input mostly adopts the structure of the shared encoder. The disadvantage of this structure is that it is difficult to balance the learning of local features and global features. If the receptive field of the convolution kernel in the convolutional layer is increased, the learning of global features can be enhanced, but some local features will be lost at the same time, and vice versa. Therefore, we adopted a multibranch structure with a separate encoder for each resolution input. The high-resolution input branch can learn the local information of the image without worrying about the loss of global information because the extraction of global information is done by the low-resolution input branch. Similarly, the low-resolution input branch is responsible for learning the global

information of the image without worrying about the loss of local information. For the selection of the number of branches, after our experiments, we chose the dual-branch structure, as shown in Figure 1. For branch 1, its input is an image with a resolution of 224×224 , and the output is the feature $M_{E,i}$ ($i = 1, 2, 3, 4, 5$) of each encoding stage. For branch 2, its input is an image with a resolution of 112×112 , and the output is the feature $N_{E,j}$ ($j = 2, 3, 4, 5, 6$) of each encoding stage.

2.2. Multiresolution Mutual Assistance Module. After obtaining the features at each stage of the input image at different resolutions, our goal is to use this information to obtain better decoded features for final segmentation prediction. The input to multiresolution mutual assistance module is the features of each stage of the branch 1 and branch 2 encoders, including $M_{E,i}$ ($i = 1, 2, 3, 4, 5$) and $N_{E,j}$ ($j = 2, 3, 4, 5, 6$). The outputs are the features of each stage of the branch 1 and branch 2 decoders, including $M_{D,i}$ ($i = 1, 2, 3, 4$) and $N_{D,j}$ ($j = 2, 3, 4$). For each branch, the input to each decoder stage consists of the complementary features generated by the previous stage and the features of the corresponding encoder stage. The output of each stage of the decoder is calculated by the following formulas.

For branch 1,

$$\begin{cases} M_{D,i} = \mathbf{A}(\mathbf{Up}(\mathbf{F}(M_{D,i+1}, N_{D,i+1})), M_{E,i}), i = 1, 2, 3, \\ M_{D,4} = \mathbf{A}(\mathbf{Up}(\mathbf{F}(\mathbf{Up}(N_{E,6}), \mathbf{AG}(\mathbf{Up}(N_{E,6}), M_{E,5}))), M_{E,4}). \end{cases} \quad (1)$$

For branch 2,

$$\begin{cases} N_{D,j} = \mathbf{A}(\mathbf{Up}(\mathbf{F}(M_{D,j+1}, N_{D,j+1})), N_{E,j}), j = 2, 3, \\ N_{D,4} = \mathbf{A}(\mathbf{Up}(\mathbf{F}(\mathbf{Up}(N_{E,6}), \mathbf{AG}(\mathbf{Up}(N_{E,6}), M_{E,5}))), N_{E,4}). \end{cases} \quad (2)$$

Here, **Up** is the up sampling; **AG** is the attention gate; **A** is the attention feature selection; and **F** is the feature fusion.

2.2.1. Attention Gate. In our network, branch 1 is mainly used to extract shallow local features, and branch 2 is mainly used to extract deep global features. Local features contain a large amount of detailed information of the target tissue, but they also introduce a lot of irrelevant background information. Global features contain information such as the location of the target tissue, and there is less detailed information, but there is also little irrelevant background information. Inspired by reference [35], we designed an attention gate that utilizes global features of the last stage

($N_{E,6}$) of branch 2 to suppress the irrelevant background information of the local features of the last stage ($M_{E,5}$) of branch 1. The structure of attention gate is shown in Figure 2.

2.2.2. Attention Feature Selection. For each branch, the input to each decoder stage consists of the complementary features generated by the previous stage and the features of the corresponding encoder stage. The feature input from the encoder stage has shallower features than the corresponding complementary features. Therefore, we also designed to use complementary features to suppress the irrelevant background information of the corresponding encoder stage input features, and the attention feature selection is shown in Figure 3.

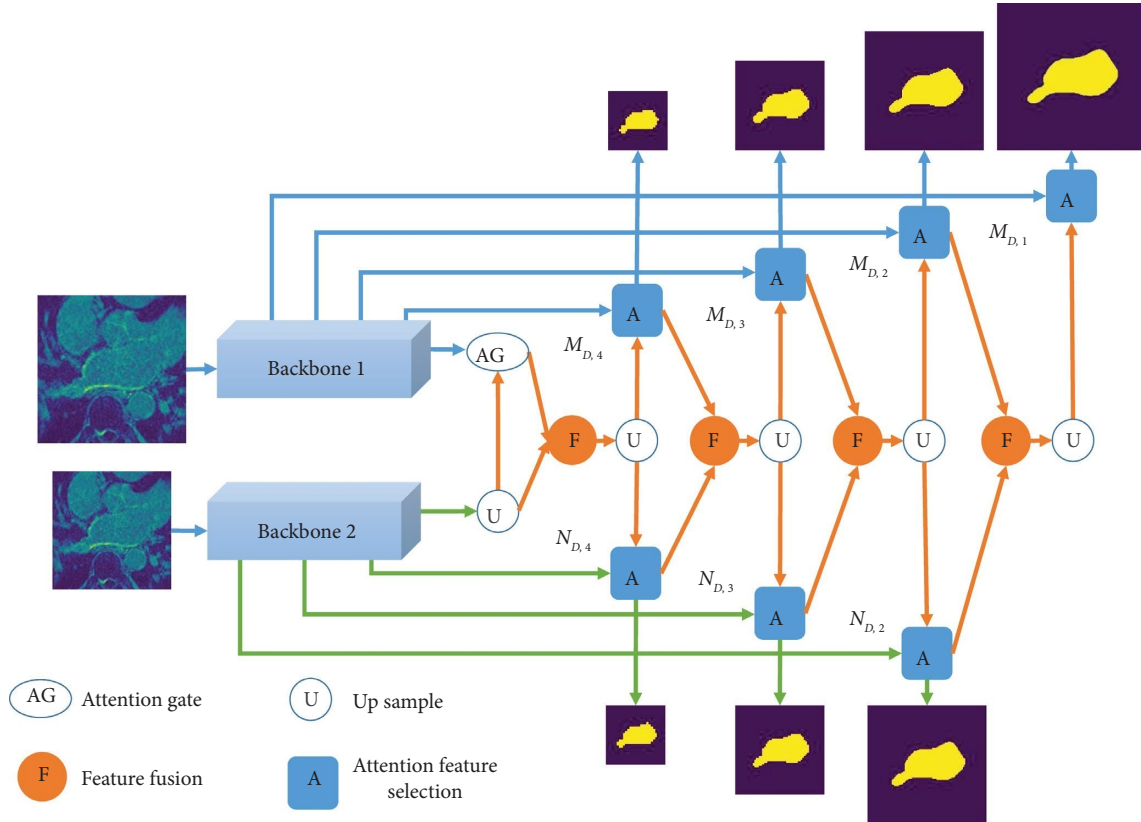


FIGURE 1: The proposed multiresolution mutual assistance network (MMA-Net).

2.2.3. Feature Fusion. It first concatenates multiple input features along the channel axis and then applies two 3×3 convolutional layers to the fusion result with the same number of output channels as a single input.

2.3. Multilabel Deep Supervision. In deep supervision, there are only labels of the same size as the original image. The prediction result of the last layer is the same as the scale of the label, and the loss can be calculated directly with the label. The prediction results of other small-scale feature layers are usually up sampling to the original image size, and then the loss is calculated with the labels. However, during the up sampling process, the prediction results become coarse, which may lead to inconsistencies between the prediction results and the labels. To solve this problem, we propose a multilabel deep supervision. Figure 4 shows the deep supervision and multilabel deep supervision of $M_{D,1}$ and $M_{D,4}$ layers. As shown in Figure 4(a), in the deep supervision, the consistency between $M_{D,1}$ results and labels is good, but the consistency between up sampling and labels is poor in $M_{D,4}$ results, which may cause the network to learn wrong information. As shown in Figure 4(b), in the multilabel deep supervision, each scale feature layer has a label that is consistent with its feature map size. The results are consistent with the label, which can well guide the network learning.

We have seven output prediction maps and the total loss function is a simple addition of the loss functions of these

seven output prediction maps. For each output prediction map, we considered the combination of binary cross-entropy and dice loss as

$$L = L_{BCE} + L_{DICE}. \quad (3)$$

Here,

$$L_{BCE} = \sum_{x,y} G_{x,y} * \log(Q_{x,y}) + (1 - G_{x,y}) * \log(1 - Q_{x,y}),$$

$$L_{DICE} = 1 - \frac{2 \sum_{x,y} G_{x,y} * Q_{x,y}}{\sum_{x,y} (G_{x,y} + Q_{x,y})}, \quad (4)$$

where L_{BCE} and L_{DICE} represent the binary cross-entropy loss and dice loss, respectively. $G_{x,y} \in \{0, 1\}$ is the area label at position (x, y) , and $Q_{x,y} \in [0, 1]$ is the area value at position (x, y) in output prediction.

3. Experiments

3.1. Datasets, Preprocessing, Implementation Details, and Evaluation Metrics

3.1.1. Datasets and Preprocessing. We evaluated our method at the medical image computing and computer-assisted intervention (MICCAI) automated cardiac diagnosis challenge (ACDC) [33] and the MICCAI atrial

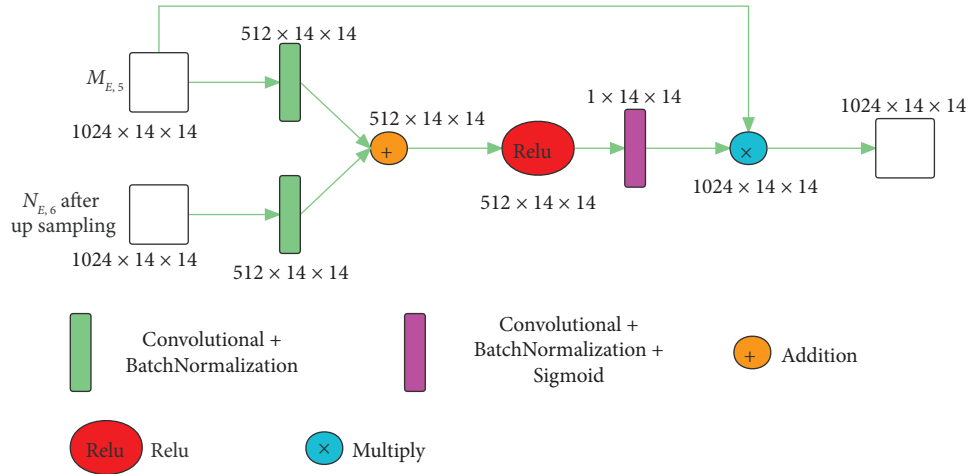


FIGURE 2: Attention gate.

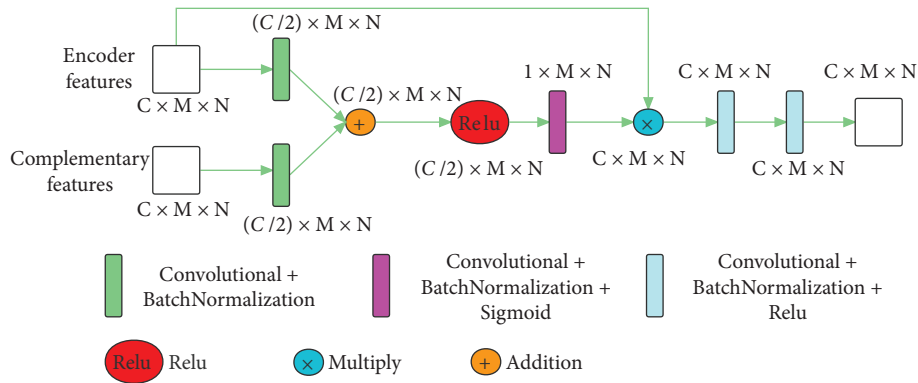


FIGURE 3: Attention feature selection.

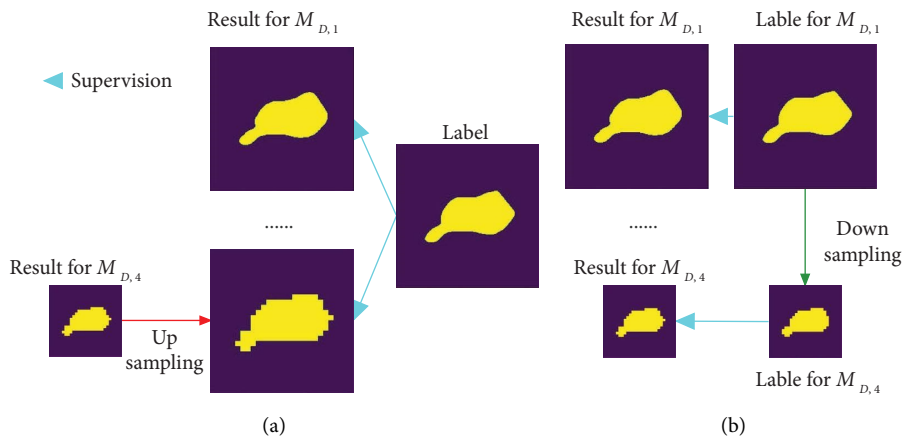


FIGURE 4: The deep supervision and multilabel deep supervision of the $M_{D,1}$ and $M_{D,4}$ layers. (a) Deep supervision. (b) Multilabel deep supervision.

segmentation challenge (ASC) [34]. The ACDC has cardiac MR images for 150 cases, but only 100 cases are annotated (right ventricle, myocardium, and left ventricle), and we only use the 100 annotated cases. The ASC has 154 cardiac MR images, all annotated. For all datasets, we additionally cropped input images at their centers to

make their size 224×224 pixels. In addition, we performed max-min normalization (0–1) for each case. For both the ACDC and ASC datasets, we augmented the training set by slightly translating, scaling, and rotating. We evaluated the performance of each model by 5-fold cross-validation.

TABLE 1: The influence of the number of branches in the network on the segmentation accuracy in the ACDC.

	Num = 1	Num = 2	Num = 3	Num = 4
Dice	0.897	0.920	0.911	0.908
Specificity	0.999	0.999	0.999	0.999
Sensitivity	0.891	0.915	0.902	0.900
F1	0.899	0.920	0.911	0.908

The best performance is shown in bold.

TABLE 2: The influence of the number of branches in the network on the segmentation accuracy in the ASC.

	Num = 1	Num = 2	Num = 3	Num = 4
Dice	0.909	0.919	0.914	0.912
Specificity	0.995	0.996	0.996	0.995
Sensitivity	0.911	0.917	0.914	0.913
F1	0.911	0.920	0.915	0.913

The best performance is shown in bold.

3.1.2. Implementation Details. Each model runs on four RTX 3090 cards. We trained our network in the multilabel deep supervision way. All models are trained with the Adam optimizer with batch size 32, learning rate 5×10^{-4} , momentum 0.9, weight decay 1×10^{-4} , and max-epoch 1000. The early stopping is set to 20. For each branch, we use VGG19 as the backbone network to extract features.

3.1.3. Evaluation Metrics. We measured the accuracy of segmentation by the dice similarity coefficient (dice), specificity, sensitivity, and F1-score (F1) by

$$\begin{aligned}
 \text{dice} &= \frac{2 * (A \cap B)}{A \cup B}, \\
 \text{specificity} &= \frac{TN}{TN + FP}, \\
 \text{sensitivity} &= \frac{TP}{TP + FN}, \\
 \text{F1} &= \frac{2 * TP}{2 * TP + FP + FN},
 \end{aligned} \tag{5}$$

where A and B represent prediction result and ground truth, respectively. TP, TN, FP, and FN represent the number of true positives, true negatives, false positives, and false negatives, respectively.

3.2. Ablation Experiments and Analyses

3.2.1. Number of Branches. We analyzed the influence of the number of branches in the network on the segmentation accuracy in the ACDC and ASC, which includes (a) $Num = 1$ (the input image size is 224×224), (b) $Num = 2$ (the input image sizes are 224×224 and 112×112 , respectively), (c) $Num = 3$ (the input image sizes are 224×224 , 112×112 and 56×56 , respectively), and (d) $Num = 4$ (the input image sizes are 224×224 , 112×112 ,

TABLE 3: The influence of the multiresolution mutual assistance module in the network on the segmentation accuracy in the ACDC.

	UFM	TFM	MMAM
Dice	0.908	0.909	0.920
Specificity	0.999	0.999	0.999
Sensitivity	0.900	0.899	0.915
F1	0.908	0.911	0.920

The best performance is shown in bold.

TABLE 4: The influence of the multiresolution mutual assistance module in the network on the segmentation accuracy in the ASC.

	UFM	TFM	MMAM
Dice	0.911	0.910	0.919
Specificity	0.995	0.995	0.996
Sensitivity	0.913	0.912	0.917
F1	0.911	0.910	0.920

The best performance is shown in bold.

TABLE 5: The influence of the multilabel deep supervision in the network on the segmentation accuracy in the ACDC.

	DS	MLDS
Dice	0.915	0.920
Specificity	0.999	0.999
Sensitivity	0.908	0.915
F1	0.915	0.920

The best performance is shown in bold.

TABLE 6: The influence of the multilabel deep supervision in the network on the segmentation accuracy in the ASC.

	DS	MLDS
Dice	0.916	0.919
Specificity	0.996	0.996
Sensitivity	0.915	0.917
F1	0.917	0.920

The best performance is shown in bold.

56×56 and 28×28 , respectively). The results are shown in Tables 1 and 2. According to our network structure rules, when $Num = 5$, branch 5 will have no decoder stage; therefore, we do not compare the case of $Num \geq 5$. As shown in Tables 1 and 2, when $Num = 2$, the segmentation performance of the network is the best, so we finally chose the dual-branch network structure.

3.2.2. Multiresolution Mutual Assistance Module. We analyzed the influence of the multiresolution mutual assistance module in the network on the segmentation accuracy in the ACDC and ASC, which includes (a) unidirectional fusion mode (UFM), (b) two-way fusion mode (TFM), and (c) multiresolution mutual assistance module (MMAM). The results are shown in Tables 3 and 4. As shown in Tables 3 and 4, compared with other modes, our multiresolution mutual assistance module achieves better performance.

TABLE 7: Comparison with state-of-the-art methods on the ACDC.

		UNet [24]	Att-UNet [35]	UNet++ [36]	UNet3+ [37]	TransUNet [38]	Swin-UNet [39]	Ours
Dice	RV	0.892	0.895	0.900	0.905	0.901	0.841	0.920
	Myo	0.861	0.851	0.858	0.854	0.845	0.775	0.881
	LV	0.939	0.940	0.936	0.947	0.942	0.909	0.960
Specificity	RV	0.998	0.999	0.999	0.999	0.999	0.997	0.999
	Myo	0.999	0.998	0.999	0.999	0.998	0.997	0.998
	LV	0.999	0.999	0.999	0.999	0.999	0.999	0.999
Sensitivity	RV	0.891	0.885	0.894	0.900	0.899	0.855	0.909
	Myo	0.852	0.849	0.840	0.843	0.868	0.800	0.883
	LV	0.929	0.928	0.931	0.939	0.950	0.894	0.953
F1	RV	0.894	0.898	0.902	0.907	0.903	0.845	0.921
	Myo	0.863	0.854	0.859	0.856	0.847	0.777	0.881
	LV	0.941	0.942	0.939	0.949	0.945	0.910	0.960

The best performance is shown in bold.

TABLE 8: Comparison with state-of-the-art methods on the ASC.

		UNet [24]	Att-UNet [35]	UNet++ [36]	UNet3+ [37]	TransUNet [38]	Swin-UNet [39]	Ours
Dice	LA	0.909	0.908	0.907	0.907	0.904	0.876	0.919
Specificity	LA	0.995	0.995	0.996	0.995	0.995	0.994	0.996
Sensitivity	LA	0.911	0.914	0.905	0.911	0.902	0.887	0.917
F1	LA	0.911	0.911	0.911	0.911	0.908	0.886	0.920

The best performance is shown in bold.

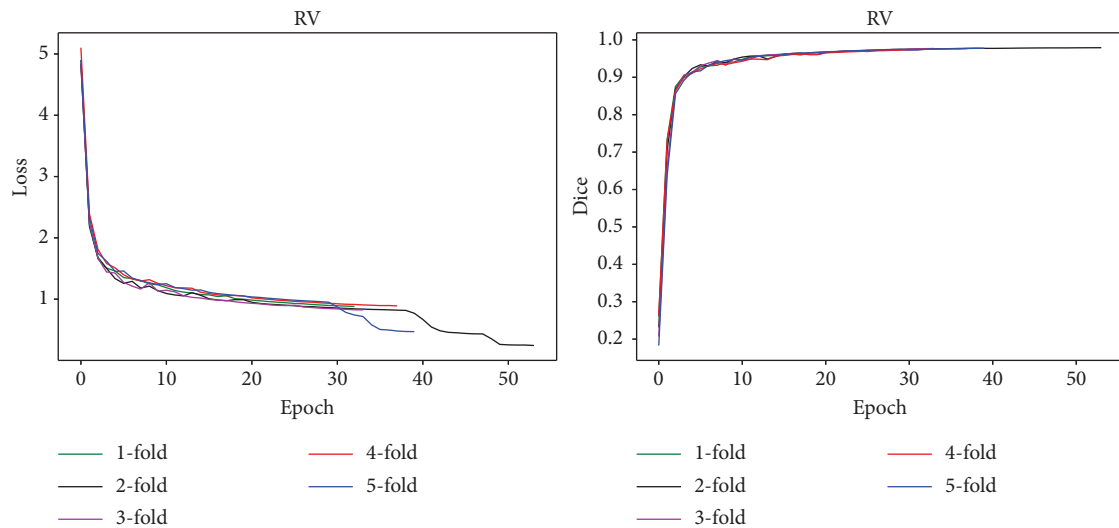


FIGURE 5: The change in loss and dice of our method on the ACDC (RV tissue).

3.2.3. Multilabel Deep Supervision. We analyzed the influence of the multilabel deep supervision in the network on the segmentation accuracy in the ACDC and ASC, which includes (a) deep supervision (DS) and (b) multilabel deep supervision (MLDS). The results are shown in Tables 5 and 6. As shown in Tables 5 and 6, compared with deep supervision, our multilabel deep supervision achieves better performance. This is because in deep supervision, the results of small-scale feature layers are inconsistent with the labels, causing the network to learn wrong information. Our multilabel deep supervision has labels of corresponding sizes for each scale feature layer. The results are consistent with the label, which can well guide the network learning.

3.3. Comparison with State-of-the-Art Methods and Discussion. In this section, we compared the proposed MMA-Net with previous state-of-the-art medical image segmentation methods on the ACDC [33] and the ASC [34].

3.3.1. Quantitative Comparison. Tables 7 and 8 show the segmentation results on the ACDC and the ASC, respectively. As shown in Tables 7 and 8, our method achieves the best performance for most of the metrics on the ACDC and the ASC. Especially in the dice, as a key indicator for evaluating the performance of medical image segmentation, our method has a great improvement compared with other methods. The

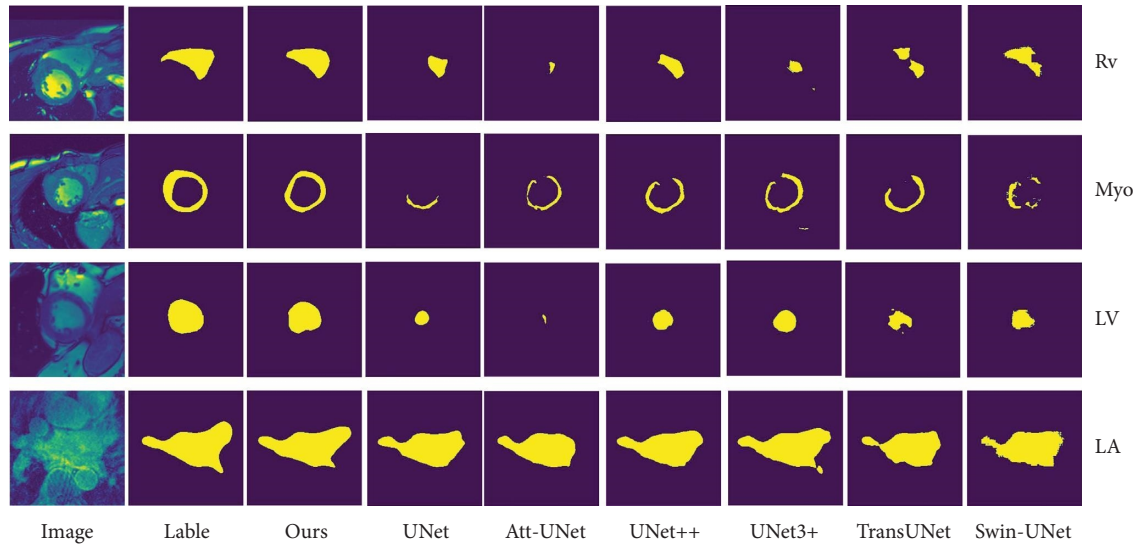


FIGURE 6: Qualitative comparisons with state-of-the-art methods.

specificity of all methods is close to 1.000 because the background is the majority, and most of the background is easily classified. Our method may be more sensitive to tissue, misclassifying many backgrounds as tissue, which may be the reason why our method does not achieve optimal performance in terms of specificity. Sensitivity is another important metric to evaluate the performance of medical image segmentation, and our method achieves a large performance improvement on RV, Myo, and LV, and a certain performance improvement on LA as well. F1 is a relatively comprehensive evaluation index for medical image segmentation performance, and our method has a certain degree of improvement compared with other methods. Figure 5 shows the change in loss and dice of our method on the ACDC (RV tissue). As the number of iterations increases, the loss function converges rapidly, proving that our network structure and training parameter design are reasonable.

3.3.2. Qualitative Comparison. Figure 6 shows the visualizations on the right ventricle (RV), myocardium (Myo), left ventricle (LV), and left atrium (LA). As shown in Figure 6, compared with other methods, our method shows significant improvement in segmentation performance. For the RV tissue, our method can localize the tissue well and segment the edges of the tissue well. For the Myo tissue, only our method formed complete rings, and none of the other methods formed complete rings. LV is an easy tissue to segment, but other methods still have some segmentation failures. Our method can segment the LV tissue more perfectly. LA is a difficult tissue to segment, and other methods are generally effective in segmenting the details of LA tissue. Our method can better segment the details of LA.

4. Conclusion

In this paper, a novel multiresolution mutual assistance network (MMA-Net) for cardiac MR images segmentation is proposed. It implements multiresolution feature interaction

and progressively improves semantic features to more completely express the information of the tissue. We compare with state-of-the-art medical image segmentation methods on the ACDC and the ASC. The mean dice score of our method in the left atrium, right ventricle, myocardium, and left ventricle are 0.919, 0.920, 0.881, and 0.960, respectively. The analysis of evaluation indicators and segmentation results shows that our method achieves the best performance in cardiac magnetic resonance images segmentation.

Data Availability

The data used to support the findings of this study are available from the corresponding author upon request.

Conflicts of Interest

The authors declare that there are no conflicts of interest regarding the publication of this paper.

Acknowledgments

This work was supported in part by the Science and Technology Planning Project of Guangdong Science and Technology Department under Grant Guangdong Key Laboratory of Advanced IntelliSense Technology (2019B121203006).

References

- [1] A. Ammar, O. Bouattane, and M. Youssfi, "Automatic cardiac cine MRI segmentation and heart disease classification," *Computerized Medical Imaging and Graphics*, vol. 88, Article ID 101864, 2021.
- [2] H. D. White, R. M. Norris, M. A. Brown, P. W. Brandt, R. M. Whitlock, and C. J. Wild, "Left ventricular end-systolic volume as the major determinant of survival after recovery from myocardial infarction," *Circulation*, vol. 76, no. 1, pp. 44–51, 1987.

- [3] M. Lynch, O. Ghita, and P. Whelan, "Left-ventricle myocardium segmentation using a coupled level-set with a priori knowledge," *Computerized Medical Imaging and Graphics*, vol. 30, no. 4, pp. 255–262, 2006.
- [4] C. Petitjean and J. N. Dacher, "A review of segmentation methods in short axis cardiac MR images," *Medical Image Analysis*, vol. 15, no. 2, pp. 169–184, 2011.
- [5] P. Peng, K. Lekadir, A. Gooya, L. Shao, S. E. Petersen, and A. F. Frangi, "A review of heart chamber segmentation for structural and functional analysis using cardiac magnetic resonance imaging," *Magnetic Resonance Materials in Physics, Biology and Medicine*, vol. 29, no. 2, pp. 155–195, 2016.
- [6] Q. Q. Tong, C. Z. Li, W. X. Si et al., "RIANet: recurrent interleaved attention network for cardiac MRI segmentation," *Computers in Biology and Medicine*, vol. 109, pp. 290–302, 2019.
- [7] S. P. O'Brien, O. Ghita, and P. F. Whelan, "A novel model-based 3D Time left ventricular segmentation technique," *IEEE Transactions on Medical Imaging*, vol. 30, no. 2, pp. 461–474, 2011.
- [8] H. Y. Lee, N. Codella, M. Cham, M. Prince, J. Weinsaft, and Y. Wang, "Left ventricle segmentation using graph searching on intensity and gradient and a priori knowledge (lvGIGA) for short-axis cardiac magnetic resonance imaging," *Journal of Magnetic Resonance Imaging*, vol. 28, no. 6, pp. 1393–1401, 2008.
- [9] L. J. Wang, M. C. Pei, N. C. F. Codella et al., "Left ventricle: fully automated segmentation based on spatiotemporal continuity and myocardium information in cine cardiac magnetic resonance imaging (LV-FAST)," *BioMed Research International*, vol. 2015, pp. 1–9, 2015.
- [10] M. Ammar, S. Mahmoudi, M. A. Chikh, and A. Abbou, "Automatic segmentation of cardiac magnetic resonance images using active appearance models and hausdorff distance," *Journal of Mechanics in Medicine and Biology*, vol. 12, no. 4, Article ID 1250059, 2012.
- [11] A. Kausar, I. Razzak, M. I. Shapiai, and A. Beheshti, "3D shallow deep neural network for fast and precise segmentation of left atrium," *Multimedia Systems*, pp. 1–11, 2021.
- [12] D. Q. Liu, Z. Jia, M. Jin et al., "Cardiac magnetic resonance image segmentation based on convolutional neural network," *Computer Methods and Programs in Biomedicine*, vol. 197, Article ID 105755, 2020.
- [13] T. A. Ngo, Z. Lu, and G. Carneiro, "Combining deep learning and level set for the automated segmentation of the left ventricle of the heart from cardiac cine magnetic resonance," *Medical Image Analysis*, vol. 35, pp. 159–171, 2017.
- [14] M. B. Khamechian and M. Saadatmand-Tarzan, "FoCA: a new framework of coupled geometric active contours for segmentation of 3D cardiac magnetic resonance images," *Magnetic Resonance in Imaging*, vol. 51, pp. 51–60, 2018.
- [15] M. Penso, S. Moccia, S. Scafuri et al., "Automated left and right ventricular chamber segmentation in cardiac magnetic resonance images using dense fully convolutional neural network," *Computer Methods and Programs in Biomedicine*, vol. 204, Article ID 106059, 2021.
- [16] S. Arirangan and K. Kottursamy, "Multi-scaled feature fusion enabled convolutional neural network for predicting fibrous dysplasia bone disorder," *Expert Systems*, Article ID E12882, 2021.
- [17] A. Saranya, K. Kottursamy, A. A. AlZubi, and A. K. Bashir, "Analyzing fibrous tissue pattern in fibrous dysplasia bone images using deep R-CNN networks for segmentation," *Soft Computing*, vol. 26, no. 16, pp. 7519–7533, 2021.
- [18] Y. Qin, X. L. Li, J. Wu, and K. Yu, "A management method of chronic diseases in the elderly based on IoT security environment," *Computers & Electrical Engineering*, vol. 102, Article ID 108188, 2022.
- [19] L. Chang, J. Wu, N. Moustafa, A. K. Bashir, and K. Yu, "AI-driven synthetic biology for non-small cell lung cancer drug effectiveness-cost analysis in intelligent assisted medical systems," *IEEE Journal of Biomedical and Health Informatics*, vol. 26, no. 10, pp. 5055–5066, 2022.
- [20] X. Huang, H. Z. Zhu, and J. X. Wang, "Adoption of snake variable model-based method in segmentation and quantitative calculation of cardiac ultrasound medical images," *Journal of Healthcare Engineering*, vol. 2021, pp. 1–13, 2021.
- [21] C. J. Liu, P. H. Gu, and Z. Y. Xiao, "Multiscale U-Net with spatial positional attention for retinal vessel segmentation," *Journal of Healthcare Engineering*, vol. 2022, pp. 1–10, Article ID 5188362, 2022.
- [22] P. Thapar, M. Rakhra, G. Cazzato, and M. S. Hossain, "A novel hybrid deep learning approach for skin lesion segmentation and classification," *Journal of Healthcare Engineering*, vol. 2022, pp. 1–21, Article ID 1709842, 2022.
- [23] M. B. Sudhan, M. Sinthuja, S. Pravinth Raja et al., "Segmentation and classification of glaucoma using U-Net with deep learning model," *Journal of Healthcare Engineering*, vol. 2022, pp. 1–10, Article ID 1601354, 2022.
- [24] O. Ronneberger, P. Fischer, and T. Brox, "U-Net: convolutional networks for biomedical image segmentation," in *Proceedings of the 2015 International Conference on Medical Image Computing and Computer Assisted Intervention (MICCAI)*, vol. 9351, pp. 234–241, Munich, Germany, October 2015.
- [25] D. P. Li, Y. J. Peng, Y. F. Guo, and J. Sun, "MFAUNet: multiscale feature attentive U-Net for cardiac MRI structural segmentation," *IET Image Processing*, vol. 16, no. 4, pp. 1227–1242, 2021.
- [26] T. S. Sharan, S. Tripathi, S. Sharma, and N. Sharma, "Encoder modified U-Net and feature pyramid network for multi-class segmentation of cardiac magnetic resonance images," *IETE Technical Review*, vol. 0, pp. 1–13, 2021.
- [27] X. W. Sun, P. Garg, S. Plein, and R. J. Geest, "SAUN: stack attention U-Net for left ventricle segmentation from cardiac cine magnetic resonance imaging," *Medical Physics*, vol. 48, no. 4, pp. 1750–1763, 2021.
- [28] H. F. Cui, L. Jiang, C. Yuwen, Y. Xia, and Y. Zhang, "Deep U-Net architecture with curriculum learning for myocardial pathology segmentation in multi-sequence cardiac magnetic resonance images," *Knowledge-Based Systems*, vol. 249, Article ID 108942, 2022.
- [29] K. N. Wang, X. Yang, J. Z. Miao et al., "AWSnet: an auto-weighted supervision attention network for myocardial scar and edema segmentation in multi-sequence cardiac magnetic resonance images," *Medical Image Analysis*, vol. 77, Article ID 102362, 2022.
- [30] H. Z. Fu, J. Cheng, Y. W. Xu, D. W. K. Wong, J. Liu, and X. Cao, "Joint optic disc and cup segmentation based on multi-Label deep network and polar transformation," *IEEE Transactions on Medical Imaging*, vol. 37, no. 7, pp. 1597–1605, 2018.
- [31] J. S. Shi, Y. G. Ye, D. X. Zhu, L. Su, Y. Huang, and J. Huang, "Automatic segmentation of cardiac magnetic resonance images based on multi-input fusion network," *Computer Methods and Programs in Biomedicine*, vol. 209, Article ID 106323, 2021.

- [32] S. L. Chen, C. Z. Qiu, W. P. Yang, and Z. Zhang, "Multi-resolution aggregation transformer UNet based on multiscale input and coordinate attention for medical image segmentation," *Sensors*, vol. 22, no. 10, p. 3820, 2022.
- [33] O. Bernard, A. Lalande, C. Zotti et al., "Deep learning techniques for automatic MRI cardiac multi-structures segmentation and diagnosis: is the problem solved," *IEEE Transactions on Medical Imaging*, vol. 37, no. 11, pp. 2514–2525, 2018.
- [34] R. Karim, R. J. Housden, M. Balasubramaniam et al., "Evaluation of current algorithms for segmentation of scar tissue from late gadolinium enhancement cardiovascular magnetic resonance of the left atrium: an open-access grand challenge," *Journal of Cardiovascular Magnetic Resonance*, vol. 15, no. 1, p. 105, 2013.
- [35] O. Oktay, J. Schlemper, L. L. Folgoc et al., "Attention U-Net: learning where to look for the pancreas," 2018, <https://arxiv.org/abs/1804.03999>.
- [36] Z. W. Zhou, M. M. R. Siddiquee, N. Tajbakhsh, and J. Liang, "UNet++: redesigning skip connections to exploit multiscale features in image segmentation," *IEEE Transactions on Medical Imaging*, vol. 39, no. 6, pp. 1856–1867, 2020.
- [37] H. M. Huang, L. F. Lin, R. F. Tong et al., "UNet 3+: a full-scale connected UNet for medical image segmentation," in *Proceedings of the ICASSP 2020 - 2020 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, pp. 1055–1059, Barcelona, Spain, May 2020.
- [38] J. Chen, Y. Lu, Q. Yu et al., "TransUNet: transformers make strong encoders for medical image segmentation," 2021, <https://arxiv.org/abs/2102.04306>.
- [39] H. Cao, Y. Y. Wang, J. Chen et al., "Swin-Unet: unet-like pure transformer for medical image segmentation," 2021, <https://arxiv.org/abs/2105.05537>.