AIP Advances

Automated segmentation of left ventricular myocardium using cascading convolutional neural networks based on echocardiography

Cite as: AIP Advances 11, 045003 (2021); https://doi.org/10.1063/5.0040863
Submitted: 16 December 2020 • Accepted: 10 March 2021 • Published Online: 01 April 2021

Shenghan Ren, Yongbing Wang, Rui Hu, et al.

Automated segmentation of left ventricular myocardium using cascading convolutional neural networks based on echocardiography

Cite as: AIP Advances 11, 045003 (2021); doi: 10.1063/5.0040863
Submitted: 16 December 2020 • Accepted: 10 March 2021 •
Published Online: 1 April 2021

Shenghan Ren,[1,2] Yongbing Wang,[1,2] Rui Hu,[1] Lei Zuo,[1] Liwen Liu,[1,a)] and Heng Zhao[2]

AFFILIATIONS

[1] Department of Ultrasound of Xijing Hospital, Xijing Hypertrophic Cardiomyopathy Center, Fourth Military Medical University, Xi'an, Shaanxi 710032, China

[2] Engineering Research Center of Molecular and Neuro Imaging of Ministry of Education, School of Life Science and Technology, Xidian University, Xi'an, Shaanxi 710071, China

[a)] **Author to whom correspondence should be addressed:** liuliwen@fmmu.edu.cn

ABSTRACT

Quickly and accurately segmenting the left ventricular (LV) myocardium from ultrasound images and measuring the thickness of the interventricular septum and LV wall play an important role in hypertrophic cardiomyopathy. However, the segmentation of the LV myocardium is a challenging task due to image blurring and individual differences. We attempted to perform LV segmentation in ultrasound images using the encoder–decoder architecture of U-Net and other networks and found it to be not accurate enough. Therefore, we propose a novel multi-task cascaded convolutional neural network (called MTC-Net) to segment the LV myocardium from echocardiography. MTC-Net contains two parts: One is pre-trained Resnet-34 followed by two decoder branches for mask and boundary detection, and the other module is pre-trained with many improved novel encoder–decoder architectures for extracting more detailed features. Both parts of the network use the atrous spatial pyramid pooling module to capture high-level text information. A hybrid loss function is engaged for mask and contour prediction. The network is trained and evaluated with echocardiographic images, which are labeled manually by doctors. The comparison study with other networks shows that MTC-Net has better accuracy and performance. MTC-Net achieves state-of-the-art performance on the test set. The mean value of the dice coefficient is 0.9442 and the mean value of intersection over union is 0.8951.

I. INTRODUCTION

Hypertrophic cardiomyopathy (HCM) is a common autosomal dominant cardiovascular disease, which is associated with thickening of the left ventricular (LV) myocardium, most commonly at the interventricular septum (IVS) below the aortic valve, caused LV outflow tract obstruction (LVOTO). It is usually without LV cavity enlargement and necessary to exclude hypertension, aortic stenosis, and congenital aortic valve septum caused by increased load such as LV wall thickening. As the most common genetic cardiomyopathy, HCM occurs in approximately 1 in 500 to 1 in 200 of the general population.[1,2] We proposed the Liwen procedure to reduce the thickness of IVS and LVOTO in patients.[3] The result proves that the Liwen procedure is a safe and effective treatment approach for HCM. However, HCM is a common disease that still poses a threat to human health.

Cardiac imaging plays a crucial role in HCM. The diagnosis of HCM rests on the detection of increased LV wall thickness by any imaging modality. According to the guidelines, in an adult, HCM is defined by a wall thickness ≥15 mm in one or more LV myocardial segments and genetic disorders can present with wall thickness ≥13 mm[1,4] as measured by any imaging technique [echocardiography, cardiac magnetic resonance imaging (CMR), or computed tomography (CT)]. Compared with other imaging methods, echocardiography has become our preferred imaging modality due to its real-time and non-radiation advantages. Standard echocardiography is an important tool for rapid assessment and quantification of left ventricular wall thickness and identification of specific details.

Therefore, the accurate assessment of pathophysiology and anatomy is essential for diagnosis.

However, the diagnosis of HCM is based on the thickness of the thickest segment of left ventricular myocardium at the end of diastole, which is measured by doctors based primarily on their clinical experience. It is a challenging task to segment the LV myocardium on echocardiography automatically. Echocardiography is highly anisotropic and position-dependent. The reflection intensity, spatial resolution, and signal-to-noise ratio depend on the depth and angle of incidence of the echocardiographic beam as well as on user-controlled depth gain settings. The backscattering of incident wavefront hitting the tissue microstructure produces speckle noise, resulting in the poor quality of the echocardiograms. The echocardiographic ultrasound images have low signal-to-noise ratios and blurred borders. Most tissues are not distinguished by their intensity value or texture. In addition, due to weak echo and patient obesity, there were missing pixels in the region of interest (ROI) of the LV, as shown in the red box in Fig. 1.

Many approaches for automatic segmentation of echocardiographic ultrasound images have been proposed in recent years. Traditional segmentation methods include the level set method,[5] active shape model,[6] and active contour method[7] etc. Lin *et al.* proposed a multi-scale level set framework for echo image segmentation.[8] The method uses different level set methods at different scale levels and fuses region and edge information with spatial scales, resulting in a good segmentation with the boundary shape constraint. Chen *et al.* built an active contour algorithm with a prior intensity and shape.[9] The segmentation results indicated that the proposed model provides close agreement with expert traced borders. However, these methods need to empirically initialize the contour model and manually adjust the parameters according to the target to obtain better segmentation results. Other approaches such as conventional machine learning methods[10,11] require features to be extracted manually first, and these methods take a long time for extracting the features of the medical image due to the large differences in left ventricular myocardial anatomy between HCM patients.

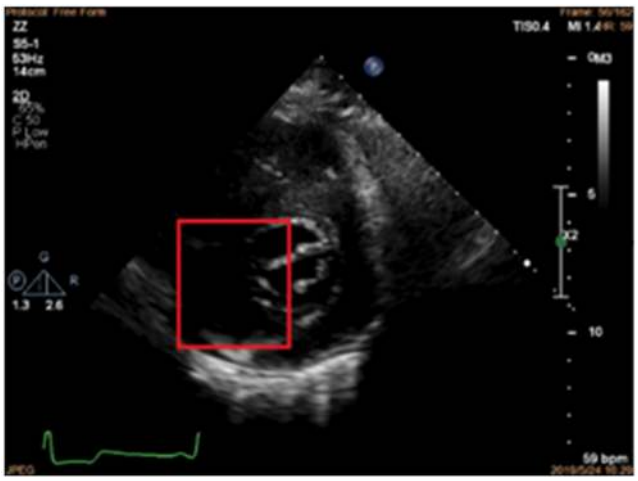

FIG. 1. The poor quality of echocardiographic ultrasound images (the red box is the part with the missing pixels).

In recent years, convolutional neural networks (CNNs) have been widely used in medical image segmentation due to their powerful feature extraction capabilities.[12,13] State-of-the-art performance has been achieved in medical image segmentation tasks based on encoder–decoder architectures such as U-Net.[14] In order to reduce the effect of information loss at the up-sampling stage,[15] U-Net fuses the features of the encoder and decoder through a cascade approach. However, due to its simple architecture, U-Net segmentation in our ultrasound image data, which was trained using the traditional cross-entropy loss function, does not give good results. This is mainly due to the following drawbacks of this type of architecture. First, encoder–decoder networks suffer from spatial information loss due to down sampling operations performed via max-pooling layers. Second, this architecture mostly uses cross-entropy loss and dice loss as loss functions, but cross-entropy loss may lead to over-segmentation and under-segmentation when the categories are unbalanced. The disadvantage of dice loss is that it penalizes false negative and false positive equally, resulting in a high precision and low recall of the segmentation.

In order to overcome the above drawbacks of U-Net, we proposed a novel multi-task cascaded CNN model called MTC-Net with a mixed loss function in this paper. The main contributions of our paper are as follows: We proposed MTC-Net cascades' two modules together. The first module uses two decoder branches for mask and contour prediction, respectively. After passing the model, the approximate area of interest is obtained. The original inputs, predicted masks, and boundaries are cascaded and passed into a second module. We use the output of the second model as the final prediction. The predicted mask boundaries are smoother due to the inclusion of boundary information during the training of the second network. By combining cross-entropy loss, structural similarity index (SSIM) loss,[16] and dice loss, we proposed a new hybrid loss function that optimizes the network at three levels: pixel, patch, and map-level, respectively, which can significantly improve the prediction result of the model. Using commonly used segmentation metrics including dice similarity coefficient (DSC) and intersection over union (IOU), we compared our model with the predictions of other models, and the results showed that our model is more accurate and robust. Our model is better than other models on all metrics with over-segmentation and under-segmentation significantly reduced.

This paper is organized as follows: The details of the proposed segmentation method are introduced in Sec. II. The dataset and experiments are presented in Sec. III. The segmentation results are analyzed in Sec. IV. Finally, discussion and conclusion are made in Sec. V.

II. METHOD

A. Architecture of the network

The overview of the architecture of MTC-Net is shown in Fig. 2. The network contains two modules: the candidate masks' prediction module and the reinforced learning module. The first module uses a pre-trained Resnet-34 for feature extraction, followed by two decoder branches: one is for mask prediction and the other is for contour prediction. The problem of spatial information loss due to max-pooling and stepwise convolution is solved by fusing the feature map of the encoder part corresponding to the decoder

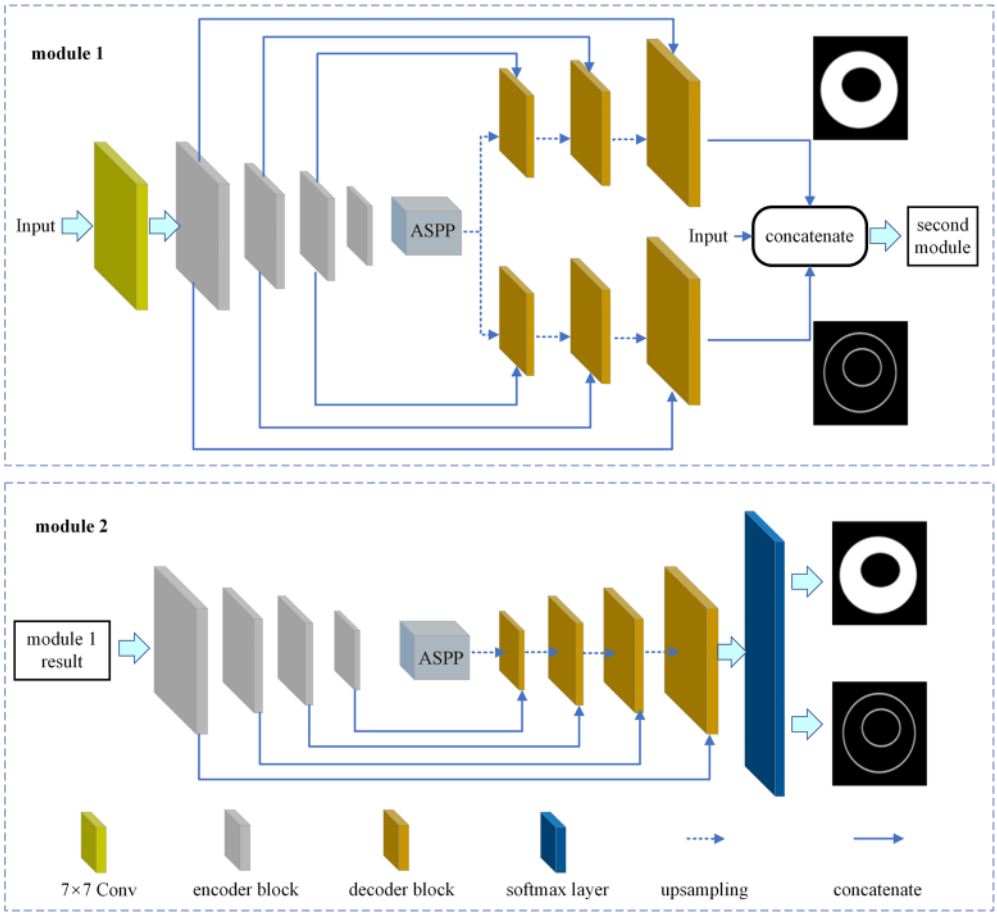

FIG. 2. Architecture of MTC-Net for segmentation of LV from echocardiographic ultrasound images.

by summing operations. The second module is a multi-task network with a typical encoder–decoder architecture. The low-level feature maps are merged with high-level feature maps by using the concatenate method in the second module. The probability ROI mask, boundary contour generated by the first module, and the input image are concatenated together and fed to the second module. Finally, a softmax layer is used for mask and contour predictions. Considering that atrous convolution can expand the feature reception field without sacrificing feature spatial resolution,[17] we use atrous convolution pyramid pooling (ASPP) in both modules. We define a hybrid loss function that allows the network to be trained end-to-end for both two modules.

B. Encoder and decoder explanation

The encoders in our first module use pre-trained Resnet-34, then followed by a ASPP module, which employs multiple filters with different rates to capture objects and context at multiple scales. Following the up-sampling, feature maps from the ASPP module are concatenated with feature maps from Resnet-34 and subsequently passed to the softmax layer for ROI mask or boundary contour prediction. The predicted boundary contour and ROI mask as well as the original image are concatenated together and passed into the second module.

Since dilated convolution[18] can effectively increase the receptive field without changing the resolution, we consider how to increase the receptive field in the encoder of the second module. In the second module, we use four encoder blocks to extract the feature maps [shown in Fig. 3(a)]. Each encoder block first performs a 3 × 3 convolution, then followed by four parallel atrous convolutions with dilated rates of 1, 4, 6, and 8 and the global average pooling layer. The parallel layers are cascaded together and passed to a 3 × 3 convolution layer. Each convolution of the encoder block is followed by a batch normalization, which can reduce the internal co-variant shift and increases the speed of convergence. The rectified Linear Unit (ReLU) is severed as the activation function. Finally, the feature map is followed by the max-pooling layer to halve the spatial dimension and pass into the next encoder block.

We use different strategies for the decoder implementation of the two modules. In the first module, the feature map is up-sampled by transposed convolution, then followed by a 3 × 3 convolution.

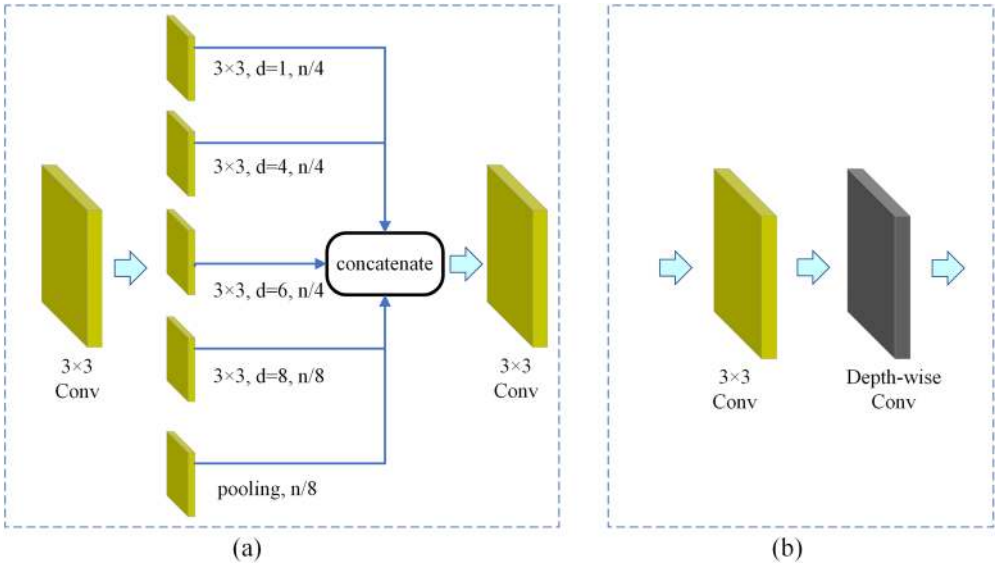

FIG. 3. (a) Structure of the encoder block in module 2 and (b) structure of the decoder block in module 2.

Finally, the feature map of the corresponding encoder part is added to enhance the feature representation further. In the second module, the decoder block first performs a 2 × 2 bi-linear up-sampling on the feature map, and then the up-sampled feature map is cascaded with the feature map of the corresponding encoder part. The cascaded feature map performs a 3 × 3 convolution and depth-wise convolution[19] [Fig. 3(b)]. Each convolution of the decoder block is followed by a batch normalization and ReLU activation function to introduce nonlinearity to the model.

C. Loss function

The loss function of MTC-Net consists of four components. The proposed hybrid loss function is used for mask, and the cross-entropy loss function is used for contour. The total loss is given by

$$L_{total} = L_{model1-mask} + L_{model1-contour} + L_{model2-mask} + L_{model2-contour}, \quad (1)$$

where $L_{model1-mask}$ and $L_{model1-coutour}$ are the losses of the ROI mask and predicted contour in the first module, respectively, while $L_{model2-mask}$ and $L_{model2-coutour}$ are the losses of the ROI mask and contour in the second module, respectively. Each part of the loss function in Eq. (1) is calculated by a hybrid loss function, which is defined as follows:

$$L_{\text{hybrid}} = L_{model*} = L_{\text{bce}} + L_{\text{dice}} + L_{\text{ssim}}, \quad (2)$$

where L_{bce}, L_{dice}, and L_{ssim} denote the losses of cross-entropy, dice coefficient, and SSIM. L_{ssim} is defined as follows:

$$L_{\text{ssim}} = 1 - \frac{(2\mu_x\mu_y + C_1)(2\sigma_{\text{xy}} + C_2)}{(\mu_{\text{x}}^2 + \mu_y^2 + C_1)(\sigma_{\text{x}}^2 + \sigma_y^2 + C_2)}, \quad (3)$$

where μ_x and μ_y are the average densities of two images, σ_x^2 and σ_y^2 are the variances of the density of two images, σ_{xy} is the covariance between two images, and C_1 and C_2 are two constants to avoid dividing by 0 and defined as $C_1 = (k_1 L)^2$ and $C_2 = (k_2 L)^2$, where $k_1 = 0.01$, $k_2 = 0.03$, and L is the dynamic range of pixel values.

The cross-entropy loss L_{bce} and dice loss L_{dice} calculate the classification error from the pixel and map-level, respectively. They are calculated as

$$L_{\text{bce}} = \sum_{i=1}^{N} g_i \log p_i + \sum_{i=1}^{N} (1 - g_i) \log(1 - p_i), \quad (4)$$

$$L_{\text{dice}} = 1 - \frac{2\sum_{i=1}^{N} p_i g_i}{\sum_{i=1}^{N} p_i^2 + \sum_{i=1}^{N} g_i^2}, \quad (5)$$

where p represents the output of the network after the sigmoid function and g is the true label.

III. DATASET AND EXPERIMENTS

A. Dataset preparation

In this study, the dataset from Xijing Hypertrophic Cardiomyopathy Center (Xijing hospital) contains a total of 200 cases of echocardiography of the horizontal short-axis view of the mitral valve and passes the ethic approvals. All the data were acquired at end diastole in order to obtain stable ultrasound images. The image is pre-processed using contrast-adaptive histogram equalization (CLAHE)[20] to improve the image contrast. In order to avoid over-fitting, we augment the training data to 10 times by using rotation. In order to eliminate the influence of background and accelerate the training process, we crop the approximate area of the LV myocardium and unify the size to 448 × 448. The dataset was divided into the training set, validation set, and a test set according to the ratio of 8:1:1. The physician performs the task of marking

the endocardium and epicardium to segment the LV myocardium using the interactive segmentation software for training. Five doctors are engaged in manual segmentation. Each doctor labeled the image data as a separate atlas. In order to maintain the inter-rater reliability and consistency, the similarity between different atlases was calculated and the atlas with the lowest similarity was removed. The ground truth is obtained from the selected atlases by using a majority voting strategy, which is implemented by SimpleITK.

B. Training and implementation details

For the training step, 20 images were randomly selected from the dataset to evaluate the performance of MTC-Net during training. Another 20 images were randomly selected for testing. The segmentation errors from these images were used to adjust the weights in different layers of the networks. We use the Adam optimizer with an initial learning rate of 1×10^{-4} and a batch size of 4 for all experiments. The model was trained with 100 epochs and periodically assessed against the validation set. All the models are implemented using PyTorch. The training was parallelly run on the graphics processing unit (GPU) (NVIDIA GeForce GTX 1070).

C. Evaluation metrics

In order to evaluate the MTC-Net quantitively, DSC, precision, and IOU are engaged as the standards for comparison. The dice coefficient and IOU are calculated as follows:

$$DSC = 2\frac{|R_G \cap R_S|}{|R_G| + |R_S|}, \tag{6}$$

$$IOU = \frac{|R_G \cap R_S|}{|R_G \cup R_S|}, \tag{7}$$

where R_G and R_S represent the ground truth region and the segmentation result, respectively, and $|\cdot|$ denotes the number of pixels.

D. Thickness measurement

The measurement of thickness of the IVS and LV wall plays an important role in HCM. As shown in Fig. 4, the thickness is measured based on the results of our model segmentation. Since our equipment obtains the ultrasound image resolution at 96 dpi in both the horizontal and vertical direction, the thickness can be calculated with the steps shown in Fig. 4. First, obtain the segmentation result of LV myocardium and find the center line of the region of LV wall. Then, calculate the Euclidean distance transformation of the center line to the boundary of region 1 [shown in Fig. 4(a)], and we can obtain the point of the maximum distance to the boundary in region 1, which is depicted by point O in Fig. 4(b). Considering each inch contains 96 pixels in our ultrasound image, the maximum thickness (mm) can be represented by line AB through point O, which is perpendicular to the center line of the LV wall and calculated as

$$Dis(AB) = Px(OA) \times 2 \times 25.4/96(\text{mm}). \tag{8}$$

IV. RESULTS

A. Overview

In this study, we compare MTC-Net with other segmentation networks, such as U-Net,[14] Conv-MCD,[21] DeepLabv3+,[22] and CE-Net.[23] We can find that our proposed method MTC-Net has DSC, IOU, and accuracy of 0.9442, 0.8951, and 0.9424, respectively, which are better than other methods in terms of DSC, IOU, and accuracy. T-test and P-value on DSC are also listed in Table I. The detailed comparison result shown in Table I proved that MTC-Net is better than all other models.

We randomly select five cases of the test result to compare MTC-Net with other models visually, which are shown in Fig. 5. The red curves are the ground truth labeled by doctors, while the green curves are the segmentation result by different models. From Fig. 5, we can find that our proposed method is much more accurate than other models. Due to that, we add the shape information of LV myocardium in our model during training, and compared with other models, over-segmentation or under-segmentation is significantly reduced. The boundaries of the segmentation results are smoother, and isolated noise is significantly reduced in MTC-Net. The predicted results by MTC-Net are more closely aligned with the ground truth labeled by doctors.

B. Evaluate with clinical metric

In order to evaluate MTC-Net clinically, we measured the thickness of LV myocardium. The number of images for

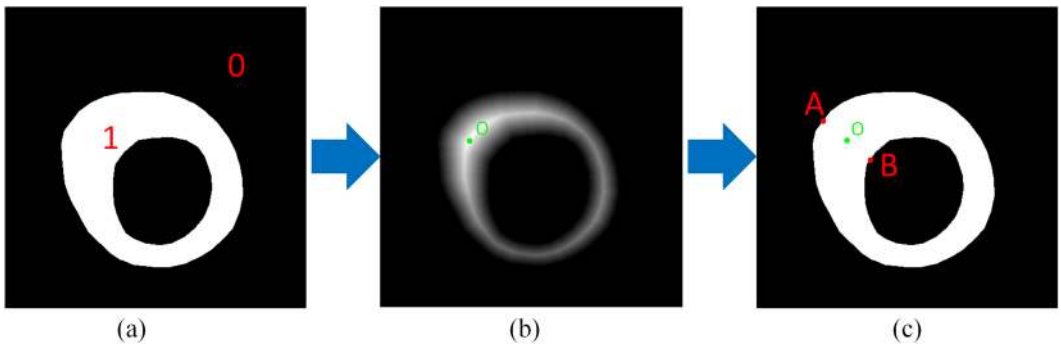

FIG. 4. The process to measure the thickness of the IVS and LV wall. (a) Segmentation result of LV myocardium, (b) result of Euclidean distance transformation, and (c) the line AB represents the maximum thickness.

TABLE I. Comparison of U-Net, Conv-MCD, DeepLabv3+, CE-Net, and MTC-Net based on our datasets in terms of DSC, IOU, and precision. The best results are highlighted in bold.

Method	DSC	IOU	Precision	T-test on DSC	*P*-value on DSC
U-net	0.9064 ± 0.0298	0.8301	0.8782	8.35	2.037×10^{-7}
Conv-MCD	0.9182 ± 0.025	0.8496	0.9005	5.42	4.577×10^{-5}
DeepLabv3+	0.9198 ± 0.0204	0.8522	0.9124	3.85	0.001 28
CE-net	0.9355 ± 0.0188	0.8794	0.9371	2.93	0.009 35
MTC-net	**0.9442 ± 0.0221**	**0.8951**	**0.9424**	...	...

evaluation is the same as the above segmentation experiment. As shown in Fig. 6, the red points stand for the ground truth labeled by doctors, while the green points stand for the results obtained by MTC-Net. We can find that the results by MTC-Net have a good consistency with the ground truth. The average error of the thickness measured by MTC-Net has only 1.278 ± 0.1425 mm distance to the ground truth. The results proved that MTC-Net has a great potential in clinical application.

C. Ablation experiments

We also performed the ablation experiments to analyze which trick is effective in MTC-Net. The ablation result shows that the boundary information and parallel atrous convolutions play an important role in our method. First, we remove the boundary information from MTC-Net to test its performance, which is called MTC-Net-BF. Specifically, the boundary branch in module 1 and

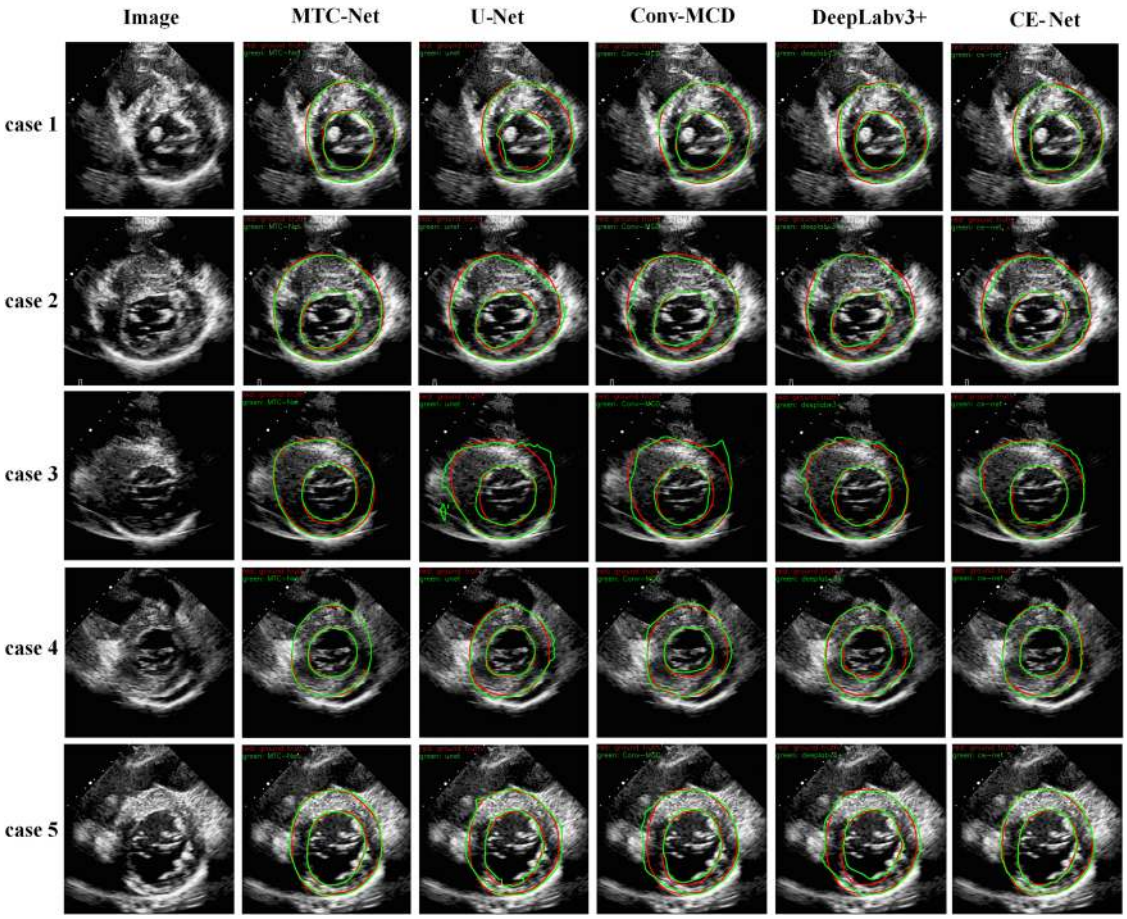

FIG. 5. Results of the LV myocardium segmentation, from left to right: original test image, result by MTC-Net, U-Net, Conv-MCD, DeepLabv3+, and CE-Net, respectively.

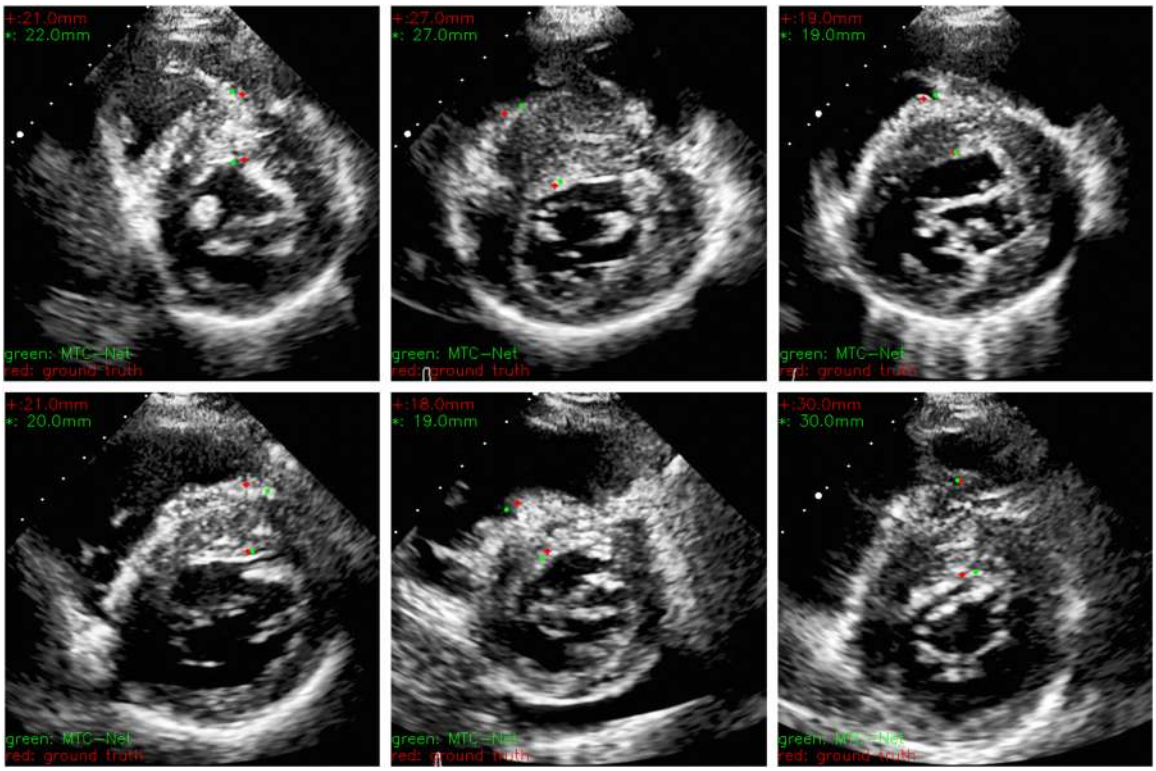

FIG. 6. Thickness measurement based on the segmentation results: The red points represent the ground truth labeled by doctors, and the green points represent the result obtained by MTC-Net.

the boundary loss in module 2 are removed in MTC-Net-BF. Second, to verify the contribution of the encoder–decoder block in module 2 to the performance improvement, we replace the encoder and decoder blocks in MTC-Net module 2 with two convolutional layers, called MTC-Net-ED. Table II shows that DSCs of both MTC-Net-BF and MTC-Net-ED drop to 0.9379 and 0.9308, respectively, compared with MTC-Net. IOU and precision of both MTC-Net-BF and MTC-Net-ED are also dropped. The MTC-Net-BF obtains a better result than MTC-Net-ED in DSC and IOU. T-test and P-value on DSC are also listed in Table II. T-test on DSC between MTC-Net-BF and MTC-Net shows that MTC-Net-BF gets almost the same performance with MTC-Net, which proves that boundary information is not obvious to improve the segmentation accuracy.

V. DISCUSSION

In this paper, we proposed a multitasking method based on cascading convolutional neural networks for echocardiographic image segmentation. The method contains two modules: a pre-trained Resnet-34 followed by two decoder branches for mask and boundary detection and the other module with many improved novel encoder–decoder architectures. Both modules are involved in ASPP for extension of the receptive field and capturing high level features. Compared to U-Net, MTC-Net has more parameters and hence a longer learning time. Therefore, we use depth-wise separable convolution to reduce the computational complexity of decoding block in the second module. MTC-Net employs a new hybrid loss function that can significantly improve the prediction results of the model.

TABLE II. Comparison of our method and ablation method on our datasets in terms of DSC, IOU, and precision. The boldface values are the results of MTC-Net.

Method	DSC	IOU	Precision	T-test on DSC	*P*-value on DSC
MTC-net-BF	0.9379 ± 0.0244	0.8840	0.9222	1.396	0.180 8
MTC-net-ED	0.9308 ± 0.0155	0.8709	0.9264	3.63	0.002 07
MTC-net	**0.9442 ± 0.0221**	**0.8951**	**0.9424**	...	...

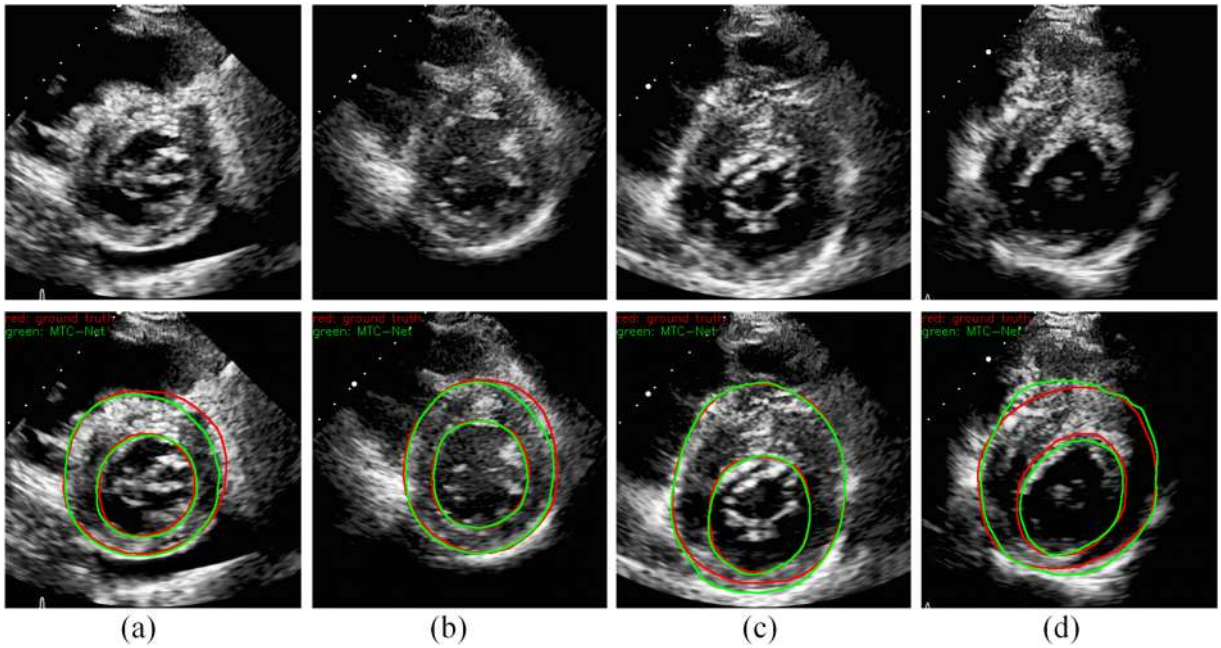

FIG. 7. Prediction results for low quality images using MTC-Net. The red curve is the ground truth, and the green curve is the result of MTC-Net. [(a)–(c)] The prediction results of over-segmentation and under-segmentation images with blurred boundaries using MTC-Net. (d) The prediction result of images with ROI missing pixels.

We compare MTC-Net with other models, and the results show that our model is more accurate and robust.

In the ablation study, we compare different MTC-Net models, including MTC-Net-BF and MTC-Net-ED. The boundary information in MTC-Net-BF is ablated, and the encoder as well as the decoder block in MTC-Net-ED is replaced by a convolutional layer. The results show that both MTC-Net-BF and MTC-Net-ED are worse than MTC-Net. We can find that the boundary information can improve the segmentation results, especially for medical images with blurred boundaries. MTC-Net-ED is worse than MTC-Net-BF, which proves that the extension of the receptive field plays an important role in the semantic segmentation.

However, we also found that MTC-Net still needs improvement in dealing with low quality ultrasound images. Figure 7 shows the prediction results for low quality images using MTC-Net, with the upper line being the test image and the lower line being the prediction result. The red curve is the ground truth, and the green curve is the prediction results of MTC-Net. Figures 7(a)–7(c) show the results of over-segmentation and under-segmentation of images with blurred boundaries using MTC-Net. Figure 7(d) shows the results of MTC-Net when processing images with ROI missing pixels. Although we have tried to add many tricks to improve segmentation performance in MTC-Net, processing images with missing pixels in the ROI is a tough task. A better solution is to acquire high-quality ultrasound images or supplement them with other types of images, such as CMR or CT.

In future study, we plan to apply our model to other medical image segmentation tasks. Compared to U-Net, our model has more parameters, leading to longer learning time. We will focus on simplifying the network model and reducing the network parameters without losing network performance. Many other types of encoder–decoder networks[24–26] such as CDED-net and M-Net also can be used for echocardiographic image segmentation. We will do more study on encoder–decoder networks in future research. We also note that recent breakthroughs in neural architecture search[27] (NAS) have motivated various applications of semantic segmentation in natural scenes. This technique is different from traditional neural networks. The architecture is obtained through search and optimization and thus performs even better than hand-crafted networks. We will try this technique in our dataset in future studies.

VI. CONCLUSION

In this paper, we proposed a novel cascaded convolutional neural network to segment the left ventricular myocardium from echocardiography. The network produces satisfactory results due to our proposed tricks, such as hybrid loss function, expansion of inception field, depth-wise convolution, and boundary information. Compared with other methods, MTC-Net is more accurate and robust.

AUTHORS' CONTRIBUTIONS

S.R. and Y.W. contributed equally to this work.

ACKNOWLEDGMENTS

This work was supported, in part, by the National Key R & D Program of China under Grant Nos. 2016YFC1300302 and

2017ZDXMSF058, the National Natural Science Foundation of China under Grant Nos. 61901338 and 82071932, and the Natural Science Basic Research Plan in Shaanxi Province of China under Grant No. 2019JQ-496.

DATA AVAILABILITY

The data that support the findings of this study are available within the article and are available from the corresponding author upon reasonable request.

REFERENCES

[1]P. M. Elliott, A. Anastasakis, M. A. Borger, M. Borggrefe, F. Cecchi, P. Charron, A. A. Hagege, A. Lafont, G. Limongelli, H. Mahrholdt, W. J. McKenna, J. Mogensen, P. Nihoyannopoulos, S. Nistri, P. G. Pieper, B. Pieske, C. Rapezzi, F. H. Rutten, C. Tillmanns, and H. Watkins, Eur. Heart J. **35**(39), 2733–2779 (2014).

[2]C. Semsarian, J. Ingles, M. S. Maron, and B. J. Maron, J. Am. College Cardiol. **65**(12), 1249–1254 (2015).

[3]L. Liu, J. Li, L. Zuo, J. Zhang, M. Zhou, B. Xu, R. T. Hahn, M. B. Leon, D. H. Hsi, J. Ge, X. Zhou, S. Ge, and L. Xiong, J. Am. College Cardiol. **72**, 1898–1909 (2018).

[4]B. J. Gersh, B. J. Maron, R. O. Bonow, J. A. Dearani, M. A. Fifer, M. S. Link, S. S. Naidu, R. A. Nishimura, S. R. Ommen, H. Rakowski, C. E. Seidman, J. A. Towbin, J. E. Udelson, and C. W. Yancy, Circulation **124**(24), 2761–2796 (2011).

[5]C. Corsi, G. Saracino, A. Sarti, and C. Lamberti, IEEE Trans. Med. Imaging **21**(9), 1202–1208 (2002).

[6]T. Cootes, A. Hill, C. Taylor, and J. Haslam, Image Vision Comput. **12**(6), 355–365 (1994).

[7]E. Debreuve, M. Barlaud, G. Aubert, I. Laurette, and J. Darcourt, IEEE Trans. Med. Imaging **20**(7), 643–659 (2001).

[8]N. Lin, W. Yu, and J. S. Duncan, in *International Conference on Medical Image Computing and Computer-Assisted Intervention* (Springer, Berlin, Heidelberg, 2002), pp. 682–689.

[9]Y. Chen, F. Huang, H. D. Tagare, and M. Rao, Int. J. Comput. Vision **71**(3), 259–272 (2007).

[10]O. Oktay, W. Shi, K. Keraudren, J. Caballero, and D. Rueckert, "Learning shape representations for multi-atlas endocardium segmentation in 3D echo images," in *Proceedings of the MICCAI Challenge on Echocardiographic Three-Dimensional Ultrasound Segmentation (CETUS)* (MIDAS Journal, Boston, 2014), pp. 57–64.

[11]K. Keraudren, O. Oktay, W. Shi, J. Hajnal, and D. Rueckert, "Endocardial 3D ultrasound segmentation using autocontext random forests," in *Proceedings of the MICCAI Challenge on Echocardiographic Three-Dimensional Ultrasound Segmentation (CETUS)* (MIDAS Journal, Boston, 2014).

[12]D. Shen, G. Wu, and H.-I. Suk, Annu. Rev. Biomed. Eng. **19**(1), 221–248 (2017).

[13]Q. Dou, L. Yu, H. Chen, Y. Jin, X. Yang, J. Qin, and P.-A. Heng, Med. Image Anal. **41**, 40–54 (2017).

[14]O. Ronneberger, P. Fischer, and T. Brox, in *International Conference on Medical Image Computing and Computer-Assisted Intervention* (Springer, Cham, 2015), pp. 234–241.

[15]X. Li, Y. Guo, F. Jiang, L. Xu, F. Shen, Z. Jin, and Y. Wang, IEEE Access **8**, 5805–5816 (2020).

[16]Z. Wang, E. P. Simoncelli, and A. C. Bovik, in *IEEE Asilomar Conference on Signals, Systems and Computers* (IEEE, 2003).

[17]L.-C. Chen, G. Papandreou, I. Kokkinos, K. Murphy, and A. L. Yuille, IEEE Trans. Pattern Anal. Mach. Intell. **40**(4), 834–848 (2018).

[18]F. Yu, V. Koltun, and T. Funkhouser, in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition* (IEEE, 2017), pp. 472–480.

[19]R. Zhang, F. Zhu, J. Liu, and G. Liu, IEEE Trans. Inf. Forensics Secur. **15**, 1138–1150 (2020).

[20]S. M. Pizer, R. E. Johnston, J. P. Ericksen, B. C. Yankaskas, and K. E. Muller, in *Proceedings of the First Conference on Visualization in Biomedical Computing* (IEEE Computer Society, 1990), pp. 337–338.

[21]B. Murugesan, K. Sarveswaran, S. M. Shankaranarayana, K. Ram, J. Joseph, and M. Sivaprakasam, in *International Workshop on Machine Learning in Medical Imaging (MLMI 2019)* (Springer, Granada, Spain, 2019), pp. 292–300.

[22]L.-C. Chen, Y. Zhu, G. Papandreou, F. Schroff, and H. Adam, in *Proceedings of the European Conference on Computer Vision (ECCV, 2018)* (Springer, Munich, Germany, 2018), pp. 801–818.

[23]Z. Gu, J. Cheng, H. Fu, K. Zhou, H. Hao, Y. Zhao, T. Zhang, S. Gao, and J. Liu, IEEE Trans. Med. Imaging **38**(10), 2281–2292 (2019).

[24]N.-Q. Nguyen and S.-W. Lee, IEEE Acess **7**, 33795–33808 (2019).

[25]M. Mohtasebi, M. Bayat, S. Ghadimi, H. Abrishami Moghaddam, and F. Wallois, IRBM **42**(1), 19–27 (2021).

[26]X. Fu, N. Cai, K. Huang, H. Wang, P. Wang, C. Liu, and H. Wang, IEEE Access **7**, 148645–148657 (2019).

[27]T. Elsken, J. Metzen, and F. Hutter, J. Mach. Learn. Res. **20**, 1–21 (2019).