2018 24th International Conference on Pattern Recognition (ICPR)
Beijing, China, August 20-24, 2018

Fully Automatic Segmentation of the Left Ventricle Using Multi-Scale Fusion Learning

Tianchen Yuan†, Qianqian Tong†, Xiangyun Liao‡, Xinling Du§, and Jianhui Zhao†
†School of Computer, Wuhan University, Wuhan, Hubei, P.R. China
‡Shenzhen Key Laboratory of Virtual Reality and Human Interaction Technology,
Shenzhen Institutes of Advanced Technology, Chinese Academy of Sciences, Shenzhen, P.R. China
§Union Hospital, Tongji College, Huazhong University of Science and Technology, Wuhan, Hubei, P.R. China
Email: ytcpcc@163.com, {qianqiantong, jianhuizhao}@whu.edu.cn, xyunliao@gmail.com, xinlingdu@hust.edu.cn

***Abstract*—Segmentation of the left ventricle (LV) is essential for quantitative calculation of clinical indices for analyzing the cardiac contractile function. However, it is challenging to automatically segment small-contour cardiac magnetic resonance (CMR) images for traditional convolutional neural networks (ConvNets) because of their low robustness to scale variation. In this paper, we propose a multi-scale fusion learning method to advance the performance of ConvNets for the LV segmentation. To realize our multi-scale fusion learning, single-scale input and multi-scale output (SIMO) networks are firstly trained to construct a SIMO-based multi-scale fusion network (SIMO-based MSF_Net). The trained SIMO networks produce different-scale coarse prediction results which are then fused into another multi-scale network. Finally, the coarse results are progressively refined to yield finer segmentation results. Our multi-scale fusion learning is evaluated on MICCAI 2009 challenging database for the LV segmentation. Experimental results demonstrate the robustness of our SIMO-based MSF_Net for the segmentation of challenging CMR images and the metric of "Good contours" achieves 98.35% on the testing set, which is greatly improved compared with the state-of-the-art methods.**

I. Introduction

Cardiovascular diseases (CVDs) are the leading cause of death globally according to the report by World Health Organization [1]. Early and accurate diagnoses are essential for the management of these disorders [2]. Accurate segmentation of left ventricle (LV) form cardiac magnetic resonance (CMR) images is important for calculation of clinical indices such as ventricular volume, ejection fraction, which can provide necessary information for quantitatively analyzing cardiac functionalities [3], [4]. However, there are many challenges to automatically segment the LV because of the inherent characteristics of CMR imaging: CMR images possess intensity inhomogeneity; the intensity of trabeculations and papillary muscles (TPM) is similar to that of the myocardium; the presence of left ventricular outflow track (LVOT) which takes the same intensity profiles as the cavity.

Level set methods [5], [6] are usually explored for the automatic LV segmentation from CMR images [4], [7]–[10]. However, these methods prone to yield boundary leakage or excessive shrinkage because of bad initial contours or inappropriate shape constraint for the difficulties of LV segmentation. Machine learning models perform LV segmentation by extracting more effective features, not only concerning the intensity difference at the border but also including the shape of the LV, which is beneficial for handling intensity inhomogeneity, the presence of TPM and LVOT in CMR images. However, it is complicated to acquire comprehensive medical data for training a highly precise model. Several researchers have tried to adopt combined models based on machine learning to improve the performance of LV segmentation [3], [11]–[14]. Avendi et al. [3] adopted convolutional neural networks (ConvNets) [15]–[17] to automatically detect the LV chamber and incorporated the inferred shape employing deep learning into a level set function. As described in [3], contours of image slices near the apex are smaller than those at the base. What is worse, small-contour images always possess blurred boundary. To improve the segmentation accuracy, Avendi et al. [3] divided the training set into large-contour and small-contour groups and trained two networks. However, it is obvious that the dataset for training each network is smaller than training a single network using all training data. Therefore, it is urgent to improve ConvNets' robustness to scale variation thus to accurately segment all slices of CMR images.

To improve the ConvNets' robustness to scale variation, researchers usually utilized features extracted from images of different scales [18] and from different layers of the standard ConvNets trunk [19]. To improve the performance of LV segmentation, Tong et al. [20] constructed a joint multi-scale ConvNet (JMS_Net) which joints the results predicted by the networks trained using different-scale images and the results predicted by another network with several branches. However, JMS_Net simply joints results of two trained networks, which can not take full advantage of their multi-scale representation ability because of their respective learning stage.

In this paper, we aim to further advance the performance of ConvNets for the accurate segmentation of the LV. To achieve this goal, multi-scale fusion learning is proposed by fusing prediction results of different-scale networks into another multi-scale network. Moreover, we utilize single-scale input and multi-scale output (SIMO) networks to produce different-scale prediction results, and then these results are fused into different-scale branches of the second multi-scale networks, yielding SIMO-based multi-scale fusion network (MSF_Net). Furthermore, multi-scale fusion networks which fuse different numbers of multi-scale prediction results are evaluated on the

978-1-5386-3788-3/18/$31.00 ©2018 IEEE

same testing dataset, which is used to further explore the performance of the proposed multi-scale fusion learning. The contributions of this paper are summarized as follows.

- To advance the performance of ConvNets for the LV segmentation, we present multi-scale fusion learning which fuses prediction results of different-scale networks into another multi-scale network. Our multi-scale fusion learning can produce better segmentation results, demonstrating its advantages of competitive multi-scale feature representation.
- SIMO-based MSF_Net is constructed to improve our multi-scale fusion learning. SIMO-based MSF_Net is compared with MIMO-based MSF_Net and SISO-based MSF_Net which produces different-scale results by multi-scale input and multi-scale output (MIMO) networks and single-scale input and single-scale output (SISO) networks, and our SIMO-based MSF_Net performs best.
- To further explore the performance of the proposed multi-scale fusion learning, different numbers of multi-scale prediction results by different-scale networks are fused into the second multi-scale network. In our experiment, the performance of our MSF_Nets gets better and better as the number of multi-scale prediction results increases.

II. Multi-scale fusion Network for LV segmentation

To improve the performance of the LV segmentation, we first detect the region of interest (ROI) containing the LV from CMR images, which can alleviate the impact of surrounding organs. The ROI is then pipelined to our constructed SIMO-based MSF_Net to yield segmentation results. This section firstly describes the preprocessing procedure including data augmentation and ROI detection. Multi-scale fusion learning method is then described and finally we construct SIMO-based MSF_Net for the LV segmentation.

A. Data Augmentation and ROI Detection

In this paper, we use the Medical Image Computing and Computer-Assisted Intervention (MICCAI) 2009 challenge database [21] to train and assess the performance of our multi-scale fusion learning method. Manual segmentation of the endocardial and epicardial by experienced experts is included in this database.

There are only 260 images annotated by experts for training in the MICCAI 2009 database. Obviously, it is challenging to utilize this limited training set to obtain a robust deep learning model. The training set is firstly augmented in this paper. The size of the raw image $\boldsymbol{I}$ is 256×256. We first denote a square window as $\boldsymbol{W}$. The size of $\boldsymbol{W}$ is $m_w \times n_w$ and $\boldsymbol{W}$ contains the LV. Firstly, we rotate $\boldsymbol{I}$ by α and the unit of α is degree. Thus we can obtain a rotated image $\boldsymbol{I}_1 = R(\boldsymbol{I}, \alpha)$. Then the square window $\boldsymbol{W}$ is translated on the image $\boldsymbol{I}_1$ with a step of ρ in the direction of d, and the translated image $\boldsymbol{I}_2 = T(\boldsymbol{I}_1, \rho, d)$. With this method, we can acquire more training data which is hundreds of times larger than the raw training data.

The raw CMR images usually include the heart and surrounding organs, which will mislead the segmentation model and produce bad results. ROI is firstly detected from the raw image $\boldsymbol{I}$ to tackle this issue [20] using a ROI detection ConvNet. The ROI detection ConvNet contains seven layers with parameters to be learned: three convolutional layers, three pooling layers and a fully connected (FC) layer.

B. Multi-scale Fusion Learning

As elaborated in Section I, JMS_Net improves the ability of multi-scale representation by jointing the prediction results of P_Net and B_Net [20]. We can see that the design philosophy of JMS_Net is similar to the circumstance that two nonspecialist persons work together to complete a task. Although their final results were jointed to yield final results, inevitably their quintessence can not be fully utilized because of their respective learning stage.

In nature, tips from experienced people play a key role on acquiring unknown knowledge for novices. Inspired by this phenomenon, we present multi-scale fusion learning to improve the performance of ConvNets for the LV segmentation. Our multi-scale fusion learning firstly trains different-scale networks (named Net_1) to produce multiple different-scale prediction results. These prediction results are then fused into another multi-scale network (named Net_2), and the fused network is called Net_2'. From the above perspective of acquiring knowledge in nature, the former trained different-scale networks are just like the experienced person and the latter multi-scale network are just like a novice.

Suppose that S-scale networks are trained firstly for Net_1, and the prediction result of the s^{th} trained network is $\boldsymbol{P}_s$. To realize our multi-scale fusion learning, $\boldsymbol{P}_s$ is fused into the s^{th} branch of Net_2. It is worth noting that different branches of Net_2 denote different-scale feature representation. The input of s^{th} branch of Net_2 is denoted as $\boldsymbol{F}_s$, and the input of s^{th} branch of Net_2' is denoted as $\boldsymbol{F}_s'$. Our fusion operation is denotes as Γ, and $\boldsymbol{F}_s'$ can be written as

$$\boldsymbol{F}_s' = \Gamma\{\boldsymbol{P}_s, \boldsymbol{F}_s\}, \tag{1}$$

where Γ denotes the concatenation operation in this paper.

C. SIMO-based Multi-scale Fusion Network

Having considered the fact that during the segmentation result of $\boldsymbol{I}_{ROI}$ is predicted using Net_2', the pooling operation is adopted to obtain larger receptive fields. Therefore, Net_2' includes multiple different-scale feature representation stages since $\boldsymbol{I}_{ROI}$ is pipelined into Net_2'. In this regard, predicting different-scale results directly from $\boldsymbol{I}_{ROI}$ is closer to the prediction way of Net_2'. Consequently, we firstly train single-scale input and multi-scale output (SIMO) networks to predict different-scale results directly from $\boldsymbol{I}_{ROI}$.

As shown in Fig.1, a SIMO-based multi-scale fusion network (SIMO-based MSF_Net) is constructed in this paper. We firstly train S different-scale networks, namely Net_1. The input of each scale network for Net_1 is $\boldsymbol{I}_{ROI}$, and the label of s-scale network are the corresponding-scale label of $\boldsymbol{I}_{ROI}$. Each single-scale network of Net_1 is trained separately so as to acquire S different-scale prediction masks of $\boldsymbol{I}_{ROI}$.

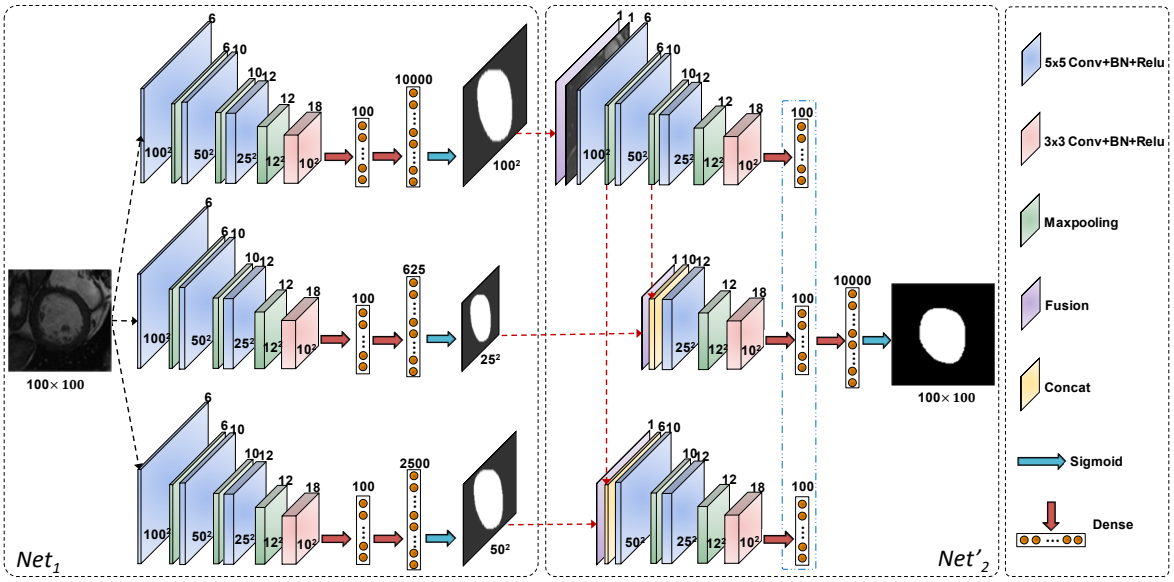

Fig. 1. Overview of our SIMO-based multi-scale fusion network. The different-scale prediction results of Net_1 are fused into different branches of Net'_2.

Suppose that the S-scale prediction results of Net_1 are $\boldsymbol{P}_1, \boldsymbol{P}_2, ..., \boldsymbol{P}_S$. As shown in Fig.1, $\boldsymbol{P}_1, \boldsymbol{P}_2, ..., \boldsymbol{P}_S$ are fused into Net'_2. The input of each branch of Net'_2 comprises the prediction results of Net_1 besides feature maps extracted from front layers. Suppose that the branches of Net'_2 are referred to as B_1, B_2, ..., B_D. Consequently, the input of B_1 is the assembly of $\boldsymbol{P}_1$ and $\boldsymbol{I}_{ROI}$, and the input of the $d^{th}(d = 2, 3, \cdots, D)$ branch B_d is given by

$$L_n^d = G\{\boldsymbol{P}_d, \psi(\sum\nolimits_m (f_m^l * h_m^d) + b_n^d)\}, \tag{2}$$

where $G\{\cdot\}$ denotes that the result is the assembly of elements inside the brace. f_m^l are the feature maps of the layer l at the trunk network of Net'_2 and its number is m. L_n^d denotes the input of the B_d and its number is $n + 1$. h is the convolution kernel. "$*$" denotes the convolution operator. $\psi(\cdot)$ is the activation function and b is the bias scalar.

Suppose that the fully connected layer of B_d is $\boldsymbol{fc}_d$. The fully connected layers of all branches are concatenated as $\boldsymbol{fc}$ which is used as the final prediction,

$$\boldsymbol{fc} = concat(\boldsymbol{fc}_1, \boldsymbol{fc}_2, \cdots, \boldsymbol{fc}_D), \tag{3}$$

where $concat(\cdot)$ denotes the concatenation operation.

During the training stage of our SIMO-based MSF_Net, the mean square error is used as the cost function. For the n^{th} training sample, the cost function J^n is denoted as:

$$J^n = \frac{1}{2} \sum_{i=1}^{m_1} \sum_{j=1}^{n_1} (p_{i,j}^n - y_{i,j}^n)^2, \tag{4}$$

where $m_1 \times n_1$ is the output image size. When $y_{i,j}^n = 1$, the pixel (i, j) belongs to the region of LV. Suppose that the feature vector error of the tail perception is $\boldsymbol{fv}$, and the length of the fully connected layer $\boldsymbol{fc}_d$ is $l_d(d = 1, 2, \cdots, D)$. In the stage of back propagation, the weights of each branch of Net'_2 are updated separately. Suppose that the feature vector error of the B_d is $\boldsymbol{\delta}_d$

$$\boldsymbol{\delta}_d = \Theta(\boldsymbol{\delta}(l_d)) - \Theta(\boldsymbol{\delta}(l_{d-1})), \tag{5}$$

where $\boldsymbol{\delta}(l_d)$ is the feature vector error of the first d branches. $\Theta(\cdot)$ denotes the elements of a vector. The operation "$-$" denotes the removal of the elements on the corresponding position of the second vector from the first vector.

The output of Net'_2 is the final prediction result of our SIMO-based MSF_Net. In the segmentation task, morphological operations including opening, dilation and erosion to eliminate isolated small regions and smooth lesion edges, are usually used [22], [23]. We utilize morphological operations to further improve the accuracy of LV segmentation.

III. Experiments and Results

A. Evaluation Metrics and Experimental Settings

We assess the performance of our multi-scale fusion learning on the MICCAI 2009 LV segmentation challenge database [21]. The segmentation accuracy is evaluated according to the following four measures: 1) percentage of "Good contours" (GC), 2) Dice metric (DM), 3) average perpendicular distance (APD), 4) the conformity coefficient (CC). A high value of APD implies that the manual contour delineated by experts and automatic segmented contour does not match closely. A segmentation result is classified as "Good contours" if its APD is less than 5mm [21]. DM measures contour overlap between the contour areas of automatically segmentation and that of manually segmentation and a higher DM indicates a better match. CC measures the ratio of the number of mis-segmented pixels to the number of correctly segmented pixels [24].

In the experiments, the parameter settings of our MSF_Nets are shown in Fig.1. Our networks are trained using two NVIDA GeForce GTX 1080 Ti GPUs, developed on a 64-bit

TABLE I
COMPARISON OF OUR SIMO-BASED MSF_NET AND TWO OTHER MODELS (MIMO-BASED MSF_NET AND SISO-BASED MSF_NET) ON VALIDATION AND ONLINE DATASETS ON MICCAI 2009 DATABASE [21]. (NUMBERS' FORMAT: MEAN VALUE (STANDARD DEVIATION))

Model	Dataset	Good contours(%)	APD(mm)	DM	CC
MIMO-based MSF_Net	Val	96.99(6.14)	1.981(0.17)	0.925(0.02)	0.838(0.05)
SISO-based MSF_Net		96.99(5.52)	1.989(0.19)	0.920(0.03)	0.824(0.07)
SIMO-based MSF_Net		**97.74(4.48)**	**1.959(0.19)**	**0.927(0.02)**	**0.841(0.06)**
MIMO-based MSF_Net	Onl	**99.28(2.08)**	2.067(0.41)	0.918(0.03)	0.821(0.08)
SISO-based MSF_Net		98.92(3.59)	2.035(0.37)	0.918(0.03)	0.820(0.07)
SIMO-based MSF_Net		98.92(2.44)	**2.041(0.40)**	**0.919(0.03)**	**0.824(0.07)**
MIMO-based MSF_Net	Val+Onl	98.17(4.06)	2.025(0.29)	0.921(0.03)	0.829(0.07)
SISO-based MSF_Net		97.98(4.53)	2.013(0.28)	0.919(0.03)	0.822(0.08)
SIMO-based MSF_Net		**98.35(3.44)**	**2.000(0.29)**	**0.923(0.03)**	**0.832(0.06)**

ubuntu 16.04 platform with Intel®Core™ i5-7640X CPU @ 4.00GHZ×4, 32GB memory (RAM).

B. Data Augmentation Results

The training dataset is firstly artificially augmented to improve the accuracy of ROI detection and LV segmentation using our data augmentation method descried in section II-A. In this experiment, we set $m_w \times n_w = 100 \times 100$, $\rho = 15$, and $d = \omega\pi/4(\omega = 0, 1, \cdots, 7)$. In addition, we increase the amount of training set using the following two methods: resize the raw image to 100×100; crop a 100×100 region from the raw image. Thus the training set can be extra enlarged by ten times. To evaluate the performance of our data augmentation method and ascertain the parameters of data augmentation, we train six ConvNet models with the same architecture using six training sets augmented by different times.

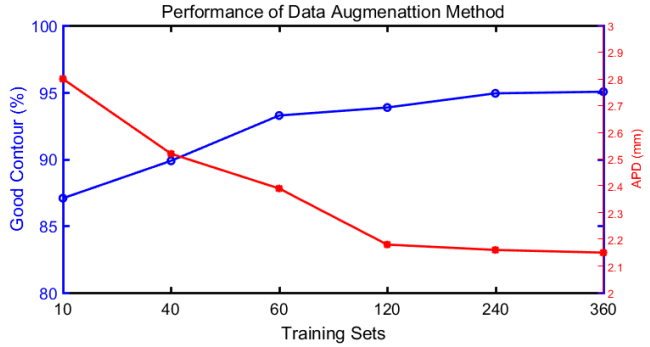

Fig. 2. Comparison results of our data augmentation of different multiples.

Fig.2 depicts the changing trend of the metrics of "Good contours" and APD as the amount of training set increases. The right vertical axis and left vertical axis of Fig.2 denote the values of "Good contours" and APD, respectively. The horizontal axis denotes six different training sets (the values of α from left to right are $0°, 90°, 60°, 30°, 15°$ and $10°$, and the corresponding multiples of data augmentation are 10, 40, 60, 120, 240 and 360, respectively). It is obvious that the two metrics get better and better as the amount of the training data increases, and the results of $\alpha = 15°$ are very close to those of $\alpha = 10°$. Therefore, we augment our training set by the factor of 240, which is used to evaluate the performance of our multi-scale fusion networks.

C. Quantitative Results of SIMO-based MSF_Net

The performance of our SIMO-based MSF_Net is compared with two other models (MIMO-based MSF_Net and SISO-based MSF_Net). It is worth noting that MIMO-based MSF_Net firstly trains multi-scale input and multi-scale output networks to produce different-scale results for the second multi-scale network, and SISO-based MSF_Net firstly trains one network to produce different-scale results by the down-sampling operation. The average value and standard deviation of the evaluation metrics (GC, APD, DM and CC) for aforementioned three models are calculated for the validation (Val) and online (Onl) datasets of MICCAI 2009 challenge database [21] and their comparison results are listed in Table I.

From Table I, we can see that the proposed SIMO-based MSF_Net outperforms MIMO-based MSF_Net and SISO-based MSF_Net on all metrics for the validation dataset. For the online dataset, MIMO-based MSF_Net achieves a higher GC than SIMO-based MSF_Net, and SISO-based MSF_Net achieves a higher APD. However, judging from the metric GC of the two testing datasets, our SIMO-based MSF_Net is more robust. What's more, from the row "Val+Onl" of Table I, we can see that all the metrics of our SIMO-based MSF_Net are better than that of MIMO_Net and SISO-based MSF_Net when considering two testing datasets, demonstrating that our SIMO-based MSF_Net is more competitive.

D. Performance of Fusing Different Numbers of Multi-scale Prediction Results

To further evaluate the performance of our multi-scale fusion learning, we fuse different numbers of multi-scale predictions results into MIMO-based MSF_Net and SIMO-based MSF_Net. The results are shown in Table II. The architecture of "Fundamental network" is the same as Net_2' with no multi-scale fusion. F100, F100+F50 and F100+F50+F25 denotes that our MSF_Nets fuse single-scale, two-scale and three-scale prediction results, respectively. From Table II, we can see that multi-scale fusion networks performs better and better as the number of multi-scale prediction results increases for both MIMO-based MSF_Nets and SIMO-based MSF_Nets. In

TABLE II
PERFORMANCE OF MULTI-SCALE FUSION NETWORKS FUSING DIFFERENT NUMBERS OF MULTI-SCALE PREDICTION RESULTS.

Model	Good contours(%)	APD(mm)	DM	CC
Fundamental network	97.07(5.62)	2.093(0.38)	0.918(0.03)	0.818(0.08)
MIMO-based MSF_Net-F100	97.61(4.91)	2.056(0.31)	0.917(0.03)	0.819(0.09)
MIMO-based MSF_Net-F100+F50	98.17(4.36)	2.036(0.30)	0.919(0.03)	0.824(0.07)
MIMO-based MSF_Net-F100+F50+F25	**98.17(4.06)**	**2.025(0.29)**	**0.921(0.03)**	**0.829(0.07)**
Fundamental network	97.07(5.62)	2.093(0.38)	0.918(0.03)	0.818(0.08)
SIMO-based MSF_Net-F100	97.61(4.91)	2.056(0.31)	0.917(0.03)	0.819(0.09)
SIMO-based MSF_Net-F100+F50	97.98(4.66)	2.017(0.30)	0.918(0.03)	0.822(0.08)
SIMO-based MSF_Net-F100+F50+F25	**98.35(3.44)**	**2.000(0.29)**	**0.923(0.03)**	**0.832(0.06)**

addition, compared with "Fundamental network", our three-scale SIMO-based MSF_Net has significant improvement on all metrics, especially for "Good contours" and APD, which improves 1.28% and reduces 0.093mm, respectively.

E. Comparison with the State of the Art

We compare our multi-scale fusion learning with the start-of-the-art methods for the LV segmentation challenge on MICCAI 2009 database [21] and the computed metrics are summarized in Table III. The comparison results reveal that our method achieves a significant improvement on "Good contours". The other three metrics of our SIMO-based MSF_Net are not as good as that of [3] which combines deep learning and deformable model. However, the performance of the deep learning method ("Good contours" = 88.46%, APD = 2.90mm, DM = 0.89, and CC = 0.77) in [3] is far behind that of our method. Our SIMO-based MSF_Net achieves 98.35% on "Good contours", showing it can detect more good contours including small-contour images. However, small-contour images always have blurred boundary, which leads to our high APD, DM and CC.

TABLE III
COMPARISON BETWEEN OUR SIMO-BASED MSF_NET AND STATE-OF-THE-ART METHODS ON MICCAI 2009 DATABASE [21].

Method	Dataset	GC (%)	APD(mm)	DM	CC
Jolly [25]	30	95.62(8.8)	2.26(0.59)	0.88(0.04)	0.73
Huang et al. [26]	45	79.2(19.0)	2.16(0.46)	0.89(0.04)	0.75
Liu et al. [27]	45	91.17(8.5)	2.36(0.39)	0.88(0.03)	0.73
Hu et al. [28]	45	91.06(9.42)	2.24(0.40)	0.89(0.03)	0.75
Ngo et al. [11]	30	93.23(9.84)	2.26(0.46)	0.89(0.03)	0.75
Avendi et al. [3]	30	96.69(5.70)	**1.81(0.44)**	**0.94(0.02)**	**0.86**
Ngo et al. [12]	45	94.55(9.31)	2.22(0.46)	0.88(0.02)	0.73
JMS_Net [20]	30	96.80(5.91)	2.08(0.53)	0.91(0.03)	0.81
Proposed	30	**98.35(3.44)**	2.00(0.29)	0.92(0.03)	0.83

F. Segmentation Results of Our SIMO-based MSF_Net

Fig.3 illustrates the automatic LV segmentation results produced by our SIMO-based MSF_Net and JMS_Net [20] for challenging CMR images described in Section I. CMR images of the first row posses TPM, and CMR images of the second row posses LVOT. Besides, the third row shows small-contour images. The adjacent columns of "Proposed" and "[20]" denotes segmentation results of the same CMR image using our SIMO-based MSF_Net and JMS_Net. We can see that the overall segmentation performance of our SIMO-based MSF_Net is better than JMS_Net. As shown in the third row of Fig.3, small-contour images always have blurred boundary, which makes the segmentation of small-contour images more challenging. It is worth noting that some small-contour images can not be accurately detected by JMS_Net, such as the fourth image in the third row. Although the automatic segmentation contours of our SIMO-based MSF_Net are not very close to the manual annotation contours for such challenging images, it can detect the rough area of the LV, demonstrating its competitive ability of multi-scale feature representation.

IV. CONCLUSION

In this paper, the multi-scale fusion learning is proposed for the automatic LV segmentation, which fuses different-scale prediction results of trained multi-scale networks into another multi-scale network thus to extract more competitive features. To advance the performance of our multi-scale fusion learning, we construct a SIMO-based MSF_Net, which firstly trains single-scale input and multi-scale output networks to produce different-scale prediction results. The proposed method has been evaluated on the MICCAI database 2009 [21]. It is worth noting that although our SIMO-based MSF_Net ranks the first on "Good contours", our APD, DM and CC all rank the second in Table III. This phenomenon results from the introduction of bad APD, DM and CC of small-contour images whose contour is difficult to be accurately segmented because of the blurred boundary. In the future work, we will focus on how to further refine the automatic segmentation results, especially for images of blurred boundary.

ACKNOWLEDGMENT

The work was supported by the National Natural Science Foundation of China under Grant No. 61373107, Wuhan Science and Technology Program under Grant No. 2016010101010022, SIAT Innovation Program for Excellent Young Researchers under Grant No. 2017059 and China Postdoctoral Science Foundation under Grant No. 2017M622831. Qianqian Tong and Jianhui Zhao are both the corresponding authors of this paper.

REFERENCES

[1] W. H. organization (WHO), "Cardiovascular diseases (cvds)," http://www.who.int/mediacentre/factsheets/fs317/en/.

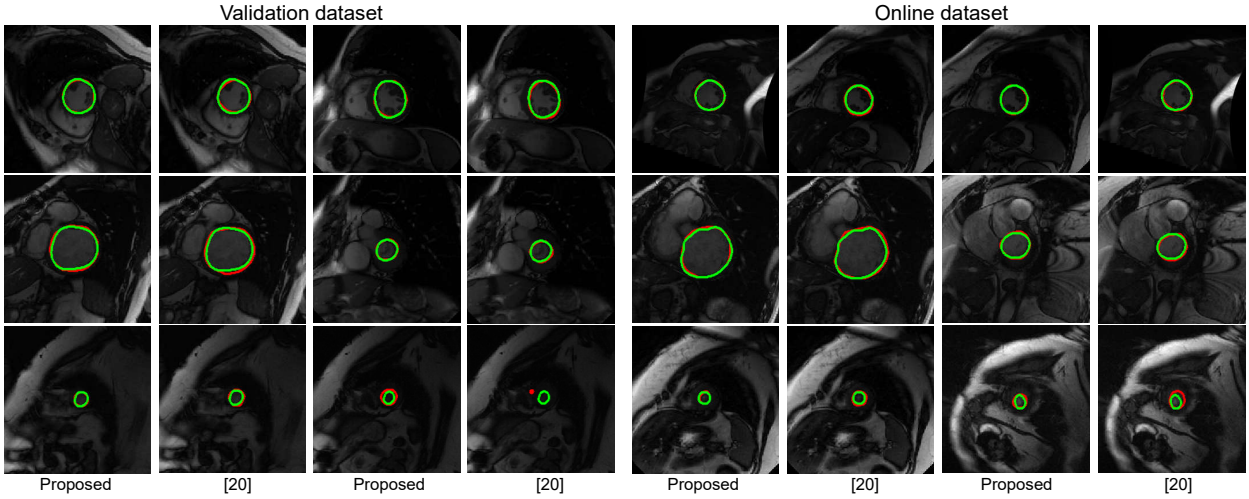

Fig. 3. Automatic segmentation results of the proposed SIMO-based MSF_Net and JMS_Net [20] for challenging images described in Section I. The green curve denotes ground truth, and the red curve denotes automated segmentation of our SIMO-based MSF_Net and JMS_Net.

[2] H. Chalian, J. K. ODonnell, M. Bolen, and P. Rajiah, "Incremental value of pet and mri in the evaluation of cardiovascular abnormalities," *Insights into imaging*, vol. 7, no. 4, pp. 485–503, 2016.

[3] M. Avendi, A. Kheradvar, and H. Jafarkhani, "A combined deep-learning and deformable-model approach to fully automatic segmentation of the left ventricle in cardiac mri," *Medical image analysis*, vol. 30, pp. 108–119, 2016.

[4] C. Yang, W. Wu, Y. Su, and S. Zhang, "Left ventricle segmentation via two-layer level sets with circular shape constraint," *Magnetic resonance imaging*, vol. 38, pp. 202–213, 2017.

[5] M. Kass, A. Witkin, and D. Terzopoulos, "Snakes: Active contour models," *International journal of computer vision*, vol. 1, no. 4, pp. 321–331, 1988.

[6] T. F. Chan and L. A. Vese, "Active contours without edges," *IEEE Transactions on image processing*, vol. 10, no. 2, pp. 266–277, 2001.

[7] T. A. Ngo and G. Carneiro, "Left ventricle segmentation from cardiac mri combining level set methods with deep belief networks," in *Image Processing (ICIP), 2013 20th IEEE International Conference on*. IEEE, 2013, pp. 695–699.

[8] T. Chen, J. Babb, P. Kellman, L. Axel, and D. Kim, "Semiautomated segmentation of myocardial contours for fast strain analysis in cine displacement-encoded mri," *IEEE Transactions on Medical Imaging*, vol. 27, no. 8, pp. 1084–1094, 2008.

[9] V.-T. Pham and T.-T. Tran, "Active contour model and nonlinear shape priors with application to left ventricle segmentation in cardiac mr images," *Optik-International Journal for Light and Electron Optics*, vol. 127, no. 3, pp. 991–1002, 2016.

[10] M. Hajiaghayi, E. M. Groves, H. Jafarkhani, and A. Kheradvar, "A 3-d active contour method for automated segmentation of the left ventricle from magnetic resonance images," *IEEE Transactions on Biomedical Engineering*, vol. 64, no. 1, pp. 134–144, 2017.

[11] T. Anh Ngo and G. Carneiro, "Fully automated non-rigid segmentation with distance regularized level set evolution initialized and constrained by deep-structured inference," in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, 2014, pp. 3118–3125.

[12] 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.

[13] D. Cobzas and M. Schmidt, "Increased discrimination in level set methods with embedded conditional random fields," in *Computer Vision and Pattern Recognition, 2009. CVPR 2009. IEEE Conference on*. IEEE, 2009, pp. 328–335.

[14] J. F. Dreijer, B. M. Herbst, and J. A. Du Preez, "Left ventricular segmentation from mri datasets with edge modelling conditional random fields," *BMC medical imaging*, vol. 13, no. 1, p. 24, 2013.

with deep convolutional neural networks," in *Advances in neural information processing systems*, 2012, pp. 1097–1105.

[15] A. Krizhevsky, I. Sutskever, and G. E. Hinton, "Imagenet classification

[16] C. Farabet, C. Couprie, L. Najman, and Y. LeCun, "Learning hierarchical features for scene labeling," *IEEE transactions on pattern analysis and machine intelligence*, vol. 35, no. 8, pp. 1915–1929, 2013.

[17] Y. LeCun, Y. Bengio, and G. Hinton, "Deep learning," *Nature*, vol. 521, no. 7553, pp. 436–444, 2015.

[18] H. Dou and X. Wu, "Coarse-to-fine trained multi-scale convolutional neural networks for image classification," in *Neural Networks (IJCNN), 2015 International Joint Conference on*. IEEE, 2015, pp. 1–7.

[19] Z. Cai, Q. Fan, R. S. Feris, and N. Vasconcelos, "A unified multi-scale deep convolutional neural network for fast object detection," in *European Conference on Computer Vision*. Springer, 2016, pp. 354–370.

[20] Q. Tong, Z. Yuan, X. Liao, M. Zheng, W. Zhu, G. Zhang, and M. Ning, "A joint multi-scale convolutional network for fully automatic segmentation of the left ventricle," in *Image Processing (ICIP), 2017 IEEE International Conference on*. IEEE, 2017, pp. 3110–3114.

[21] P. Radau, Y. Lu, K. Connelly, G. Paul, A. Dick *et al.*, "Evaluation framework for algorithms segmenting short axis cardiac mri," *The MIDAS Journal-Cardiac MR Left Ventricle Segmentation Challenge*, vol. 49, 2009.

[22] Y. Wang, A. K. Katsaggelos, X. Wang, and T. B. Parrish, "A deep symmetry convnet for stroke lesion segmentation," in *Image Processing (ICIP), 2016 IEEE International Conference on*. IEEE, 2016, pp. 111–115.

[23] M. H. Jafari, N. Karimi, E. Nasr-Esfahani, S. Samavi, S. M. R. Soroushmehr, K. Ward, and K. Najarian, "Skin lesion segmentation in clinical images using deep learning," in *International Conference on Pattern Recognition*, 2017.

[24] H.-H. Chang, A. H. Zhuang, D. J. Valentino, and W.-C. Chu, "Performance measure characterization for evaluating neuroimage segmentation algorithms," *Neuroimage*, vol. 47, no. 1, pp. 122–135, 2009.

[25] M. Jolly, "Fully automatic left ventricle segmentation in cardiac cine mr images using registration and minimum surfaces," *The MIDAS Journal-Cardiac MR Left Ventricle Segmentation Challenge*, vol. 4, 2009.

[26] S. Huang, J. Liu, L. C. Lee, S. K. Venkatesh, L. L. San Teo, C. Au, and W. L. Nowinski, "An image-based comprehensive approach for automatic segmentation of left ventricle from cardiac short axis cine mr images," *Journal of digital imaging*, vol. 24, no. 4, pp. 598–608, 2011.

[27] H. Liu, H. Hu, X. Xu, and E. Song, "Automatic left ventricle segmentation in cardiac mri using topological stable-state thresholding and region restricted dynamic programming," *Academic radiology*, vol. 19, no. 6, pp. 723–731, 2012.

[28] H. Hu, H. Liu, Z. Gao, and L. Huang, "Hybrid segmentation of left ventricle in cardiac mri using gaussian-mixture model and region restricted dynamic programming," *Magnetic resonance imaging*, vol. 31, no. 4, pp. 575–584, 2013.