# Cascaded Conditional Generative Adversarial Networks With Multi-Scale Attention Fusion for Automated Bi-Ventricle Segmentation in Cardiac MRI

**LIN QI**[1, 2]**, (Member, IEEE), HAORAN ZHANG**[1]**, WENJUN TAN**[3]**, SHOULIANG QI**[1, 2]**,
LISHENG XU**[1, 2, 3]**, (Senior Member, IEEE), YUDONG YAO**[4]**, (Fellow, IEEE),
AND WEI QIAN**[5]

[1]College of Medicine and Biological Information Engineering, Northeastern University, Shenyang 110169, China
[2]Engineering Research Center of Medical Imaging and Intelligent Analysis, Ministry of Education, Shenyang 110169, China
[3]Key Laboratory of Medical Image Computing, Ministry of Education, Northeastern University, Shenyang 110169, China
[4]Electrical and Computer Engineering Department, Stevens Institute of Technology, Hoboken, NJ 07030, USA
[5]Department of Electrical and Computer Engineering, College of Engineering, The University of Texas at El Paso, El Paso, TX 79968, USA

Corresponding author: Lisheng Xu (xuls@bmie.neu.edu.cn)

**ABSTRACT** Accurate segmentation of bi-ventricle from cardiac magnetic resonance images (MRI) is a critical step in cardiac function analysis and disease diagnosis. Due to the morphological diversification of the heart and the factors of MRI itself, fully automated and concurrent bi-ventricle segmentation is a well-known challenge. In this paper, we propose cascaded conditional generative adversarial networks (C-cGANs) to divide the problem into two segmentation subtasks: binary segmentation for region of interest (ROI) extraction and bi-ventricle segmentation. In both subtasks, we adopt adversarial training that makes discriminator network to discriminate segmentation maps either from generator network or ground-truth which aims to detect and correct pixel-wise inconsistency between the sources of segmentation maps. For capturing more spatial information with multi-scale semantic features, in the generator network, we insert a multi-scale attention fusion (MSAF) module between the encoder and decoder paths. The experiment on ACDC 2017 dataset shows that the proposed model outperforms other state-of-the-art methods in most metrics. Moreover, we validate the generalization capability of this model on MS-CMRSeg 2019 and RVSC 2012 datasets without fine-tuning, and the results demonstrate the effectiveness and robustness of the proposed method for bi-ventricle segmentation.

**INDEX TERMS** Bi-ventricle segmentation, ROI extraction, cascaded conditional generative adversarial networks (C-cGANs), MSAF module.

## I. INTRODUCTION

The World Health Organization reports that cardiovascular diseases (CVDs) with a high incidence have been the leading cause of death worldwide [1]. Cardiac image analysis plays an important role in the diagnosis and treatment of CVDs. As is known to all, cardiac magnetic resonance images (MRI) are considered as the most accurate method

The associate editor coordinating the review of this manuscript and approving it for publication was Md. Kamrul Hasan.

to estimate clinical indicators, such as ventricular volume, ejection fraction, myocardial mass, etc. [2]. The accurate and reliable method for calculating these indicators is to delineate the contours of blood pool of left ventricle (LV), left ventricular epicardium (or myocardium, MYO) and blood pool of right ventricle (RV) in end-diastole (ED) and end-systole (ES) phases which is also called bi-ventricle segmentation. However, manual segmentation is not only cumbersome and time consuming, but also prone to intra- and inter-observer variability [3]. Therefore, accurate and automated algorithms
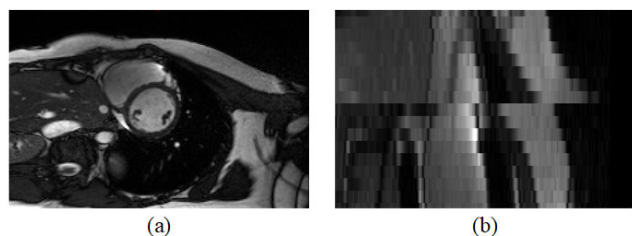
**FIGURE 1.** (a) Short-axis slice of cardiac MRI. (b) Cross section obtained by reconstructing multiple slices along the short axis. Note: Slice misalignment is due to different breath-holds during successive image acquisitions.

for bi-ventricular segmentation are urgently needed in clinical practice.

The difficulties and challenges of heart segmentation have been clearly demonstrated in previous studies [4]–[6]. Among them, as shown in Fig. 1 (a), the following points are more notable: 1) local poor contrast between heart and surrounding structures especially in apex and base; 2) similar pixel intensity in papillary muscles and MYO; 3) inherent noise due to motion artifacts; 4) diversity of cardiac morphology and pixel intensity especially in RV; 5) imbalance of area ratio of the heart and the background. Some previous works have overcome partial challenges on the segmentation of the single-ventricle (more on LV rather than on RV) [6]–[11]. Although the bi-ventricle contours can be obtained by putting the results of a single-ventricle segmentation model together in previous research [12], this practice cannot learn the related anatomical information of the bi-ventricle, ignoring the morphological correlations and resulting in the accumulation of errors. Therefore, concurrent segmentation of the two ventricles makes more sense and becomes a challenging task.

Over the past decades, some conventional methods that performed well in single-ventricle segmentation have also been tried for bi-ventricle. These can be roughly divided into strong prior and weak prior methods. Specifically, strong prior methods include: 1) active shape and appearance models [13], [14] whose results are influenced by the number of cardiac MRI images; 2) atlas-based methods [15] using atlas as frame of reference. Due to anatomical variability, it is difficult to accurately segment RV with variable shapes. Weak prior methods include: 1) pixel-based and image-based methods such as Threshold [16] and clustering [17]. However, they cannot distinguish between the myocardium and the blood pool leading to failure when the two ventricles and adjacent structures exhibit similar intensity distributions; 2) deformable model methods such as active contour [18] and level-set [19]. They are sensitive to diverse boundary variations due to pathologies and behave badly to local weak and no boundary due to noise and artifacts, as well as with high computational cost. In general, strong prior methods excessively rely on prior information such as handcrafted features which have limited expressive power for rare pathological cases. On the other hand, weak prior methods have relatively low accuracy and robustness so as to need more

interactive operation. Automated bi-ventricle segmentation still has more room for improvement.

Deep learning has been applied to various pattern recognition and computer vision tasks in recent years. In particular, convolutional neural networks (CNNs) have achieved great success in the field of medical image segmentation compared to traditional methods because of its stronger hierarchical learning and representation capabilities. CNN derive representative variants such as 2D U-Net [20] and 3D U-Net [21], which guides the new development direction for medical image segmentation. Up to now, researchers have developed many automated deep learning methods in medical image segmentation which are divided into three categories: 2D (a slice as unit), 3D (a volume as unit) and combination of 2D and 3D (multiple slices as unit). 2D methods have a wide range of applications because they have fewer parameters and require less training data. However, they often perform fail on some special slices due to the lack of contextual information within slices. For example, Zotti *et al.* [22] used CNN-based model with shape prior to segment bi-ventricle in every slice, and the results showed that they performed well on most slices, but struggled on apical and basal slices. 3D methods have high accuracy on segmentation of some organs with global context relation [23]–[25]. However, in the study [26], the bi-ventricle segmentation result of 2D CNN is better than that of 3D CNN. This indicates that slice thickness has a great influence on 3D convolution (cardiac MRI slices thickness is larger than the slice resolution, usually 5∼10 mm). Moreover, more training data and larger GPU memory space are required for 3D segmentation. Furthermore, some studies used the combination of 2D and 3D methods which absorbs the advantages of both. Du *et al.* [27] combined CNN with a recurrent unit (Long Short-Term Memory, LSTM) to segment all slices in one sequence from the base to apex. The information of all slices was learned as spatial context to segment the current slice. As shown in Fig. 1 (b), slice misalignment usually occurs due to different breath-holds during successive acquisitions, thus causing slices to propagate erroneous information. On the other hand, large slice thickness can also reduce the correlation between slices and even fail to segment.

In addition to the above problems, an important concern in the application of deep learning is the number of medical images for training. Although Bai *et al.* [28] and Zheng *et al.* [29] have already tried to use the UK Biobank [30] dataset which is the largest cardiac dataset including more than 3000 cases for 2D bi-ventricle segmentation, this dataset is not open to researchers. They also tried to validate generalization capability in other datasets using trained models on large datasets, but the results didn't meet expectations. Fortunately, in 2017 MICCAI, Automated Cardiac Diagnosis Challenge (ACDC 2017) provides a certain number of cardiac MRI to measure the performance of the most advanced bi-ventricle segmentation methods and to find out 'the reasons for the failure of the segmentation'. Therefore, the current goal is probably to find the well-turned

method on a limited amount of data to achieve the best results and good generalization capability at the same level with other methods, rather than simply relying on increase of the amount of data to maintain and improve accuracy.

Through the above analysis, 2D slice segmentation method based on deep learning is still relatively suitable to identify bi-ventricle. In order to improve the performance of cardiac MRI segmentation under a limited amount of data, we propose a novel 2D method which overcomes the bottleneck of current 2D segmentation. The proposed work is to detect cardiac region of interest (ROI) and intently segment the cardiac bi-ventricle through adversarial training, inspired by generative adversarial networks (GANs) [31] in the field of medical image analysis in recent years [32-34]. A module based on attention mechanism is introduced to enhance the feature representations of the network. Specifically, the main contributions of this work can be summarized as follows:

- We propose cascaded conditional generative adversarial networks (C-cGANs) to detect cardiac ROI and segment bi-ventricle. The ROI obtained by the 1st cGAN (C1-cGAN) is used as the input of the 2nd cGAN (C2-cGAN) to eliminate the interference from other organs and tissues and thereby reduce false positives. Adversarial training alternately improves the segmentation and discrimination capabilities under labeled data. Accurate pixel-wise classification is only yielded by the generator network, without increasing complexity to the model used in the testing phase.
- We insert the multi-scale attention fusion (MSAF) module into encoder-decoder architecture as the generative network to generate high-level attention maps at different scales. The module consists of two blocks: position attention mechanism (PAM) block for paying more attention to the contribution of the foreground and weakening the influence of the background in the image and multi-scale dilated fusion MDF) block for fusing attention-based maps with different receptive fields.
- We train and test our proposed method on ACDC 2017 and the testing results show that we achieve competitive segmentation compared to other the state-of-the-art approaches for bi-ventricle segmentation. In addition, we validate generalization capability through testing on two open cardiac datasets. The results of experiments demonstrate the possibility of accurate bi-ventricle segmentation.

The rest of this work is listed as follows: the proposed method is presented in Section II. The experiments and results are shown in Section III. At last, Section IV states the discussion and conclusion.

## II. METHOD

### A. OVERVIEW

The ultimate goal of our work is to concurrently segment bi-ventricle objectives (LV/RV/MYO) from $X$ short-axis slices separated by 3D $H \times W \times X$ raw cine MR images volume. We propose a C-cGANs model to divide complex
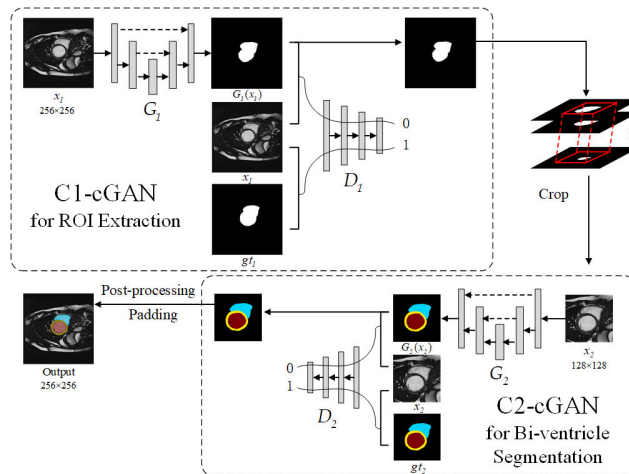


**FIGURE 2.** Flowchart of the proposed C-cGANs model.

segmentation task into cascaded two segmentation subtasks: foreground segmentation for cardiac ROI extraction and multi-objective segmentation for bi-ventricle. The cascade operation divides complex problems into two simple ones, making the model easy to converge by sequentially training. Cascaded architecture can also eliminate the interference from other organs and tissues and thereby reduce false positives for accurate segmentation. As shown in Fig. 2, The ROI extraction subtask aims to obtain the cardiac region size-cropped $x_1$ to increase the attention of the cardiac architecture. It is considered as binary classification (background and cardiac morphology) essentially. First, the input image with a certain size ($256 \times 256$) is obtained from $H \times W$ raw slices by pre-processing. The trained ROI extraction model (C1-cGAN) outputs the predicted binary mask $G_1(x_1)$ for $x_1$. To ensure that each ROI image has the same size and contains the corresponding entire heart structure, we crop each image $x_1$ directly at the center of the largest 3D connected domain based on the binary segmented result of cine MR images volume. At the end, $x_2$ is fed to C2-cGAN model to produce accurate contours of LV, RV and MYO concurrently. We consider that two network models are trained separately in turns for avoiding deeper network layers so that they can reduce training difficulty and avoid over-fitting.

In the next, we present C-cGANs model for bi-ventricle pixel-wise segmentation in section II.B. The core network architectures of generator and discriminator are introduced in section II.C. The proposed adversarial training steps are described in section II.D.

### B. CASCADED CONDITIONAL GENERATIVE ADVERSARIAL NETWORKS

#### 1) CONDITIONAL GENERATIVE ADVERSARIAL NETWORK

GAN has a wide range of research and applications in the field of image and vision. The conventional GAN [31] consists of two parts: a generator network $D$ and a discriminator network $D$. $G$ is mainly responsible for generating

samples $G(z)$ obeying the distribution of real data $x$ through random noise $z$. $D$ is responsible for identifying the input data samples from the generator or ground-truth. Both of them are iteratively optimized to improve their respective performances. Until $D$ cannot discriminate the source of input data, we assume that $G$ has learned the distribution of ground-truth. The GAN is trained in a competitive manner which is like a two-player mini-max game. Thus the value function $V(D, G)$ is defined as:

$$\min_{G} \max_{D} V(D, G)$$
$$= E_x [logD(x)] + E_z [log(1 - D(G(z)))] \quad (1)$$

However, the conventional GAN infers generative samples only from random noise, thus its loss is not easy to converge and it will require more training iterations to obtain acceptable segmentation results than conditional generative adversarial network (cGAN) [35].

In cGAN, the generator which takes auxiliary information $y$ and random vector $z$ as input learns mapping from joint hidden layer characteristics to probability distribution of real data $x$. Moreover, $D$ tries to discriminate real data $x$ or generative data $G(z \mid y)$. Similar to GAN, the objective function of cGAN is a two-player mini-max game with conditional probability. The value function $V(D, G)$ is defined like this:

$$\min_{G} \max_{D} V(D, G)$$
$$= E_x [logD(x \mid y)] + E_z [log(1 - D(G(z \mid y)))] \quad (2)$$

### 2) CASCADED CONDITIONAL GENERATIVE ADVERSARIAL NETWORKS

Luc *et al.* [36] first applied the adversarial training for semantic segmentation, they constructed a generator as generative network whose input is the image to be segmented. Based on previous researches, we propose cascaded cGANs (C-cGANs) for cardiac ROI extraction and accurate pixel-wise bi-ventricle segmentation.

The proposed C-cGANs is a cascade-connected architecture which consists of C1-cGAN and C2-cGAN. C1-cGAN is followed by C2-cGAN. In each cGAN, $G$ takes MR image as input to generate segmentation mask by feature extraction capability of convolution network. Concatenation of the generative mask and corresponding MR image (or concatenation of ground-truth mask and MR image) are input to $D$ for discriminating. Until the $D$ fails to discriminate the input source, the adversarial training can be assumed to achieve a Nash equilibrium between $G$ and $D$. More specifically, for the C1-cGAN, image $x_1$ with size of $256 \times 256$ is taken as condition information to $G_1$ and $D_1$. $G_1$ is trained by $x_1$ to produce predicted binary mask $G_1(x_1)$ which is supposed to close to ground-truth binary mask $gt_1$. Trained $D_1$ for input $D_1(x_1, gt_1)$ tends to output 1 (real) and for $D_1(x_1, G_1(x_1))$ tends to output 0 (fake). Nevertheless, $G_1$ tries to control $D_1$ to output 1. The optimization goal of $G_1$ will minimize the objective function and $D_1$ is opposite. The objective function

for C1-cGAN is defined as:

$$\mathcal{L}_{adv}(D_1, G_1) \leftarrow \min_{G} \max_{D} V(D_1, G_1)$$
$$= E_{x_1, gt_1} [logD_1(x_1, gt_1)]$$
$$+ E_{x_1} [log(1 - D_1(x_1, G_1(x_1)))] \quad (3)$$

Similar to C1-cGAN, $x_2$ denotes ROI image, $gt_2$ represents the ground-truth of bi-ventricle mask in C2-cGAN. We define similar formulations for C2-cGAN:

$$\mathcal{L}_{adv}(D_2, G_2) \leftarrow \min_{G} \max_{D} V(D_2, G_2)$$
$$= E_{x_2, gt_2} [logD_2(x_2, gt_2)]$$
$$+ E_{x_2} [log(1 - D_2(x_2, G_2(x_2)))] \quad (4)$$

Beyond the adversarial loss, we mix the cross entropy with weighted Dice coefficient (WDC) as generative loss. The general Dice coefficient loss [37] is disadvantageous to the small target in segmentation. Part of the prediction error of the small target will lead to a large change of Dice, resulting in drastic gradient change and unstable training. In view of the small proportion of each category in cardiac MRI, we propose WDC loss as:

$$\mathcal{L}_{WDC} = \sum_{k=1}^{K} \lambda_k \cdot \left(1 - \frac{2 \cdot \sum_{n=1}^{N} g_{kn} \cdot p_{kn}}{\sum_{n=1}^{N} (g_{kn} + p_{kn}) + \varepsilon}\right) \quad (5)$$

where $K$ is the number of classes, $N$ is the number of total pixels in the image, $g_{kn}$ represents ground-truth of class $k$ in the $n^{th}$ pixel (0 or 1), and $p_{kn}$ represents corresponding predicted value (from 0 to 1), $\varepsilon$ equals to $10^{-5}$ for numerical stability, $\lambda_k$ denotes the weighted factor of class $k$ which varies with the number of class of pixels and defined as:

$$\lambda_k = 1 - \frac{\sum_{n=1}^{N} g_{kn}}{\sum_{k=1}^{K} \sum_{n=1}^{N} g_{kn}} \quad (6)$$

The WDC loss improves the segmentation contour, while the cross entropy loss optimizes each class of accuracy on pixel-level. For two tasks with different segmentation objectives, the binary cross entropy (BCE) and categorical cross entropy (CCE) are used for $G_1$ and $G_2$. We get respective generative losses:

$$\mathcal{L}_{gen}(G_1) = \underbrace{\sum_{k=1}^{K_1} \lambda_k \cdot \left(1 - \frac{2 \cdot \sum_{n=1}^{N_1} g_{kn} \cdot p_{kn}}{\sum_{n=1}^{N_1} (g_{kn} + p_{kn}) + \varepsilon}\right)}_{\mathcal{L}_{WDC}}$$
$$+ \underbrace{\left(-\frac{1}{N_1} \left(\sum_{n=1}^{N_1} \sum_{k=1}^{K_1} g_{kn} \log(p_{kn})\right)\right)}_{\mathcal{L}_{BCE}} \quad (7)$$

where $K_1 = 2$ (2 classes: heart and background), and $N_1 = H \times W = 256 \times 256$ which is the number of total pixels of original image;

$$\mathcal{L}_{gen}(G_2) = \underbrace{\sum_{k=1}^{K_2} \lambda_k \cdot \left(1 - \frac{2 \cdot \sum_{n=1}^{N_2} g_{kn} \cdot p_{kn}}{\sum_{n=1}^{N_2} (g_{kn} + p_{kn}) + \varepsilon}\right)}_{\mathcal{L}_{WDC}}$$
$$+ \underbrace{\left(-\frac{1}{N_2} \left(\sum_{n=1}^{N_2} \sum_{k=1}^{K_2} g_{kn} \log(p_{kn})\right)\right)}_{\mathcal{L}_{CCE}} \quad (8)$$
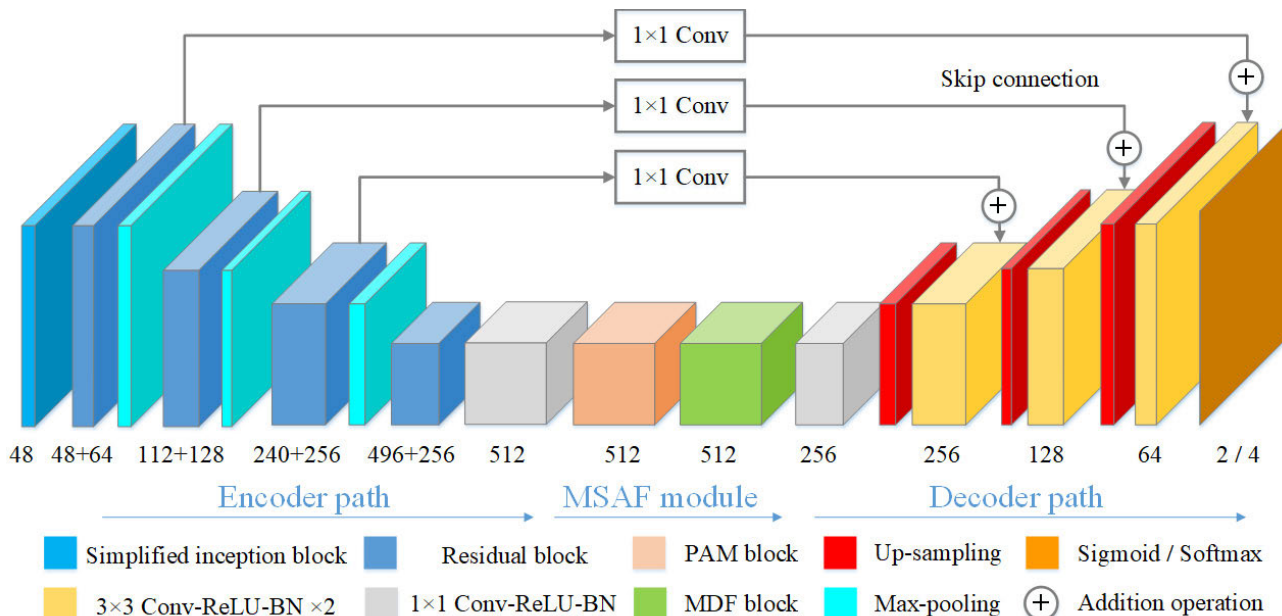
**FIGURE 3.** An illustration of proposed generator network. The input images are followed by the encoder path, MSAF module and decoder path to acquire the corresponding contours. The number of output channel is shown under each block. In the last layer, Sigmoid/Softmax functions are applied for 2 and 4 classes classification of ROI extraction subtask and bi-ventricle segmentation subtask, respectively.

where $K_2 = 4$ (4 classes: LV, RV, MYO, and background), and $N_2 = H \times W = 128 \times 128$ which is the number of total pixels of ROI image.

To the end, we obtain the final losses for both ROI extraction and bi-ventricle segmentation tasks as follows:

$$\mathcal{L}_{C1-cGAN}(D_1, G_1) = \mathcal{L}_{adv}(D_1, G_1) + \lambda \cdot \mathcal{L}_{gen}(G_1) \quad (9)$$

$$\mathcal{L}_{C2-cGAN}(D_2, G_2) = \mathcal{L}_{adv}(D_2, G_2) + \lambda \cdot \mathcal{L}_{gen}(G_2) \quad (10)$$

where $\lambda$ is set to 10 empirically for magnitude balance. In order to reduce the complexity of training and to segment the bi-ventricle with accurate cardiac ROI, we separately train the two networks in sequence.

## C. NETWORK ARCHITECTURES

### 1) GENERATOR NETWORK

As shown in Fig. 3, the generator network in C-cGANs includes three parts: the encoder path, the MSAF module and decoder path. Compared with the 31 million (M) parameters of the U-Net, the number of parameters of the proposed generator network is only 6.3 M.

#### a: ENCODER PATH AND DECODER PATH

In the encoder path, we take a simplified inception block as the first layer for processing the input image by channel concatenation with multi-scales. Each group of filter number is 16 and kernels have three sizes of $1 \times 1$, $3 \times 3$ and $5 \times 5$ (see Fig. 4(a)). Unlike U-Net, we adopt residual block before each max-pooling layer instead of two simple convolution layers in encoder path. As shown in Fig. 4 (b), this block concatenates the previous layer and the layer obtained after two convolutions as the next input. Thus,
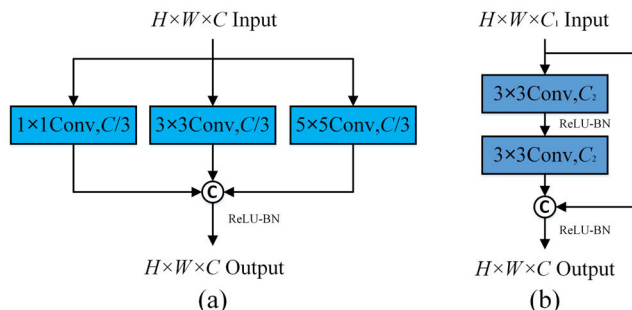


**FIGURE 4.** (a) Simplified Inception block. We concatenate three $C/3$ channels produced by $1 \times 1$, $3 \times 3$ and $5 \times 5$ convolution layers. (b) Residual block. The number of output channel $C = C_1 + C_2$.

more significant original information can be preserved, and vanishing gradient and exploding gradient problems can be avoided. The feature map is resized into 1/8 of the input image by three max-pooling operations to extract high-level semantic information. In the decoder path, we utilize the up-sampling operation to reshape the feature map to original size. In order to alleviate the number explosions of feature map in the encoder path, we employ skip connection to combine max-pooling layers and corresponding up-sampling layers using pixel-wise addition rather than channel concatenation. A $1 \times 1$ convolution which is considered as bottleneck layer is used in the skip connection to math channel dimension. This operation contributes to reduce parameter and memory usage compared to the concatenation operation without affecting the segmented quality [38]. Finally, we apply Sigmoid or Softmax function as the last layer for the two tasks to get

the pixel-wise classification mask. In addition, after each convolution layer of the entire network, we add Rectified Linear Unit (ReLU) [39] and Batch Normalization (BN) [40] operation.

#### b: MULTI-SCALE ATTENTION FUSION (MSAF) MODULE

MSAF module is obtained by connecting PAM block and MDF block. It can enhance the relationship between pixel features through attention mechanism and extract the features of ventricles with various scale by combining multi-scale fusion mechanism. Specifically, the module first captures the spatial dependence between different locations of the high-level feature map to enhance feature representation. Then dilated convolutions with increasing rates is used for multi-scale context aggregation according to spatial attention mapping. Finally, multi-scale feature maps and attention-based maps are fused to avoid gradient disappearance and explosion, and we acquire the feature maps of the same size and dimension as the input.

*Position Attention Mechanism (PAM) Block:* Previous works [41], [42] presented that local features extracted from general convolutional networks could lead to misclassification of objects. The PAM block is designed to search abundant contextual correlations over local features[43], thus we use it to enhance the representation of each feature by utilizing the association between any two position features. More specifically, as shown in Fig. 5, firstly we utilize input local feature $I \in \mathbb{R}^{H \times W \times C}$ to generate feature maps $A$ and $B$ ($A, B \in \mathbb{R}^{H \times W \times C/4}$) by dimension reduction of $1 \times 1$ convolution and reshape them to $\mathbb{R}^{N \times C/4}$, where $N = H \times W$ is the total elements of one feature map. The new feature maps $C \in \mathbb{R}^{H \times W \times C}$ are also generated from $I$ by convolution operation and reshaped into $\mathbb{R}^{N \times C}$. Then, we feed $B \times A^T$ into Softmax function to calculate spatial attention map $P \in \mathbb{R}^{N \times N}$ (see Equation (11), where $S_{ji}$ denotes the impact of the $i^{th}$ position on the $j^{th}$ position). After that, we multiply $C$ and $P^T$, and reshape the matrix into $Q \in \mathbb{R}^{H \times W \times C}$. At last, we sum $Q$ and $I$ by pixel-wise addition operation to get the final output $O \in \mathbb{R}^{H \times W \times C}$. According to this, the output $O$ at each position is considered as a weighted sum of features across all positions and original features. The PAM
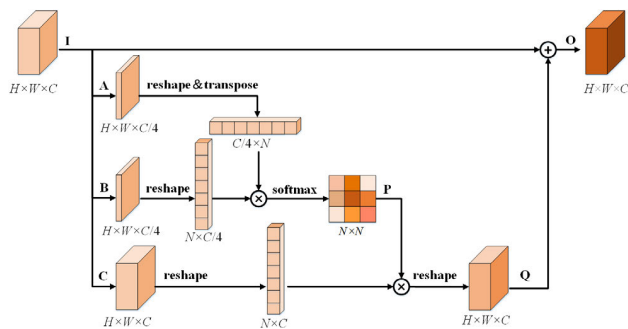


**FIGURE 5.** The details of PAM block. $N = H \times W$ is the total elements of one feature map.
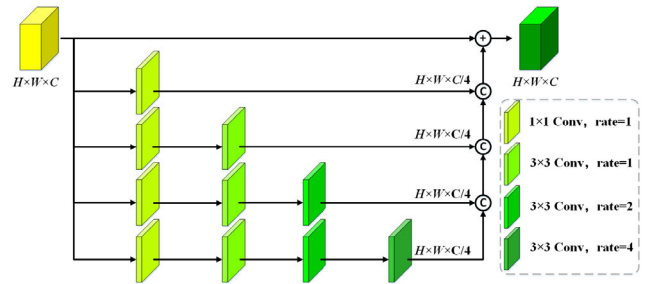


**FIGURE 6.** The details of MDF block. This block has one main branch and four sub-branches of which the receptive field is 1, 3, 7, and 15 in turn.

module enables similar semantic features to achieve mutual position benefits, thereby improving intra-class correlation and semantic consistency.

$$s_{ji} = \frac{e^{(B_i, C_j)}}{\sum_{i=1}^{N} e^{(B_i, C_j)}} \qquad (11)$$

*Multi-Scale Dilated Fusion (MDF) Block:* After PAM block, we design the MDF block which consists of dilated convolution with different rates to encode output maps based on spatial attention mechanism. The dilated convolution introduced by [44] is widely used in object extraction and semantic segmentation. Compared with pooling layer, it can expand receptive fields by changing the rate value without loss of resolution or coverage. Rate = 1 means standard convolution. As shown in Fig. 6, this block has one main branch and four sub-branches. In sub-branches, we firstly use $1 \times 1$ convolution to reduce the dimension of the input channel to a quarter. Then we gradually increase dilated convolution layers with increasing rate of 1, 2 and 4. Accordingly, the receptive field of each sub-branch is 1, 3, 7, and 15 in turn. Finally, the multi-scale feature maps of each sub-branch are fused by concatenation operation. In the main branch, we directly add the original input with concatenation of multi-scale features for stable feature propagation, similar to a residual connection. Small receptive field can contribute to detect small target regions and large receptive field can have a larger spatial area to capture high-level abstract features. By fusion of different receptive fields on attention-based maps, we get multi-scale context features as input of decoder path.

#### 2) DISCRIMINATOR NETWORK

As listed in Table 1, the discriminator network uses a general CNN and acts as a classifier with positive and negative feedback. We encode and downsample the feature maps by four convolution layers with kernel size of $5 \times 5$, strides 2 and reduce dimensionality from 3D to 1D which represents one response for every heat map by global maxpooling operation. LeakyReLU [45] (0.2 slope) and BN [40] operation follow each convolution layer. After that we flatten the heat map and output a classification value by Sigmoid function. The value close to 0 represents false segmentation, whereas value close

**TABLE 1.** The architecture of discriminator network. The value close to 0 represents false segmentation, whereas value close to 1 represents successful segmentation.

| Layers | Input size | Output size | Operational details |
|---|---|---|---|
| Convolution layer | 128×128×(K+1) | 64×64×32 | 5×5 Conv, stride 2, 32 Filters |
| Convolution layer | 64×64×32 | 32×32×64 | 5×5 Conv, stride 2, 64 Filters |
| Convolution layer | 32×32×64 | 16×16×128 | 5×5 Conv, stride 2, 128 Filters |
| Convolution layer | 16×16×128 | 8×8×256 | 5×5 Conv, stride 2, 256 Filters |
| Pooling layer | 8×8×256 | 1×1×256 | Global MaxPooling |
| Classification layer | 1×1×256 | 1×1 | Sigmoid |

Note: K is the number of classes.

to 1 represents successful segmentation. The discriminator network has only 1.1 M parameters.

### D. ADVERSARIAL TRAINING STEPS

We take C2-cGAN as an example to describe the process of adversarial training since the training processes of two tasks are in the same pattern.

The adversarial training process contains four steps of updates for the parameters $\theta_G$ of $G_2$ and parameters $\theta_D$ of $D_2$: (1) $G_2$ generates segmentation map $G_2(x_2)$ similar to $gt_2$ to update parameters $\theta_G$; (2) $D_2$ attempts to output 1 to update parameters $\theta_D$ using $gt_2$ and ROI image as input; (3) $D_2$ attempts to output 0 to update parameters $\theta_D$ using $G_2(x_2)$ and ROI image as input; (4) $G_2$ and $D_2$ work together (the former generates $G_2(x_2)$ and the later makes a discrimination) to update parameters $\theta_G$ based on fixed parameters $\theta_D$. These four updates of parameters are performed alternately for mutual iteration between $G$ and $D$.

### III. EXPERIMENTS, RESULTS AND DISCUSSION
### A. DATASET AND EVALUATION CRITERIA

In this study, we train and test the proposed network model on ACDC 2017 [46] (provided for MICCAI 2017 challenge). This open dataset includes 150 exams of individual patients divided into 5 evenly distributed subgroups (4 pathological plus 1 healthy subject groups). The data is acquired over 6 years using two MRI scanners of different magnetic strengths (1.5 T and 3.0 T) from Hospital of Dijon (France). Images are acquired in breath hold with a retrospective or prospective gating and a steady state free precession (SSFP) sequence in short axis orientation. In particular, a series of short-axis slices cover from the base to the apex of LV, with a thickness of 5 or 8 mm and an interslice gap of 5 or 10 mm. The spatial resolution goes from 1.37 to 1.68 mm$^2$/pixel. Each exam contains 28∼40 images that cover completely or partially the cardiac cycle. The dataset has been divided into two parts: 100 exams for training and 50 exams for testing. The pixel-level manual annotations of bi-ventricle (LV/RV/MYO) were manually drawn on short-axis slices in ED and ES phases by two cardiologists. It should be noted that manual annotations are only provided for the training set.

The final segmentation results on testing set need to be submitted online for evaluation.[1]

Segmentation accuracy is measured by geometrical metrics: Dice [47] and Hausdorff distance (HD) [48] at the ED and ES phases. Dice measures the overlap between both predicted contour $P$ and manual contour $G$ by varying from 0 (without overlap) to 1 (full overlap) in Equation (12). HD measures symmetric distance between both two contours in Equation (13), where $p$ and $g$ are the pixel in $P$ and $G$, respectively.

$$D(P, G) = 2\frac{P \cap G}{P + G} \tag{12}$$

$$H(P, G) = \max\left(\max_{p \in P}\left(\min_{g \in G} d(p, g)\right),\right.$$
$$\left.\max_{g \in G}\left(\min_{p \in P} d(p, g)\right)\right) \tag{13}$$

In addition, clinical metrics are also introduced which are the most widely used in cardiac clinical practice. The clinical metrics include the correlation, the bias and the standard deviation from the ejection fraction (EF, for LV and RV), the ED volume (EDV, for LV and RV) and the myocardial mass. Let $V_{ED}$ and $V_{ES}$ be the volume at ED and ES phases, the calculation of EF is shown in Equation (14).

$$EF = \frac{V_{ED} - V_{ES}}{V_{ED}} \cdot 100\% \tag{14}$$

### B. IMPLEMENTATION DETAILS

We implemented our code using Keras[2] library with Tensorflow backend based on Ubuntu 16.04 system. The proposed model was trained and tested using GPU of NVidia GeForce GTX 1080Ti with 11G memory.

Before training, we adopted mean-variance normalization to normalize the pixel intensity distribution of each input image. In addition, we utilized horizontal flip, vertical flip, random scaling (0.2 scale) and random rotation transformation between ±90° operations to generate new images for data augmentation. During the training stage, the hyper-parameter

---

[1] https://www.creatis.insa-lyon.fr/Challenge/acdc/index.html
[2] https://keras.io/

**TABLE 2.** Intra-comparison of ablation experiments on Dice metric for bi-ventricle segmentation performance on ACDC 2017. The bold values refer to the best scores for each metric. Number format: mean value (standard deviation).

| Method | LV | | RV | | MYO | |
|---|---|---|---|---|---|---|
| | ED | ES | ED | ES | ED | ES |
| U-Net (baseline) | 0.950 (0.030) | 0.908 (0.042) | 0.913 (0.045) | 0.820 (0.081) | 0.874 (0.034) | 0.880 (0.034) |
| cGAN | 0.963 (0.017) | 0.916 (0.055) | 0.930 (0.040) | 0.839 (0.082) | 0.886 (0.026) | 0.893 (0.021) |
| C-cGANs without MSAF | **0.965 (0.017)** | 0.920 (0.055) | 0.938 (0.039) | 0.859 (0.064) | 0.893 (0.029) | 0.907 (0.024) |
| C-cGANs | **0.965 (0.015)** | 0.929 (0.041) | **0.949 (0.014)** | 0.879 (0.064) | **0.895 (0.027)** | 0.908 (0.026) |
| C-cGANs + Post-processing | **0.965 (0.015)** | 0.930 (0.039) | **0.949 (0.013)** | 0.881 (0.063) | 0.893 (0.027) | **0.910 (0.026)** |

**TABLE 3.** Intra-comparison of ablation experiments on HD metric for bi-ventricle segmentation performance on ACDC 2017. The bold values refer to the best scores for each metric. Number format: mean value (standard deviation).

| Method | LV | | RV | | MYO | |
|---|---|---|---|---|---|---|
| | ED | ES | ED | ES | ED | ES |
| U-Net (baseline) | 36.47 (67.14) | 52.79 (70.33) | 53.70 (67.19) | 52.26 (68.06) | 68.04 (71.39) | 84.29 (76.46) |
| cGAN | 7.69 (10.70) | 18.44 (21.87) | 11.70 (5.62) | 15.64 (5.50) | 28.91 (36.04) | 37.53 (33.06) |
| C-cGANs without MSAF | 7.94 (15.06) | 14.41 (21.11) | 18.67 (24.65) | 20.53 (17.99) | 13.15 (16.06) | 23.12 (22.70) |
| C-cGANs | **4.94 (2.73)** | 11.76 (20.51) | 19.37 (18.54) | 14.44 (13.33) | 9.53 (6.85) | 17.20 (17.14) |
| C-cGANs + Post-processing | **4.94 (2.73)** | **6.34 (3.06)** | **9.27 (5.97)** | **12.82 (5.54)** | **8.78 (6.00)** | **9.88 (4.52)** |

setting is consistent for both ROI extraction and segmentation tasks of the network. We employed "Xavier" [49] initializer to randomly initialize network weights. We used Adam [50] optimizer ($\beta_1 = 0.9, \beta_2 = 0.999$ and $\varepsilon = 10^8$) with the initial learning rate (*initial_lr*) of $10^4$ for $G$ network and $10^5$ for $D$ network. This is because $D$ can have sufficient impact on the $G$ network by making $D$ adapts more slowly than the $G$[51]. We set the decay policy for learning rate as $initial\_lr \times \left(1 - \frac{iter}{max\_iter}\right)^{power}$, where *power* is set to 0.5, *iter* is the current iteration and *max_iter* is the maximum iteration. We trained for 200 epochs with batch-size 1. In the testing stage, the discriminator network did not work and the generator network generated accurate segmentation contours automatically.

Moreover, based on the fact of prior information that each target (LV/RV/MYO) has only one connected component and the heart has two complete blood pools, we employed morphological post-processing including maximum connected component preserving and binary hole filling operations for bi-ventricle segmentation results.

### C. ABLATION EXPERIMENTS ANALYSIS
In order to evaluate the effectiveness of the main components of the proposed model, we have performed an internal comparison by ablation experiments. Note that manual contours of testing set of ACDC 2017 are not provided, so we select 4 cases from each subgroup to construct the validation set (20 cases), and construct the training set with the left 80 cases. As shown in Table 2 and Table 3, we design five methods by controlling variates for intra-comparison. Specifically, we use original U-Net [20] as the standard segmentation baseline.

cGAN means that the model without cascaded network (without ROI extraction part) uses $256 \times 256$ size of original cardiac MRI as input. C-cGANs without MSAF means that we remove the MSAF module from the two generative networks in C-cGANs. We also use post-processing to process our results of C-cGANs for comparison.

The segmentation results are shown in Table 2 and Table 3. We observe that cGAN frame obtains huge improvements for all metrics compared with U-Net. C-cGANs avoid false positive segmentation caused by other similar tissues and organs by ROI extraction firstly, making a substantial decline of HD and 0.2~4% rise of Dice for each value compared with cGAN. False positives only results in a slight decrease of Dice, while it does bring an explosion of HD. The success rate of ROI extraction is 100%. The reason is that the fixed size ROI can completely contain the heart region, and on the other hand, occasional failures of slice do not affect the heart center positioning of the whole sequence. Since the MSAF module also plays an important role in segmentation, especially in Dice of RV, we compare the C-cGANs with and without MSAF. This shows that the proposed module can robustly learn complex and variable RV geometric structure. In addition, post-processing including two morphological operations yields the correct cardiac anatomy to reduce HD while maintaining good Dice. At the bottom rows of Table 2 and Table 3, the final results of proposed method are presented: for ES phase, we obtain the Dice of 0.965±0.015, 0.949±0.013, 0.893±0.027 and the HD of 4.94±2.73 mm, 9.27±5.97 mm, and 8.78±6.00 mm for LV, RV, and MYO, respectively; for ED phase, we obtain the Dice of 0.930±0.039, 0.881±0.063, 0.910±0.026 and the HD of
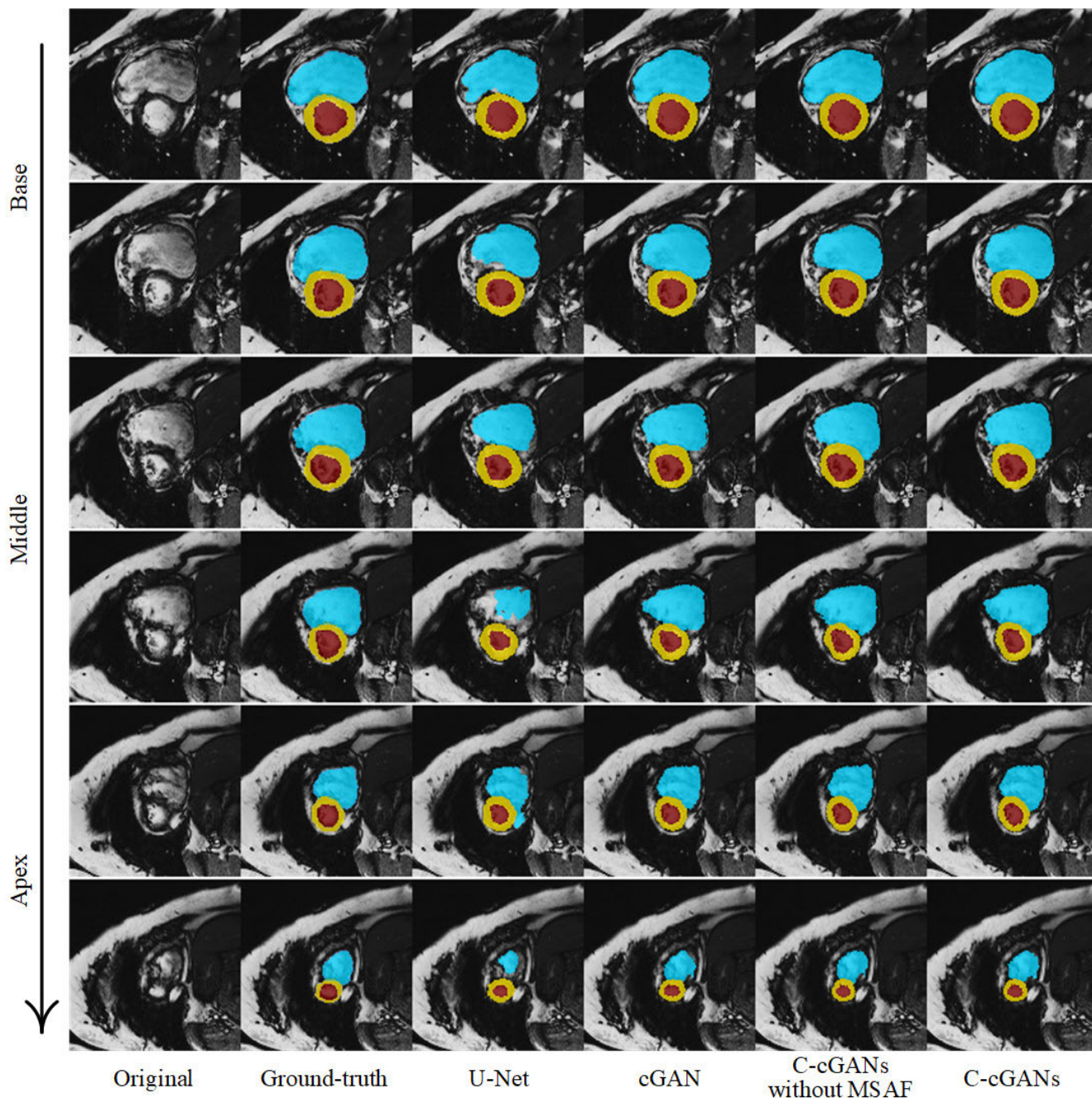
**FIGURE 7.** ED example of cardiac MRI segmentation results from ablation experiments. Images from top to down represent slices located from base to apex.

6.34±3.06 mm, 12.82±5.54 mm, and 9.88±4.52 mm for LV, RV, and MYO, respectively. The quantitative mean value and standard deviation demonstrate the effectiveness of each part in the ablation experiment.

Representative examples in ED and ES phases are shown in Fig. 7 and Fig. 8 (enlarged view). Through the intuitive and qualitative comparison, the adversarial training has been sufficiently improved rather than U-Net. The addition of MSAF module makes the model more focused on the diversity

of ventricular morphology and achieves better effects in bi-ventricle segmentation, especially on the base and apex slices.

### D. COMPARISON WITH THE STATE-OF-THE-ART METHODS

In order to measure the quantitative evaluation metrics of our model, we upload the segmentation results of ACDC 2017 testing set for online evaluation. Fig. 9 shows the
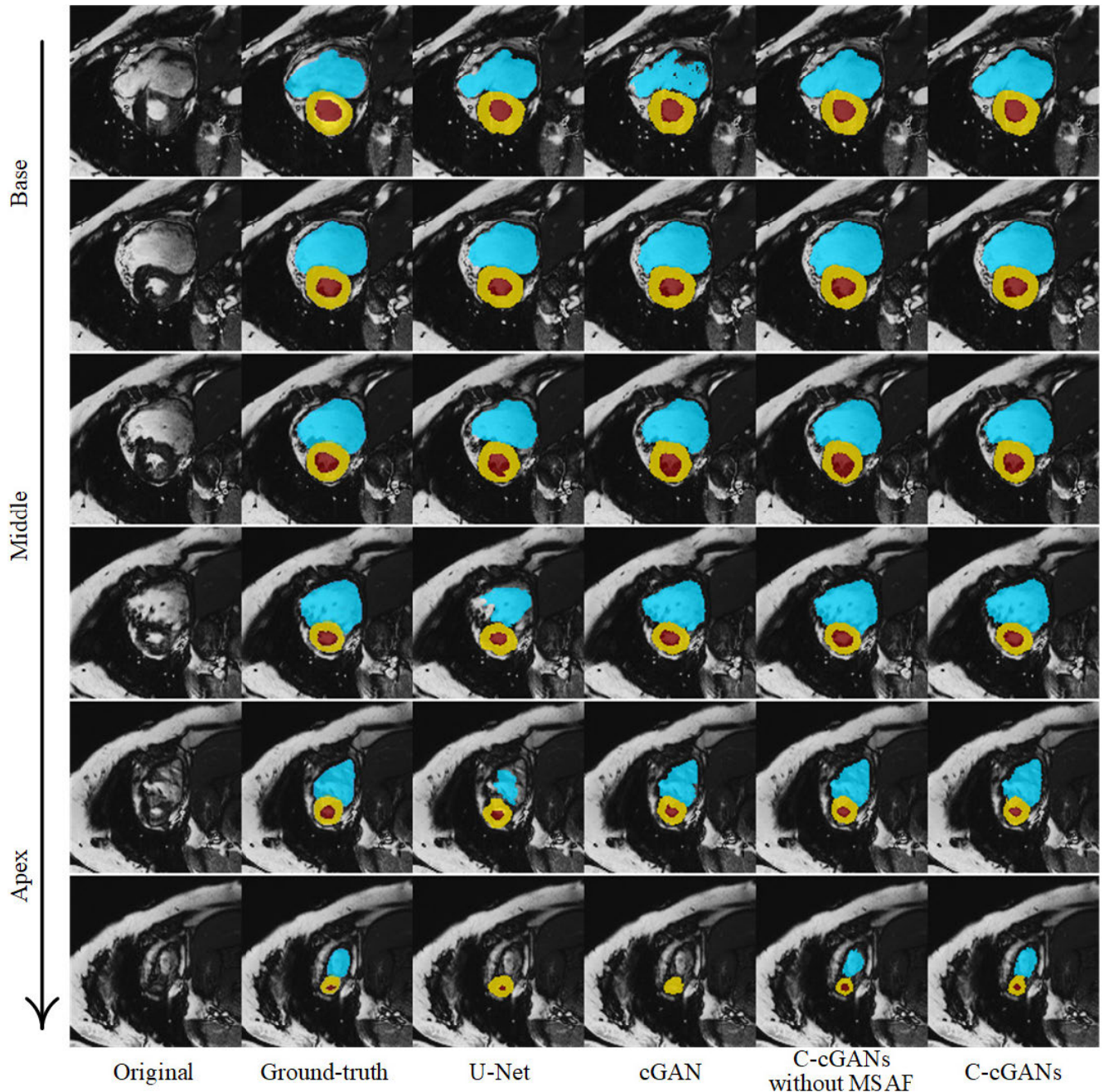
**FIGURE 8.** ES example of cardiac MRI segmentation results from ablation experiments. Images from top to down represent slices located from base to apex.

geometric metrics for ED/ES. Average Dice of LV/RV (0.968/0.951) in ED tends to be higher than that (0.936/0.894) in ES, however, the MYO is opposite (0.904/0.921 in ED/ES). This is because that the relatively large area is more conducive to the segmentation of the contour. HD of RV are the largest in both phases due to complex variable shape. Geometric results with fewer outliers are promising and need to be compared with other methods. Moreover, we quantify three main clinical metrics: EF for LV, EF for RV and mass for MYO. In the left column of Fig. 10, linear regression analyses indicate the

correlation of clinical metrics between the manual contours and automated segmentation. We can see most red points are close to the blue line (y = x), indicating that the predicted results are highly correlated to manual metrics. Right column of Fig. 10 shows consistency between the automated and the manual segmentation by using bland-Altman analysis. We observe that most points locate in the range of the Mean $\pm 1.96SD$, which reports the 95% limits of consistency. They demonstrate the predicted results of our method are highly consistent with manual metrics.

**TABLE 4.** Comparison of the geometric metrics for published methods on ACDC 2017 testing set. No. 1∼9 are online leaderboard, No. 10∼12 are recent reported studies. Red is the best score and blue is the second-best score. Number format: mean value.

| No. | Methods | LV Dice ED | LV Dice ES | LV HD (mm) ED | LV HD (mm) ES | RV Dice ED | RV Dice ES | RV HD (mm) ED | RV HD (mm) ES | MYO Dice ED | MYO Dice ES | MYO HD (mm) ED | MYO HD (mm) ES |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| 1 | Isensee et al. [56] | 0.968 | 0.931 | 7.4 | 6.9 | 0.946 | 0.899 | 10.1 | 12.1 | 0.902 | 0.919 | 8.7 | 8.7 |
| 2 | Baumgartner et al. [26] | 0.963 | 0.911 | 6.6 | 9.2 | 0.932 | 0.883 | 12.7 | 14.7 | 0.892 | 0.901 | 8.7 | 10.6 |
| 3 | Jang et al. [57] | 0.959 | 0.921 | 7.7 | 7.1 | 0.929 | 0.885 | 12.9 | 11.8 | 0.875 | 0.895 | 9.9 | 8.9 |
| 4 | Zotti et al. [58] | 0.957 | 0.905 | 6.6 | 8.7 | 0.941 | 0.882 | 10.3 | 14.1 | 0.884 | 0.896 | 8.7 | 9.3 |
| 5 | Khened et al. [59] | 0.964 | 0.917 | 8.1 | 9.0 | 0.935 | 0.879 | 14.0 | 13.9 | 0.889 | 0.898 | 9.8 | 12.6 |
| 6 | Wolterink et al. [60] | 0.961 | 0.918 | 7.5 | 9.6 | 0.928 | 0.872 | 11.9 | 13.4 | 0.875 | 0.894 | 11.1 | 10.7 |
| 7 | Patravali et al. [61] | 0.955 | 0.885 | 8.2 | 10.9 | 0.911 | 0.819 | 13.5 | 18.7 | 0.882 | 0.897 | 9.8 | 11.3 |
| 8 | Rohé et al. [62] | 0.957 | 0.900 | 7.5 | 10.7 | 0.916 | 0.845 | 14.0 | 15.9 | 0.867 | 0.869 | 11.5 | 13.0 |
| 9 | Grinias et al. [63] | 0.948 | 0.865 | 8.9 | 11.6 | 0.863 | 0.743 | 21.0 | 25.7 | 0.794 | 0.801 | 12.6 | 14.8 |
| 10 | Tong et al. [52] | 0.966 | 0.918 | 8.0 | 8.2 | 0.948 | 0.898 | 11.7 | 12.8 | 0.905 | 0.915 | 9.5 | 9.6 |
| 11 | Zotti et al. [22] | 0.963 | 0.912 | 6.2 | 8.4 | 0.934 | 0.885 | 11.1 | 12.7 | 0.886 | 0.902 | 9.6 | 9.3 |
| 12 | Khened et al. [53] | 0.964 | 0.917 | 8.1 | 9.0 | 0.935 | 0.879 | 14.0 | 13.9 | 0.889 | 0.898 | 9.8 | 12.6 |
| 13 | Proposed method | 0.968 | 0.936 | 5.6 | 6.2 | 0.951 | 0.894 | 10.5 | 12.0 | 0.904 | 0.921 | 7.0 | 7.9 |

**TABLE 5.** Comparison of the clinical metrics for published methods on ACDC 2017 testing set. No. 1∼9 are online leaderboard, No. 10∼12 are recent reported studies. Red is the best score and blue is the second-best score.

| No. | Methods | $LV_{EDV}$ Corr val. | $LV_{EDV}$ Bias ml. | $LV_{EDV}$ Std ml. | $LV_{EF}$ Corr val. | $LV_{EF}$ Bias % | $LV_{EF}$ Std % | $RV_{EDV}$ Corr val. | $RV_{EDV}$ Bias ml. | $RV_{EDV}$ Std ml. | $RV_{EF}$ Corr val. | $RV_{EF}$ Bias % | $RV_{EF}$ Std % | $MY_{Mass}$ Corr val. | $MY_{Mass}$ Bias g. | $MY_{Mass}$ Std g. |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| 1 | Isensee et al. [56] | 0.997 | 2.7 | 5.1 | 0.991 | 0.2 | 2.1 | 0.988 | 4.4 | 7.9 | 0.901 | -2.7 | 4.7 | 0.989 | -4.8 | 7.3 |
| 2 | Baumgartner et al. [26] | 0.995 | 1.4 | 6.1 | 0.988 | 0.6 | 2.6 | 0.977 | -2.3 | 11.1 | 0.851 | 1.2 | 5.7 | 0.982 | -6.9 | 9.8 |
| 3 | Jang et al. [57] | 0.993 | -0.4 | 6.0 | 0.989 | -0.3 | 2.3 | 0.989 | -10.8 | 12.1 | 0.793 | -3.2 | 6.3 | 0.968 | 11.5 | 14.1 |
| 4 | Zotti et al. [58] | 0.997 | 9.6 | 10.3 | 0.987 | -1.2 | 2.7 | 0.991 | -3.7 | 7.4 | 0.872 | -2.2 | 5.4 | 0.984 | -12.4 | 13.1 |
| 5 | Khened et al. [59] | 0.997 | 0.6 | 4.2 | 0.989 | -0.5 | 2.5 | 0.982 | -2.9 | 8.4 | 0.858 | -2.2 | 5.3 | 0.990 | -2.9 | 6.3 |
| 6 | Wolterink et al. [60] | 0.993 | 3.0 | 6.8 | 0.988 | -0.5 | 2.5 | 0.980 | 3.6 | 10.9 | 0.852 | -4.6 | 6.6 | 0.963 | -1.0 | 10.0 |
| 7 | Patravali et al. [61] | 0.997 | 9.9 | 10.8 | 0.971 | 1.7 | 4.1 | 0.945 | 5.6 | 15.0 | 0.791 | 6.8 | 8.3 | 0.989 | 11.6 | 11.9 |
| 8 | Rohé et al. [62] | 0.993 | 4.2 | 7.5 | 0.989 | -0.1 | 2.6 | 0.983 | 7.3 | 11.7 | 0.781 | -0.7 | 7.8 | 0.967 | -3.4 | 10.3 |
| 9 | Grinias et al. [63] | 0.992 | 2.0 | 8.5 | 0.975 | -1.6 | 4.3 | 0.930 | 18.6 | 24.8 | 0.785 | -0.5 | 7.1 | 0.942 | -28.9 | 30.3 |
| 10 | Tong et al. [52] | 0.995 | 0.1 | 5.4 | 0.982 | 1.5 | 2.9 | 0.990 | 1.6 | 6.5 | 0.868 | -1.7 | 5.4 | 0.991 | -6.7 | 8.1 |
| 11 | Zotti et al. [22] | 0.997 | 3.7 | 5.1 | 0.990 | -0.5 | 3.1 | 0.986 | 2.4 | 11.5 | 0.869 | -0.9 | 6.8 | 0.986 | -1.8 | 8.6 |
| 12 | Khened et al. [53] | 0.997 | 0.6 | 5.5 | 0.989 | -0.5 | 3.4 | 0.982 | -2.9 | 12.7 | 0.858 | -2.2 | 6.8 | 0.990 | -2.9 | 7.5 |
| 13 | Proposed method | 0.996 | 2.5 | 6.2 | 0.991 | 0.6 | 3.1 | 0.993 | 0.5 | 8.1 | 0.844 | -1.82 | 7.7 | 0.990 | -2.8 | 7.2 |
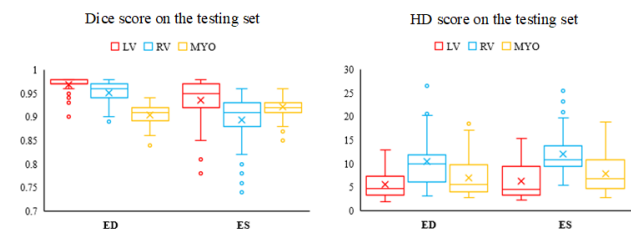


**FIGURE 9.** The box-plot results of Dice (left) and HD (right) on the ACDC 2017 testing set for both ED and ES phases.

Table 4 and Table 5 show the geometric and clinical metrics from the ACDC 2017 online leaderboard (No. 1∼9) and latest researches (No. 10∼12). For each metric, the red color

represents the best score, the blue color is the second-best score for reference. The paper [46] describes the leaderboard methods of this challenge as details, and eight of the top nine (No. 1∼8) use deep learning methods so that we can anticipate the importance of deep learning and the accuracy compared with traditional methods. In addition, we also collect the latest state-of-the-art researches on ACDC 2017 testing set for comparison. Tong *et al.* [52] proposed a recurrent interleaved attention network based on recurrent feedback blocks and interleaved attention mechanism. Zotti *et al.* [22] implemented CNN architecture which is an extension of the U-Net that embeds a cardiac shape prior and involves a loss function tailored to the cardiac anatomy. Khened *et al.* [53] presented a novel DenseNet based on FCN with incorporating
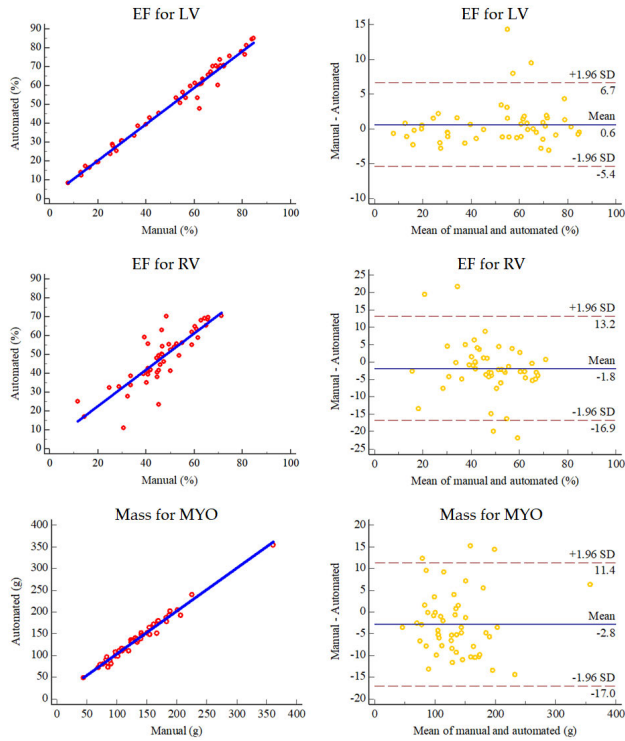
**FIGURE 10.** The comparison of EF for LV, EF for RV and mass for MYO between manual (ground-truth) and automated (our method) segmentation on the ACDC 2017 testing set. Left column: linear regression analysis. Right column: Bland-Altman analysis.



**FIGURE 11.** Comparison of ground-truth between MS-CMRSeg 2019 and ACDC 2017. (a) RV is delineated by MS-CMRSeg 2019, but not by ACDC 2017 for the similar base position. (b) The segmentation convention of the slices below the base seems roughly consistent for both datasets.

long skip and short-cut connections to overcome the feature map explosion. In geometric metrics, we perform best or second-best in the LV and MYO, and also Dice of ED and HD of ES in RV. In clinical metrics, we get the best correlation in EF for LV and EDV for RV, and second-best result in EDV for LV and Mass for MYO, except for EF for RV. The highest accuracy in ED segmentation and the general accuracy in ES segmentation result in relatively low EF for RV compared with other studies (calculation in Equation 14). As shown in Table 4 and Table 5, the proposed method yields most colored scores. This shows that we get competitive results of ACDC 2017 testing set. In general, we conclude that the proposed model is a more efficient bi-ventricle segmentation method than other existing methods. This indicates the potential of the proposed model in clinical practice.

### E. GENERALIZATION CAPABILITY ON OTHER DATASETS

To test the generalization capability and performance of the model we proposed on the untrained dataset, we directly apply the trained model without fine-tuning on other two datasets: Multi-sequence Cardiac MR Segmentation Challenge (MS-CMRSeg 2019) [54] and Right Ventricle Segmentation Challenge (RVSC 2012) [55]. Although the motivation of MS-CMRSeg 2019 is not exactly the same as ours, it is the only open dataset containing balanced-SSFP (bSSFP) short- axis cardiac sequences with bi-ventricle contours delineation besides ACDC 2017. Furthermore, since the
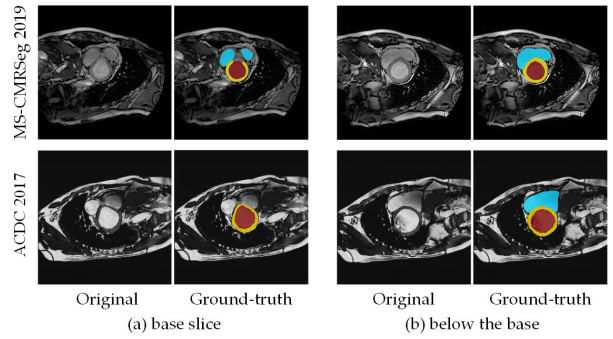
segmentation of RV with complex variable shape is more difficult than that of LV, we also extend the generalization test on RVSC 2012.

#### 1) GENERALIZATION EXPERIMENT ON MS-CMRSeg 2019

MS-CMRSeg 2019 consists of 35 bSSFP cardiac sequences with manual delineation at ED phase. Consistency of segmentation convention between different datasets is critical to the performance of segmentation [29]. As shown in Fig. 11 (a), for the similar base position in two datasets, RV is delineated by MS-CMRSeg 2019, but not by ACDC 2017. By contrast, the segmentation convention of the slices below the base seems roughly consistent (see Fig. 11 (b)). Therefore, we select the slices below the base to eliminate the effects caused by different segmentation conventions.

The performance of bi-ventricle segmentation on MS-CMRSeg 2019 is listed in Table 6, with the segmentation results of ACDC 2017 testing set at ED phase as a reference. We can see that Dice and HD for LV are very close to that on ACDC 2017. Specifically, Dice of RV and MYO is about 4% lower than ACDC 2017, which is probably due to some inappropriate ground-truth provided by MS-CMRSeg 2019. As shown in Fig. 12 (a), for some slices in MS-CMRSeg 2019, the predicted contours are more reasonable and consistent with the anatomical structure. Incomplete or missing contours of MYO and RV may reduce the accuracy of evaluation metrics. The results of LV also indicate that the segmentation of LV is easier due to its geometric shape. The similar standard deviations show the robustness of generalization for bi-ventricle segmentation. Examples of ground-truth and prediction on MS-CMRSeg 2019 are shown in Fig. 12 (b).

#### 2) GENERALIZATION EXPERIMENT ON RVSC 2012

In RVSC 2012, 16 cases of training set with the manual delineation of RV are all located between the base and apex (Test1 and Test2 sets are no longer accepted for online evaluation, so we use training set only). Cardiac images given by RVSC 2012 have been zoomed and cropped to $256 \times 216$ (or $216 \times 256$) sized ROI [55]. Therefore, the images are

**TABLE 6.** The generalization results of our trained model on the MS-CMRSeg 2019 (without fine-tuning). The bottom reference shows the metric values of the ED phase on ACDC 2017 testing set. Number format: mean value (standard deviation).

| Reference | LV | | RV | | MYO | |
|---|---|---|---|---|---|---|
| | Dice | HD (mm) | Dice | HD (mm) | Dice | HD (mm) |
| MS-CMRSeg 2019 | 0.963 (0.010) | 6.38 (3.09) | 0.912 (0.031) | 15.91 (7.01) | 0.863 (0.030) | 11.54 (5.49) |
| ACDC 2017 | 0.968 (0.014) | 5.64 (2.93) | 0.951 (0.025) | 10.47 (6.06) | 0.904 (0.024) | 7.03 (4.11) |

**TABLE 7.** The generalization results of our trained model on the RVSC 2012 (without fine-tuning). The reference list includes the recent state-of-the-art researches on RVSC 2012. Number format: mean value (standard deviation).

| Reference | Datasets involved | ED | | ES | |
|---|---|---|---|---|---|
| | | Dice | HD (mm) | Dice | HD (mm) |
| Proposed method | Training | 0.87 (0.14) | 9.28 (4.69) | 0.84 (0.07) | 7.56 (2.90) |
| Ringenberg et al. [64] | Training | 0.89 (0.10) | 6.95 (5.14) | 0.78 (0.17) | 8.74 (5.61) |
| Punithakumar et al. [65] | Training | — | — | 0.82 (0.15) | 7.19 (4.05) |
| Zuluaga et al. [66] | Test1 | 0.83 (0.17) | 9.77 (7.88) | 0.72 (0.27) | 11.41 (10.49) |
| Bai et al. [67] | Test1 | 0.86 (0.11) | 7.70 (3.74) | 0.69 (0.25) | 11.16 (5.53) |
| Guo et al.[10] | Test1 | 0.88 (0.08) | 8.03 (5.57) | 0.81 (0.14) | 9.46 (6.93) |
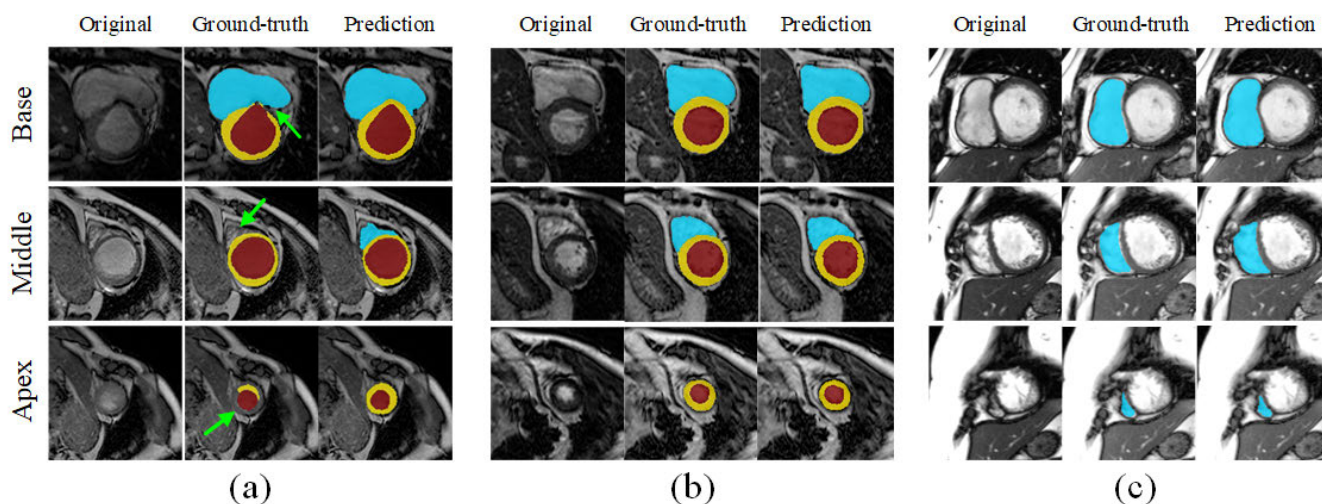


**FIGURE 12.** Examples of ground-truth vs prediction on two datasets. (a) Inappropriate ground-truth of certain slice. The green arrows indicate some examples with incomplete MYO contours and missing RV contours on MS-CMRSeg 2019. (b) Examples of bi-ventricle segmentation on MS-CMRSeg 2019 (without fine-tuning). (c) Examples of RV segmentation on RVSC 2012(without fine-tuning).

directly zero-padded to square ($256 \times 256$) and then reshaped to $128 \times 128$ as the input of trained C2-cGAN.

Table 7 presents the comparison with other state-of-the-art researches on RVSC 2012. We achieve better scores such as Dice at ES phase. In addition, we also get the smaller standard deviation at ES phase which demonstrates the robustness of our model for RV segmentation. Examples of ground-truth and prediction on RVSC 2012 are shown in Fig. 12 (c).

## IV. CONCLUSION

Bi-ventricle segmentation is the first step in cardiac function analysis. In this work, we propose a cascaded network frame named C-cGANs to detect cardiac ROI and concurrently

segment LV, RV and MYO. The generator generates segmentation mask, meanwhile the discriminator distinguishes the source of the segmentation mask of input, and finally both are optimized by competing with each other. In order to capture more spatial information with multi-scale semantic features, we design a MSAF module embedded in the encoder-decoder architecture as the generator network. The proposed model has several advantages: 1) The cascaded framework helps to reduce false positives because the formal bi-ventricle segmentation works on the cardiac region by ROI extraction; 2) adversarial training between generator and discriminator mitigates the class imbalance of the heart, improving the segmentation accuracy of the general "difficult" slices such

as the base and apex; 3) our model with fewer parameters get better performance compared with other methods in small datasets.

Finally, on the testing set of ACDC 2017, the proposed method is fairly accurate with geometric metrics and clinical metrics for LV, RV and MYO segmentation. To the best of our knowledge, our approach is generally superior to all advanced methods on ACDC 2017. Besides, the proposed model shows a powerful generalization performance without any fine-tuning. This brings hope for clinical application in automated bi-ventricle segmentation under the restriction of expensive annotation cost for medical images.

## REFERENCES

[1] E. J. Benjamin, "Heart disease and stroke statistics-2018 update: A report from the american heart association," *Circulation*, vol. 137, no. 12, pp. e467–e492, Mar. 2018.

[2] F. H. Epstein, "MRI of left ventricular function," *J. Nucl. Cardiol.*, vol. 14, no. 5, pp. 729–744, Sep. 2007.

[3] L. Bonnemains, D. Mandry, P.-Y. Marie, E. Micard, B. Chen, and P.-A. Vuissoz, "Assessment of right ventricle volumes and function by cardiac MRI: Quantification of the regional and global interobserver variability," *Magn. Reson. Med.*, vol. 67, no. 6, pp. 1740–1746, Jun. 2012.

[4] C. Petitjean and J.-N. Dacher, "A review of segmentation methods in short axis cardiac MR images," *Med. Image Anal.*, vol. 15, no. 2, pp. 169–184, 2011.

[5] V. Tavakoli and A. A. Amini, "A survey of shaped-based registration and segmentation techniques for cardiac images," *Comput. Vis. Image Understand.*, vol. 117, no. 9, pp. 966–989, Sep. 2013.

[6] S. Queirós, D. Barbosa, B. Heyde, P. Morais, J. L. Vilaça, D. Friboulet, O. Bernard, and J. D'hooge, "Fast automatic myocardial segmentation in 4D cine CMR datasets," *Med. Image Anal.*, vol. 18, no. 7, pp. 1115–1131, Oct. 2014.

[7] T. A. Ngo and G. Carneiro, "Left ventricle segmentation from cardiac MRI combining level set methods with deep belief networks," in *Proc. IEEE Int. Conf. Image Process.*, Sep. 2013, pp. 695–699.

[8] 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," *Magn. Reson. Imag.*, vol. 31, no. 4, pp. 575–584, 2013.

[9] M. R. 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," *Med. Image Anal.*, vol. 30, pp. 108–119, May 2016.

[10] Z. Guo, W. Tan, L. Wang, L. Xu, X. Wang, B. Yang, and Y. Yao, "Local motion intensity clustering (LMIC) model for segmentation of right ventricle in cardiac MRI images," *IEEE J. Biomed. Health Inform.*, vol. 23, no. 2, pp. 723–730, Mar. 2018.

[11] J. Chen, H. Zhang, W. Zhang, X. Du, Y. Zhang, and S. Li, "Correlated regression feature learning for automated right ventricle segmentation," *IEEE J. Transl. Eng. Health Med.*, vol. 6, 2018, Art. no. 1800610.

[12] P. V. Tran, "A fully convolutional neural network for cardiac segmentation in short-axis MRI," Apr. 2016, *arXiv:1604.00494*. [Online]. Available: https://arxiv.org/abs/1604.00494

[13] S. C. Mitchell, J. G. Bosch, B. P. F. Lelieveldt, R. J. Van der Geest, J. H. C. Reiber, and M. Sonka, "3-D active appearance models: Segmentation of cardiac MR and ultrasound images," *IEEE Trans. Med. Imag.*, vol. 21, no. 9, pp. 1167–1178, Sep. 2002.

[14] H. Zhang, A. Wahle, R. K. Johnson, T. D. Scholz, and M. Sonka, "4-D cardiac MR image analysis: Left and right ventricular morphology and function," *IEEE Trans. Med. Imag.*, vol. 29, no. 2, pp. 350–364, Feb. 2010.

[15] M. Lorenzo-Valdés, G. I. Sanchez-Ortiz, A. G. Elkington, R. H. Mohiaddin, and D. Rueckert, "Segmentation of 4D cardiac MR images using a probabilistic atlas and the EM algorithm," *Med. Image Anal.*, vol. 8, no. 3, pp. 255–265, Sep. 2004.

[16] A. Katouzian, A. Prakash, and E. Konofagou, "A new automated technique for left- and right-ventricular segmentation in magnetic resonance imaging," in *Proc. Int. Conf. IEEE Eng. Med. Biol. Soc.*, New York, NY, USA, Aug./Sep. 2006, pp. 3074–3077.

[17] J. Ulen, P. Strandmark, and F. Kahl, "An efficient optimization framework for multi-region segmentation based on Lagrangian duality," *IEEE Trans. Med. Imag.*, vol. 32, no. 2, pp. 178–188, Feb. 2013.

[18] G. Hautvast, S. Lobregt, M. Breeuwer, and F. Gerritsen, "Automatic contour propagation in cine cardiac magnetic resonance images," *IEEE Trans. Med. Imag.*, vol. 25, no. 11, pp. 1472–1482, Nov. 2006.

[19] F. Billet, M. Sermesant, H. Delingette, and N. Ayache, "Cardiac motion recovery and boundary conditions estimation by coupling an electromechanical model and cine-MRI data," in *Proc. Int. Conf. Funct. Imag. Modeling Heart*. Nice, France: Springer, Jun. 2009, pp. 376–385.

[20] O. Ronneberger, P. Fischer, and T. Brox, "U-net: Convolutional networks for biomedical image segmentation," in *Proc. Int. Conf. Med. Image Comput. Comput.-Assist. Intervent.* Granada, Spain: Springer, Sep. 2015, pp. 234–241.

[21] Ö. Çiçek, A. Abdulkadir, S. S. Lienkamp, T. Brox, and O. Ronneberger, "3D U-net: Learning dense volumetric segmentation from sparse annotation," in *Proc. Int. Conf. Med. Image Comput. Comput.-Assist. Intervent.*, Athens, Greece: Springer, Oct. 2016, pp. 424–432.

[22] C. Zotti, Z. Luo, A. Lalande, and P.-M. Jodoin, "Convolutional neural network with shape prior applied to cardiac MRI segmentation," *IEEE J. Biomed. Health Inform.*, vol. 23, no. 3, pp. 1119–1128, May 2018.

[23] Q. Dou, H. Chen, Y. Jin, L. Yu, J. Qin, and P.-A. Heng, "3D deeply supervised network for automatic liver segmentation from CT volumes," in *Proc. Int. Conf. Med. Image Comput. Comput.-Assist. Intervent.* Athens, Greece: Springer, Oct. 2016, pp. 149–157.

[24] C. Ye, W. Wang, S. Zhang, and K. Wang, "Multi-depth fusion network for whole-heart CT image segmentation," *IEEE Access*, vol. 7, pp. 23421–23429, 2019.

[25] L. Yu, X. Yang, H. Chen, J. Qin, and P. A. Heng, "Volumetric convnets with mixed residual connections for automated prostate segmentation from 3D MR images," in *Proc. 31st AAAI Conf. Artif. Intell.*, San Francisco, CA, USA, Feb. 2017, pp. 4–9.

[26] C. F. Baumgartner, L. M. Koch, M. Pollefeys, and E. Konukoglu, "An exploration of 2D and 3D deep learning techniques for cardiac MR image segmentation," in *Proc. Int. Workshop Stat. Atlases Comput. Models Heart*. Montreal, QC, Canada: Springer, Sep. 2017, pp. 111–119.

[27] X. Du, S. Yin, R. Tang, Y. Zhang, and S. Li, "Cardiac-DeepIED: Automatic pixel-level deep segmentation for cardiac bi-ventricle using improved end-to-end encoder-decoder network," *IEEE J. Transl. Eng. Health Med.*, vol. 7, 2019, Art. no. 1900110.

[28] W. Bai *et al.*, "Automated cardiovascular magnetic resonance image analysis with fully convolutional networks," *J. Cardiovascular Magn. Reson.*, vol. 20, no. 1, p. 65, 2018.

[29] Q. Zheng, N. Delingette, H. Duchateau, and N. Ayache, "3-D consistent and robust segmentation of cardiac images by deep learning with spatial propagation," *IEEE Trans. Med. Imag.*, vol. 37, no. 9, pp. 2137–2148, Sep. 2018.

[30] S. E. Petersen, P. M. Matthews, J. M. Francis, M. D. Robson, F. Zemrak, R. Boubertakh, A. A. Young, S. Hudson, P. Weale, S. Garratt, R. Collins, S. Piechnik, and S. Neubauer, "UK Biobank's cardiovascular magnetic resonance protocol," *J. Cardiovascular Magn. Reson.*, vol. 18, no. 1, Feb. 2016, Art. no. 8.

[31] I. Goodfellow, J. Pouget-Abadie, M. Mirza, B. Xu, D. Warde-Farley, S. Ozair, A. Courville, and Y. Bengio "Generative adversarial nets," in *Proc. Adv. Neural Inf. Process. Syst.*, Montreal, QC, Canada, Dec. 2014, pp. 2672–2680.

[32] S. Kazeminia, C. Baur, A. Kuijper, B. van Ginneken, N. Navab, S. Albarqouni, and A. Mukhopadhyay, "GANs for medical image analysis," Sep. 2018, *arXiv:1809.06222*. [Online]. Available: https://arxiv.org/abs/1809.06222

[33] M. Rezaei, H. Yang, and C. Meinel, "Whole heart and great vessel segmentation with context-aware of generative adversarial networks," in *Bildverarbeitung für die Medizin*. Berlin, Germany: Springer, 2018, pp. 353–358.

[34] M. Rezaei, H. Yang, and C. Meinel, "Recurrent generative adversarial network for learning imbalanced medical image semantic segmentation," *Multimedia Tools Appl.*, pp. 1–20, Feb. 2019.

[35] M. Mirza and S. Osindero, "Conditional generative adversarial nets," Nov. 2014, *arXiv:1411.1784*. [Online]. Available: https://arxiv.org/abs/1411.1784

[36] P. Luc, C. Couprie, S. Chintala, and J. Verbeek, "Semantic segmentation using adversarial networks," Nov. 2016, *arXiv:1611.08408*. [Online]. Available: https://arxiv.org/abs/1611.08408

[37] F. Milletari, N. Navab, and S.-A. Ahmadi, "V-net: Fully convolutional neural networks for volumetric medical image segmentation," in *Proc. 4th Int. Conf. 3D Vis.*, Stanford, CA, USA, Oct. 2016, pp. 565–571.

[38] M. Lin, Q. Chen, and S. Yan, "Network in network," Dec. 2013, *arXiv:1312.4400*. [Online]. Available: https://arxiv.org/abs/1312.4400

[39] V. Nair and G. E. Hinton, "Rectified linear units improve restricted Boltzmann machines," in *Proc. 27th Int. Conf. Mach. Learn.*, Haifa, Israel, Jun. 2010, pp. 807–814.

[40] S. Ioffe and C. Szegedy, "Batch normalization: Accelerating deep network training by reducing internal covariate shift," Feb. 2015, *arXiv:1502.03167*. [Online]. Available: https://arxiv.org/abs/1502.03167

[41] C. Peng, X. Zhang, G. Yu, G. Luo, and J. Sun, "Large kernel matters—Improve semantic segmentation by global convolutional network," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, Honolulu, HI, USA, Jul. 2017, pp. 4353–4361.

[42] H. Zhao, J. Shi, X. Qi, X. Wang, and J. Jia, "Pyramid scene parsing network," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, Honolulu, HI, USA, Jul. 2017, pp. 2881–2890.

[43] J. Fu, J. Liu, H. Tian, Y. Li, Y. Bao, Z. Fang, and H. Lu, "Dual attention network for scene segmentation," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, Long Beach, CA, USA, Jun. 2019, pp. 3146–3154.

[44] F. Yu and V. Koltun, "Multi-scale context aggregation by dilated convolutions," Nov. 2015, *arXiv:1511.07122*. [Online]. Available: https://arxiv.org/abs/1511.07122

[45] A. L. Maas, A. Y. Hannun, and A. Y. Ng, "Rectifier nonlinearities improve neural network acoustic models," in *Proc. Int. Conf. Mach. Learn.*, Atlanta, GA, USA, Jun. 2013, vol. 30, no. 1, p. 3.

[46] O. Bernard *et al.*, "Deep learning techniques for automatic MRI cardiac multi-structures segmentation and diagnosis: Is the problem solved?" *IEEE Trans. Med. Imag.*, vol. 37, no. 11, pp. 2514–2525, Nov. 2018.

[47] L. R. Dice, "Measures of the amount of ecologic association between species," *Ecology*, vol. 26, no. 3, pp. 297–302, 1945.

[48] D. P. Huttenlocher, G. A. Klanderman, and W. A. Rucklidge, "Comparing images using the Hausdorff distance," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 15, no. 9, pp. 850–863, Sep. 1993.

[49] X. Glorot and Y. Bengio, "Understanding the difficulty of training deep feedforward neural networks," in *Proc. Int. Conf. Artif. Intell. Statist.*, Setti Ballas CA, Italy, May 2010, pp. 249–256.

[50] D. P. Kingma and J. Ba, "Adam: A method for stochastic optimization," Dec. 2014, *arXiv:1412.6980*. [Online]. Available: https://arxiv.org/abs/1412.6980

[51] P. Moeskops, M. Veta, M. W. Lafarge, K. A. Eppenhof, and J. P. Pluim, "Adversarial training and dilated convolutions for brain MRI segmentation," in *Proc. Int. Workshop Deep Learn. Med. Image Anal.* Montreal, QC, Canada: Springer, Sep. 2017, pp. 56–64.

[52] Q. Tong, C. Li, W. Si, X. Liao, Y. Tong, Z. Yuan, and P. A. Heng, "RIANet: Recurrent interleaved attention network for cardiac MRI segmentation," *Comput. Biol. Med.*, vol. 109, pp. 290–302, Jun. 2019.

[53] M. Khened, V. A. Kollerathu, and G. Krishnamurthi, "Fully convolutional multi-scale residual densenets for cardiac segmentation and automated cardiac diagnosis using ensemble of classifiers," *Med. Image Anal.*, vol. 51, pp. 21–45, Jan. 2019.

[54] X. Zhuang, "Multivariate mixture model for myocardial segmentation combining multi-source images," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 41, no. 12, pp. 2933–2946, Dec. 2019.

[55] C. Petitjean *et al.*, "Right ventricle segmentation from cardiac MRI: A collation study," *Med. Image Anal.*, vol. 19, no. 1, pp. 187–202, 2015.

[56] F. Isensee, P.-F. Jaeger, P. M. Full, I. Wolf, S. Engelhardt, and K. H. Maier-Hein, "Automatic cardiac disease assessment on cine-MRI via time-series segmentation and domain specific features," in *Proc. Int. Workshop Stat. Atlases Comput. Models Heart*. Montreal, QC, Canada: Springer, Sep. 2017, pp. 120–129.

[57] Y. Jang, Y. Hong, S. Ha, S. Kim, and H.-J. Chang, "Automatic segmentation of LV and RV in cardiac MRI," in *Proc. Int. Workshop Stat. Atlases Comput. Models Heart*. Montreal, QC, Canada: Springer, Sep. 2017, pp. 161–169.

[58] C. Zotti, Z. Luo, O. Humbert, A. Lalande, and P.-M. Jodoin, "GridNet with automatic shape prior registration for automatic MRI cardiac segmentation," in *Proc. Int. Workshop Stat. Atlases Comput. Models Heart*. Montreal, QC, Canada: Springer, Sep. 2017, pp. 73–81.

[59] M. Khened, V. Alex, and G. Krishnamurthi, "Densely connected fully convolutional network for short-axis cardiac cine MR image segmentation and heart diagnosis using random forest," in *Proc. Int. Workshop Stat. Atlases Comput. Models Heart*. Montreal, QC, Canada: Springer, Sep. 2017, pp. 140–151.

[60] J. M. Wolterink, T. Leiner, M. A. Viergever, and I. Išgum, "Automatic segmentation and disease classification using cardiac cine MR images," in *Proc. Int. Workshop Stat. Atlases Comput. Models Heart*. Montreal, QC, Canada: Springer, Sep. 2017, pp. 101–110.

[61] J. Patravali, S. Jain, and S. Chilamkurthy, "2D-3D fully convolutional neural networks for cardiac MR segmentation," in *Proc. Int. Workshop Stat. Atlases Comput. Models Heart*. Montreal, QC, Canada: Springer, Sep. 2017, pp. 130–139.

[62] M.-M. Rohé, M. Sermesant, and X. Pennec, "Automatic multi-atlas segmentation of myocardium with SVF-Net," in *Proc. Int. Workshop Stat. Atlases Comput. Models Heart*. Montreal, QC, Canada: Springer, Sep. 2017, pp. 170–177.

[63] E. Grinias and G. Tziritas, "Fast fully-automatic cardiac segmentation in MRI using MRF model optimization, substructures tracking and B-spline smoothing," in *Proc. Int. Workshop Stat. Atlases Comput. Models Heart*. Montreal, QC, Canada: Springer, Sep. 2017, pp. 91–100.

[64] J. Ringenberg, M. Deo, V. Devabhaktuni, O. Berenfeld, P. Boyers, and J. Gold, "Fast, accurate, and fully automatic segmentation of the right ventricle in short-axis cardiac MRI," *Comput. Med. Imag. Graph.*, vol. 38, no. 3, pp. 190–201, Apr. 2014.

[65] K. Punithakumar, P. Boulanger, and M. Noga, "A GPU-accelerated deformable image registration algorithm with applications to right ventricular segmentation," *IEEE Access*, vol. 5, pp. 20374–20382, 2017.

[66] M. A. Zuluaga, M. J. Cardoso, and S. Ourselin, "Automatic right ventricle segmentation using multi-label fusion in cardiac MRI," in *Proc. Med. Image Comput. Comput.-Assist. Intervent., Workshop RV Segmentation Challenge in Cardiac MRI*, Nice, France, Oct. 2012, pp. 1–5.

[67] W. Bai, W. Shi, H. Wang, N. S. Peters, and D. Rueckert, "Multiatlas based segmentation with local label fusion for right ventricle MR images," in *Proc. Med. Image Comput. Comput.-Assist. Intervent., Workshop RV Segmentation Challenge Cardiac MRI*, Nice, France, Oct. 2012, vol. 6, p. 9.

**LIN QI** (M'08) received the B.S. and M.S. degrees from the School of Medicine, Tsinghua University, all in biomedical engineering, and the Ph.D. degree from the Department of Electronic Engineering, The Chinese University of Hong Kong. He is currently an Associate Professor with the Sino-Dutch Biomedical and Information Engineering School, Northeastern University, China. His current research interests include medical robotics and biomedical signal processing.
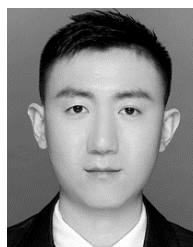
**HAORAN ZHANG** is currently pursuing the M.S. degree with the Sino-Dutch Biomedical and Information Engineering School, Northeastern University, China. His current research interests include medical imaging analysis and deep learning.
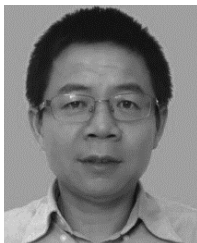
**WENJUN TAN** received the M.S. and the Ph.D. degrees in pattern recognition and intelligent system from Northeastern University, Shenyang, China, in 2007 and 2010, respectively. He is currently an Associate Professor with the School of Computer Science and Engineering, Northeastern University, China. He has authored or coauthored more than 50 international research articles, and holds 16 patents. His current research interests include medical image processing and computer aided diagnosis.

**SHOULIANG QI** received the Ph.D. degree in philosophy from Shanghai Jiao Tong University, in 2007. He joined the GE Global Research Center, where he was responsible for designing innovative magnetic resonance imaging (MRI) system. From 2014 to 2015, he was a Visiting Scholar with the Eindhoven University of Technology and the Epilepsy Center Kempenhaeghe, The Netherlands. He is currently an Associate Professor with the Sino-Dutch Biomedical and Information Engineering School, Northeastern University, China. His current research interests include the intelligent medical imaging computing and modeling, machine learning, brain networks, and brain models. He has published over 80 articles in peer-reviewed journals and international conferences. He received academic awards, such as the Chinese Excellent Ph.D. Dissertation Nomination Award and the Award for Outstanding Achievement in Scientific Research from the Ministry of Education.

**LISHENG XU** (SM'15) received the B.S. degree in electrical power system automation, the M.S. degree in mechanical electronics, and the Ph.D. degree in computer science and technology from the Harbin Institute of Technology, Harbin, China, in 1998, 2000, and 2006, respectively. He is currently a Full Professor with the Sino-Dutch Biomedical and Information Engineering School, Northeastern University, China. He has authored or coauthored more than 100 international research articles, and holds 15 patents and four pending patents. His current research interests include nonlinear medical signal processing, computational electromagnetic simulation, medical imaging, and pattern recognition. Prof. Xu is a Senior Member of the Chinese Society of Biomedical Engineering. He is the member of the editor board of many international journals, such as *Physiological Measurement*, *Biomedical Engineering Online*, and *Computers in Biology and Medicine*. He is the Director of Theory and Education Professional Committee of the China Medical Informatics Association.

**YUDONG YAO** (S'88–M'88–SM'94–F'11) received the B.Eng. and M.Eng. degrees from the Nanjing University of Posts and Telecommunications, Nanjing, China, in 1982 and 1985, respectively, and the Ph.D. degree from Southeast University, Nanjing, in 1988, all in electrical engineering. From 1987 to 1988, he was a Visiting Student with Carleton University, Ottawa, ON, Canada. From 1989 to 2000, he was with Carleton University, Spar Aerospace Ltd., Montreal, Canada, and Qualcomm Inc., San Diego, CA, USA. Since 2000, he has been with the Stevens Institute of Technology, Hoboken, NJ, USA, where he is currently a Professor and the Chair of the Department of Electrical and Computer Engineering. He holds one Chinese patent and 13 U.S. patents. His current research interests include wireless communications, machine learning and deep learning techniques, and healthcare and medical applications. He served as an Associate Editor for the IEEE Communications Letters, from 2000 to 2008, and the IEEE Transactions on Vehicular Technology, from 2001 to 2006, and an Editor for the IEEE Transactions on Wireless Communications, from 2001 to 2005. For his contributions to wireless communications systems, he was elected as a Fellow of the IEEE, in 2011, the National Academy of Inventors, in 2015, and the Canadian Academy of Engineering, in 2017.

**WEI QIAN** has many years of experience in biomedical imaging for breast cancer and lung cancer detection and diagnosis while he was with the H. Lee Moffitt Cancer Center and Research Institute. He is currently a Professor with the College of Engineering, The University of Texas at El Paso. He has published widely in the below mentioned research areas with more than 100 articles in peer-reviewed journals and premier international conferences with an H-index of 29, which is an important index for the translational research (the successful rate is 50%). He has six USA patents in which three of them have been licensed and commercialized by Kodak Company (now Carestream Health, Inc.). His current research interests include medical imaging, biomedical imaging, and molecular imaging, particularly, in computer-aided cancer detection and diagnosis or remote detection and diagnosis for digital mammography, X-ray lung images, lung CT images, cellular images and molecular images, and integrating intelligent techniques, including the use of neural networks, fuzzy logic, genetic algorithms, and evolution algorithms in simulation, modeling, and the design of high performance and robust systems.

• • •