IEEE Access
Multidisciplinary ⋮ Rapid Review ⋮ Open Access Journal

Received August 15, 2021, accepted September 15, 2021, date of publication November 18, 2021,
date of current version November 30, 2021.

Digital Object Identifier 10.1109/ACCESS.2021.3129333

Global Context and Enhanced Feature Guided Residual Refinement Network for 3D Cardiovascular Image Segmentation

JINGJING LIU[1], AO WEI[2], ZHIGANG GUO[1], AND CHANG TANG[3], (Member, IEEE)
[1]Department of Cardiac Surgery, Tianjin Chest Hospital, Tianjin 300222, China
[2]Department of Cardiology, Tianjin Chest Hospital, Tianjin 300222, China
[3]School of Computer Science, China University of Geosciences, Wuhan 430074, China

Corresponding author: Zhigang Guo (zhigangguo@yahoo.com)

This work was supported in part by the National Natural Science Foundation of Hubei Province under Grant 2020CFB644.

ABSTRACT As an important pre-processing step in clinical applications, automatic and accurate 3D cardiovascular image segmentation has attracted more and more attention. However, cardiovascular structures are often with high diversity, blood pool and myocardium shapes are also with large variability, and ambiguous cardiac borders make the segmentation task very challenging. In this paper, a novel deep neural network to segment the blood pool and myocardium from three dimensional cardiovascular images is introduced by fully exploiting the global context and complementary information encoded in different feature extraction layers, referred to as GCEFG-R^{2}Net briefly. In order to semantically locate the two kinds of regions in a global manner, we design a global context pooling module which can effectively learn context information in a global manner from the deep features extracted from the last two deep layers. Instead of directly using or combining different levels of deep features, we develop an interactive feature aggregation strategy to enhance different levels of deep features by embedding a series of interactive feature aggregation modules. By using the enhanced features, a residual feature refining branch is designed for refining the side outputs in a top-down stream with the guidance of global context features. Finally, the refined side outputs of different layers and the enhanced deep features are combined to generate the final segmentation result by using a feature fusion module. Extensive experiments on two challenge datasets are conducted to demonstrate that the proposed GCEFG-R^{2}Net can obtain appealing segmentation results for the blood pool and myocardium and performs better than other state-of-the-art methods.

INDEX TERMS Cardiovascular image segmentation, deep neural network, blood pool and myocardium segmentation.

I. INTRODUCTION

There are a large number of people that face the cardiovascular diseases each year in the world. Therefore, timely cardiovascular disease diagnosing and treatment is crucial [1]. As an intuitive manner, cardiovascular images can give detailed visual morphology presentation for the blood pool and the corresponding surrounding myocardium. Segmenting the heart in cardiovascular images plays an important and crucial role in cardiovascular disease diagnosing and treatment planning [2]–[4]. However, manually accomplishing this task is laborious, tedious and much time is needed, especially when medical resources are scarce. As a result, designing effective algorithms for accurately segmenting 3D cardiovascular images in an automatic manner is imperious.

In the past few decades, there are many methods proposed for segmenting the blood pool and surrounding myocardium from cardiovascular images. In general, there are two mainly kinds of methods for this task. One prominent family of methods that focus on multiple atlases and traditional deformable models [5]–[11] and the other family of methods based on deep neural networks (DNNs) [12]–[16]. As to the first kind of methods, the high anatomical variations in different parts should be taken into consideration. As to DNNs based methods, learning discriminative deep features is critical, and sufficient number of training data is also necessary to train an

The associate editor coordinating the review of this manuscript and approving it for publication was Vishal Srivastava.

This work is licensed under a Creative Commons Attribution 4.0 License. For more information, see https://creativecommons.org/licenses/by/4.0/

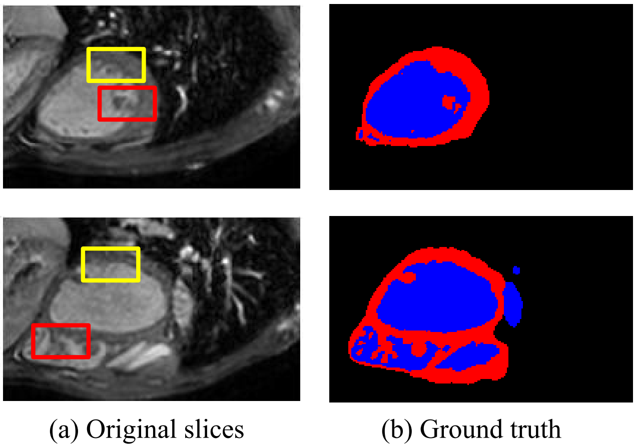


FIGURE 1. Some intuitive examples to show the challenging cases in cardiovascular image segmentation. In the ground truth images, the blue and red color denotes blood pool and myocardium, respectively.

effective network. Although great success has been achieved by previous proposed methods, some challenging issues still significantly influence the performance of different models and hinder their practical applications. E.g., the diversity of cardiovascular structures is often very high (As shown in the red rectangles of Figure 1), the shapes of blood pool and myocardium also vary widely, and ambiguous cardiac borders (as shown in the yellow rectangles of Figure 1), and the cardiac borders are ambiguous since the contrast between cardiac and the surrounding tissues is often very low.

In order to boost the performance of existing 3D cardiovascular image segmentation methods, we propose a novel deep neural network (GCEFG-R^{2}Net) which can automatically segment the blood pool and myocardium from cardiovascular images more accurately by fully exploiting the global context and complementary information encoded in different feature extraction layers. For capturing the heterogeneity of different parts of the cardiovascular image in a global manner, a global context pooling module is designed for learning image content context information from the deep features extracted from the last two deep layers. Instead of using original different levels of deep features, we develop an interactive feature aggregation strategy to enhance different levels of deep features, which can sufficiently obtain more efficient multi-scale information. Then, a hierarchical residual feature refining branch is designed by using the enhanced features to refine the side outputs in a top-down stream with the guidance of global context features generated from the global context pooling module. At last, the refined side outputs from each layer are fused by a feature fusion module to generate final segmentation result. During the fusion process, the enhanced deep features are also leveraged to boost the final segmentation map. In a nutshell, the major contributions of this work are as follows:

- We propose a new deep neural network for blood pool and myocardium segmentation from 3D cardiovascular images;
- In order to capture the heterogeneity from different parts of the cardiovascular image, we design a context pooling module to learn the cardiovascular image content context information from deep features; An interactive feature aggregation strategy is introduced to enhance different levels of deep features, which aims to obtain more efficient multi-scale information;
- A residual feature refining branch is designed for refining the segmentation result in a hierarchical and top-down manner. In addition, the learned global context features are used as a guidance for fusing the side output of each layer to get the final result;
- Extensive experiments are conducted to validate the superiority of the proposed network when compared with other state-of-the-art methods.

The rest of this paper are arranged as follows. Some related works about cardiovascular image segmentation will be firstly introduced in Section II. The detailed illustration of our proposed modules and network construction are elaborated in Section III and the experimental results with analysis are shown in Section V. In Section VI, we draw the conclusion of this work.

II. RELATED WORK

Medical images provide abundant information to help disease diagnosing and treatment, a large number of medical image segmentation methods have been designed during the past decades. In this work, we focus on automatic blood pool and myocardium segmentation from 3D cardiovascular images.

In the earlier years, segmentation methods rely heavily on multiple atlases and traditional deformable models. In [2], an interactive method is developed to accurately segment the cardiac chambers and vessels. However, since this method needs to be performed in an interactive manner, it is very slow and laborious. By combining the active appearance model and active shape model together, Mitchell *et al.* [6] introduced a hybrid approach to separate the right ventricle and left ventricle automatically. By using the Markov random field, a mixed flow model is designed to segment the right ventricle as well as generate the target shape priority [11]. As a classic model used for natural image segmentation, graph cut is also deployed for right ventricle segmentation [17] and myocardium segmentation [18]. Motivated by the prior knowledge from atlas and the 4D Markov random field, Lorenzo-Valds *et al.* [19] utilized the space-time background information for ventricle and myocardium segmentation from the cardiac MR image volume. Inspired by the shape discrepancy compensation principle, Liu *et al.* [20] found that the cross-constraint shape can be used to help segment the myocardium over delayed enhancement and T2-weight images. In order to exploit the multi-dimensional information [21], 3D random field model [22], multi-atlas as well as level sets [23] are also utilized (Atlas). In [24], random forest is used to learn context information, and the segmentation is obtained via appearance priori. Although these traditional methods based on atlases and deformable models make remarkable progress for cardiovascular image segmentation, their results are still not satisfactory due to the limitation of existing priori and extracted features.

Due to the powerful feature representation learning capability, DNNs boost the performance of a tremendous number of tasks such as computer vision, natural language processing and biomedical data analysis in a bid step [25], [26]. Therefore, DNN based methods have also been put forward for cardiovascular image segmentation. In [27], dilated CNN is used to demarcate blood pool and myocardium, while 3D volumetric information is neglected. In order to tackle this issue, Xu *et al.* [28] combined convolutional neural network (CNN) and recurrent neural network (RNN) to detect and segment the myocardial from fraction areas, which considers the 3D volumetric structure information. In [29], Payer *et al.* proposed a pipeline of two fully convolutional networks for automatic multi-label whole heart segmentation from CT and MRI volumes (MLWHS), which learns from the relative positions among labels and focuses on anatomically feasible configurations. Considering that fully convolutional neural network (FCN) can also obtain appealing performance for image segmentation, Qin *et al.* [30] proposed to use FCN for ventricles and myocardium segmentation. In addition, the motion state of the heart can be also well estimated. In order to preserve the maximum information flow between different deep feature extraction layers, Yu *et al.* [15] added the densely-connected mechanism into their network, and extra auxiliary side paths are embedded to strengthen the gradient propagation as well as stabilize the learning process. Since there are plenty of complementary information contained in multiple views of 3D cardiac data, Zheng *et al.* [31] utilized asymmetrical 3D kernels and pooling to capture contextual information. In order to avoid the domain shift in the field of biomedical image analysis, Dou *et al.* [32] proposed an unsupervised domain adaptation framework with adversarial learning for cross-modality biomedical image segmentation (UCMDA). By combining hybrid pyramid pooling and dilated residual learning, Du *et al.* [16] proposed a multi-task framework for joint blood pool and myocardium segmentation. In [33], both up-sampling and down-sampling strategies are used for blood vessels and the myocardium segmentation. Compared to traditional methods that rely on hand-crafted features or some pre-defined priori models, it is the fact methods based on deep leaning predominate the field of medical image segmentation and obtain better performance. To this end, we also focus on deep learning and propose a deep neural network to segment the blood pool and myocardium from 3D cardiovascular images.

III. PROPOSED NETWORK

In this section, we will illustrate the details of our proposed modules and the construction of the whole GCEFG-R^{2}Net, which consists of four main components including a global context pooling module (GCPM), an interactive feature aggregation module (IFAM), a hierarchical feature refining module (RFRM) and a deep feature guided feature fusion module (DFGFM). In Figure 2, we give an overview of our proposed GCEFG-R^{2}Net. During our network implementation, we use the 3D ResNeXt structure [34] as our feature extraction backbone and obtain five feature extraction layers. For simplicity, we denote the features extracted from the five layers as F_1, F_2, F_3, F_4 and F_5, respectively. Since the size of the slices used in the experiments is often small, in order to obtain the final accurate result, we first upsample all of the feature maps to the size of original input slice. In the following sections, we will elaborate each module of the proposed GCEFG-R^{2}Net in detail.

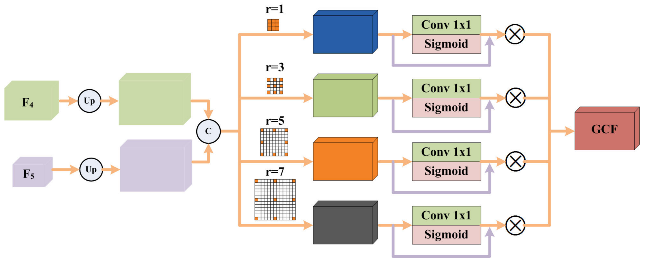


FIGURE 2. Flowchart of our proposed GCEFG-R^{2}Net. Considering that the blood pool and myocardium in an slice are often scattered with varying shapes, we first design a global context pooling module (GCPM) which can capture the global distribution information of image content. In order to fully exploit the complementary information of different layers of features, an interactive feature aggregation module (IFAM) is developed and embedded into the network for deep feature enhancing. Then a series of residual feature refining modules are designed and embedded in a hierarchical manner to refine the side outputs of different layers. Finally, the segmentation result is obtained by fusing all of the side outputs.

A. GCPM

As can be seen from the example images in Figure 1, the spatial distribution of different parts of an slice is often scattered with varying shapes. Therefore, it is important to capture the global context information to help locate different parts in the whole slice. Considering that the higher layers of the backbone network contains abundant semantic and context information [35]–[37], we use F_4 and F_5 to learn the global context information, as shown by Figure 3. Firstly, we concatenate the upsampled F_4 and F_5 together. Then, a series of convolutions with a hybrid dilation rate are used to learn the global context features (GCF). The ‘‘hybrid dilation rate assembled convolution’’ can be regarded as a manner to aggregate the input features locally. From Figure 3, we can see that the proposed GCPM is motivated from Atrous Spatial Pyramid Pooling (ASPP) [38]. However, our proposed GCPM differs ASPP significantly at least from following two aspects:

- Firstly, the channel attention is embedded into the proposed GCPM to adaptively fuse multi-scale information.
- Secondly, feature channel selection and receptive fields enlarging are performed simultaneously in our GCPM to exploit the global information for feature interaction. In addition, the GCF learned from GCPM is used as a

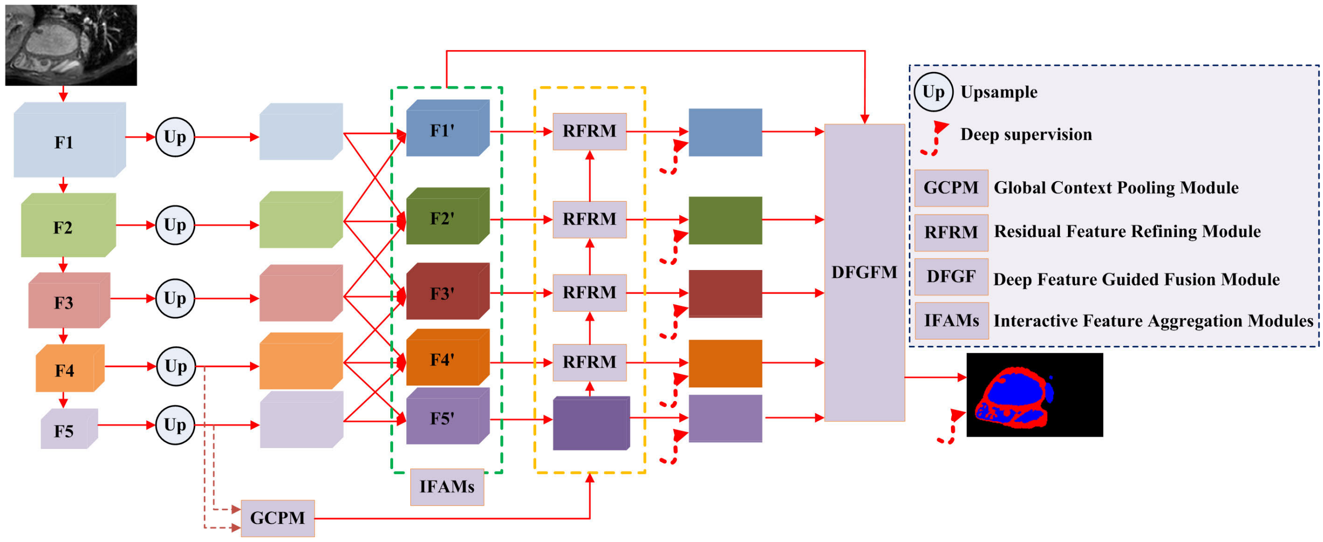

F1
F2
F3
F4
F5
Up
F1'
F2'
F3'
F4'
F5'
RFRM
DFGFM
IFAMs
GCPM
Up Upsample
Deep supervision
GCPM Global Context Pooling Module
RFRM Residual Feature Refining Module
DFGF Deep Feature Guided Fusion Module
IFAMs Interactive Feature Aggregation Modules

FIGURE 3. The architecture of the proposed GCPM.

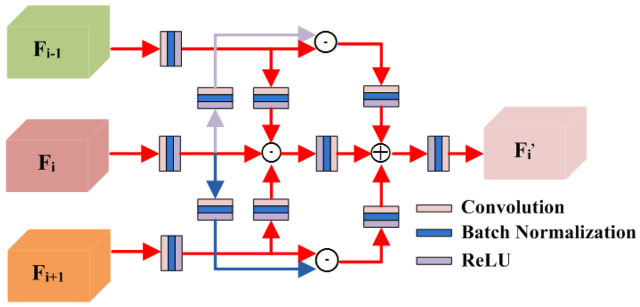


FIGURE 4. The architecture of the proposed IFAM.

guidance for feature refining which will be introduced in the later section.

B. IFAM

For the feature extraction backbone network, different layers of features reflect different degree of feature abstraction for original image slice. The shallow layers extract most of the details, the top layers often contain sufficient semantic and global context information, and the middle layers often contain both semantic and detailed information. It can help enhance the representation capability of different layers of features by exploiting their complementary information. Therefore, we design a series of interactive feature aggregation modules to aggregate the deep features in an interactive manner. Figure 4 gives a brief architecture of a IFAM corresponding to the *i*-th layer. As can be seen, except for the first and fifth layers, there are three input channels for each IFAM.

Without loss of generality, the input of the IFAM corresponding to the *i*-th layer consists of features F_i, F_{i+1} and F_{i-1}. For each input, we implement an initial transformation by a combination of a convolutional operation, a batch normalization operation and a ReLU operation, the channel number of initial features can be reduced. For feature aggregation, F_i, F_{i+1} and F_{i-1} interact with their corresponding layer of features. Finally, the three feature interaction branches are fused together to obtain the enhanced features, i.e., $F_i{}'$, which will be used for segmentation refinement.

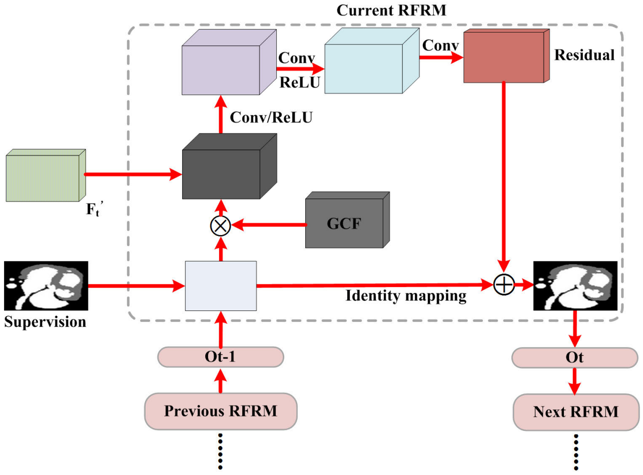


FIGURE 5. The architecture of the proposed RFRM.

As to the first and fifth IFAM, there are only two input branch, i.e., F_5 and F_4 for the first IFAM, F_1 and F_2 for the fifth IFAM. The enhanced features $F_5{}'$ and $F_1{}'$ can be learned in a similar way. In such a manner, each IFAM performs feature crossing to mitigate the discrepancy between different layers of features. The common parts of continuous feature layers are firstly extracted by element-wise multiplication and then original features are combined to capture complementary information by element-wise addition.

C. RFRM

By implementing a convolution operation on the enhanced feature maps $F_5{}'$, we can get an initial segmentation result, i.e., O_0. However, the resolution of the initial segmentation map is very low due to a series of pooling operation, and some detailed information such as the edges of different parts of original image content could be lost due to a series of continual pooling operations consisted in different feature extraction layers. Therefore, we design an RFRM and embed it into the proposed network for segmentation refinement in a hierarchical manner. In each RFRM, residual feature learning is used for refining the stage-wise segmentation map. The motivation of the proposed RFRM lies in two points. Firstly, previous literatures demonstrate that deep neural network embedded with residual feature learning can obtain better results in many computer vision problems when compared to commonly used plain network blocks. Secondly, gradient vanishing problem during deep neural network training can be effectively avoided and the training process of deep neural network can reach convergence faster, especially for the medical image segmentation task that the training samples are limited. In Figure 5, we present the detailed structure of an specific RFRM. In detail, we embed four RFRMs in our GCEFG-R^{2}Net. As to the t-th RFRM ($t = 1, 2, \cdots, 4$), there are three inputs, including the output of the $(t-1)$-th RFRM, i.e., O_{t-1}, enhanced feature maps F'_{5-t} and the GCF learned from the GCPM. Mathematically, the residual learned from the t-th RFRM can be formulated as follows:

$$R_t = \Psi(Cat(O_{t-1} \otimes GCF, F'_{5-t})), \;\; t = 1, 2, \cdots, 4, \tag{1}$$

where O_{t-1} denotes the output obtained from the $(t-1)$-th RFRM, $\Psi(\cdot)$ represents a mapping process which consists of a set of convolution and ReLU operations, $\otimes$ represents element-wise multiplication of feature maps and *Cat* is the channel-wise concatenation operation. By adding R_t with O_{t-1}, the output of RFRM can be calculated as follows:

$$O_t = R_t \oplus O_{t-1}, \quad t = 1, 2, \cdots, 4, \tag{2}$$

In addition, in order to improve the side output of each refining step during the training process, we add the supervision signal to each RFRM [39].

D. DFGFM

Since different layers of features reflect different abstract levels of original image slices, we develop a feature fusion module to fuse the segmentation maps generated from different RFRMs to obtain the final segmentation result. In addition, considering that information in original images is also important to help segmentation, we produce guiding deep features by using original enhanced features for guiding the final fusion process. The deep guided features **DGF** can be obtained as follows:

$$DGF = ReLU(W * Cat(F'_1, F'_2, \cdots, F'_5) + b), \tag{3}$$

where the enhanced feature maps of the i-th layer are denoted as F'_i. W and b are the convolution parameters that need to be learned during the training process and *ReLU* is the ReLU activation function [25]. Then, the final segmentation map O can be generated by following operations:

$$O = ReLU(W' * Cat(DGF, O_1, O_2, \cdots, O_5) + b'), \tag{4}$$

where W' and b' are also the convolution parameters.

IV. IMPLEMENTATION DETAILS

In our experiments, we implement the proposed GCEFG-R^{2}Net by using the PyTorch framework and we use the 3D ResNeXt [34] as backbone network for feature extraction. In this work, the mean square error (MSE) is used to compute the loss between the ground-truth G and outputs of the network, and the final loss function is formulated as follows:

$$\mathcal{L}(O, O_1, \cdots, O_5, G; \boldsymbol{\theta}) = \mathcal{L}_{mse}(O, G) + \sum_{i=1}^{5} \mathcal{L}^i_{mse}(O_i, G), \tag{5}$$

where $\mathcal{L}_{mse}(\cdot, \cdot)$ is a function to compute the MSE between two segmentation maps, $\mathcal{L}_{mse}(O, G)$ is the MSE between the final fused output and the ground-truth, $\mathcal{L}^i_{mse}(O_i, G)$ is the MSE between the i-th layer-wise side-output and the ground-truth. The definition of MSE can be formulated as follows:

$$\mathcal{L}_{mse}(O, G) = \frac{1}{WH} \sum_{i=1}^{W} \sum_{j=1}^{H} |O(i, j) - G(i, j)| \tag{6}$$

Our network is trained in an end-to-end manner by using the Adam algorithm with the initial learning rate of 0.001 on a single Nvidia Titan V GPU with 12Gb memory. We train the network with the "poly" learning rate policy and the training data are also augmented to reduce over-fitting, the training batch size is fixed to 4.

V. EXPERIMENTAL RESULTS

In this section, we report the segmentation results of the proposed network on two datasets including the 2016 HVSMR dataset [2] and the 2017 MM-WHS CT dataset [40]. In addition, we also compare our network with other state-of-the-art ones to validate its superiority.

TABLE 1. The statistic information of the two datasets used in our experiments. "Yes" means the ground-truth of the data are publicly released while "No" means the ground-truth of the data are not publicly released.

Dataset	Training		Testing		# Class
	# stack	GT	# stack	GT	
2016 HVSMR	10	Yes	10	No	2
2017 MM-WHS CT	16	Yes	4	Yes	7

A. DATASETS

The 2016 HVSMR dataset [2] aims to segment myocardium and blood pool from cardiovascular MR images. There are 10 patients with 3D Cardiovascular MR images including 10 training sets and 10 testing sets. For different patients, the

TABLE 2. Quantitative segmentation evaluation results of different methods on the 2016 HVSMR dataset.

			Blood Pool			
Method	Dice	Jac	PPV	Sens	Spec	HD
Atlas	0.873±0.015	0.818±0.021	0.878±0.024	0.867±0.026	0.936±0.004	25.647±7.215
U-Net	0.927±0.013	0.864±0.023	0.927±0.026	0.928±0.021	0.985±0.005	7.993±6.121
SDNet	0.901±0.021	0.820±0.034	0.897±0.046	0.907±0.035	0.979±0.010	20.929±7.278
SSLLN	0.908±0.017	0.832±0.029	0.909±0.036	0.908±0.030	0.982±0.007	19.656±9.367
HFA-Net	0.931±0.017	0.886±0.029	0.932±0.032	0.926±0.029	0.987±0.006	4.106±3.213
DRHPPN	0.946±0.008	0.898±0.014	0.948±0.017	0.945±0.011	0.989±0.003	2.082±1.035
GCEFG-R^{2}Net	0.958±0.011	0.914±0.021	0.956±0.023	0.956±0.014	0.994±0.004	1.847±1.003
			Myocardium			
Atlas	0.731±0.035	0.624±0.047	0.773±0.041	0.704±0.35	0.938±0.003	16.528±1.497
U-Net	0.788±0.034	0.651±0.046	0.838±0.042	0.743±0.031	0.990±0.002	5.627±1.542
SDNet	0.722±0.044	0.566±0.054	0.817±0.061	0.649±0.054	0.989±0.004	13.263±5.120
SSLLN	0.745±0.045	0.596±0.056	0.800±0.041	0.699±0.057	0.987±0.002	9.063±4.492
HFA-Net	0.792±0.034	0.683±0.042	0.833±0.064	0.752±0.049	0.991±0.003	12.246±3.854
DRHPPN	0.824±0.027	0.701±0.039	0.883±0.020	0.772±0.037	0.992±0.001	11.297±2.682
GCEFG-R^{2}Net	0.836±0.031	0.721±0.028	0.896±0.024	0.791±0.039	0.996±0.002	9.834±3.817

numbers of MR images are different. This dataset can be obtained from http://segchd.csail.mit.edu/index.html.

The 2017 MM-WHS CT dataset [40] aims to evaluate algorithms that segment seven cardiac structures, i.e., the left/right ventricle blood cavity (LV/RV), left/right atrium blood cavity (LA/RA), myocardium of the left ventricle (LV-myo), ascending aorta (AO), and pulmonary artery (PA). Similar to [31], we randomly split the dataset into the training subset (16 subjects) and testing subset (4 subjects) by following the work in [32].

In Table 1, we present the details of the two datasets.

B. EXPERIMENTS ON THE 2016 HVSMR DATASET

Since the segmentation ground-truth of testing sets are not available and the challenge submission system also dose not work for online testing, we divide the training sets into training subsets and a validation subset. In this experiment, we use leave-one-out setting for performance evaluation of our network. For each patient, the corresponding images are used for testing and the images of other patients are used for training. Therefore, there are totally 10 repeated training and testing times. Six indicators including Dice coefficient (Dice), Jaccard coefficient (Jac), positive predictive value (PPV), sensitivity (Sens), specificity (Spec), and Hausdorff distance of boundaries (HD) are used to evaluate the performance of our different methods.

In Figure 6, we show the six indicators of the segmentation results of 10 subjects in this dataset. As can be seen, for the blood pool, the segmentation accuracy is much higher than the myocardium, which indicates that segmenting the myocardium is more difficult than the blood pool. This is an also challenging problem faced by previous segmentation methods, which is caused by the fact that the myocardium areas in medical images are often small, scattered and with varying shapes. As to the HD indicator, the score of the blood pool is much smaller than that of the myocardium. In most cases, our network can obtain stable results for different patients for the blood pool.

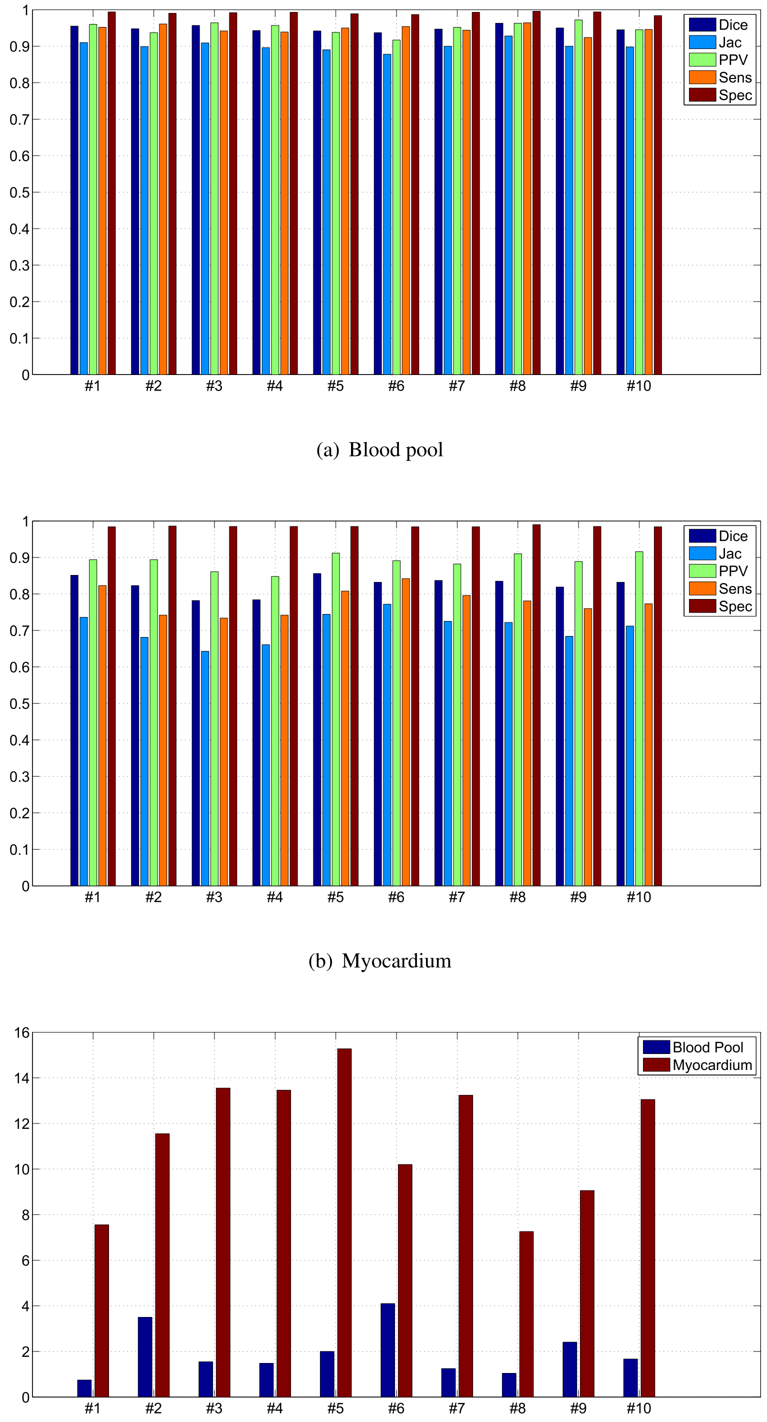
In order to demonstrate our proposed network can obtain better segmentation results, we also compare it with traditional method Atlas [23] and some classical segmentation network, i.e., U-Net [13]. In addition, previous cardiac image segmentation networks including the SSLLN [41], SDNet [42], HFA-Net [31] and DRHPPN [16] are also used for comparison. In Table 2, we report the results of different methods in terms of different indicators and the results also demonstrate that our proposed network performs better than other state-of-the-art ones.

C. EXPERIMENTS ON THE 2017 mm-WHS CT DATASET

For this dataset, the ground-truth segmentation maps for both training subjects and testing subjects are available. As a result, we use the training samples for network training and test it on the testing subset. Four indicators including Dice coefficient (Dice), Jaccard coefficient (Jac), average surface distance (ADB) and Hausdorff distance of boundaries (HD) are used for evaluation. We compare the proposed GCEFG-R^{2}Net with HFA-Net [31] as well as its baselines. In addition, we also compare the Dice indicator with [32] and [29] which can be obtained from [31]. The results of different method on this dataset are shown in Table 3, which also validate the efficacy of our proposed network.

D. ABLATION STUDIES

As mentioned in previous sections, there are two critical modules for our proposed network, i.e., GCPM which learns the global context information that can be used to guide the feature refining and the IFAM which enhances layer-wise features in a cross layer manner. In order to validate the influence of the two modules for the final results, we remove GCPM and IFAMs respectively from GCEFG-R^{2}Net (denoted as noGCPM and noIFAM) and perform experiments on the 2016 HVSMR dataset and report the results in Table 4. As can be seen, when GCPM is removed, the results degenerate significantly, which validate that the global context information is critical for final segmentation. In order to give a more intuitive



(a) Blood pool

(b) Myocardium

(c) HD

FIGURE 6. Six indicators of the segmentation results of 10 subjects in the 2016 HVSMR dataset.

demonstration, we also show the visual segmentation results without the two modules in Figure 7. As can be seen, when the GCPM module is removed, the segmentation refining process is similar to tradition U-Net. Without GCPM module, the global context information cannot be well embedded for feature refining, which produces some missed regions in the final results, as shown by the row titled "noGCPM" in Figure 7. When the IFAM module is removed, shallow features and deep features are not well aggregated, which induces incomplete segmentation results, especially for the

TABLE 3. Segmentation results on the 2017 MM-WHS CT dataset in terms of four indicators.

Method	Indicators	Structures						
		LV	RV	LA	RA	LV-myo	AO	PA
UCMDA	Dice	0.888	–	0.891	–	0.733	0.813	–
MLWHS	Dice	0.918	0.909	0.929	0.888	0.881	0.933	0.900
DVN	Dice	0.942	0.891	0.933	0.879	0.908	0.959	0.824
	Jac	0.891	0.806	0.874	0.786	0.832	0.922	0.713
	ADB	0.084	0.448	0.199	0.459	0.180	0.132	1.710
	HD	6.752	39.156	71.189	101.57	35.422	27.81	59.982
S-DVN	Dice	0.929	0.89	0.914	0.899	0.895	0.956	0.828
	Jac	0.870	0.805	0.843	0.817	0.811	0.916	0.718
	ADB	0.610	0.666	1.384	0.307	0.362	0.210	1.594
	HD	21.214	55.473	85.726	73.757	62.053	80.511	77.181
HFA-Net	Dice	0.946	0.893	0.925	0.897	0.91	0.964	0.830
	Jac	0.898	0.810	0.861	0.816	0.836	0.930	0.722
	ADB	0.076	0.562	0.210	0.334	0.225	0.103	1.685
	HD	7.148	33.128	42.173	22.903	36.954	12.075	37.845
GCEFG-R^{2}Net	Dice	0.951	0.902	0.937	0.903	0.918	0.968	0.840
	Jac	0.901	0.818	0.874	0.819	0.845	0.937	0.728
	ADB	0.071	0.435	0.187	0.301	0.177	0.100	1.583
	HD	6.725	31.059	40.169	21.097	33.024	11.206	35.894

TABLE 4. Ablation experimental results on the 2016 HVSMR dataset.

Blood Pool						
Method	Dice	Jac	PPV	Sens	Spec	HD
L0	0.928±0.006	0.834±0.014	0.901±0.016	0.915±0.023	0.947±0.004	6.312±1.258
L1	0.897±0.007	0.847±0.014	0.893±0.015	0.906±0.021	0.958±0.004	7.341±1.241
L2	0.913±0.007	0.839±0.013	0.911±0.014	0.897±0.022	0.965±0.003	8.671±1.247
L3	0.899±0.006	0.849±0.012	0.920±0.016	0.903±0.021	0.976±0.004	7.457±1.247
L4	0.904±0.007	0.857±0.012	0.924±0.014	0.917±0.023	0.977±0.005	6.237±1.246
noGCPM	0.931±0.007	0.876±0.013	0.932±0.015	0.928±0.021	0.981±0.005	5.654±1.249
noIFAM	0.940±0.008	0.885±0.012	0.939±0.014	0.934±0.017	0.986±0.003	3.201±1.871
GCEFG-R^{2}Net	0.958±0.011	0.914±0.021	0.956±0.023	0.956±0.014	0.994±0.004	1.847±1.003
Myocardium						
L0	0.754±0.027	0.697±0.025	0.835±0.045	0.748±0.031	0.947±0.005	13.314±2.874
L1	0.743±0.028	0.684±0.026	0.834±0.048	0.735±0.033	0.932±0.003	14.547±2.367
L2	0.771±0.026	0.676±0.025	0.842±0.054	0.741±0.034	0.956±0.004	13.624±2.894
L3	0.752±0.027	0.702±0.027	0.851±0.048	0.756±0.031	0.966±0.002	13.614±2.678
L4	0.765±0.027	0.698±0.028	0.857±0.049	0.757±0.032	0.968±0.005	13.542±2.732
noGCPM	0.784±0.028	0.691±0.027	0.866±0.056	0.764±0.033	0.978±0.004	12.206±2.942
noIFAM	0.806±0.026	0.706±0.024	0.871±0.030	0.785±0.032	0.981±0.003	10.657±2.481
GCEFG-R^{2}Net	0.836±0.031	0.721±0.028	0.896±0.024	0.791±0.039	0.996±0.002	9.834±3.817

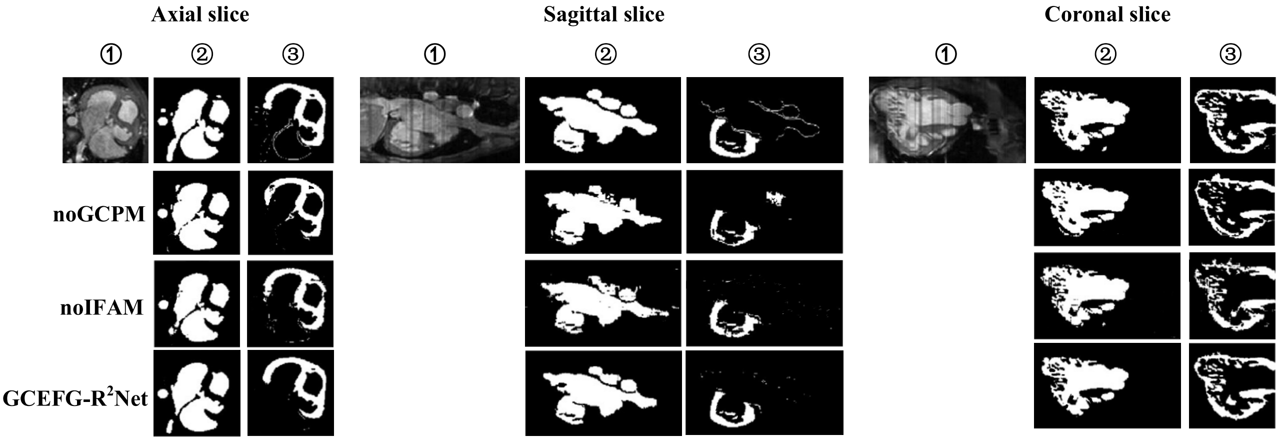


FIGURE 7. Intuitive segmentation results of ablation studies. ① represent original images from different views, ② represent the separated binary results of blood pool and ③ represent the segmentation results of myocardium.

details information, as shown by the row titled "noIFAM" in Figure 7.

In addition, In order to demonstrate the efficacy of different layers of our proposed network, we report the results of different layers before the final DFGF module, which are denoted by L0, L1, L2, L3 and L4, respectively. We show the results in above Table 4. As can been seen, by aggregating the final segmentation results of different layers, the DFGF module can capture the complementary information of layer-wise features and side-output results to generate better final segmentation map.

VI. CONCLUSION

In this paper, we introduce a deep neural network for segmenting blood pool and myocardium from 3D cardiovascular images. In order to capture the global context information of the two kinds of regions, a global context pooling module is designed to learn the context information from the deep features extracted from the last two deep layers of backbone network. Rather than directly using or combining different levels of deep features, we design an interactive feature aggregation strategy to enhance different levels of deep features by embedding a series of interactive feature aggregation modules. Extensive experiments as well as ablation analysis are conducted on two public datasets to validate the efficacy of the proposed network, which can obtain higher segmentation accuracy.

AVAILABILITY OF DATA AND MATERIALS

The datasets used in our experiments are available from http://segchd.csail.mit.edu/ and https://zmiclab.github.io/projects/mmwhs/, respectively.

REFERENCES

[1] 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.

[2] D. F. Pace, A. V. Dalca, T. Geva, A. J. Powell, M. H. Moghari, and P. Golland, "Interactive whole-heart segmentation in congenital heart disease," in *Proc. Int. Conf. Med. Image Comput. Comput.-Assist. Intervent.* Springer, 2015, pp. 80–88.

[3] T. Liu, Y. Tian, S. Zhao, X. Huang, and Q. Wang, "Residual convolutional neural network for cardiac image segmentation and heart disease diagnosis," *IEEE Access*, vol. 8, pp. 82153–82161, 2020.

[4] M. Jacobs, M. Benovoy, L.-C. Chang, D. Corcoran, C. Berry, A. E. Arai, and L.-Y. Hsu, "Automated segmental analysis of fully quantitative myocardial blood flow maps by first-pass perfusion cardiovascular magnetic resonance," *IEEE Access*, vol. 9, pp. 52796–52811, 2021.

[5] X. Zhuang, "Challenges and methodologies of fully automatic whole heart segmentation: A review," *J. Healthcare Eng.*, vol. 4, no. 3, pp. 371–407, 2013.

[6] S. C. Mitchell, B. P. F. Lelieveldt, R. J. van der Geest, H. G. Bosch, J. H. C. Reiver, and M. Sonka, "Multistage hybrid active appearance model matching: Segmentation of left and right ventricles in cardiac MR images," *IEEE Trans. Med. Imag.*, vol. 20, no. 5, pp. 415–423, May 2001.

[7] 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.

[8] O. M. Maier, D. Jiménez, A. Santos, and M. J. Ledesma-Carbayo, "Segmentation of RV in 4D cardiac MR volumes using region-merging graph cuts," in *Proc. Comput. Cardiol.*, 2012, pp. 697–700.

[9] D. Mahapatra and J. M. Buhmann, "Automatic cardiac RV segmentation using semantic information with graph cuts," in *Proc. IEEE 10th Int. Symp. Biomed. Imag.*, Apr. 2013, pp. 1106–1109.

[10] J. Woo, P. J. Slomka, C.-C.-J. Kuo, and B.-W. Hong, "Multiphase segmentation using an implicit dual shape prior: Application to detection of left ventricle in cardiac MRI," *Comput. Vis. Image Understand.*, vol. 117, no. 9, pp. 1084–1094, Sep. 2013.

[11] O. Moolan-Feroze, M. Mirmehdi, M. Hamilton, and C. Bucciarelli-Ducci, "Segmentation of the right ventricle using diffusion maps and Markov random fields," in *Proc. Int. Conf. Med. Image Comput. Comput.-Assist. Intervent.* Springer, 2014, pp. 682–689.

[12] O. 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.* Springer, 2016, pp. 424–432.

[13] 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.* Springer, 2015, pp. 234–241.

[14] 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.* Springer, 2016, pp. 149–157.

[15] L. Yu, J.-Z. Cheng, Q. Dou, X. Yang, H. Chen, J. Qin, and P.-A. Heng, "Automatic 3D medical image segmentation with densely-connected volumetric ConvNets," in *Proc. Int. Conf. Med. Image Comput. Comput.-Assist. Intervent.* Springer, 2017, pp. 287–295.

[16] X. Du, Y. Song, Y. Liu, Y. Zhang, H. Liu, B. Chen, and S. Li, "An integrated deep learning framework for joint segmentation of blood pool and myocardium," *Med. Image Anal.*, vol. 62, May 2020, Art. no. 101685.

[17] D. Grosgeorge, C. Petitjean, J.-N. Dacher, and S. Ruan, "Graph cut segmentation with a statistical shape model in cardiac MRI," *Comput. Vis. Image Understand.*, vol. 117, no. 9, pp. 1027–1035, 2013.

[18] X. Song, Y. Wang, Q. Feng, and Q. Wang, "Improved graph cut model with features of superpixels and neighborhood patches for myocardium segmentation from ultrasound image," *Math. Biosci. Eng.*, vol. 16, no. 3, p. 1115, 2019.

[19] 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.

[20] J. Liu, H. Xie, S. Zhang, and L. Gu, "Multi-sequence myocardium segmentation with cross-constrained shape and neural network-based initialization," *Computerized Med. Imag. Graph.*, vol. 71, pp. 49–57, Jan. 2019.

[21] C. Tang, X. Zheng, X. Liu, W. Zhang, J. Zhang, J. Xiong, and L. Wang, "Cross-view locality preserved diversity and consensus learning for multi-view unsupervised feature selection," *IEEE Trans. Knowl. Data Eng.*, early access, Jan. 1, 2021, doi: 10.1109/TKDE.2020.3048678.

[22] G. Tziritas, "Fully-automatic segmentation of cardiac images using 3-D MRF model optimization and substructures tracking," in *Reconstruction, Segmentation, and Analysis of Medical Images*. Springer, 2016, pp. 129–136.

[23] R. Shahzad, S. Gao, Q. Tao, O. Dzyubachyk, and R. van der Geest, "Automated cardiovascular segmentation in patients with congenital heart disease from 3D CMR scans: Combining multi-atlases and level-sets," in *Reconstruction, Segmentation, and Analysis of Medical Images*. Springer, 2016, pp. 147–155.

[24] A. Mukhopadhyay, "Total variation random forest: Fully automatic MRI segmentation in congenital heart diseases," in *Reconstruction, Segmentation, and Analysis of Medical Images*. Springer, 2016, pp. 165–171.

[25] A. Krizhevsky, I. Sutskever, and G. E. Hinton, "ImageNet classification with deep convolutional neural networks," in *Proc. Adv. Neural Inf. Process. Syst. (NIPS)*, 2012, pp. 1097–1105.

[26] K. Simonyan and A. Zisserman, "Very deep convolutional networks for large-scale image recognition," in *Proc. Int. Conf. Represent. Learn.*, 2015, pp. 1–14.

[27] J. M. Wolterink, T. Leiner, M. A. Viergever, and I. Išgum, "Dilated convolutional neural networks for cardiovascular mr segmentation in congenital heart disease," in *Reconstruction, Segmentation, and Analysis of Medical Images*. Springer, 2016, pp. 95–102.

[28] C. Xu, L. Xu, Z. Gao, S. Zhao, H. Zhang, Y. Zhang, X. Du, S. Zhao, D. Ghista, and S. Li, "Direct detection of pixel-level myocardial infarction areas via a deep-learning algorithm," in *Proc. Int. Conf. Med. Image Comput. Comput.-Assist. Intervent.* Springer, 2017, pp. 240–249.

[29] C. Payer, D. Štern, H. Bischof, and M. Urschler, "Multi-label whole heart segmentation using CNNs and anatomical label configurations," in *Proc. Int. Workshop Stat. Atlases Comput. Models Heart*. Springer, 2017, pp. 190–198.

[30] C. Qin, W. Bai, J. Schlemper, S. E. Petersen, S. K. Piechnik, S. Neubauer, and D. Rueckert, "Joint learning of motion estimation and segmentation for cardiac MR image sequences," in *Proc. Int. Conf. Med. Image Comput. Comput.-Assist. Intervent.* Springer, 2018, pp. 472–480.

[31] H. Zheng, L. Yang, J. Han, Y. Zhang, P. Liang, Z. Zhao, C. Wang, and D. Z. Chen, "HFA-Net: 3D cardiovascular image segmentation with asymmetrical pooling and content-aware fusion," in *Proc. Int. Conf. Med. Image Comput. Comput.-Assist. Intervent.* Springer, 2019, pp. 759–767.

[32] Q. Dou, C. Ouyang, C. Chen, H. Chen, and P.-A. Heng, "Unsupervised cross-modality domain adaptation of ConvNets for biomedical image segmentations with adversarial loss," in *Proc. 27th Int. Joint Conf. Artif. Intell.*, Jul. 2018, pp. 691–697.

[33] A. Taha, P. Lo, J. Li, and T. Zhao, "Kid-Net: Convolution networks for kidney vessels segmentation from CT-volumes," in *Proc. Int. Conf. Med. Image Comput. Comput.-Assist. Intervent.* Springer, 2018, pp. 463–471.

[34] S. Xie, R. Girshick, P. Dollár, Z. Tu, and K. He, "Aggregated residual transformations for deep neural networks," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jul. 2017, pp. 1492–1500.

[35] C. Tang, X. Liu, X. Zheng, W. Li, J. Xiong, L. Wang, A. Zomaya, and A. Longo, "DeFusionNET: Defocus blur detection via recurrently fusing and refining discriminative multi-scale deep features," *IEEE Trans. Pattern Anal. Mach. Intell.*, early access, Aug. 6, 2020, doi: 10.1109/TPAMI.2020.3014629.

[36] C. Tang, X. Liu, X. Zhu, E. Zhu, K. Sun, P. Wang, L. Wang, and A. Zomaya, "R^{2}MRF: Defocus blur detection via recurrently refining multi-scale residual features," in *Proc. AAAI Conf. Artif. Intell.*, 2020, vol. 34, no. 7, pp. 12063–12070.

[37] C. Tang, X. Liu, S. An, and P. Wang, "BR^{2}Net: Defocus blur detection via a bidirectional channel attention residual refining network," *IEEE Trans. Multimedia*, vol. 23, pp. 624–635, 2020.

[38] L.-C. Chen, G. Papandreou, F. Schroff, and H. Adam, "Rethinking atrous convolution for semantic image segmentation," 2017, *arXiv:1706.05587*.

[39] S. Xie and Z. Tu, "Holistically-nested edge detection," in *Proc. IEEE Int. Conf. Comput. Vis. (ICCV)*, Dec. 2015, pp. 1395–1403.

[40] X. Zhuang and J. Shen, "Multi-scale patch and multi-modality atlases for whole heart segmentation of MRI," *Med. Image Anal.*, vol. 31, pp. 77–87, Jul. 2016.

[41] J. Duan, "Automatic 3D bi-ventricular segmentation of cardiac images by a shape-refined multi-task deep learning approach," *IEEE Trans. Med. Imag.*, vol. 38, no. 9, pp. 2151–2164, Jan. 2019.

[42] A. Chartsias, T. Joyce, G. Papanastasiou, S. Semple, M. Williams, D. E. Newby, R. Dharmakumar, and S. A. Tsaftaris, "Disentangled representation learning in cardiac image analysis," *Med. Image Anal.*, vol. 58, Dec. 2019, Art. no. 101535.

JINGJING LIU received the B.S. degree in clinical medicine and the master's degree in cardiosurgery from Tianjin Medical University, Tianjin, China, in 2012 and 2018, respectively. She is currently an Attending Doctor with the Department of Cardiac Surgery, Tianjin Chest Hospital, Tianjin. Her current research interests include congenital heart disease, structural heart disease, and translational medicine.

AO WEI received the B.S. degree in clinical medicine and the master's degree in cardiology from Tianjin Medical University, Tianjin, China, in 2011 and 2019, respectively. He is currently an Attending Doctor with the Department of Cardiology, Tianjin Chest Hospital, Tianjin. His current research interests include coronary atherosclerotic heart disease and arrhythmia.

ZHIGANG GUO received the B.S. degree in clinical medicine from Tianjin Medical University, Tianjin, China, in 1989, and the master's degree in cardiosurgery from the Peking Union Medical College Hospital, Beijing, China, in 2008. He is currently the President of Tianjin Chest Hospital, Tianjin, and also a Doctoral Supervisor of Tianjin Medical University and Tianjin University. His current research interests include coronary atherosclerotic heart disease, structural heart disease, and translational medicine.

CHANG TANG (Member, IEEE) received the Ph.D. degree from Tianjin University, Tianjin, China, in 2016. He joined the AMRL Laboratory, University of Wollongong, between September 2014 and September 2015. He is currently a Full Professor with the School of Computer Science, China University of Geosciences, Wuhan, China. He has published more than 50 peer-reviewed papers, including those in highly regarded journals and conferences, such as IEEE T-PAMI, IEEE T-MM, IEEE T-KDE, IEEE T-HMS, ICCV, CVPR, ICML, IJCAI, AAAI, and ACM MM. His current research interests include machine learning and computer vision. He serves as an Associate Editor of *BioMed Research International*, *BMC Bioinformatics*, and a Young Editor of *CAAI Transactions on Intelligence Technology and Computer Engineering*. He often serves on the Technical Program Committees of some top conferences, such as NIPS, CVPR, ICML, ICCV, ECCV, IJCAI, ICME, and AAAI.