HFA-Net: 3D Cardiovascular Image Segmentation with Asymmetrical Pooling and Content-Aware Fusion

Hao Zheng, Lin Yang, Jun Han, Yizhe Zhang,
Peixian Liang, Zhuo Zhao, Chaoli Wang, and Danny Z. Chen

Department of Computer Science and Engineering,
University of Notre Dame, Notre Dame, IN 46556, USA
hzheng3@nd.edu

Abstract. Automatic and accurate cardiovascular image segmentation is important in clinical applications. However, due to ambiguous borders and subtle structures (e.g., thin myocardium), parsing fine-grained structures in 3D cardiovascular images is very challenging. In this paper, we propose a novel deep *heterogeneous feature aggregation network* (HFA-Net) to fully exploit complementary information from multiple views of 3D cardiac data. First, we utilize asymmetrical 3D kernels and pooling to obtain heterogeneous features in parallel encoding paths. Thus, from a specific view, distinguishable features are extracted and indispensable contextual information is kept (rather than quickly diminished after symmetrical convolution and pooling operations). Then, we employ a content-aware multi-planar fusion module to aggregate meaningful features to boost segmentation performance. Further, to reduce the model size, we devise a new DenseVoxNet model by sparsifying residual connections, which can be trained in an end-to-end manner. We show the effectiveness of our new HFA-Net on the 2016 HVSMR and 2017 MM-WHS CT datasets, achieving state-of-the-art performance. In addition, HFA-Net obtains competitive results on the 2017 AAPM CT dataset, especially on segmenting subtle structures among multi-objects with large variations, illustrating the robustness of our new segmentation approach.

1 Introduction

Cardiovascular diseases are a leading cause of death globally. Segmenting the whole heart in cardiovascular images is a prerequisite for morphological and pathological analysis, disease diagnosis, and surgical planning [6]. However, automatic and accurate cardiovascular image segmentation remains very challenging due to large variations in different subjects, missing/ambiguous borders, and inhomogeneous appearance and image quality (e.g., see Fig. 1(a-b)).

Recent studies showed that deep learning based methods [11, 2–4, 12] can learn robust contextual and semantic features and achieve state-of-the-art segmentation performance. 3D fully convolutional networks (FCNs) are a mainstream approach for cardiac segmentation due to their ability to integrate both

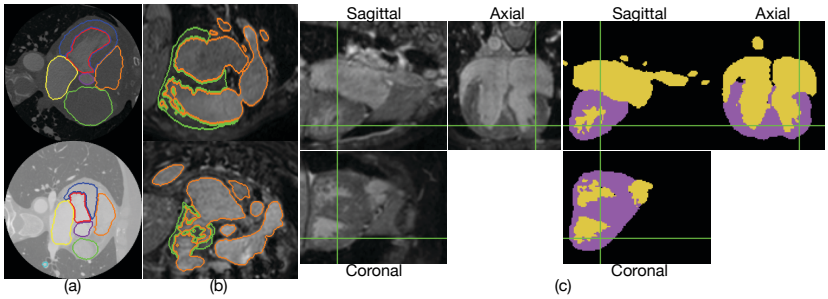

Fig. 1. Examples of cardiovascular images from (a) the MM-WHS CT dataset [14] in the axial plane and (b) the HVSMR dataset [6] in the sagittal plane. (c) Myocardium boundaries in the axial plane are easier to recognize.

inter- and intra-slice information in 3D images. However, two key factors have not been well explored: (1) the imaging qualities in different anatomical planes are not the same, and thus the degrees of segmentation difficulty from different views are unequal; (2) subtle structures (e.g., myocardium, pulmonary artery) have different orientations in different anatomical planes. Symmetrical convolutional and pooling operations may cause quick diminishment of subtle structures or boundaries, incurring segmentation errors. As shown in Fig. 1(c), myocardium boundaries in the axial plane are easier to recognize; with asymmetrical pooling along the longitudinal axis, more complementary inter-slice information can be kept which in return benefits segmentation in the axial plane.

Many recent studies tried to tackle the anisotropic issue of 3D biomedical images. But still, they could not segment myocardium or pulmonary artery well. Known methods that explored anisotropic 3D kernels in FCNs can be categorized into two types. (1) The methods in [8, 2] focused on designing repeatable cell structures and replaced all 3D convolutions systematically, called *short-range asymmetrical cell*. However, symmetrical pooling was used and deep features were fused periodically (with distinctive features vanishing quickly). (2) The methods in [3, 5] dealt with the anisotropic problem in 3D images using 2D FCNs to extract intra-slice features and 3D FCNs to aggregate inter-slice features. But, they did not exploit the fact that complementary information in the other planes (xz- and yz-planes) can also benefit the xy-plane, especially in less anisotropic 3D data (e.g., when the spacing resolution in the z-axis is only $3 \sim 5\times$ larger than that of the x- and y-axes).

To address the above two key factors, we propose a new *heterogeneous feature aggregation network* (HFA-Net), which is able to fully exploit complementary information in multiple views of 3D cardiac images and aggregate heterogeneous features to boost segmentation performance. To handle the issue in [8, 2], we utilize long-range asymmetrical branches to maintain distinguishable features associated with a specific view. Besides asymmetrical convolutional operations, we also apply asymmetrical pooling operations to maintain spatial resolution in the other planes. To address the issue in [3, 5], we utilize parallel encoding paths

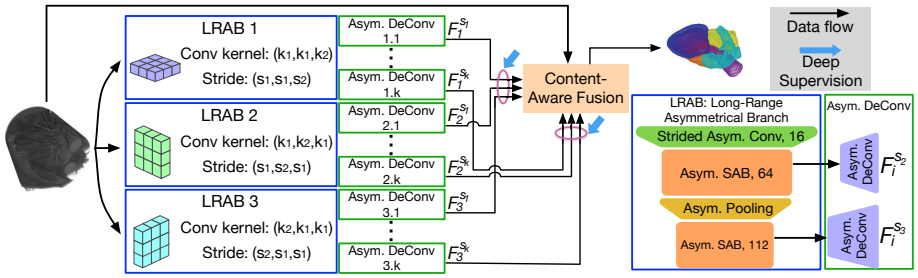

Fig. 2. An overview of our new HFA-Net framework.

to extract heterogeneous features from multiple geometric views of the 3D data (i.e., the axial, coronal, and sagittal planes). There is a good chance that an object can be distinguished from at least one of the geometric views. Thus, we encourage richer contextual and semantic features to be extracted. Further, to improve the parameter-performance efficiency and reduce GPU memory usage, we devise a sparsified densely-connected convolutional block for our model, and our HFA-Net thus designed can be trained end-to-end.

Experiments on three public challenge datasets [6, 14, 10] show that our new method achieves competitive segmentation results over state-of-the-art methods.

2 Method

Our HFA-Net has three main components (see Fig. 2): (1) Long-range asymmetrical branches (LRABs) that preserve subtle structures via asymmetrical convolutions and poolings; (2) a content-aware fusion module (CAFM) that combines multiple asymmetrical branches together, utilizing both raw images and feature maps from LRABs; (3) a new 3D sparse aggregation block (SAB) to reduce GPU memory usage and enable end-to-end training of the entire network.

2.1 Long-Range Asymmetric Branch (LRAB)

A straightforward way to exploit multiple geometric views of 3D images is to replace conventional 3D convolutional (Conv) layers by *short-range asymmetrical cell* (SRAC) [8, 2]. As shown in Fig. 3(a), a 3D Conv kernel is decomposed into m parallel streams, each having n pseudo 2D kernels and a corresponding orthogonal pseudo 1D kernel. But, the typical decompositions they exploited are $\{m = 1, 2; n = 1, 2\}$, which may not make the best out of all geometric properties of 3D data. Further, such SRAC only governs the specific layer-wise computation but neglects the outer branch/network level which controls spatial resolution changes. Most importantly, feature maps are added together periodically after each SRAC, which causes homogeneous feature maps in deeper layers and that parallel streams do not benefit richer feature extraction anymore. To address these issues, our method aims to fully exploit all the three orthogonal views and

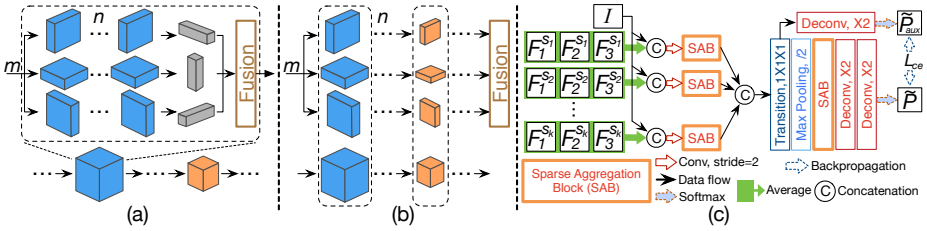

Fig. 3. (a) Short-Range Asymmetric Cell; (b) Long-Range Asymmetric Branch; (c) Content-Aware Fusion Module. I: raw image; $F_i^{s_j}$: feature maps (see Sect. 2.2).

encourage extracting heterogeneous features from different scales. For this goal, we need to carefully design both the layer-level and branch-level operations.

Notation. We denote a 3D Conv layer as $\text{Conv}(\mathcal{K}_{k_1,k_2,k_3}/\mathcal{S}_{s_1,s_2,s_3})$, where k_i and s_i are the kernel size and stride step size in each direction. Conventionally, $k_1 = k_2 = k_3$ and $s_1 = s_2 = s_3$. A 3D kernel $\mathcal{K}_{3,3,3}$ can be decomposed into an SRAC (with $m = 1$ and $n = 1$) by $\mathcal{K}_{3,3,1} \otimes \mathcal{K}_{1,1,3}$, $\mathcal{K}_{3,1,3} \otimes \mathcal{K}_{1,3,1}$, or $\mathcal{K}_{1,3,3} \otimes \mathcal{K}_{3,1,1}$, where $\otimes$ is convolution. Similarly, we denote a 3D deconvolutional (DeConv) layer as $\text{DeConv}(\mathcal{K}_{k_1,k_2,k_3} \times \mathcal{S}_{s_1,s_2,s_3})$. A pooling layer is denoted as $\mathcal{P}_{s_1,s_2,s_3}$.

Fig. 3(b) shows the concept of our *long-range asymmetrical branch* (LRAB). We utilize three LRABs ($m = 3$) to operate on three orthogonal geometric views separately, thus increasing the independency among m parallel encoding paths. The original symmetrical $\text{Conv}(\mathcal{K}_{k_a,k_a,k_a}/\mathcal{S}_{s_a,s_a,s_a})$ is replaced by an asymmetrical counterpart in each branch (i.e., $(\mathcal{K}_{k_a,k_a,1}/\mathcal{S}_{s_a,s_a,1})$, $(\mathcal{K}_{k_a,1,k_a}/\mathcal{S}_{s_a,1,s_a})$, or $(\mathcal{K}_{1,k_a,k_a}/\mathcal{S}_{1,s_a,s_a})$). Also, the consecutive 3D Conv kernel $(\mathcal{K}_{k_b,k_b,k_b}/\mathcal{S}_{s_b,s_b,s_b})$ is decomposed in the same orientation in each branch. Besides, since in each LRAB, Conv kernels are along the same orientation, conventional symmetrical pooling is no longer suitable (otherwise, inter-slice features may vanish quickly before being extracted). In our problem, cardiovascular segmentation is highly challenging especially due to the missing/ambiguous boundaries between the regions of interest and background or among various sub-structures. Thus, asymmetrical pooling (i.e., $\mathcal{P}_{s,s,1}$, $\mathcal{P}_{s,1,s}$, or $\mathcal{P}_{1,s,s}$) is utilized to maintain spatial resolution in the orthogonal direction so that there is a bigger chance that subtle distinguishable features can be kept in at least one of the geometric views.

For example, a $T \times T \times T$ tensor after three $\mathcal{P}_{2,2,2}$ becomes a $\frac{T}{8} \times \frac{T}{8} \times \frac{T}{8}$ tensor but becomes $\frac{T}{8} \times \frac{T}{8} \times T$ after three $\mathcal{P}_{2,2,1}$. Hence, additional information of subtle structures along the z-axis is kept and will be utilized by subsequent processing. Observe that the designs in [3, 5] can be viewed as special cases of our LRAB since these methods only used (pre-trained) 2D FCN to extract deep feature maps from 3D data slice by slice independently with $m = 1$. Thus, our method is more cautious in heterogeneous feature aggregation. Specifically, as shown in Fig. 2, our first LRAB is composed of stacking layers of $\text{Conv}(\mathcal{K}_{3,3,1}/\mathcal{S}_{2,2,1})$, $\text{SAB}(\mathcal{K}_{3,3,1}/\mathcal{S}_{1,1,1})$, $\mathcal{P}_{2,2,1}$, and $\text{SAB}(\mathcal{K}_{3,3,1}/\mathcal{S}_{1,1,1})$, where $\text{SAB}(\mathcal{K}_{3,3,1}/\mathcal{S}_{1,1,1})$ refers to sparse aggregation block (SAB) composed of stacked

Conv($\mathcal{K}_{3,3,1}/\mathcal{S}_{1,1,1}$). We will present SAB in Sect. 2.3. In the i^{th} LRAB, feature maps from different scales ($s_j, j = 1, 2, \ldots, k$) are recovered by asymmetrical DeConv layers accordingly, denoted by $F_i^{s_j}$. We will discuss how to aggregate useful information from these heterogeneous feature maps in Sect. 2.2.

2.2 Content-Aware Fusion Module (CAFM)

To maximally exploit the extracted heterogeneous features maps $F_i^{s_j}$ from parallel LRABs, we need to selectively leverage the correct information and suppress the incorrect one. It is quite possible that each voxel is correctly classified in at least one geometric view; thus, a key challenge is how to deal with agreement and disagreement in different views. For this, we present a content-aware fusion module (CAFM, see Fig. 3(c)) to generate aggregated deep features.

The input of CAFM includes two parts: a raw image I and heterogeneous feature maps $F_i^{S_j}$ of the same shape, where i is for the i^{th} LRAB and S_j is for the selected scales in LRABs. HFA-Net has $m = 3$ LRABs; thus $i \in \{1, 2, 3\}$. There are three scales in each LRAB and we choose the last two scales; thus $j \in \{2, 3\}$. To recover the asymmetrical feature maps to the original resolution of the input image I, we use asymmetrical DeConv accordingly (e.g., we use stacked $\{\text{DeConv}(\mathcal{K}_{4,4,1} \times \mathcal{S}_{2,2,1}), \text{DeConv}(\mathcal{K}_{4,4,1} \times \mathcal{S}_{2,2,1})\}$ to obtain $F_1^{S_3}$ for the 1^{st} LRAB). Then we average the feature maps from the same scale but different branches together to obtain hierarchical features $F^{S_j} = \frac{1}{m}\sum_{i=1}^{m} F_i^{S_j}$. This averaging provides a compact representation of all $F_i^{S_j}$'s while still showing the image areas where the heterogeneous features have agreement or disagreement. Next, each F^{S_j} is concatenated with the raw image I and fed to an encoder SAB, and all the intermediate feature maps are integrated in the middle of CAFM for extracting better representations. The raw image I provides a reference for helping further find detailed features and guide the feature aggregation process.

The loss function is computed as $\ell(X, Y; \theta) = \ell_{mse}(\widetilde{P}, Y) + \lambda_1 \ell_{mse}(\widetilde{P}_{aux}, Y) + \sum_i \sum_j \lambda_{ij} \ell_{mse}(S(F_i^{S_j}), Y)$, where Y is the corresponding ground truth of each training sample X, ℓ_{mse} is the multi-class cross-entropy loss and $S(\cdot)$ is the softmax function. See supplementary material for more details on HFA-Net.

2.3 Sparse Aggregation Block (SAB)

DenseVoxNet [11] is a state-of-the-art model for cardiovascular image segmentation, built on DenseBlock with dense residual connections. It aggregates all the previously computed features to each subsequent layer, computed as $x_\ell = H_\ell([x_0, x_1, \ldots, x_{\ell-1}])$, where x_0 is the input, x_ℓ is the output of layer ℓ, $[\cdot]$ is the concatenation operation, and $H_\ell(\cdot)$ is a composite of operations such as Conv, Pooling, BN, and ReLU. The dense connections help transfer useful features from shallower to deeper layers, and in turn, allow each shallow layer to receive direct supervision signal, thus alleviating the gradient vanishment issue in training deep ConvNets and achieving better parameter-performance efficiency.

Table 1. Datasets and training details. "GT = ✗": the ground truth of the data is kept by the organizers for fair comparison. The initial learning rate $L_r = 5 \times 10^{-4}$.

Dataset	Train		Test		# Class	Optimizer	# Iter.	Learning rate policy
	# stack	GT	# stack	GT				
2016 HVSMR [6]	10	✓	10	✗	2	Adam: $\beta_1 = 0.9$, $\beta_2 = 0.999$, $\epsilon =$ 1e-10	45,000	$L_r \times (1 - \frac{iter}{\#iter})^{0.9}$
2017 MM-WHS CT [14]	16	✓	4	✓	7		60,000	
2017 AAPM CT [10]	36	✓	12	✗	5		60,000	

However, for a DenseBlock of depth N, the number of skip connections and parameters grows quadratically asymptotically (i.e., $O(N^2)$). This means that each layer generates only a few new outputs to an ever-widening concatenation of previously seen feature representations. Thus, it is hard for the model to make full use of all the parameters and dense skip connections [13].

To further ease the training of our HFA-Net, we devise a new sparsified densely-connected convolutional block, called sparse aggregation block (SAB), to improve parameter-performance efficiency. The output x_ℓ of layer ℓ is computed as $x_\ell = H_\ell([x_{\ell-c^0}, x_{\ell-c^1}, x_{\ell-c^2}, x_{\ell-c^3}, \ldots, x_{\ell-c^k}])$, where $c > 1$ is an integer and $k \geq 0$ is the largest integer such that $c^k \leq \ell$. For an SAB of total depth N, this sparse aggregation introduces no more than $\log_c(N)$ incoming links per layer, for a total of $O(N \log(N))$ connections and parameters. We use $c = 2$ and $N = 12$ in all experiments. See supplementary material for more details.

3 Experiments and Results

Three 3D Datasets. (1) The **2016 HVSMR dataset** [6] aims to segment myocardium and great vessels (blood pool) in cardiovascular MRIs. The results are evaluated using three criteria: Dice coefficient, average surface distance (ADB), and symmetric Hausdorff distance. A score $S = \sum_{class}(\frac{1}{2}Dice - \frac{1}{4}ADB - \frac{1}{30}Hausdorff)$ is used to measure the overall accuracy of the results and for ranking. (2) The **2017 MM-WHS CT dataset** [14] aims to segment seven cardiac structures (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)). Following the setting in [1], we randomly split the dataset into the training (16 subjects) and testing (4 subjects) sets, which are fixed throughout all experiments. (3) The **2017 AAPM CT dataset** [10] aims to segment five thoracic structures (esophagus, spinal cord, left/right lung, and heart); esophagus and spinal cord are highly difficult cases.

Implementation Details. Our proposed method is implemented with Python using the TensorFlow framework and trained on an NVIDIA Tesla V100 graphics card with 32GB GPU memory. All the models are initialized using a Gaussian distribution and trained with the "poly" learning rate policy. We perform data augmentation to reduce overfitting. More details can be found in Table 1.

Quantitative Results. Table 2 (top) shows quantitative comparison of HFA-Net against other methods from the 2016 HVSMR Challenge Leaderboard, including a conventional atlas-based method [9] and 3D FCN based methods [4, 11]. First, our re-implementation of DVN achieves the state-of-the-art perfor-

Table 2. Segmentation results on the 2016 HVSMR dataset (top), 2017 MM-WHS CT dataset (middle), and 2017 CT AAPM dataset (bottom).

Method	Myocardium			Blood pool			Overall score
	Dice	ADB $[mm]$	Hausdorff $[mm]$	Dice	ADB $[mm]$	Hausdorff $[mm]$	
Shahzad et al. [9]	0.747	1.099	5.091	0.885	1.553	9.408	-0.330
3D Unet [4]	0.762	0.943	5.618	0.932	0.826	7.015	-0.016
DVN [11]	0.821	0.964	7.294	0.931	0.938	9.533	-0.161
DVN (ours)	0.829	0.701	3.431	0.933	0.921	8.489	0.078
S-DVN	0.822	0.689	3.729	0.936	0.900	8.770	0.065
Gonda et al. [2]	0.793	0.783	4.002	0.934	0.853	7.043	0.087
Li et al. [3]	0.802	0.876	4.243	0.930	0.978	7.481	0.012
HFA-Net	**0.837**	**0.627**	**3.301**	**0.942**	**0.751**	**5.875**	**0.239**

Model	Metrics	Structures							mean
		LV	RV	LA	RA	LV-myo	AO	PA	
Payer et al. [7]	Dice	0.918	0.909	0.929	0.888	0.881	0.933	**0.840**	0.900
Dou et al. [1]	Dice	0.888	-	0.891	-	0.733	0.813	-	-
DVN	Dice	0.942	0.891	**0.933**	0.879	0.908	0.959	0.824	0.905
	Jacard	0.891	0.806	**0.874**	0.786	0.832	0.922	0.713	0.832
	ADB[voxel]	0.084	**0.448**	**0.199**	0.459	**0.180**	0.132	1.710	0.459
	Hausdorff[voxel]	**6.752**	39.156	71.189	101.570	**35.422**	27.810	59.982	48.840
S-DVN	Dice	0.929	0.890	0.914	**0.899**	0.895	0.956	0.828	0.902
	Jaccard	0.870	0.805	0.843	**0.817**	0.811	0.916	0.718	0.826
	ADB[voxel]	0.610	0.666	1.384	**0.307**	0.362	0.210	**1.594**	0.733
	Hausdorff[voxel]	21.214	55.473	85.726	73.757	62.053	80.511	77.181	65.131
HFA-Net	Dice	**0.946**	**0.893**	0.925	0.897	**0.910**	**0.964**	0.830	**0.909**
	Jaccard	**0.898**	**0.810**	0.861	0.816	**0.836**	**0.930**	**0.722**	**0.839**
	ADB[voxel]	**0.076**	0.562	0.210	0.334	0.225	**0.103**	1.685	**0.456**
	Hausdorff[voxel]	7.148	**33.128**	**42.173**	**22.903**	36.954	**12.075**	**37.845**	**27.461**

Model	Metrics	Structures					mean
		Esophagus	Spinal Cord	Lung_R	Lung_L	Heart	
DVN [4]	Dice	0.676	0.851	0.960	0.960	0.917	0.873
	ADB[mm]	2.227	0.867	**1.212**	1.295	2.418	1.604
	Hausdorff[mm]	7.748	2.298	**3.938**	4.100	6.781	4.973
HFA-Net	Dice	**0.697**	**0.874**	**0.962**	**0.964**	**0.920**	**0.883**
	ADB[mm]	**1.974**	**0.766**	1.266	**0.967**	**2.336**	**1.462**
	Hausdorff[mm]	**5.883**	**2.190**	4.149	**3.370**	**6.557**	**4.430**

mance and our S-DVN with SAB achieves competitive results while reducing the number of parameters by $\sim 60\%$ (4.3M *vs.* 1.6M). Second, recall the two types of the known anisotropic 3D methods (see Sect. 1). We choose at least one typical method from each type for comparison. The method [2] is based on the short-range asymmetrical cell design, which utilizes 3D kernel decomposition on the orthogonal planes to predict a class label for each voxel. The method [3] extracts features from the xy-plane by a 2D FCN and applies a 3D FCN to fuse inter-slice information. Our HFA-Net outperforms these methods across nearly all the metrics with a very high overall score of 0.239. The results for the 2017 MM-WHS CT dataset are given in Table 2 (middle). First, our baselines (DVN and S-DVN) already achieve better results than the known state-of-the-art methods [7, 1]. Second, our HFA-Net further improves the accuracy on most the categories across nearly all the metrics, especially for subtle structures such as LV-myo and AO. To further show that our method is robust and effective in delineating subtle structures, we experiment with HFA-Net on the 2017 AAPM CT dataset. Quantitative results in Table 2 (bottom) show promising performance gain, especially for esophagus and spinal cord (2% gain in Dice coefficient).

Qualitative Results. As shown in Fig. 4, our HFA-Net attains better results and shows a strong capability of delineating missing/ambiguous boundaries. More qualitative results can be found in supplementary material.

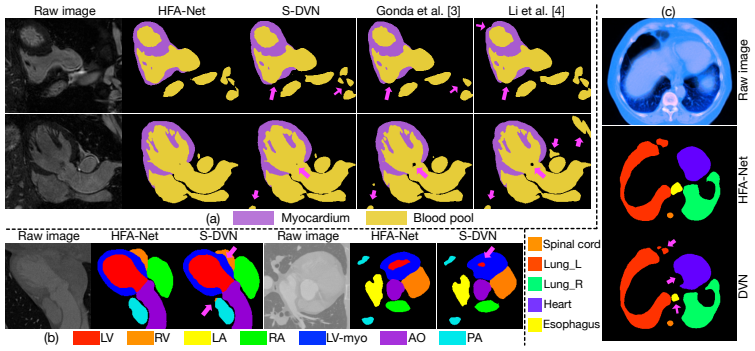

Fig. 4. Visual qualitative results: the 2016 HVSMR dataset (a), 2017 MM-WHS CT dataset (b), and 2017 CT AAPM dataset (c) (some errors marked by magenta arrows).

4 Conclusions

In this paper, we presented a new deep *heterogeneous feature aggregation network* (HFA-Net) for cardiovascular segmentation in 3D CT/MR images. Our proposed HFA-Net extracts rich heterogeneous features using long-range asymmetrical branches and aggregates diverse contextual and semantic deep features using a content-aware fusion module. Sparse aggregation block is utilized to give HFA-Net a better parameter-performance efficiency. Comprehensive experiments on three open challenge datasets demonstrated the efficacy of our new method.

Acknowledgement. This research was supported in part by the U.S. National Science Foundation through grants IIS-1455886, CCF-1617735, CNS-1629914, DUE-1833129 and NIH grant R01 DE027677-01.

References

1. Dou, Q., Ouyang, C., Chen, C., Chen, H., Heng, P.A.: Unsupervised cross-modality domain adaptation of convnets for biomedical image segmentations with adversarial loss. In: Twenty-Seventh International Joint Conference on Artificial Intelligence. pp. 691–697 (2018)
2. Gonda, F., Wei, D., Parag, T., Pfister, H.: Parallel separable 3D convolution for video and volumetric data understanding. arXiv preprint arXiv:1809.04096 (2018)
3. Li, X., Chen, H., Qi, X., Dou, Q., Fu, C.W., Heng, P.A.: H-DenseUNet: Hybrid densely connected UNet for liver and tumor segmentation from CT volumes. IEEE Transactions on Medical Imaging **37**(12), 2663–2674 (2018)
4. Liang, P., Chen, J., Zheng, H., Yang, L., Zhang, Y., Chen, D.Z.: Cascade decoder: A universal decoding method for biomedical image segmentation. In: IEEE, ISBI 2019. pp. 339–342 (2019)
5. Liu, S., Xu, D., Zhou, S.K., Pauly, O., Grbic, S., Mertelmeier, T., Wicklein, J., Jerebko, A., Cai, W., Comaniciu, D.: 3D anisotropic hybrid network: Transferring

convolutional features from 2D images to 3D anisotropic volumes. In: MICCAI. pp. 851–858 (2018)
6. Pace, D.F., Dalca, A.V., Geva, T., Powell, A.J., Moghari, M.H., Golland, P.: Interactive whole-heart segmentation in congenital heart disease. In: MICCAI. pp. 80–88. Springer (2015)
7. Payer, C., Štern, D., Bischof, H., Urschler, M.: Multi-label whole heart segmentation using CNNs and anatomical label configurations. In: International Workshop on Statistical Atlases and Computational Models of the Heart. pp. 190–198 (2017)
8. Qiu, Z., Yao, T., Mei, T.: Learning spatio-temporal representation with pseudo-3D residual networks. In: ICCV. pp. 5533–5541 (2017)
9. Shahzad, R., Gao, S., Tao, Q., Dzyubachyk, O., van der Geest, R.: 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, pp. 147–155. Springer (2016)
10. Yang, J., Sharp, G., Veeraraghavan, H., van Elmpt, W., Dekker, A., Lustberg, T., Gooding, M.: Lung CT segmentation challenge 2017 — the cancer imaging archive. http://doi.org/10.7937/k9/tcia.2017.3r3fvz08 (2017)
11. Yu, L., Cheng, J.Z., Dou, Q., Yang, X., Chen, H., Qin, J., Heng, P.A.: Automatic 3D cardiovascular MR segmentation with densely-connected volumetric ConvNets. In: MICCAI. pp. 287–295 (2017)
12. Zheng, H., Zhang, Y., Yang, L., Liang, P., Zhao, Z., Wang, C., Chen, D.Z.: A new ensemble learning framework for 3D biomedical image segmentation. In: Thirty-Third AAAI Conference on Artificial Intelligence (2019)
13. Zhu, L., Deng, R., Maire, M., Deng, Z., Mori, G., Tan, P.: Sparsely aggregated convolutional networks. In: ECCV. pp. 186–201 (2018)
14. Zhuang, X., Shen, J.: Multi-scale patch and multi-modality atlases for whole heart segmentation of MRI. Medical Image Analysis **31**, 77–87 (2016)