Automatic 3D Cardiovascular MR Segmentation with Densely-Connected Volumetric ConvNets

Lequan Yu[1(✉)], Jie-Zhi Cheng[2], Qi Dou[1], Xin Yang[1], Hao Chen[1], Jing Qin[3], and Pheng-Ann Heng[1,4]

[1] Department of Computer Science and Engineering, The Chinese University of Hong Kong, Shatin, Hong Kong
lqyu@cse.cuhk.edu.hk
[2] Department of Electrical Engineering, Chang Gung University, Taoyuan, Taiwan
[3] Centre for Smart Health, School of Nursing, The Hong Kong Polytechnic University, Kowloon, Hong Kong
[4] Guangdong Provincial Key Laboratory of Computer Vision and Virtual Reality Technology, Shenzhen Institutes of Advanced Technology, Chinese Academy of Sciences, Shenzhen, China

Abstract. Automatic and accurate whole-heart and great vessel segmentation from 3D cardiac magnetic resonance (MR) images plays an important role in the computer-assisted diagnosis and treatment of cardiovascular disease. However, this task is very challenging due to ambiguous cardiac borders and large anatomical variations among different subjects. In this paper, we propose a novel densely-connected volumetric convolutional neural network, referred as *DenseVoxNet*, to automatically segment the cardiac and vascular structures from 3D cardiac MR images. The DenseVoxNet adopts the 3D fully convolutional architecture for effective volume-to-volume prediction. From the learning perspective, our DenseVoxNet has three compelling advantages. First, it preserves the maximum information flow between layers by a densely-connected mechanism and hence eases the network training. Second, it avoids learning redundant feature maps by encouraging feature reuse and hence requires fewer parameters to achieve high performance, which is essential for medical applications with limited training data. Third, we add auxiliary side paths to strengthen the gradient propagation and stabilize the learning process. We demonstrate the effectiveness of DenseVoxNet by comparing it with the state-of-the-art approaches from HVSMR 2016 challenge in conjunction with MICCAI, and our network achieves the best dice coefficient. We also show that our network can achieve better performance than other 3D ConvNets but with fewer parameters.

1 Introduction

Accurate segmentation of cardiac structures in 3D cardiac MR images is crucial for the diagnosis and treatment planning of cardiovascular disease. For example, the segmentation results can support the building of patient-specific 3D heart model for the surgical planning of the severe congenital heart disease [9]. The

M. Descoteaux et al. (Eds.): MICCAI 2017, Part II, LNCS 10434, pp. 287–295, 2017.
DOI: 10.1007/978-3-319-66185-8_33

manual segmentation on every MR slice can be very tedious and time-consuming, and subjects to inter- and intra-observer variability. Accordingly, an automatic segmentation scheme is highly demanded in clinical practice.

However, the automatic segmentation is by no means a trivial task, as some parts of cardiac borders are not very well defined due to the low contrast to the surrounding tissues. Meanwhile, the inter-subject variation of cardiac structures may impose more difficulty for the segmentation task. One prominent family of approaches are based on multiple atlases and deformable models [15]. These approaches needs to well consider the high anatomical variations in different subjects and useful atlases need to be built from a relatively large dataset. Pace et al. [9] developed an interactive method for the accurate segmentation of cardiac chambers and vessels, but this method is very slow. Recently, convolutional neural networks (ConvNets) significantly improve the segmentation performance for medical images [2,3,10]. As for this task, Wolterink et al. [13] employed a dilated ConvNet to demarcate the myocardium and blood pool, but the 3D volumetric information was not fully used in the study. Yu et al. [14] proposed the 3D FractalNet to consider the 3D image information. However, this network and other 3D ConvNets (e.g., 3D U-Net [2], VoxResNet [1]) usually generate a large number of feature channels in each layer and they have plenty of parameters to be tuned during training. Although these networks introduce different skip connections to ease the training, the training of an effective model with the limited MR images for heart segmentation is still very challenging.

In order to ease the training of 3D ConvNets with limited data, we propose a novel densely-connected volumetric ConvNet, namely *DenseVoxNet*, to segment the cardiac and vascular structures in cardiac MR images. The DenseVoxNet adopts 3D fully convolutional architecture, and thus can fully incorporate the 3D image and geometric cues for effective volume-to-volume prediction. More importantly, the DenseVoxNet incorporates the concept of dense connectivity [5] and enjoys three advantages from the learning perspective. First, it implements direct connections from a layer to all its subsequent layers. Each layer can receive additional supervision from the loss function through the shorter connections, and thus make the network much easier to train. Second, the DenseVoxNet has fewer parameters than the other 3D ConvNets. Since layers can access to feature maps from all of its preceding layers, the learning of redundant feature maps can be possibly avoided. Therefore, the DenseVoxNet has fewer feature maps in each layer, which is essential for training ConvNets with limited images as it has less chance to encounter the overfitting problem. Third, we further improve the gradient flow within the network and stabilize the learning process via auxiliary side paths. We extensively evaluate the DenseVoxNet on the HVSMR 2016 challenge dataset. The results demonstrate that DenseVoxNet can outperform other state-of-the-art methods for the segmentation of myocardium and blood pool in 3D cardiac MR images, corroborating its advantages over existing methods.

2 Method

In this section, we first introduce the concept of dense connection. Then, we elaborate the architecture of our DenseVoxNet bearing the spirit of dense connection. The training procedure is detailed in the last subsection.

2.1 Dense Connection

In a ConvNet, we denote $\mathbf{x}_\ell$ as the output of the ℓ^{th} layer, and $\mathbf{x}_\ell$ can be computed by a transformation $H_l(\mathbf{x})$ from the output of the previous layer, $\mathbf{x}_{\ell-1}$ as:

$$\mathbf{x}_\ell = H_\ell(\mathbf{x}_{\ell-1}), \tag{1}$$

where $H_\ell(\mathbf{x})$ can be a composite of operations such as Convolution (Conv), Pooling, Batch Normalization (BN) or rectified linear unit (ReLU), etc. To boost the training against the vanishing gradients, ResNet [4] introduces a kind of skip connection which integrates the response of $H_\ell(\mathbf{x})$ with the identity mapping of the features from the previous layer to augment the information propagation as:

$$\mathbf{x}_\ell = H_\ell(\mathbf{x}_{\ell-1}) + \mathbf{x}_{\ell-1}. \tag{2}$$

However, the identity function and the output of H_ℓ are combined by summation, which may impede the information flow in the network.

To further improve the information flow within the network, the dense connectivity [5] exercises the idea of skip connections to the extreme by implementing the connections from a layer to all its subsequent layers. Specifically, the $\mathbf{x}_\ell$ is defined as:

$$\mathbf{x}_\ell = H_\ell([\mathbf{x}_0, \mathbf{x}_1, ..., \mathbf{x}_{\ell-1}]), \tag{3}$$

where [...] refers to the concatenation operation. The dense connectivity, as illustrated at the left bottom of Fig. 1, makes all layers receive direct supervision signal. More importantly, such a mechanism can encourage the reuse of features among all these connected layers. Suppose that if the output of each layer has k feature maps, then the k, referred as *growth rate*, can be set to a small number to reduce the number of parameters since there is no need to re-learn redundant feature maps. This characteristic is quite compelling to medical image analysis tasks, where it is usually difficult to train an effective network with a lot of parameters with limited training data.

2.2 The Architecture of DenseVoxNet

Figure 1 illustrates the architecture of our proposed DenseVoxNet. It adopts the 3D fully convolutional network architecture [1–3] and has the down- and up-sampling components to achieve end-to-end training. Note that the Eq. 3 is not applicable when the feature maps have different sizes; on the another hand, we need to reduce the feature map size for better efficiency of memory space and increase the receptive field to enclose more information when prediction.

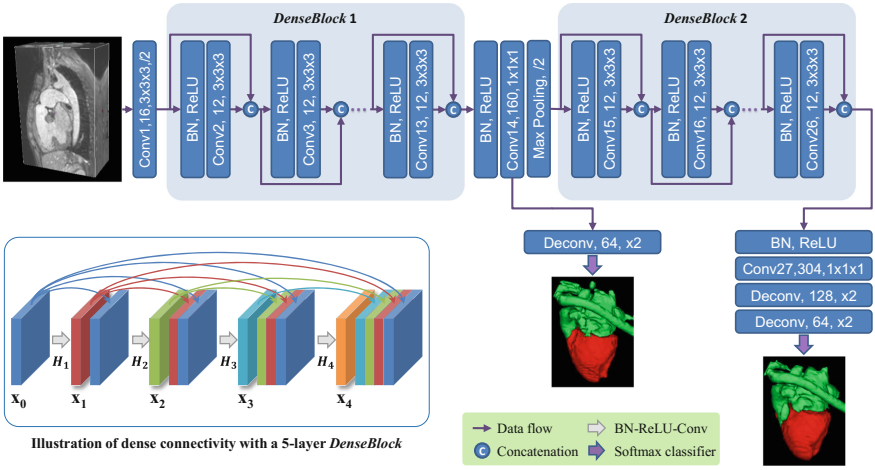

Fig. 1. The architecture of our DenseVoxNet. It consists of two *DenseBlocks* and all operations are implemented in a 3D manner. The green and red color denotes the output of blood pool and myocardium. The graph in left bottom illustrates the dense connectivity scheme taking a 5-layer *DenseBlock* as an example.

We, therefore, divide the down-sampling components into two densely-connected blocks, referred as *DenseBlock*, and each *DenseBlock* is comprised of 12 transformation layers with dense connections (Only draw 3 layers in the figure for simplicity). Each transformation layer is sequentially composed of a BN, a ReLU, and a $3 \times 3 \times 3$ Conv and the growth rate, k, of our DenseVoxNet is 12. The first *DenseBlock* is prefixed with a Conv with 16 output channels and stride of 2 to learn primitive features. In-between the two *DenseBlocks* is the transition block which consists of a BN, a ReLU, a $1 \times 1 \times 1$ Conv and a $2 \times 2 \times 2$ max pooling layers.

The up-sampling component is composed of a BN, a ReLU, a $1 \times 1 \times 1$ Conv and two $2 \times 2 \times 2$ deconvolutional (Deconv) layers to ensure the sizes of segmentation prediction map consistent with the size of input images. The up-sampling component is then followed with a $1 \times 1 \times 1$ Conv layer and soft-max layer to generate the final label map of the segmentation. To equip the DenseVoxNet with the robustness against the overfitting problem, the dropout layer is implemented following each Conv layer with the dropout rate of 0.2.

To further boost the information flow within the network, we implement a kind of long skip connection to connect the transition layer to the output layer with a $2 \times 2 \times 2$ Deconv layer. This skip connection shares the similar idea of deep supervision [3] to strengthen the gradient propagation and stabilize the learning process. In addition, this long skip connection may further tap the potential of the limited training data to learn more discriminative features. Our DenseVoxNet has about 1.8M parameters in total, which is much fewer than 3D U-Net [2] with 19.0M parameters and VoxResNet [1] with 4.0M parameters.

2.3 Training Procedure

The DenseVoxNet is implemented with Caffe [6] library[1]. The weights were randomly initialized with a Gaussian distribution ($\mu = 0$, $\sigma = 0.01$). The optimization is realized with the stochastic gradient descend algorithm (batch size = 3, weight decay = 0.0005, momentum = 0.9). The initial learning rate was set to 0.05. We use the "poly" learning rate policy (i.e., the learning rate is multiplied by $(1 - \frac{iter}{max\_iter})^{power}$) for the decay of learning rate along the training iteration. The power variable was set to 0.9 and maximum iteration number (max_iter) was set as 15000. To fit the limited 12 GB GPU memory, the input of the DenseVoxNet is sub-volumes with size of $64 \times 64 \times 64$, which were randomly cropped from the training images. The final segmentation results were obtained with the major voting strategy [7] from the predictions of the overlapped sub-volumes.

3 Experiments and Results

Dataset and Pre-processing. The DenseVoxNet is evaluated with the dataset of HVSMR 2016 Challenge. There are in total 10 3D cardiac MR scans for training and 10 scans for testing. The scans have low quality as they were acquired with a 1.5T scanner. All cardiac MR images were scanned from the patients with congenital heart diseases (CHD). The HVSMR 2016 dataset contains the annotations for the myocardium and great vessel, and the testing data annotations are held by organizers for fair comparison. Due to the large intensity variance among different images, all cardiac MR images were normalized to have zero mean and unit variance. We did not employ spatial resampling. To leverage the limited training data, simple data augmentation was employed to enlarge the training data. The augmentation operations include the rotation with 90, 180 and 270°, as well as image flipping along the axial plane.

Qualitative Results. In Fig. 2, we demonstrate 4 typical segmentation results on training images (the first two samples, via cross validation) and testing images (the last two samples). The four slices are from different subjects but with the same coronal plane view. The blue and purple color denotes our segmentation results for blood pool and myocardium, respectively, and segmentation ground truth is also presented in white and gray regions in the first two samples. As can be observed, there exists large variation of cardiac structures among different subjects in both training and testing images. Our method can still successfully demarcate myocardium and blood pool from the low-intensity contrast cardiac MR images, demonstrating the effectiveness of the proposed DenseVoxNet.

Comparison with Other Methods. The quantitative comparison between DenseVoxNet and other approaches from the participating teams in this challenge is shown in Table 1. According to the rules of the challenge, methods were

[1] https://github.com/yulequan/HeartSeg.

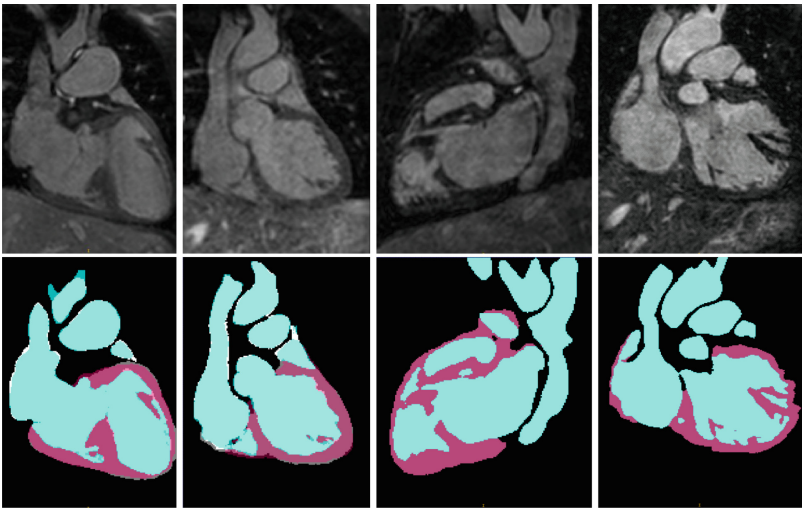

Fig. 2. Segmentation results on training images (the first two) and testing images (the last two). The blue and purple color denotes our segmentation results for blood pool and myocardium, respectively, and segmentation ground truth is also presented in white and gray regions in the first two samples.

ranked based on Dice coefficient (Dice). Meanwhile, other ancillary measures like average surface distance (ABD) and symmetric Hausdorff distance (Hausdorff) are also computed for reference. Higher Dice values suggest a higher agreement between segmentation results and ground truth, while lower ABD and Hausdorff values indicate higher boundary similarity. Three of the six approaches employed traditional methods based on hand-crafted features, including random forest [8], 3D Markov random field and substructure tracking [12] and level-set method driven by multiple atlases [11]. The other three methods, include ours, are based on ConvNet. Wolterink et al. [13] employed 2D dilated ConvNets to segment the myocardium and blood pool, while Yu et al. [14] utilized 3D ConvNets.

Table 1. Comparison with different approaches on HVSMR2016 dataset.

Method	Myocardium			Blood pool		
	Dice	ADB [mm]	Hausdorff [mm]	Dice	ADB [mm]	Hausdorff [mm]
Mukhopadhyay [8]	0.495 ± 0.126	2.596 ± 1.358	12.796 ± 4.435	0.794 ± 0.053	2.550 ± 0.996	14.634 ± 8.200
Tziritas [12]	0.612 ± 0.153	2.041 ± 1.022	13.199 ± 6.025	0.867 ± 0.047	2.157 ± 0.503	19.723 ± 4.078
Shahzad et al. [11]	0.747 ± 0.075	1.099 ± 0.204	$\mathbf{5.091 \pm 1.658}$	0.885 ± 0.028	1.553 ± 0.376	9.408 ± 3.059
Wolterink et al. [13]	0.802 ± 0.060	$\mathbf{0.957 \pm 0.302}$	6.126 ± 3.565	0.926 ± 0.018	0.885 ± 0.223	7.069 ± 2.857
Yu et al. [14]	0.786 ± 0.064	0.997 ± 0.353	6.419 ± 2.574	0.931 ± 0.016	$\mathbf{0.868 \pm 0.218}$	$\mathbf{7.013 \pm 3.269}$
DenseVoxNet (Ours)	$\mathbf{0.821 \pm 0.041}$	0.964 ± 0.292	7.294 ± 3.340	$\mathbf{0.931 \pm 0.011}$	0.938 ± 0.224	9.533 ± 4.194

Table 1 reports the results of different methods. It can be observed that the ConvNet-based methods (the last three rows) can generally achieve better

performance than the other methods do, suggesting that ConvNets can generate more discriminative features in a data-driven manner to better tackle the large anatomical variability of patients with CHD. Regarding the segmentation of myocardium, our method achieves the best performance with the Dice, i.e., the ranking metric in the Challenge, of 0.821 ± 0.041 and outperforms the second one by around 2%. For the segmentation of blood pool, our method also achieves the best Dice score of 0.931 ± 0.011 with a small deviation. The ADB and Hausdorff scores of our method are also competitive compared to the best performance. It is worth noting that the dice scores of myocardium in all methods are lower than the Dice scores of blood pool, suggesting that the segmentation of myocardium is relatively more challenging due to the ambiguous borders of the myocardium in the low-resolution MR images. While other two ConvNet-based approaches achieve quite close Dice scores to our DenseVoxNet in blood pool segmentation, our method is obviously better than these two methods in the dice scores of the myocardium, demonstrating our densely-connected network with auxiliary long side paths has the capability to tackle hard myocardium segmentation problem.

We further implement other two state-of-the-art 3D ConvNets, 3D U-Net [2] and VoxResNet [1], for comparison. We also compare the performance of the proposed DenseVoxNet with and without auxiliary side paths. The quantitative comparison can be found in Table 2, where "DenseVoxNet-A" denotes the DenseVoxNet without the auxiliary side paths. As can be observed, our DenseVoxNet achieves much better performance than the other two 3D ConvNets in both myocardium and blood pool segmentation. It suggests that our DenseVoxNet can benefit from the improved information flow throughout the network with the dense connections. In addition, our method achieves better performance with much fewer parameters than our competitors, corroborating the effectiveness of the feature map reusing mechanism encoded in the densely-connected architecture, which is quite important to enhance the capability of ConvNet models under limited training data. It is also observed that the auxiliary side path can further improve the segmentation performance, especially for the myocardium.

Table 2. Quantitative analysis of our network

Method	Parameters	Myocardium			Blood pool		
		Dice	ADB[mm]	Hausdorff[mm]	Dice	ADB[mm]	Hausdorff[mm]
3D U-Net [2]	19.0M	0.694 ± 0.076	1.461 ± 0.397	10.221 ± 4.339	0.926 ± 0.016	0.940 ± 0.192	$\mathbf{8.628 \pm 3.390}$
VoxResNet [1]	4.0M	0.774 ± 0.067	1.026 ± 0.400	$\mathbf{6.572 \pm 3.551}$	0.929 ± 0.013	0.981 ± 0.186	9.966 ± 3.021
DenseVoxNet-A	1.7M	0.787 ± 0.042	1.811 ± 0.752	17.534 ± 7.838	0.917 ± 0.018	1.451 ± 0.537	15.892 ± 6.772
DenseVoxNet	1.8M	$\mathbf{0.821 \pm 0.041}$	$\mathbf{0.964 \pm 0.292}$	7.294 ± 3.340	$\mathbf{0.931 \pm 0.011}$	$\mathbf{0.938 \pm 0.224}$	9.533 ± 4.194

4 Discussion and Conclusion

A DenseVoxNet is proposed to automatically segment the cardiac structures in the 3D cardiac MR images. The DenseVoxNet is equipped with dense connectivity and spares network architecture from a large number of redundant features.

It is because the learned features from previous layers can be reused. Therefore, the DenseVoxNet may enjoy better parameter efficiency and has less chance to encounter the overfitting problem when training with limited data. We use lots of Conv layers in downsampling path and hence equip the network with large receptive fields to learn sufficient higher level features. The denseVoxNet can attain best Dice scores for the segmentation of myocardium and blood pool on the challenge dataset. On the other hand, it is also interesting to observe that the 2D ConvNet method [13] can outperform some 3D ConvNet methods on some metrics. It may be because the dataset in the HVSMR 2016 challenge is quite limited and it is very difficult to train an effective 3D network with such limited data. On the other hand, the DenseVoxNet can achieve better segmentation performance than the three 3D ConvNets do. Therefore, the efficacy of the DenseVoxNet can then be well corroborated.

Acknowledgments. The work described in this paper was supported by the grants from the Research Grants Council of the Hong Kong Special Administrative Region (Project No. CUHK 412513 and CUHK 14203115) and the National Natural Science Foundation of China (Project No. 61233012).

References

1. Chen, H., Dou, Q., Yu, L., Qin, J., Heng, P.A.: Voxresnet: deep voxelwise residual networks for brain segmentation from 3D MR images. NeuroImage (2017). ISSN 1053-8119. http://dx.doi.org/10.1016/j.neuroimage.2017.04.041
2. Çiçek, Ö., Abdulkadir, A., Lienkamp, S.S., Brox, T., Ronneberger, O.: 3D U-Net: learning dense volumetric segmentation from sparse annotation. In: Ourselin, S., Joskowicz, L., Sabuncu, M.R., Unal, G., Wells, W. (eds.) MICCAI 2016. LNCS, vol. 9901, pp. 424–432. Springer, Cham (2016). doi:10.1007/978-3-319-46723-8_49
3. Dou, Q., Chen, H., Jin, Y., Yu, L., Qin, J., Heng, P.-A.: 3D deeply supervised network for automatic liver segmentation from CT volumes. In: Ourselin, S., Joskowicz, L., Sabuncu, M.R., Unal, G., Wells, W. (eds.) MICCAI 2016. LNCS, vol. 9901, pp. 149–157. Springer, Cham (2016). doi:10.1007/978-3-319-46723-8_18
4. He, K., Zhang, X., Ren, S., Sun, J.: Deep residual learning for image recognition. In: CVPR, pp. 770–778 (2016)
5. Huang, G., Liu, Z., Weinberger, K.Q., van der Maaten, L.: Densely connected convolutional networks. arXiv preprint arXiv:1608.06993 (2016)
6. Jia, Y., Shelhamer, E., Donahue, J., et al.: Caffe: convolutional architecture for fast feature embedding. arXiv preprint arXiv:1408.5093 (2014)
7. Kontschieder, P., Bulo, S.R., Bischof, H., Pelillo, M.: Structured class-labels in random forests for semantic image labelling. In: ICCV, pp. 2190–2197 (2011)
8. Mukhopadhyay, A.: Total variation random forest: fully automatic MRI segmentation in congenital heart diseases. In: Zuluaga, M.A., Bhatia, K., Kainz, B., Moghari, M.H., Pace, D.F. (eds.) RAMBO/HVSMR -2016. LNCS, vol. 10129, pp. 165–171. Springer, Cham (2017). doi:10.1007/978-3-319-52280-7_17
9. 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: Navab, N., Hornegger, J., Wells, W.M., Frangi, A.F. (eds.) MICCAI 2015. LNCS, vol. 9351, pp. 80–88. Springer, Cham (2015). doi:10.1007/978-3-319-24574-4_10

10. Ronneberger, O., Fischer, P., Brox, T.: U-Net: convolutional networks for biomedical image segmentation. In: Navab, N., Hornegger, J., Wells, W.M., Frangi, A.F. (eds.) MICCAI 2015. LNCS, vol. 9351, pp. 234–241. Springer, Cham (2015). doi:10.1007/978-3-319-24574-4_28
11. Shahzad, R., Gao, S., Tao, Q., Dzyubachyk, O., Geest, R.: Automated cardiovascular segmentation in patients with congenital heart disease from 3D CMR scans: combining multi-atlases and level-sets. In: Zuluaga, M.A., Bhatia, K., Kainz, B., Moghari, M.H., Pace, D.F. (eds.) RAMBO/HVSMR -2016. LNCS, vol. 10129, pp. 147–155. Springer, Cham (2017). doi:10.1007/978-3-319-52280-7_15
12. Tziritas, G.: Fully-automatic segmentation of cardiac images using 3-D MRF model optimization and substructures tracking. In: Zuluaga, M.A., Bhatia, K., Kainz, B., Moghari, M.H., Pace, D.F. (eds.) RAMBO/HVSMR -2016. LNCS, vol. 10129, pp. 129–136. Springer, Cham (2017). doi:10.1007/978-3-319-52280-7_13
13. Wolterink, J.M., Leiner, T., Viergever, M.A., Išgum, I.: Dilated convolutional neural networks for cardiovascular MR segmentation in congenital heart disease. In: Zuluaga, M.A., Bhatia, K., Kainz, B., Moghari, M.H., Pace, D.F. (eds.) RAMBO/HVSMR -2016. LNCS, vol. 10129, pp. 95–102. Springer, Cham (2017). doi:10.1007/978-3-319-52280-7_9
14. Yu, L., Yang, X., Qin, J., Heng, P.-A.: 3D FractalNet: dense volumetric segmentation for cardiovascular MRI volumes. In: Zuluaga, M.A., Bhatia, K., Kainz, B., Moghari, M.H., Pace, D.F. (eds.) RAMBO/HVSMR -2016. LNCS, vol. 10129, pp. 103–110. Springer, Cham (2017). doi:10.1007/978-3-319-52280-7_10
15. Zhuang, X.: Challenges and methodologies of fully automatic whole heart segmentation: a review. J. Healthcare Eng. **4**(3), 371–407 (2013)