



Automatic Segmentation of LV and RV in Cardiac MRI

Yeonggul Jang¹, Yoonmi Hong²(✉), Seongmin Ha², Sekeun Kim²,
and Hyuk-Jae Chang^{2,3}

¹ Brain Korea 21 PLUS Project for Medical Science,
Yonsei University, Seoul, South Korea
jyg1722@gmail.com

² Integrative Cardiovascular Imaging Research Center,
Yonsei University College of Medicine, Seoul, South Korea
yoonmhong@gmail.com

³ Division of Cardiology, Severance Cardiovascular Hospital,
Yonsei University College of Medicine, Seoul, South Korea

Abstract. Automatic and accurate segmentation of Left Ventricle (LV) and Right Ventricle (RV) in cine-MRI is required to analyze cardiac function and viability. We present a fully convolutional neural network to efficiently segment LV and RV as well as myocardium. The network is trained end-to-end from scratch. Average dice scores from five-fold cross-validation on the ACDC training dataset were 0.94, 0.89, and 0.88 for LV, RV, and myocardium. Experimental results show the robustness of the proposed architecture.

Keywords: Cardiac segmentation · Convolutional neural network
Cardiac MRI · Automated Cardiac Diagnosis Challenge

1 Introduction

Cardiac image segmentation plays an important role for the diagnosis of cardiac diseases, quantification of volume, and image-guided interventions [1]. Due to the advancement of the echocardiogram, Computed Tomography (CT), Magnetic Resonance Imaging (MRI), quantitative and qualitative measurements from the cardiac imaging can be proceeded easily. Accurate segmentation of Left Ventricle (LV) and Right Ventricle (RV) are particularly valuable for the extraction of ventricular function information such as stroke volume and ejection fraction. In clinic, MRI becomes a reference modality to evaluate the cardiac function.

Various methods are developed to automate the segmentation of the ventricles in cardiac MRI [2–4]. Recently, the advancement of the deep learning shows high performances in object detection, recognition, as well as segmentation. There are many attempts to use deep learning technique, especially Convolutional Neural Networks

Y. Jang and Y. Hong—both authors contributed equally.

(CNN), in cardiac image segmentation problems. 2D CNN was applied in cardiac images for LV segmentation with auto-encoder [5]. Also, 3D CNN is applied to detect coronary calcium in gated cardiac CT Angiography [6]. Recurrent Neural Network (RNN) is applied for LV segmentation in multi-slice MR images [7]. Multi-scale convolutional deep belief network is proposed to estimate bi-ventricular volume in cardiac MRI [8]. The largest challenge in the cardiac imaging was the 2015 Kaggle Data Science Bowl that aims to automatically measure end-systolic and end-diastolic volumes in cardiac MRI [9].

In this paper, we introduce a fully automated segmentation method for LV, LV myocardium, and RV in cine MRI. Our Architecture is based on the M-net where we only use 2D CNNs without 3D-to-2D converter [10]. The M-net architecture proposed in [10] has 3D filtering layer to utilize 3D information, and the authors applied the architecture to Brain MR images with slice thickness 1 mm to 1.5 mm. However, we observe that our training dataset has relatively large slice thickness from 5 mm to 10 mm. The experimental results also show that utilizing 3D information degrades the performance.

This paper is organized as follows. In Sect. 2, we present the proposed architecture as well as the pre-processing method. We show the experimental results from five-fold cross-validation using given 100 training dataset in Sect. 3. Conclusions and discussions are given in Sect. 4.

2 Methods

2.1 Dataset

The training datasets come from clinical exams acquired at the University Hospital of Dijon (Dijon, France), Automated Cardiac Diagnosis Challenge (ACDC) in MICCAI challenge 2017. This datasets contain cardiac short-axis MRI images with the corresponding manual reference images of LV, LV myocardium, and RV for 100 patients. Each case contains all phases of 4D images; however, manual reference images are provided only in ED (end-diastole) and ES (end-systole) phases.

The dataset is divided into 5 evenly distributed subgroups: normal case (NOR), heart failure with infarction (MINF), dilated cardiomyopathy (DCM), hypertrophic cardiomyopathy (HCM), and abnormal right ventricle (ARV).

2.2 Preprocessing

We observe that the MRI datasets provided by ACDC challenge have a wide range of in-plane dimensions from 154×224 to 428×512 . We re-scale all dataset to 256×256 by fitting maximum size of X and Y to 256 and padding residual regions with minimum value of each image. Also, the 16-bit MRI datasets have a wide range of voxel intensity that results from different scanner types or acquisition protocols. This variety can affect the performance of a segmentation model. We normalize the voxel intensity of each image by subtracting its mean then dividing it by its standard deviation.

2.3 Architecture

The provided datasets in this challenge have a large slice-thickness (5 to 10 mm) and the connectivity between adjacent slices is insufficient in this case. In short, 3D information is not considered necessarily since it can impede generalization of a model. We propose an end-to-end fully convolutional network (FCN) architecture, which is based on M-net [10]. This architecture is inspired by U-net [11]. The proposed FCN architecture has the same layers with M-net excluding the 3D convolution filter. Figure 1 illustrates our proposed FCN architecture.

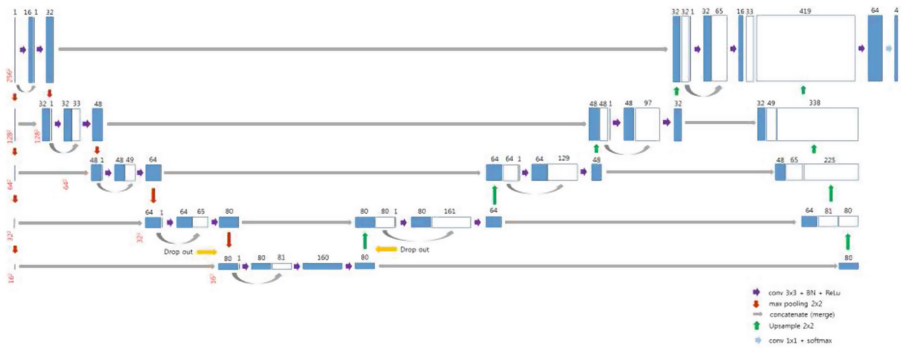


Fig. 1. The proposed FCN architecture

Our architecture has two main paths in common with M-net: Contraction and Expansion paths. Contraction path has 5 cascade steps. Each step in this path has 2 convolution layers of size 3×3 and max-pooling layer of size 2×2 . This path reduces the size of input by half and allows network to capture contextual information. As input size is reduced by max-pooling, the number of filters gradually increases to avoid bottleneck (information loss). Expansion path has the symmetric steps and layers but for replacing max-pooling with up-sampling to double the size of input. For precise localization, previous feature maps are concatenated to the corresponding next feature maps. The final layer is processed by 1×1 convolution layer with 4 channels (Background, RV endocardium, LV myocardium, and LV endocardium) and pixel-wise softmax which gives the probability of 4 classes to every pixel. The final segmentation labels are assigned to the classes with maximum probability for every pixel. Batch normalization layers are applied after each convolution layer before Relu activation. Dropout with probability 0.5 is applied to contraction and expansion only once, respectively.

To resolve bad training of a certain class due to the class imbalance (especially LV-Myo), we use weighted cross-entropy as loss function. The weight of loss function is defined based on the number of voxels in a certain class.

3 Experimental Results

3.1 Implemented Details

We divide the ACDC datasets which have ED and ES volumes of 100 patients into 80 training sets and 20 test sets to train our FCN and test its performance with five-fold cross validation. Five-fold cross validation was performed with the following details: (1) select 4 volumes sequentially in each of 5 disease classes, which gives 20 volumes in total, (2) The selected 20 volumes and the remaining 80 ones are considered as test sets and training sets, respectively, (3) Training and Testing, (4) conduct the whole process iteratively five times for ED and ES, respectively.

Considering each image consists of approximately 10 slices, it is not enough to train a model without overfitting. Also, a CNN architecture is not invariant to rotation though it is partially invariant to translation. Therefore, we perform rotation transformation from -60° to 60° at uniform intervals of 15° to augment the training datasets. For post-processing, we apply morphological operations to fill the small gap or to remove small volumes. We also apply convex hull to remove concavities only for LV.

The proposed FCN was trained on NVIDIA TITAN X, with 12 GB of RAM for 150 epochs through the training set of about 6,800 images (80 volumes), which took 18 h for training. The FCN was implemented by Tensorflow r0.11 and trained using RMSprop Optimizer with following hyper parameters: learning rate = 10^{-3} , decay = 0.9, momentum = 0.0, and epsilon = 10^{-10} .

3.2 Results and Quantitative Analysis with Other Methods

The segmentation performance is evaluated with the mean Dice Similarity Coefficient (DSC) and Hausdorff distance (HD). Let S_R and S_{GT} be the segmentation result and ground truth, respectively. The $DSC(S_R, S_{GT})$ is defined as $\frac{2|S_R \cap S_{GT}|}{|S_R| + |S_{GT}|}$, where 0 signifies the zero overlap between the ground truth and the derived segmentation result, and 1 signifies the complete overlap between ground truth and segmentation result in both the foreground and background. The $HD(S_R, S_{GT})$ is defined as $\max(\max_{x \in C_R} \min_{y \in C_{GT}} d(x, y), \max_{x \in C_{GT}} \min_{y \in C_R} d(x, y))$, where C_R and C_{GT} are contour point sets of S_R and S_{GT} , respectively, and $d(x, y)$ is the distance between two points. It is the longest distances of all which are measured from a contour point in one to the closest contour point in the other.

We compared the segmentation results of the proposed FCN with U-net, and U-net with 3D to 2D converter. U-net with converter includes 3D to 2D convolution layer in front of existing U-net layers. It takes $2n + 1$ slices as input image, a central slice and its neighboring $2n$ slices, for using 3D context. In this paper, the number of neighboring slices n was empirically assigned to one in the light of large slice thickness.

For fair comparison, we did not consider data augmentation, and the comparison is conducted only for one cross-validation subset at ED phase. The mean DSC values of each structure for three architectures are listed in Table 1. Based on these results, it is shown that the proposed FCN architecture segments three structures of interest on MRI images slightly better than two other architectures.

Table 1. Comparison of segmentation results of the proposed FCN with U-net and U-net with 3D to 2D converter.

	U-net			U-net with converter			The proposed FCN		
	RV	LV-Myo	LV	RV	LV-Myo	LV	RV	LV-Myo	LV
DSC	0.900	0.878	0.967	0.878	0.829	0.959	0.908	0.883	0.962

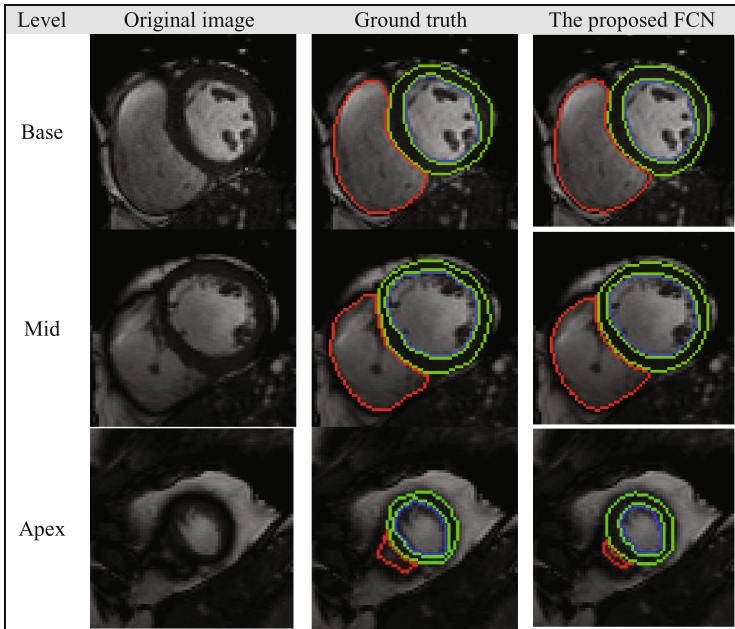
It should be noted that U-net with 3D to 2D converter for using 3D context produced lower mean DSC values than two other architectures. It is due to relatively large slice thickness of the provided datasets or image-shift by different breath-hold during acquisition. As previously mentioned, 3D information is not considered necessarily in images with thick slices and it can impede generalization of a model.

Finally, we trained the proposed FCN with the augmented datasets by rotation and gained the increased segmentation results for RV, LV-Myo and LV. Table 2 shows mean DSC values and Hausdorff distances of the proposed FCN with augmentation for each subset of five-fold cross validation and its average. There is major improvement in DSC for RV and minor improvement for LV-myocardium and LV on the same datasets (ED CV#5). It is noted that average Hausdorff distances for RV are higher than LV-Myo and LV due to many false positives especially in basal slices. We also evaluated group-based cross-validation, and the results are summarized in Table 3.

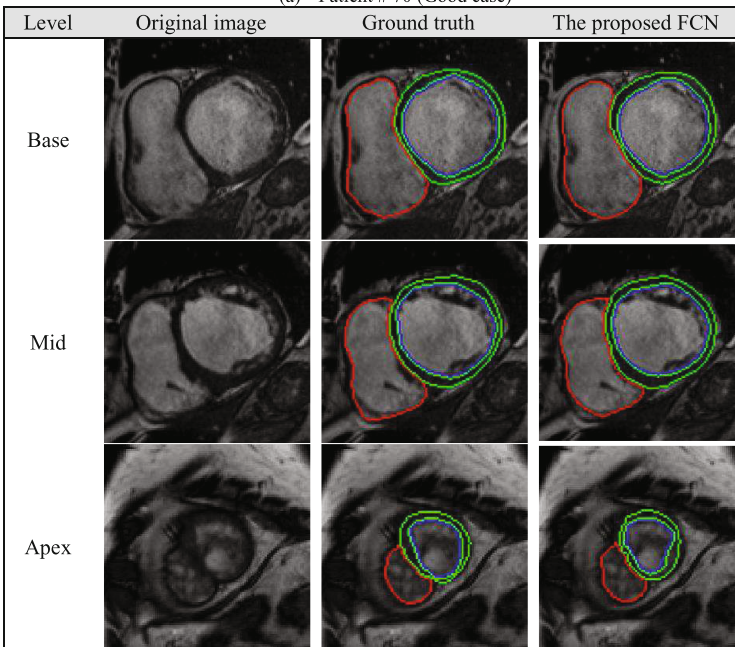
Table 2. Cross-validation (CV) results of our model on the 100 cases (training datasets are 80, test datasets are 20). Values correspond to the mean and standard deviation.

		RV		LV-Myo		LV	
		ED	ES	ED	ES	ED	ES
DSC	CV#1	0.934 ± 0.03	0.837 ± 0.09	0.880 ± 0.02	0.903 ± 0.02	0.960 ± 0.02	0.924 ± 0.05
	CV#2	0.924 ± 0.03	0.853 ± 0.07	0.866 ± 0.03	0.882 ± 0.04	0.959 ± 0.01	0.919 ± 0.04
	CV#3	0.917 ± 0.03	0.851 ± 0.10	0.861 ± 0.05	0.884 ± 0.04	0.956 ± 0.02	0.898 ± 0.09
	CV#4	0.918 ± 0.06	0.843 ± 0.09	0.853 ± 0.04	0.873 ± 0.04	0.958 ± 0.02	0.911 ± 0.08
	CV#5	0.932 ± 0.04	0.891 ± 0.04	0.892 ± 0.03	0.900 ± 0.03	0.970 ± 0.02	0.920 ± 0.05
	Avg	0.925 ± 0.04	0.855 ± 0.08	0.870 ± 0.04	0.888 ± 0.04	0.961 ± 0.02	0.915 ± 0.06
	Total	0.890 ± 0.07		0.879 ± 0.04		0.938 ± 0.05	
	HD (mm)	CV#1	12.49 ± 5.87	15.54 ± 5.05	9.23 ± 5.93	8.68 ± 3.91	7.29 ± 4.36
	CV#2	14.63 ± 5.55	16.20 ± 5.89	10.88 ± 9.02	11.86 ± 6.78	6.94 ± 4.50	9.52 ± 4.10
	CV#3	15.61 ± 7.62	15.34 ± 5.57	9.52 ± 5.27	10.49 ± 3.80	8.73 ± 5.66	8.78 ± 4.69
	CV#4	15.30 ± 7.60	16.98 ± 8.80	9.68 ± 6.21	11.83 ± 6.28	7.49 ± 8.29	9.33 ± 5.94
	CV#5	11.82 ± 4.91	14.64 ± 7.49	7.84 ± 6.50	9.22 ± 5.36	3.99 ± 2.83	8.90 ± 6.55
	Avg	12.60 ± 6.02	14.78 ± 6.36	9.47 ± 6.64	10.05 ± 5.27	6.27 ± 4.74	8.28 ± 4.67
	Total	13.69 ± 6.30		9.76 ± 6.02		7.27 ± 4.83	

Figure 2 shows segmentation results for three different levels (position) of two sample volumes aligned by short axis of heart. Figure 2(a) and (b) represent a good case without any and a case with LV trabeculations and partial volume effect at apical level, respectively. The segmentation results by our FCN are coterminous with the provided ground truth for both two cases as shown in Fig. 2.



(a) Patient # 70 (Good case)



(b) Patient#17 (with LV trabeculations at apex)

Fig. 2. The segmentation results for three different levels (base, middle and apex) of two sample slices aligned by short axis of heart. (a) and (b) represent a good case without any and a case with LV trabeculations and partial volume effect at apical level, respectively. Red: Right Ventricle, Blue: Left Ventricle, Green: Myocardium. (Color figure online)

Table 3. Group-base analysis results of our model on the 100 cases with five-fold cross-validation (training datasets are 80, test datasets are 20).

Group	DSC			HD (mm)		
	RV	LV-Myo	LV	RV	LV-Myo	LV
DCM	0.898 ± 0.07	0.879 ± 0.03	0.966 ± 0.01	15.37 ± 8.28	8.84 ± 4.72	5.83 ± 2.58
HCM	0.867 ± 0.09	0.903 ± 0.03	0.889 ± 0.08	12.89 ± 7.43	11.98 ± 6.73	9.57 ± 5.15
MINF	0.865 ± 0.08	0.863 ± 0.04	0.952 ± 0.02	15.06 ± 6.71	9.40 ± 4.25	7.02 ± 3.32
NOR	0.909 ± 0.07	0.896 ± 0.03	0.945 ± 0.04	10.86 ± 4.80	7.55 ± 4.59	6.31 ± 4.12
ARV	0.907 ± 0.04	0.857 ± 0.04	0.930 ± 0.05	13.23 ± 4.85	11.47 ± 8.18	7.96 ± 6.98

We note that the mean DSC values for RV and LV on ES are relatively degraded compared with ED as shown in Table 2. The issues about it and LV trabeculations will be discussed on Sect. 4 in detail.

The CPU and GPU run times to segment one volume using the proposed FCN are approximately 8.09 s and 0.62 s, respectively.

4 Conclusion and Discussion

In this paper, we proposed a new FCN architecture for three structures (RV, LV-Myo and LV) segmentation on MRI images. It has the same layers as M-net excluding 3D-to-2D converter layer. As we observe that the datasets provided by ACDC challenge have large slice thickness and image-shift due to different breath-hold during acquisition, we think that considering 3D information can impede generalization of a model. Therefore, the proposed FCN architecture combines U-net architecture and skip connections of M-net to learn better features. Experimental results on the provided datasets showed that the proposed FCN has better performance for RV, LV-Myo and LV segmentation than other current state-of-the-art models. It is well known that CNN is not invariant to rotation. We found that the orientation of the provided datasets varies approximately from -60° to 60° by volume. For this reason, we applied data augmentation for rotation and the DSC values were slightly improved especially for RV, which has variant shape to rotation unlike LV-Myo and LV.

Most segmentation errors occur at basal and apical slices of volumes aligned by short-axis of heart as shown in Figs. 2 and 3. Usually, LV contours are approximately delineated as an ellipse which includes LV Trabeculations. These are located near boundary at basal and middle level, but near center at apical level. As shown in Fig. 2(b), it seems that there exist two different structures in LV cavity, which become faint with partial volume effect. These reasons make it difficult to segment structures of interest at apical level.

On the other hand, there are some cases without RV labels for ground truth at basal level as shown in Fig. 3. These mismatch between RV labels by our FCN and ground truth at basal level causes performance decreases in DSC and Hausdorff measures. RV and LV are connected to pulmonary artery and ascending aorta, respectively, and these connected regions are usually observed at basal level. Although regions which belong to RV or LV still remain at basal level, these regions are occasionally not included in

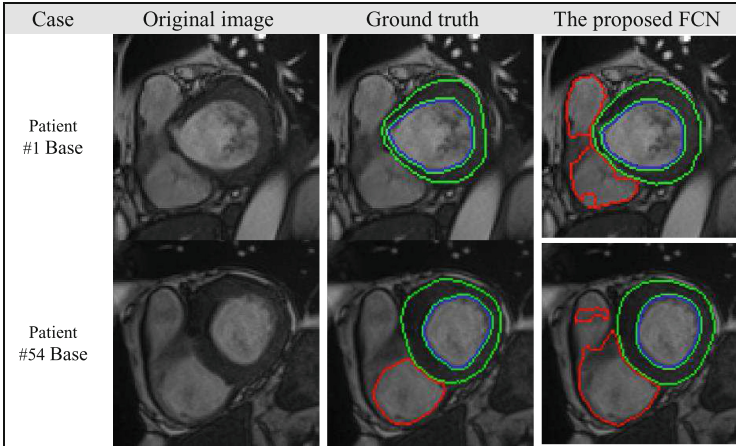


Fig. 3. The mismatch between RV labels by our FCN and ground truth at basal level.

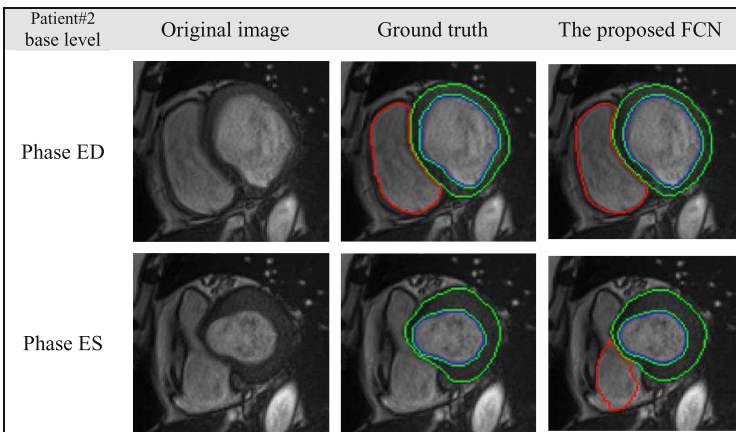


Fig. 4. The different shapes of structures of interest on ED and ES at the same basal level.

clinical setting. It is very difficult to make the distinction on a static slice, although the regions belonging to RV look similar in the first and second rows in Fig. 3. Also, it can be explained why the mean DSC values for RV and LV on ES are relatively poor from the similar reason. In case of patient#2, at the same basal level of ED and ES, these connected regions are observed only on the image on ES as shown in Fig. 4. Thus, it needs to consider for the additional processes at basal level.

In the future, we will apply further sophisticated post-processing algorithms using the probability maps for structures of interest produced by the proposed FCN as well as level classification methods for additional performance improvement at basal and apical level.

Acknowledgement. This work was supported by Institute for Information & communications Technology Promotion (IITP) grant funded by the Korea government (MSIT) (2017-0-00255, Autonomous digital companion framework and application).

References

1. Kang, D., et al.: Heart chambers and whole heart segmentation techniques. *J. Electr. Imaging* **21**(1), 010901-1–010901-16 (2012)
2. Petitjean, C., et al.: A review of segmentation methods in short axis cardiac MR images. *Med. Image Anal.* **15**(2), 169–184 (2011)
3. Bai, W., et al.: A probabilistic patch-based label fusion model for multi-atlas segmentation with registration refinement: application to cardiac MR images. *IEEE Trans. Med. Imaging* **32**(7), 1302–1315 (2013)
4. Bai, W., et al.: Multi-atlas segmentation with augmented features for cardiac MR images. *Med. Image Anal.* **19**(1), 98–109 (2015)
5. Avendi, M., et al.: A combined deep-learning and deformable-model approach to fully automatic segmentation of the left ventricle in cardiac MRI. *Med. Image Anal.* **30**, 108–119 (2016)
6. Wolterink, J.M., Leiner, T., Viergever, M.A., Išgum, I.: Automatic coronary calcium scoring in cardiac CT angiography using convolutional neural networks. In: Navab, N., Hornegger, J., Wells, W.M., Frangi, A.F. (eds.) MICCAI 2015. LNCS, vol. 9349, pp. 589–596. Springer, Cham (2015). https://doi.org/10.1007/978-3-319-24553-9_72
7. Poudel, R.P.K., Lamata, P., Montana, G.: Recurrent fully convolutional neural networks for multi-slice MRI cardiac segmentation. In: Zuluaga, M.A., Bhatia, K., Kainz, B., Moghari, M.H., Pace, D.F. (eds.) RAMBO/HVSMR -2016. LNCS, vol. 10129, pp. 83–94. Springer, Cham (2017). https://doi.org/10.1007/978-3-319-52280-7_8
8. Zhen, X., et al.: Multi-scale deep networks and regression forests for direct bi-ventricular volume estimation. *Med. Image Anal.* **30**, 120–129 (2016)
9. Litjens, G., et al.: A survey on deep learning in medical image analysis. [arXiv:1702.05747](https://arxiv.org/abs/1702.05747) (2017)
10. Mehta, R., et al.: M-Net: a convolutional neural network for deep brain structure segmentation. In: International Symposium on Biomedical Imaging, pp. 437–440 (2017)
11. 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). https://doi.org/10.1007/978-3-319-24574-4_28