

Automated Left and Right Chamber Segmentation in Cardiac MRI Using Dense Fully Convolutional Neural Network

Marco Penso¹, Sara Moccia^{2,3}, Stefano Scafuri¹, Giuseppe Muscogiuri¹, Gianluca Pontone¹, Mauro Pepi¹, Enrico Gianluca Caiani^{4,5}

¹Clinical Cardiology Unit and Department of Cardiovascular Imaging, Centro Cardiologico Monzino IRCCS, Milan, Italy

²Department of Information Engineering, Università Politecnica delle Marche, Ancona, Italy

³Department of Advanced Robotics, Istituto Italiano di Tecnologia, Genoa, Italy

⁴Department of Electronics, Information and Bioengineering, Politecnico di Milano, Milan, Italy

⁵National Council of Research (CNR) IEIIT, Milan, Italy

Abstract

Cardiac magnetic resonance (CMR) represents the gold standard for the diagnosis of cardiovascular diseases. We developed a deep learning approach for the automatic detection and segmentation of left and right ventricles and myocardium (Myo) on short-axis cine CMR images, including all clinically relevant slices. A dataset of 210 studies (3 pathology groups) was considered: Images were acquired and manually segmented (gold standard, GS) at Centro Cardiologico Monzino (Milan, Italy). Automatic segmentation was performed with a U-Net inspired architecture where two loss functions were used: weighted cross entropy (WCE) and its combination with the Dice loss function (WCE+Dice). Two experiments were conducted: A) all the slices were included; ii) slices where the Myo did not completely surround the LV were removed. To evaluate the clinical relevance of our approach, the predicted segmentation was reviewed and corrected by an expert physician. The two loss functions performed similarly, with slightly better results for WCE, resulting in a strong correlation with the manually-adjusted segmentation.

1. Introduction

Cardiovascular disease is one of the principal causes of death in the world. Cardiac imaging is used to estimate structural and functional cardiac parameters for both diagnosis and disease evaluation. Cardiac Magnetic Resonance (CMR) is the reference technique for estimating parameters such as left (LV) and right (RV) ventricular end-diastolic (EDV) and end-systolic (ESV) volumes, stroke volume (SV), ejection fraction (EF), and mass [1], as a result of the endocardial and epicardial

contour segmentation process. Usually, this operation is manually performed, resulting in a tedious, time-consuming and error-prone analysis. For this reason, in the last decade several approaches to automatize this process were investigated, lately with a particular interest for machine learning (ML) algorithms, that are emerging as the new state of the art [2]. To possibly compare different methods in terms of performance, over the past years, different databases of CMR images were made available through many challenges [3], but generally they do not include the most apical and basal slices, or they are limited to only one of the two ventricles.

In this paper we propose a dense fully convolutional neural network (FCN) architecture, for semantic segmentation of LV, RV and LV myocardium (Myo) on short-axis (SAX) cine CMR images, applied to all the slices relevant in clinical practice, including the most basal and apical ones.

2. Method

2.1. Patient population

A retrospectively selected dataset consisting of SAX cine (steady-state free-precession) CMR images (size 512x512 pixels, in-plane pixel resolution 0.59-0.70 mm) acquired (GE, 1.5 T) from 210 patients at Centro Cardiologico Monzino (Milan, Italy), was analysed. For each patient, the gold standard (GS) manual tracings of the LV and RV endocardium and epicardium at the end-diastolic (ED) and end-systolic (ES) frames, obtained by an expert cardiologist, were available. The dataset includes patients of three medical groups, equally distributed, namely: hypertrophic cardiomyopathy (HCM), dilated cardiomyopathy (DCM) and ventricular arrhythmias (VA).

2.2. Network architecture

The proposed network architecture was inspired from U-Net [4], one of the most successful FCN for medical segmentation tasks, which has symmetric encoder and decoder and skip connections between corresponding layers in the encoder and decoder. Each step of the down-sampling path consists of two 3x3 convolutions with batch normalization and rectified linear unit (ReLU) as activation function. This sequence is called convolutional block (conv block). Each block is followed by a max-pooling operation with 2x2 size and stride of 2. In this paper, the up-sampling path was modified, using transposed convolution instead of up-convolution [5] and reducing the number of filters in the transposed convolution layer, with a number of feature maps equal to the number of classes. As reported in [6], the full complexity of the up-sampling path is not necessary, with benefits in term of learnable parameters reduction. A series of 4x4 transposed convolution with stride 2 and conv block were repeated at each layer. A final 1x1 convolution layer with softmax activation function is used to generate the final segmentation maps. Each conv block had a different number of filters, from 48 (first layer) to 768 (bottom layer). Skip connections were modified incorporating dense blocks [7], to alleviate the semantic gap [10] between the up- and down-sampling path. Each layer of the dense block consisted of 3x3 convolutions with ReLU and batch normalization. Each layer concatenated its own features maps with the feature maps of all preceding layers and transferred its own feature maps to all following, adding and preserving information.

2.3. Pre-processing and network setting

Images were cropped and rescaled to 176x176 pixels size, voxel were normalized in the [0, 1] range. In addition, images were rotated. In cine CMR images, anatomical structures of interest are generally highly imbalanced, posing challenges for FCN training. To overcome this problem, two different loss functions were tested: the weighted cross-entropy [4], (WCE) to manage the class imbalance between the background and the foreground classes, and a combination of WCE with *Dice* loss (WCE+*Dice*). The *Dice* loss can be calculated as:

$$Dice\ loss = 1 - Dice \quad (1)$$

where the *Dice* calculates the overlap between foreground regions:

$$Dice = \frac{2 \sum p_{h,w} t_{h,w}}{\sum p_{h,w} + \sum t_{h,w}} \quad (2)$$

where h is the number of classes of the foreground, w is

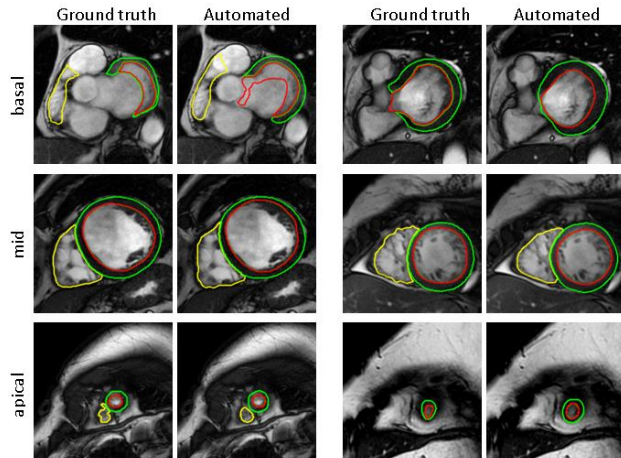


Figure 1. Example segmentation using the proposed FCN compared to the gold standard (RV endocardium: yellow; LV endocardium: red; LV epicardium: green).

the number of pixels, t is the ground truth label per location and p is prediction of the sigmoid function. Therefore, the WCE+*Dice* loss was:

$$loss = WCE + 0.2\ Dice\ loss \quad (3)$$

Online data augmentation was implemented using gamma correction, blurring, and random rotation in $[-15^\circ, 15^\circ]$ range. The dataset was randomly split into training (70%), validation (15%) and testing sets (15%). The best model among epochs was selected as the one with the highest *Dice* score on the validation set.

2.4. Experimental protocol

Performance of the FCN was measured using sensitivity (Se), *Dice* and the Hausdorff distance (HD) [8] compared to the ground-truth mask. To assess the influence of the basal slice, the network was trained: 1) using all the slices including basal slices according to guidelines [9]; 2) removing slices where the Myo did not completely surrounded the LV. Network performance was evaluated considering the two different loss functions (WCE, WCE+*Dice*). To evaluate a real clinical scenario, a physician reviewed the prediction on the test set, correcting manually when needed the contours. The adjusted segmentations and the automated results were compared respect to the ground-truth, in term of actual LV and RV volumes (in ml) and LV mass (in g). Correlation and Bland-Altman analyses were reported.

3. Results

Table 1 reports the performance of the FCN segmentation of LV, RV and Myo for the two defined criteria. Performance of the two loss function were compared: WCE reached slightly better results than

Table 1. Segmentation results of the proposed FCN using two different loss function, measured as Sensitivity (Se), $Dice$ coefficient, and Hausdorff distance (HD). Se and $Dice$ are reported as mean, while HD is reported as median.

Basal slices ($\geq 50\%$ Myo around the LV)			
LV			
	Dice	HD [mm]	Se
WCE	0.944	7.2	0.948
WCE+Dice	0.942	7.4	0.966
RV			
	Dice	HD [mm]	Se
WCE	0.908	7.3	0.881
WCE+Dice	0.892	9.3	0.844
Myo			
	Dice	HD [mm]	Se
WCE	0.851	5.9	0.828
WCE+Dice	0.849	5.8	0.815
Only basal slices with 100% Myo around the LV			
LV			
	Dice	HD [mm]	Se
WCE	0.946	6.2	0.950
WCE+Dice	0.944	6.5	0.969
RV			
	Dice	HD [mm]	Se
WCE	0.909	7.1	0.882
WCE+Dice	0.897	8.2	0.853
Myo			
	Dice	HD [mm]	Se
WCE	0.853	5.4	0.831
WCE+Dice	0.851	5.5	0.817

WCE+ $Dice$ loss. Considering cumulatively all the slices, no significant differences between the two basal selection criteria were observed. Further analyses were conducted using WCE loss function only.

In Table 2, results of the FCN before and after manual correction, at the basal level according to the two defined selection criteria. No significant differences were found, highlighting the high inter-observer variability for basal contours.

Table 3 reports the correlation and the Bland-Altman analysis of clinical parameters, for the predicted and the manually adjusted segmentation, compared to the GS. For LV ED and ES volumes, as well as for RV ED the FCN resulted in a strong correlation with no bias and narrow limits of confidence, both with and without correction, while a slight overestimation measured for RV ES compared to the GS was observed. For Myo, mass a very good correlation was visible, with a slight overestimation of both the FCN and the following correction compared to the GS, and wider (but still acceptable) limits of agreement. In Figure 1, some examples of the predicted contours compared with the corresponding GS are shown.

When the Myo did not completely surround the LV, the predicted segmentation sometimes failed including the aortic root in the LV. Similarly, for the RV, sometimes the right atrium was included in the ventricular segmentation. These errors on the basal slices did not significantly affect the cumulative performance.

Table 2: Segmentation performance measured as Dice coefficient, and Hausdorff distance (HD) reported as median, before and after manual correction at the basal level.

Basal ($\geq 50\%$ Myo around the LV)				
LV				
	Dice	HD [mm]	Myo	
	Dice	HD [mm]	Dice	HD [mm]
FCN+corr	0.927	7.5	0.811	13.5
FCN	0.909	9.7	0.808	13.0
Basal (100% Myo around the LV)				
LV				
	Dice	HD [mm]	Dice	HD [mm]
FCN+corr	0.958	4.7	0.874	4.5
FCN	0.959	5.0	0.873	4.5

Table 3: Clinical metrics before and after manual correction reported in term of Correlation (R^2) and bias ± 2 *standard deviation (Bias $\pm 2\sigma$). Results considering all the slices. EDV: End diastolic volume; ESV: End systolic volume.

LV EDV (ESV)		
	R^2	Bias $\pm 2\sigma$ [ml]
FCN+corr	0.99 (0.98)	-2 \pm 13 (-2 \pm 14)
FCN	0.99 (0.98)	-1 \pm 12 (-1 \pm 14)
RV EDV (ESV)		
	R^2	Bias $\pm 2\sigma$ [ml]
FCN+corr	0.98 (0.96)	4 \pm 14 (7 \pm 12)
FCN	0.99 (0.95)	6 \pm 12 (6 \pm 14)
Myo ED Mass (ES Mass)		
	R^2	Bias $\pm 2\sigma$ [ml]
FCN+corr	0.92 (0.91)	7 \pm 26 (6 \pm 28)
FCN	0.92 (0.92)	6 \pm 26 (5 \pm 26)

4. Discussion

We presented a FCN for the automated segmentations of LV, RV and Myo on cardiac SAX cine-MRI images. The network was based on a U-Net with dense skip connections, in order to improve its performance. Skip connections are known to be effective in recovering information lost during the down-sampling path. Directly concatenating semantically different feature maps, from the down-sampling to the up-sampling, it might generate a semantic gap which could negatively affect the segmentation accuracy [10]. To deal with this problem, we took advantage from dense skip connections with

benefit even on the optimization of the network. The two compared loss functions performed similarly, with the WCE able to reach a slightly better performance in terms of mean *Dice* and *HD*. We limited the impact of the *Dice* loss because even if it can consider global loss information, WCE presents better differentiable properties which results in an easier optimization process. Further analysis considering different weighting strategies between *Dice* and WCE loss are needed to evaluate more in depth their potential contribution to the optimal results.

During the past years different CMR datasets based on SAX cine images were made accessible in different challenges for testing segmentation algorithms, but generally they do not include the most apical and basal slices, they are limited to only one of the two ventricles, or refer to specific clinical conditions. This affects the representation of real clinical scenario, limiting the potential of deep learning (DL) algorithms that are developed, trained and tested on such CMR datasets. The database of images used in this study represents a step forward towards a more realistic training and testing of the DL algorithm for clinical applicability.

To evaluate the impact of the inclusion criteria of the basal slice, performance needs to be reported separately for those slices, while considering all slices it does not significantly affect the global results. Accordingly, for a better comparison of performance of different segmentation algorithms, inclusion criteria for basal slices should be always reported, as well as the corresponding specific results. As expected the segmentation for basal slices is tedious due to the presence of valves and outflow tract, which contribute to increase the variability, with consequences on the stability of the training phase. Despite these complications, performance of our FCN on basal slices revealed to be good. The segmentation of the RV resulted more complicated than for the LV, as indicated by the lower *Dice* and higher *HD*, probably due to the more complicated geometry, higher variability from the base to the apical section and a general lower contrast of the RV. Myo was also segmented, with appreciable results, but a general greater difficult when compared to the LV and RV. When considering the correlation and Bland-Altman in the computed clinical parameters from the automated FCN contours, and after manual supervision and correction when needed, the results showed a strong agreement, with minimal contribution of the correction to the final results, thus highlighting the potential of the developed method in generating results comparable with an expert operator variability.

5. Conclusions

We presented a FCN for semantic segmentation of LV, RV and Myo in CMR images, reaching high correlation and volumes estimation with manual gold standard, able

to manage the high real world variability and complexity of cardiac chambers morphology, even when basal slices were considered according to clinical analysis guidelines. These results demonstrate the potential of the DL for accurate and fast cardiac segmentation in real clinical applications.

References

- [1] SE. Petersen et al., "Reference ranges for cardiac structure and function using cardiovascular magnetic resonance (CMR) in Caucasians from the UK Biobank population cohort," *J Cardiovasc Magn Reson*, vol.19, no. 18, Feb. 2017.
- [2] G. Litjens et al., "A survey on deep learning in medical image analysis," *Med Image Anal*, vol. 42, pp. 60-88, Dec. 2017.
- [3] O. Bernard et al., "Deep learning techniques for automatic MRI cardiac multi-structures segmentation and diagnosis: is the problem solved?," *IEEE Trans Med Imaging*, vol. 37, no. 11, pp. 2514-2525, Nov. 2018.
- [4] O. Ronneberger et al., "U-net: convolutional networks for biomedical image segmentation," *Miccai*, pp. 234-241, Oct. 2015.
- [5] MD Zeiler et al., "Deconvolutional networks," *Proc IEEE Comput Soc Conf Comput Vis Pattern Recognit*, pp. 2528-2535, 2010.
- [6] CF Baumgartner et al., "An exploration of 2D and 3D deep learning techniques for cardiac MR image segmentation," *Proc. STACOM-MICCAI, LNCS*, vol. 10663, pp. 111-119, Sep. 2017.
- [7] G. Huang et al., "Densely connected convolutional networks," *IEEE Comput Soc Conf Comput Vis Pattern Recognit*, pp. 2261-2269, 2017.
- [8] A. Suinesiaputra et al., "A collaborative resource to build consensus for automated left ventricular segmentation of cardiac MR images," *Medical image analysis*, vol. 18, no. 11, pp. 50-62, 2014.
- [9] J. Schulz-Menger et al., "Standardized image interpretation and post processing in cardiovascular magnetic resonance: society for cardiovascular magnetic resonance (SCMR) board of trustees task force on standardized post processing," *J Cardiovasc Magn Reson*, vol 15, no.35, pp. 1167-1186, May. 2013
- [10] N. Ibtehaz et al., "MultiResUNet: rethinking the U-Net architecture for multimodal biomedical image segmentation," *Neural Networks*, vol. 121, pp.74-87,2020

Address for correspondence:

Prof. Enrico G. Caiani
 Politecnico di Milano
 Department of Electronics, Information and Bioengineering,
 20133 Milan, Italy
 enrico.caiani@polimi.it