Deep Learning Segmentation of the Left Ventricle in Structural CMR: Towards a Fully Automatic Multi-scan Analysis

Hakim Fadil, John J. Totman, and Stephanie Marchesseau(✉)

Clinical Imaging Research Centre, A*STAR-NUS, Singapore, Singapore
stephanie.marchesseau@circ.a-star.edu.sg

Abstract. In the past three years, with the novel use of artificial intelligence for medical image analysis, many authors have focused their efforts on defining automatically the ventricular contours in cardiac Cine MRI. The accuracy reached by deep learning methods is now high enough for routine clinical use. However, integration of other cardiac MR sequences that are routinely acquired along with the functional Cine MR has not been investigated. Namely, T1 maps are static T1-based images that encode in each pixel the T1 relaxation time of the tissue, enabling the definition of local and diffuse fibrosis; T2 maps are static T2-based images that highlight excess water (edema) within the muscle; Late Gadolinium Enhancement (LGE) images are acquired 10 min after injection of a contrast agent that will linger in infarct areas. These sequences are acquired in short-axis plane similar to the 2D Cine MRI, and therefore contain similar anatomical features. In this paper we focus on segmenting the left ventricle in these structural images for further physiological quantification. We first evaluate the use of transfer learning from a model trained on Cine data to analyze these short-axis structural sequences. We also develop an automatic slice selection method to avoid over-segmentation which can be critical in scar/fibrosis/edema delineation. We report good accuracy with dice scores around 0.9 for T1 and T2 maps and correlation of the physiological parameters above 0.9 using only 40 scans and executed in less than 15 s on CPU.

1 Introduction

Cardiac imaging is an active research field with increasingly new imaging techniques being proposed to better understand, diagnose, treat and anticipate the evolution of cardiac diseases, especially after myocardial infarction. Manual analysis of these images made of thousands of slices is extremely tedious, and more importantly leads to high variability between observers which impacts diagnosis as well as research methodology (e.g. increased recruitment is needed to account for measurement variability). Hence, automatic and reproducible pipelines to process all imaging data are strongly needed. With the emergence of high computational power and the availability of massive labeled image data, recent advances in deep neural networks have made fully automatic cardiac MRI segmentation possible.

M. Pop et al. (Eds.): STACOM 2018, LNCS 11395, pp. 40–48, 2019.
https://doi.org/10.1007/978-3-030-12029-0_5

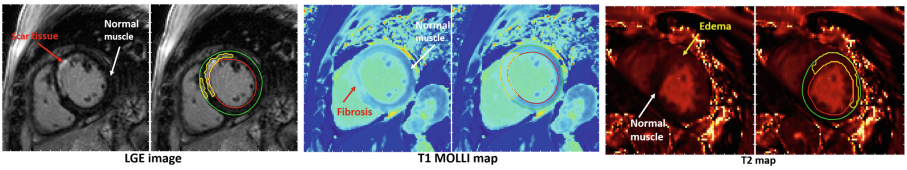

Fig. 1. Illustration of the three studied structural MRI sequences for segmentation of the scar tissue (in LGE), fibrosis (in T1 maps) and edema (in T2 maps). (Color figure online)

1.1 Deep Learning for Cardiac Cine MRI

Last year's Automated Cardiac Diagnosis Challenge [10] allowed the comparison of several deep learning models for the segmentation of the left and right ventricles from Cine MRI. Very high accuracies were reached, for instance by models derived from the DenseNets model [6] which connects each layer to every other layer; or the U-Net architecture [1,4,11] made of a series of symmetric downsampling convolutional layers followed by upsampling convolutional layers and skip connections for each resolution.

A modified version of the 2D U-Net was optimized for cardiac segmentation [1] in which the authors reduced the up-sampling path complexity by setting the number of filters in the up-convolution layers to the number of classes. Since this model gives very accurate results on our Cine data, we based our implementation on this modified 2D U-Net model.

1.2 Structural MRI Sequences

In addition to Cine MRI, which allows the evaluation of functional parameters, most MRI scans also include structural images. In this paper we will focus on three of the available sequences that are routinely used in cardiac scans.

Late Gadolinium Enhancement - Acquired in most imaging centers, short-axis Late Gadolinium Enhancement (LGE) images (also known as delayed enhancement images) are T1-weighted images acquired 10–15 min after injection of a contrast agent (gadolinium) that washes-out more slowly in infarct tissue than healthy tissue, resulting in areas of high intensities in the presence of scarring (see example Fig. 1 (left)). Despite many studies, no consensus has yet been reached to determine the best segmentation method to extract scar tissue once the myocardium is segmented [8]. On the other hand, only very few papers focus their energy on segmenting the myocardium in these sequences, using either deformable models [2], or Cine to LGE registration [12]. As one can notice on Fig. 1 (left), the intensity difference between the blood pool and the area of interest is rather low, leading to high difficulty in distinguishing accurately between endocardium and scar, making automatic segmentation of the myocardium in LGE images for scar quantification an open challenge.

T1 maps - Short-axis T1 maps are routinely acquired since the development of fast acquisition methods such as the Modified Look-Locker imaging (MOLLI), for its ability to measure T1 relaxation times without contrast. More specifically, myocardial T1 relaxation time is prolonged in most forms of pathology. Unlike LGE images for which only local abnormalities can be measured, T1 MOLLI maps can distinguish diffuse as well as local fibrosis by comparison with known reference T1 values. An example of local fibrosis definition (in yellow) on T1 MOLLI map is shown Fig. 1 (middle). In order to assess diffuse fibrosis, the T1 value of the entire myocardium must be compared to healthy muscle, therefore in T1 maps the accuracy of the LV segmentation is crucial for diffuse fibrosis quantification as well as further thresholding for local fibrosis quantification.

T2 maps - Similarly, T2 maps are acquired to measure myocardial edema, which manifests as free water content with prolonged T2 relaxation times. Differentiating edema from infarction has been shown to be useful to characterize myocardial salvage by comparison with LGE, and hence quantify the success of coronary revascularization surgery. Prolonged T2 value is also an indicator of diffuse pathology such as cardiomyopathy and is therefore a key biomarker to measure. An example of T2 maps is presented Fig. 1 (right) where a zone with edema is delineated in yellow.

In this paper, we propose to automatically segment the endocardium and epicardium contours on the three structural images described above. To the best of our knowledge, this is the first attempt to automatically segment the left ventricle in T1 and T2 maps. Since a large amount of Cine data is available for training compared to available structural data, we will test a transfer learning approach from the Cine model as well as a full U-Net architecture on our dataset made of 40 patients.

2 Methods

2.1 Available Data: Pre and Post-processing

The T1 and T2 maps being relatively new sequences, and varying from scanner to scanner, only 40 patient scans were available at the time of this study. The same number of LGE data was used for fair comparison. The train/validation partition was set to 80/20% for each experiment without data augmentation. All images were acquired post-myocardial infarction on Siemens scanners (3T mMR and 3T Prisma). To account for various image sizes and resolutions, all 2D slices were rescaled to 212×212 pixels of $1.37\,\text{mm} \times 1.37\,\text{mm}$ resolution, and intensities were normalized. Softmax predictions were rescaled to the original resolution and size, and 3D masks were created choosing the labels with the highest score. Finally, only large connected components were kept, and a convex hull was defined to smooth the endocardium and epicardium contours. Since no consensus exist on the best way to define scar, fibrosis or edema once the left ventricle contours exist, we decided to use 2-dimensional Otsu automatic threshold [5] to estimate the amount of damaged tissue for both the ground truth contours (obtained manually) and the predicted contours, instead of relying on a manual ground truth.

2.2 Transfer Learning from U-Net vs Full U-Net Model

Transfer learning is based on the hypothesis that in a deep neural network, the first layers extract global features that are similar among many types of images, hence do not require to be retrained for every set of similar images. In our case, all images are short-axis slices showing similar shapes for which only the contrast between the various structures changes. We therefore hypothesize that training only the last few layers will allow a smaller database to be used for training while giving accuracy similar to a full U-Net model. A transfer model also has the advantage of requiring less computational memory and time and can therefore be trained on CPU. To apply transfer learning, we base our code on the open-source 2D U-Net python code [1]. We apply our pretrained Cine model up to the last 3 layers to create the input image and define a small model (named *3 Layers*) only made of the last 3 layers of U-Net (3 convolution blocks with 64, 64 and 3 filters). This model as well as the full U-Net model (named *Full Model*) are optimized to maximize the foreground dice (combining endocardium and myocardium contours) with Adam optimizer (learning rate of 0.01, $\beta_1 = 0.9$, $\beta_2 = 0.999$, batch size = 5). The training of the *3 Layers* and *Full Model* took respectively 5 h and 20 h for 40K steps on a GPU NVIDIA K40. The volume segmentation takes less than 14 s on an Intel Xeon E5-1650 CPU.

2.3 Automatic Image Selection

Imbalance between foreground and background data led to over-segmentation of basal and apical slices that should not be included in the segmentation (do not contain myocardium), as previously noticed in right ventricle segmentation using the full 2D U-Net on Cine MRI. To cope with this issue, we decided to independently train a binary classification network that automatically selects for each scan the slices that include the myocardium. We designed a neural network (see Fig. 2) made of 7 convolutional blocks, a flatten layer and two fully connected blocks, inspired by the AlexNet [7] architecture. Each convolution block consists of a 3×3 convolution without padding, followed by a 2×2 max pooling layer and a ReLU activation. The model was optimized to fit weighted cross-entropy with a weight of 0.2 to decrease false negative rate since removing a slice to segment is more critical than over-segmenting. The decision to remove a slice was made if the probability to remove it was 10 times higher than the probability to keep it. We used Adam optimizer (learning rate of 0.0001, $\beta_1 = 0.9$, $\beta_2 = 0.999$, batch size = 16) and dropout layers after each block, to regularize the model that initially showed high variance, with the keep prob adjusted for each sequence from 0.7 to 0.9. The model was trained in approximately 2000 steps which took 20 min on CPU. The classification of each scan took less than 1 s on CPU. Each 3D scan is therefore first segmented through the U-Net model and then over-segmentation are deleted using the results of this selection network applied on the original scan.

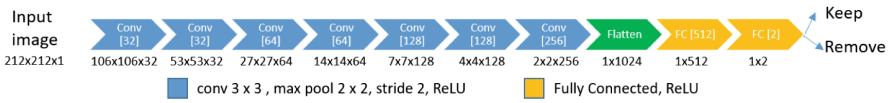

Fig. 2. Representation of the designed CNN, inspired by AlexNet [7] architecture. Each blue arrow represents a convolution layer made of a 3×3 convolution, a 2×2 max pooling and a ReLU activation. The number of filters is written in each arrow. A flatten layer and two fully connected layers terminate the network. Resulting image sizes are written below each block. (Color figure online)

3 Results

3.1 Accuracy of Image Selection for Structural Images

Table 1 reports the sensitivity (true positive rate, where positive means keep a slice), specificity (true negative rate, where negative means remove a slice) and total accuracy for the training and validation sets. Dropout coefficients were adjusted to each sequence to balance the high variance problem we observed in our first experiments. Notice the very high sensitivity rates in the validation set that guarantees that all slices to be segmented will be rightly kept. The model also manages to remove between 78% (for T2 maps) and 91% (for T1 maps) of all unwanted slices.

Table 1. Sensitivity, specificity and accuracy of the designed method to select the slices to keep for further segmentation. KP: keep prob in the dropout regularization

	Training (n = 384 slices)			Validation (n = 81 slices)		
	Sensitivity	Specificity	Accuracy	Sensitivity	Specificity	Accuracy
LGE (KP = 0.7)	98%	74%	91%	100%	85%	96%
T1 maps (KP = 0.9)	100%	100%	100%	97%	91%	94%
T2 maps (KP = 0.8)	89%	76%	86%	100%	78%	91%

3.2 Comparison of the Methods

Using the image selection results, corrections were applied (slices were removed according to the classification) after prediction, for the *3 Layers* and the *Full Model* leading to a comparison of four methods. Results were measured in terms of dice scores for the endocardium and the myocardium masks. Additionally, using the Otsu thresholding to define damaged zones, Pearson's correlation (R), bias and reproducibility error (RPC) were assessed for the scar size. For T1 and T2 maps, since the actual relaxation times are also predictors of cardiac dysfunction, the mean value within the myocardium was also measured and a Pearson's correlation reported.

Table 2. Segmentation accuracy and statistical results for the estimation of infarct size using the various proposed methods on LGE images. Bold values represent best values in the validation set.

	Training(n = 32)					Validation (n=8)				
	Dice		Scar %LVM			Dice		Scar %LVM		
	Endo	Myo	R	Bias	RPC	Endo	Myo	R	Bias	RPC
3 Layers	0.93	0.84	0.97	1.96	3.63	0.83	0.68	0.94	0.09	5.58
Corrected 3L	0.95	0.85	0.96	2.11	4.38	0.87	**0.69**	**0.95**	0.16	**4.92**
Full Model	0.84	0.88	0.97	−0.82	3.85	0.78	0.64	0.91	−1.11	7.34
Corrected FM	0.96	0.96	0.99	−0.03	2.06	**0.89**	0.66	0.94	**0.05**	5.67

LGE Image Segmentation. As shown in Table 2, training accuracy was extremely high for the *3 Layers* (with or without correction) and the *Corrected Full Model*, which proves the capacity of these models to estimate myocardium contours in LGE images despite the close signal intensity values between blood pool and scar tissue. However, the validation set did not reach such performance in terms of dice scores despite a good correlation in the scar size. Fair and poor results are mainly due to a wrong endocardium segmentation that includes the infarct zone as shown Fig. 4 (left). Since scar appearance in MRI varies considerably in shape, location and intensity from patient to patient, and depends on the contrast injected and the acquisition method, we believe a larger database is mandatory to reduce the high variance we obtained.

T1 Maps Segmentation. As illustrated in Fig. 3 (left), best dice scores and correlations are reported with the *Corrected Full Model* (dice = 0.96 for endocardium, dice = 0.92 for myocardium, R = 0.996 for Mean T1 and R = 0.983 for fibrosis size) while the *3 Layers* model seems unreliable for fibrosis detection. Figure 4 (middle) gives three example segmentations of local fibrosis and

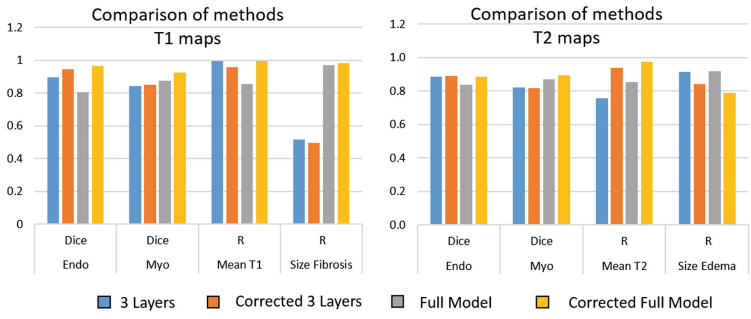

Fig. 3. Segmentation accuracy and correlation of T1 and T2 measures using the various proposed methods for the validation set made of 8 patient scans.

endocardium in T1 maps. Fair results correspond to slight underestimation of the fibrosis due to over-segmentation of the endocardium. Poor results arise in cases of important artifacts (probably due to breathing motion that perturbed T1 reconstruction).

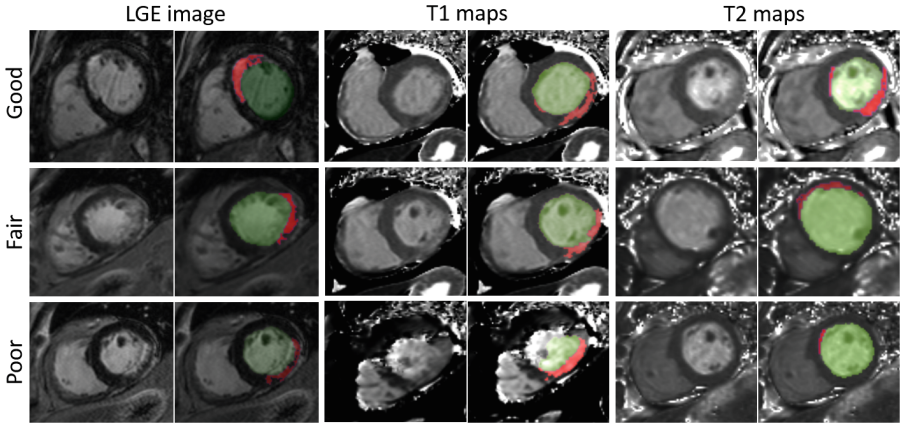

Fig. 4. Example endocardium (green) and scar, fibrosis or edema (red) segmentations obtained using the *Corrected Full Model* on validation data. Good, median and bad results are shown for LGE, T1 and T2 images. (Color figure online)

T2 Maps Segmentation. As shown Fig. 3 (right), the *Corrected 3 Layers* showed results similar to the *Corrected Full Model* in terms of dice scores for the endocardium (0.889 and 0.885) and correlation for Mean T2 values (0.937 and 0.977 respectively). Surprisingly, the edema size showed better correlation without the added correction (R = 0.919). Figure 4 (right) illustrates examples of endocardium and edema delineations showing that poor results were due to wrong segmentation of non-existing edema, while fair results, similarly to T1 and LGE, were due to over-segmentation of the endocardium.

4 Discussion and Conclusions

In this paper, we used deep neural network to automatically segment the myocardium contours in structural images, very accurately for T1 and T2 maps, and reasonably for LGE. Despite a fair dice score in the myocardium delineation of the LGE images, the scar size was well estimated with an error around 5% and a high correlation (R = 0.94) on our small validation set. These correlations and error levels are comparable to reported reproducibility errors obtained between two manual segmentations [3].

We evaluated the use of transfer learning to segment the structural maps since these maps share features with Cine data for which our pretrained model reach high performance. Results were extremely comparable to the full U-Net

model in terms of dice scores or correlations, apart from the estimation of the fibrosis size. It can therefore be a good alternative when working with restricted memory or computation power.

Better results were achieved after correction using our proposed CNN image selection model. High accuracies were reached, preventing most over-segmentation without deleting required slices. Cine MRI segmentation also suffers from this type of over-segmentation as the definition of the basal slice varies among observers. Recent active shape algorithms have been proposed to automatically define the basal slice [9] but once again deep learning methods, such as the one proposed here, showed more promising results. For a fully automatic cardiac analysis pipeline, we will extend the proposed slice selection neural network to our Cine data in combination with the current U-Net model that we are routinely using.

Future work will extend the current database and validate the robustness of the proposed pipelines. Similar concepts will also be applied to other structural images such as T2*, T2 black blood, post-contrast T1 maps or even ECV maps. Estimating the scar/fibrosis/edema zones directly from a deep learning network would also be appreciated once a reliable ground truth is given, or a consensus among experts has been reached.

After further validation, these models and the Cine model will be incorporated into our in-house cardiac analysis software. The amount of post-prediction corrections required will be measured in order to obtain confidence intervals for a fully automatic multi-scan cardiac MRI analysis pipeline in the near future.

References

1. Baumgartner, C.F., Koch, L.M., Pollefeys, M., Konukoglu, E.: An exploration of 2D and 3D deep learning techniques for cardiac MR image segmentation. In: Pop, M., et al. (eds.) STACOM 2017. LNCS, vol. 10663, pp. 111–119. Springer, Cham (2018). https://doi.org/10.1007/978-3-319-75541-0_12
2. Ciofolo, C., Fradkin, M., Mory, B., Hautvast, G., Breeuwer, M.: Automatic myocardium segmentation in late-enhancement MRI. In: ISBI 2008, pp. 225–228. IEEE (2008)
3. Flett, A.S., et al.: Evaluation of techniques for the quantification of myocardial scar of differing etiology using cardiac magnetic resonance. JACC Cardiovasc. Imaging. **4**(2), 150–156 (2011)
4. Isensee, F., Jaeger, P.F., Full, P.M., Wolf, I., Engelhardt, S., Maier-Hein, K.H.: Automatic cardiac disease assessment on cine-MRI via time-series segmentation and domain specific features. In: Pop, M., et al. (eds.) STACOM 2017. LNCS, vol. 10663, pp. 120–129. Springer, Cham (2018). https://doi.org/10.1007/978-3-319-75541-0_13
5. Liu, J., Li, W., Tian, Y.: Automatic thresholding of gray-level pictures using two-dimension Otsu method. In: International Conference on Circuits and Systems, pp. 325–327. IEEE (1991)
6. Khened, M., Alex, V., Krishnamurthi, G.: Densely connected fully convolutional network for short-axis cardiac cine MR image segmentation and heart diagnosis using random forest. In: Pop, M., et al. (eds.) STACOM 2017. LNCS, vol. 10663, pp. 140–151. Springer, Cham (2018). https://doi.org/10.1007/978-3-319-75541-0_15

7. Krizhevsky, A., Sutskever, I., Hinton, G.E.: ImageNet classification with deep convolutional neural networks. In: Advances in Neural Information Processing Systems, pp. 1097–1105 (2012)
8. McAlindon, E., Pufulete, M., Lawton, C., Angelini, G.D., Bucciarelli-Ducci, C.: Quantification of infarct size and myocardium at risk: evaluation of different techniques and its implications. EHJ-CVI **16**(7), 738–746 (2015)
9. Paknezhad, M., Marchesseau, S., Brown, M.S.: Automatic basal slice detection for cardiac analysis. J. Med. Imaging **3**(3) (2016). https://doi.org/10.1117/1.JMI.3.3.034004
10. Pop, M., Sermesant, M., Jodoin, P.-M., Lalande, A., Zhuang, X., Yang, G., Young, A., Bernard, O. (eds.): STACOM 2017. LNCS, vol. 10663. Springer, Cham (2018). https://doi.org/10.1007/978-3-319-75541-0
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
12. Wei, D., Sun, Y., Chai, P., Low, A., Ong, S.H.: Myocardial segmentation of late Gadolinium enhanced MR images by propagation of contours from cine MR images. In: Fichtinger, G., Martel, A., Peters, T. (eds.) MICCAI 2011. LNCS, vol. 6893, pp. 428–435. Springer, Heidelberg (2011). https://doi.org/10.1007/978-3-642-23626-6_53