

# Recurrent Fully Convolutional Neural Networks for Multi-slice MRI Cardiac Segmentation

Rudra P K Poudel, Pablo Lamata and Giovanni Montana

Department of Biomedical Engineering, King's College London, SE1 7EH, UK

**Abstract.** In cardiac magnetic resonance imaging, fully-automatic segmentation of the heart enables precise structural and functional measurements to be taken, e.g. from short-axis MR images of the left-ventricle. In this work we propose a recurrent fully-convolutional network (RFCN) that learns image representations from the full stack of 2D slices and has the ability to leverage inter-slice spatial dependencies through internal memory units. RFCN combines anatomical detection and segmentation into a single architecture that is trained end-to-end thus significantly reducing computational time, simplifying the segmentation pipeline, and potentially enabling real-time applications. We report on an investigation of RFCN using two datasets, including the publicly available MICCAI 2009 Challenge dataset. Comparisons have been carried out between fully convolutional networks and deep restricted Boltzmann machines, including a recurrent version that leverages inter-slice spatial correlation. Our studies suggest that RFCN produces state-of-the-art results and can substantially improve the delineation of contours near the apex of the heart.

**Keywords:** recurrent fully convolutional networks, recurrent restricted Boltzmann machine, left ventricle segmentation

## 1 Introduction

Cardiovascular disease is one of the major causes of death in the world. Physicians use imaging technologies such as magnetic resonance imaging (MRI) to estimate structural (e.g. volume) and functional (e.g. ejection fraction) cardiac parameters for both diagnosis and disease management. Fully-automated estimation of such parameters can facilitate early diagnosis of the disease and has the potential to remove the more mechanistic aspects of a radiologist's assessment. As such, lately there has been increasing interest in machine learning algorithms for fully automatic left-ventricle (LV) segmentation [1, 17, 8, 10, 12]. This is a challenging task due to the variability of LV shape across slices, cardiac phases, patients and scanning machines as well as weak boundaries of the LV due to the presence of blood flow, papillary muscles and trabeculations. A review of LV segmentation methods in short-axis cardiac MR images can be found in [20].

The main image analysis approaches to LV segmentation can be grouped into three broad categories: active contour models, machine learning models, and

hybrid versions that combine elements of the two approaches. Active contour models with either explicit [13] or implicit [15] contour representations minimize an energy function composed of internal and external constraints. The internal constraints represent continuity and smoothness of the contour and external constraints represent appearance and shape of the target object. However, designing appropriate energy functions that can handle all sources of variability is challenging. Also, the quality of the segmentations produced by these methods typically depends on the region-of-interest (ROI) used to initialise the algorithms. Machine learning approaches have been proposed to circumvent some of these issues [9, 18, 17, 1] at the expense of collecting large training datasets with a sufficient number of examples. Investigating hybrid methods that combine some elements of both approaches is an active research area [4].

Current state-of-the-art LV segmentation approaches rely on deep artificial neural networks [18, 17, 1]. Typically, these solutions consist of three distinct stages carried out sequentially. Initially, the LV is localised within each two-dimensional slice; then the LV is segmented, and finally the segmentation is further refined to improve its quality. For instance, a pipeline consisting of Deep Belief Networks (DBNs) for both localisation and segmentation, followed by a level-set methodology, has shown to generate high-quality segmentations [17]. In more recent work, a different pipeline has been proposed that consists of convolutional neural networks for initial LV detection, followed by a segmentation step deploying stacked autoencoders, and a fine-tuning strategy also based on level-sets methodology [1]. The latter approach has been proven to produce state-of-the-art results on the MICCAI 2009 LV segmentation challenge [21]. Both approaches share a number of common features. First, the segmentation is carried out using two-dimensional patches that are independently extracted from each MRI slice. Second, they use a separate architecture for the two tasks, localization and segmentation. Third, different neural network architectures are trained for cardiac MR slices containing the base and apex of the heart, due to the observed heterogeneity in local shape variability.

In this work we investigate a neural network architecture, trained end-to-end, that learns to detect and segment the LV jointly from the the entire stack of short-axis images rather than operating on individual slices. Recently, fully convolutional networks (FCN) have been proposed for the segmentation of 2D images [16]. They take arbitrarily sized input images, and use feature pooling coupled with an upsampling step to produce same size outputs delivering the segmentation. Compared to more traditional sliding-window approaches, FCNs are more efficient. They have received increasing interest lately as they unify object localization and segmentation in a single process by extracting both global and local context effectively [16, 22]. Applications of FCNs to medical imaging segmentation problems have also started to appear, for instance for the identification of neuronal structures in electron microscopic recordings [22]. In independent work, Valipour et al. [25] have recently adapted recurrent fully convolutional networks for video segmentation.

Here we propose an extension of FCNs, called Recurrent Fully-Convolutional Networks (RFCN), to directly address the segmentation problem in multi-slice MR images. We are motivated by the desire to exploit the spatial dependencies that are observed across adjacent slices and learn image features that capture the global anatomical structure of the heart from the full image stack. We investigate whether exploiting this information is beneficial for accurate anatomical segmentation, especially for cardiac regions with weak boundaries, e.g. poor structural contrast due to the presence of blood flow, papillary muscles and trabeculations.

## 2 Datasets

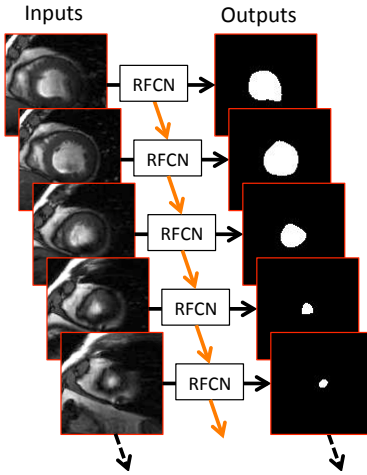
Our experiments are based on two independent datasets consisting of short-axis cardiac MR images for which the endocardium has been manually segmented by expert radiologists in each axial slice. Further details are provided below.

### 2.1 MICCAI dataset

The MICCAI 2009 LV Segmentation Challenge [21] dataset was made publicly available by the Sunnybrook Health Sciences Center (Canada) and has been extensively used to compare a number of LV segmentation algorithms [12, 10, 8, 18, 17, 1]. It consists of 45 CINE MRI images from a number of different pathologies. The individual exams have been pre-grouped into training, validation and online testing subsets. Each subset contains 15 cases of which there are 4 heart failure with infarction (HF-I), 4 heart failure without infarction (HF), 4 LV hypertrophy (HYP) and 3 healthy subjects. However, the clinical information has not been used by any of the algorithms discussed here and in our experiments. All the images were obtained during breath-hold sessions lasting 10 – 15 seconds with a temporal resolution of 20 cardiac phases over the heart cycle. A typical phase, end diastole (ED) or end systole (ES), contains 6 – 12 short-axis slices obtained from the base to apex. In all the images, the slice thickness is 8mm, the inter-slice gap is 8mm, the field of view is 320mm  $\times$  320mm and the pixel size is 256  $\times$  256. In all 45 samples, LV endocardial contours were drawn by an experienced cardiologist by taking 2D slices at both the end-systolic and end-diastolic phases, and then independently confirmed by a second reader. The manual segmentations were used as ground truth for the evaluation of the proposed models. Each set consists of 30 sequences (15 samples for each one of the two cardiac phases) with an average sequence length 8.94 slices.

### 2.2 PRETERM dataset

A second and larger dataset was used for an independent evaluation of all the cardiac segmentation algorithms. The dataset consists of 234 subjects used to study perinatal factors modifying the left ventricular parameter [14]. All the individuals are between 20 to 39 years of age. Of these, 102 were followed prospectively since preterm birth, and are characterised by an average gestational age of

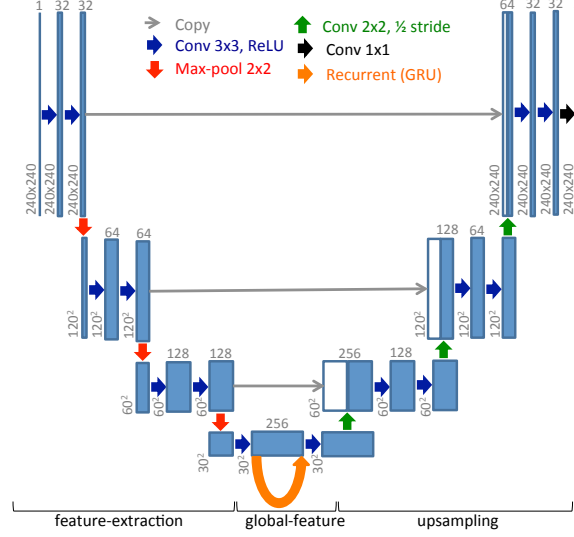


**Fig. 1.** A stack of short-axis cardiac MR slices (left) with corresponding left-ventricular binary masks (right) for a cardiac phase. The proposed RFCN leverages the spatial correlations that can be observed moving from the base of the heart to the apex.

$30.3 \pm 2.5$  weeks and a birth weight of  $1.3 \pm 0.3$  kg. The remaining 132 subjects were born at term to uncomplicated pregnancies. Short-axis CINE MRI stacks were acquired with a 1.5-T Siemens Sonata scanner. All images have a 7mm slice thickness and 3mm inter-slice gap, the in-plane resolution is  $1.43 \pm 0.29$  mm (min. 0.57, max 2.17). All cardiovascular MRI was prospectively ECG gated and acquired during end-expiration breath holding. LV slices and endocardial masks were resampled into a homogeneous in-plane resolution of  $2\text{mm}$ , which yield slice pixel size of  $212 \times 212$ . Left ventricular short-axis endocardial borders were manually contoured by an expert reader at ES and ED using Siemens analytic software (Argus, Siemens Medical Solutions, Germany). The dataset was randomly divided into training, validation and testing sets of sizes 194, 20 and 20, respectively.

### 3 Recurrent fully-convolutional networks

The proposed recurrent fully-convolutional network (RFCN) is an extension of the architecture originally introduced in [16] for predicting pixel-wise, dense outputs from arbitrarily-sized inputs. The main idea underlying FCNs is to extend a contracting path, in which a sequence of pooling operators progressively reduces the size of the network, by adding successive layers where pooling operators are replaced by upsampling operators. In this respect, our architecture is similar to U-net [22] where the expanding path is characterised by a large number of feature channels allowing the network to propagate context information to higher resolution layers.



**Fig. 2.** Overview of the RFCN architecture. Blue boxes represent feature maps and white boxes represent copied feature maps. The number of feature maps and their dimensions are displayed above each box and on the left-side, respectively. Arrows represent network operations: gray arrows indicate copy operations, blue arrows indicate convolutional operations, red arrows indicate max-pooling operations, green arrows indicate convolutional operations with 1/2 stride, black arrow indicates  $1 \times 1$  convolutional operation and orange arrow indicates a recurrent connection to handle inter-slice dependencies learned through GRU.

Our purpose is to model the full stack of short-axis images extracted from cardiac MRI and improve the segmentation of the left ventricle in each slice by leveraging inter-slice spatial dependencies. The input is the entire sequence of  $S$  slices obtained at a particular cardiac phase (ED or ES) and the output is the sequence of corresponding (manually produced) left-ventricular masks. Each input and output image is assumed to have equal size. A schematic illustration is given in Figure 1. As can be seen there, slices around the base of the heart (at the top) cover larger LV regions and show relatively clear boundaries whereas slices around the apex (at the bottom) cover smaller LV regions and present more blurred boundaries. Learning the typical shape deformations that are observed as we move from the base towards the apex is expected to improve the overall quality of the segmentation in challenging regions around the apex.

Three main building blocks characterise the proposed RFCN as illustrated in Figure 2: a feature-extraction (contracting) path, a global-feature component and an upsampling (expanding) path. The feature-extraction component, which is independently applied to each image in the stack, deploys successive convolution and max-pooling operations to learn higher level features and remove local

redundancy. In our architecture, this component consists of a repeated block of two  $(3 \times 3)$  convolutional layers (with stride of 1) followed by a rectified linear unit (ReLU) and a  $(2 \times 2)$  max pooling layer (with stride of 2). We doubled the number of feature channels  $c$  after each max pooling layer to maintain enough context, i.e each block takes an input of size  $(c \times h \times w)$  and generates output feature maps of size  $(2c \times h/2 \times w/2)$ .

At the end of this contracting path the network has extracted the most compressed features carrying global context. The global feature component starts here with a  $(3 \times 3)$  convolutional layer (with stride of 1) followed by a ReLU. We denote  $\mathbf{e}_s$  the output of this layer where  $s$  indicates the slice index, i.e.  $s \in \{1, \dots, S\}$ . This output consists of  $(256 \times 30 \times 30)$  feature maps. In an attempt to extract global features that capture the spatial changes observed when moving from the base to the apex of the heart, we introduce a recurrent mechanism mapping  $\mathbf{e}_s$  into a new set of features,  $\mathbf{h}_s = \phi(\mathbf{h}_{s-1}, \mathbf{e}_s)$ , where  $\phi(\cdot)$  is a non-linear function, and the size of  $\mathbf{h}_s$  is the same as the size of  $\mathbf{e}_s$ . Another  $(3 \times 3)$  convolutional layer (with stride of 1) followed by a ReLU is then applied to complete the global-feature extraction block. Given that training recurrent architectures is particularly difficult due to the well-documented vanishing gradient problem, several options were considered for the implementation of recurrent function  $\phi$ , including a Long Short-Term Memory (LSTM) [7] and Gated Recurrent Units (GRUs) module [2]. GRUs in particular have been shown to achieve a performance comparable to LSTM on a number of tasks involving sequential data whilst requiring fewer parameters and less memory [3]. Here we have chosen to use a convolutional variant of GRU so that the local correlation of the input images is preserved whilst achieving a notable reduction in the number of parameters compared to its non-convolutional counter part.

For every slice, the dense feature maps that have been learned by the convolutional GRU module are then upsampled to compensate for the input size reduction caused by the max-pooling operations. The upsampled features are concatenated with a high resolution parallel layer aligned to the feature-extraction component, similarly to the U-net architecture [22]. Our upsampling component consists of a repeated block of a convolutional layer (with a fractional stride of  $1/2$ ), a feature map concatenation module and two  $3 \times 3$  convolutional layers (with stride of 1) followed by ReLU. The feature map concatenation module combines the outputs of the upsample layer and parallel feature-extraction block. Each block of the upsampling component takes a three-dimensional input  $c \times h \times w$  and outputs a  $c/2 \times 2h \times 2w$  dimensional tensor. A convolutional operation with fractional stride is employed to compensate for the reduction in input size due to the max pooling operation. Even though the upsampling procedure smooths out the boundaries of the object to be segmented, the concatenation of up-sampled feature maps with high-resolution feature maps helps mitigate this smoothing problem by providing better local and boundary information. The final segmentation is obtained by using a  $1 \times 1$  convolutional layer, which maps the output of the upsampling component onto the two classes, i.e. LV and

background. The probability for each class is given by a softmax function across all pixel locations.

## 4 Other architectures and model training

Recently, deep belief networks (DBNs) have been proposed for automatic LV detection and segmentation using short-axis MR images [18, 17]. A DBN was first used to detect the region of interest containing the LV. Anatomical segmentation was then carried out using distance-regularised level sets, which were modified to leverage prior shape information inferred by a separate DBN. In these models, as in FCNs, each slice in the short-axis stack is segmented independently of all the others. The main building block of a DBN model is a restricted Boltzmann machine (RBM), typically trained using the contrast divergence algorithm [5]. In some of our experiments, we have assessed the performance of DBNs for LV segmentation comparably to FCNs and the proposed RFCNs.

In order to further investigate whether modelling the dependence across slices typically yields improved performance, and motivated by the existing body of work on DBNs, we have also assessed the performance of a recurrent version of restricted Boltzmann machines (RRBM), originally proposed to learn human body motion [23], but never used for LV segmentation. RRBMs are stacked together to form what we call a recurrent deep belief network (RDBN). Similarly to the proposed RFCN, RDBN takes the entire sequence of short-axis slices as input and leverages the spatial correlations through additional bias units. For further information, we refer the reader to the original work [23].

The two convolutional architectures, FCN and RFCN, were trained by minimizing the cross-entropy objective function. FCN was trained using a stochastic gradient descent algorithm with momentum whereas RFCN was trained using a stochastic gradient descent algorithm with RMSProp [24]. Back-propagation was used to compute the gradient of the cross-entropy objective function with respect to all parameters of the model, including the GRU component in the case of RFCN. We also learned  $\mathbf{h}_0$  as required by the first slice of the sequence. In each block, batch normalization [11] was added after each convolutional layer, i.e. just before the max-pooling and upsampling layers. All reported results refer to the best out of 5 experiments in which the models were initialised with random parameters. RFCN was initialised using weights obtained from FCN, which reduces the training time and provided the good initial weights. Both the DBN and RDBN architectures were trained using the contrast divergence algorithm [5]. Dropout [6] was found to improve their overall performance. For all these models, best results were achieved using a learning rate of 0.01 with constant decay of 3% after each epoch, a momentum of 0.9 and weight decay of 0.00005. At the training phase, both the MICCAI and PRETERM datasets were augmented by generating additional artificial training images to prevent model overfitting. During training, we performed translation ( $\pm 16$  pixels) and rotation ( $\pm 40^\circ$ ) data augmentation, which was found to yield better performance.

## 5 Experimental results

This section presents an empirical evaluation of several LV endocardium segmentation algorithms using three performance metrics: good contours (GC) [1], Dice index, and average perpendicular distance (APD) between manually drawn and predicted contours [21]. In order to make our experimental results comparable with published studies on the MICCAI dataset, all models were validated using the online set, and we report results obtained on the validation set. Table 1 summarises the experimental results. On the MICCAI dataset, the DBN-based results presented in [17] include a Dice index of 0.88, a GC of 95.71% and an APD of 2.34 mm whereas the pipeline described in [1] results in a Dice index of 0.90, a GC of 90% and an APD of 2.84 mm (before further post-processing). A comparable Dice index is obtained by both FCN and RFCN, which yield higher GC and smaller ADP. Here RFCN outperforms FCN with a substantially improved ADP of 2.05 mm.

The PRETERM dataset was modelled using the the same architectures, without further customisation. The results of this application are also summarised in Table 1. For this dataset, we compared the performance of four different architectures: FCN, RFCN, DBN and RDBN. The latter two models were given as input a region of interest containing the LV thus conferring them an advantage compared to FCN and RFCN. On this dataset we were not able to test the recently proposed pipeline described in [1], which relies on multiple stages. As in the MICCAI dataset, the fully convolutional architectures have achieved superior performance. RFCN has outperformed all other architectures in terms of Dice index and APD, which was found to be as small as 1.56 mm. In comparison, DBN with known LV location yields an APD of 2.05 mm. RDBN yields higher GC and lower APD compared to DBN thus providing additional evidence that performance gains can be obtained by modelling intra-slice dependencies.

In order to shed insights into the regional improvements introduced by RFCN, the Dice index was computed separately for different local regions of the LV, and the results are summarised in Table 2. Here, Base-1, Base-2 and Base-3 indicates that 1, 2 and 3 slices were taken starting from the base of the heart and moving towards the middle, and analogously for the apex. All the remaining slices contributed towards the Central class. In all cases, the Dice index is calculated using all the samples at once to reflect overall pixel accuracy. In both datasets, RFCN outperforms FCN around the central slices and around the apex, as expected. However, in the MICCAI dataset, FCN yields better performance around the base of the heart. On the PRETERM dataset, both DBN and RDBN gave the worst performance, compared to FCN and RFCN, despite using focused region of interest instead of full-sized images. Here again it can be observed that RDBN improves upon DBN across all cardiac locations.

## 6 Conclusions

In this paper we have investigated whether a single neural network architecture, trained end-to-end, can deliver a fully-automated and accurate segmentation of



**Table 1.** Performance assessment on MICCAI and PRETERM datasets. For the MICCAI dataset, for completeness we also report on published results after a post-processing stage based on level-sets. The DBN and RDBN models only performed endocardium segmentation, not detection, i.e. they were applied to focused regions of interest centered around the left-ventricle. All other architectures performed LV detection and endocardium segmentation from full short-axis slices.

Dataset	Model	GC	Dice	APD
MICCAI with level-sets	[17]	95.91 (5.28)	0.880 (0.03)	2.34 (0.46)
	[1]	97.80 (4.70)	0.94 (0.02)	1.70 (0.37)
	[19]	95.91 (5.28)	0.880 (0.03)	2.34 (0.46)
MICCAI without level-sets	[17]	<b>95.71 (6.96)</b>	0.880 (0.03)	2.34 (0.45)
	[1]	90.00 (10.00)	0.900 (0.10)	2.84 (0.29)
	[19]	90.29 (12.73)	0.880 (0.03)	2.42 (0.36)
	FCN	94.78 (06.27)	<b>0.902 (0.04)</b>	2.14 (0.38)
	RFCN	95.34 (07.20)	0.900 (0.04)	<b>2.05 (0.29)</b>
PRETERM	DBN	92.01 (8.36)	0.913 (0.02)	2.05 (0.38)
	RDBN	97.50 (6.77)	0.909 (0.02)	1.94 (0.23)
	FCN	<b>97.59 (4.82)</b>	0.916 (0.03)	1.80 (0.41)
	RFCN	95.37 (5.69)	<b>0.935 (0.03)</b>	<b>1.56 (0.31)</b>

the left ventricle using a stack of MR short-axis images. The proposed architecture, RFCN, learns image features that are important for the localisation of the LV in a sequential manner, going from the base to the apex of the heart, through a recurrent modification of fully convolutional networks.

Experimental findings obtained from two independent applications demonstrate that propagating information from adjacent slices can help extract improved context information with positive effect on the resulting segmentation quality. The hypothetical value of the large inter-slice correlation has been further tested by introducing a recurrent version of deep belief networks, and verified with our results showing that RDBNs generally outperform DBNs on the segmentation task, assuming the LV has already been localised. As expected, notable improvements can be seen in the delineation of cardiac contours around the apex, which are notoriously more difficult to identify.

One surprising finding was to note that performance of RFCN in apical slices was better for MICCAI than for PRETERM cohort (0.85 vs. 0.76 Dice index in the most apical slice, see Table 2), when one could expect the opposite: a regular and homogeneous cohort, PRETERM, should lead to a better performance when leveraging the inter-slice spatial dependence. This aspect will warrant further investigations.

Compared to other models, RFCN has the advantage of carrying out both LV detection and segmentation in a single architecture with clear computational benefits and the potential for real-time application. In future work, we are

**Table 2.** Breakdown of the Dice index by LV regions on both MICCAI and PRETERM datasets. Base-X and Apex-X represent number of slice(s) included starting from the base and apex of the left ventricle, respectively. The index is calculated using all slices from the all samples at once hence measuring overall pixels accuracy. DBN and RDBN performed the segmentation task using a pre-defined region of interest containing the LV region.

Dataset	Model	Base-1	Base-2	Base-3	Central	Apex-3	Apex-2	Apex-1
MICCAI	FCN	<b>0.9313</b>	<b>0.9314</b>	<b>0.9342</b>	0.9367	0.8751	0.8441	0.7581
	RFCN	0.9040	0.9178	0.9268	<b>0.9433</b>	<b>0.9112</b>	<b>0.8917</b>	<b>0.8468</b>
PRETERM	DBN	0.9285	0.9374	0.9413	0.9385	0.8465	0.7809	0.6139
	RDBN	0.9319	0.9379	0.9420	0.9433	0.8856	0.8409	0.7542
	FCN	0.9486	0.9536	0.9559	0.9610	0.9051	0.8686	0.7468
	RFCN	<b>0.9576</b>	<b>0.9621</b>	<b>0.9631</b>	<b>0.9625</b>	<b>0.9178</b>	<b>0.8800</b>	<b>0.7571</b>

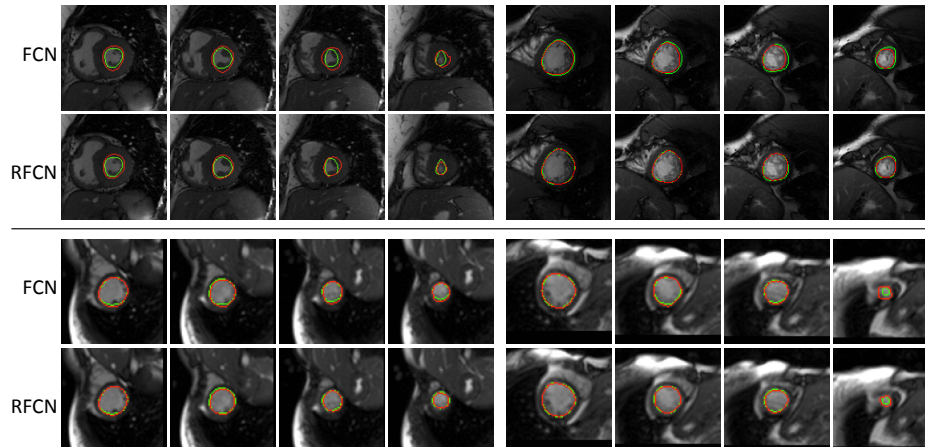
planning to investigate alternative operations that can capture inter-slice correlations, such as 3D convolutions, and further extend RFCN by incorporating a bi-directional mechanism for the inclusion of an inverse path (from the apex to the base of the heart) as well as a temporal extension to handle all cardiac phases at once.

## Acknowledgements

The authors would like to thank Paul Leeson and Adam Lewandowski from Oxford University for their assistance with the PRETERM dataset.

## References

1. Avendi, M.R., Kheradvar, A., Jafarkhani, H.: A combined deep-learning and deformable-model approach to fully automatic segmentation of the left ventricle in cardiac MRI. *Medical Image Analysis* 30, 108–119 (May 2016)
2. Cho, K., van Merriënboer, B., Gulcehre, C., Bahdanau, D., Bougares, F., Schwenk, H., Bengio, Y.: Learning Phrase Representations using RNN Encoder-Decoder for Statistical Machine Translation. *arXiv:1406.1078* (2014)
3. Chung, J., Gulcehre, C., Cho, K., Bengio, Y.: Empirical Evaluation of Gated Recurrent Neural Networks on Sequence Modeling. *arXiv:1412.3555* (2014)
4. Georgescu, B., Zhou, X.S., Comaniciu, D., Gupta, A.: Database-guided segmentation of anatomical structures with complex appearance. In: *CVPR*. vol. 2, pp. 429–436 (2005)
5. Hinton, G.E., Salakhutdinov, R.R.: Reducing the Dimensionality of Data with Neural Networks. *Science* 313(5786), 504–507 (2006)
6. Hinton, G.E., Srivastava, N., Krizhevsky, A., Sutskever, I., Salakhutdinov, R.R.: Improving neural networks by preventing co-adaptation of feature detectors. *arXiv:1207.0580 [cs]* (2012)



**Fig. 3.** Example of segmented left ventricle using RFCN and FCN architectures from MICCAI dataset (top two rows) [21] and PRE/TERM dataset (bottom two rows). Green contours represent the ground truth and red contours are the predicted contours. RFCN is often able to better delineate the left-ventricle contours with weaker boundaries compared to FCN.

7. Hochreiter, S., Schmidhuber, J.: Long Short-Term Memory. *Neural Computation* 9(8), 1735–1780 (1997)
8. Hu, H., Liu, H., Gao, Z., Huang, L.: Hybrid segmentation of left ventricle in cardiac MRI using gaussian-mixture model and region restricted dynamic programming. *Magnetic Resonance Imaging* 31(4), 575–584 (2013)
9. Huang, R., Pavlovic, V., Metaxas, D.N.: A graphical model framework for coupling MRFs and deformable models. vol. 2, pp. 739–746 (Jun 2004)
10. Huang, S., Liu, J., Lee, L.C., Venkatesh, S.K., Teo, L.L.S., Au, C., Nowinski, W.L.: An image-based comprehensive approach for automatic segmentation of left ventricle from cardiac short axis cine MR images. *Journal of Digital Imaging* 24(4), 598–608 (2011)
11. Ioffe, S., Szegedy, C.: Batch Normalization: Accelerating Deep Network Training by Reducing Internal Covariate Shift. *arXiv:1502.03167 [cs]* (Feb 2015)
12. Jolly, M.: Fully Automatic Left Ventricle Segmentation in Cardiac Cine MR Images Using Registration and Minimum Surfaces. *MIDAS Journal* 49 (2009)
13. Kass, M., Witkin, A., Terzopoulos, D.: Snakes: Active contour models. *International Journal of Computer Vision* 1(4), 321–331 (1988)
14. Lewandowski, A.J., Augustine, D., Lamata, P., Davis, E.F., Lazdam, M., Francis, J., McCormick, K., Wilkinson, A.R., Singhal, A., Lucas, A., Smith, N.P., Neubauer, S., Leeson, P.: Preterm heart in adult life: cardiovascular magnetic resonance reveals distinct differences in left ventricular mass, geometry, and function. *Circulation* 127(2), 197–206 (2013)
15. Li, C., Xu, C., Gui, C., Fox, M.: Distance Regularized Level Set Evolution and Its Application to Image Segmentation. *IEEE Transactions on Image Processing* 19(12), 3243–3254 (2010)

16. Long, J., Shelhamer, E., Darrell, T.: Fully Convolutional Networks for Semantic Segmentation. In: CVPR (2015)
17. Ngo, T.A., Carneiro, G.: Fully Automated Non-rigid Segmentation with Distance Regularized Level Set Evolution Initialized and Constrained by Deep-Structured Inference. In: CVPR. pp. 3118–3125 (2014)
18. Ngo, T.A., Carneiro, G.: Left ventricle segmentation from cardiac MRI combining level set methods with deep belief networks. In: ICIP. pp. 695–699 (2013)
19. Ngo, T.A., Lu, Z., Carneiro, G.: Combining deep learning and level set for the automated segmentation of the left ventricle of the heart from cardiac cine magnetic resonance. *Medical Image Analysis* 35, 159–171 (2017)
20. Petitjean, C., Dacher, J.N.: A review of segmentation methods in short axis cardiac MR images. *Medical Image Analysis* 15(2), 169–184 (2011)
21. Radau, P., Lu, Y., Connelly, K., Paul, G., Dick, A.J., Wright, G.A.: Evaluation Framework for Algorithms Segmenting Short Axis Cardiac MRI. *The MIDAS Journal Cardiac MR Left Ventricle Segmentation Challenge* (2009)
22. Ronneberger, O., Fischer, P., Brox, T.: U-Net: Convolutional Networks for Biomedical Image Segmentation. In: MICCAI (2015)
23. Sutskever, I., Hinton, G.E., Taylor, G.W.: The Recurrent Temporal Restricted Boltzmann Machine. In: NIPS, pp. 1601–1608 (2009)
24. Tieleman, T., Hinton, G.: Lecture 6.5-rmsprop: Divide the gradient by a running average of its recent magnitude. *COURSERA: Neural Networks for Machine Learning* 4 (2012)
25. Valipour, S., Siam, M., Jagersand, M., Ray, N.: Recurrent Fully Convolutional Networks for Video Segmentation. *arXiv:1606.00487 [cs]* (2016)