Cardiac Left Ventricle Segmentation using Convolutional Neural Network Regression

Li Kuo Tan[1], Yih Miin Liew[2], Einly Lim[2]
[1]Faculty of Medicine, [2]Faculty of Engineering
University of Malaya
Kuala Lumpur, Malaysia

Robert A. McLaughlin
ARC Centre of Excellence for Nanoscale Biophotonics
University of Adelaide
Adelaide, Australia

***Abstract*—Cardiac MRI is important for the diagnosis and assessment of various cardiovascular diseases. Automated segmentation of the left ventricular (LV) endocardium at end-diastole (ED) and end-systole (ES) enables automated quantification of various clinical parameters including ejection fraction. Neural networks have been used for general image segmentation, usually via per-pixel categorization e.g. "foreground" and "background". In this paper we propose that the generally circular LV endocardium can be parameterized and the endocardial contour determined via neural network *regression*. We designed two convolutional neural networks (CNN), one for localization of the LV, and the other for determining the endocardial radius. We trained the networks against 100 datasets from the Medical Image Computing and Computer Assisted Intervention (MICCAI) 2011 challenge, and tested the networks against 45 datasets from the MICCAI 2009 challenge. The networks achieved 0.88 average Dice metric, 2.30 mm average perpendicular distance, and 97.9% good contours, the latter being the highest published result to date. These results demonstrate that CNN regression is a viable and highly promising method for automated LV endocardial segmentation at ED and ES phases, and is capable of generalizing learning between highly distinct training and testing data sets.**

Keywords—cardiac MRI; left ventricle; segmentation; neural network regression

I. Introduction

Cardiac MRI is the current gold standard for the diagnosis and assessment of various cardiovascular diseases [1]. A standard cardiac MRI study usually begins with the assessment of left ventricular (LV) structure and function from a stack of steady state free precession (SSFP) short axis cine images captured from the mitral valve plane through to the apex. These images are non-isotropic, with in-plane resolutions of about 1.5 mm/pixel and slice thickness of 10 mm. The scans are gated to cover a complete cardiac cycle, typically at ≈50 ms temporal resolution which results in ≈20 cardiac phases [2].

Key LV parameters for the diagnosis of cardiac diseases include ejection fraction and cardiac output. Quantification requires the delineation of myocardial boundaries; most crucially the endocardial wall at end-diastole (ED) and end-systole (ES), whereby the heart is at rest and at its maximum contraction respectively. Manual delineation of the boundaries is time consuming and inconsistent among clinicians. Automated LV segmentation to facilitate delineation is therefore desirable.

Petitjean and Dacher have performed a comprehensive review of automated and semi-automated LV segmentation algorithms [3]. In brief, the techniques used included pixel-based methods (e.g. thresholding), pixel classification (e.g. clustering), deformable models (e.g. active contours), shape priors (e.g. PCA), and probabilistic atlases.

Neural networks for classification of data have seen development dating back to the 1960s with a history of alternate waxing and waning popularity. In recent years the popularity of neural network techniques has dramatically increased, largely due to its dominating performance in some key image processing challenges [4]. The enabling factors prompting this resurgence in popularity include the introduction of convolutional neural networks (CNNs), the availability of large quantities of training data, as well as the advances in computational capability, particularly graphical processing units (GPUs).

To date, CNN architectures are best known for whole image categorization, e.g. labelling an entire image as containing a "flower" or "car". However, CNNs can also be used for image segmentation. The standard approach is via per-pixel categorization. For example, if a 100×100 image is to be segmented into three labels: *blood pool*, *myocardium*, and *background*, then the output of the CNN is a 100×100×3 matrix; a 3-element vector representing the individual label's probability for each pixel.

Avendi et al. presented the most recent effort in applying neural network per-pixel categorization to ED/ES LV endocardium segmentation. They used three separate networks: one for initial LV localization, one for segmentation at basal and mid-LV slices, and one for segmentation at apical slices. A separate deformable model was applied post-network to moderate the contours [5].

In this work, we propose that the generally circular LV endocardium can be readily parameterized into a polar radial system. From there, LV boundary delineation was determined via neural network *regression* of the radial distances as opposed to the standard classification of each pixel. This method reduces the dimensionality of the problem, enabling a simpler network design while attaining good results without the need for complicated post-processing.

This research is supported by University of Malaya Research Grant (RP028A/B/C-14HTM). Author RAM is funded by the National Health and Medical Research Council, Australia, the Australian Research Council, and a South Australian Premier's Research and Industry Fund Fellowship.

II. Materials and Methods

A. Data

Multi-layer CNNs require large quantities of labelled data to be adequately trained. They can require roughly 10,000 to 1,000,000 images depending on the variation within the data. In this paper we focus on the segmentation of the LV endocardial wall at the ED and ES cardiac phases, captured via SSFP MRI. Contrast between the higher intensity blood pool and lower intensity myocardium is relatively strong and consistent, which helps in reducing the data required for training.

Medical Image Computing and Computer Assisted Intervention (MICCAI) is a large medical image processing-focused conference. It often hosts public image processing challenges as part of the conference activities. In 2009, one of the MICCAI challenges was for LV segmentation. In this challenge, 45 datasets were provided with manually drawn gold standard endocardial contours at the ED and ES phases. This matched our requirements, but with only around 900 labelled images, we felt it would be insufficient even if used completely as a training set.

In 2011, MICCAI held another similar challenge, this time for full 4D LV segmentation. A hundred labelled datasets were provided, with gold standard endo- and epicardial contours at all cardiac phases. This represented a much larger training set of around 20,000 images. The cardiac images of both the 2009 and 2011 challenges are of a similar nature; we thus were able to utilize the MICCAI 2011 data to train our CNN, and we then tested it against the MICCAI 2009 data.

The MICCAI 2009 images were obtained with a 1.5T GE Signa MRI and the following imaging parameters: image matrix 256×256, 1.25 mm/pixel in-plane resolution, slice spacing 8 mm, and 20 cardiac phases [6]. The MICCAI 2011 images were obtained from multiple scanners, with varying parameters: image matrix 156×192 to 512×512, 0.7 to 2.1 mm/pixel in-plane resolution, slice spacing 7 to 10 mm, and 18 to 35 cardiac phases [7].

B. Network Architecture

Our network operates on 2D images. To segment the endocardial wall, we performed a two-step process: step one localizes the (x, y) position of the LV centroid within the slice, and step two determines the endocardial contour in the form of 96 points evenly spaced radially around the LV centroid. This required two separate CNNs: the first for LV center point localization (referred to as the CPL network), and the second for delineation of the endocardial border (referred to as the EB network).

The SSFP image acquisition is gated, and the resulting image sequence covers a single cardiac cycle. We thus included temporal data by calculating a 1D Fast Fourier Transform (FFT) across the temporal dimension, and extracting the magnitude image of the 1st harmonic, I_F. This FFT image suppresses static regions, and was stacked together with the corresponding SSFP intensity image frame, I_M, to form the input to the CPL and EB networks. To elaborate, for a particular LV slice, there are two different intensity images, $I_{M(ED)}$ and $I_{M(ES)}$ at the ED and ES phases respectively, but only one I_F FFT image; I_F is used as the input twice, once for each phase.

To recap, for a single, stacked $I_M + I_F$ image input, the final output is a single (x, y) position indicating the LV centroid (from the CPL network), and 96 values representing radial distances from the centroid (from the EB network). We did not determine all 96 values simultaneously. Rather, the image was first remapped into a polar coordinate system centered at the LV centroid, and 96 overlapping segments were fed into the CPL network separately as detailed in the following section.

C. Data Preprocessing and Augmentation

1) LV Centroid – CPL Network

The I_M and I_F images were first resampled to 2 mm/pixel to provide consistency across data from different scanners, and cropped to 84×84 pixels centered at the image midpoint. The cropped images were normalized to a mean of 0 and standard deviation (SD) of 1. The resulting images were fed into the CPL network, which outputs the estimated LV center point, C_{LV}.

During training, we artificially augmented the data by randomly shifting the crop window up to ±35mm (simulating translation), scaling the images up to ±15%, rotating the images up to ±180°, and distorting the image mean and SD by up to ±0.15.

2) Endocardial Border – EB Network

For detection of the LV boundary, the source I_M and I_F images were resampled to 1 mm/pixel, then cropped and remapped to a 96 angular sector polar space centered around C_{LV}. Specifically, a radial coordinate map was interpolated from the source images to form polar images of 56×96 pixels (radius × angular sector) (Fig. 1). The cropped and remapped images were normalized to a mean of 0 and standard deviation (SD) of 1. The resulting images were fed into the EB network in 96 wrapped and overlapping 56×64 sections. This means the EB network is run 96 times to output 96 radius values for a single image, analogous to the sliding window operation in standard convolution.

During training, we artificially augmented the data by randomly shifting the LV center point by ±50% of the endocardial radius, scaling the images up to ±15%, and distorting the image mean and SD by up to ±0.15.

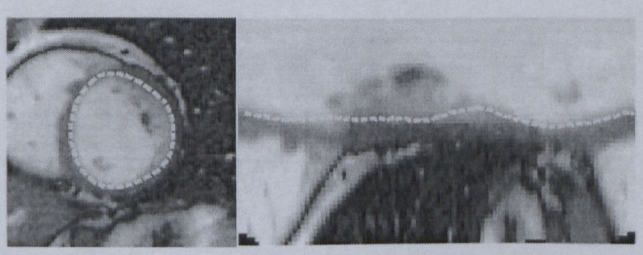

Fig. 1. (Left) Example LV at ED phase. (Right) polar remapped image: vertical axis is endocardial radius (blood pool at top), and horizontal axis is angular sector. Yellow dotted lines are the endocardial boundaries.

D. Network Design and Training

The network design for the CPL and EB networks are summarized in Table I and Table II respectively.

For the EB network, we utilized a two-resolution fine and coarse parallel design in the CNN layers. One sub-section of the network worked on an overall, coarse view of the image, while the other sub-section worked on a narrow, fine view of the image. To do this, the input polar image of 56×64 pixels in size were cropped to a single central 56×3 strip for the fine sub-network, whereas every other pixel along the angular dimension was omitted to form a 56×32 down-sampled image for the coarse sub-network.

Maxout activations with 3 maxout units are used in all cases except the output layers. Dropout with 0.33 dropout probability was used for regularization in the fully connected layers. Valid padding was used for all convolutional layers in the CPL network. We use the Adam stochastic optimization algorithm to minimize the mean-squared error loss function. The mini-batch size was 24 and the initial learning rate was 0.001. Training was stopped when the cross-validation loss was manually observed to have plateaued.

The network design was implemented using the TensorFlow r0.8 machine learning framework (Google Inc., California, U.S.). The network was trained and executed on a NVIDIA GeForce GTX 980 4 GB GPU.

E. Post-processing

For each slice, the 96 endocardial radius output values were filtered to exclude erroneous results. We excluded points which differed by >2 mm from their neighbors, and replaced them via linear interpolation.

TABLE I. CPL NETWORK DESIGN. CN[1,2,3] ARE CONVOLUTIONAL LAYERS, FC[1,2,3] ARE FULLY CONNECTED LAYERS.

Layer	In	Weights	Pool	Out
CN1	84×84×2	5×5×72	2×2	40×40×24
CN2	40×40×24	5×5×72	2×2	18×18×24
CN3	18×18×24	5×5×72	2×2	7×7×24
FC1	7×7×24	7×7×768	–	1×1×256
FC2	256	256×768	–	256
FC3	256	256×768	–	256
Out	256	256×2	–	2

TABLE II. EB NETWORK DESIGN. CN[1,2,3] ARE CONVOLUTIONAL LAYERS, FC[1,2,3] ARE FULLY CONNECTED LAYERS. …-C IS THE COARSE SUB-NETWORK, WHEREAS…-F IS THE FINE SUB-NETWORK. THE RESULTS OF BOTH SUB-NETWORKS ARE CONCATENATED IN FC2.

Layer	In	Weights	Pool	Out
CN1-c	56×32×2	3×3×18	2×2	28×16×6
CN2-c	28×16×6	3×3×18	2×2	14×8×6
CN3-c	14×8×6	3×3×18	2×2	7×4×6
FC1-c	7×7×24	7×7×768	–	1×1×256
CN1-f	56×3×2	3×3×54	–	56×3×18
CN2-f	56×3×18	3×3×54	–	56×3×18
CN3-f	56×3×18	3×3×54	–	56×3×18
FC1-f	56×3×18	56×3×576	–	1×1×192
FC2	64+192	256×768	–	256
FC3	256	256×768	–	256
Out	256	256×1	–	1

F. Evaluation

The MICCAI 2009 dataset is divided by the challenge organizer into three groups: 15 for training, 15 for testing (validation), and 15 for the final contest (online). Since we exclusively trained our CNN against the MICCAI 2011 datasets, we thus utilized all 45 MICCAI 2009 datasets for final testing and evaluation.

Performance was assessed via the Dice metric and average perpendicular distance (APD) between the automatically delineated and gold standard manual contours. Additionally, a segmentation (per slice or per subject) was classified as "good" if the mean APD was less than 5 mm [6], corresponding to an offset of 4 pixels at 1.25 mm/pixel in-plane resolution.

III. RESULTS AND DISCUSSION

Fig. 2 and Fig. 3 show results from the best and worst cases as measured by the average Dice metric. Fig. 2 shows that our proposed CNN is able to provide reasonable delineation of endocardial contours, mimicking expert contours with the inclusion of papillary muscles and trabeculae within the LV blood pool. In the worst case (Fig. 3), the network appears to have difficulty returning lower endocardium radius values at ES; possibly being confounded by the relatively small size of the heart as well as the thick myocardium. There are multiple other cases within the MICCAI 2009 datasets with similar characteristics (smaller heart or blood pool), but Fig. 3 was the only case where the average Dice metric fell below 0.8.

Table III compares our results against other published work. Using two of the assessment criteria, the Dice metric and APD, the results are competitive or superior to most published methods, although the approaches of Avendi et al. [5] and Queiros et al. [8] reported slightly higher values. In contrast,

TABLE III. COMPARISON OF RESULTS AGAINST OTHER PUBLISHED WORKS EVALUATED USING MICCAI 2009 DATASET. VALUES FOR DICE METRIC AND APD ARE MEAN (STANDARD DEVIATION). RESULTS FROM OTHER PUBLISHED WORKS ADAPTED FROM TABLE 2 OF AVENDI ET AL.[6]

Method	#[a]	Good[b] (%)	Dice	APD[c] (mm)
This paper	**45**	**97.9**	**.88(.10)**	**2.30(1.11)**
Avendi et al.	30	96.7	.94(.02)	1.81(0.44)
Queiros et al.	45	92.7	.9 (.05)	1.76(0.45)
Ngo et al.	30	93.2	.89(.03)	2.26(0.46)
Hu et al.	45	91.1	.89(.03)	2.24(0.4)
Constantinides et al.	45	80	.86(.05)	2.44(0.56)
Liu et al.	45	91.2	.88(.03)	2.36(0.39)
Huang et al.	45	79.2	.89(.04)	2.16(0.46)
Schaerer et al.	45	-	.87(.04)	2.97(0.38)
Jolly et al.	30	95.6	.88(.04)	2.26(0.59)

[a] Number of datasets evaluated. 30 – validation and online datasets, 45 – all datasets

[b] Good contours, where per slice average perpendicular distance < 5 mm

[c] Average perpendicular distance (APD)

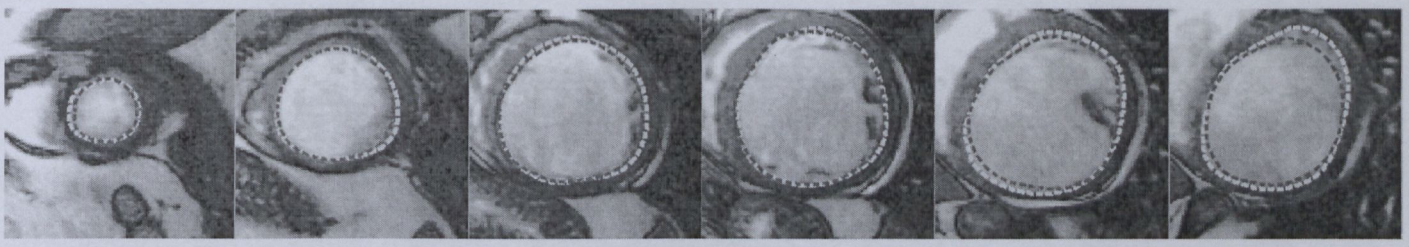

Fig. 2. Good results from apex (left) to base (right) at ED phase. Yellow and red contours are gold standard and network results, respectively.

Fig. 3. Bad results from apex (left) to base (right) at ES phase. Yellow and red contours are gold standard and network results respectively. The network appears to have difficulty returning lower endocardium radius values, especially near the apex.

we reported 97.9% good contours (averaged per slice) or 100% good contours (averaged per subject). This is the best reported result to date, which suggests that the slightly deficient Dice and APD results reflect different conditions in the training data as opposed to actual deficiencies in the algorithm. To elaborate, even amongst human experts there is variance in the exact definition of the endocardial boundary. E.g., how best to encapsulate the papillary muscles and trabeculae at the endocardial wall. As our network was trained against the MICCAI 2011 dataset, the results will reflect the assumptions made by the gold standard assessors in that set as opposed to the assessors in the MICCAI 2009 set. Importantly, these results demonstrate that the algorithm is highly transferable across varying datasets.

We make special comparison to Avendi et al., who utilized CNNs for per-pixel categorization [5]. Avendi et al. manually filtered out apical slices for separate processing, and integrated a comprehensive deformable model for post-processing. We did neither, while still retaining highly competitive performance (superior in the case of good contour percentage).

The network performs weaker at end-systole (ES) and at the apex (Fig. 4). If we omit the most apical slices, the Dice metric and APD improves to 0.90 and 2.27 mm, respectively.

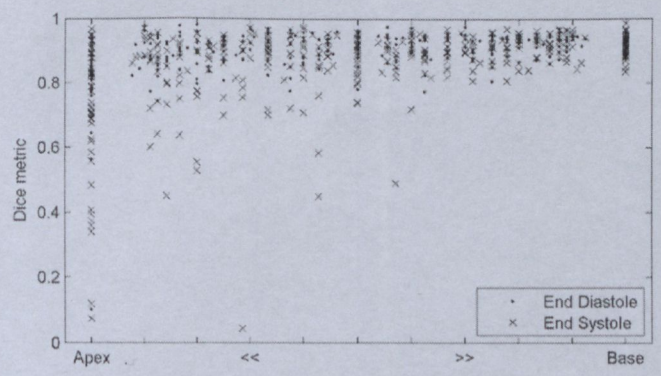

Fig. 4. Dice metric at ED and ES phases, from apex to basal slices. Performance is noticeably weaker at ES and at the LV apex.

This is notable, as the most apical slice is usually negligible in the calculation of clinical parameters such as ejection fraction (EF), due to its small volume relative to the other slices.

In conclusion, we have demonstrated that CNN regression is a viable and highly promising method for automated LV endocardial segmentation at ED and ES phases, and is capable of generalizing learning between highly distinct training and testing data sets.

References

[1] Y. Faridah Abdul Aziz, F. Fadzli, R. Rizal Azman, F. Mohamed Sani, A. Vijayananthan, and M. Nazri, "State of the Heart: CMR in Coronary Artery Disease," *Current Medical Imaging Reviews*, vol. 9, no. 3, pp. 201–213, Aug. 2013.

[2] C. M. Kramer, J. Barkhausen, S. D. Flamm, R. J. Kim, and E. Nagel, "Standardized cardiovascular magnetic resonance (CMR) protocols 2013 update," *J Cardiovasc Magn Reson*, vol. 15, no. 1, pp. 1–10, Dec. 2013.

[3] C. Petitjean and J.-N. Dacher, "A review of segmentation methods in short axis cardiac MR images," *Med. Image Anal.*, vol. 15, no. 2, pp. 169–184, Apr. 2011.

[4] J. Schmidhuber, "Deep learning in neural networks: An overview," *Neural Networks*, vol. 61, pp. 85–117, Jan. 2015.

[5] M. R. Avendi, A. Kheradvar, and H. Jafarkhani, "A combined deep-learning and deformable-model approach to fully automatic segmentation of the left ventricle in cardiac MRI," *Medical Image Analysis*, vol. 30, pp. 108–119, May 2016.

[6] P. Radau, Y. Lu, K. Connelly, G. Paul, A. J. Dick, and G. A. Wright, "Evaluation Framework for Algorithms Segmenting Short Axis Cardiac MRI," *MIDAS J. - Card. MR Left Ventricle Segmentation Chall.*, Jul. 2009.

[7] A. Suinesiaputra, B. R. Cowan, J. P. Finn, C. G. Fonseca, A. H. Kadish, D. C. Lee, P. Medrano-Gracia, S. K. Warfield, W. Tao, and A. A. Young, "Left Ventricular Segmentation Challenge from Cardiac MRI: A Collation Study," in *Statistical Atlases and Computational Models of the Heart. Imaging and Modelling Challenges*, O. Camara, E. Konukoglu, M. Pop, K. Rhode, M. Sermesant, and A. Young, Eds. Springer Berlin Heidelberg, 2012, pp. 88–97.

[8] S. Queirós, D. Barbosa, B. Heyde, P. Morais, J. L. Vilaça, D. Friboulet, O. Bernard, and J. D'hooge, "Fast automatic myocardial segmentation in 4D cine CMR datasets," *Medical Image Analysis*, vol. 18, no. 7, pp. 1115–1131, Oct. 2014.