CardiacNET: Segmentation of Left Atrium and Proximal Pulmonary Veins from MRI Using Multi-view CNN

Aliasghar Mortazi[1(✉)], Rashed Karim[2], Kawal Rhode[2], Jeremy Burt[3], and Ulas Bagci[1]

[1] Center for Research in Computer Vision (CRCV), University of Central Florida, Orlando, FL, USA
a.mortazi@knights.ucf.edu

[2] Division of Imaging Sciences and Biomedical Engineering, King's College London, London, UK

[3] Diagnostic Radiology Department, Florida Hospital, Orlando, FL, USA

Abstract. Anatomical and biophysical modeling of left atrium (LA) and proximal pulmonary veins (PPVs) is important for clinical management of several cardiac diseases. Magnetic resonance imaging (MRI) allows qualitative assessment of LA and PPVs through visualization. However, there is a strong need for an advanced image segmentation method to be applied to cardiac MRI for quantitative analysis of LA and PPVs. In this study, we address this unmet clinical need by exploring a new deep learning-based segmentation strategy for quantification of LA and PPVs with high accuracy and heightened efficiency. Our approach is based on a multi-view convolutional neural network (CNN) with an adaptive fusion strategy and a new loss function that allows fast and more accurate convergence of the backpropagation based optimization. After training our network from scratch by using more than 60K 2D MRI images (slices), we have evaluated our segmentation strategy to the STACOM 2013 cardiac segmentation challenge benchmark. Qualitative and quantitative evaluations, obtained from the segmentation challenge, indicate that the proposed method achieved the state-of-the-art sensitivity (90%), specificity (99%), precision (94%), and efficiency levels (10s in GPU, and 7.5 min in CPU).

Keywords: Left atrium · Pulmonary veins · Deep learning · Cardiac magnetic resonance · MRI · Image segmentation · CardiacNET

1 Introduction

Atrial fibrillation (AF) is a cardiac arrhythmia caused by abnormal electrical discharges in the atrium, often beginning with hemodynamic and/or structural changes in the left atrium (LA) [1]. AF is clinically associated with LA strain, and MRI is shown to be a promising imaging method for assessing the disease

M. Descoteaux et al. (Eds.): MICCAI 2017, Part II, LNCS 10434, pp. 377–385, 2017.
DOI: 10.1007/978-3-319-66185-8_43

state and predicting adverse clinical outcomes. The LA also has an important role in patients with ventricular dysfunction as a booster pump to augment ventricular volume [2]. Computed tomography (CT) imaging of the heart is frequently performed when managing AF and prior to pulmonary vein ablation (isolation) therapy due to its rapid processing time. In recent years, there is an increasing interest in shifting towards cardiac MRI due to its excellent soft tissue contrast properties and lack of radiation exposure. For pulmonary vein ablation therapy planning in AF, precise segmentation of the LA and PPVs is essential. However, this task is non-trivial because of multiple anatomical variations of LA and PPV.

Historically, statistical shape and atlas-based methods have been the state-of-the-art cardiac segmentation approaches due to their ability to handle large shape/appearance variations. One significant challenge for such approaches is their limited efficiency: an average of 50 min processing time per volume [3]. Statistical shape models are faster than atlas-based methods, and a high degree uncertainties in the accuracy of such models is inevitable [4]. To alleviate this problem and accomplish the segmentation of LA and PPVs from 3D cardiac MRI with high *accuracy* and *efficiency*, we propose to a new deep CNN. Our proposed method is fully automated, and largely different from previous methods of LA and PPVs segmentation. The summary of these differences and key novelties of the proposed method, named as *CardiacNET*, are listed as follows:

- Training CNN from scratch for 3D cardiac MRI is not feasible with insufficient 3D training data (with ground truth) and limited computer memory. Instead, we parsed 3D data into 2D components (axial (A), sagittal (S), and coronal (C)), and utilized a separate deep learning architecture for each component. The proposed *CardiacNET* was trained using more than 60K 2D slices of cardiac MR images without relying on a pre-training network of non-medical data.
- We have combined three CNN networks through an adaptive fusion mechanism where complementary information of each CNN was utilized to improve segmentation results. The proposed adaptive fusion mechanism is based on a new strategy; called *robust region*, which measures (roughly) the reliability of segmentation results without the need for ground truth.
- We devised a new loss function in the proposed network, based on a modified z-loss, to provide fast convergence of network parameters. This not only improved segmentation results due to fast and reliable allocation of network parameters, but it also provided a significant acceleration of the segmentation process. The overall segmentation process for a given 3D cardiac MRI takes at most 10s in GPU, and 7.5 min in CPU on a normal workstation.

2 Proposed Multi-view Convolutional Neural Network (CNN) Architecture

The proposed pipeline for deep learning based segmentation of the LA and PPVs is summarized in Fig. 1. We used the same CNN architecture for each view of the

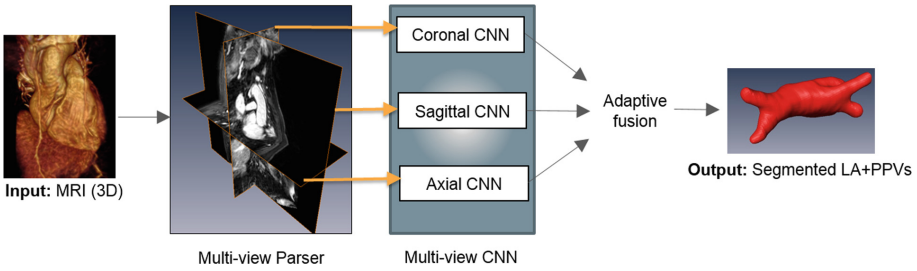

Fig. 1. High-level overview of the proposed multi-view CNN architecture.

3D cardiac MRI after parsing them into axial, sagittal, and coronal views. The rationale behind this decision is based on the limitation of computer memory and insufficient 3D data for training on 3D cardiac MRI from scratch. Instead, we reduced the computational burden of the CNN training by constraining the problem into a 2D domain. The resulting pixel-wise segmentations from each CNN are combined through an adaptive fusion strategy. The fusion operation was designed to maximize the information content from different views. The details of the pipeline are given in the following subsections.

Encoder-Decoder CNN: We constructed an encoder-decoder CNN architecture, similar to that of Noh et al. [5]. The network includes 23 layers (11 in encoder, 12 in decoder units). Two max-pooling layers in encoder units reduce the image dimensions by half, and a total of 19 convolutional (9 in encoder, 10 in decoder), 18 batch normalization, and 18 ReLU (rectified linear unit) layers are used. Specific to the decoder unit, two upsampling layers are used to convert the images back into original sizes. Also, the kernel size of all filters are considered as 3×3. The final layer of the network includes a softmax function (logistic) for generating a probability score for each pixel. Details of these layers, and associated filter size and numbers are given in Fig. 2.

Loss Function: We used a new loss function that can estimate the parameters of the proposed network at a much faster rate. We trained end-to-end mapping with a loss function $L(\mathbf{o}, c) = \text{softplus}(a(b - z_c))/a$, called z-loss [6], where $\mathbf{o}$ denotes output of the network, c denotes the ground truth label, and z_c indicate z-normalized label, obtained as $z_c = (o_c - \mu)/\sigma$ where mean (μ) and standard deviation σ are obtained from $\mathbf{o}$. z-loss is simply obtained with the reparametrization of *soft-plus* (SP) function (i.e., $SP(x) = ln(1+e^x)$) through two hyperparameters: a and b. Herein, we kept these hyperparameters fixed, and trained the network with a reduced z-loss function. The rationale behind this choice is the following: the z-loss function provides an efficient training performance as it belongs to spherical loss family, and it is invariant to scale and shift changes in the output, avoiding output parameters to deviate from extreme values.

Training *CardiacNET* from Scratch: 3D cardiac MRI images along with its corresponding expert annotated ground truths were used to train the CNN

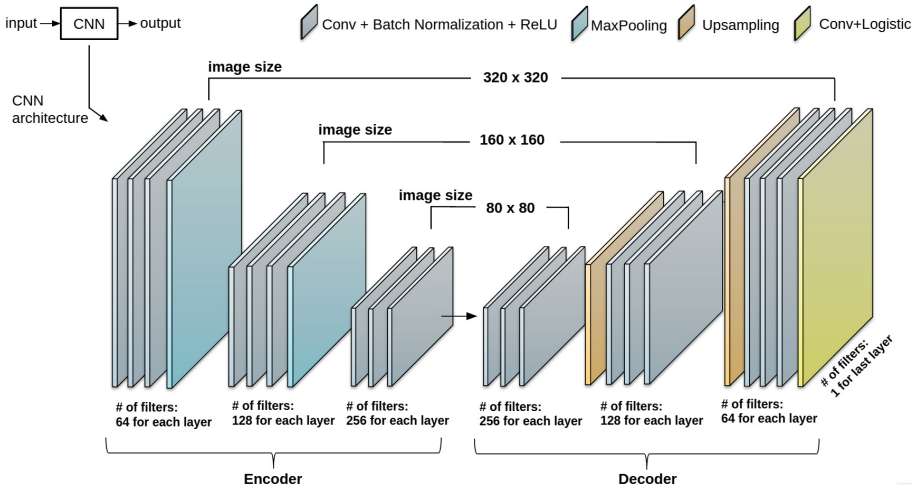

Fig. 2. Details of the CNN architecture. Note that image size is not necessarily fixed for each view's CNN.

after the images are parsed into three views (A, S, C). Data augmentation has been conducted on the training dataset with translation and rotation operation as indicated in Table 1. Obtained 3D images were parsed into A, S, and C views, and more than 60K 2D images were obtained to feed training of the CNN (approximately 30K for A and C views, around 11K for S view). The 9 of the subjects and their corresponding augmented data are considered as a training and 1 subject and its corresponding augmented data is considered as validation. As a preprocessing step, all images have undergone anisotropic smoothing filtering and histogram matching.

Multi-view Information Fusion. Since cardiac MRI is often not reconstructed with isotropic resolution, we expected varying segmentation accuracy in different views. In order to alleviate potential adverse effects caused by non-isotropic spatial resolutions of a particular view, it is desirable to reduce the contribution of that view into final segmentation. We have achieved this with the adaptive fusion strategy as described next. For a given MRI volume **I**, and its corresponding segmentation **o**, we proposed a new strategy, called *robust region*, that roughly determined the reliability of the output segmentation **o** by assessing its object distribution. To achieve this, we hypothesized that the output

Table 1. Data augmentation parameters and number of training images

Data augmentation		
Methods	Parameters	
Translations	$(x + trans, y = 0), trans\ \varepsilon[-20, 20]$	
	$(x = 0, y + trans), trans\ \varepsilon[-20, 20]$	
Rotation	$k \times 45, k\,\varepsilon[-2, -1, 1, 2]$	
Training images		
CNN	# of images	Image size
Sagittal	10,800	320×0
Axial	28,800	110×0
Coronal	28,800	110×0

should include only one connected object when the segmentation is successful, and if there was more than a single connected object available, these can be considered as false positives. Accordingly, respective performance of segmentation performance in A, S, and C views can be compared and weighted. To this end, we utilized connected component analysis (CCA) to rank output segmentations and reduced the contribution of CNN for a particular view when false positive findings (non-trusted objects/components) were large and true positive findings (trusted object/component) were small. Figure 3 describes the adaptive fusion strategy as $CCA(\mathbf{o}) = \{o_1, \ldots, o_n | \cup\ o_i = \mathbf{o}, \text{and} \cap\ o_i = \phi\}$. Thus, the contribution of each view's CNN was computed based on a weighting $w = max_i\{|o_i|\}/\sum_i |o_i|$, indicating that higher weights were assigned when the component with largest volume dominated the whole output volume. Note that this block has been used only in the test phase. Complementary to this strategy, we also used simple linear fusion of each views for comparison (See Experimental Results section).

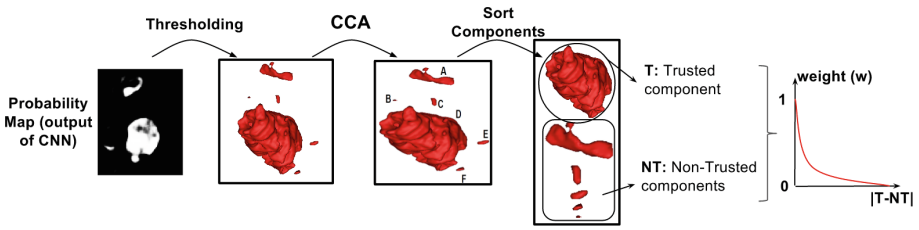

Fig. 3. Connected components obtained from each view were computed and the residual volume (T-NT) was used to determine the strength for fusion with the other views.

3 Experimental Results

Data sets: Thirty cardiac MRI data sets were provided by the STACOM 2013 challenge organizers [3]. Ten training data were provided with ground truth labels, and the remaining twenty were provided as a test set. It is important to note that not the complete PVs are considered in the segmentation challenge, but only the proximal segments of the PVs up to the first branching vessel or after 10 mm from the vein ostium were included in the segmentation. MR images were obtained from a 1.5T Achieva (Philips Healtcare, The Neatherlands) scanner with an ECG-gated 3D balanced steady-state free precession acquisition [3] with TR/TE = 4.4/2.4 ms, and Flip-angle = 90°. Typical acquisition time for the cardiac volume imaging was 10 min. In-plane resolution was recorded as $1.25 \times 25\,\text{mm}^2$, slice thickness was measured as 2.7 mm. Further details on the data acquisition, and image properties can be found in [3].

Evaluations. For evaluation and comparison with other state-of-the-art method, we have used the same evaluation metrics, provided by the STACOM

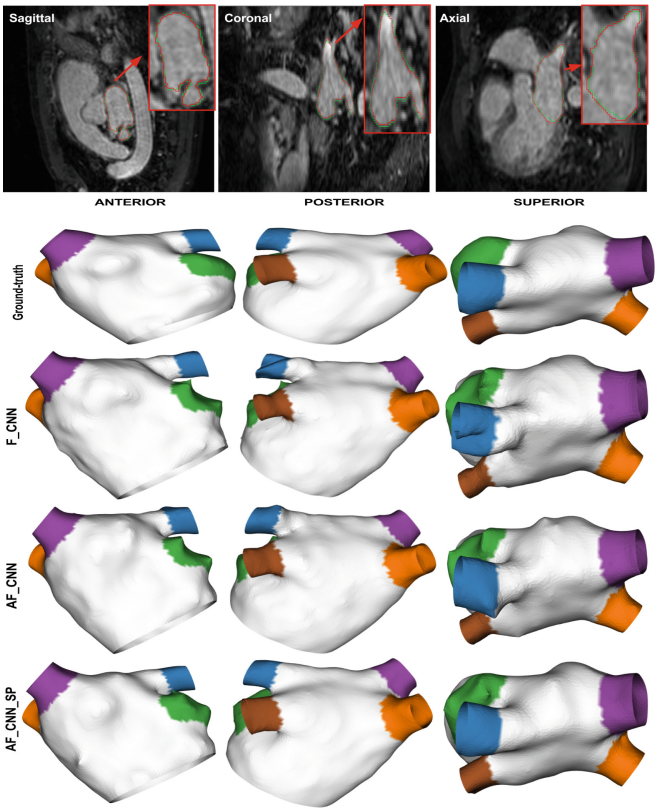

Fig. 4. First row shows sample MRI slices from S, C, and A views (red contour is ground-truth and green one is output of proposed method). Second-to-fifth rows: 3D surface visualization for the ground-truth and the output generated by the proposed method w.r.t simple fusion (F), adaptive fusion (AF), and the new loss function (SP).

2013 challenge: Dice index and surface-to-surface (S2S) metrics. In addition, we calculated Dice index and S2S for the LA and PPVs separately. To provide a comprehensive evaluation and comparisons, sensitivity (true positive rate), specificity (true negative rate), precision (positive prediction value), and Dice index values for the combined LA and PPVs were included too. Table 2 summarizes all these evaluation metrics along with efficiency comparisons where we tested our algorithm both in GPU and CPU. LTSI-VRG, UCL-1C, and UCL-4C are three atlas-based method which their output were published publicly as a part of STACOM 2013 challenge. Also, OBS-2 is the result from human observer which its output was available as a part of STACOM 2013 challenge. Using leave-one-out cross-validation strategy on training dataset, we achieved high sensitivity (0.92) and Dice value (0.93). Similarly, in almost all evaluation metrics in the test set, the proposed method out-performed the state-of-the-art approaches by large margins. Table 2 indicates the results of varying combinations using *Cardiac-*

Table 2. The evaluation metrics for state-of-the-art and proposed methods. **: the running time on CPU *: the running time on NVIDIA TitanX GPU

Methods	LTSI_VRG	UCL_1C	UCL_4C	OBS_2	A_CNN	C_CNN	S_CNN	F-CNN	AF-CNN	AF-CNN-SP
Dice(LA)	0.910	0.938	0.859	0.908	0.903	0.804	0.787	0.873	0.928	**0.951**
Dice(PPVs)	0.653	0.609	0.646	0.751	0.561	0.478	0.398	0.506	0.616	**0.685**
S2S(LA) in mm	1.640	1.086	2.136	1.538	1.592	2.679	2.853	1.771	1.359	**1.045**
S2S(PPVs) in mm	1.994	1.623	2.375	1.594	1.928	2.878	3.581	2.121	1.718	**1.427**
Sensitivity	**0.926**	0.828	0.832	0.894	0.806	0.658	0.663	0.743	0.883	0.895
Specificity	0.998	0.999	0.999	0.997	0.996	0.994	0.997	0.997	0.999	**0.999**
Precision	0.815	**0.957**	0.814	0.936	0.905	0.774	0.880	0.953	0.936	0.938
Dice (all)	0.862	0.886	0.819	0.911	0.845	0.695	0.734	0.820	0.887	**0.905**
Running	3100**	1200**	1200**	-	170**	170**	155**	**450****	**450****	**450****
Time (sec)	-	-	-	-	3.5*	3.5*	3*	**10***	**10***	**10***

NET such as single CNN in particular view (i.e., *S_CNN*), with simple linear fusion F-CNN, adaptive fusion AF-CNN, and with the new loss function AF-CNN-SP. In AF-CNN, the loss function was cross-entropy. The best method in the challenge data set was reported to have a Dice index of 0.94 for LA and 0.65 for PPVs (combined LA and PPVs was less than 0.9). In our proposed method, the Dice index for combined LA and PPVs was well above 0.90. For efficiency comparison, our approach only takes at most 10s on a Nvidia TitanX GPU and 7.5 min in a CPU with Octa-core processor (2.4 GHz) configuration. The method in [7] required 30–45 min of processing times (with Quad-core processor (2.13 GHz)). For qualitative evaluation, we have used surface rendering of output segmentations compared to ground truth in Fig. 4. Sample axial, sagittal, and coronal MRI slices are given in the same figure with ground truth annotations overlaid with the segmented LA and PPVs.

4 Discussions and Concluding Remarks

The advantage of *CardiacNET* is accurate and efficient method for both LA and PPVs segmentation in atrial fibrillation patients: combined segmentation of the LA and PPVs. Precise segmentation of the LA and PPVs is needed for ablation therapy planning and clinical guidance in AF patients. PPVs have a greater number of anatomical variations than the LA-body, leading to challenges with accurate segmentation. Joint segmentation the LA and PPVs is even more challenging compared to sole LA-body segmentation. Nevertheless, with all available quantitative metrics, the proposed method has been shown to greatly improve the segmentation accuracy on the existing benchmark for LA and PPVs segmentation. The benchmark evaluation has also allowed the method and its variations to be cross-compared on the same dataset with other existing methods in literature (Fig. 5).

Despite the efficacy of the proposed method, there are several possibilities that our work can be extended in future studies. Firstly, the new method will be tested, evaluated, and validated our in more diverse data sets from several independent cohorts, and at the different imaging resolution and noise levels,

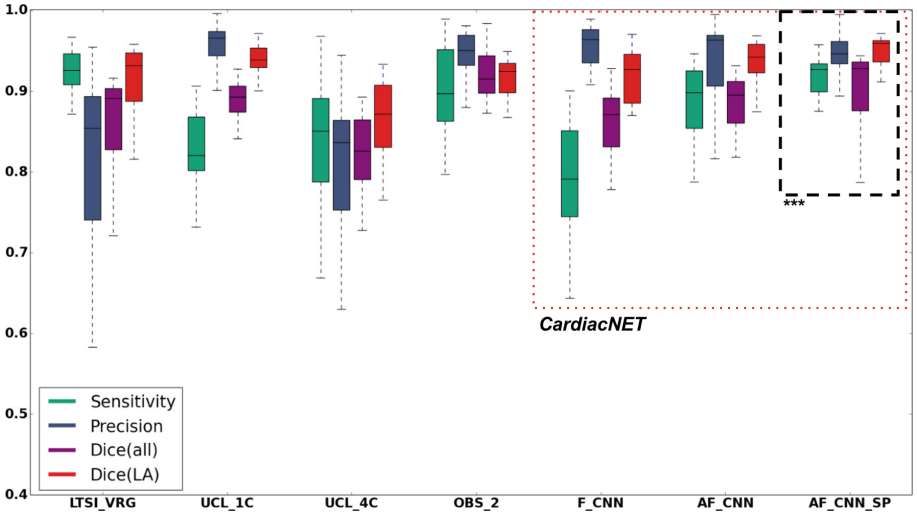

Fig. 5. Box plots for sensitivity, precision, and Dice index for state-of-the-art (LTSI_VRG,UCL_1C, UCL_4C, OBS_2) and proposed methods (F_CNN, AF_CNN, AF_CNN_SP) on the LA segmentation benchmark

and even across different scanner vendors. Secondly, extending our framework into 4D (i.e. motion) analysis of cardiac images can be possible by extending our parsing strategy. Thirdly, we aim to explore the feasibility of training completely 3D cardiac MRI based on the availability of multiple GPUs, or developing sparse CNNs to alleviate the segmentation problem. Fourthly, with low-dose cardiac CT technology on the rise; it is desirable to have similar network structure trained on CT scans. This notable efficacy of the deep learning strategies presented in this work promises a similar performance on CT scans.

In conclusion, the proposed method has utilized the strength of deeply trained CNN to segment LA and PPVs from cardiac MRI. We have shown combining information from different views of MRI by using an adaptive fusion strategy and a new loss function improves segmentation accuracy and efficiency significantly.

Acknowledgment. Thanks to Nvidia for donating a GPU for deep learning experiments. All CNN experiments have been conducted using Tensorflow.

References

1. Kuppahally, S.S., et al.: Left atrial strain and strain rate in patients with paroxysmal and persistent atrial fibrillation relationship to left atrial structural remodeling detected by delayed-enhancement MRI. Circ.: Cardiovasc. Imaging **3**, 231–239 (2010)
2. Daoudi, A., Mahmoudi, S., Chikh, M.A.: Automatic segmentation of the left atrium on CT images. In: Camara, O., Mansi, T., Pop, M., Rhode, K., Sermesant, M., Young, A. (eds.) STACOM 2013. LNCS, vol. 8330, pp. 14–23. Springer, Heidelberg (2014). doi:10.1007/978-3-642-54268-8_2

3. Tobon-Gomez, C., et al.: Benchmark for algorithms segmenting the left atrium from 3D CT and MRI datasets. IEEE TMI **34**(7), 1460–1473 (2015)
4. Stender, B., Blanck, O., Wang, B., Schlaefer, A.: Model-based segmentation of the left atrium in CT and MRI scans. In: Camara, O., Mansi, T., Pop, M., Rhode, K., Sermesant, M., Young, A. (eds.) STACOM 2013. LNCS, vol. 8330, pp. 31–41. Springer, Heidelberg (2014). doi:10.1007/978-3-642-54268-8_4
5. Noh, H., Hong, S., Han, B.: Learning deconvolution network for semantic segmentation. In: Proceedings of the IEEE International Conference on Computer Vision, pp. 1520–1528 (2015)
6. de Brébisson, A., Vincent, P.: The Z-loss: a shift and scale invariant classification loss belonging to the spherical family. arXiv preprint arXiv:1604.08859 (2016)
7. Zuluaga, M.A., Cardoso, M.J., Modat, M., Ourselin, S.: Multi-atlas propagation whole heart segmentation from MRI and CTA using a local normalised correlation coefficient criterion. In: Ourselin, S., Rueckert, D., Smith, N. (eds.) FIMH 2013. LNCS, vol. 7945, pp. 174–181. Springer, Heidelberg (2013). doi:10.1007/978-3-642-38899-6_21