MULTI-COMPONENT DEFORMABLE MODELS COUPLED WITH 2D-3D U-NET FOR AUTOMATED PROBABILISTIC SEGMENTATION OF CARDIAC WALLS AND BLOOD

Dong Yang[1]*, Qiaoying Huang[1]*, Leon Axel[2], , Dimitris Metaxas[1]

[1] CBIM, Rutgers University, Piscataway, NJ 08854, USA
[2] Department of Radiology, New York University, New York, NY 10016, USA

ABSTRACT

The segmentation of the ventricular wall and the blood pool in cardiac magnetic resonance imaging (MRI) has been investigated for decades, given its important role for delineation of cardiac functioning and diagnosis of heart diseases. One of the major challenges is that the inner epicardium boundary is not always visible in the image domain, due to the mixture of blood and muscle structures, especially at the end of contraction, or systole. To address it, we propose a novel approach for the cardiac segmentation in the short-axis (SAX) MRI: coupled deep neural networks and deformable models. First, a 2D U-Net is adopted for each magnetic resonance (MR) slice, and a 3D U-Net refines the segmentation results along the temporal dimension. Then, we propose a multi-component deformable model to extract accurate contours for both endo- and epicardium with global and local constraints. Finally, a partial blood classification is explored to estimate the presence of boundary pixels near the trabeculae and solid wall, and to avoid moving the endocardium boundary inward. Quantitative evaluation demonstrates the high accuracy, robustness, and efficiency of our approach for the slices acquired at different locations and different cardiac phases.

Index Terms— Cardiac MRI, ventricular wall segmentation, U-Net, multi-component deformable model, partial blood classification

1. INTRODUCTION

Cardiovascular disease, such as cardiac dyssynchrony, hypertrophic cardiomyopathy and myocardial infarction, is one of the major causes of human death. It is vitally important to understand cardiac function comprehensively for disease prevention, diagnosis and treatment. In order to analyze cardiac motion, one of the most essential clinical tasks is extracting left ventricle (LV) contours for myocardium muscle layers at both end-diastolic volume (EDV) and end-systolic volume (ESV) in cardiac magnetic resonance imaging (MRI). The contour extraction, used for computing ventricular global

---

functions, is equivalent to the heart wall segmentation. In the conventional wall segmentation, the task is done either manually, interactively placing a contour at the best visual estimate of the boundaries of the solid wall, or more automatically, using mathematical optimization and learning-based methods with some smooth outer hull being placed around the "blood" and "muscle" pixels. There are effectively three concentric "zones" in the ventricle: **1)** solid muscle zone consisting of the outer wall and the endocardium wall, **2)** transitional zone with mixed blood and muscle structures, and **3)** mostly blood zone (with possibly a few muscle bundles running through it). The principal challenge associated with the conventional segmentation methods [1] is the difficulty in reliably distinguishing the trabeculae (and papillary muscles) from the underlying solid muscle wall, especially in cardiac phases near end-systole, when the blood is largely squeezed out from between the trabeculae, making them blend with each other and the wall. This causes conventional approaches to tend to fail for the end-systole determination in many cases, especially when there is hypertrophy of the wall and trabecular structures, with a resulting under-estimation for the end-systolic cavity volume and an associated over-estimation of the ejection fraction (EF).
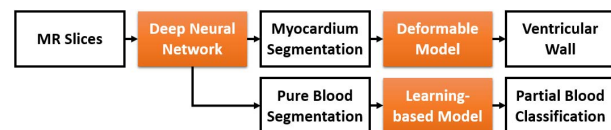

Fig. 1. The flowchart of the proposed approach.

We propose an automatic heart wall (myocardial muscle) segmentation approach using the deep neural networks coupled with a new multi-component deformable model. First, the 2D-3D neural network model provides fine segmentation masks of muscle layers with temporal continuity. Then, the multi-component deformable model is adapted to extract contours dynamically along the cardiac cycle, for both inner and outer heart walls, from the segmentation masks. The neural networks provide external force for the deformable models. The global and local constraints in the deformable model help avoid having the apparent detected boundary move arti-

factually inward, especially for epicardium/inner wall. Furthermore, we explore the "partial blood content" approach to estimate the presence of boundary pixels near the trabeculae (and papillary muscles) as well as the solid wall. Adaptively adjusting the associated threshold according to the cardiac cycle phase, we should be able to improve our sensitivity to the smaller bits of blood between the trabeculae near end-systole.

2. RELATED WORK

The LV myocardium segmentation has been addressed by many researchers in the past decades, in order to alleviate the human effort for the time compsuming annotation procedure. For example, Paragios developed a segmentation pipeline using level-set optimization and gradient vector flow (geodesic active contour) [2]. Jolly proposed a multi-stage graph-cut method for cardiac segmentation in both MRI and CT [3]. These methods rely on the image appearance, and they would probably fail when the image contrast is changed. Zhu et al. [4] introduced a subject-specific dynamic model to delineate the ventricular shape variance. However, their results tend to move a bit inward to the blood pool, which may introduce errors for the global function estimation. Recently, the cardiac segmentation has been addressed using deep neural networks [5, 6, 7, 8], benefiting from its advanced feature learning capacities. However, most of these learning models are trained and used to infer the boundaries for individual images, which tend to lead to lack of temporal continuity in the segmentation.

As an alternative, Codella et al. tried to get the best estimate of the total "true" blood volume in the chamber, by weighting each candidate voxel by its fractional blood content and then summing them [9, 10]. This would then be used for calculation of the conventional global function measures (based on differences in the blood in the ventricle between end-diastole and end-systole), including stroke volume and ejection fraction (EF). It was presented as an alternative to the conventional method of defining the "cavity," the trabeculae, and the papillary muscles, in order to segment out the "solid" wall only. In their approach, the voxel probability distribution was computed directly from the intensity scale, which may cause errors when the imaging quality is compromised. Our approach addresses previous limitations and correctly segments the 3 zones based on the coupled U-Net and the multi-component deformable model.

3. METHODOLOGY

Figure 1 shows the flowchart of our proposed approach for myocardium segmentation in cardiac MRI. Initially, coarse segmentation results are generated from the 2D U-Net for individual images. Then, the previous results are stacked into 3D volumes according to their order of cardiac phases and cropped into a centralized region-of-interest (ROI) according to the coarse segmentation for segmentation refinement with a

3D U-Net. Finally, we utilize a multi-component deformable model to determine the myocardium boundaries with global and local constraints. We also compute the probability of pixels belonging to blood, based on the features learned from deep neural networks.

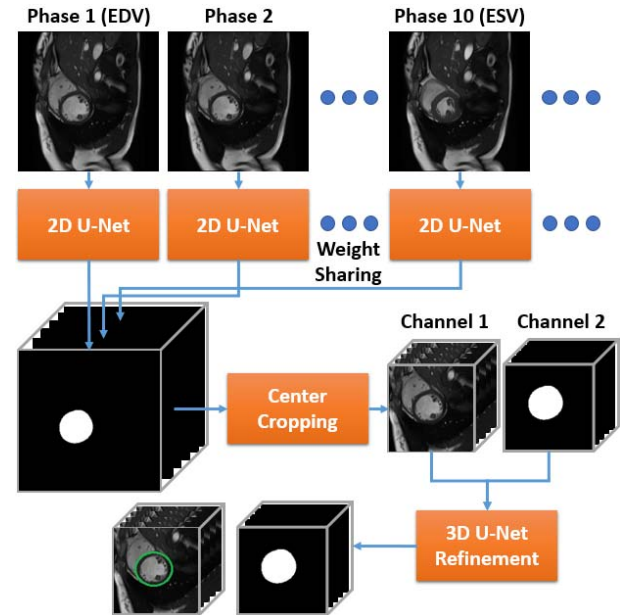### 3.1. 2D-3D U-Net Model

Fig. 2. The flowchart of the proposed 2D-3D U-Net method. The 2D U-Net is used for generating segmentation priors at each individual cardiac phase, and the 3D one further refines the segmentation results along the temporal dimension with a small cropping region.

The concept of U-Net, first proposed in [11] has been successfully applied in many applications of medical image analysis. It has been validated to possess good generalization capacity with few annotated samples. The network consists of a convolutional encoder and decoder, and its U-shape generates multi-scale features and computes them with multi-step convolution and up-sampling. The output of the U-Net shares the same size as input. There are skip connections in between the encoder and decoder to concatenate multi-level feature maps, allowing the decoder to store back the relative features that are lost in the prior stage.

Our 2D-3D U-Net model is described in the Fig. 2. For individual phases of MR sequences, we adopt two 2D U-Nets [11] to segment epicardium and endocaridium masks, respectively. Direct predicting myocadium muscle using only one network is also possible. However, in that scenario, the positive samples in the gold standard are generally much less than the negative samples, which makes the learning procedure biased and affects the performance later. After achieving preliminary segmentation using 2D U-Nets, we crop the region-of-interest (ROI) according to the center of segmentation for

both image and segmentation. The same cropping region is applied for all phases in one MR sequence. Then, we stack all the cropped images at the same location, and segment them along the temporal dimension into 3D volumes as input, and adopt 3D U-Nets [12] to refine the previous segmentation. The 3D U-Nets would enforce the smooth prediction in-between consecutive cardiac phases. It may not be easy to directly apply 3D U-Net to the entire image and segmentation regions, since the input with the original size would have large memory consumption and become a bottleneck during computation with GPUs.

Our key point here is to use a convolutional model instead of a recurrent model to handle the temporal data. Although the recurrent model is well established for time-series problems, recent research shows that a fully convolutional model could outperform a recurrent model in some sequential problems, for instance, language translation [13], and video segmentation [14]. Furthermore, similar 2D-3D network models have been adopted for a few medical image segmentation applications and achieved excellent results [15].

3.2. Multi-Component Deformable Model

In clinical practice, doctors and physicians often manually correct overestimated regions of the inner wall contour, relying on playing the serial frames of the cardiac cycle as a movie of the imaged slice to locate trabeculae pretty reliably at end-systole. Then their associated motion can be estimated over the cardiac cycle from the moving animated display, including when they are too close together to reliably detect in an isolated frame. Similarly, we propose a multi-component deformable model to finalize the contours of endo- and epicardium, to simulate the manual correction. At each cardiac phase, the energy function of the deformable model for epicardium can be written as follows.

$$E_{\text{epi}} = \alpha E_{\text{external}} + \beta E_{\text{continuity}} + \gamma E_{\text{smooth}} \quad (1)$$

The energy function of the deformable model for endocardium is different from Eq. 1, as follows.

$$E_{\text{endo}} = \alpha E_{\text{external}} + \beta E_{\text{continuity}} + \gamma E_{\text{smooth}} + \phi E_{\Delta\text{area}} \quad (2)$$

The external energy is $E_{\text{external}} = \int \|v_s - v_s'\|_2^2 \, ds$. v_s and v_s' are the corresponding points of the current deformable model and targeting contour from previous deep neural networks. The continuity term $E_{\text{continuity}} = \int \left\|\frac{dv_s}{ds}\right\|_2^2 \, ds$ ensures that the neighboring points are close. The smoothness term $E_{\text{smooth}} = \int \left\|\frac{d^2 v_s}{ds^2}\right\|_2^2 \, ds$ guarantees that the model is always a convex smooth shape. Here, $\alpha, \beta, \gamma, \phi$ are all positive constants. In practice $\alpha = 1.0$, $\beta = \gamma = \phi = 0.2$. The epicardium deformable model has an extra energy term $E_{\Delta\text{area}} = \left| \int \|v_s - w_s\|_2^2 \, ds - \int \|v_s'' - w_s''\|_2^2 \, ds \right|$. w_s are the points of the fixed epicardium contour in the current phase. v_s'' and w_s'' are the corresponding points from the previous

endo- and epicardium deformable models. $\int \|v_s'' - w_s''\|_2^2 \, ds$ is equivalent to the myocardium area in the previous phase. The assumption is that the area of myocardium muscle can only change within a limit range among neighboring cardiac phases because the muscle volume is almost unchanged during the cardiac cycle. At inference, we start from the contours of the first phase (normally EDV). The epicardium contour of the next phase is computed by solving Eq. 1 and moving contour points along their normal directions, till reaching a minimum status. Then the endocardium contour is computed by minimizing Eq. 2 together with the updated epicardium contour. The myocardium contours are derived phase-by-phase in sequence; and the sample results are shown in Fig. 3.
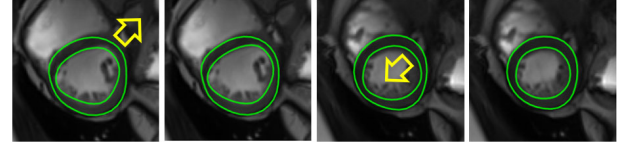

Fig. 3. The contours before and after applying deformable models. Left: the contours from previous frame, right: the updated contours. The yellow arrows indicate the updating direction of contours.

In order to calculate the probability of being blood for pixels inside myocardium wall, we apply extra 2D U-Nets to segment pure muscle and blood regions, respectively, and extract the features (length 64) from the second last layers of networks. A logistic regression model is learned with pixels from regions of pure muscle/blood and their extracted features. The pixels of pure blood are labeled as 0, and those of pure muscle are labeled as 1. The probabilities of the other pixels inside the ventricle would be computed using the learned model. For cases where the apparent blood spaces in the transitional zone seem to entirely disappear, we are able to estimate where the corresponding transitional zone was moving to at the times (in earlier systole and later diastole) when we could more reliably see and track it. While this is inherently uncertain, it should still be better than purely relying on image intensity, which can result in significant errors in the estimation of the end-systolic volumes, due to the effective initial assignment of the "wall" contour to the solid wall-transitional zone boundary at end-diastole, but then moving it inward toward the transitional zone-blood boundary at end-systole.

4. EXPERIMENTS

4.1. Dataset and Myocardium Segmentation

We adopted a cardiac MRI dataset consisting of 22 normal volunteers and 3 patients with cardiac dyssynchrony disease. All LV contours of these SAX images over different spatial locations and different cardiac phases are manually annotated by experts. In-plane resolution of images ranged from 1.17 mm to 1.43 mm, and size varied from 224×204 pixels to 240×198 pixels. Each cardiac cycle contains 25 phases. As

for [16, 17], we conducted our evaluation procedure in the following way. We run the 5-fold cross validation, and make sure that each of the 25 subjects (containing both normal subjects and patient, around 4000 2D slices with manual annotation) in the test set exactly once.

In order to boost the robustness of our model, we used data augmentation by 90 degree rotations and mirroring. All the images were scaled to resolution of $1.25\ mm$ and padded with 0s to gain the same image size. The filter numbers of 3D U-Nets was reduced by half in order to fit the data and reduce GPU memory consumption. We adopted the soft Dice loss [18] during training models. For both tasks, we used ADAM optimizer with a fixed learning rate of 0.001 and weight decay of 2.0^{-5}. The results reported below were gained after training for 30 epochs with batch size 32 for 2D U-Net and 15 epochs with batch size 16 for 3D U-Net. Training one model takes 12 hours on a single NVIDIA K80 GPU, and inference takes about 1 second for a complete cardiac cycle.

Table 1. Evaluation of endo- and epicardium segmentation, A, B, C represents 2D U-Net, 2D-3D U-Net (Ours), and 2D-3D U-Net + Deformable Model (Ours), respectively.

	Method	Dice	Jaccard	APD (mm)
Endo.	FCN8	0.855 ± 0.218	0.759 ± 0.195	3.845 ± 4.950
	FCN16	0.848 ± 0.212	0.738 ± 0.214	3.134 ± 5.595
	FCN32	0.639 ± 0.274	0.475 ± 0.238	8.094 ± 7.534
	[6]	0.850 ± 0.204	0.742 ± 0.219	6.278 ± 17.801
	[8]	0.859 ± 0.203	0.758 ± 0.213	2.470 ± 3.967
	A	0.864 ± 0.180	0.764 ± 0.196	3.799 ± 8.930
	B	0.886 ± 0.035	0.821 ± 0.038	2.768 ± 1.946
	C	$\mathbf{0.902} \pm 0.035$	$\mathbf{0.847} \pm 0.037$	$\mathbf{1.647} \pm 0.609$
Epi.	FCN8	0.877 ± 0.177	0.731 ± 0.276	1.964 ± 4.986
	FCN16	0.883 ± 0.244	0.763 ± 0.2553	2.994 ± 4.213
	FCN32	0.833 ± 0.165	0.716 ± 0.172	4.318 ± 3.619
	[6]	0.857 ± 0.194	0.746 ± 0.204	3.050 ± 7.291
	[8]	0.821 ± 0.231	0.712 ± 0.243	4.071 ± 4.976
	A	0.886 ± 0.168	0.797 ± 0.187	2.821 ± 4.922
	B	0.895 ± 0.327	0.839 ± 0.035	2.387 ± 0.839
	C	$\mathbf{0.905} \pm 0.037$	$\mathbf{0.855} \pm 0.039$	$\mathbf{2.094} \pm 0.535$

In order to make the fair comparison, we re-implemented the state-of-the-art methods[6, 6], and ran them on the same dataset with our proposed method. For endocardium, the average Dice' score of the original 2D U-Net is 0.864 better than the FCN8 (0.855), FCN16 (0.848) and FCN32 (0.639). Our proposed method gains much better results, i.e., 2D-3D U-Net is 0.886 and 2D-3D U-Net + Deformable Model is 0.902 which is the highest one to date. In addition, our proposed methods outperform others in terms of the Jaccard index and APD. To further evaluate our method, we also calculated the percentage of good contours (a perentage of the predicted contours, out of all contours, that have APD less than 5 mm from the gold standard [17]). Among all the segmentation results, Our best model 2D-3D U-Net + Deformable had 97.5% good contours. In the epicardium experiment, both our proposed model 2D-3D U-Net (0.895) and 2D-3D U-Net + Deformable Model (0.905) achieved very high Dice's score as well. Also, it's obvious that the average Jaccard index of the 2D-3D U-Net + Deformable Model (0.855) is higher than that of both 2D U-Net (0.797) and 2D-3D U-Net (0.839), and the good contour percentage of the 2D-3D

U-Net + Deformable Model is 93.3%. Some good segmentation examples of our 2D-3D U-Net + Deformable Model are shown in Fig. 4. As the 2D-3D U-Net utilizes three dimensional information, it outperforms the traditional 2D U-Net in all three evaluation metrics. Moreover, the 2D-3D U-Net + Deformable Model uses the temporal information to further refine the result. Overall, our methods generate results very close to the gold standard compared with other methods for both endo- and epicardium.
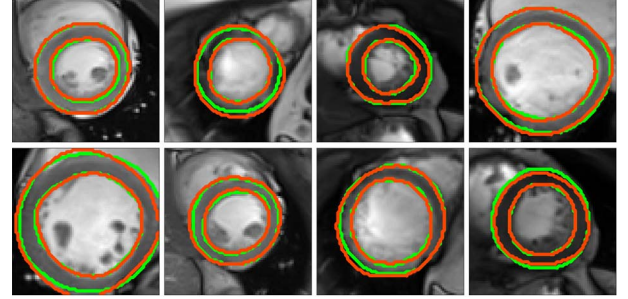

Fig. 4. Sample results of proposed methods. Green contours are the gold standard, and red contours are the prediction.

4.2. Partial Blood Estimation

Since it is hard to define a gold standard for the probability maps of partial blood estimation, we only evaluated our results visually, as shown in Fig. 5. We can see that the region where muscle and blood mix are assigned with a probability value between 0 and 1, not purely affected by the local appearance.
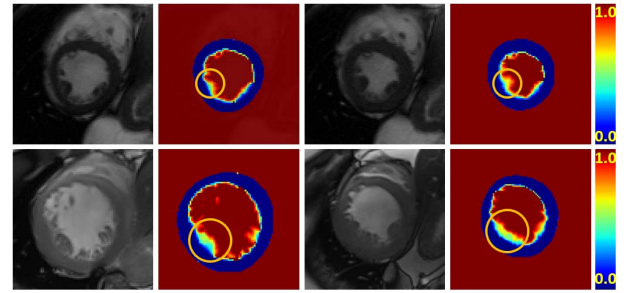

Fig. 5. Sample results of partial blood segmentation. Left: original image, right: probability map of blood. The yellow circles denote the transition zone.

5. CONCLUSIONS

We have proposed a robust and efficient method for short-axis cardiac MRI segmentation, using both deep neural networks and multi-component deformable models. The evaluation results demonstrated that our method outperforms other methods. Future directions could include applying the proposed approach for the long-axis cardiac MRI or the segmentation of other heart chambers.

6. REFERENCES

[1] Caroline Petitjean and Jean-Nicolas Dacher, "A review of segmentation methods in short axis cardiac mr images," *Medical image analysis*, vol. 15, no. 2, pp. 169–184, 2011.

[2] Nikos Paragios, "A variational approach for the segmentation of the left ventricle in cardiac image analysis," *International Journal of Computer Vision*, vol. 50, no. 3, pp. 345–362, 2002.

[3] Marie-Pierre Jolly, "Automatic segmentation of the left ventricle in cardiac mr and ct images," *International Journal of Computer Vision*, vol. 70, no. 2, pp. 151–163, 2006.

[4] Yun Zhu, Xenophon Papademetris, Albert J Sinusas, and James S Duncan, "Segmentation of the left ventricle from cardiac mr images using a subject-specific dynamical model," *IEEE Transactions on Medical Imaging*, vol. 29, no. 3, pp. 669–687, 2010.

[5] MR Avendi, Arash Kheradvar, and Hamid 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, 2016.

[6] Phi Vu Tran, "A fully convolutional neural network for cardiac segmentation in short-axis mri," *arXiv preprint arXiv:1604.00494*, 2016.

[7] Dong Yang, Pengxiang Wu, Chaowei Tan, Kilian M Pohl, Leon Axel, and Dimitris Metaxas, "3d motion modeling and reconstruction of left ventricle wall in cardiac mri," in *International Conference on Functional Imaging and Modeling of the Heart*. Springer, 2017, pp. 481–492.

[8] Matthew Sinclair, Wenjia Bai, Esther Puyol-Antón, Ozan Oktay, Daniel Rueckert, and Andrew P King, "Fully automated segmentation-based respiratory motion correction of multiplanar cardiac magnetic resonance images for large-scale datasets," in *International Conference on Medical Image Computing and Computer-Assisted Intervention*. Springer, 2017, pp. 332–340.

[9] Noel CF Codella, Hae Yeoun Lee, David S Fieno, Debbie W Chen, Sandra Hurtado-Rua, Minisha Kochar, John Paul Finn, Robert Judd, Parag Goyal, Jesse Schenendorf, et al., "Improved left ventricular mass quantification with partial voxel interpolation," *Circulation: Cardiovascular Imaging*, vol. 5, no. 1, pp. 137–146, 2012.

[10] Noel CF Codella, Jonathan W Weinsaft, Matthew D Cham, Matthew Janik, Martin R Prince, and Yi Wang, "Left ventricle: automated segmentation by using myocardial effusion threshold reduction and intravoxel computation at mr imaging," *Radiology*, vol. 248, no. 3, pp. 1004–1012, 2008.

[11] Olaf Ronneberger, Philipp Fischer, and Thomas Brox, "U-net: Convolutional networks for biomedical image segmentation," in *International Conference on Medical Image Computing and Computer-Assisted Intervention*. Springer, 2015, pp. 234–241.

[12] Özgün Çiçek, Ahmed Abdulkadir, Soeren S Lienkamp, Thomas Brox, and Olaf Ronneberger, "3d u-net: learning dense volumetric segmentation from sparse annotation," in *International Conference on Medical Image Computing and Computer-Assisted Intervention*. Springer, 2016, pp. 424–432.

[13] Jonas Gehring, Michael Auli, David Grangier, Denis Yarats, and Yann N Dauphin, "Convolutional sequence to sequence learning," *arXiv preprint arXiv:1705.03122*, 2017.

[14] Du Tran, Lubomir Bourdev, Rob Fergus, Lorenzo Torresani, and Manohar Paluri, "Deep end2end voxel2voxel prediction," in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition Workshops*, 2016, pp. 17–24.

[15] Xiaomeng Li, Hao Chen, Xiaojuan Qi, Qi Dou, Chi-Wing Fu, and Pheng Ann Heng, "H-denseunet: Hybrid densely connected unet for liver and liver tumor segmentation from ct volumes," *arXiv preprint arXiv:1709.07330*, 2017.

[16] Avan Suinesiaputra, Brett R Cowan, Ahmed O Al-Agamy, Mustafa A Elattar, Nicholas Ayache, Ahmed S Fahmy, Ayman M Khalifa, Pau Medrano-Gracia, Marie-Pierre Jolly, Alan H Kadish, et al., "A collaborative resource to build consensus for automated left ventricular segmentation of cardiac mr images," *Medical image analysis*, vol. 18, no. 1, pp. 50–62, 2014.

[17] P Radau, Y Lu, K Connelly, G Paul, A Dick, and G Wright, "Evaluation framework for algorithms segmenting short axis cardiac mri," *The MIDAS Journal-Cardiac MR Left Ventricle Segmentation Challenge*, vol. 49, 2009.

[18] Fausto Milletari, Nassir Navab, and Seyed-Ahmad Ahmadi, "V-net: Fully convolutional neural networks for volumetric medical image segmentation," in *3D Vision (3DV), 2016 Fourth International Conference on*. IEEE, 2016, pp. 565–571.