# A Fully Convolutional Neural Network for Cardiac Segmentation in Short-Axis MRI

**Phi Vu Tran**

Strategic Innovation Group
Booz | Allen | Hamilton
McLean, VA USA

tran_phi@bah.com

**Abstract:** Automated cardiac segmentation from magnetic resonance imaging datasets is an essential step in the timely diagnosis and management of cardiac pathologies. We propose to tackle the problem of automated left and right ventricle segmentation through the application of a deep fully convolutional neural network architecture. Our model is efficiently trained end-to-end in a single learning stage from whole-image inputs and ground truths to make inference at every pixel. To our knowledge, this is the first application of a fully convolutional neural network architecture for pixel-wise labeling in cardiac magnetic resonance imaging. Numerical experiments demonstrate that our model is robust to outperform previous fully automated methods across multiple evaluation measures on a range of cardiac datasets. Moreover, our model is fast and can leverage commodity compute resources such as the graphics processing unit to enable state-of-the-art cardiac segmentation at massive scales. The models and code are available at `https://github.com/vuptran/cardiac-segmentation`.

*Keywords*: convolutional neural networks; cardiac segmentation

# 1 Introduction

Cardiovascular diseases are the number one cause of death globally, according to the World Health Organization[*] . Management of cardiac pathologies typically relies on numerous cardiac imaging modalities, which include echocardiogram, computerized tomography, and magnetic resonance imaging (MRI). The current gold standard is to leverage non-invasive cine MRI to quantitatively analyze global and regional cardiac function through the derivation of clinical parameters such as ventricular volume, stroke volume, ejection fraction, and myocardial mass. Calculation of these parameters depends upon accurate manual delineation of endocardial and epicardial contours of the left ventricle (LV) and right ventricle (RV) in short-axis stacks. Manual delineation is a time-consuming and tedious task that is also prone to high intra- and inter-observer variability (Petitjean and Dacher, 2011; Miller et al., 2013; Tavakoli and Amini, 2013; Suinesiaputra et al., 2014). Thus, there exists a need for a fast, accurate, reproducible, and fully automated cardiac segmentation method to help facilitate the diagnosis of cardiovascular diseases.

There are a number of open technical challenges in automated LV and RV segmentation (Petitjean and Dacher, 2011; Tavakoli and Amini, 2013; Queirós et al., 2014):

- The overlap of pixel intensity distributions between cardiac objects and surrounding background structures;
- The shape variability of the endocardial and epicardial contours across slices and phases;
- Extreme imbalance in the number of pixels belonging to object class versus background;
- Fuzzy boundary and edge information, especially in basal and apical slices;
- Variability in cine MRI from different institutions, scanners, and populations;
- Inherent noise associated with cine MRI.

Although research over the past decade has addressed some of the above technical difficulties to achieve incremental progress on automated ventricle segmentation from short-axis cine MRI, the resulting automated segmentation contours still need to be significantly improved in order to be useable in the clinical setting (Petitjean and Dacher, 2011). Furthermore, the evaluation of previous research has been limited in scope, on small benchmark datasets that may not represent the real-world variability in image quality and cardiac anatomical and functional characteristics across sites, institutions, scanners, and populations. In addition, previously proposed methods require some *a priori* knowledge about the cardiac ventricles in order

[*] http://www.who.int/cardiovascular_diseases/en/. Accessed February 8 2016.

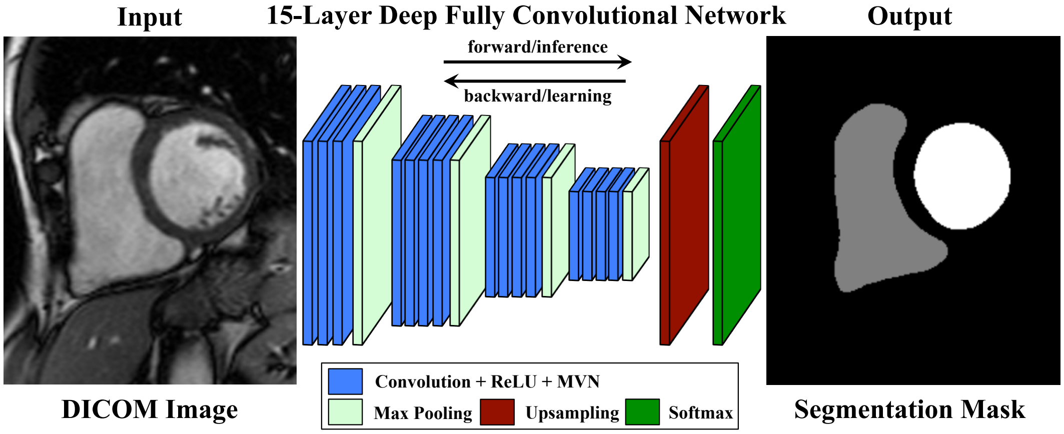

**Figure 1.** A schematic of our proposed fully convolutional neural network architecture. Acronyms: ReLU – Rectified Linear Unit; MVN – Mean-Variance Normalization.

to increase their accuracy and robustness (Petitjean and Dacher, 2011). For semi-automated approaches, direct user interaction is a form of *a priori* knowledge. For fully automated methods, *a priori* information includes hand-engineered features about the spatial relationships of the LV and RV objects and their surrounding structures, knowledge of the heart biomechanics, or anatomical assumptions about the statistical shapes of the objects (e.g., circular geometry of the LV, and complex crescent shape of the RV). Such assumptions about the LV and RV objects, through either weak or strong priors, may contribute to the propensity of previous methods to overfit on a particular training dataset.

Our contribution in this paper is a comprehensive evaluation of a convolutional neural network (CNN) architecture on multiple benchmark MRI datasets consisting of the left and right ventricle. The basic components of a CNN architecture include trainable filters that can automatically learn intricate features and concepts from a training set in a supervised manner, without the need for feature engineering or to hard-code *a priori* knowledge. CNNs are also amenable to transfer learning (Donahue et al., 2013; Zeiler and Fergus, 2014; Oquab et al., 2014; Yosinski et al., 2014; Razavian et al., 2014), a task that proves to be valuable in the absence of abundant training data. We show that the creative application of a CNN variant, the *fully convolutional neural network* (FCN), achieves state-of-the-art semantic segmentation in short-axis cardiac MRI acquired at multiple sites and from different scanners. The proposed FCN architecture is efficiently trained end-to-end on a graphics processing unit (GPU) in a single learning stage to make inference at every pixel, a task commonly known as pixel-wise labeling or per-pixel classification. At test time, the model independently segments each image in milliseconds, so it can be deployed in parallel on clusters of CPUs, GPUs, or both for scalable and accurate ventricle segmentation. To our knowledge, this is the first application of a CNN architecture for pixel-wise labeling in cardiac MRI.

The remainder of this paper is structured as follows: Section 2 briefly surveys previous research on LV and RV segmentation and fully convolutional neural networks. Section 3 presents an experimental framework to demonstrate and evaluate the efficacy of our FCN model on a range of publicly available benchmark cardiac MRI datasets. Section 4 reports results and analysis of the evaluation. Finally, Section 5 concludes with a summary and parting remarks.

## 2 Previous Work

### 2.1 Left Ventricle Segmentation

The task of delineating the left ventricle endocardium and epicardium from cine MRI throughout the cardiac cycle has received much research focus and attention over the past decade. Two grand challenges, Medical Image Computing and Computer Assisted Intervention (MICCAI) 2009 LV Segmentation Challenge (Radau et al., 2009) and Statistical Atlases and Computational Modeling of the Heart (STACOM) 2011 LV Segmentation Challenge (Suinesiaputra et al., 2014), have emerged with the goal of advancing the state of the art in automated LV segmentation. To facilitate research and development in this arena, the challenges provide benchmark datasets that come with expert ground truth contours and standard evaluation measures to assess automated segmentation performance. Petitjean and Dacher (2011) provide a comprehensive survey of previous methods for LV segmentation that include multi-level Otsu thresholding, deformable models and level sets, graph cuts, knowledge-based approaches such as active and appearance shape models, and atlas-based methods. Although these methods have achieved limited success on small benchmark LV datasets, they suffer from low robustness, low accuracy, and limited capacity to generalize over subjects with heart conditions outside of the training set.

More recently, Ngo and Carneiro (2013) couple Restricted Boltzmann Machines and a level set method to produce competitive results on a small benchmark LV dataset (Radau et al., 2009). However, their method is semi-automated that requires user input. Queirós et al. (2014) propose a novel automated 3D+time LV segmentation framework that combines automated 2D and 3D segmentation with contour propagation to yield accurate endocardial and epicardial contours on the same LV dataset. Avendi et al. (2015) integrate recent advances in machine learning such as stacked autoencoders and convolutional neural networks with a deformable model to establish a new state of the art on LV endocardium segmentation. The main limitation of the recent methods is that they are multi-stage approaches that require manual offline training and extensive hyper-parameter tuning, which can be cumbersome. Furthermore, the evaluation of the methods by Ngo and Carneiro (2013) and Avendi et al. (2015) is limited to only LV endocardial contours. None of these methods is evaluated for the task of automated right ventricle segmentation.

## 2.2 Right Ventricle Segmentation

The task of delineating the right ventricle endocardium and epicardium from short-axis cine MRI at various phases of the cardiac cycle shares similar goals, clinical motivations, and inherent technical difficulties as its LV counterpart. However, it does not receive as much research attention, partly because RV segmentation algorithms have never had access to a common database with expert ground truth contours. In an effort to advance the research and development of RV segmentation towards clinical applications, the MICCAI 2012 Right Ventricle Segmentation Challenge proposes a benchmark MRI dataset that comes with expert ground truth segmentation contours and standard evaluation measures, following the formats of prior LV segmentation competitions. Petitjean et al. (2015) survey the automated and semi-automated approaches presented by the seven challenger teams that include three atlas-based methods, two prior-based methods, and two prior-free, image-driven methods that make use of the temporal dimension of the data. This competition highlights the current interest in methods based on the multi-atlas segmentation framework (Rohlfing et al., 2004; Klein et al., 2005; Heckemann et al., 2006), which is becoming one of the most widely used and successful image segmentation techniques in medical imaging applications (Iglesias and Sabuncu, 2015).

Although the methods presented in the MICCAI 2012 challenge achieve reasonable segmentation accuracy, there is still much room left for improvement, especially if the methods are to be utilized in the clinical setting. The main limitation of previous methods based on statistical shape modeling, feature engineering, and multi-atlas registration is that they tend to overfit on one particular dataset and may not generalize well to other datasets nor are amenable to transfer learning.

## 2.3 Fully Convolutional Neural Networks

Convolutional neural networks (CNNs) continue to achieve record-breaking accuracy performances on many visual recognition benchmarks across research domains. CNNs are supervised models trained end-to-end to learn hierarchies of features automatically–without resorting to sometimes complicated input preprocessing, output postprocessing, and feature engineering schemes–yielding robust classification and regression performances. Recent successes of deep CNN architectures like AlexNet (Krizhevsky et al., 2012), VGGNet (Simonyan and Zisserman, 2015), GoogLeNet (Ioffe and Szegedy, 2015), and ResNet (He et al., 2015) have made CNNs the *de facto* standard for whole-image classification. In addition, high-performance deep CNNs have been adapted to advance the state of the art on other visual recognition tasks such as bounding box object detection (Girshick et al., 2014; Girshick, 2015; Ren et al., 2016) and semantic segmentation (Long et al., 2015; Liu et al., 2015; Zheng et al., 2015).

A standard deep CNN architecture for whole-image classification typically consists of convolution layer, nonlinear activation function, pooling layer, and fully connected layer as basic building blocks. Long et al. (2015) adapt and extend deep classification architectures by removing the fully connected layers and introducing fractional convolution layers to learn per-pixel labels end-to-end from whole-image inputs and corresponding whole-image ground truths. Long et al. (2015) describe their fractional convolution layer as useful for learning (nonlinear) upsampling filters in order to map or connect coarse outputs to the dense pixel space. Their key to success is to leverage large-scale image classification (Deng et al., 2009) as supervised pre-training, and fine-tune *fully convolutionally* via transfer learning. Others expand upon the FCN idea by adding global context (Liu et al., 2015) and coupling Conditional Random Field learning (Zheng et al., 2015) to push the performance boundaries in semantic segmentation.

## 3 Experimental Framework

We propose to tackle the problem of automated LV and RV segmentation through the application of an FCN architecture. Numerical experiments demonstrate that our FCN model is robust to outperform previous methods across multiple evaluation measures on a range of cardiac MRI datasets. All MRI datasets are in anonymized DICOM format.

### 3.1 Datasets

**Sunnybrook Cardiac Data (Radau et al., 2009):** The Sunnybrook dataset comprises cine MRI from 45 patients, or cases, having a mix of cardiac conditions: healthy, hypertrophy, heart failure with infarction, and heart failure without infarction. Expert manual segmentation contours for the endocardium, epicardium, and papillary muscles are provided for basal through apical slices at both end-diastole (ED) and end-systole (ES) phases. This dataset was made available as part of the MICCAI 2009 challenge on automated LV segmentation from short-axis cardiac MRI. The Sunnybrook dataset is available through the Cardiac Atlas Project (CAP)[†] with a public domain license.

The Sunnybrook dataset is divided into three disjoint sets of 15 cases each: training, validation, and online. Ground truth contours are provided for training, validation, and online sets. We use the training set to train an FCN model for LV endocardium and epicardium segmentation, and evaluate model performance on the validation and online sets. We do not investigate the segmentation of the papillary muscles because few researchers have done so in the past and therefore it is hard to compare results. It is important to note that we take great care to perform model selection and hyper-parameter tuning on a *development* subset derived from randomly splitting the training set into 0.9/0.1 folds. This procedure is standard protocol to ensure that we do not peek into the validation and online sets that can result in

[†] http://www.cardiacatlas.org/studies/sunnybrook-cardiac-data/

overfitting, and that we stay consistent with how the challenge was conducted.

**Left Ventricle Segmentation Challenge (Suinesiaputra et al., 2014):** This dataset, denoted here as LVSC, was made publicly available as part of the STACOM 2011 challenge on automated LV myocardium segmentation from short-axis cine MRI. The dataset is derived from the DETERMINE cohort (Kadish et al., 2009) consisting of 200 patients with coronary artery disease and myocardial infarction. The LVSC dataset comes with expert-guided semi-automated segmentation contours for the myocardium, a region composed of pixels inside the epicardium and outside the endocardium, derived from the Guide-Point Modeling technique (Li et al., 2010). This approach involves expert input to refine the segmentation contours by interactively positioning a small number of guide points on a subset of slices and frames. The contours are available for basal through apical slices at both ED and ES phases. The LVSC dataset can be downloaded from the LV Segmentation Challenge website via the CAP[‡].

The LVSC dataset is divided into two disjoint sets of 100 cases each: testing and validation. We use the testing set with the provided expert-guided contours to train an FCN model to segment the LV myocardium, and evaluate model performance on the validation set. We split the testing set into 0.95/0.05 training/development folds for experimentation, model selection, hyper-parameter tuning. There are no absolute ground truth contours for the validation set. Instead, the challenge organizers estimate reference consensus images from a set of five segmentation algorithms (two fully automated and three semi-automated requiring user input) using the STAPLE (Simultaneous Truth and Performance Level Estimation) method (Warfield et al., 2004). The idea is to establish a large community resource of ground truth images based on common data for the development, validation, and benchmarking of LV segmentation algorithms (Suinesiaputra et al., 2014). Reference consensus images are not provided for the validation set, so we submit our predicted myocardial contours to the challenge organizers for independent evaluation.

**Right Ventricle Segmentation Challenge (Petitjean et al., 2015):** This dataset, denoted here as RVSC, was provided as part of the MICCAI 2012 challenge on automated RV endocardium and epicardium segmentation from short-axis cine MRI. The dataset comprises 48 cases having various cardiac pathologies. Expert manual endocardial and epicardial contours are provided for basal through apical slices at both ED and ES phases. The RVSC dataset is available for download from the LITIS lab at the University of Rouen[§].

The RVSC dataset is divided into three disjoint sets of 16 cases each: training, test1, and test2. Expert manual contours are provided for the training set only. We split the training set into 0.9/0.1 training/development subsets for experimentation,

[‡] `http://www.cardiacatlas.org/challenges/lv-segmentation-challenge/`

[§] `http://www.litislab.fr/?projet=1rvsc` model selection, and hyper-parameter tuning. At test time, we submit our predicted RV endocardial and epicardial contours for test1 and test2 sets to the challenge organizers for independent evaluation.

### 3.2 Data Preparation and Augmentation

We observe that the heart cavity containing both the left and right ventricle appears roughly at the center of each short-axis slice. We proceed to take the square center crop of each image to define our region of interest (ROI). The size of the ROI can have an impact on the accuracy performance of the FCN model. We choose a multi-resolution approach to crop the ROI at multiple sizes that wholly contain the ventricles. Multi-scale cropping provides the following benefits:

- Augment the training set by providing multiple views of the same image at multiple resolutions;
- Capture the ROI while providing a “zooming” effect for enhanced feature learning;
- Mitigate class imbalance by removing unnecessary background pixels;
- Accelerate computations via the reduction of input spatial dimensions.

The 16-bit MRI datasets have a wide range of pixel intensities that directly influence the accuracy performance of automated segmentation models, especially if the acquired images come from multiple sites using different scanner types or manufacturers. We normalize the pixel intensity distribution of each input image by subtracting its mean and dividing the resulting difference by its standard deviation. The normalized output is an image with pixel values having zero mean and unit variance. Mean-variance normalization (MVN) is a simple yet effective technique that significantly enhances the learning capacity of our FCN model during training and segmentation performance during testing across datasets. No further preprocessing is done on the input image pixels. We also perform affine transformations (rotation, vertical flipping, and horizontal flipping) to augment the training set in an effort to mitigate overfitting and further improve model generalization. Multi-scale center cropping and affine transformations artificially inflate the training set by 12-fold, although the resulting augmented dataset is highly correlated. Table 1 summarizes our data augmentation strategy for each dataset.

### 3.3 FCN Architecture

Figure 1 illustrates our proposed FCN architecture, which is selected via cross-validation on a development set. The FCN architecture comprises 15 stacked convolution layers and three layers of overlapping, two-pixel strided max pooling. Each

**Table 1.** Data augmentation strategy for each dataset during training. At test time, we standardize input images via center cropping and mean-variance normalization. The tuple $(h, w)$ denotes the height and width of the input image, respectively.

| | **Training** | | | | **Testing** |
|---|---|---|---|---|---|
| **Dataset** | **Center Crop** | **Rotation** | **Vertical Flip** | **Horizontal Flip** | **Center Crop** |
| Sunnybrook | $dim \in [100, 120]$ | $k \times 90, k \in [1, 2, 3]$ | Yes | Yes | $dim = 100$ |
| LVSC | $dim = \text{int}(\min(h, w) \times 0.6)$ | No | No | No | $dim = \text{int}(\min(h, w) \times 0.6)$ |
| RVSC | $dim \in [200, 216]$ | $k \times 90, k \in [1, 2, 3]$ | Yes | Yes | $dim = 200$ |

convolution layer is followed by Rectified Linear Unit (ReLU) activation function (Nair and Hinton, 2010) and MVN operation. The architecture has roughly 11 million parameters to be estimated. Such a high-dimensional model is prone to overfit on the relatively small MRI datasets under consideration; we take great care to mitigate overfitting through data augmentation during data preparation, and dropout and regularization during training.

We employ the "skip" architecture of Long et al. (2015) to combine coarse semantic information at deep layers and fine appearance information at shallow layers to learn filters for output upsampling. The end result is a dense heatmap predicting class membership of each pixel in the input image. For time benchmarking purposes, the FCN model takes an average of 61 milliseconds to segment one image of $256 \times 256$ pixels using a single NVIDIA GeForce GTX TITAN X GPU. We apply and evaluate a single instantiation of this FCN architecture on all MRI datasets under consideration.

## 3.4 Training Protocol

We leverage the Caffe deep learning framework (Jia et al., 2014) for the design, implementation, and experimentation of our deep FCN architecture. We employ stochastic gradient descent with momentum of 0.9 to minimize the multinomial logistic loss on per-pixel softmax probabilities from whole-image inputs and ground truths. We randomly initialize parameter weights according to the "Xavier" scheme (Glorot and Bengio, 2010). We further combat the adverse effects of overfitting by using dropout ratio of 0.5 (Srivastava et al., 2014) and $L_2$ weight decay regularization of 0.0005. We train for 10 epochs, or passes over the training set, and anneal the learning rate according to the polynomial decay of the form: $\text{base\_lr} \times \left(1 - \frac{\text{iter}}{\text{max\_iter}}\right)^{\text{power}}$, where $\text{base\_lr} = 0.01$ is the initial learning rate, iter is the current iteration, max_iter is the dataset-specific maximum number of iterations approximately equal to 10 epochs, and $\text{power} = 0.5$ controls the rate of decay.

## 3.5 Transfer Learning

We also explore the benefits of transfer learning in training deep FCN models with limited data. We first train an FCN model using "Xavier" random initialization on the relatively large LVSC testing set of roughly 22,000 DICOM images that come with expert-guided semi-automated contours to obtain a set of learned convolution filters. We call the pre-trained FCN model the source model, the LVSC testing set the source dataset, and the task of estimating LV contours the source task. In transfer learning, we initialize a second FCN model (the target model) with the learned weights from the source model by copying or transferring from selected convolution and upsampling layers. The remaining layers of the target model are then randomly initialized and trained toward a target task using a target dataset via supervised fine-tuning.

We experiment with transferring the learned feature representation from the source task of LV segmentation on the LVSC dataset to the target task of LV segmentation on the Sunnybrook dataset and to the target task of RV segmentation on the RVSC dataset. The transferred weights serve as supervised pre-training that enable training a large target model on small target datasets without severe overfitting. In transfer learning, we set the initial learning rate to be small, base_lr = 0.001, in order to refine the update of the learned weights during backpropagation. Transfer learning and supervised fine-tuning offer the following benefits:

- Domain adaptation – transfer learning allows the learned model on the source task to be adapted to a different, but related, target task. For example, we can enable a source model that learns LV contours to train on estimating RV contours. Yosinski et al. (2014) document that the transferability of features decreases as the distance between the source task and target task increases, but that transferring features even from distant tasks can be better than using random features;

- The source and target datasets need not be from the same distribution. For example, the Sunnybrook Cardiac Dataset follows a different distribution than the LVSC dataset because they were acquired from different institutions, even though both datasets represent the LV object;

- Better convergence and accuracy performance even with limited training data.

# 4 Empirical Evaluation

## 4.1 Metrics

Let $a$ and $m$ be the predicted (automated) and ground truth (manual) contours delineating the object class in short-axis MRI, respectively. Let $A$ and $M$ be the corresponding areas enclosed by contours $a$ and $m$, respectively. The following evaluation metrics are used to assess the accuracy of automated segmentation methods using the ground truth as reference. Different challenges use different measures for the respective dataset; we provide an overview of the main metrics reported in the literature for comparative purposes.

**Sunnybrook Cardiac Dataset**:

- Average perpendicular distance (APD) measures the distance between contours $a$ and $m$, averaged over all contour points. A high value implies that the two contours *do not* closely match (Radau et al., 2009). APD is computed in millimeter with spatial resolution obtained from the `PixelSpacing` DICOM field.

- The Dice index (Dice, 1945) is a measure of overlap or similarity between two contour areas, and is defined as:

  $$\mathcal{D}(A, M) = 2\frac{A \cap M}{A + M}.$$

  The Dice index varies from zero (total mismatch) to unity (perfect match).

- Percentage of good contours is a fraction of the predicted contours, out of the total number of contours, that have APD less than 5 millimeters away from the ground truth contours (Radau et al., 2009).

**LVSC Dataset**:

- Sensitivity ($p$), specificity ($q$), positive predictive value ($PPV$), and negative predictive value ($NPV$) are defined as:

  $$p = \frac{T_1}{N_1}, \quad q = \frac{T_0}{N_0}, \quad PPV = \frac{T_1}{T_1 + F_1}, \quad NPV = \frac{T_0}{T_0 + F_0},$$

  where $T_1$ and $T_0$ are the number of correctly predicted pixels as belonging to the object and background class, while $F_1$ and $F_0$ are the number of misclassified pixels as object and background, respectively. The total number of object and background pixels are denoted by $N_1$ and $N_0$, respectively (Suinesiaputra et al., 2014).

- The Jaccard index (Jaccard, 1912) is a measure of overlap or similarity between two contour areas, and is defined as:

  $$\mathcal{J}(A, M) = \frac{A \cap M}{A \cup M} = \frac{A \cap M}{A + M - (A \cap M)}$$

  Similar to the Dice index, the Jaccard index varies from zero to unity, with unity representing perfect correspondence with the ground truth.

**RVSC Dataset**:

- The Hausdorff distance is a symmetric measure of distance between two contours (Huttenlocher et al., 1993), and is defined as:

$$\mathcal{H}(a, m) = \max\left(\max_{i \in a}\left(\min_{j \in m} d(i, j)\right), \max_{j \in m}\left(\min_{i \in a} d(i, j)\right)\right),$$

  where $d(\cdot, \cdot)$ denotes Euclidean distance. Similar to APD, a high Hausdorff value implies that the two contours *do not* closely match. The Hausdorff distance is computed in millimeter with spatial resolution obtained from the `PixelSpacing` DICOM field.

- The Dice index $\mathcal{D}(A, M)$ as defined above.

### 4.2 Results and Analysis

Tables 2, 3, and 4 summarize our automated segmentation results and compare them to the previous state of the art. On the combined Sunnybrook validation and online sets, our FCN model achieves comparable Dice index with that of the method by Avendi et al. (2015) for automated LV endocardium segmentation. For all other evaluation measures, our model obtains the best scores across the board. Note that fine-tuning the FCN model does have an accuracy improvement over the same FCN model initialized with random features, a result that has been consistently corroborated in many previous studies. At test time, our model segments the endocardium and epicardium in both validation and online sets (a total of 830 images) in less than 25 seconds. Figure 2 illustrates example predicted endocardial and epicardial contours for the left ventricle using the Sunnybrook dataset.

For the task of predicting myocardial contours on the LVSC validation set, our FCN model achieves the best scores in three out of five metrics including the Jaccard index, specificity, and negative predictive value in comparison to previous fully automated methods. Note that we compare our segmentation results against previous methods based on Table 2 of Suinesiaputra et al. (2014) using the `CS*` consensus.

When compared against the expert-guided semi-automated method of Li et al. (2010), which was used to generate reference ground truth contours for the LVSC testing set, the FCN model performs significantly worse. It is important to note that the FCN model segments each DICOM image independently using the contextual cues of the image pixels as features, while the Guide-Point Modeling technique of Li et al. (2010) requires human expert input to refine and approve the segmentation results for all slices and for all frames. There are several difficult cases where our model cannot detect the presence of the LV object in apical/basal slices, mainly because they exhibit ambiguous or imperceptible object boundaries. These cases necessitate user intervention to improve segmentation, which is the reason why the Guide-Point Modeling technique performs so well. However, the main limitation of the guide-point approach is the slow processing time associated with user intervention, giving rise to the crux of the problem in scalability. In contrast, the FCN

**Table 2.** Comparison of LV endocardium and epicardium segmentation performance between our proposed FCN model and previous research using the Sunnybrook Cardiac Dataset. We distinguish performance of the FCN model through either transfer learning and supervised fine-tuning from the source LVSC dataset or Xavier random initialization. Number format: mean value (standard deviation).

| **Method** | #[¶] | **Dice Index** | | **APD[‖] (mm)** | | **Good Contours (%)** | |
|---|---|---|---|---|---|---|---|
| | | **Endo** | **Epi** | **Endo** | **Epi** | **Endo** | **Epi** |
| Our FCN model w/ finetune | 30 | 0.92 (0.03) | **0.96 (0.01)** | **1.73 (0.35)** | **1.65 (0.31)** | **98.48 (4.06)** | **99.17 (2.20)** |
| Our FCN model w/ Xavier init | 30 | 0.92 (0.03) | 0.95 (0.02) | 1.74 (0.43) | 1.69 (0.34) | 97.42 (5.86) | 98.00 (3.78) |
| (Avendi et al., 2015) | 30 | **0.94 (0.02)** | – | 1.81 (0.44) | – | 96.69 (5.7) | – |
| (Queirós et al., 2014) | 45 | 0.90 (0.05) | 0.94 (0.02) | 1.76 (0.45) | 1.80 (0.41) | 92.70 (9.5) | 95.40 (9.6) |
| (Ngo and Carneiro, 2013) | 45 | 0.90 (0.03) | – | 2.08 (0.40) | – | 97.91 (6.2) | – |
| (Hu et al., 2013) | 45 | 0.89 (0.03) | 0.94 (0.02) | 2.24 (0.40) | 2.19 (0.49) | 91.06 (9.4) | 91.21 (8.5) |
| (Liu et al., 2012) | 45 | 0.88 (0.03) | 0.94 (0.02) | 2.36 (0.39) | 2.19 (0.49) | 91.17 (8.5) | 90.78 (10.7) |
| (Huang et al., 2011) | 45 | 0.89 (0.04) | 0.93 (0.02) | 2.16 (0.46) | 2.22 (0.43) | 79.20 (19.0) | 83.90 (16.8) |
| (Constantinides et al., 2009) | 30 | 0.89 (0.04) | 0.92 (0.02) | 2.04 (0.47) | 2.35 (0.57) | 90.35 | 92.56 |
| (Jolly, 2009) | 30 | 0.88 (0.04) | 0.93 (0.02) | 2.26 (0.59) | 1.97 (0.48) | 95.62 (8.8) | 97.29 (5.8) |

[¶] Number of test cases: 30 – validation and online cases; 45 – training, validation, and online cases.

[‖] Average Perpendicular Distance.

**Table 3.** Comparison of LV myocardium segmentation performance between our proposed FCN model and previous research using the LVSC validation set based on the `CS*` consensus. Values are taken from Table 2 of Suinesiaputra et al. (2014). Number format: mean value (standard deviation).

| **Method** | **FA/SA[**]** | **Jaccard Index** | **Sensitivity** | **Specificity** | **PPV** | **NPV** |
|---|---|---|---|---|---|---|
| Our FCN model | FA | **0.74 (0.13)** | 0.83 (0.12) | **0.96 (0.03)** | 0.86 (0.10) | **0.95 (0.03)** |
| (Jolly et al., 2012) | FA | 0.69 (0.23) | 0.74 (0.23) | 0.96 (0.05) | **0.87 (0.16)** | 0.89 (0.09) |
| (Margeta et al., 2012) | FA | 0.43 (0.10) | **0.89 (0.17)** | 0.56 (0.15) | 0.50 (0.10) | 0.93 (0.09) |
| (Li et al., 2010) (Expert guided) | SA | **0.84 (0.17)** | **0.89 (0.13)** | **0.96 (0.06)** | **0.91 (0.13)** | **0.95 (0.06)** |
| (Fahmy et al., 2012) | SA | 0.74 (0.16) | 0.88 (0.15) | 0.91 (0.06) | 0.82 (0.12) | 0.94 (0.06) |
| (Ourselin et al., 2000) | SA | 0.64 (0.18) | 0.80 (0.17) | 0.86 (0.08) | 0.74 (0.15) | 0.90 (0.08) |

[**] Fully Automated / Semi-Automated

model is fully automated and scales to massive datasets; at test time, our model segments all 29,859 short-axis images, for 100 cases total, in the LVSC validation set in under 19 minutes.

We also outperform previous fully automated and semi-automated methods on the task of RV endocardium and epicardium segmentation across all evaluation metrics. At test time, our model predicts endocardial and epicardial contours for both test1 and test2 sets (a total of 1,028 images) in less than a minute. Figure 3 illustrates some example endocardium and epicardium segmentation results for the right ventricle using the RVSC dataset. Again, note that fine-tuning the FCN model results in a significant accuracy boost over the same FCN model initialized with random features, thus establishing a new state of the art for right ventricle segmentation.

**Table 4.** Comparison of RV endocardium and epicardium segmentation performance between our proposed FCN model and previous research using the RVSC dataset. We distinguish performance of the FCN model through either transfer learning and supervised fine-tuning from the source LVSC dataset or Xavier random initialization. Values are averaged over test1 and test2 sets in format: mean value (standard deviation).

| **Method** | **FA/SA**** | **Dice Index** | | **Hausdorff Dist (mm)** | |
|---|---|---|---|---|---|
| | | **Endo** | **Epi** | **Endo** | **Epi** |
| Our FCN model w/ finetune | FA | **0.84 (0.21)** | **0.86 (0.20)** | **8.86 (11.27)** | **9.33 (10.79)** |
| Our FCN model w/ Xavier init | FA | 0.80 (0.27) | 0.84 (0.24) | 11.41 (15.25) | 11.27 (15.04) |
| (Zuluaga et al., 2013) | FA | 0.76 (0.25) | 0.80 (0.22) | 11.51 (10.06) | 11.82 (9.38) |
| (Wang et al., 2012) | FA | 0.59 (0.34) | 0.63 (0.35) | 25.32 (22.66) | 24.43 (22.26) |
| (Ou et al., 2012) | FA | 0.58 (0.31) | 0.63 (0.27) | 19.12 (14.39) | 18.85 (13.47) |
| (Grosgeorge et al., 2013) | SA | **0.79 (0.18)** | **0.84 (0.12)** | **8.63 (4.54)** | **9.36 (4.58)** |
| (Bai et al., 2013) | SA | 0.77 (0.22) | 0.82 (0.16) | 9.52 (5.26) | 9.99 (5.18) |
| (Maier et al., 2012) | SA | 0.79 (0.22) | – | 10.47 (6.00) | – |
| (Nambakhsh et al., 2013) | SA | 0.58 (0.24) | – | 21.21 (9.71) | – |

** Fully Automated / Semi-Automated

Overall, the main limitation of the proposed FCN model lies in its inability to segment cardiac objects in difficult slices of the heart, especially at the apex. Petitjean et al. (2015) report that the accuracy of previous RV segmentation methods also depends on slice level. Their analysis reveals that error is most prominent in apical slices. For example, the Dice index for the endocardial contour decreases by 0.20 from base to apex. **This analysis is also consistent in left ventricle segmentation**, where our FCN model fails to detect the presence of the cardiac object at the apex in several cases. Figure 4 shows some examples of poor FCN segmentation on these difficult apical slices. Petitjean et al. (2015) suggest that the improvement of segmentation accuracy could be searched in apical slices, by emphasizing the model over the image content for these slices. Segmentation error on apical slices has minor impact on the volume computation, but it can be a limiting factor in other research fields such as the study of fiber structure (Petitjean et al., 2015).

# 5 Conclusion

In this paper, we demonstrated the utility and efficacy of a fully convolutional neural network architecture for semantic segmentation in cardiac MRI. We showed that a single FCN model can be trained end-to-end to learn intricate features useful for segmenting both the left *and* right ventricle. Comprehensive empirical evaluations revealed that our FCN model achieves state-of-the-art segmentation accuracy on multiple metrics and benchmark MRI datasets exhibiting real-world variability in image quality and cardiac anatomical and functional characteristics across sites, institutions, scanners, populations, and heart conditions. Moreover, the FCN model is fast, and can run on commodity compute resources such as the GPU to enable cardiac segmentation at massive scales.

The proposed FCN model can be further improved, in light of discovered limitations. The power of the FCN model lies in its capacity to learn millions of parameters on an abundance of training data. In order to improve segmentation accuracy of cardiac objects in difficult heart locations that exhibit ambiguous or imperceptible object boundaries such as apical and basal slices, the research community could dedicate effort in collecting more labeled or annotated examples at these locations. The demonstrated potential of the proposed FCN model is merely the tip of the iceberg. With more cardiac data to feed and train powerful, large-scale networks, FCN models could become the workhorse in advancing automated cardiac segmentation toward clinical applications with speed, accuracy, and reliability.

## Acknowledgement

The author thanks Alex Newton for some helpful code and gracious reviewers for constructive feedback on the paper.

## References

Avendi, M.R., Kheradvar A., Jafarkhani H. 2015. A combined deep-learning and deformable-model approach to fully automatic segmentation of the left ventricle in cardiac MRI. http://arxiv.org/abs/1512.07951.

Bai, W., Shi W., O'Regan D.P., Tong T., Wang H., Jamil-Copley S. et al. 2013. A probabilistic patch-based label fusion model for multi-atlas segmentation with registration refinement: Application to cardiac MR images. *IEEE Transactions on Medical Imaging* **32**(7) 1302–1315.

Constantinides, M., Chenoune Y., Kachenoura N., Roullot E. et al. 2009. Semi-automated cardiac segmentation on cine magnetic resonance images using GVF-Snake deformable models. *The MIDAS Journal – Cardiac MR Left Ventricle Segmentation Challenge* .

Deng, J., Dong W., Socher R., Li L.J. 2009. ImageNet: A large-scale hierarchical image database. *IEEE Conference on Computer Vision and Pattern Recognition*. IEEE, 248–255. http://image-net.org.

Dice, L.R. 1945. Measures of the amount of ecologic association between species. *Ecology* **26**(3) 297–302.

Donahue, J., Jia Y., Vinyals O., Hoffman J., Zhang N. et al. 2013. DeCAF: A deep convolutional activation feature for generic visual recognition. http://arxiv.org/abs/1310.1531.

Fahmy, A., Al-Agamy A., Khalifa A. 2012. Myocardial segmentation using contour-constrained optical flow tracking. *Statistical Atlases and Computational Models of the Heart. Imaging and Modeling Challenges*. Springer, 120–128.

Girshick, R. 2015. Fast R-CNN. http://arxiv.org/abs/1504.08083.

Girshick, R., Donahue J., Darrell T., Malik J. 2014. Rich feature hierarchies for accurate object detection and semantic segmentation. http://arxiv.org/abs/1311.2524.

Glorot, X., Y. Bengio. 2010. Understanding the difficulty of training deep feedforward neural networks. *13th International Conference on Artificial Intelligence and Statistics*.

Grosgeorge, D., Petitjean C., Dacher J.N., Ruan S. 2013. Graph cut segmentation with a statistical shape model in cardiac MRI. *Computer Vision and Image Understanding* **117**.

He, K., Zhang X., Ren S., Sun J. 2015. Deep residual learning for image recognition. `http://arxiv.org/abs/1512.03385`.

Heckemann, R.A., Hajnal J.V., Aljabar P., Rueckert D., Hammers A. 2006. Automatic anatomical brain MRI segmentation combining label propagation and decision fusion. *Neuroimage* **33** 115–126.

Hu, H., Liu H., Gao Z., Huang L. 2013. Hybrid segmentation of left ventricle in cardiac MRI using gaussian-mixture model and region restricted dynamic programming. *Magnetic Resonance Imaging* **31** 575–584.

Huang, S., Liu J., Lee L.C., Venkatesh S., Teo L., Au C., Nowinski W. 2011. An image-based comprehensive approach for automatic segmentation of left ventricle from cardiac short axis cine MR images. *Journal of Digital Imaging* **24** 598–608.

Huttenlocher, D.P., Klanderman G.A., Rucklidge W.J. 1993. Comparing images using the Hausdorff distance. *IEEE Transactions on Pattern Analysis and Machine Intelligence* **15**(9).

Iglesias, J.E., M.R. Sabuncu. 2015. Multi-atlas segmentation of biomedical images: A survey. *Medical Image Analysis* **24**(1) 205–219.

Ioffe, S., C. Szegedy. 2015. Batch normalization: Accelerating deep network training by reducing internal covariate shift. `http://arxiv.org/abs/1502.03167`.

Jaccard, P. 1912. The distribution of the flora in the alpine zone. *New Phytologist* **11** 37–50.

Jia, Y., Shelhamer E., Donahue J., Karayev S., Long J. et al. 2014. Caffe: Convolutional Architecture for Fast Feature Embedding. *arXiv preprint arXiv:10408.5093*. `http://arxiv.org/abs/1408.5093`.

Jolly, M. 2009. Fully automatic left ventricle segmentation in cardiac cine MR images using registration and minimum surfaces. *The MIDAS Journal – Cardiac MR Left Ventricle Segmentation Challenge* .

Jolly, M.P., Guetter C., Lu X., Xue H., Guehring J. 2012. Automatic segmentation of the myocardium in cine MR images using deformable registration. *Statistical Atlases and Computational Models of the Heart. Imaging and Modeling Challenges*. Springer, 98–108.

Kadish, A., Bello D., Finn J.P., Bonow R.O., Schaechter A. et al. 2009. Rationale and design for the Defibrillators to Reduce Risk by Magnetic Resonance Imaging Evaluation (DETERMINE) trial. *Journal of Cardiovascular Electrophysiology* **20** 982–987.

Klein, A., Mensh B., Ghosh S., Tourville J., Hirsch J. 2005. Mindboggle: automated brain labeling with multiple atlases. *BMC Medical Imaging* **5**(7).

Krizhevsky, A., Sutskever I., Hinton G.E. 2012. ImageNet classification with deep convolutional neural networks. *Neural Information Processing Systems*.

Li, B., Liu Y., Occleshaw C.J., Cowan B.R., Young A.A. 2010. In-line automated tracking for ventricular function with magnetic resonance imaging. *JACC Cardiovasc. Imag.* **3** 860–866.

Liu, H., Hu H., Xu X., Song E. 2012. Automatic left ventricle segmentation in cardiac MRI using topological stable-state thresholding and region restricted dynamic programming. *Academic Radiology* .

Liu, W., Rabinovich A., Berg A.C. 2015. ParseNet: Looking wider to see better. `http://arxiv.`

org/abs/1506.04579.

Long, J., Shelhamer E., Darrell T. 2015. Fully convolutional networks for semantic segmentation. http://arxiv.org/abs/1411.4038.

Maier, O., Jimenez D., Santos A., Ledesma-Carbayo M. 2012. Segmentation of RV in 4D cardiac MR volumes using Region-Merging Graph Cuts. *Computing in Cardiology*. IEEE, 697–700.

Margeta, J., Geremia E., Criminisi A., Ayache N. 2012. Layered spatio-temporal forests for left ventricle segmentation from 4D cardiac MRI data. *Statistical Atlases and Computational Models of the Heart. Imaging and Modeling Challenges*. Springer, 109–119.

Miller, C.A., Jordan P., Borg A., Argyle R., Clark D., Pearce K., Schmitt M. 2013. Quantification of left ventricular indices from SSFP cine imaging: Impact of real-world variability in analysis methodology and utility of geometric modeling. *Journal of Magnetic Resonance Imaging* **37**(5) 1213–1222.

Nair, V., G.E. Hinton. 2010. Rectified Linear Units Improve Restricted Boltzmann Machines. *27th International Conference on Machine Learning*.

Nambakhsh, C.M., Yuan J., Punithakumar K., Goelaa A., Rajchl M. et al. 2013. Left ventricle segmentation in MRI via convex relaxed distribution matching. *Medical Image Analysis* **17** 1010–1024.

Ngo, T.A., G. Carneiro. 2013. Left ventricle segmentation from cardiac MRI combining level set methods with deep belief networks. *20th IEEE International Conference on Image Processing*. IEEE, 695–699.

Oquab, M., Bottou L., Laptev I., Sivic J. 2014. Learning and transferring mid-level image representations using convolutional neural networks. *IEEE Conference on Computer Vision and Pattern Recognition*. IEEE, 1717–1724.

Ou, Y., Doshi J., Erus G., Davatzikos C. 2012. Multi-atlas segmentation of the cardiac MR right ventricle. *Proceedings of 3D Cardiovascular Imaging: a MICCAI Segmentation Challenge*.

Ourselin, S., Roche A., Prima S., Ayache N. 2000. Block matching: a general framework to improve robustness of rigid registration of medical images. *Medical Image Computing and Computer-Assisted Intervention – MICCAI 2000*. Springer, Berlin/Heidelberg, 557–566.

Petitjean, C., J.N. Dacher. 2011. A review of segmentation methods in short axis cardiac MR images. *Medical Image Analysis* **15**(2) 169–184.

Petitjean, C., Zuluaga M.A., Bai W., Dacher J.N., Grosgeorge D. et al. 2015. Right ventricle segmentation from cardiac MRI: A collation study. *Medical Image Analysis* **19**(1) 187–202.

Queirós, S., Barbosa D., Heyde B., Morais P. et al. 2014. Fast automatic myocardial segmentation in 4D cine CMR datasets. *Medical Image Analysis* **18**(7) 1115–1131.

Radau, P., Lu Y., Connelly K., Paul G., Dick A.J., Wright G.A. 2009. Evaluation framework for algorithms segmenting short axis cardiac MRI. *The MIDAS Journal – Cardiac MR Left Ventricle Segmentation Challenge*. http://hdl.handle.net/10380/3070.

Razavian, A.S., Azizpour H., Sullivan J., Carlsson S. 2014. CNN features off-the-shelf: An astounding baseline for recognition. *IEEE Conference on Computer Vision and Pattern Recognition*. 806–813.

Ren, S., He K., Girshick R., Sun J. 2016. Faster R-CNN: Towards real-time object detection with Region Proposal Networks. http://arxiv.org/abs/1506.01497.

Rohlfing, T., Brandt R., Menzel R., Maurer Jr C.R. 2004. Evaluation of atlas selection strategies for atlas-based image segmentation with application to confocal microscopy images of bee brains.

*Neuroimage* **21** 1428–1442.

Simonyan, K., A. Zisserman. 2015. Very deep convolutional networks for large-scale image recognition. `http://arxiv.org/abs/1409.1556`.

Srivastava, N., Hinton G., Krizhevsky A., Sutskever I., Salakhutdinov R. 2014. Dropout: A simple way to prevent neural networks from overfitting. *Journal of Machine Learning Research* **15** 1929–1958.

Suinesiaputra, A., Cowan B.R., Al-Agamy A.O., Elattar M.A. et al. 2014. A collaborative resource to build consensus for automated left ventricular segmentation of cardiac MR images. *Medical Image Analysis* **18**(1) 50–62.

Tavakoli, V., A.A. Amini. 2013. A survey of shaped-based registration and segmentation techniques for cardiac images. *Computer Vision and Image Understanding* **117** 966–989.

Wang, C.W., Peng C.W., Chen H.C. 2012. A simple and fully automatic right ventricle segmentation method for 4-dimensional cardiac MR images. *Proceedings of 3D Cardiovascular Imaging: a MICCAI Segmentation Challenge*.

Warfield, S.K., Zou K.H., Wells W.M. 2004. Simultaneous truth and performance level estimation (STAPLE): an algorithm for the validation of image segmentation. *IEEE Trans. Med. Imag.* (23) 903–921.

Yosinski, J., Clune J., Bengio Y., Lipson H. 2014. How transferable are features in deep neural networks? *Advances in Neural Information Processing Systems 27*. 3320–3328.

Zeiler, M.D., R. Fergus. 2014. Visualizing and understanding convolutional networks. *Computer Vision – ECCV 2014*. 818–833.

Zheng, S., Jayasumana S., Bernardino R.P., Vineet V., Su Z., et al. 2015. Conditional Random Fields as Recurrent Neural Networks. *International Conference on Computer Vision*.

Zuluaga, M., Cardoso M., Modat M., Ourselin S. 2013. Multi-atlas propagation whole heart segmentation from MRI and CTA using a local normalised correlation coefficient criterion. *Functional Imaging and Modeling of the Heart*. Lecture Notes in Computer Science, 172–180.

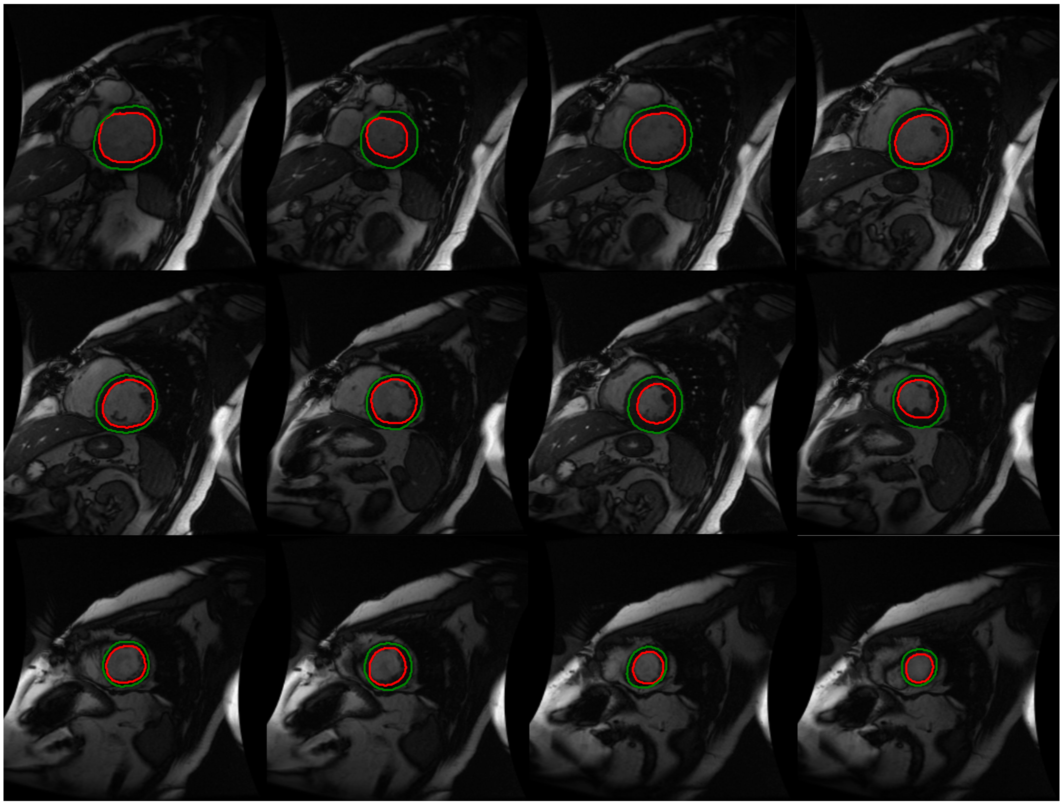

**Figure 2.** FCN segmentation result of an example test case in the Sunnybrook dataset for both ED and ES phases. Colors: red – endocardium; green – epicardium.

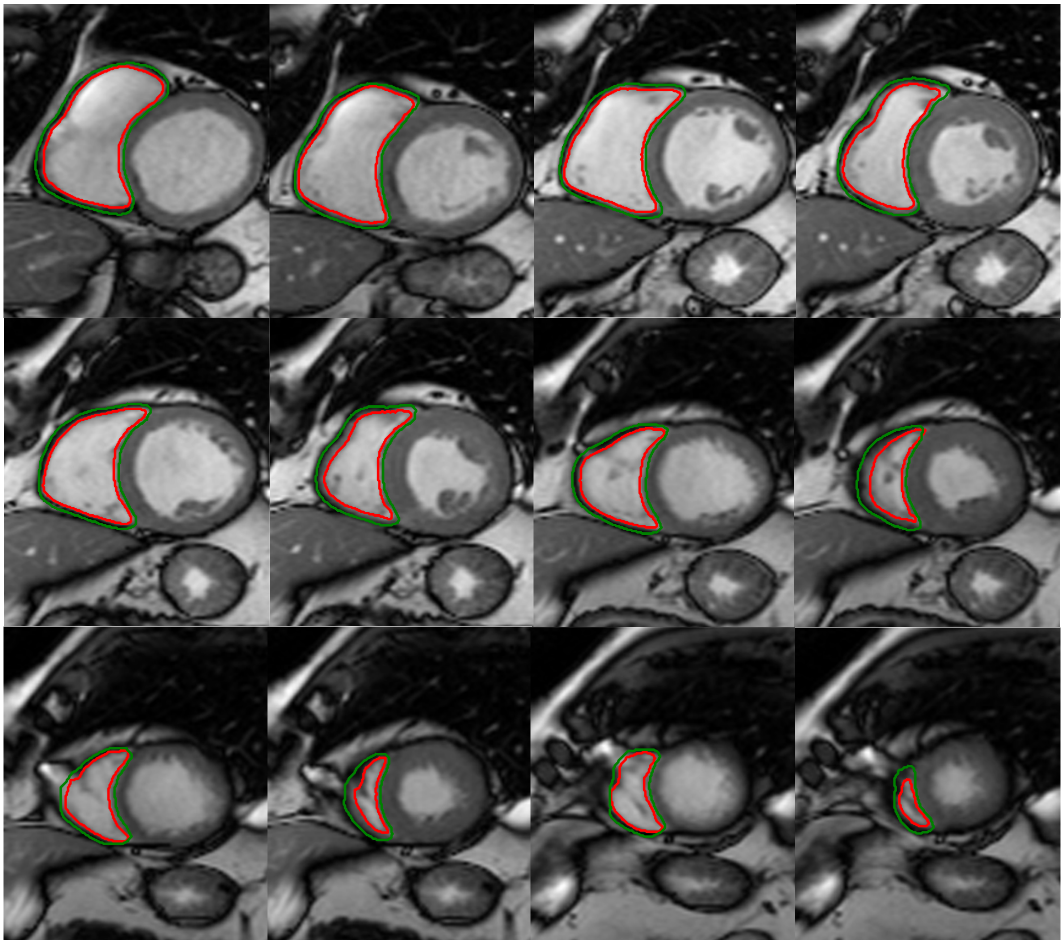

**Figure 3.** FCN segmentation result of an example test case in the RVSC dataset for both ED and ES phases. Colors: red – endocardium; green – epicardium.

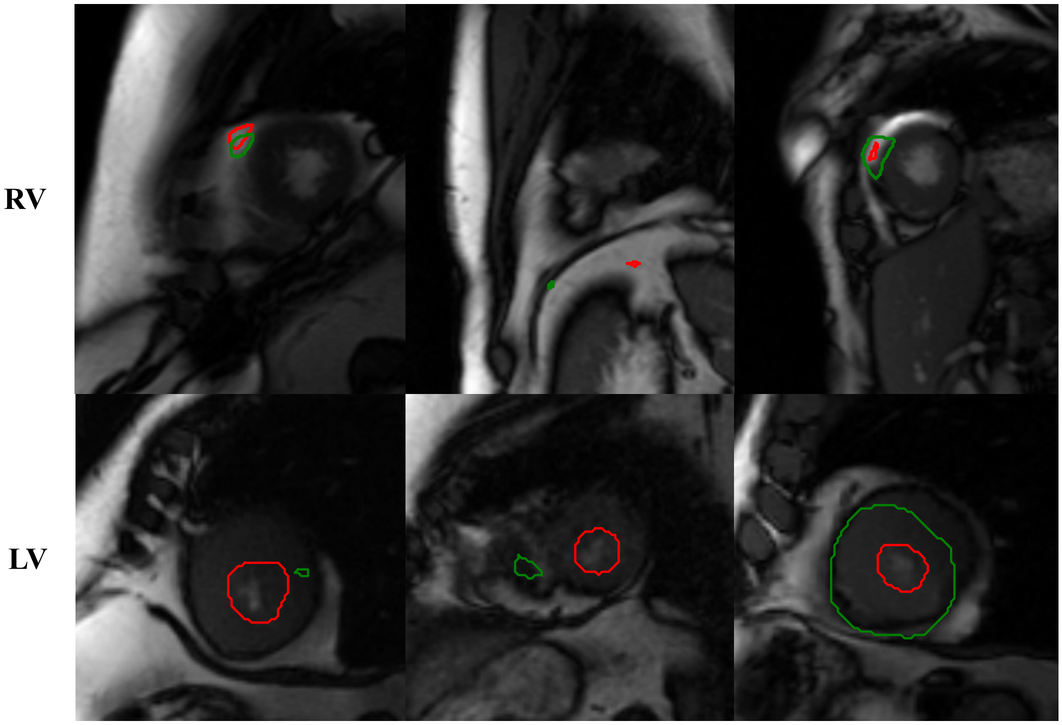

**Figure 4.** Examples of poor FCN segmentation on difficult apical slices having ambiguous or imperceptible object boundaries. Cropped and zoomed in for better viewing. Colors: red – endocardium; green – epicardium.