Year 7
Final Project Report
2018-2019

Project 7a.021 - Early Detection of Myocardial Infarction Using Echocardiogram Images

A National Science Foundation
University Cooperative Research Center

Project 7a.021 - Early Detection of Myocardial Infarction Using Echocardiogram Images

Contents

Personnel

PI First and Last Name (PI):

- Moncef Gabbouj

Other team member's first and last name (project role):

- Serkan Kiranyaz (Co-PI)
- Morteza Zabihi (Researcher)
- Aysen Degerli (Research Assistant)

Sponsoring IAB member's first and last name (company name):

- **Tieto**

Executive Summary/Abstract

The significant proportion (20-30%) of emergency department admissions are related to patients with acute chest pain. These patients are required to have a rapid assessment due to their time-critical condition. It has been shown that parameters such as changes in ECG characteristics or elevation of troponin level may detect only 30% of acute ischemic events. Here, echocardiography can play a valuable role as an alternative diagnostic tool in an appropriate triage of patients with acute chest pain. Echocardiography is a reliable method for revealing the anomalies in the regional heart wall motion. Due to the early manifestation of Myocardial Infarction (MI) symptoms in echocardiogram, this imaging modality is now included in the universal definition of acute MI and in international guidelines regarding the management of cardiac arrest. In this project, the ultimate goal is to design an automatic model, which traces the movement of the heart's wall using the echocardiogram images and detects the anomalies in the wall motion. The initial focus of this work is detecting the Left ventricle (LV) muscle in each echo frames.

In this project, we evaluate our methods using a benchmark dataset including 152 echocardiogram videos from healthy and patients with MI. We first provide a pseudo labeling process to generate ground truth mask for LV location in each frame. Then, we used two state-of-the-art Convolutional Neural Networks (CNN) to segment the LV muscle. To the best of our knowledge, this is the first application of CNN for labeling the complete LV muscle at pixel-level. Numerical experiments demonstrate that both of the models have a robust and reliable performance for LV detection. Moreover, we design a flickering detection to be able to choose the minimum flickering among the predictions of different CNNs. Flickering evaluation is an essential step because it can easily lead to MI misclassification. Currently, we are waiting for MI labels from medical doctors. We shall continue this project in the next year to complete our end-to-end automatic MI detection.

Goals and Objectives

1. Provide the ground truth mask for the LV in the echo dataset using deep learning.
    a. Using a semi-automated process we drastically decrease the doctors' burden
2. Fully automated segmentation of the full LV muscle in echo videos with 4 chamber views
3. Comparative analysis of two different CNN topologies for LV detection
4. Propose an ensemble of CNNs and flickering detection to increase the MI detection accuracy
5. Motion tracking of the detected LV muscle in echo video for early MI detection (*Continuation of this project for the year 2019-2020*)

Differences from Current State of Art

1. Generating high-quality ground truth for the unannotated dataset using pseudo labeling

2. To the best of our knowledge, this is the first attempt to automatically detect and segment the full LV wall in echocardiography. In several studies (e.g., [1] and [2]), the focus is to detect only the edge of the muscle using active contours based methods. In this work, we segment the full LV muscle wall. This can provide more information and accuracy for various applications such as MI detection or any anatomical deficiency diagnosis associated with LV.

Methods and Datasets

Dataset: We use a collection of 152 echocardiogram videos from healthy subjects and patients with MI. We focus our study on the videos with the 4-chamber view. Based on clinical guidelines for MI detection, we designed methods for tracking the heart motion of the LV [3]. The dataset does not include annotated LV masks. Therefore, we use a pseudo labeling process to overcome this issue. This approach also gives us information regarding noisy videos. We shall explain this approach in the next section. Moreover, all frames are resized to the same size of 224×224. In addition, these dimensions are suitable for many state-of-the-art CNN algorithms. This makes the dataset appropriate for several experiments and comparative analysis.

Pseudo Labeling: For supervised leaning approach, we need to have enough labels for training. Due to the number of echo frames in the dataset, it is not practical to have expert annotation for each frame. To solve this problem, we use a pseudo labeling technique. Instead of manually labeling the whole dataset, ground truth masks for only a few frames were drawn manually by medical doctors and then the rest of the dataset is labeled based on the annotated ones. This process consists of the following steps:

(1) Build a CNN model based on the available annotated data
(2) Predict the mask for the unlabeled data
(3) Evaluate the predicted masks and exclude the incorrect masks
(4) Build a new model using the combination of labeled data from steps (1) and (3)

The aforementioned four steps continue until the whole data is labeled or a predefined constraint is met (see Figure 1). In this project, for the steps (1) and (4) two CNN topologies are used (these models are explained in Section *CNN Topology*). In step (3), first, the model is chosen which has less flickering. Then, based on visual inspection, incorrectly predicted masks are removed. For flickering detection, the difference between every two consecutive masks in the video is measured and then the variance of the calculated differences is obtained. The model with minimum variance is chosen as the candidate model (more information provided in Section *An ensemble of CNN*). The whole process is repeated eight times to assure high-quality LV masks.

It is obvious that after each iteration, only the challenging videos remain. On the other hand, the models' performance increase as they are fed with more training data. Therefore, this process can also be used for detecting noisy or problematic videos. In this study, 10 videos still after eight iterations cannot be segmented correctly. Further investigation confirms that the 10 remaining echo

videos are either too noisy or part of the LV muscle is not visible. Two examples of such noisy frames are shown in Figure 2.

LV Segmentation: Once the segmentation masks for the dataset created using the pseudo labeling technique, we use two different CNN topologies and evaluate their capabilities in learning the LV segmentation. Moreover, an ensemble of CNNs is designed in order to create more accurate and robust MI detection.

❖ CNN Topology: We used two encoder-decoder networks, whose structures are formed by stacking convolutional and deconvolutional blocks (see Figure 3 and 4). *Model 1* is inspired by UNet [4] and all the blocks are designed with shortcut connections which provide a more efficient training procedure [5]. For *Model 2* we used FCN [6] topology. Both of the selected topologies have shown state-of-the-art performances in segmentation of medical images. Further detailed information regarding the models is provided in the *Appendix*.

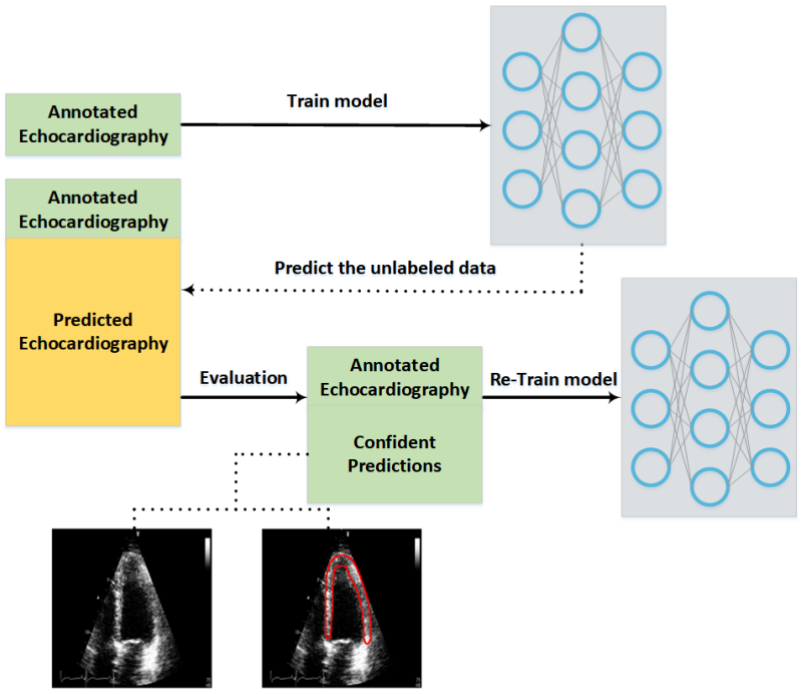

Figure 1. The pseudo labeling procedure

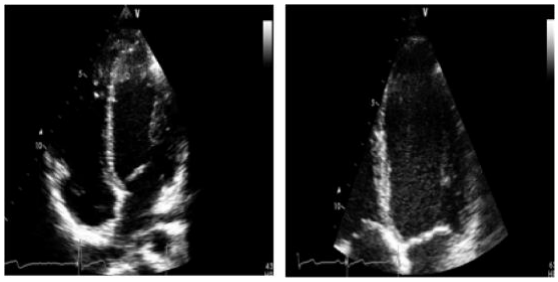

Figure 2. Two examples of noisy echo frames. The right side of LV, in both *a)* and *b)*, is missing and not visible.

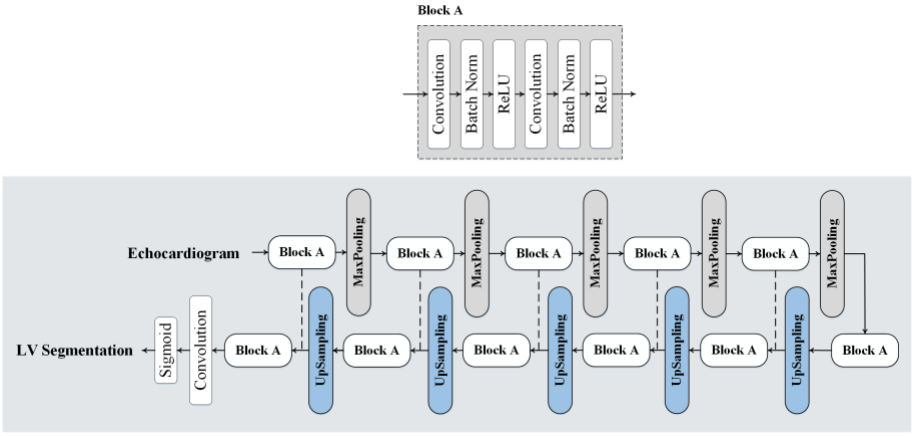

Figure 3. The topology of *Model 1* [4]

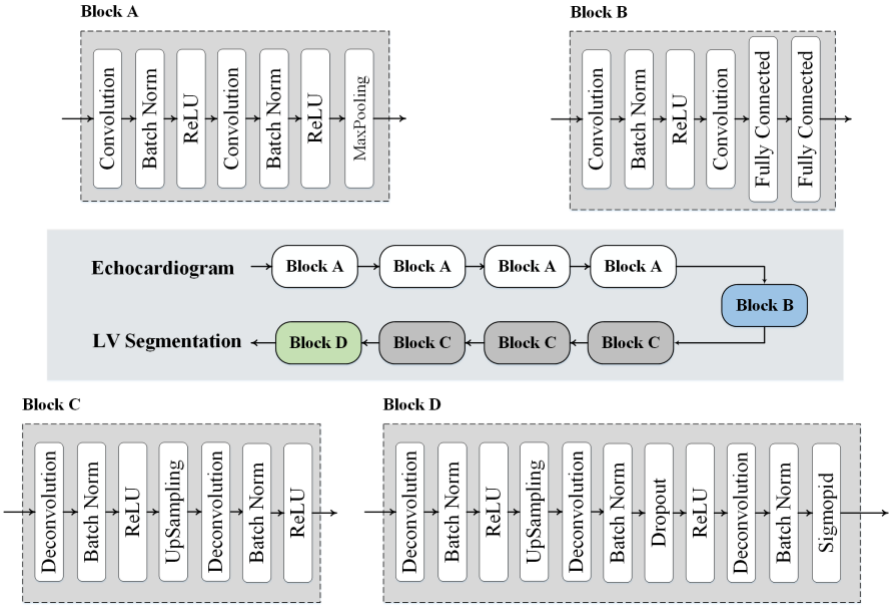

Figure 4. The topology of *Model 2* [6]

❖ Training Procedure: For evaluation, we followed the 5-fold cross-validation scheme. To be more specific, we train our model using 80% of the echo videos available in the dataset and test our model using the 20% holdout echo videos. In this work, the Adam optimization algorithm [7] and cross-entropy loss function are used to train the CNN with 32 mini-batch size, 25 epochs, and learning rate of 0.001. These hyper-parameters are tuned empirically. For each model, 20% of the training frames are randomly chosen as the validation set. The models are implemented in Keras using Tensorflow backend, and the experiments were performed on a workstation with NVIDIA TITAN V GPU and 160 GB memory.

❖ Postprocessing: For postprocessing, we remove noise and false small-detected objects (i.e., false positives) from the CNN predictions while preserving the shape and size of the detected LV. For this purpose, we use the morphological opening operation, which is erosion followed by dilation, using a kernel with values of 1 and a size of 3×3.

❖ An ensemble of CNN: We also combine *Model 1* and *Model 2* to achieve more robust results for MI detection. To be more specific, we generate LV masks using each of the models and then chose the one with a less flickering effect. Existence of flickering can drastically decrease the MI detection accuracy. MI can be diagnosed based on the amount of the heart muscles movement in a cardiac cycle. Therefore, we choose the minimum amount of flickering as the main criterion between several masks.

❖ Evaluation Metrics: For evaluating the performance of the proposed method, we consider segmentation as a binary problem. The LV muscle is determined as class 1 and the background is class 0. Thus, *True Negative (TN)* is defined as if the background correctly detected, *False Negative (FN)* is if the LV muscle is misclassified as background, *False Positive (FP)* is if the background is misclassified as the LV muscle, and the *True Positive (TP)* is if the LV muscle is correctly detected. For each frame in each echo videos, we calculate the *Sensitivity*, *Specificity*, *Precision*, *Negative Predictive Value*, *F1 measure*, and *Accuracy* as follows:

$$Sensitivity = \frac{TP}{TP+FN}, Specificity = \frac{TN}{TN+FP},$$
$$Precision = \frac{TP}{TP+FP}, Negative\ Predictive\ Value = \frac{TN}{TN+FN},$$
$$F1 = \frac{2TP}{2TP+FP+FN}, Accuracy = \frac{TP+TN}{TP+TN+FP+FN}.$$

Results

Tables 1, 2, and 3 summarize our experiments. For the LV segmentation, both models achieved robust and high performances. It is worth mentioning that these evaluation measures reflect the performance of the LV segmentation at pixel-level. As can be seen in Tables 1 and 2, *Model 1* achieved a relatively higher *Precision* and *F1* compared to *Model 2*. Table 3 shows the performance of the ensemble of CNNs. It should be mentioned that our main goal of using ensemble learning is to decrease the flickering effect on the predicted LV masks. That said, the

ensemble model improves the *Precision* compared to *Model 2*. The last column of Table 3, indicating the proportion of the times each model is chosen. As can be seen, overall no signle model shows significant superiority in terms of minimum flickering. This confirms that an ensemble of CNNs can outperform each of the two models in terms of MI detection. However, we need to access the MI annotation of the dataset to confirm this.

Table 1. The average (and standard deviation) of the evaluation metrics for unseen test set in a 5-fold cross-validation scheme using *Model 1*.

Fold	Sens.	Spec.	Prec.	NPV.	F1	Acc.
1	0.9453 (0.0313)	0.9970 (0.0024)	0.9369 (0.0460)	0.9974 (0.0017)	0.9404 (0.0308)	0.9947 (0.0032)
2	0.9334 (0.0434)	0.9977 (0.0019)	0.9437 (0.0503)	0.9970 (0.0025)	0.9373 (0.0343)	0.9950 (0.0029)
3	0.9561 (0.0310)	0.9965 (0.0027)	0.9233 (0.0529)	0.9980 (0.0015)	0.9382 (0.0305)	0.9948 (0.0027)
4	0.9862 (0.0236)	0.9940 (0.0024)	0.8771 (0.0460)	0.9994 (0.0012)	0.9279 (0.0314)	0.9937 (0.0029)
5	0.9477 (0.0368)	0.9973 (0.0022)	0.9396 (0.0487)	0.9976 (0.0017)	0.9426 (0.0307)	0.9951 (0.0026)
Average	**0.9538 (0.0332)**	**0.9965 (0.0023)**	**0.9241 (0.0488)**	**0.9979 (0.0017)**	**0.9373 (0.0315)**	**0.9947 (0.0029)**

Table 2. The average (and standard deviation) of the evaluation metrics for unseen test set in a 5-fold cross-validation scheme using *Model 2*.

Fold	Sens.	Spec.	Prec.	NPV.	F1	Acc.
1	0.9872 (0.0216)	0.9926 (0.0032)	0.8594 (0.0475)	0.9994 (0.0010)	0.9182 (0.0308)	0.9924 (0.0034)
2	0.9706 (0.0360)	0.9949 (0.0018)	0.8847 (0.0472)	0.9986 (0.0019)	0.9246 (0.0297)	0.9938 (0.0024)
3	0.9313 (0.0433)	0.9966 (0.0021)	0.9228 (0.0467)	0.9971 (0.0017)	0.9257 (0.0293)	0.9940 (0.0023)
4	0.9178 (0.0482)	0.9982 (0.0014)	0.9571 (0.0342)	0.9966 (0.0019)	0.9360 (0.0303)	0.9950 (0.0023)
5	0.9812 (0.0252)	0.9928 (0.0030)	0.8580 (0.0578)	0.9991 (0.0011)	0.9144 (0.0362)	0.9923 (0.0032)
Average	**0.9576 (0.0348)**	**0.9950 (0.0023)**	**0.8964 (0.0467)**	**0.9981 (0.0015)**	**0.9238 (0.0313)**	**0.9935 (0.0027)**

Table 3. The average (and standard deviation) of the evaluation metrics for unseen test set in a 5-fold cross-validation scheme using the ensemble model.

Fold	Sens.	Spec.	Prec.	NPV.	F1	Acc.	*Model 1/ Model 2*
1	0.9462 (0.0362)	0.9963 (0.0025)	0.9219 (0.0485)	0.9975 (0.0019)	0.9331 (0.0338)	0.9941 (0.0033)	0.45/0.55
2	0.9394 (0.0556)	0.9969 (0.0024)	0.9259 (0.0535)	0.9973 (0.0031)	0.9312 (0.0427)	0.9944 (0.0039)	0.52/0.48
3	0.9632 (0.0503)	0.9950 (0.0030)	0.8952 (0.0597)	0.9985 (0.0019)	0.9259 (0.0375)	0.9938 (0.0032)	0.64/0.36
4	0.9461 (0.0390)	0.9957 (0.0022)	0.9049 (0.0502)	0.9977 (0.0013)	0.9244 (0.0392)	0.9938 (0.0028)	0.25/0.75
5	0.9571 (0.0469)	0.995 (0.0028)	0.9075 (0.0583)	0.9980 (0.0022)	0.9306 (0.0442)	0.9940 (0.0040)	0.39/0.61
Average	**0.9504 (0.0456)**	**0.9959 (0.0026)**	**0.9111 (0.0540)**	**0.9978 (0.0021)**	**0.9291 (0.0395)**	**0.9940 (0.0034)**	-

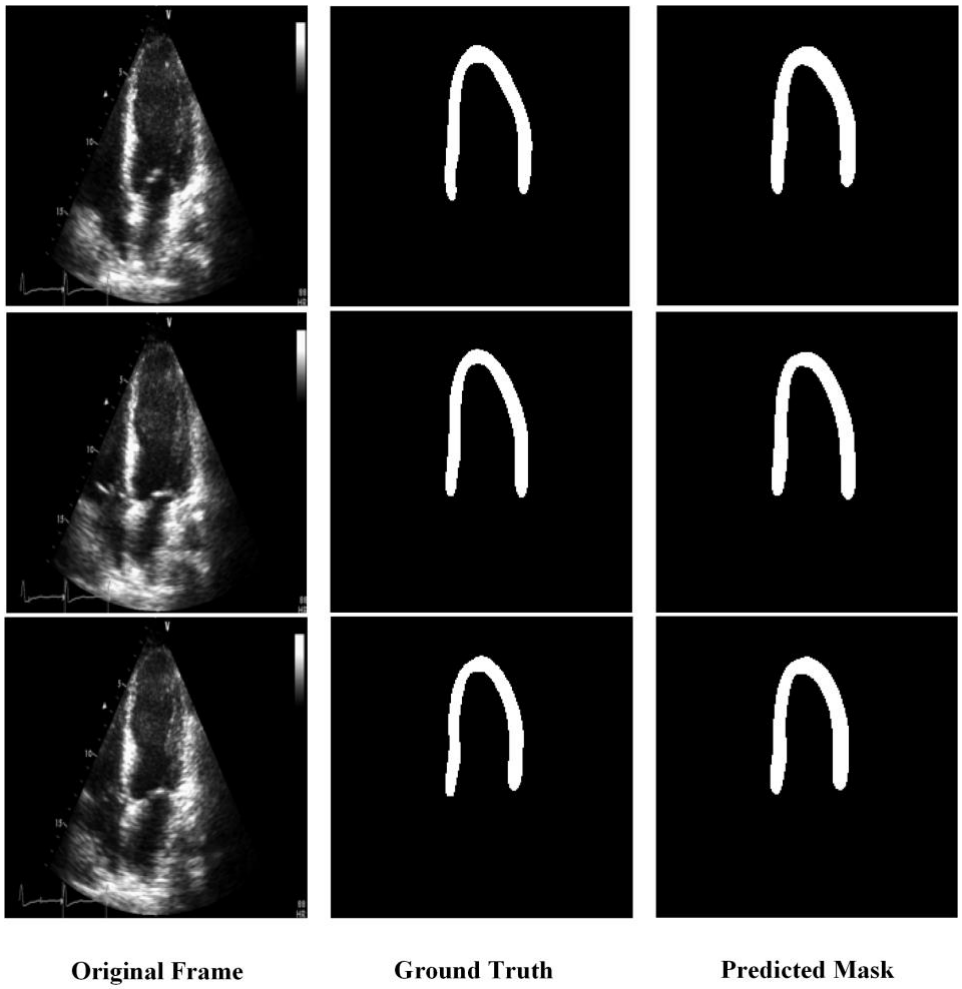

Figure 5. Example of segmentation results generated by *Model 1*. First, second, and third rows are representing the first, fourth, and ninth frame of an echo video, respectively.

Functionality of Innovation(s)

1. The first novelty of the proposed method is using the pseudo labeling process to overcome the lack of annotation data. In addition, this method is used to detect noisy data. The proposed method shows significant impacts on decreasing the expert's burden and improvement segmentation performance.
2. The achieved performance of LV segmentation shows the reliability and accuracy of the proposed method.
3. Using an ensemble of CNNs and choosing the masks with less flickering can improve MI detection.

Conclusions and Recommendations

In this project, we demonstrate the efficacy and high performance of two state-of-the-art CNN architectures for echocardiogram videos. We first showed that using the proposed pseudo labeling we can generate high-quality ground truth for the unannotated dataset. Empirical evaluation of the two CNNs revealed that both *Model 1* and *Model 2* achieved high accuracy on multiple evaluation metrics for LV segmentation. For MI detection not only the accuracy of segmentation is important but also minimum flickering between frames is also crucial. For this purpose, we designed a flickering detection method based on the differences between every two consecutive frames in an echo video. Afterwards, we designed an ensemble of CNNs formed by the two used models. The main reason for the ensemble learning is to assure the minimum flickering in an echo video, which is essential for MI detection. We are now writing a scientific paper based on the achieved results. This project will continue on the next year to provide an end-to-end automatic method for early MI detection.

Impact and Uses/Benefits

1. The proposed pseudo labeling process can be used for any other dataset with only a few annotated data. The constraint can be adapted for the application of interest.
2. To the best of our knowledge, this study provides the annotations for the first and largest MI echo dataset available in this domain.
3. The trained models in this study can be used for other dataset using transfer learning.
4. The provided LV masks can be used for further studies on the anatomical structure of heart muscle.
5. The proposed method can be used along with echocardiogram devices to provide a fast and reliable LV segmentation for cardiologists. This information can lead to improvement of the MI diagnosis.

List of References

[1] G. Dharanibai, A. Chandrasekharan, Z. C. Alex, "Automated Segmentation of Left Ventricle Using Local and Global Intensity Based Active Contour and Dynamic Programming," *International Journal of Automation and Computing*, pp. 673-688, 2018.

[2] Y. Niu, L. Qin, X. Wang, "Structured graph regularized shape prior and cross-entropy induced active contour model for myocardium segmentation in CTA images," *Neurocomputing*, vol. 357, pp. 215-230, 2019.

[3] C. F. P. Wharton, C. S. Smithen, E. Sowton, "Changes in Left Ventricular Wall Movement after Acute Myocardial Infarction Measured by Reflected Ultrasound," *Br Med J*, vol. 5779, no. 4, p. 75–77, 1971.

[4] O. Ronneberger, P. Fischer, T. Brox, "U-Net: Convolutional Networks for Biomedical Image Segmentation," in *International Conf. on Medical Image Computing and Computer-assisted Intervention*, 2015.

[5] K. He, X. Zhang, S. Ren, J Sun, "Deep Residual Learning for Image Recognition," in *IEEE Conf. on Computer Vision and Pattern Recognition (CVPR)*, Las Vegas, NV, USA, 2016.

[6] P. V. Tran, "A Fully Convolutional Neural Network for Cardiac Segmentation in Short-Axis MRI," *ArXiv*, 2016.

[7] D. P. Kingma, and J. Ba, "Adam: a Method for Stochastic Optimization," in *3rd International Conf. on Learning Representations (ICLR)*, 2014.

Appendix

Table 4. Implementation details of *Model 1*

	Encoder		Decoder	
Kernel size	Filters	MaxPooling	Filters	UpSampling
3×3	32	2×2	512	2×2
3×3	64	2×2	256	2×2
3×3	128	2×2	128	2×2
3×3	256	2×2	64	2×2
3×3	512	2×2	32	2×2
3×3	1024	-	-	-

Table 5. Implementation details of *Model 2*

Encoder			Decoder		
Filters	Kernel size	MaxPooling	Filters	Kernel size	UpSampling
16	3×3	-	256	3×3	2×2
32	3×3	2×2	256	3×3	-
64	3×3	-	128	3×3	2×2
64	3×3	2×2	128	4×4	-
128	3×3	-	128	3×3	2×2
128	4×4	2×2	64	3×3	-
256	3×3	-	32	3×3	2×2
256	3×3	2×2	16	3×3	-
512	3×3	-	1	3×3	-
Fully Connected (1024) Fully Connected (1024)					