An Adversarial Network Architecture Using 2D U-Net Models for Segmentation of Left Ventricle from Cine Cardiac MRI

Roshan Reddy Upendra[1], Shusil Dangi[1], and Cristian A. Linte[1,2]

[1] Center for Imaging Science, Rochester Institute of Technology, Rochester, NY USA
[2] Biomedical Engineering, Rochester Institute of Technology, Rochester, NY USA
{ru6928,sxd7257,calbme}@rit.edu

Abstract. Cardiac magnetic resonance imaging (CMRI) provides high resolution images ideal for assessing cardiac function and diagnosis of cardiovascular diseases. To assess cardiac function, estimation of ejection fraction, ventricular volume, mass and stroke volume are crucial, and the segmentation of left ventricle from CMRI is the first critical step. Fully convolutional neural network architectures have proved to be very efficient for medical image segmentation, with U-Net inspired architecture as the current state-of-the-art. Generative adversarial networks (GAN) inspired architectures have recently gained popularity in medical image segmentation with one of them being SegAN, a novel end-to-end adversarial neural network architecture. In this paper, we investigate SegAN with three different types of U-Net inspired architectures for left ventricle segmentation from cardiac MRI data. We performed our experiments on the 2017 ACDC segmentation challenge dataset. Our results show that the performance of U-Net architectures is better when trained in the SegAN framework than when trained stand-alone. The mean Dice scores achieved for three different U-Net architectures trained in the SegAN framework was on the order of 93.62%, 92.49% and 94.57%, showing a significant improvement over their Dice scores following stand-alone training - 92.58%, 91.46% and 93.81%, respectively.

Keywords: Image Segmentation - Deep Learning · Cine Magnetic Resonance Image · Cardiac Image Analysis · Left Ventricle Segmentation.

1 Introduction

Cardiac magnetic resonance imaging (CMRI) is considered as the benchmark for analysis of the cardiac function and quantification of the ventricular volume [6]. The analysis of the function of the ventricles, described by their ejection fraction (EF), stroke volume, mass and wall thickness are crucial in clinical cardiology for diagnosis, evaluating risks and planning therapy [2] [14]. In particular, the function of left ventricle is a very good predictor of myocardial damage, cardiac failure, etc. Therefore, accurate and robust segmentation of left ventricle from the MRI data plays an important role in a large number of cardiac problems.

Manual segmentation can be a very laborious task prone to significant user variability. Therefore, semi-automatic or fully automated segmentation methods

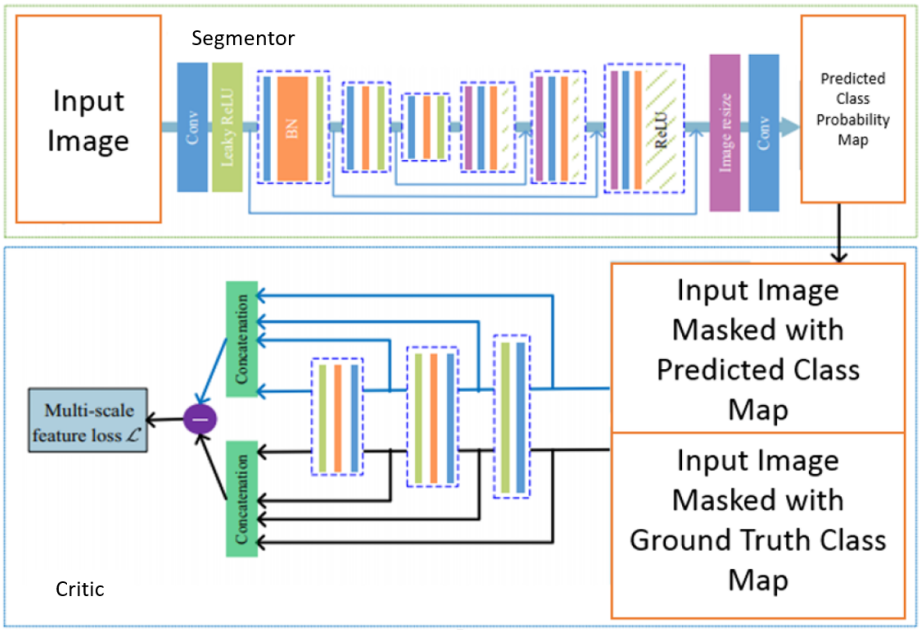

Fig. 1. SegAN Architecture Inspired from GAN [15]

would be very useful to cardiologists in the decision making process [8]. Some of the challenges in automated segmentation of the left ventricles are delineation between the myocardium of the left ventricle and other surrounding chambers, high contrast plus brightness heterogeneity in the ventricular cavity due to the presence of blood, the presence of papillary muscles, noise due to motion artifacts, and the variable structure of the heart [2].

A number of methods for automatic segmentation of the left ventricle from MRI images have been proposed. Traditional algorithms such as thresholding, edge detection, region growing, clustering, etc., were proposed initially [8]. These algorithms work decently for mid-ventricle slices, but often fail in the basal and apical slices. Also, they require considerable user-intervention. In graph based segmentation algorithms [7], graphs are created and a cost is assigned to each pixel or node. A minimum cost path is found using a graph searching algorithm to segment the left ventricle. These methods fail in complex cardiac structures, like papillary and trabecular muscles (PTMs). In [11], active shape models (ASM) are used to segment the left ventricles using the energy minimization of rigidity and elasticity internally, and edges externally. It is difficult to segment left ventricle from low contrast images using ASMs. In spite of all these research in left ventricle segmentation from MRI images, the accuracy of existing algorithms is not sufficient for clinical applications.

With the increase in popularity of deep learning due to the availability of resources for training, medical image segmentation has benefited immensely. Convolutional neural networks (CNNs) work outstandingly well for image classification where the output of the CNN to an image is a class label. The availability of large number of CMRI images enabled the use of deep learning for left ventricular segmentation. Several international challenges have been organized in the past few years to develop and evaluate segmentation algorithms for both ventricles [13], [12], [9]. The top ten results of the 2017 automated cardiac diagnosis challenge (ACDC) have all used CNNs for segmentation of the ventricles [2].

Ronneberger et al. [10] proposed U-Net, an elegant network architecture using fully convolutional network. U-Net is void of any fully connected layers and the convolutional layer labels each pixel in the image allowing segmentation with fewer training images [10].

Generative adversarial networks (GAN) [3] are becoming increasingly popular for medical image segmentation as training these generative models enable latent representations, which can serve as useful features. Eule et al. [4] show that segmentation of epithelial tissue using cycle-GAN outperforms the state-of-the-art U-Net. Xue et al. [15] proposed a GAN inspired end-to-end architecture, called segmentation adversarial network (SegAN), for semantic segmentation. They achieved better Dice score and precision compared to the state-of-the-art U-Net architecture in the segmentation of the MICCAI BRATS (2013 and 2015) brain tumor segmentation challenge [15].

In this paper, we combine three U-Net models with SegAN adversarial architecture to segment the left ventricle on the 2017 automated cardiac diagnosis challenge (ACDC) dataset. The objective of this paper is to test if SegAN, when combined with different U-Net architectures, produces better segmentation results than when stand-alone U-Net architectures are trained.

2 Methodology

Inspired by GAN, Xue et al. have come up with SegAN, an adversarial network that has two networks, segmentor and critic, analagous to generator and discriminator in GAN, respectively. The segmentor, a fully convolutional neural network, takes in raw images as input and outputs a probability label map. The critic network, which is the encoder part of the fully convolutional neural network needs two inputs - the masked image by the ground truth labels and the masked image by predicted labels obtained from the segmentor. The aim of the segmentor network is to minimize the L_1 loss function and that of critic network is to maximize the L_1 loss function [15].

2.1 Conventional GAN Models

In GANs, the loss function is defined as -

$$\min_{\theta_G} \max_{\theta_D} L(\theta_G, \theta_D) = E_{x \sim P_{data}}[logD(x)] + E_{z \sim P_z}[log(1 - D(G(z)))]. \quad (1)$$

In the above equation, θ_G and θ_D are the parameters of generator G and discriminator D, respectively. x and z are real image from unknown distribution P_{data} and random input for G from probability distribution P_z, respectively. The generator G outputs a high dimensional vector which is the input to the discriminator D. The discriminator D is trained to maximize the probability of assigning the correct label to the training data and the data generated from G. The generator G is simultaneously trained to minimize the objective function $log(1 - D(G(z)))$ to generate images that are difficult to differentiate for D [3]. The aim of the generator is to produce images that are as similar as possible to the real image and the aim of the discriminator is to successfully distinguish between the real image and the fake image produced by the generator.

2.2 Loss Function in SegAN

In SegAN, the aim is to solve the mapping between input images and their segmentation masks. The loss function L for SegAN is given by -

$$\min_{\theta_S} \max_{\theta_C} L(\theta_S, \theta_C) = \frac{1}{N} \sum_{n=1}^{N} l_{mae}(f_C(x_n \circ S(x_n)), f_C(x_n \circ y_n)). \tag{2}$$

In this equation, θ_S and θ_C are the parameters of segmentor S and critic C, respectively and N represents the number of training images. $(x_n \circ S(x_n))$ and $(x_n \circ y_n)$ are input images masked with segmentor predicted label map and ground truth, respectively. $f_c(x)$ are the features extracted from image x by critic and l_{mae} is the mean absolute error (MAE) given by -

$$l_{mae}(f_C(x), f_C(x')) = \frac{1}{L} \sum_{i=1}^{L} ||f_C^i(x) - f_C^i(x')||_1, \tag{3}$$

with L representing the number of layers in the critic network [15].

The segmentor and critic networks are trained alternatively, just like GAN. The difference between GAN and SegAN is that GAN has two seperate losses for generator and discriminator, while, the SegAN has only one multi-scale L_1 loss funtion for both segmentor and critic.

2.3 Segmentor and Critic

We use three different segmentor networks to predict the segmented mask and compare their results. The first one is the original ***U-Net*** [10]. The second one is a U-Net architecture with skip connection used in [15] (***U-Net A***). The third segmentor used is a modified version of the U-Net architecture inspired from [5] (***U-Net B***). The input to all these three networks are raw images and the output is a predicted mask.

For the critic network, we used a similar structure to the downsampling part of the corresponding segmentor network to extract hierarchical features from

multiple layers of the network. We then concatenated all these features extracted across multiple layers and computed the overall L_1 loss using the concatenated feature vector [15]. The input to the critic network are two images - input image masked with predicted class map and input image masked with the ground truth class map; and output is a feature vector.

2.4 Training U-Net and SegAN Models

Our experiments involve three different U-Net architectures - the original ***U-Net*** from [10], the encoder-decoder network used as segmentor in [15] (***U-Net A***), and a modified U-Net inspired from [5] (***U-Net B***), which is the current state-of-the-art for left ventricle segmentation in the ACDC 2017 dataset. The U-Net models are trained with cross entropy loss function. We experimented with Dice loss but cross entropy loss gave us better results. This is supported by results in [1], too.

The segmentor and the critic network are trained alternately using back-propagation and the loss function. First, the segmentor outputs a predicted class map. Then, the segmentor is fixed and the critic is trained in the next step using gradients calculated from the loss function. After that, the critic is fixed and the segmentor is trained using gradients from the loss function passed to the segmentor from the critic [15]. As explained in GANs, this process resembles a min-max game, where the segmentor aims to minimize the loss and the critic tries to maximize it. Provided additional data and more epochs, the segmentor will produce segmented masks i.e. labelled maps that are similar to the ground truth. For each U-Net model we use as segmentor, we use the encoder part of that particular U-Net model as critic.

We train the U-Net and SegAN models by resizing each slice to a 224x224 image and feeding it into the network with a learning rate of 0.0008, a batch size of 8, a decay of 0.5, a beta value of 0.5, one GPU and 50 epochs.

2.5 Dataset

The **A**utomated **C**ardiac **D**iagnosis **C**hallenge (ACDC) dataset was released during the MICCAI 2017 conference in conjunction with the STACOM workshop. The images were acquired using two different MRI scanners with different magnetic strength - 1.5 T and 3.0 T. The short axis slices cover the left ventricle from base to apex such that we get one image every 5 mm to 10 mm. A complete cardiac cycle is usually covered by 28 to 40 images. Their spatial resolution is 1.37 to 1.68 $mm^2/pixel$ [2]. The training dataset is composed of 100 subjects and the test dataset is composed of 50 subjects.

The image dataset corresponding to each subject consists of two image volumes, one at end-diastole and one at end-systole, with each containing 10 slices, therefore leading to a total of $1,902$ images. Since we do not have the ground truth for the 50 test subjects, we divide the training dataset into 80 subjects for training and 20 subjects for validation. The Dice score and the IoU in this paper are the result of 5-fold cross validation of the training dataset.

Table 1. Segmentation evaluation, mean (std-dev) for end diastole (ED) and end systole (ES) left ventricle segmentation in the 2017 ACDC dataset. Statistical significance (T-test) of the results of SegAn architecture compared against U-Net models are represented by * for p < 0.1 and ** for p < 0.05. The best Dice values achieved are labeled in **bold**.

	Dice (ED) (%)	IoU (ED) (%)	Dice (ES) (%)	IoU (ES) (%)
U-Net	93.41 (4.23)	87.25 (3.12)	91.75 (2.26)	83.64 (4.01)
SegAN + U-Net	*94.71 (1.24)***	*89.55 (2.46)***	*92.54 (3.89)*	*84.91 (5.75)*
U-Net A	92.62 (2.75)	85.27 (1.81)	90.30 (7.11)	81.58 (5.60)
SegAN + U-Net A	*93.88 (2.86)**	*88.54 (1.12)*	*91.10 (4.15)**	*82.74 (5.71)*
U-Net B	94.91 (2.40)	91.55 (3.23)	92.72 (4.71)	87.44 (3.81)
SegAN + U-Net B	***95.87 (1.71)****	***92.94 (3.27)****	***93.14 (2.56)****	***88.94 (3.92)***

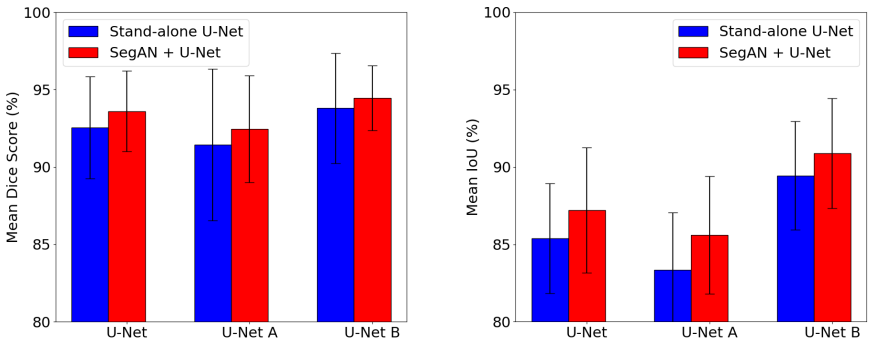

Fig. 2. Comparison of (a) mean Dice scores and (b) mean IoU values U-Net models and its corresponding SegAN architecture

3 Experiments and Results

The focus of our experiment is to compare the results of a stand-alone 2D U-Net architecture with a SegAN architecture. For example, we obtain segmentation results using U-Net [10] with cross entropy loss as cost function. Then, we use this U-Net [10] as segmentor and the downsampling part of the U-Net as critic in the SegAN architecture with multi-scale L_1 loss as cost function. The results of these two networks are compared to determine if the SegAN architecture improves the segmentation results of the U-Net model. Experiments are performed with three variants of 2D U-Net architectures - an original ***U-Net*** [10], the encoder-decoder network used as segmentor in [15] (***U-Net A***) and a modified U-Net inspired from [5] (***U-Net B***), the current state-of-the-art for left ventricle segmentation in the ACDC 2017 dataset.

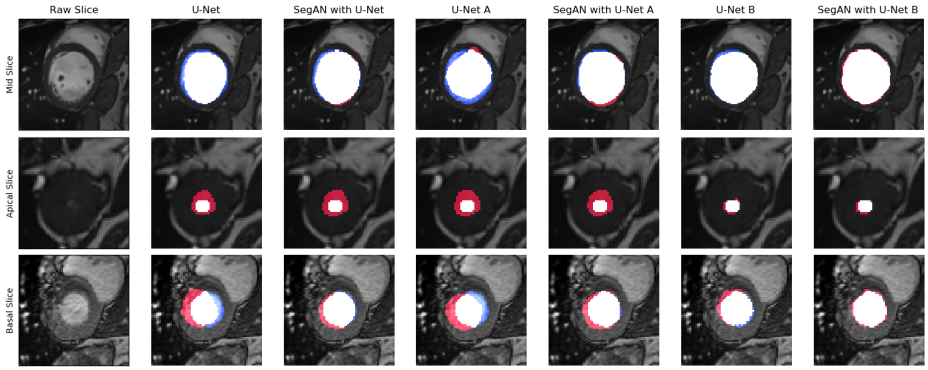

Fig. 3. Examples of segmentation of the left ventricle in mid, apical and basal slice (top to bottom). The white, red and blue regions represent true positives, false positives and false negatives, respectively.

Table 1 summarizes the segmentation performance of the investigated frameworks with and without the SegAN integration. The results are obtained using 80% of the training subjects as training dataset and the 20% of the training subjects as validation dataset.

In Table 1, we can observe that the mean Dice scores and mean IoU values of the three SegAN architectures are higher than their corresponding U-Net models. To compare the performance the three stand-alone U-Net models with their SegAN frameworks, we conducted a statistical significance (T-test) test. The mean Dice score showed significant improvement ($p < 0.05$) from 93.41% (***U-Net***) to 94.71% (***SegAN + U-Net***), ($p < 0.1$) from 92.62% (***U-Net A***) to 93.88% (***SegAN + U-Net A***) and ($p < 0.05$) from 94.91% (***U-Net B***) to 95.87% (***SegAN + U-Net B***) in end diastole, and ($p < 0.1$) from 90.30% (***U-Net A***) to 91.10% (***SegAN + U-Net A***) and ($p < 0.1$) from 92.72% (***U-Net B***) to 93.14% (***SegAN + U-Net B***) in end systole. The mean IoU values showed significant improvement ($p < 0.05$) from 87.25% (***U-Net***) to 89.55% (***SegAN + U-Net***) and ($p < 0.1$) from 91.55% (***U-Net B***) to 92.94% (***SegAN + U-Net B***) in end diastole

The highest mean Dice score and mean IoU in our experiments are obtained using the SegAN architecture with U-Net B as its segmentor network and the U-Net B's encoder as the critic network. The ***SegAN + U-Net B*** outperforms ***U-Net*** by 2.46% (Dice) in end diastole and 1.40% in end systole.

In Fig. 2, it can be observed that the segmentation performance of SegAN frameworks (shown in red) is better than the performance of the corresponding stand-alone U-Net architectures (shown in blue). When we use these U-Net models as segmentor, we see significant improvement in both Dice score and IoU, for ED and ES.

Fig. 3 shows examples of mid, apical and basal slices of the heart and the corresponding segmented masks using the six architectures. The white regions

represent the overlap between the ground truth mask and the tested mask. The red and blue regions represent the false positive (pixels predicted as left ventricle by the tested algorithm, but not annotated in the ground truth), and false negative (pixels not predicted as left ventricle by the tested algorithm, but annotated in the ground truth) regions, respectively.

4 Discussion

In this paper, the integration of three different U-Net models into the SegAN framework is evaluated on the 2017 ACDC segmentation challenge dataset. Our goal was to investigate if the SegAN framework improves the segmentation performance of U-Net models. Our experiments reveal that U-Net models, when trained in the SegAN framework, produces significantly better segmentation results than when trained stand-alone, consistently. The features extracted across multiple layers of the critic network and concatenated into the feature vector used to compute the multi-scale L_1 loss captures pixel-, low-, mid- and high-level features. This multi-resolution approach to feature extraction enables the SegAN model to learn the dissimilarities between the generated and the ground truth segmentation maps across the multiple layers of the critic network.

We use cross entropy loss as cost function for training U-Net models. We also experimented with training all three U-Net variants with a Dice loss cost function, however the results indicated a consistently lower performance than that achieved using cross entropy loss. We also experimented with multi-scale L_2 loss as cost function for training the SegAN models, however, the results were not very consistent. Further investigation with multi-scale L_2 loss as cost function will be conducted, to determine if it can outperform multi-scale L_1 loss as cost function.

To evaluate our method, we used a 5-fold cross-validation strategy, in which we employed five different combinations of 80 training and 20 testing datasets from the available 100 datasets. This is a common approach used to validate novel deep learning techniques, as it enables testing the robustness of the method across different training datasets, while also removing the bias associated with a single 80 training - 20 testing data split.

One of the disadvantage of SegAN is that it needs more computational time than the stand-alone U-Net models. It takes around 110 seconds to run each epoch of a regular U-Net model, but when the same U-Net model is used in the SegAN framework, it needs around 320 seconds.

5 Conclusion and Future Work

SegAN is a promising algorithm that was recently shown to outperform the current state-of-the-art methods in brain tumor segmentation [15]. Moreover, in this work we showed that the SegAN integration alongside three different U-Net variants led to significant improvement in Dice score for left ventricle segmentation.

In light of the improved performance of the SegAN addition in both brain tumor and left ventricle segmentation, further investigation into these methods and their further refinement is justified. As such we will employ the current state-of-the-art methods from the leaderboard of the 2017 ACDC segmentation challenge as segmentor and show their performance using adversarial regularization. One of the variants of SegAN that we would like to experiment in our future work would be to use cross entropy loss as cost function for training the segmentor and multi-scale L_1 loss to train the critic network, instead of using only the multi-scale L_1 loss for both segmentor and critic.

Lastly, we will also aim to answer the question whether this adversarial regularization method may improve the segmentation results of any fully convolutional network when employed as a segmentor along with a critic network.

Acknowledgement

Research reported in this publication was supported by the National Institute of General Medical Sciences of the National Institutes of Health under Award No. R35GM128877 and by the Office of Advanced Cyber-infrastructure of the National Science Foundation under Award No. 1808530.

References

1. Baumgartner, C.F., Koch, L.M., Pollefeys, M., Konukoglu, E.: An exploration of 2d and 3d deep learning techniques for cardiac MR image segmentation. In: International Workshop on Statistical Atlases and Computational Models of the Heart. Lecture notes in Computer Science. vol. 10663, pp. 111–119. Springer (2017)
2. Bernard, O., Lalande, A., Zotti, C., Cervenansky, F., Yang, X., Heng, P.A., Cetin, I., Lekadir, K., Camara, O., Ballester, M.A.G., et al.: Deep learning techniques for automatic MRI cardiac multi-structures segmentation and diagnosis: Is the problem solved? IEEE transactions on medical imaging **37**(11), 2514–2525 (2018)
3. Goodfellow, I., Pouget-Abadie, J., Mirza, M., Xu, B., Warde-Farley, D., Ozair, S., Courville, A., Bengio, Y.: Generative adversarial nets. In: Advances in neural information processing systems. pp. 2672–2680 (2014)
4. Haering, M., Grosshans, J., Wolf, F., Eule, S.: Automated segmentation of epithelial tissue using cycle-consistent generative adversarial networks. bioRxiv p. 311373 (2018)
5. Isensee, F., Jaeger, P.F., Full, P.M., Wolf, I., Engelhardt, S., Maier-Hein, K.H.: Automatic cardiac disease assessment on cine-MRI via time-series segmentation and domain specific features. In: International Workshop on Statistical Atlases and Computational Models of the Heart. Lecture notes in Computer Science. vol. 10663, pp. 120–129. Springer (2017)
6. La, A.G., Claessen, G., de Bruaene Van, A., Pattyn, N., Van, J.C., Gewillig, M., Bogaert, J., Dymarkowski, S., Claus, P., Heidbuchel, H.: Cardiac MRI: a new gold standard for ventricular volume quantification during high-intensity exercise. Circulation. Cardiovascular imaging **6**(2), 329–338 (2013)

7. Lin, X., Cowan, B., Young, A.: Model-based graph cut method for segmentation of the left ventricle. In: Engineering in Medicine and Biology Society, 2005. IEEE-EMBS 2005. 27th Annual International Conference of the. pp. 3059–3062. IEEE (2006)
8. Nasr-Esfahani, M., Mohrekesh, M., Akbari, M., Soroushmehr, S., Nasr-Esfahani, E., Karimi, N., Samavi, S., Najarian, K.: Left ventricle segmentation in cardiac MR images using fully convolutional network. arXiv preprint arXiv:1802.07778 (2018)
9. Petitjean, C., Zuluaga, M.A., Bai, W., Dacher, J.N., Grosgeorge, D., Caudron, J., Ruan, S., Ayed, I.B., Cardoso, M.J., Chen, H.C., et al.: Right ventricle segmentation from cardiac MRI: a collation study. Medical image analysis **19**(1), 187–202 (2015)
10. Ronneberger, O., Fischer, P., Brox, T.: U-net: Convolutional networks for biomedical image segmentation. In: International Conference on Medical image computing and computer-assisted intervention. Lecture Notes in Computer Science. vol. 9351, pp. 234–241. Springer (2015)
11. Santiago, C., Nascimento, J.C., Marques, J.S.: A new ASM framework for left ventricle segmentation exploring slice variability in cardiac MRI volumes. Neural Computing and Applications **28**(9), 2489–2500 (2017)
12. Suinesiaputra, A., Cowan, B.R., Al-Agamy, A.O., Elattar, M.A., Ayache, N., Fahmy, A.S., Khalifa, A.M., Medrano-Gracia, P., Jolly, M.P., Kadish, A.H., et al.: A collaborative resource to build consensus for automated left ventricular segmentation of cardiac MR images. Medical image analysis **18**(1), 50–62 (2014)
13. Suinesiaputra, A., Cowan, B.R., Finn, J.P., Fonseca, C.G., Kadish, A.H., Lee, D.C., Medrano-Gracia, P., Warfield, S.K., Tao, W., Young, A.A.: Left ventricular segmentation challenge from cardiac MRI: a collation study. In: International Workshop on Statistical Atlases and Computational Models of the Heart. Lecture Notes in Computer Science. vol. 7085, pp. 88–97. Springer Berlin Heidelberg, Berlin, Heidelberg (2011)
14. White, H.D., Norris, R.M., Brown, M.A., Brandt, P.W., Whitlock, R., Wild, C.J.: Left ventricular end-systolic volume as the major determinant of survival after recovery from myocardial infarction. Circulation **76**(1), 44–51 (1987)
15. Xue, Y., Xu, T., Zhang, H., Long, L.R., Huang, X.: Segan: Adversarial network with multi-scale l 1 loss for medical image segmentation. Neuroinformatics pp. 1–10 (2018)