A JOINT MULTI-SCALE CONVOLUTIONAL NETWORK FOR FULLY AUTOMATIC SEGMENTATION OF THE LEFT VENTRICLE

Qianqian Tong[1], *Zhiyong Yuan*[1*], *Xiangyun Liao*[2], *Mianlun Zheng*[1], *Weixu Zhu*[1], *Guian Zhang*[1], *Munan Ning*[1]

[1]School of Computer, Wuhan University, Hubei, China
[2]Shenzhen Key Laboratory of Virtual Reality and Human Interaction Technology, Shenzhen
Institutes of Advanced Technology, Chinese Academy of Sciences, Shenzhen, Guangdong, China

ABSTRACT

Left ventricle (LV) segmentation is crucial for quantitative analysis of the cardiac contractile function. In this paper, we propose a joint multi-scale convolutional neural network to fully automatically segment the LV. Our method adopts two kinds of multi-scale features of cardiac magnetic resonance (CMR) images, including multi-scale features directly extracted from CMR images with different scales and multi-scale features constructed by intermediate layers of standard CNN architecture. We take advantage of these two strategies and fuse their prediction results to produce more accurate segmentation results. Qualitative results demonstrate the effectiveness and robustness of our method, and quantitative evaluation indicates our method achieves LV segmentation with higher accuracy than state-of-the-art approaches.

Index Terms— LV segmentation, convolutional neural network, multi-scale representation.

1. INTRODUCTION

Cardiovascular diseases are the leading cause of death globally [1]. Early and accurate diagnoses are essential for the management of these disorders [2]. LV segmentation in CMR images is crucial for calculating ventricular volume and ejection fraction. Manual delineation is time-consuming and tedious, thus it is necessary to perform this task automatically, which helps to improve the efficiency of diagnosis.

Nevertheless, there exist numerous difficulties for automatic LV segmentation in CMR image datasets [3, 4, 5], including gray level inhomogeneity because of the presence of the blood flow, influence of papillary muscles inside the heart chambers, etc. Many researches focus on automatic LV segmentation, including traditional models (image-based methods, deformable models etc.) [5, 6, 7], machine learning models [8, 9], and combined models [3, 4, 10]. Compared with traditional models, machine learning models can automatically adjusting model's parameters without amounts of user interaction. Contours of image slices near the apex are smaller than those at the base. To improve the segmentation accuracy, Avendi et al. [3] divided the training set into the large-contour and small-contour groups and trained two networks, respectively. In fact, it is not easy to distinguish large-contour and small-contour images automatically.

Convolutional Networks (ConvNets) [11, 12, 13] possess biologically-inspired architectures that can automatically learn hierarchical features. However, their ability in scale invariance is limited. To tackle the segmentation difficulty of small-contour images, we propose a joint multi-scale convolutional neural network (JMS_CNN) for fully automatic LV segmentation. JMS_CNN utilizes two kinds of multi-scale features, which are from CMR images of different scales and from different layers of a standard CNN trunk, respectively.

Our contributions mainly lie in: (1) We utilize the multi-scale information of CMR images for LV segmentation, which improves the segmentation accuracy without distinguishing large-contour and small-contour images; (2) We propose the JMS_CNN framework for LV segmentation, which takes advantages of two kinds of feature representation and achieves more accurate results.

2. RELATED WORK

2.1. LV segmentation

There are many methods that have been proposed for LV segmentation in the past decades. Petitjean et al. [5] demonstrated that image-based methods [14, 15] could produce satisfactory results in many cases, while they required amounts of user interaction. Ayed et al. [6] evolved two curves toward the LV endocardium and epicardium boundaries and they derived the curve evolution equations. Hajiaghayi et al. [7] combined internal and external energy terms for active contour to improve the segmentation performance. Dreijer et al. [9] presented a conditional random field (CRF) implementation for the automated LV segmentation. Ngo et al. [10] proposed a semi-automated method combining a level set method with a

[*] Corresponding author: zhiyongyuan@whu.edu.cn. This work was supported by Wuhan Science and Technology Program under Grant No. 2016010101010022 and the National Natural Science Foundation of China under Grant No.61373107

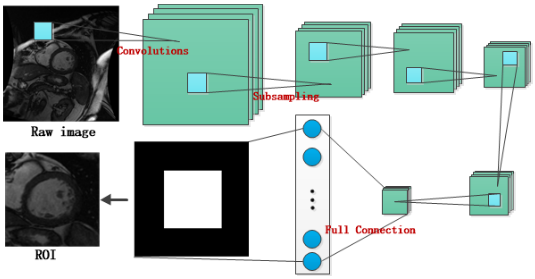

Fig. 1. Diagram of ROI detection network for CMR image.

deep belief network (DBN) for LV segmentation. They extended their work [4] by combining an active contour model with DBN to realize fully automatic LV segmentation. Avendi et al. [3] adopted stacked autoencoders to infer the LV shape and the inferred shape is introduced into deformable models to improve the accuracy and robustness of the segmentation.

2.2. Multi-Scale Convolutional Neural Networks

Many methods have been proposed to make CNNs suitable for segmenting different sizes of images. Sermanet et al. [13] introduced a multi-scale convolutional network for traffic sign recognition, and their network could extract both local and global features. Dou et al. [16] proposed a multi-scale CNN model with a depth-decreasing multi-column structure. Chen et al. [17] proposed an attention mechanism that learns to softly weight the multi-scale features at each pixel location for semantic image segmentation. Cai et al. [18] presented a multi-scale CNN (MS-CNN) for fast multi-scale object detection. Their sub-network performed detection at multiple output layers. The idea to utilize multi-scale features of depth-decreasing multi-scale CNN [16] is different from MS-CNN [18]. The former extracted multi-scale features from images of different scales, and the latter from different layers of a standard CNN trunk. Inspired by these two strategies, we propose a joint multi-scale convolutional neural network by combining multi-scale information extracted from images of different scales and that from different layers of CNN for fully automatic LV segmentation.

3. MATERIAL AND METHOD

The raw CMR imaging datasets usually include the heart and surrounding organs. The first step in our method is to segment a ROI which contains the LV using ConvNets, then the LV contour is delineated by the trained JMS_CNN.

3.1. ROI detection

As shown in Fig. 1, we detect ROI from the raw input images using CNN, which has 7 layers with weights: three convolutional layers, three pooling layers and a fully connected (FC) layer. The response of each convolution layer is given by

$$f_n^{l+1} = \psi(\sum\nolimits_m (f_m^l * h_m^{l+1}) + b_n^{l+1}) \qquad (1)$$

where, f_n^{l+1} and f_m^l are the feature maps of the current layer l and the next layer $l+1$, respectively . h is the convolution kernel. m and n are the number of feature maps of layer l and layer $l+1$. "$*$" denotes the convolution operator. $\psi(\cdot)$ is the activation function and b is the bias scalar. Following each convolutional layer, a max-pooling layer with a downsampling factor of 2 is introduced to select feature subsets.

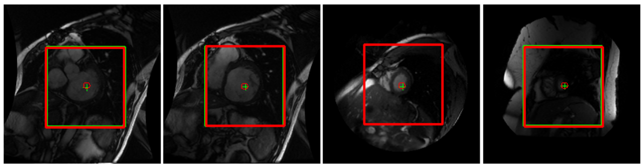

Fig. 2. Results of ROI detection.

Avendi et al. [3] initialized their convolutional network using a sparse auto-encoder (AE), which is used to tackle the difficulty that the number of training and labeled data is limited. In this paper, we augment the training dataset simply using image translation and image rotation. The size of the raw input image is 256 × 256. Firstly, we rotate the image with step of 15 degree, then the rotated image is cropped to 200 × 200 by translating. Thus we augment the training dataset by a factor of 240. Finally, the cropped image is downsampled to 100 × 100, which is conducive to reduce the complexity of models. The results of ROI detection is depicted in Fig. 2. The green square denotes real ROI and the green "+" is its central position. The red square denotes the corresponding predicted ROI, and the red "o" is its central position. We can see that the predicted ROI is very consistent with the real one for the apical slice images and basal slice images.

3.2. Joint multi-scale convolutional neural network

The detailed architecture of our JMS_CNN is shown in Fig.3. To implement multi-scale representation of the input ROI image $\boldsymbol{I}$ and enable the trained network to be suitable for different slice images, we propose two strategies inspired by [16] and [18]. For the first strategy, we utilize different scales of the input image to train S networks in parallel, where S is the scale number in total. We refer to the S CNNs as P_CNN$_1$, P_CNN$_2$,. . ., P_CNN$_S$ because of its parallel characteristics. To improve the robustness of the entire network, we do not train the CNNs separately, but train them assembly and simultaneously. Supposing that the fully connected layer of P_CNN$_S$ is $\boldsymbol{fc}_S$, the fully connected layers of all scales are concatenated as $\boldsymbol{fc}$ and it is used to final prediction.

$$\boldsymbol{fc} = [\boldsymbol{fc}_1, \boldsymbol{fc}_2, \cdots, \boldsymbol{fc}_S] \qquad (2)$$

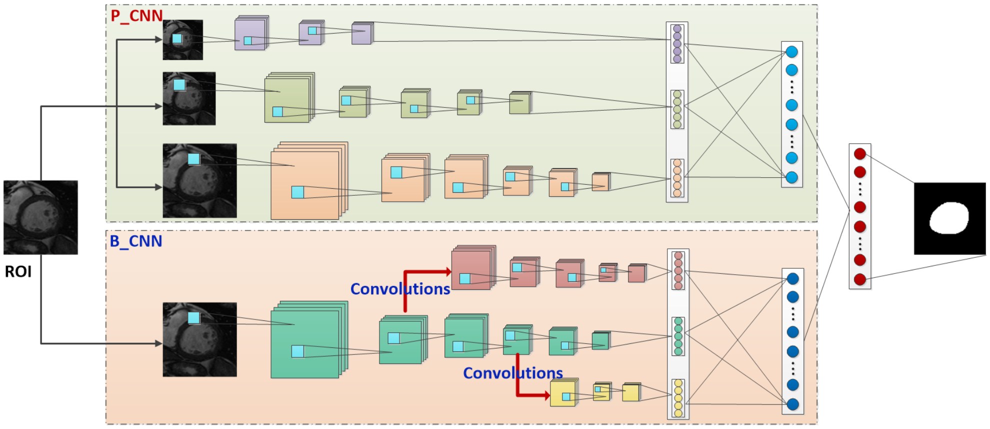

Fig. 3. Overview of the proposed joint multi-scale convolutional neural network.

During the training stage of the CNNs, square error is used as the cost function. In our application, we implement end-to-end prediction for all pixels in $\boldsymbol{I}$. For the n^{th} training sample, the cost function J^n is denoted as:

$$J^n = \frac{1}{2}\sum_{i=1}^{m1}\sum_{j=1}^{n1}(p_{i,j}^n - y_{i,j}^n)^2 = \frac{1}{2}\|\boldsymbol{p}^n - \boldsymbol{y}^n\|_2^2 \qquad (3)$$

where, $[m1, n1]$ is the output image size, $\boldsymbol{p}^n$ and $\boldsymbol{y}^n$ are the predicted mask and the ground truth mask of the n^{th} training sample, respectively. When $y_{i,j}^n = 1$, the pixel $[i,j]$ belongs to the region of LV. Supposing that the feature vector error of the tail perceptron is $\boldsymbol{fv}$, the length of fully connected layer $\boldsymbol{fc}_s$ is $l_s(s = 1, 2, \cdots, S)$, $\boldsymbol{fv}$ is calculated by assembling all the multi-scale P_CNNs. In the stage of back propagation, we update weights of each P_CNN separately. The feature vector error of the tail perceptron for the P_CNNs is $\boldsymbol{fv}_s$

$$\boldsymbol{fv}_s = \Phi(\boldsymbol{fv}(l_s)) - \Phi(\boldsymbol{fv}(l_{s-1})) \qquad (4)$$

where, $\boldsymbol{fv}(l_s)$ is the feature vector error of the first s scales. $\Phi(\cdot)$ denotes the elements of vector. The operation "$-$" denotes the removal of the elements on the corresponding position of the second vector from the first vector.

The second strategy just uses a single scale of the input image. This multi-scale network has a trunk network, which has a standard CNN architecture. To obtain variable receptive field sizes and improve the ability to tackle the segmentation task of different object sizes, we utilize the feather maps emanating from intermediate layers of the trunk network to accomplish multi-scale representation. Thus the trunk network has many branches, and we called the entire network as B_CNN. For each branch, its first convolutional layer follows the pooling layer of the trunk network. Different branches are referred to as B_CNN$_1$, B_CNN$_2$, . . . , B_CNN$_D$, where B_CNN$_1$ denotes the trunk network and D is the numbers of branches in the trunk network. The ways of concatenation and training are the same with P_CNN as described above.

It is worth noting that P_CNN and B_CNN both accomplish multi-scale representation of the input image $\boldsymbol{I}$. In this paper, we fuse the predicted results of P_CNN and B_CNN for each pixel. Supposing that $\boldsymbol{I}_p(i,j)$ and $\boldsymbol{I}_B(i,j)$ are the predicted results of P_CNN and B_CNN for pixel $[i,j]$, the binary mask of the input ROI image is calculated as follows

$$B_{mask} = \mathop{\Gamma}_{i=1,j=1}^{i=m1,j=n1}(average(\mathbf{I}_P(i,j), \mathbf{I}_B(i,j))) \qquad (5)$$

where, $average(\cdot)$ is the average function. $[m1, n1]$ is the size of the predicted binary mask B_{mask}. $\Gamma(\cdot)$ denotes to match the binary results to the input ROI image so as to obtain the contour of the LV.

To further improve the LV segmentation, we utilize morphological operations including opening, dilation and erosion to eliminate isolated small regions and smooth lesion edges, which is usually used in segmentation task [19].

4. EXPERIMENTAL RESULTS

4.1. Datasets and experimental settings

We assess the performance of our method on the MICCAI 2009 challenge database [20]. The database contains three data (training, validation and online) sets and manual segmentation by experienced experts is included in the database. In this paper, we focus on the endocardium segmentation. We augment the training dataset using the method in section 3.1.

To evaluate the performance of the proposed JMS_CNN, we set the scale number of P_CNN is $S = 3$. The input image size of the three scales is 100×100, 50×50 and 25×25. The down-sampling factor of all the P_CNNs and B_CNNs is 2. The settings of convolutional layers are as follows: P_CNN$_1$ $\{6(5\times5)$, $10(5\times5)$, $12(5\times5)\}$, P_CNN$_2$ $\{6(5\times5)$, $10(4\times4)$, $12(3\times3)\}$, P_CNN$_3$ $\{6(4\times4)$, $10(4\times4)\}$, B_CNN$_1$ $\{6(5\times5)$, $10(5\times5)$, $12(5\times5)\}$, B_CNN$_2$ $\{6(5\times5)$, $10(3\times3)$, $12(3\times3)\}$ and B_CNN$_3$ $\{6(5\times5)$, $10(3\times3)\}$. For $6(5\times5)$, 6 is the numbers of the convolutional kernels 5×5.

4.2. Segmentation results

We assess the performance of our method by comparing its segmentation contours with that of manual annotation by experts. The automatic and manual LV segmentation results are shown in Fig.4, and yellow line indicates better matching. It can be seen that the LV are accurately segmented even if the size of the LV is very small, which is depicted in the third row of Fig.4. In addition, our JMS_CNN can overcome the difficulties of gray level inhomogeneity, presence of papillary muscles and influence of surrounding areas, which can be observed in the first image of the third row, the first image of the first row and the third image of the first row, respectively.

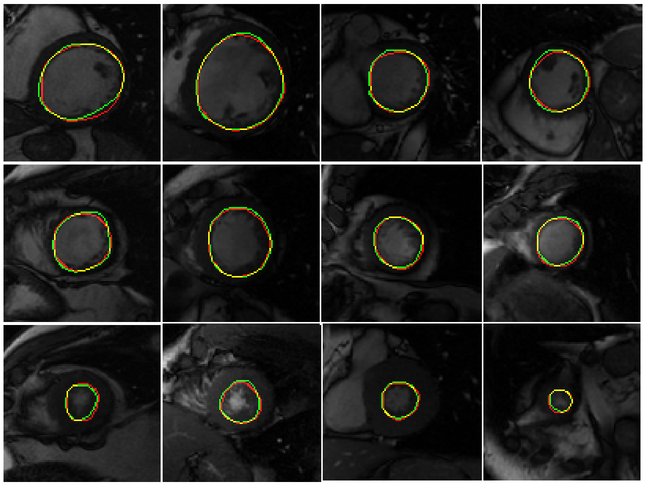

Fig. 4. Comparison of the automatic segmentation contour (red line) and the manual contour (green line) of the LV.

4.3. Quantitative analysis

In order to further assess the performance of our method, we evaluate the following four metrics: 1) percentage of good contours (GC), 2) Dice metric (DM), 3) average perpendicular distance (APD), 4) the conformity coefficient (CC). The APD measures the distance from the automatically segmented contour to the corresponding manually drawn expert contour, averaged over all contour points. A segmentation result is classified as good contour if the APD is less than 5mm [20]. A high value of APD implies that the two contours do not match closely. DM measures contour overlap considering the contour areas automatically segmented A_a, manually segmented A_m, and their intersection A_{am}. $\mathrm{DM} = 2(A_{am})/(A_a + A_m)(0 \leq \mathrm{DM} \leq 1)$ and higher DM indicates better match between automated and manual segmentation results. CC measures the ratio of the number of mis-segmented pixels to the number of correctly segmented pixels and it is defined as $\mathrm{CC} = (3\mathrm{DM} - 2)/\mathrm{DM}$ [21].

We compare the segmentation performance between our JMS_CNN and other machine learning methods, including: Avendi et al. [3], standard CNN that is single scale CNN (S_CNN), P_CNN and B_CNN. P_CNN and B_CNN are described in section 3.2. In Table 1, the average values and the standard deviation of the computed metrics are listed for the same validation (V) and online (O) datasets (DT). Table 2 compares the general segmentation performance between our method and the other four machine learning methods. Table 1 and Table 2 show that the proposed JMS_CNN outperforms four other machine learning methods on all the metrics. Our percentage of good contours is 96.80, which is much higher than Avendi et al. [3]. It is worth noting that S_CNN cannot detect all the CMR images, and it fails to detect some images containing smaller LV. As expected, the proposed JMS_CNN can deal with LV segmentation of different sizes thus to achieve better "good contours" and "APD" by taking advantages of two kinds of multi-scale feature representation.

Table 1. Comparison of the evaluation metrics between JMS_CNN and other methods for validation and online datasets.

Method	DT	GC(%)	APD	DM	CC
[3]	V	90(10)	2.84(0.29)	0.90(0.10)	0.78(0.03)
	O	87(12)	2.95(0.54)	0.89(0.03)	0.76(0.07)
S_CNN	V	97.10(5.03)	2.12(0.28)	0.91(0.03)	0.81(0.07)
	O	92.83(10.14)	2.21(0.49)	0.91(0.04)	0.80(0.10)
P_CNN	V	96.15(6.13)	2.12(0.30)	0.91(0.03)	0.80(0.08)
	O	93.86(9.40)	2.18(0.54)	0.91(0.03)	0.80(0.09)
B_CNN	V	97.03(5.75)	2.12(0.27)	0.91(0.03)	0.81(0.07)
	O	94.10(9.67)	2.15(0.40)	0.91(0.03)	0.81(0.07)
Proposed	V	**97.86(4.67)**	**2.04(0.28)**	**0.91(0.03)**	**0.81(0.09)**
	O	**95.79(7.10)**	**2.11(0.53)**	**0.91(0.03)**	**0.81(0.09)**

Table 2. Comparison of the general segmentation performance between JMS_CNN and other methods.

Method	GC(%)	APD	DM	CC
[3]	88.46	2.90	0.89	0.77
S_CNN	94.92	2.16	0.91	0.80
P_CNN	94.98	2.15	0.91	0.80
B_CNN	95.53	2.13	0.91	0.81
Proposed	**96.80**	**2.08**	**0.91**	**0.81**

5. CONCLUSION

In this paper, we proposed a joint multi-scale convolutional neural network for accurate segmentation of the LV in CMR images. The proposed method can cover different LV sizes by utilizing multi-scale representation not only from images of different scales but also that from multiple intermediate layers of a standard CNN architecture. The experimental results demonstrate that our method can yield accurate segmentation. In the future work, we will try to accomplish multi-scale representation on other networks such as fully convolutional networks and evaluate our method on larger data sets.

6. REFERENCES

[1] Shanthi Mendis, Pekka Puska, Bo Norrving, et al., *Global atlas on cardiovascular disease prevention and control.*, World Health Organization, 2011.

[2] Hamid Chalian, James K ODonnell, Michael Bolen, and Prabhakar Rajiah, "Incremental value of pet and mri in the evaluation of cardiovascular abnormalities," *Insights into imaging*, vol. 7, no. 4, pp. 485–503, 2016.

[3] MR Avendi, Arash Kheradvar, and Hamid Jafarkhani, "A combined deep-learning and deformable-model approach to fully automatic segmentation of the left ventricle in cardiac mri," *Medical image analysis*, vol. 30, pp. 108–119, 2016.

[4] Tuan Anh Ngo and Gustavo Carneiro, "Fully automated non-rigid segmentation with distance regularized level set evolution initialized and constrained by deep-structured inference," in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, 2014, pp. 3118–3125.

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

[6] Ismail Ben Ayed, Shuo Li, and Ian Ross, "Embedding overlap priors in variational left ventricle tracking," *IEEE Transactions on Medical Imaging*, vol. 28, no. 12, pp. 1902–1913, 2009.

[7] Mahdi Hajiaghayi, Elliott M Groves, Hamid Jafarkhani, and Arash Kheradvar, "A 3-d active contour method for automated segmentation of the left ventricle from magnetic resonance images," *IEEE Transactions on Biomedical Engineering*, vol. 64, no. 1, pp. 134–144, 2017.

[8] Ján Margeta, Ezequiel Geremia, Antonio Criminisi, et al., "Layered spatio-temporal forests for left ventricle segmentation from 4d cardiac mri data," in *International Workshop on Statistical Atlases and Computational Models of the Heart*. Springer, 2011, pp. 109–119.

[9] Janto F Dreijer, Ben M Herbst, and Johan A Du Preez, "Left ventricular segmentation from mri datasets with edge modelling conditional random fields," *BMC medical imaging*, vol. 13, no. 1, pp. 24, 2013.

[10] Tuan Anh Ngo and Gustavo Carneiro, "Left ventricle segmentation from cardiac mri combining level set methods with deep belief networks," in *Image Processing (ICIP), 2013 20th IEEE International Conference on*. IEEE, 2013, pp. 695–699.

[11] Yann LeCun, Yoshua Bengio, and Geoffrey Hinton, "Deep learning," *Nature*, vol. 521, no. 7553, pp. 436–444, 2015.

[12] Kevin Jarrett, Koray Kavukcuoglu, Yann LeCun, et al., "What is the best multi-stage architecture for object recognition?," in *Computer Vision, 2009 IEEE 12th International Conference on*. IEEE, 2009, pp. 2146–2153.

[13] Pierre Sermanet and Yann LeCun, "Traffic sign recognition with multi-scale convolutional networks," in *Neural Networks (IJCNN), The 2011 International Joint Conference on*. IEEE, 2011, pp. 2809–2813.

[14] Su Huang, Jimin Liu, Looi Chow Lee, et al., "An image-based comprehensive approach for automatic segmentation of left ventricle from cardiac short axis cine mr images," *Journal of digital imaging*, vol. 24, no. 4, pp. 598–608, 2011.

[15] Hong Liu, Huaifei Hu, Xiangyang Xu, and Enmin Song, "Automatic left ventricle segmentation in cardiac mri using topological stable-state thresholding and region restricted dynamic programming," *Academic radiology*, vol. 19, no. 6, pp. 723–731, 2012.

[16] Haobin Dou and Xihong Wu, "Coarse-to-fine trained multi-scale convolutional neural networks for image classification," in *Neural Networks (IJCNN), 2015 International Joint Conference on*. IEEE, 2015, pp. 1–7.

[17] Liang-Chieh Chen, Yi Yang, Jiang Wang, Wei Xu, and Alan L Yuille, "Attention to scale: Scale-aware semantic image segmentation," in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, 2016, pp. 3640–3649.

[18] Zhaowei Cai, Quanfu Fan, Rogerio S Feris, and Nuno Vasconcelos, "A unified multi-scale deep convolutional neural network for fast object detection," in *European Conference on Computer Vision*. Springer, 2016, pp. 354–370.

[19] Yanran Wang, Aggelos K Katsaggelos, Xue Wang, and Todd B Parrish, "A deep symmetry convnet for stroke lesion segmentation," in *Image Processing (ICIP), 2016 IEEE International Conference on*. IEEE, 2016, pp. 111–115.

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

[21] Herng-Hua Chang, Audrey H Zhuang, Daniel J Valentino, and Woei-Chyn Chu, "Performance measure characterization for evaluating neuroimage segmentation algorithms," *Neuroimage*, vol. 47, no. 1, pp. 122–135, 2009.