Multiple Convolutional Neural Networks for Robust Myocardial Segmentation

Klas Berggren*, Erik Hedström*†, Katarina Steding Ehrenborg*, Marcus Carlsson*,
Henrik Engblom*, Ellen Ostenfeld*, Jonas Jögi*, Dan Atar‡, Ulf Ekelund§, Håkan Arheden*, Einar Heiberg*¶
*Department of Clinical Sciences Lund, Clinical Physiology, Lund University, Skane University Hospital, Lund, Sweden
†Department of Clinical Sciences Lund, Diagnostic Radiology, Lund University, Skane University Hospital, Lund, Sweden
‡Department of Cardiology, Oslo University Hospital Ullevål, and
Institute of Clinical Sciences, University of Oslo, Oslo, Norway
§Emergency Medicine, Department of Clinical Sciences Lund, Lund University, Skane University Hospital, Lund, Sweden
¶Wallenberg Center for Molecular Medicine, Lund University, Lund, Sweden

***Abstract*—Delineation of the left ventricle in cardiac MRI images is time consuming task when performed manually. Deep convolutional neural networks have shown excellent results in performing left ventricle segmentation. But the datasets used are often limited in terms of variability. In this work we used multiple convolutional neural networks, trained on a highly heterogenous cohort, to obtain a robust segmentation model.**

I. Introduction

Cardiovascular disease is the most common cause of death [1]. In clinical practice cardiovascular magnetic resonance imaging (MRI) is used to obtain quantitative measures of cardiac function. The most important parameters are based on endocardial and epicardial delineations of the left and right ventricles. These delineations are commonly performed in short-axis stacks which consist of images slicing the heart from base to apex. Manually segmenting these stacks is a time consuming task. Convolutional neural networks have shown excellent results in performing this segmentation [2] but may be limited to homogenous data, without variability in field of view, resolution, number of slices, number of timeframes, pathology, and different imaging sequences between vendors.

Therefore, the aim was to generate a robust left ventricular segmentation tool by coupling three networks for a) myocardial detection, b) bounding-box estimation and c) segmentation trained on a multi-vendor, multi-site dataset with a large variability in pathology.

II. Methods

The segmentation algorithm was implemented in the freely available software Segment [3]. Code was written in Matlab and training was performed on an Nvidia Titan RTX. All models were trained using the Adam [4] optimizer. All convolutional layers have a kernel size 3 otherwise.

A. Material

Two datasets were used for training, a research dataset and a clinical dataset. Research data included images delineated in a number of research projects. The dataset contained healthy subjects [5], athletes [5], subjects from three multicentre trials (CHILL-MI, MITOCARE, SOCCER) [6], [7], [8], patients with pulmonary hypertension, cardiac syndrome X, atrial septal defects [9], systemic sclerosis [10], and a broad range of other diagnoses. The clinical dataset consisted of routine cardiovascular examinations at Skåne University Hospital, Lund, Sweden.

A population of $n = 49$ was used as a test set from [11]. These patients were carefully delineated in consensus by three experienced observers in the end-diastolic (ED) and end-systolic (ES) timeframes.

B. Slice selection

The slice selection network was trained using the research dataset. The clinical dataset was not included in training the slice selection network due to unknown consistency in slice selection. Slices were automatically labelled based on existing manual delineations and divided into three classes. The classes being: above, in or below the left ventricle. The model used was based on the Darknet-19 architecture [12]. The layers past and including the 1000 filter convolutional layer was replaced. Instead the network ended with a 3 filter convolutional layer followed by a global averaging layer with softmax activation. This resulted in a pseudo-probability output of the image being in any of the three classes. Training was performed for 120 epochs using cross entropy loss.

C. Myocardial bounding box

Both the research and clinical datasets were used. Clinical images were augmented in scale and rotation, and horizontal flipping was applied. Bounding boxes were derived from manual segmentation. This model was also based on the Darknet-19 architecture [12]. In the same manner as for myocardial detection, the last layers were replaced by a 4 filter convolutional layer followed by a global averaging layer with ReLU activation, yielding 4 output values corresponding to the parameters of a bounding box. The network loss function L_{bb} was defined as

$$L_{bb}(X, \hat{X}) = \frac{1}{n_b} \sum_{i}^{n_b} \sum_{i}^{4} (\hat{X} - X)^2 \qquad (1)$$

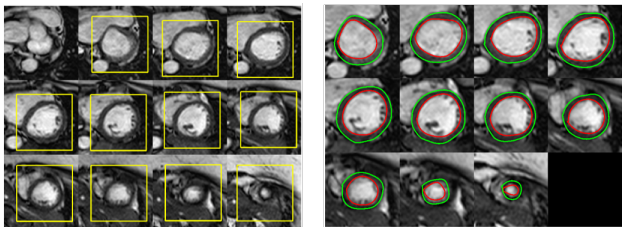

Fig. 1. The left image shows slice selection by the the myocardial detection network. Subsequently the myocardial bounding box network found a region containing the myocardium in all the slices shown as a yellow box. The image within the bounding box network were fed to the segmentation network which generated the contours shown in the right image.

where $\hat{X}$ contains the network-estimated ventricle centre, myocardial height and width making up the bounding box. The variable X denotes the ground truth values. The model was trained for 100 epochs.

D. Myocardial segmentation

Images used for training the segmentation network were cropped to match square bounding boxes covering the largest manually segmented slice with an additional $10-50\%$. The network architecture proposed by Bai et. al [2] was applied to myocardial segmentation. The model was trained for 30 epochs using binary cross entropy loss.

E. Full pipeline

To summarize; first the slice selection network determined slices that should be segmented. These slices were fed to the bounding box network. After a bounding box was detected it was made square by using only the longest side of the suggested bounding box. The square bounding box sides were increased by 30% in length to assure that the entire myocardium was included. Thereafter the image was cropped according to the bounding box and used as input to the segmentation network. The masks generated by the segmentation network were post-processed, picking out the largest joint segmented region. This region was converted into endocardial and epicardial contours shown as red and green lines in Fig. 3. A second degree Savitsky Golay filter with filter length 45 was applied for smoothing the contours [13]. The entire procedure is described in Fig. 1.

F. Evaluation

To evaluate the model the clinical measures left ventricular mass (LVM), end-diastolic volume (EDV) and end-systolic volume (ESV) were used. The latter two are the ventricular volumes of the heart in its most relaxed and most contracted state, respectively. Dice score for delineation match between network and experts was computed. For myocardial detection accuracy and number of missed or extra slices included was computed.

III. Results

The slice selection network had an accuracy of 0.92 in end-diastole and 0.90 in end-systole. Fig. 2 shows the error in slice selection, where missed or extra slices are shown as negative and positive bars, respectively. The slice selection framework was prone to discarding slices. End-diastolic and end-systolic volumes were slightly underestimated (Table I). Automatic segmentation showed less bias and standard deviation in end-systolic volume compared to end-diastolic volume (Table I). Fig. 3 and Fig. 4 show the test cases with the largest error in terms of end-diastolic volume and left ventricular mass. In Fig. 5, 6, 7, 8, 9, 10 regression analysis and Bland Altman comparisons of the manual and automatic method are shown. In Table II Dice similarity coefficient is shown.

TABLE I
Mean $\pm$ SD of cardiac measures for manual and automatic segmentation.

	Automatic-Manual	Automatic	Manual
EDV [ml]	-9 ± 9	183 ± 39	192 ± 41
ESV [ml]	-4 ± 7	77 ± 27	81 ± 29
LVM [g]	1 ± 9	109 ± 32	108 ± 28

TABLE II
Mean $\pm$ SD Dice score for manual and automatic segmentation.

	Endo	Epi	Myo
DSC	0.91 ± 0.04	0.94 ± 0.03	0.82 ± 0.04

IV. Discussion

Previous methods achieved absolute differences in end-diastolic volume -11 ± 11 ml, end-systolic volume 1 ± 10 ml and left ventricular mass 4 ± 15 g [11] on the same test set the proposed model was applied to, which can be compared to the first column in Table I. The left ventricular mass error in Fig. 4 was positive and not related to the missing apical slice. It is therefore likely related to inclusion of too much papillary muscle/trabeculation. The error in end-diastolic volume for the case shown in Fig. 3 was negative. This indicates that it is mainly due to the missing basal slice.

V. Conclusions

An automated left ventricular segmentation algorithm based on deep convolutional neural networks was implemented. Comparing to the previous algorithm applied to the test set [11], the proposed model performed on par in terms of bias with a slightly lower standard deviation in all cardiac measures when comparing absolute difference between manual and automatic methods. Furthermore it requires no user input and is designed to handle variable field of views, number of slices and pathologies.

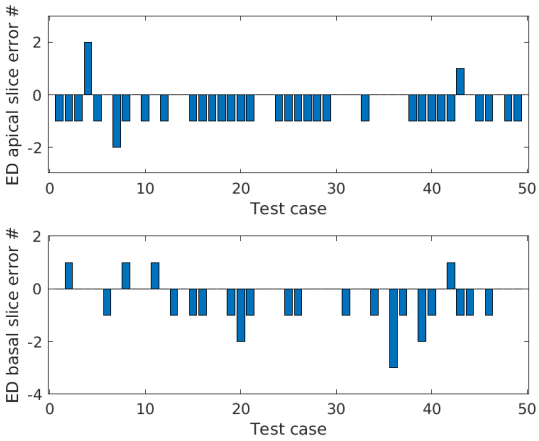

Fig. 2. Apical and basal slice differences for each subject in the test set. A negative number refers to a missing apical or basal slice and a positive means that an extra slice was included. The plot shows that the method is prone to excluding both basal and apical slices.

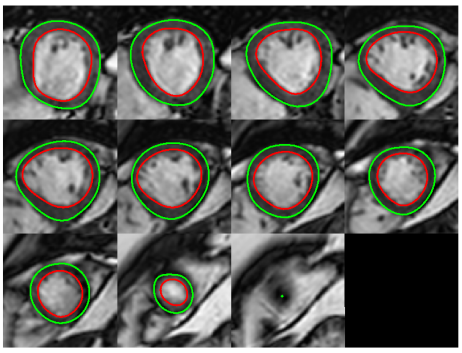

Fig. 3. Test set patient with largest error in terms of end-diastolic volume. This is likely due to that the network missed to include one basal slice. The segmented apex is also too small.

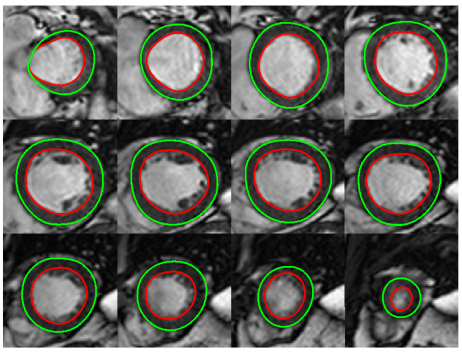

Fig. 4. Test set patient with largest error in terms of left ventricular mass. The error is probably due to including too much papillary muscle/trabeculation.

REFERENCES

[1] "World health organisation fact sheet: Cardiovascular diseases," 2017.

[2] W. Bai, M. Sinclair, G. Tarroni, O. Oktay, M. Rajchl, G. Vaillant, A. M. Lee, N. Aung, E. Lukaschuk, M. M. Sanghvi *et al.*, "Automated cardiovascular magnetic resonance image analysis with fully convolutional networks," *Journal of Cardiovascular Magnetic Resonance*, vol. 20, p. 65, 2018.

[3] E. Heiberg, J. Sjögren, M. Ugander, M. Carlsson, H. Engblom, and H. Arheden, "Design and validation of segment-freely available software for cardiovascular image analysis," *BMC medical imaging*, vol. 10, no. 1, p. 1, 2010.

[4] D. P. Kingma and J. Ba, "Adam: A method for stochastic optimization," *arXiv preprint arXiv:1412.6980*, 2014.

[5] K. Steding, H. Engblom, T. t. Buhre, M. Carlsson, H. Mosén, B. Wohlfart, and H. Arheden, *Journal of Cardiovascular Magnetic Resonance*, vol. 12, no. 1, p. 8, 2010.

[6] D. Atar, H. Arheden, A. Berdeaux, J.-L. Bonnet, M. Carlsson, P. Clemmensen, V. Cuvier, N. Danchin, J.-L. Dubois-Randé, H. Engblom *et al.*, "Effect of intravenous tro40303 as an adjunct to primary percutaneous coronary intervention for acute st-elevation myocardial infarction: Mitocare study results," *European heart journal*, vol. 36, no. 2, pp. 112–119, 2015.

[7] A. Khoshnood, M. Carlsson, M. Akbarzadeh, P. Bhiladvala, A. Roijer, D. Nordlund, P. Höglund, D. Zughaft, L. Todorova, A. Mokhtari *et al.*, "Effect of oxygen therapy on myocardial salvage in st elevation myocardial infarction: the randomized soccer trial," *European Journal of Emergency Medicine*, vol. 25, no. 2, pp. 78–84, 2018.

[8] D. Erlinge, M. Götberg, I. Lang, M. Holzer, M. Noc, P. Clemmensen, U. Jensen, B. Metzler, S. James, H. E. Bötker *et al.*, "Rapid endovascular catheter core cooling combined with cold saline as an adjunct to percutaneous coronary intervention for the treatment of acute myocardial infarction: the chill-mi trial: a randomized controlled study of the use of central venous catheter core cooling combined with cold saline as an adjunct to percutaneous coronary intervention for the treatment of acute myocardial infarction," *Journal of the American College of Cardiology*, vol. 63, no. 18, pp. 1857–1865, 2014.

[9] S. S. Stephensen, E. Ostenfeld, K. Steding-Ehrenborg, U. Thilén, E. Heiberg, H. Arheden, and M. Carlsson, "Alterations in ventricular pumping in patients with atrial septal defect at rest, during dobutamine stress and after defect closure," *Clinical physiology and functional imaging*, vol. 38, no. 5, pp. 830–839, 2018.

[10] T. Gyllenhammar, M. Kanski, H. Engblom, D. M. Wuttge, M. Carlsson, R. Hesselstrand, and H. Arheden, "Decreased global myocardial perfusion at adenosine stress as a potential new biomarker for microvascular disease in systemic sclerosis: a magnetic resonance study," *BMC cardiovascular disorders*, vol. 18, no. 1, p. 16, 2018.

[11] J. Tufvesson, E. Hedström, K. Steding-Ehrenborg, M. Carlsson, H. Arheden, and E. Heiberg, "Validation and development of a new automatic algorithm for time-resolved segmentation of the left ventricle in magnetic resonance imaging," *BioMed research international*, vol. 2015, 2015.

[12] J. Redmon and A. Farhadi, "Yolo9000: better, faster, stronger," in *Proceedings of the IEEE conference on computer vision and pattern recognition*, 2017, pp. 7263–7271.

[13] A. Savitzky and M. J. Golay, "Smoothing and differentiation of data by simplified least squares procedures." *Analytical chemistry*, vol. 36, no. 8, pp. 1627–1639, 1964.

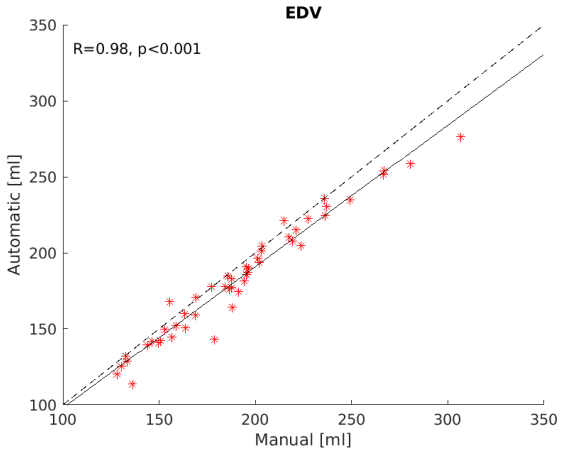

Fig. 5. Comparison of end-diastolic volume between automatic and manual segmentation methods. Dashed line is the line of identity and solid line is the regression line.

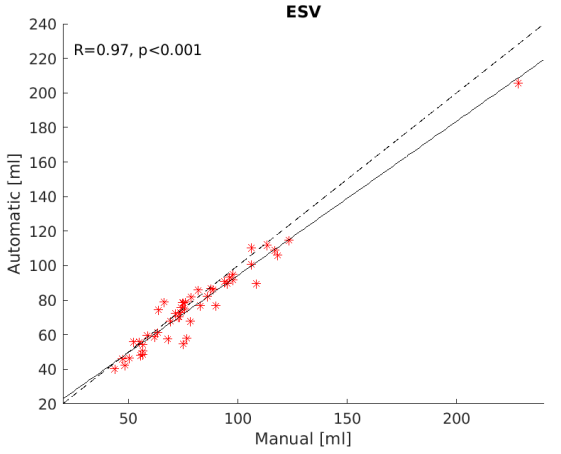

Fig. 6. Comparison of end-systolic volume between automatic and manual segmentation methods. Dashed line is the line of identity and solid line is the regression line.

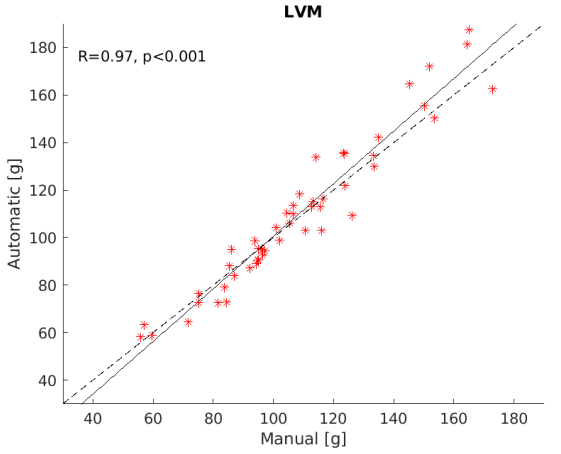

Fig. 7. Comparison of left ventricular mass between automatic and manual segmentation methods. Dashed line is the line of identity and solid line is the regression line.

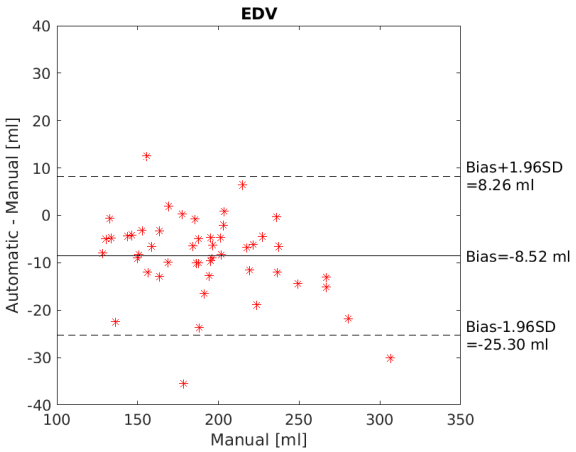

Fig. 8. Bland-Altman comparison of end-diastolic volumes from manual and automatic segmentation.

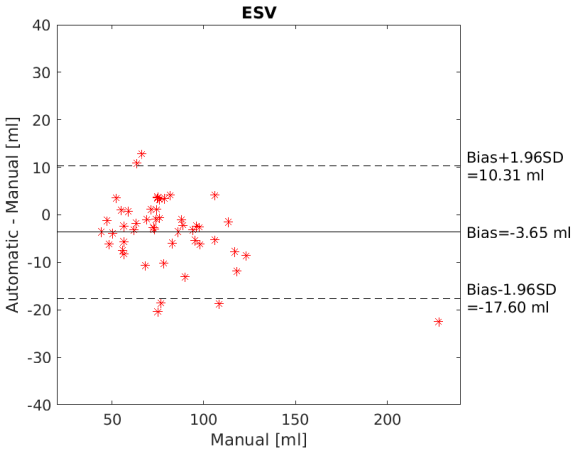

Fig. 9. Bland-Altman comparison between end-systolic volumes from manual and automatic segmentation.

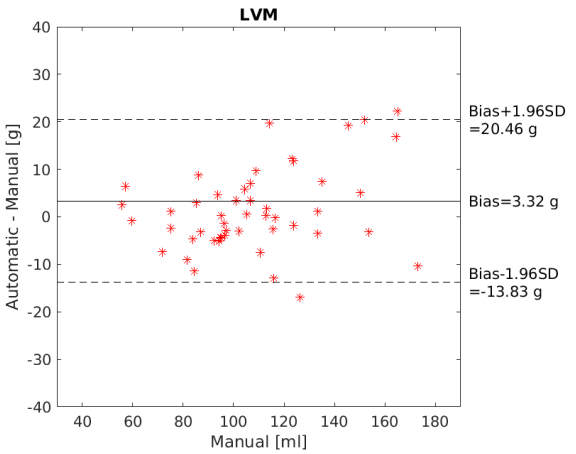

Fig. 10. Bland-Altman comparison of left ventricular mass from manual and automatic segmentation.