2D left ventricle segmentation using deep learning

Erik Smistad
Centre for Innovative Ultrasound Solutions (CIUS)
Norwegian University of Science and Technology (NTNU)
Trondheim, Norway
ersmistad@gmail.com

Andreas Østvik
Centre for Innovative Ultrasound Solutions (CIUS)
Norwegian University of Science and Technology (NTNU)
Trondheim, Norway
andreas.ostvik@ntnu.no

Bjørn Olav Haugen
Centre for Innovative Ultrasound Solutions (CIUS)
Norwegian University of Science and Technology (NTNU)
Trondheim, Norway
bjorn.o.haugen@ntnu.no

Lasse Løvstakken
Centre for Innovative Ultrasound Solutions (CIUS)
Norwegian University of Science and Technology (NTNU)
Trondheim, Norway
lasse.lovstakken@ntnu.no

***Abstract*—Automatic segmentation of the left ventricle (LV) can become a useful tool in echocardiography. Deep convolutional neural networks (CNNs) have shown promising results for image classification and segmentation on several domains, however CNNs seem to require a lot of training data. In this work, CNNs are investigated for LV ultrasound image segmentation. We study if the need for manual annotation can be reduced by pretraining a CNN using a previously published automatic Kalman filter (KF) based segmentation method. The results show that a CNN is able to achieve similar accuracy to that of the automatic method, by only training with generated data. The dice similarity coefficient was measured to be 0.86 ± 0.06 for the CNN versus 0.87 ± 0.06, while the Hausdorff distance was better at 5.9 ± 2.9 mm for the CNN versus 7.5 ± 5.6 mm for the KF method. In future work, this may enable CNNs to exceed state-of-the-art with a small set of expert annotations for fine-tuning.**

***Keywords*—left ventricle, segmentation, deep learning, neural networks**

I. Introduction

Automatic segmentation of the left ventricle (LV) can become a useful tool in echocardiography, for instance to provide automatic ejection fraction measurements or to initialize deformation imaging algorithms.

There are many proposed methods for 2D left ventricle segmentation, such as active contours, level sets, active shape models and Kalman filter. A review of such methods was conducted by Noble and Boukerroui [1]. Neural network based approaches have also been proposed. Carneiro et al. [2] proposed using deep belief networks to identify the LV region of interest and adapt a spline to edges. Deep convolutional neural networks have recently shown very promising results for improving image classification and segmentation. These methods learn using only a set of input and output data, but may require a large and representative amount of annotated data to be successful. This means an expert has to draw the LV border in potentially thousands of images, which is highly tedious and time consuming. Oktay et al. [3] demonstrated the use of these neural networks on 3D left ventricle segmentation. They solved the issue of limited training data by regularizing the training with an accurate anatomical 3D model created from a large database of annotated cardiac magnetic resonance images.

In this work, we investigate deep convolutional neural networks for segmentation of the LV from 2D ultrasound images. More specifically, we study if the need for manual annotation can be reduced by pretraining a deep convolutional neural network (CNN) using an automatic Kalman filter (KF) segmentation approach as the teacher. The hypothesis is that the CNN is able to achieve comparable accuracy with the KF method.

II. Methodology

This section first describes the dataset and how it was processed to generate training data for the neural network. Next, a description of the neural network architecture and training is provided. Finally, the evaluation procedure is described.

A. Dataset

Over 1,500 ultrasound recordings of 100 patients referenced to the outpatient clinic was collected. These recordings consisted of about 100,000 2D image frames acquired through the apical window.

The recordings were processed with an automatic model-based segmentation method [5] adapted to 2D. This segmentation method models the left ventricle as a cubic hermite spline. The spline may be translated, rotated and scaled, and each control point may move independently to enable local deformation. All these transformation state parameters are estimated and predicted over time using a Kalman filter (KF). The state is updated using edge detection measurements along the normal of several points on the spline. This method outputs a label image for each input ultrasound image frame. The

DOI: 10.1109/ULTSYM.2017.8092573.

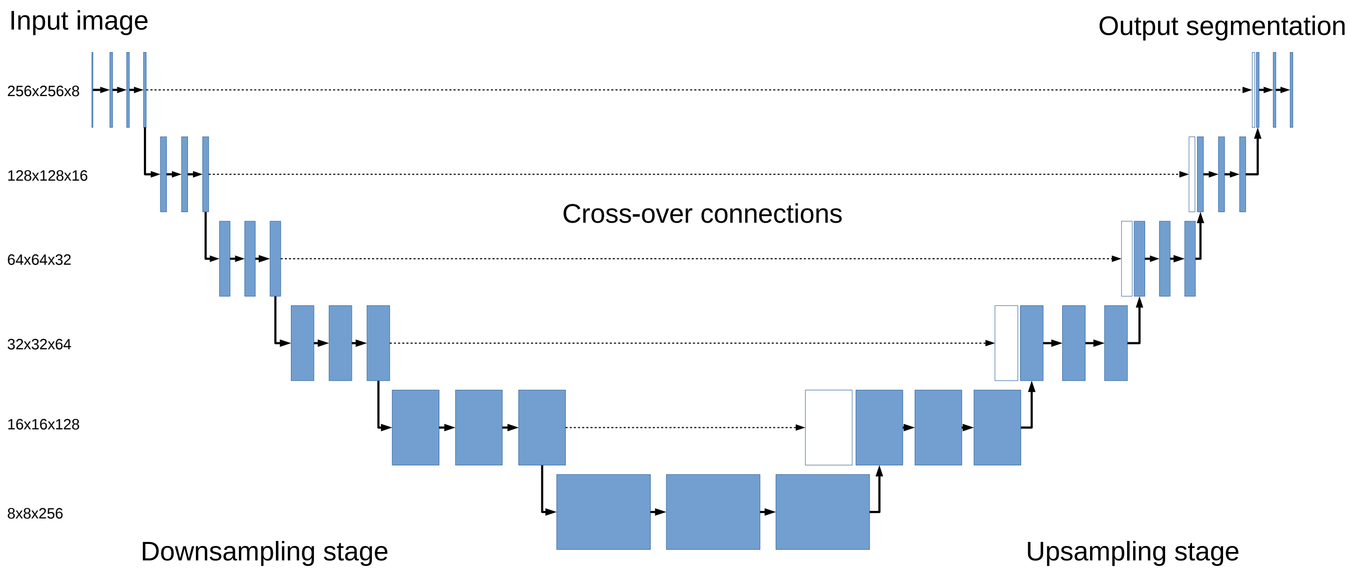

Fig. 1. U-net architecture [4]. Downward and upward arrows indicate downsampling and upsampling operations respectively. The straight filled arrows are 3×3 convolutions, and the dashed arrows are concatenation operations where a tensor is merged with another tensor. The number of features are doubled for each downsampling, and halved for each upsampling.

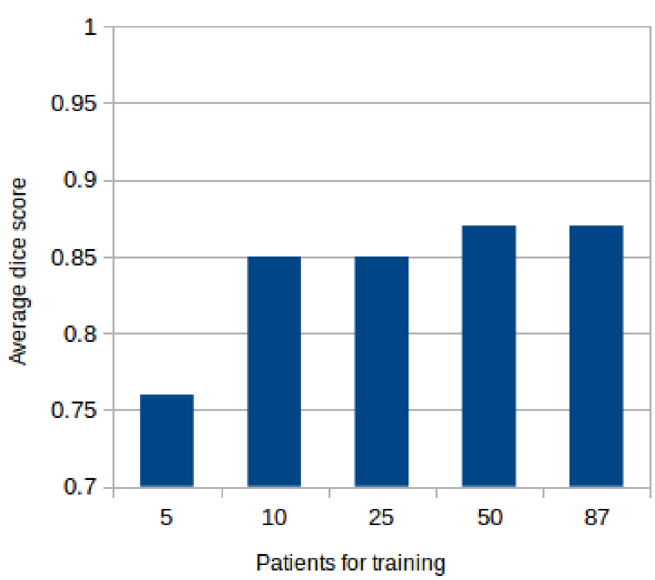

Fig. 2. The average dice score on fixed validation data versus the number of patients used for training the neural network.

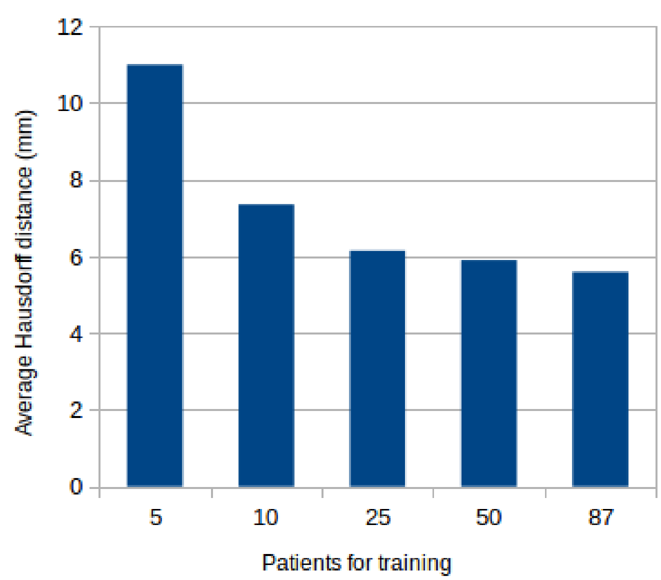

Fig. 3. The average Hausdorff distance on fixed validation data versus the number of patients used for training the neural network.

training dataset for the neural network consists of all the input ultrasound images and the corresponding label images from the KF segmentation method.

B. Neural network

A U-net neural network segmentation architecture was used as shown in Fig. 1. This architecture has shown to be applicable to multiple medical image segmentation problems [4]. The first part of U-net consists of several convolutional layers and downsampling steps while gradually increasing the number of features. The second part perform multiple upsampling steps to recover the original image resolution. To recover fine-grained features that may be lost in the downsampling stage, cross-over connections are used by concatenating equally sized feature maps. The last part of U-net are the cross-over connection, which are used to recover fine-grained features in the image that may be lost in the downsampling stages. All convolutions had a filter size of 3×3.

The network requires a fixed size input of 256×256 pixels, thus all input ultrasound images and output segmentations from the dataset were resized to this size. The pixel intensity of the ultrasound images were rescaled to unit scale $[0, 1]$. The neural

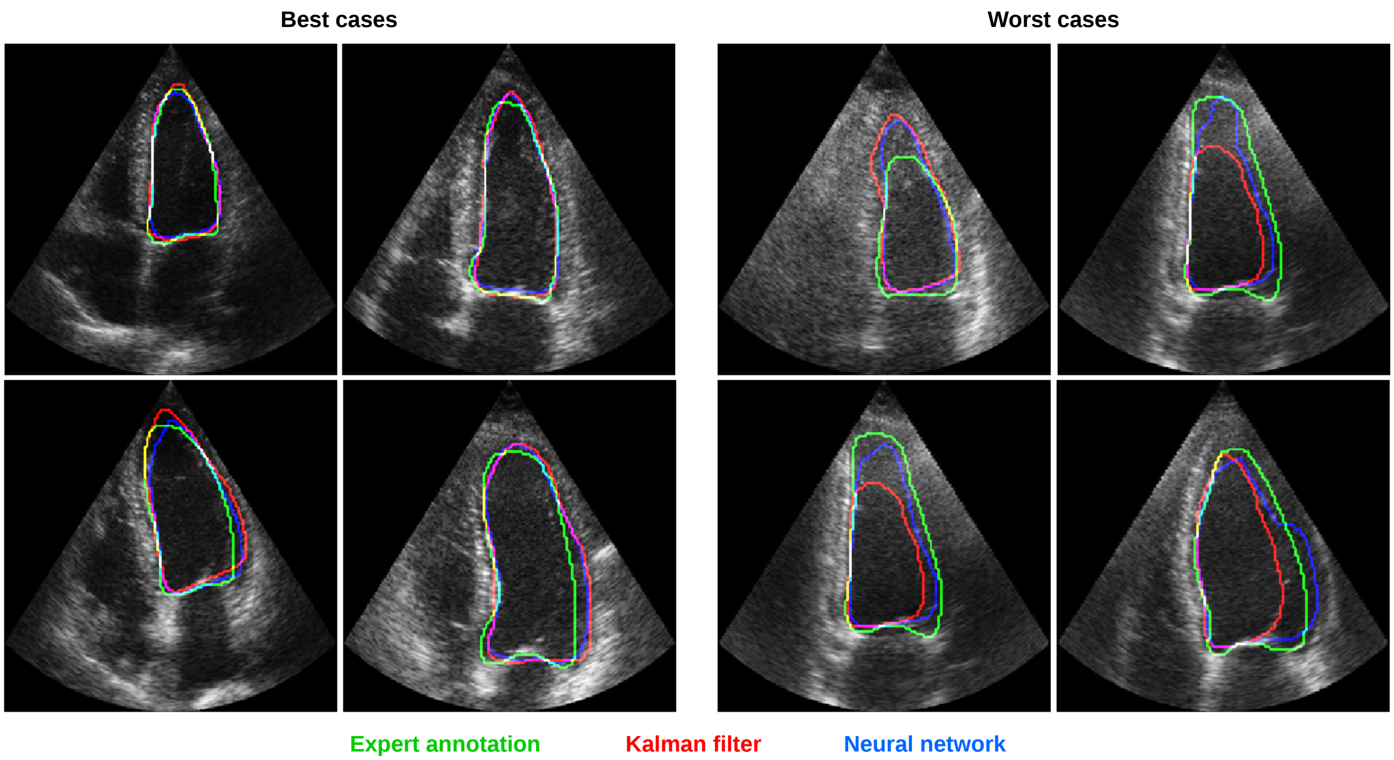

Fig. 4. Ultrasound images annotated by an expert cardiologist. The expert segmentation contour is shown in green, the Kalman filter in red and the neural network in blue. The four images to the left are the best cases, where the contours were closest to the expert annotation. The four images to the right show the opposite, the worst cases.

network was implemented using Tensorflow and Keras [6]. For the network training, stochastic gradient descent was used with 10 epochs, learning rate 0.01 and momentum 0.9.

C. Evaluation

For the evaluation, the segmentation accuracy of the model-based Kalman filter method was compared to that of the trained neural network. 13 of the 100 patients were randomly selected, excluded from training and used for validation. 52 images from these patient's recordings where manually segmented by an expert cardiologist.

The dice similarity coefficient D [7] was calculated to measure the overlapping regions of the segmentation S and the ground truth G.

$$D = \frac{2|S \cap G|}{|S| + |G|} \tag{1}$$

For the contour of the segmentation, the Hausdorff distance H was calculated in millimeters.

$$H = \max\left(\max_{i \in [0, O-1]} d(i, G, S), \max_{i \in [0, M-1]} d(i, S, G) \right) \tag{2}$$

Here, $d(i, G, S)$ is the distance from contour point i in G to the closest contour point in S. O and M are the number of pixels on the contour of G and S respectively.

III. Results

The dice similarity coefficient (DSC) and the Hausdorff distance was calculated for the KF and CNN method. The average DSC was 0.86 ± 0.06 for the KF and 0.87 ± 0.06 for the CNN. Thus, the CNN achieved comparable performance to the KF by training it on output data from the KF. The average Hausdorff distance, representing the maximum error, was higher for the KF with 7.5 ± 5.6 mm, compared to 5.9 ± 2.9 mm for the CNN. The DSC was observed to increase, and the Hausdorff distance decrease when varying the number of patients included in training as shown in Fig. 2 and Fig. 3. Fig. 4 show 8 of the expert annotated images and the output contour of the KF and the CNN.

The inference runtime of the network was 90 ± 8 ms using a CPU and 31 ± 11 using a GPU, but may be reduced by optimizing the network architecture and computation graph.

IV. Discussion

The results show that a deep convolutional neural network can learn to segment the left ventricle directly from another segmentation algorithm. The DSC of each method is quite similar, however for the Hausdorff distance the neural network is actually performing better than the KF.

Ideally, we wish to improve on the accuracy of the KF method, not just copy it. This may be done by first pretraining the neural network using the proposed approach, and then fine tune the network afterwards using a small number of manual expert annotations. A pretraining approach for the U-net architecture was proposed by Wiehman et al. [8]. This approach may be applicable to other segmentation tasks and image modalities.

It is a common belief that large amounts of data is required for deep neural networks to learn a task. However, this seems to depend on the task at hand. In the ImageNet challenge, over one million images are used for training [9]. In this challenge, the task is to classify ordinary color photographs as one of one thousand classes. We argue that this task is much more complex than the task of LV segmentation in ultrasound images, as the image variation is much smaller. For the specific task of left ventricle segmentation, it seems according to the graphs in Fig. 2 and Fig. 3 that a neural network can learn this task with a reasonable accuracy with less data than we initially thought. Fig. 4 show that for the worst cases, the neural network segmentation contour is less smooth and thereby less anatomical correct. The KF avoids this by forcing the contour spline to be smooth.

V. Conclusion

A convolutional neural network was trained using no manual annotations, only an automatic segmentation method and a large ultrasound dataset. The resulting network model was able to achieve similar accuracy to the automatic segmentation method. In future work, this may enable deep convolutional neural networks to exceed state-of-the-art with a small set of expert annotations for fine-tuning.

References

[1] J. A. Noble and D. Boukerroui, "Ultrasound image segmentation: A survey," *IEEE Transactions on Medical Imaging*, vol. 25, no. 8, pp. 987–1010, 2006.

[2] G. Carneiro, J. C. Nascimento, and A. Freitas, "The Segmentation of the Left Ventricle of the Heart From Ultrasound Data Using Deep Learning Architectures and Derivative-Based Search Methods," *IEEE Transactions on Image Processing*, vol. 21, no. 3, pp. 968–982, mar 2012.

[3] O. Oktay, E. Ferrante, K. Kamnitsas, M. Heinrich, W. Bai, J. Caballero, R. Guerrero, S. Cook, A. de Marvao, T. Dawes, D. O'Regan, B. Kainz, B. Glocker, and D. Rueckert, "Anatomically Constrained Neural Networks (ACNN): Application to Cardiac Image Enhancement and Segmentation," pp. 1–10, may 2017. [Online]. Available: http://arxiv.org/abs/1705.08302

[4] O. Ronneberger, P. Fischer, and T. Brox, "U-Net: Convolutional Networks for Biomedical Image Segmentation," *MICCAI*, pp. 234–241, 2015.

[5] E. Smistad and F. Lindseth, "Real-time Tracking of the Left Ventricle in 3D Ultrasound Using Kalman Filter and Mean Value Coordinates," in *Proceedings MICCAI Challenge on Echocardiographic Three-Dimensional Ultrasound Segmentation (CETUS)*, Boston, 2014, pp. 65–72.

[6] M. Abadi, A. Agarwal, P. Barham, E. Brevdo, Z. Chen, C. Citro, G. S. Corrado, A. Davis, J. Dean, M. Devin, S. Ghemawat, I. Goodfellow, A. Harp, G. Irving, M. Isard, Y. Jia, R. Jozefowicz, L. Kaiser, M. Kudlur, J. Levenberg, D. Mane, R. Monga, S. Moore, D. Murray, C. Olah, M. Schuster, J. Shlens, B. Steiner, I. Sutskever, K. Talwar, P. Tucker, V. Vanhoucke, V. Vasudevan, F. Viegas, O. Vinyals, P. Warden, M. Wattenberg, M. Wicke, Y. Yu, and X. Zheng, "TensorFlow: Large-Scale Machine Learning on Heterogeneous Distributed Systems," mar 2016. [Online]. Available: http://arxiv.org/abs/1603.04467

[7] L. R. . Dice, "Measures of the Amount of Ecologic Association Between Species," *Ecology*, vol. 26, no. 3, pp. 297–302, 1945.

[8] S. Wiehman, S. Kroon, and H. De Villiers, "Unsupervised pre-Training for fully convolutional neural networks," *2016 Pattern Recognition Association of South Africa and Robotics and Mechatronics International Conference, PRASA-RobMech 2016*, 2017.

[9] O. Russakovsky, J. Deng, H. Su, J. Krause, S. Satheesh, S. Ma, Z. Huang, A. Karpathy, A. Khosla, M. Bernstein, A. C. Berg, and L. Fei-Fei, "ImageNet Large Scale Visual Recognition Challenge," *International Journal of Computer Vision*, vol. 115, no. 3, pp. 211–252, 2015.