Real-time 3D left ventricle segmentation and ejection fraction using deep learning

Erik Smistad[*†‡], Erik Nikolai Steinsland[†], Lasse Løvstakken[*,†]
[*]*Centre for Innovative Ultrasound Solutions (CIUS)*
Trondheim, Norway
[†]*Norwegian University of Science and Technology (NTNU)*
Dept. of Circulation and Medical Imaging
Trondheim, Norway
[‡]*SINTEF Medical Technology*
Trondheim, Norway

***Abstract*—Supervised learning for 3D left ventricle (LV) ultrasound segmentation is difficult due to the challenges of acquiring large amounts annotated data. In this work, pre-training on a weakly labeled dataset, combined with augmentations and fine-tuning on a limited dataset using a straightforward 3D convolutional U-net type neural network was investigated. The results indicate that an accuracy close to both state-of-the-art and inter-observer can be achieved with such an approach. The resulting neural network was highly efficient (17 ms on laptop GPU) and was used to create a real-time application for fully automatic LV volume and ejection fraction measurements over multiple heartbeats to enhance practical use in the echo lab.**

I. Introduction

Deep learning is state-of-the-art for left ventricle (LV) segmentation in both 2D and 3D ultrasound, but requires large amounts of annotated data. Annotating 3D ultrasound data is complicated and highly time consuming, and only a small dataset is publicly available from the CETUS MICCAI challenge held in 2014 [1], [2].

After the 2014 CETUS challenge, several groups have used deep convolutional neural networks (NN) to segment the LV in 3D ultrasound. Typically fully convolutional 3D encoder-decoder (e.g. U-net) type architectures have been used with 3D convolutions. To deal with the limited amount of data, Oktay et al. [3] used in 2017 an autoencoder trained on anatomically correct LV shapes. The autoencoder was then used to guide the training of a segmentation network on limited data. While this method can guide a NN to segment more anatomically correct shapes, it does not guarantee anatomically correct shapes. This method was called *"Anatomically Constrained Neural networks (ACNN)"*, and although it doesn't require a large ultrasound dataset, it still requires a large dataset of anatomically correct LV shapes. Dong et al. [4] presented in 2018 a method called *VoxelAtlasGAN* which uses a trained conditional generative adversional network (GAN) to guide the segmentation. The generator of the GAN is used as an atlas which is deformed and used to generate the output segmentation. This method was later improved in [5] and named *AtlasNet*.

This research was funded by the Research Council of Norway under project 237887.

In this work, we investigated the possibility of combining an initial automatic non-machine learning method to generate 3D annotations for pre-training, and a limited manually annotated dataset with image augmentation techniques for fine-tuning. Further, we aimed to develop a real-time 3D NN and application for automatic LV volume and ejection fraction (EF) measurements to enhance practical use in the echo lab.

II. Methods

A. Dataset and annotation

The public CETUS 3D ultrasound LV segmentation dataset was used [2]. This dataset consists of 3D ultrasound recordings from 45 patients using three different ultrasound scanners. In this dataset, only annotations of 15 of the patients were available. And they are only annotated at end-diastole (ED) and end-systole (ES) frames resulting in a total of 30 annotated volumes, while is quite small for training and testing.

To deal with the lack of annotated volumes, an automatic non-machine learning Kalman filter method (CETUS 2014, rank 2) [6] was used to segment the LV for every frame of all 45 3D recordings resulting in 1157 annotated volumes. This weakly annotated dataset was used to pretrain the neural network.

B. Neural network architecture

A fully convolutional encoder-decoder U-net type network was used. The network uses 2×2 max pooling in the encoder stage, and 2×2 repeat upsampling in the decoder stage. Each level has 3D convolution layers with ReLU activation and cross-over connections. The input and output size of the network was $64 \times 64 \times 64$. The input/output image size impacts the number of parameters and thereby the inference runtime, and was thus kept small to facilitate real-time deployment. The network has 2.5 million parameters, just 0.5 million parameters more than the 2D equivalent for LV segmentation used in [7].

C. Training

The NN was pretrained using the weakly automatic annotated dataset, and fine-tuned on the expert annotation of 15

2021 IEEE International Ultrasonics Symposium (IUS) | 978-1-6654-0355-9/21/$31.00 ©2021 IEEE | DOI: 10.1109/IUS52206.2021.9593301

978-0-7381-1209-1/21/$31.00 ©2021 IEEE

Authorized licensed use limited to: KINGS COLLEGE LONDON. Downloaded on January 11,2022 at 19:03:55 UTC from IEEE Xplore. Restrictions apply.

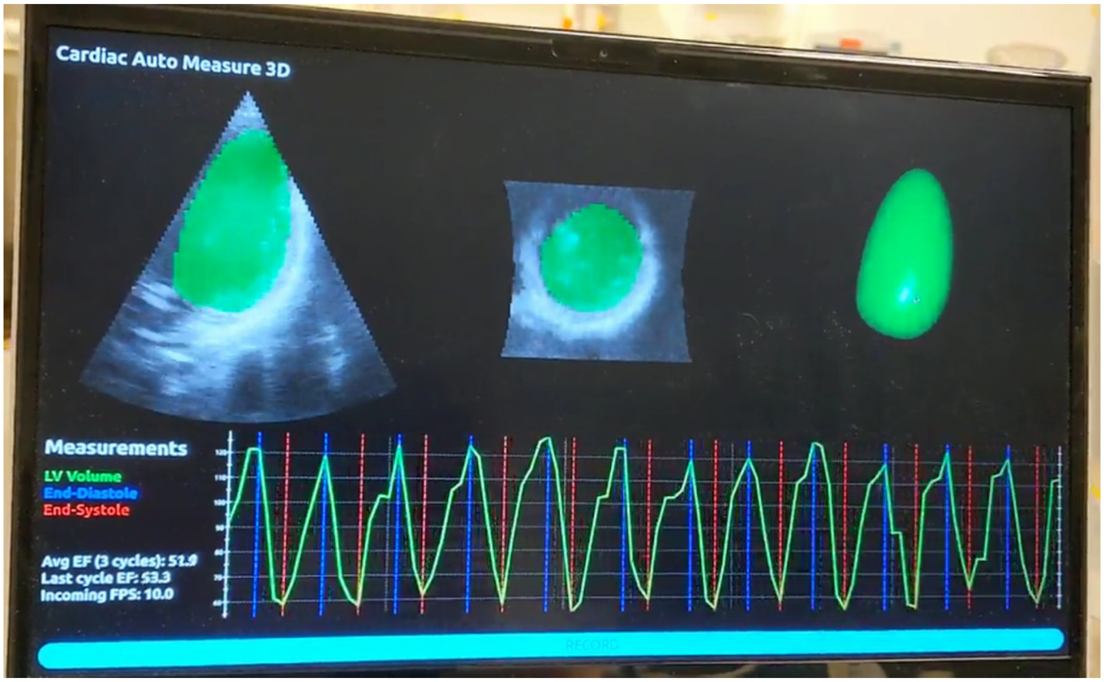

Fig. 1. Photo of the real-time 3D LV segmentation and ejection fraction application in use. 3D ultrasound volumes are streamed from a GE Vivid E95 scanner to a laptop. The application uses the trained neural network to perform segmentation, which can be seen as a green overlay on the ultrasound image. The 3D mesh at the right is extracted from the segmentation using marching cubes surface extraction on the GPU. The green curve at the bottom is the LV volume over time. The blue and red vertical lines are the estimated ED and ES time points. From the LV volume curve the ejection fraction is calculated and averaged over multiple heartbeats.

patients. The network was trained with batch size 4 and a Dice loss function.

To further avoid overfitting due to the limited dataset, the following random image augmentations were applied during training:

- Gamma intensity transformation
- Rotation around the depth axis
- Depth cropping
- Elastic deformation

D. Real-time application

A real-time application was created using the FAST framework[1] [8], [9]. The application streams 3D ultrasound data in real-time from a GE E95 ultrasound scanner over an ethernet connection. The ultrasound data is processed with the NN, and the segmentation is displayed and used to calculate the LV volume. The LV volume is plotted over time, and used to estimate ED and ES as the time with maximum and minimum LV volume. Using the ED and ES volume, the ejection fraction is calculated and averaged over multiple heart beats as shown in Fig. 1.

III. Results

In order to properly test the method with the very limited ground truth data, leave-one-subject-out cross-validation was used, which resulted in 15 models. Two random patients were used for validation and thus selecting the final model during training. The Dice scores were calculated for both ED and ES with and without pre-training and is summarized in Table I. The table also includes the inter-observer variability measured on the CETUS dataset between three experts, and the ACNN method [3] which was also validated on the CETUS dataset. Note however, that since the CETUS challenge evaluation platform has been offline for a while, the inter-observer, ACNN and the proposed methods have not been evaluated on the same subset of the CETUS dataset. Accuracy metrics from the VoxelAtlasGAN/AtlasNet method by Dong et al. [4], [5]

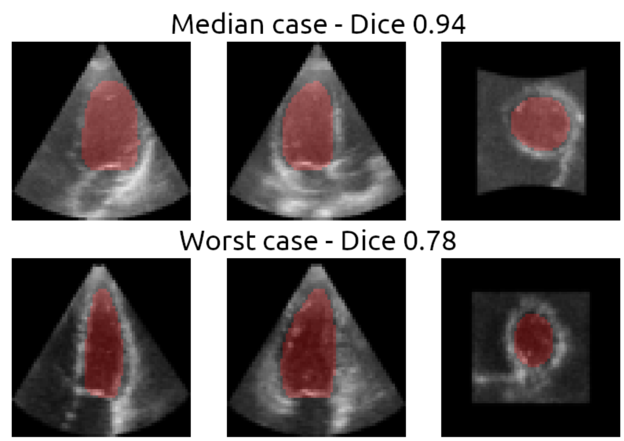

Fig. 2. Median and worse case segmentation in terms of Dice score on the CETUS dataset with the proposed method.

[1]https://fast.eriksmistad.no

Authorized licensed use limited to: KINGS COLLEGE LONDON. Downloaded on January 11,2022 at 19:03:55 UTC from IEEE Xplore. Restrictions apply.

TABLE I

LEAVE-ONE-SUBJECT-OUT CROSS-VALIDATION RESULTS ON THE 15 PATIENTS CETUS DATASET WITH MEAN DICE SCORE AND HAUSDORFF DISTANCE IN MILLIMETERS. REPORTED INTER-OBSERVER AND ACCURACY FROM RELATED WORK AT THE BOTTOM FOR COMPARISON.

Method	**Dice (ED)**	**Dice (ES)**	**Hausdorff (ED)**	**Hausdorff (ES)**
No Pre-training	0.919 ± 0.024	0.867 ± 0.092	8.48 ± 4.01	9.91 ± 10.31
With Pre-training	0.933 ± 0.027	0.909 ± 0.057	7.75 ± 4.41	6.31 ± 1.81
Inter-observer (CETUS) [2]	0.931 ± 0.021	0.920 ± 0.021	4.70 ± 1.27	4.70 ± 1.15
ACNN 2017 [3]	0.912 ± 0.023	0.873 ± 0.051	6.96 ± 1.75	7.75 ± 2.65
VoxelAtlasGAN 2018 [4]	0.953 ± 0.019		7.26 ± 2.3	
AtlasNet 2020 [5]	0.97 ± 0.012		5.6 ± 1.35	

which was tested on a small private dataset are also included in the table.

With the proposed method, the segmentation accuracy was within the inter-observer variability in ED with a Dice score of 0.933, while slightly lower for ES 0.909. Without pre-training the Dice scores were lower: 0.919 and 0.867, thus indicating that the proposed pre-training is useful. Fig. 2 shows examples of median and worst case segmentation in terms of Dice score on the CETUS 15 patients dataset with the fine-tuned models.

An inference runtime of 17 ms per volume was achieved using FAST, NVIDIA TensorRT and a RTX 2080 laptop GPU while at the same time streaming, visualizing and calculating EF in real-time. The average frames per second is thus limited by the slow network 3D image streaming.

IV. DISCUSSION

A major drawback of this study is the small size of the dataset with ground truth annotations. Nevertheless, we argue that the accuracy achieved with such a limited dataset is intriguing, demonstrating that large amounts of data is not necessarily needed to achieve good accuracy. Real-time tests also seem to indicate that the models are not overfitted and are able to generalize to new subjects and ultrasound scanners. Clinically speaking this method and the real-time application makes it possible to get fairly accurate volume and ejection fraction measurements in just a few seconds which should be very beneficial in a busy clinical practice. Future work will include acquiring a large annotated dataset, which can be used for further training and evaluation of the method and including both the myocardium and the left atrium. The CETUS challenge included a website were results could be uploaded and evaluated on the ground truth annotations for all 45 patients. However, since this evaluation platform has been offline for a while, we were only able to use the 15 patients annotations which are still available for download and direct comparison with other work is therefore difficult.

Comparing different methods in Table I, one can see that the NN atlas methods of Dong et al. [4], [5] report a much better Dice score than the proposed method, the ACNN method, as well as inter-observer. Still, one has to keep in mind that they have used a private dataset acquired with only one scanner, while all the other methods in the Table are tested on the public CETUS dataset where three different scanners have been used. Thus, direct comparison of these accuracy scores is difficult.

V. CONCLUSION

Results showed that a simple 3D NN with accuracy comparable to state-of-the-art and inter-observer can be achieved with very little ground truth data. The application was able to measure and average EF automatically over several heart beats in real-time, beneficial in the hectic echo lab.

REFERENCES

[1] O. Bernard, B. Heyde, M. Alessandrini, D. Barbosa, S. Camarasu-pop, F. Cervenansky, S. Valette, E. Galli, M. Geleijnse, A. Papachristidis, G. Johan, and D. Jan, “Challenge on Endocardial Three-dimensional Ultrasound Segmentation (CETUS),” in *Proceedings MICCAI Challenge on Echocardiographic Three-Dimensional Ultrasound Segmentation (CETUS)*, Boston, 2014, pp. 1–8.

[2] O. Bernard, J. G. Bosch, B. Heyde, M. Alessandrini, D. Barbosa, S. Camarasu-Pop, F. Cervenansky, S. Valette, O. Mirea, M. Bernier, P.-M. Jodoin, J. S. Domingos, R. V. Stebbing, K. Keraudren, O. Oktay, J. Caballero, W. Shi, D. Rueckert, F. Milletari, S.-A. Ahmadi, E. Smistad, F. Lindseth, M. van Stralen, C. Wang, O. Smedby, E. Donal, M. Monaghan, A. Papachristidis, M. L. Geleijnse, E. Galli, and J. D’hooge, “Standardized Evaluation System for Left Ventricular Segmentation Algorithms in 3D Echocardiography,” *IEEE Transactions on Medical Imaging*, vol. 35, no. 4, pp. 967–977, 4 2016.

[3] O. Oktay, E. Ferrante, K. Kamnitsas, M. Heinrich, W. Bai, J. Caballero, S. A. Cook, A. de Marvao, T. Dawes, D. P. O’Regan, B. Kainz, B. Glocker, and D. Rueckert, “Anatomically Constrained Neural Networks (ACNNs): Application to Cardiac Image Enhancement and Segmentation,” *IEEE Transactions on Medical Imaging*, vol. 37, no. 2, pp. 384–395, 2 2018.

[4] S. Dong, G. Luo, K. Wang, S. Cao, A. Mercado, O. Shmuilovich, H. Zhang, and S. Li, *VoxelAtlasGAN: 3D Left Ventricle Segmentation on Echocardiography with Atlas Guided Generation and Voxel-to-Voxel Discrimination*. Springer International Publishing, 2018, vol. 11073 LNCS.

[5] S. Dong, G. Luo, C. Tam, W. Wang, K. Wang, S. Cao, B. Chen, H. Zhang, and S. Li, “Deep Atlas Network for Efficient 3D Left Ventricle Segmentation on Echocardiography,” *Medical Image Analysis*, vol. 61, p. 101638, 2020.

[6] 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.

[7] S. Leclerc, E. Smistad, J. Pedrosa, A. Ostvik, F. Cervenansky, F. Espinosa, T. Espeland, E. A. R. Berg, P.-M. Jodoin, T. Grenier, C. Lartizien, J. Drhooge, L. Lovstakken, and O. Bernard, “Deep Learning for Segmentation using an Open Large-Scale Dataset in 2D Echocardiography,” *IEEE Transactions on Medical Imaging*, pp. 1–1, 2019.

[8] E. Smistad, M. Bozorgi, and F. Lindseth, “FAST: framework for heterogeneous medical image computing and visualization,” *International Journal of Computer Assisted Radiology and Surgery*, vol. 10, no. 11, pp. 1811–1822, 2015.

[9] E. Smistad, A. Østvik, and A. Pedersen, “High Performance Neural Network Inference, Streaming and Visualization of Medical Images using FAST,” *IEEE Access*, vol. 7, pp. 1–1, 2019.

Authorized licensed use limited to: KINGS COLLEGE LONDON. Downloaded on January 11,2022 at 19:03:55 UTC from IEEE Xplore. Restrictions apply.