Left ventricular and atrial segmentation of 2D echocardiography with convolutional neural networks

Joshua V. Stough[a,b], Sushravya Raghunath[b], John M. Pfeifer[b]
Brandon K. Fornwalt[b], and Christopher M. Haggerty[b]

[a]Computer Science, Bucknell University, Lewisburg, PA;
[b]Imaging Science and Innovation, Geisinger, Danville, PA

ABSTRACT

Segmentation of heart substructures in 2D echocardiography images is an important step in diagnosis and management of cardiovascular disease. Given the ubiquity of echocardiography in routine cardiology practice, the time–consuming nature of manual segmentation, and the high degree of inter-observer variability, fully automatic segmentation is a goal common to both clinicians and researchers. The recent publication of the annotated CAMUS dataset will help catalyze these efforts. In this work we develop and validate against this dataset a deep fully convolutional neural network architecture for the multi-structure segmentation of echocardiography, including the left ventricular endocardium and epicardium, and the left atrium. In ten-fold cross validation with data augmentation, we obtain mean Dice overlaps of 0.93, 0.95, and 0.89 on the three structures respectively, representing state of the art on this dataset. We further report small biases and narrow limits of agreement between the automatic and manual segmentations in derived clinical indices, including median absolute errors for left ventricular diastolic (7.3mL) and systolic volumes (4.9mL), and ejection fraction (3.8%), within previously reported inter-observer variability. These encouraging results must still be validated against large-scale independent clinical data.

Keywords: Echocardiography, Segmentation, Quantitative Image Analysis, Neural Networks

1. INTRODUCTION

Echocardiography is a ubiquitous imaging modality for diagnosing and managing patients with cardiovascular disease.[1] A single 2D echocardiography study may contain between 80-120 B-mode and M-mode video echocardiograms from more than 20 standard views varying in transducer positions and settings, each providing different clinical indices. Among the most common indices of cardiac structure and function extracted from a study is the left ventricular (LV) ejection fraction (EF). The most precise protocol for measuring EF from 2-dimensional echocardiograms requires the manual delineation of the LV endocardium (blood pool) in apical two chamber (AP2) and apical four chamber (AP4) views at both the end-diastolic (ED) and end-systolic (ES) phases of the cardiac cycle. This is a time-consuming task for clinicians that is also subject to a high degree of inter-observer variability,[2] motivating the development of automatic techniques.[3]

Numerous automatic echocardiogram segmentation methods have been proposed in both 2D[4] and 3D.[5] These methods can be broadly categorized as image-driven or model-driven approaches, differentiated by the use of strong prior information on shape and intensity variability from manually annotated training data. Image-driven approaches can be thought of as providing an empirical segmentation for a particular image, and such techniques include level set[6] and pixel/voxel graph cut[7] methods. Alternatively, model-driven approaches find the most likely segmentation given the learned prior, and include methods such as active appearance modeling[8] and multi-atlas registration.[9]

Deep learning and convolutional neural network (CNN) techniques have recently shown considerable promise as well. Caneiro et al.[10] use deep belief networks to first find regions containing the LV in AP4 echocardiograms and then find the LV contour within those regions. Smistad et al.[11] trained a 2D U-Net CNN architecture[12]

Corresponding author: joshua.stough@bucknell.edu

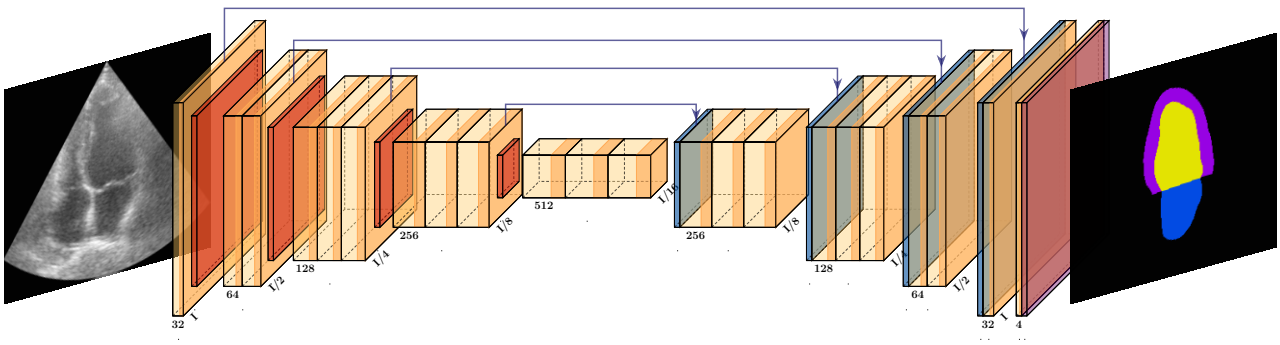

Figure 1. Deep CNN architecture for the multi-structure segmentation of echocardiography. Each block consists of chained convolution, group normalization, and non-linear activation. Convolutional down- and up-sampling halves (doubles) the feature map resolution during encoding (decoding). A softmax layer at the end normalizes the multi-label pixel map.

from the output of a previously published automatic Kalman filter-based segmentation method, showing similar results on a manually segmented test set. Oktay et al.[13] proposed an anatomically constrained CNN (ACNN) in 3D echo, where the training is regularized by an additional loss based on compact encoding of the ground-truth labeled images.

In order to catalyze further development in this field, LeClerc et al. have published the large annotated CAMUS dataset,[14] providing expert manual annotations in AP2 and AP4 views of LV endocardium (LV_{endo}), myocardium and LV epicardium (LV_{epi}), and the left atrium (LA). The authors also tested numerous deep learning and prior segmentation techniques, reporting that deep CNNs produced the best results. However, their best performing model had a large number of parameters (30M) and they did not use data augmentation to regularize model output, which has been shown to be effective in this context.[12]

We hypothesize that data augmentation will allow a more efficient deep CNN model to achieve improved results on the CAMUS dataset. Here we develop and validate such a model, using less than half the parameters of the previously reported best.[14] In ten-fold cross validation with data augmentation, we show state of the art results both in Dice overlap on the 2D model output and on derived LV_{endo} volumes and ejection fraction. In the next section we describe in turn the model architecture and training, data augmentation, and methods for evaluating the segmentation results.

2. ARCHITECTURE AND METHODS

Our model architecture (Fig. 1) is a variant of the popular U-Net CNN developed for biomedical image applications.[12] In this encoder-decoder style network, the contracting path (encoder) maps the input echo frame to a lower dimensional feature representation through convolutional and downsampling layers. The initial layers may learn to highlight local features such as oriented edges, while the deeper layers recognize combinations of features at increasing levels of abstraction and larger aperture, such as regions or objects. The expanding path (decoder) maps this abstract feature representation back to an output that is the same extent as the input frame, but now representing class probabilities per pixel for each label in training. Further, skip connections between corresponding layers in the contracting and expanding paths allow optimizing gradient information to flow directly between distant layers, leading to improved convergence and performance properties.

A large family of variants based on this design can be constructed through modifications to layer properties (e.g., kernel size, normalization, max-pooling or convolutional downsampling in the encoding path) and optimization strategies (e.g., chosen loss function, weight initialization).[14] Our model is distinct from other U-Net variants in two regards. First, we model skip connections that are additive, such that the feature maps of an encoding block are element-wise added to the upsampled inputs to the corresponding decoding block. Second, we use group normalization[15] for improved stability. The model is trained using the Adam optimizer with cross-entropy loss and weight decay. A learning rate scheduler is used to reduce the learning rate as validation loss plateaus.

2.1 Data Augmentation

To regularize the output we apply data augmentation reflecting the variability observed in the CAMUS set and other echocardiography studies. First, to reflect variable contrast across echocardiograms, we perform intensity windowing on the source frame, where a random subinterval of the input range is linearly mapped to the whole range: $I' = (1 - A)I_{min} + AI_{max}$, for a coefficient matrix A the same size as I, $A = (I - a)/(b - a)$, and the subinterval $[a, b] \subset [0, 1]$, with intensities outside the subinterval clipped. Second, we apply a small random rotation sampled from $\mathcal{N}(0, \sigma^2_{rot})$ about the transducer position in each frame, reflecting variable orientation of the ultrasound cone and heart. Last, we add Gaussian-distributed noise within the cone ($\sigma \sim \mathcal{U}\{0, \eta_{max}\}$) and clip the result.

2.2 Evaluation

To evaluate our method, we report both Dice overlap and derived ventricular volumes and EF. The Dice overlap measures the discrepancy between the automatically obtained segmentation and the expert manual delineation in a 2D frame. For S_{auto} and S_{ref} representing the areas enclosed by the respective object contours, Dice overlap measures the intersection area divided by the average, $D(S_{auto}, S_{ref}) = 2(|S_{auto} \cap S_{ref}|)/(|S_{auto}| + |S_{ref}|)$.[16]

Simpsons modified biplane method of disks is used to measure left ventricular volumes in the AP2 and AP4 views that are included in the CAMUS dataset.[17,18] In this protocol, image frames corresponding with the ED and ES phases are selected from the two views. The left ventricle is delineated in each image with a midline (from apex to base) and a series of parallel ventricular widths are specified at equal intervals along the midline and orthogonal to it. The volume is then estimated as the sum of the constituent elliptical cylinders where corresponding radii are assigned from the two views (which are ideally orthogonal to one another).

To automatically approximate this protocol, we use Principal Component Analysis to estimate the midline (first eigenvector) and its perpendicular direction (second eigenvector) in each view separately, using the boundary of the binary mask output of the model. The boundary pixels are then partitioned according to projection onto the midline and the orthogonal radii are determined as the median projection distance over each partition. To validate this automatic approach, we applied our approximation to the binary masks provided in the CAMUS dataset. We found tight agreement with the reported ED/ES left ventricular volumes ($r^2 > .98, p \ll 0.001$). EF is computed as the relative change in LV_{endo} volume between the ED and ES phases (i.e., EF = [ED-ES]/ED).

3. EXPERIMENTAL RESULTS

The publicly-released portion of the CAMUS dataset consists of 450 patients, two (AP2/AP4) views per patient, and two annotated (ED/ES) phases per view, totalling 1800 echocardiographic frames and corresponding label masks (background, LV_{endo}, LV_{epi}, LA). Additional information for each patient includes age, sex, and reported ED/ES LV volumes and EF, along with the observed image quality for each view. We perform a ten-fold cross-validation experiment with test folds stratified on patient EF range ($\leq 45\%, \geq 55\%, else$) and reported AP2 image quality (good, medium, poor), as suggested.[14] We train each view separately, saving the best performing model on a validation split that is another test fold.

A single set of model hyperparameters was chosen for all models. These include batch size 16, an initial learning rate of 0.002, and weight decay coefficient of $1e^{-6}$. Data augmentation included windowing a random subinterval up to half the input range, σ_{rot} of 5 degrees for random rotation about the transducer, and up to $\eta_{max} = 0.15$ additive Gaussian noise on the $[0, 1]$-normalized frames. Training continued to convergence, halted by a patience limit on validation loss. Developed in PyTorch, the model contains ∼13M parameters and one test fold requires ∼30 minutes training on a single Nvidia RTX 2080.

Table 1. Dice overlap scores accumulated over AP2 and AP4 views, showing median, mean, and standard deviation overall and for each of ED and ES phases separately.

Label/Score	D_{median}	$\text{D}_{\text{mean}} \pm\sigma$	ED D_{median}	ED $\text{D}_{\text{mean}} \pm\sigma$	ES D_{median}	ES $\text{D}_{\text{mean}} \pm\sigma$
LV_{endo}	0.939	0.928 ± 0.045	0.948	0.941 ± 0.029	0.928	0.916 ± 0.053
LV_{epi}	0.958	0.952 ± 0.028	0.960	0.955 ± 0.021	0.955	0.948 ± 0.033
LA	0.919	0.891 ± 0.094	0.904	0.872 ± 0.110	0.928	0.910 ± 0.071

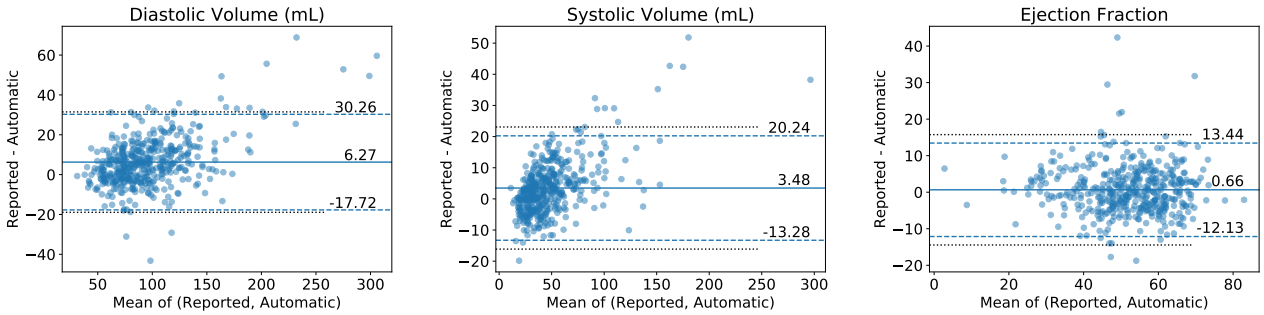

Figure 2. Bland-Altman plots comparing the automatically-obtained heart structure to the manual annotations for left ventricle volumes and ejection fraction. From left to right: diastolic volume, systolic volume, and EF, with mean bias and $\pm 1.96\sigma$ (dashed blue). Each are within reported inter-observer variability for 2D echocardiography[2] (dotted black).

Dice results are shown in Table 1. We obtain mean overlaps of 0.93 (0.94/0.92 for ED/ES) on LV_{endo} and 0.95 (0.96/0.95) on LV_{epi}. LA overlaps of 0.89 (0.87/0.91) reflect atrial filling that is opposite the LV, resulting in higher dice at (LV) ES, when the LA is largest. Despite not excluding poor quality echocardiograms, the LV overlaps show improvements over previously reported results,[14] though without identical test folds.

We further derive the LV volume measurements and EF using the corresponding AP2 and AP4 views for each patient. We compare against the values reported in the CAMUS dataset using Bland-Altman analysis (Fig. 2).[19] We obtain bias$\pm 1.96\sigma$ of 6.3mL$\pm$24.0, 3.5mL$\pm$16.8 and 0.7%$\pm$12.8 for ED and ES volumes and EF respectively, and mean absolute errors of 9.9mL, 6.5mL, and 4.8%. These measurements also represent improvement over previous results on this dataset.[14] Importantly, these biases and limits of agreement are all within reported inter-observer variability using Simpsons biplane in non-contrast 2D echocardiography ($\pm$25.2 mL ED; $\pm$19.6 mL ES; $\pm$15.1% EF).[2]

Lastly, we segmented an additional 50 held-out patients for independent evaluation through the online challenge platform.[14] In segmenting a frame from one of these patients, we accumulate the outputs from all ten-fold training models before choosing the most likely label. Our mean ED/ES overlaps of 0.95/0.93 (versus 0.94/0.91) for LV_{endo}, 0.96/0.96 (vs 0.96/0.95) for LV_{epi}, and 0.90/0.93 (vs 0.89/0.92) for LA also beat the best currently reported on the platform. Additionally-reported metrics (Hausdorff distance, mean absolute distance) also show meaningful improvement (supplemental figure to be provided in the full paper).

4. DISCUSSION

Here we have developed and validated a deep convolutional neural network model for 2D echocardiography segmentation, utilizing the recently released CAMUS large-scale dataset.[14] With a relatively efficient model and data augmentation, we obtain state of the art results in both Dice overlap and derived left ventricular volumes and ejection fraction, improving upon previously published work with this dataset, and within inter-observer variability both for this dataset and previous studies with clinical echocardiography.[2]

While the release of the CAMUS set has been an unqualified gain for the field, a significant difficulty remains in generalizing the models learned on these hundreds of patients to the tens of thousands available within our clinical system. There is large variability in acquisition settings, image quality, and even burned-in view and patient information that is common in the clinic but absent in such curated datasets; previously published fixed models[3] have not been shown to generalize in this context. We see further principled data augmentation coupled with shape constraints[13] as key to regularizing these models against such complexity.

REFERENCES

[1] Virnig, B. A., Shippee, N. D., O'Donnell, B., Zeglin, J., and Parashuram, S., "Trends in the use of echocardiography, 2007 to 2011." `https://www.ncbi.nlm.nih.gov/books/NBK208663/`.

[2] Wood, P. W., Choy, J. B., Nanda, N. C., and Becher, H., "Left ventricular ejection fraction and volumes: It depends on the imaging method," *Echocardiography* **31**(1), 87–100 (2014). https://doi.org/10.1111/echo.12331.

[3] Zhang, J., Gajjala, S., Agrawal, P., Tison, G. H., Hallock, L. A., Beussink-Nelson, L., et al., "Fully automated echocardiogram interpretation in clinical practice," *Circulation* **136**(16), 1623–1635 (2018). https://doi.org/10.1161/CIRCULATIONAHA.118.034338.

[4] Noble, J. and Boukerroui, D., "Ultrasound image segmentation: a survey," *IEEE Trans Med Imaging* **25**, 987–1010 (2006). https://www.ncbi.nlm.nih.gov/pubmed/16894993.

[5] Bernard, O., Bosch, J., Heyde, B., Alessandrini, M., Barbosa, D., Camarasu-Pop, S., Cervenansky, F., Valette, S., Mirea, O., et al., "Standardized evaluation system for left ventricular segmentation algorithms in 3d echocardiography," *IEEE Trans Med Imaging* **35**(4), 967–77 (2016). https://doi.org/10.1109/TMI.2015.2503890.

[6] Lin, N., Yu, W., and Duncan, J. S., "Combinative multi-scale level set framework for echocardiographic image segmentation," *Medical Image Analysis* **7**, 529–537 (2003). https://doi.org/10.1016/S1361-8415(03)00035-5.

[7] Bernier, M., Jodoin, P., Humbert, O., and Lalande, A., "Graph cut-based method for segmenting the left ventricle from mri or echocardiographic images," *Comput. Med Imaging Graph* **58** (2017). https://doi.org/10.1016/j.compmedimag.2017.03.004.

[8] van Stralen, M., Haak, A., Leung, K. E., van Burken, G., Bos, C., and Bosch, J. G., "Full-cycle left ventricular segmentation and tracking in 3d echocardiography using active appearance models," *Proc. IEEE International Ultrasonics Symposium (IUS)* (2015). https://doi.org/10.1109/ULTSYM.2015.0389.

[9] Oktay, O., Shi, W., Keraudren, K., Caballero, J., and Rueckert, D., "Learning shape representations for multi-atlas endocardium segmentation in 3d echo images," *The MIDAS Journal - Challenge on Endocardial Three-dimensional Ultrasound Segmentation* (2014). http://hdl.handle.net/10380/3486.

[10] Carneiro, G., Nascimento, J., and Freitas, A., "The segmentation of the left ventricle of the heart from ultrasound data using deep learning architectures and derivative-based search methods," *IEEE Trans Imag Proc* **21**, 968–82 (2012). https://doi.org/10.1109/TIP.2011.2169273.

[11] Smistad, E., Østvik, A., Haugen, B. O., and Løvstakken, L., "2d left ventricle segmentation using deep learning," *Proc. IEEE International Ultrasonics Symposium (IUS)* (2017). https://doi.org/10.1109/ULTSYM.2017.8092573.

[12] Ronneberger, O., Fischer, P., and Brox, T., "U-net: Convolutional networks for biomedical image segmentation," *Proc. Int. Conf. on Medical Image Computing and Computer Assisted Intervention (MICCAI)* , 234–241 (2015). https://doi.org/10.1007/978-3-319-24574-4_28.

[13] Oktay, O., Ferrante, E., Kamnitsas, K., Heinrich, M., et al., "Anatomically constrained neural networks (acnns): Application to cardiac image enhancement and segmentation," *IEEE Trans Med Imaging* **37**, 384–395 (2017). https://doi.org/10.1109/TMI.2017.2743464.

[14] Leclerc, S., Smistad, E., Pedrosa, J., Østvik, A., Cervenansky, F., Espinosa, F., et al., "Deep learning for segmentation using an open large-scale dataset in 2d echocardiography," *IEEE Trans Med Imaging* (2019). https://doi.org/10.1109/TMI.2019.2900516; https://www.creatis.insa-lyon.fr/Challenge/camus/index.html.

[15] Wu, Y. and He, K., "Group normalization," *Proc. The European Conference on Computer Vision (ECCV)* (2018). https://doi.org/10.1007/978-3-030-01261-8_1.

[16] Dice, L., "Measures of the amount of ecologic association between species," *Ecology* **26**, 297–302 (1945).

[17] "Lets talk left ventricle bi-plane volume measurements!." https://www.cardioserv.net/echo-lets-talk-lv-bi-plane-measurements/.

[18] Folland, E., Parisi, A., Moynihan, P., Jones, D., Feldman, C. L., and Tow, D. E., "Assessment of left ventricular ejection fraction and volumes by real-time, two-dimensional echocardiography. a comparison of cineangiographic and radionuclide techniques," *Circulation* **60**, 760–766 (1979). https://doi.org/10.1161/01.cir.60.4.760.

[19] Bland, J. M. and Altman, D. G., "Comparing methods of measurement: why plotting difference against standard method is misleading," *The lancet* **346**(8982), 1085–1087 (1995). http://dx.doi.org/10.1016/j.ijnurstu.2009.10.001.