Convolutional Neural Networks for Automated Fetal Cardiac Assessment using 4D B-mode Ultrasound

Manna E. Philip[1] Arcot Sowmya[1] Hagai Avnet[2] Ana Ferreira[3] Gordon Stevenson[3] Alec Welsh[3,4]

[1] School of Computer Science, Faculty of Engineering, UNSW, Australia
{mannaelizabeth.philip,a.sowmya}@unsw.edu.au
[2] Institute of Obstetrics and Gynecological Imaging and Fetal Therapy, Sheba Medical Center, Israel
[3] School of Womens and Childrens Health, Faculty of Medicine, UNSW, Australia
[4] Department of Maternal-Fetal Medicine, Royal Hospital for Women, Randwick, Australia

Technical Report
UNSW-CSE-TR-201805
October 2018

UNSW SYDNEY

School of Computer Science and Engineering
The University of New South Wales
Sydney 2052, Australia

Abstract

Structural and functional assessment of the fetal heart is currently performed using ultrasound biomarkers that require annotation by a trained clinician. This requires experience and expertise and is subject to inter- and intra-observer variability. In this work, a Convolutional Neural Network (CNN) was implemented for segmentation of the fetal annulus and automated measurements of the excursion of the mitral and tricuspid valve annular planes (TAPSE/MAPSE) were evaluated against manual annotation. After training, the network was able to achieve a Dice score of 0.78 ($\pm 0.02\sigma$) and the excursion measures had an RMSE of < 0.16cm. Results show the feasibility of a CNN to detect the fetal annulus and subsequently for measuring these imaging biomarkers for cardiac functional assessment.

1 Introduction

Recent advances in ultrasound have enabled the examination and assessment of the fetal circulation. This has provided a number of new ultrasound modalities (tissue Doppler, 3D ultrasound) which can be used to assess structural and functional performance of the heart *in-utero* [2]. These tools are well-established in paediatric and adult clinics, but are not studied widely in the fetus [7]. This is due to the increased difficulty in acquisition of a cardiac sequence from a moving small fetus, however technological advances in acquisition are closing the gap between these different clinical settings. In particular, the use of spatial-temporal image correlation (STIC) allows the reconstruction of a 4D (3D + time) ultrasound acquisition of the fetal heart [3]. Within these volumes, by measuring the axial displacement of the tricuspid valve annulus, an assessment of the right heart function can be performed. This distance has been shown to reflect longitudinal function and be an indicator of disease progression [7]. As well as tricuspid measurement (Tricuspid Annular Plane Systolic Excursion (TAPSE)), the left heart can be assessed by Mitral Annular Plane Systolic Excursion (MAPSE).

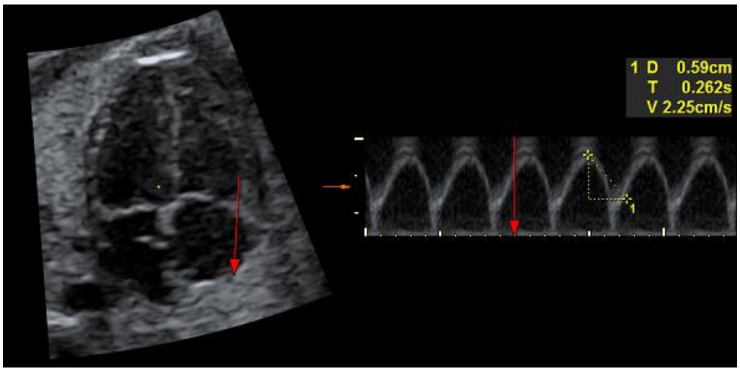


Figure 1.1: 2D slice of 4D STIC ultrasound volume of fetal heart in 4-chamber view and TAPSE measurement using M-mode (red-arrow) and calculated displacement

Currently, the TAPSE/MAPSE measurements are performed manually by an operator, either off-line or at the the time of scan. They measure the vertical distance that the annulus travels up and down in one cardiac cycle. The TAPSE/MAPSE measurements are taken from M-mode data, which requires identifying the location of the annulus and annotating the M-mode data using a cursor or trackball as shown in Figure 1.1. All these stages introduce potential error and require time, which points to the need for automation of the measurement of these imaging biomarkers.

CNN have found success in many computer vision tasks and are now being explored by several researchers for the segmentation of heart structures. Costa et al.[1] use a CNN-based U-net architecture for the segmentation of mitral valve leaflets from adult echo-cardiograms. Volumetric CNNs have also been implemented and successfully applied in 3D ultrasound with good performance compared to manual segmentation [12, 5].

Automated mitral and tricuspid annular excursion have been explored in adult hearts by Storve et al. [10] and [11] using additional M-mode and ECG information to obtain the results. A semi-automated algorithm has been proposed [6], based on cross correlation to track the tricuspid annulus (TA) points to obtain the TAPSE. In this work, we aim to segment the fetal annuli using deep learning techniques for both open and closed valve states and so compute TAPSE/MAPSE automatically. We aim to show that by automating the entire process, we can build a tool to help with better assessment of fetal cardiac measurement.

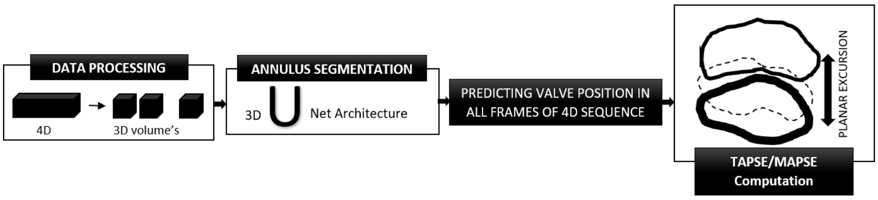


Figure 1.2: Flowchart outlining the TAPSE/MAPSE measurement algorithm

2 Method

The flowchart shown in Figure 1.2, shows the different steps involved in this work and each block is explained in detailed in the sub-sections below. This prospective observational study was approved by the Human Research Ethics Committees of The Royal Hospital for Women, Randwick and the University of New South Wales (UNSW). Following informed, written consent, participants $>$ 18 years old with a singleton, healthy pregnancy were recruited for an ultrasound scan which included the capture of 4D STIC ultrasound data.

A total of 295 fetal heart 4D STIC volumes were acquired at 4 chamber view from 95 different participants on a GE Voluson E8 ultrasound machine and a RAB 6-D, 3D transducer (GE Medical Systems, Zipf, Austria). Mean gestational age of the data was 30.7 weeks (min–max; 22.9-38.0). Raw volumes were exported as previously described [9] and converted to NIFTI format for subsequent image analysis.

2.1 Data Processing

The data set used in this work was acquired by three different sonographers. To ensure that sub-optimal performance of any designed method was not the result of large variations in data acquisition, a quality assurance scoring system was designed in consultation with the experts. Each 4D volume was evaluated manually and assigned a score out of ten. The scores were further subdivided into sections as shown in Table 2.1, depending on the relevance to the task.

Volumes with a score < 6 were distorted by speckle noise or affected by shadow, blocking the visibility of the annulus. Only volumes with a score $\geq$ of 6 and above were considered for this study. 150 4D volumes satisfied the above criteria. The 4D NIFTI files were manually rotated to an apex up/down position in parallel with the z-axis to ensure consistent visibility of the annulus for ground truth labelling. These 4D files were divided into 3D frames. Manual labelling

Table 2.1: Data Scoring

Scoring Parameter	Highest Score
Annulus Visibility	4
4 Chamber View	2
Speckle Noise	2
Shadowing Effect	1
Apex Position	1
Total Score	**10**

was performed on 250 based on an annulus segmentation protocol designed by experts to ensure labelling uniformity. The 250 volumes were made up of the first frame of the 150 chosen above plus the 20^{th} frame of a randomly chosen 100 of the 150. The labelled data set was divided into training, test and validation sets. As the size of the data set was considered small to train a network from scratch, data augmentation methods were used to create more samples, using flip up/down, flip left/right and rotation at $\pm 5°$ and $\pm 10°$ in the y-direction. Speckle noise was also added (variance of multiplicative noise: 0.01) as a data augmentation method, to increase network robustness to the presence of noise.

2.2 Annulus Segmentation

The processed training data was z-normalized (0 mean intensity; unary σ) and fed into a U-net architecture for training which was implemented using Keras with a TensorFlow backend and trained on two, NVIDIA GeForce GTX 1080 GPU (NVidia Corporation, Santa Clara, CA, USA). The network has four up-blocks and four down-blocks in accordance with the original U-net architecture [8]. The only changes were the addition of batch normalization and dropout to overcome over-fitting, and the 2D layers were converted to 3D for segmentation. All layers used a ReLU activation function, except the output layer which used a sigmoid. The U-net was chosen as we believe its architecture of an expanding path in addition to the contracting path may aid in better localization of the annulus [8]. Parameters used during training is summarized in Table 2.2.

Table 2.2: Training Parameters

Parameter	**Value**
Optimizer	Adam
Loss function	Dice Loss
Accuracy measure	Dice Coefficient
Epochs	100
Batch Size	5
Validation Split	10%

2.3 Valve positions in 4D sequence

Two separate networks were used, one for Mitral Annulus (MA) segmentation and the other for TA segmentation. Before feeding the test data, they were normalized using the same method used to normalize the training data. Since we are interested in locating the MA/TA on all frames of a given 4D sequence, the 3D frames that make up the 4D data were fed into the network for testing. The corresponding MA/TA training weights were loaded to perform the MA/TA segmentation. The final outputs were binarized as the sigmoid activation function can provide output values in the range $[0, 1]$.

2.4 TAPSE/MAPSE Measurement

Typically, planar excursion measurements are performed by subtracting the position of the annulus at End Systole (ES) from its position at End Diastole (ED). This requires manual identification of the particular ES and ED frames for each dataset. We took a different approach which does not require identification of the ES and ED frames. The centroid of the annulus segmentation was computed for all the frames. The two frames with the largest displacement from the centroid of the first frame in the positive / negative direction were considered to correspond to the ED / ES frame. The displacement in the z-plane of the centroid (cm) was considered the particular annular excursion value (TAPSE/MAPSE).

3 Results

The U-net CNN took approximately 8 hours to train, on TA/MA samples obtained using data augmentation techniques on 150 of the manually annotated samples and tested on 100 volumes. Dice Similarity Co-efficient values of 0.78 for TA segmentation and 0.77 for MA segmentation were achieved. Figure 3.1 shows the TA and MA segmentations obtained at ES and ED for two samples from the test set. For samples with a quality score of ≥ 8, Dice Similarity was 0.85 for both TA and MA segmentation.

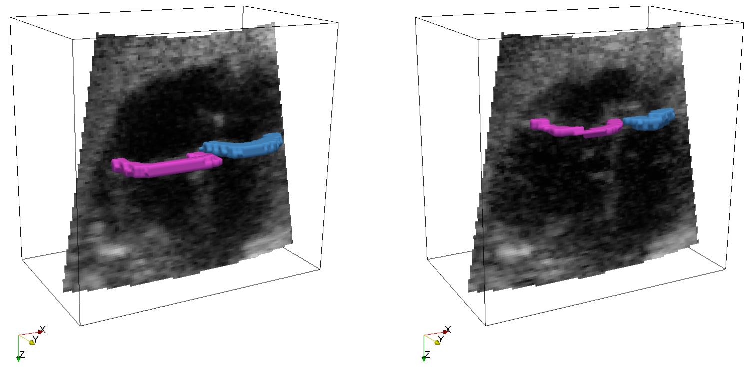
Once the segmentation of the annulus over all frames of a 4D volume was obtained, TAPSE/MAPSE were calculated for a set of 24 volumes randomly chosen from the whole dataset. Of these, outliers were removed using the median absolute deviation (MAD) * 2.5 to provide a cut-off as previously described [4]. Measurements greater than 1.13cm and 0.74cm for TAPSE and MAPSE were discarded, which removed 2 and 5 data volumes, from each group respectively.

Bland-Altman plots were used to analyse the measurement difference between two experts (Figure 3.2) and between the automated method and the average expert measurement (Figure3.3). For the automated TAPSE measurement we obtained a correlation coefficient $r = 0.61$, while that between experts was $r = 0.89$. The root mean squared error (RMSE) between the automated and reference measurements was 0.14cm.

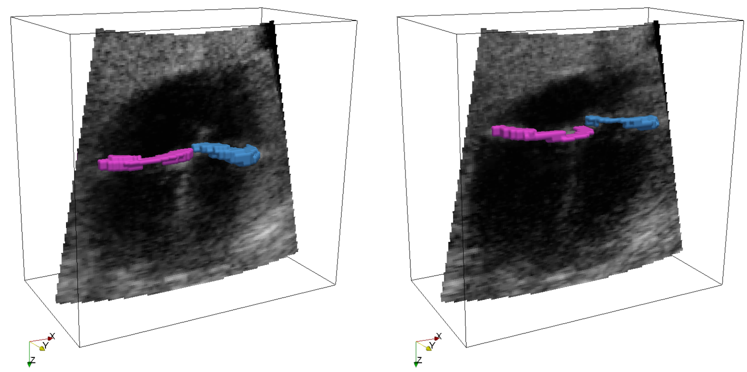
The r for automated MAPSE measurement and reference was 0.30 and between experts was 0.77 and the RMSE was 0.18. It is observed that the correlation coefficient for experts as well as proposed method for MAPSE is lower than those for TAPSE, mainly owing to the rotational movement of the MA due to the circular fiber orientation in left ventricle, in addition to the planar excursion which makes the MAPSE measurement more challenging than TAPSE.

4 Discussion

This work has demonstrated the feasibility of automated TAPSE/MAPSE measurement using 4D fetal cardiac data. Results point to the potential for a CNN



(a)



(b)

Figure 3.1: Visualisation of two different datasets are shown. 3D tricuspid (purple) and mitral annulus (blue) segmentation at end-diastolic (left) and end-systolic (right) in four-chamber view of a 2D B-mode slice of the whole 3D volume. Data was presented and rendered using ParaView (v. 5.4.1; Kitware Inc, Clifton Park, NY, USA)

.

to be used for the automated segmentation of the fetal annulus and subsequent measurement of the imaging biomarkers TAPSE/MAPSE for estimating fetal cardiac function. This work has shown promise but there are a number of steps that can be taken to improve the accuracy and repeatability of this technique.

The TAPSE/MAPSE values for good quality data were shown to be promising. For a cohort, they do seem to over-estimate displacement compared to human measurement. We speculate that the major reason for these results is due to fetal movements that can shift orientation of the valves over the cardiac cycle, which was not taken into account when computing either planar excursion. This meant that some outliers were excluded as described as these what were acquired in a five chamber view of the heart (including the aorta) which caused a large segmentation error. Another major reason for error is the fetal position during acquisition. Typically, a sonographer will acquire a scan in an "apex-up" or "apex-down" position , as they are two positions at which the annulus is clearly visible. However, in most cases the apex is slightly rotated. This is another source of potential error in the segmentation as well as excursion computation, which may be solved by either compensating for septum deviation

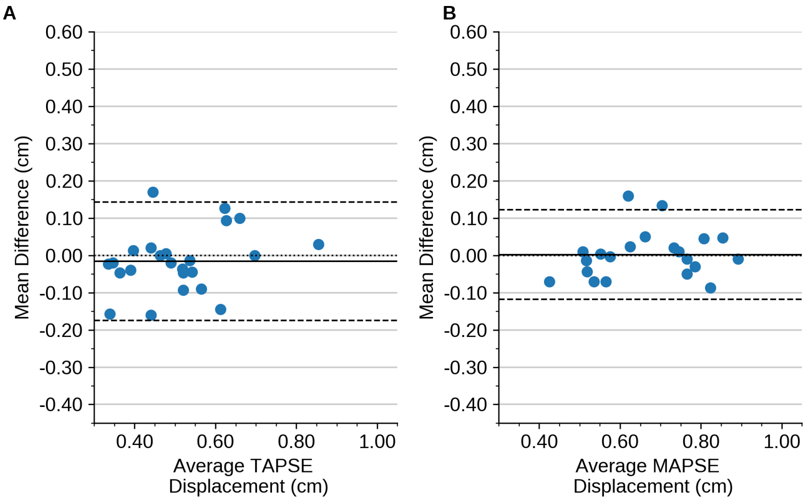


Figure 3.2: Bland-Altman plots comparing TAPSE (A) and MAPSE (B) measurements between experts.

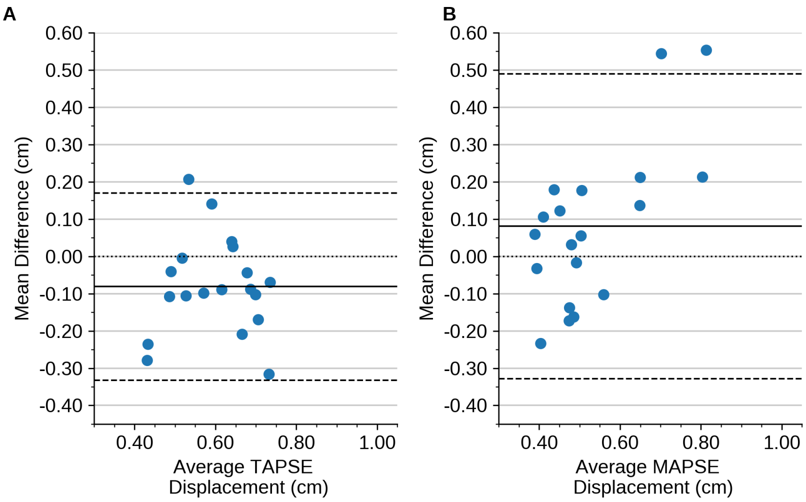


Figure 3.3: Bland-Altman plots comparing automated TAPSE (A) and MAPSE (B) measurements to average expert measurement.

or rigid registration of data to the position of the heart with respect to the first frame.

This work has that automation of TAPSE and MAPSE measurements is possible which is important as it can be a tool to reduce operator variability and increase the clinical throughput by reducing time required to measure this displacement. As such, this tool can potentially provide functional cardiac measurements where training is limited and skills lacking. The data we used for this study was clinically acquired on a system that has since been superseded and with the advent of matrix transducers and improved image acquisition we speculate that higher quality acquisitions are possible using the latest models. Limited pre- or post-processing was performed to the data which shows the potential for this technique with technical improvements and improved testing. These would include the use of a multi class CNN architecture rather than using than two separate binary segmentation networks and measurement using multiple operators.

5 Conclusion

We present a fully automated method of mitral and tricuspid annulus segmentation and measurement of its planar excursion over a cardiac cycle. The performance of the method is promising and strong for high quality data, pointing towards the feasibility of the work. Registration of the fetal heart orientation to an either apex-up or apex-down position would solve most of the problems and give better segmentation and hence TAPSE/MAPSE measurements. In this study, we do not take into account the temporal information available in the 4D data, which can be explored using LSTM's to possibly improve performance. Additional work may include modelling the entire fetal heart (ventricle volume) using CNN to study the development of the heart through the gestational weeks, which would give more insight into global cardiac function and developmental anomalies, thus assisting medical experts in faster diagnosis, assessment and treatment.

Bibliography

[1] Eva Costa, Nelson Martins, Malik Saad Sultan, Diana Veiga, Manuel Ferreira, Sandra Mattos, and Miguel Coimbra. Mitral Valve Leaflets Segmentation in Echocardiography using Convolutional Neural Networks. *MIDL*, (Midl):1–9, 2018.

[2] Fatima Crispi, Brenda Valenzuela-Alcaraz, Monica Cruz-Lemini, and Eduard Gratacos. Ultrasound assessment of fetal cardiac function. *Australasian Journal of Ultrasound in Medicine*, 16:158–67, Nov 2013.

[3] G. R. DeVore, P. Falkensammer, M. S. Sklansky, and L. D. Platt. Spatio-temporal image correlation (STIC): new technology for evaluation of the fetal heart. *Ultrasound in Obstetrics & Gynecology*, 22(4):380–387.

[4] Christophe Leys, Christophe Ley, Olivier Klein, Philippe Bernard, and Laurent Licata. Detecting outliers: do not use standard deviation around the mean, use absolute deviation around the median. *Journal of Experimental Social Psychology*, 49(4):764–766, 2013.

[5] Pádraig Looney, Gordon N Stevenson, Kypros H Nicolaides, Walter Plasencia, Malid Molloholli, Stavros Natsis, and Sally L Collins. Fully automated, real-time 3d ultrasound segmentation to estimate first trimester placental volume using deep learning. *JCI Insight*, 3(11), 2018.

[6] Francesco Maffessanti, Paola Gripari, Gianluca Pontone, Daniele Andreini, Maria C Carminati, Mauro Pepi, Enrico G Caiani, Biomedical Eng Dpt, and Politecnico Milano. Evaluation of a Semi-Automatic Algorithm for Tracking Tricuspid Valve Annulus on Magnetic Resonance Images. *Cycle*, pages 373–376, 2011.

[7] B. Messing, Y. Gilboa, M. Lipschuetz, D. V. Valsky, S. M. Cohen, and S. Yagel. Fetal tricuspid annular plane systolic excursion (f-tapse): evaluation of fetal right heart systolic function with conventional m-mode ultrasound and spatiotemporal image correlation (stic) m-mode. *Ultrasound in Obstetrics & Gynecology*, 42(2):182–188.

[8] Olaf Ronneberger, Philipp Fischer, and Thomas Brox. U-Net: Convolutional Networks for Biomedical Image Segmentation. *International Conference on Medical Image Computing and Computer-Assisted Intervention. (MICCAI)*, pages 234–241, 2015.

[9] Gordon N Stevenson, Sally L Collins, Alec W Welsh, Lawrence W Impey, and J Alison Noble. A technique for the estimation of fractional moving blood volume by using three-dimensional power doppler US. *Radiology*, 274(1):230–237, 2014.

[10] Sigurd Storve, Jahn Frederik Grue, Stein Samstad, Håvard Dalen, Bjørn Olav Haugen, and Hans Torp. Realtime automatic detection of heart failure in echocardiography. In *Ultrasonics Symposium (IUS), 2014 IEEE International*, pages 1069–1072. IEEE, 2014.

[11] Sigurd Storve, Jahn Frederik Grue, Stein Samstad, Havard Dalen, Bjørn Olav Haugen, and Hans Torp. Realtime Automatic Assessment of Cardiac Function in Echocardiography. *IEEE Transactions on Ultrasonics, Ferroelectrics, and Frequency Control*, 63(3):358–368, 2016.

[12] Xin Yang, Lequan Yu, Shengli Li, Xu Wang, Na Wang, Jing Qin, Dong Ni, and Pheng-Ann Heng. Towards automatic semantic segmentation in volumetric ultrasound. In *International Conference on Medical Image Computing and Computer-Assisted Intervention*, pages 711–719. Springer, 2017.