AutoFB: Automating Fetal Biometry Estimation from Standard Ultrasound Planes

Paper ID: 1866

*** *** ***

Abstract. During pregnancy, ultrasound examination in the second trimester can assess fetal growth according to standardised growth charts. To achieve a reproducible and accurate measurement, a sonographer needs to identify three standard 2D planes of the fetus anatomy (head, abdomen, femur) and manually mark the key anatomical landmarks on the image for accurate biometry and fetal weight calculation. This can be a time-consuming operator-dependent task, especially for a trainee sonographer. Computer-assisted techniques can help in automating the fetal biometry computation process. In this paper, we present a unified automated framework for estimating all measurements needed for the fetal weight assessment. The proposed framework semantically segment the key fetus anatomies using state-of-the-art segmentation models, followed by region fitting and scale recovery for the biometry estimation. We present an ablation study of segmentation algorithms to show their robustness through 4-fold cross-validation on a dataset of 349 ultrasound standard plane images from 42 pregnancies. Moreover, we show that the network with the best segmentation performance tends to be more accurate for biometry estimation. Furthermore, we demonstrate that the error between clinically measured and predicted fetal biometry is lower than the permissible error during routine clinical measurements.

Keywords: Fetal biometry estimation · Fetal ultrasound · Fetus anatomy segmentation · computer-assisted diagnosis.

1 Introduction

There is little global consensus on how to train, assess and evaluate skills in prenatal second trimester ultrasound screening. Recommended assessment and quality control metrics varying across countries and institutions [5]. Despite this, standardised ultrasound planes and metrics to assess fetal growth are well established [20]. In particular, weight estimation is routinely used to assess fetal well-being, both in terms of its absolute value and its growth trajectory during pregnancy. It is considered by Obstetricians for scheduling birth and by neonatologists when counselling parents on likely outcomes for their baby. There are three key planes, namely, transventricular plane in the head, the transabdominal plane in the abdomen and the femur in the leg, which are used for the estimation of fetal weight (Fig. 1). The acquisition of these standard planes is subject to

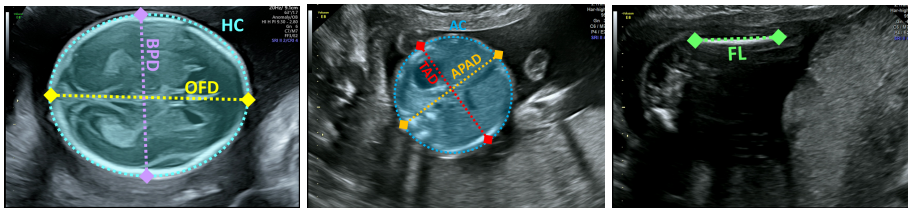


Fig. 1: Fetal biometry from (left) transventricular plane in the head, (middle) transabdominal plane in the abdomen and (right) femur plane.

intraoperator and interoperator variabilities [22] which introduces some degree of uncertainty in the obtained weight measurements and consequently requires a degree of caution when clinicians are interpreting fetal growth reports. Sonography expertise has a significant impact on minimising variability of image quality and fetal biometry [5]. Consequently, training and competence assessment are of great importance to ensure effective, reproducible and safe clinical practice. Automating fetal biometry on the standardised planes can help in minimising the variability specially in the case of less experienced sonographers and may also serves as expert for trainees.

There is extensive work on segmentation of anatomical structures in standard ultrasound planes, specifically those concerning second and third trimester screening [19]. These techniques can support automated fetal biometry, including measurements on the head [24, 13, 16, 23, 15, 4], femur [15, 12], and abdominal section [14]. These methods, however, rely on prior knowledge of which measurement to perform on a given image. A fully automated biometry system should both identify which standard plane is being imaged and whether it is of sufficient quality to perform the relevant measurements. Automatic image quality assessment has been investigated, including adequate magnification, symmetry and the visibility of relevant anatomical structures within the image [17, 15]. Such methods together with classification of standard planes [1] can be used to extract appropriate planes for fetal biometry from ultrasound video or image collections [9]. Alternative approaches involve obtaining standard planes from 3D ultrasound volumes [10], in which the extracted planes approach those of an experienced sonographer but results are so far limited to the fetal head measurements. Standard plane classification has also been further developed to provide active guidance during freehand operation [6].

In this paper, we propose to perform all the relevant measurements for fetal weight assessment within a unified automated system, which is our main contribution. The proposed AutoFB framework involves processing the three standard planes (transventricular, transabdominal, and femur) to segment the head, abdomen and femur, and the extraction of the following measurements: biparietal diameter (BPD), occipito-frontal diameter (OFD), head circumference (HC), transverse abdominal diameter (TAD), anterior-posterior abdominal diameter (APAD), abdominal circumference (AC), and femur length (FL). We achieve

this by training a multi-class segmentation neural network that automatically identifies and segments the relevant anatomy structures within any of these three standard planes. The corresponding biometry is then extracted by applying scale recovery and using ellipse fitting (head or abdomen) and bounding box fitting (femur). To the best of our knowledge, AutoFB is the first framework to automate fetal biometry estimation from all three standard planes. Both segmentation model evaluation and fetal biometry analysis are performed using the ultrasound data we acquired which contains 346 2D ultrasound planes from 42 pregnancies.

2 Fetal Biometry

To obtain fetal weight estimates the sonographer must navigate the ultrasound probe to localise a view of each of the three standard planes.While this task is subject to operator variability, there are established guidelines on which features should be visible within each standard plane [2]. They must then lock the display and manually place calipers on key landmarks from which biometric measurements are extracted. The Biparietal Diameter (BPD) and Occipitofrontal Diameter (OFD) measurements are required for the Head Circumference (HC) measurement on the transventricular plane (Fig. 1(a)). Transverse Abdominal Diameter (TAD) and Anterio-Posterior Abdominal Diameter (APAD) are required for the Abdominal Circumference (AC) measurement on the transabdominal plane (Fig. 1(b)). HC and AC are then computed using, $\pi(d_1 + d_2)/2$, where d_1 and d_2 are the BPD and OFD in the case of head and TAD and APAD in the case of abdomen measurements, respectively. Alternatively, an ellipse fitting function can be used for head and abdominal measurements, however, its usage will largely depend on operator choice or established practice within a specific clinical site. This feature is not routinely used in the context of data acquired and presented within this work. To measure the length of the femur (FL) the extreme lateral edges, including both of the epiphyses must be visualised. A single measurement along the long axis of the femur is saved as FL (Fig. 1(c)).

3 Methodology

An overview of the proposed framework is presented in Fig. 2. The framework involves training state-of-the-art segmentation models for identifying the head, abdomen and femur anatomies and selecting the best performing architecture (Sec. 3.1). This is followed by shape fitting on the segmented regions, automated image scale retrieval and biometry estimation in millimetres units (Sec. 3.2).

3.1 Multi-class Image Segmentation

In order to build a unified system, we define our problem as semantic segmentation between 4 specific classes: head, abdomen, femur, and background. With groundtruth data, each standard plane will only contain background and one of

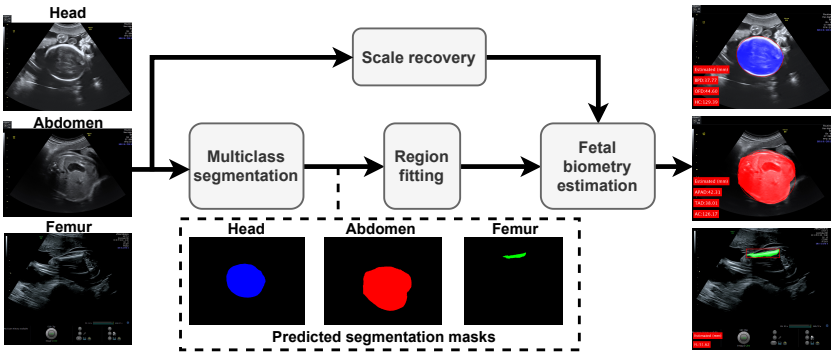


Fig. 2: Overview of the proposed AutoFB framework. Given an ultrasound image of a standard plane, the AutoFB performs multi-class segmentation, followed by shape fitting and scale recovery for the biometry estimation.

the other 3 classes. We experimented with two state-of-the-art image segmentation models, namely, U-Net [18] and Deeplabv3+ [3]. U-Net can be regarded as the most commonly used architecture for biomedical image segmentation and is recommended when the training data is limited. Deeplabv3+ has achieved state-of-the art performance on large-scale semantic segmentation datasets (PASCAL VOC 2012). Both U-Net and Deeplabv3+ are encoder-decoder networks, where U-Net is a special case in which the decoder component is connected with the encoder through skip connections and is not decoupled from the encoder.We briefly introduce these architectures and refer the reader to [18, 3] for specific details.

U-Net is a type of fully convolutional network which consists of a contraction path and an expansion path. The contraction path can be a pretrained encoder which captures the context while limiting the feature map size. The expansion path is a symmetric decoder network which also performs up-sampling to recover the segmentation map size. The encoder and decoder paths are connected through skip connections for sharing localisation information. We used the ResNet50 [11] as the encoder architecture for U-Net. We also experimented with Mobilenetv2 [21] to have a fair comparison of the two segmentation architectures under analysis. Deeplabv3+ [3] uses several parallel atrous convolutions (also known as dilated convolutions) with different rates to capture the contextual information at multiple scales without losing image resolution. This approach is referred as Atrous Spatial Pyramid Pooling. Moreover, Deeplabv3+ recovers the detailed object boundaries through a simple yet effective decoder module [3]. We used MobileNetv2 as the backbone for DeeplabV3+ as this backbone is both light-weight and effective.

We use cross entropy (CE) as loss function. From Table 1, we can observe that the data is highly imbalanced, with the femur class having much fewer samples compared to head, abdomen and background classes due to its comparatively small segmentation area. To handle this issue, we also use weighted CE (wCE) where given the total number of pixels per class, $[c_i]_i^4$, weight w_i for the i^{th} class

is given by, $w_i = \frac{max([c_i]_i^4)}{c_i}$. The obtained results are presented in Table 2 and are discussed Sec. 5.

3.2 Fetal Biometry Estimation

Different standard planes require different biometry measurements, and therefore the first step is to identify the visualised plane. This is defined as the largest segmented area provided by the networks described in the previous section. We later show experimentally that this strategy correctly identifies all planes in our test data. It is known a priori that the head and abdomen are elliptical while the femur is longitudinal (Fig. 1). Thus, ellipse fitting is performed on the segmented head and abdomen masks using [8], where the major and minor axes of the fitted ellipse represent BPD and OFD for the head and TAD and APAD for the abdomen. These are in turn used to calculate the circumference of the fitted ellipses, providing HC and AC measurements. On the femur plane, a bounding box with zero orientation is fitted on the segmented mask, where the length of its diagonal gives the FL estimate. Finally, lengths in pixels are scaled to millimetres to obtain results that are directly comparable to clinically obtained measurements.

While the metric scale of the ultrasound images (in px/mm) is usually trivial to obtain during operation, the automatic extraction of this parameter from retrospectively acquired data proved useful to fully automate the hundreds of measurements obtained in this work. To achieve this we exploit the consistent interface of the ultrasound machine used to acquire our dataset (GE Voluson), namely the caliper visible on the left-hand side of the ultrasound images. The ruler markers are detected with simple template matching and their smallest interval (can be either 5mm or 10mm) is determined from the relative size of the markers. This automatic scale recovery could be bypassed if the system was deployed on an ultrasound machine with direct access to its parameters.

4 Dataset and Experimental Setup

The data collection process has been reviewed and approved by the local research ethics committee. Patients attending the hospital for ultrasound examination were enrolled and pseudoanonomysed by the clinical research staff. The complete image library from each ultrasound was transferred to the research database. All ultrasound images saved by the operator were considered to be the optimal image for that scan and of diagnostic quality. The measurement calipers were applied by the ultrasound operator and in most cases, image with and without the measurement calipers were saved. A subset of images relevant to fetal biometry were extracted from the database by a clinical research fellow. A total of 346 images were included from 42 pregnancies. Each image in the set data was classified as AC, HC or FL. The VIA annotation tool [7] was used to annotate the head, abdomen or femur within each image for the segmentation task. The obtained fully anonymised standard ultrasound plane images have large intra-class variability.

Table 1: Total number of sample in each segmentation class and in each cross-validation fold and average pixels per class per frames.

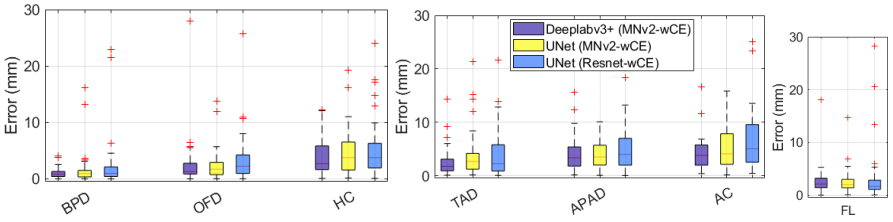
	All images	Fold 1	Fold 2	Fold 3	Fold 4	Avg. pixels per class per frame	
Total subjects	42	10	9	12	11		
Total Images	346	87	86	89	84	**Background**	816239
Head	135	26	44	29	36	**Head**	74127
Abdomen	103	32	22	26	23	**Abdomen**	44691
Femur	108	29	20	34	25	**Femur**	3833

Table 2: Four fold cross-validation results showing comparison of Deeplabv3+ and UNet having different configurations. Mean and standard deviation of mIoU across all folds is reported. Key: BG- background; H - head; A - abdomen; F - femur; CE - cross entropy; wCE - weighted cross entropy; MNv2 - Mobilenetv2.

Method	mIoU	mIoU-BG	mIoU-H	mIoU-A	mIoU-L
Deeplabv3+ (MNv2-CE)	0.87 ± 0.02	0.95 ± 0.02	0.93 ± 0.02	0.89 ± 0.03	0.61 ± 0.03
Deeplabv3+ (MNv2-wCE)	**0.88±0.01**	**0.95±0.01**	**0.93±0.02**	**0.89±0.02**	**0.61±0.02**
UNet (MNv2-CE)	0.82 ± 0.05	0.93 ± 0.03	0.89 ± 0.05	0.85 ± 0.05	0.56 ± 0.03
UNet (MNv2-wCE)	0.86 ± 0.01	0.94 ± 0.01	0.91 ± 0.02	0.86 ± 0.02	0.58 ± 0.01
UNet (Resnet-CE)	0.75 ± 0.06	0.88 ± 0.05	0.84 ± 0.07	0.77 ± 0.05	0.53 ± 0.03
UNet (Resnet-wCE)	0.78 ± 0.04	0.87 ± 0.03	0.83 ± 0.04	0.75 ± 0.06	0.53 ± 0.02

For example, in some cases the femur is well aligned to the centre of the plane while in other cases it appears as a small and distant object.

We divide the data into 4 folds such that each fold contains at least 80 images and unique subject IDs, thus the data in a fold is unseen for all other folds. The dataset distribution is mentioned in Table 1. We performed 4-fold cross-validation for testing the robustness of the segmentation networks. Mean Intersection over Union (mIoU) is used for evaluating the segmentation models and absolute error between the clinically obtained and predicted fetal biometry is used for evaluating the proposed autoFB. All images are of varying sizes (resolution) as they were cropped to remove any identifiable information. Therefore, we resized all images to 1024×1024 pixel resolution before model training. Data augmentation is applied by introducing random scale, rotation, shift, flipping, brightness and contrast changes before obtaining an image crop of size 512×512 pixel at a random location which is used as the input for training the segmentation network. Data augmentation helped in avoiding model over-fitting. An initial learning rate $10e^{-3}$ with a step decay by a factor of 1/10 at 75^{th} and 150^{th} is used with the ADAM optimiser. The model is trained for 600 epochs with early stopping based on the criteria of no improvement of the training set with patience of 50 epoch is used. The weights that captured the best performance on the training data are used to evaluate the segmentation model on the holdout fold. The segmentation networks are implemented in PyTorch and trained using a single Tesla V100-DGXS-32GB GPU of an NVIDIA DGX-station.



(a) Head (b) Abdomen (c) Femur

Fig. 3: Boxplots for the absolute error between the clinically measured and predicted fetal biometry. The comparison is only shown for the best performing model in each configuration.

5 Results and Discussion

We perform comparison of the Deeplabv3+ and U-Net having two commonly used backbones and used both CE and wCE losses (refer to Sec. 3.1). The quantitative comparison using 4-fold cross-validation is presented in Table 2. Both configurations of Deeplabv3+ are comparable (overall mIoU = 0.88) though the standard deviation is lower when wCE is used. Deeplabv3+ also outperformed the UNet configurations. The effect of introducing wCE loss for handling class imbalance problem is more evident from the different UNet configurations. Mobilenetv2 backbone, which has significantly less number of network parameters (3.5M), showed superior performance than the Resnet50 (26M parameters) backbone. Selecting an efficient and robust backbone architecture is essential and can significantly improve the overall segmentation network performance. From Table 2, we can observe that mIoU-F is particularly low compared to the mIoU-BG, mIoU-H and mIoU-A. This is because (1) the number of per-pixel samples in the femur class are very small (Table. 1); (2) a small error in predicted segmentation vs the ground-truth results in a significantly low IoU value when the object size is small; (3) of large intraclass variability.

Figure 3 shows the boxplots for the absolute error between the clinically measured and predicted biometry. The error in head measurements are the lowest, with a median of 0.80mm for BPD, 1.30mm for OFD and 2.67mm for HC and fewer outliers compared to other methods when segmentation masks from Deeplabv3+ (Mobilev2+wCE) are used (Fig. 3(a)). A similar trend is observed for the abdomen measurements, with a median of 2.39mm for TAD, 3.82mm for APAD and 3.77mm for AC (Fig. 3(b)). FL showed comparable results with a median of 2.1mm for Deeplabv3+ (Mobilenet-v2+wCE) but with fewer outliers (Fig. 3(b)). It is worth mentioning that the obtained error is less than the $\pm 15\%$ error permissible in the ultrasound assessment [22]. Figure 4 presents the qualitative comparison of the segmentation methods, depicting cases where either one or all methods fail in estimating the biometry due to inaccurate segmentation.

From a clinical point of view, successful interpretation of clinical ultrasound images requires an understanding that the fetus, a 3D object, fixed in neither

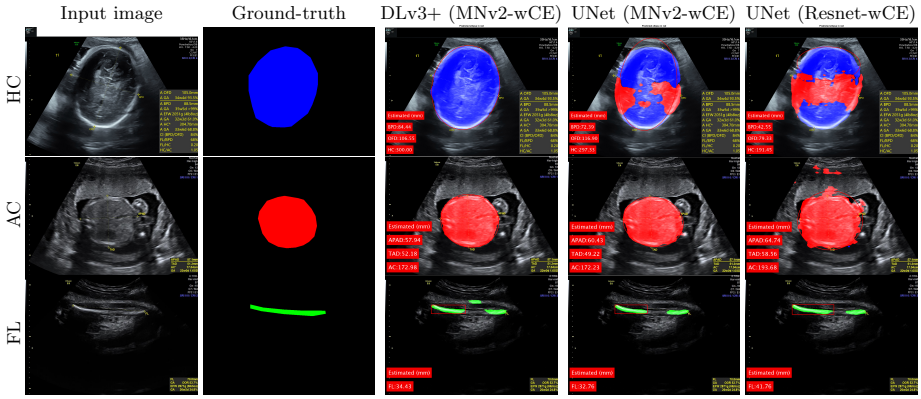


Fig. 4: Qualitative comparison of segmentation methods showing scenarios where inaccurate segmentation resulted in fetal biometry estimation failure. (Row 1 and 2) HC and AC examples where UNet resulted in inaccurate segmentation. (Row 3) FL example where all three methods failed. This image corresponds to the only outlier which is visible in Fig. 3(c) Deeplabv3+ error plot.

time nor space is being represented on a 2D grey-scale. Operator experience, combined with the effects of probe motion and homogeneity of ultrasound images contributes to high inter- and intra-operator variability. Ultrasound is used extensively in the assessment and management of pregnancies at high risk of fetal growth disorders. Appropriate management of these cases requires high quality assessment and reproducible assessment of fetal weight which can be achieved through AutoFB as demonstrated from the obtained results.

6 Conclusion

We proposed AutoFB, a unified framework for estimating fetal biometry given the three standard ultrasound planes. The proposed framework exploited the existing segmentation networks for obtaining the segmentation masks for the head, abdomen and femur. Head and abdomen were modelled as an ellipse with their major and minor axes and circumference providing an estimate for the respective measurements. Femur length was modelled as the diagonal on a rectangle fitted onto the segmentation mask. Through retrospective scale recovery and shape fitting, we obtained the fetal biometry estimates. Comparison of the predicted versus clinically measured fetal biometry showed that the error in HC (2.67mm), AC (3.77mm) and FL (2.10mm) were minimal and was better than the $\pm 15\%$ error that is typically acceptable in fetus ultrasound assessment. Future work involves increasing the training data size for further improving the segmentation and integrating AutoFB with the standard ultrasound plane detection [1] framework. Moreover, comparing experts and novices performance with the AutoFB can provide evidence supporting its clinical translation.

References

1. Baumgartner, C.F., Kamnitsas, K., Matthew, J., Fletcher, T.P., Smith, S., Koch, L.M., Kainz, B., Rueckert, D.: Sononet: real-time detection and localisation of fetal standard scan planes in freehand ultrasound. IEEE transactions on medical imaging **36**(11), 2204–2215 (2017)
2. Cavallaro, A., Ash, S.T., Napolitano, R., Wanyonyi, S., Ohuma, E.O., Molloholli, M., Sande, J., Sarris, I., Ioannou, C., Norris, T., et al.: Quality control of ultrasound for fetal biometry: results from the intergrowth-21st project. Ultrasound in Obstetrics & Gynecology **52**(3), 332–339 (2018)
3. Chen, L.C., Zhu, Y., Papandreou, G., Schroff, F., Adam, H.: Encoder-decoder with atrous separable convolution for semantic image segmentation. In: Proceedings of the European conference on computer vision (ECCV). pp. 801–818 (2018)
4. Chen, X., He, M., Dan, T., Wang, N., Lin, M., Zhang, L., Xian, J., Cai, H., Xie, H.: Automatic measurements of fetal lateral ventricles in 2d ultrasound images using deep learning. Frontiers in neurology **11**, 526 (2020)
5. Dromey, B.P., Ahmed, S., Vasconcelos, F., Mazomenos, E., Kunpalin, Y., Ourselin, S., Deprest, J., David, A.L., Stoyanov, D., Peebles, D.M.: Dimensionless squared jerk: An objective differential to assess experienced and novice probe movement in obstetric ultrasound. Prenatal Diagnosis (2020)
6. Droste, R., Drukker, L., Papageorghiou, A.T., Noble, J.A.: Automatic probe movement guidance for freehand obstetric ultrasound. In: International Conference on Medical Image Computing and Computer-Assisted Intervention. pp. 583–592. Springer (2020)
7. Dutta, A., Zisserman, A.: The VIA annotation software for images, audio and video. In: Proceedings of the 27th ACM International Conference on Multimedia. MM '19, ACM, New York, NY, USA (2019)
8. Fitzgibbon, A., Pilu, M., Fisher, R.B.: Direct least square fitting of ellipses. IEEE Transactions on pattern analysis and machine intelligence **21**(5), 476–480 (1999)
9. Gao, Y., Beriwal, S., Craik, R., Papageorghiou, A.T., Noble, J.A.: Label efficient localization of fetal brain biometry planes in ultrasound through metric learning. In: Medical Ultrasound, and Preterm, Perinatal and Paediatric Image Analysis, pp. 126–135. Springer (2020)
10. Grandjean, G.A., Hossu, G., Bertholdt, C., Noble, P., Morel, O., Grangé, G.: Artificial intelligence assistance for fetal head biometry: assessment of automated measurement software. Diagnostic and interventional imaging **99**(11), 709–716 (2018)
11. He, K., Zhang, X., Ren, S., Sun, J.: Deep residual learning for image recognition. In: Proceedings of the IEEE conference on computer vision and pattern recognition. pp. 770–778 (2016)
12. Hermawati, F., Tjandrasa, H., Sari, G.P., Azis, A., et al.: Automatic femur length measurement for fetal ultrasound image using localizing region-based active contour method. In: Journal of Physics: Conference Series. vol. 1230, p. 012002. IOP Publishing (2019)
13. van den Heuvel, T.L., de Bruijn, D., de Korte, C.L., Ginneken, B.v.: Automated measurement of fetal head circumference using 2d ultrasound images. PloS one **13**(8), e0200412 (2018)
14. Khan, N.H., Tegnander, E., Dreier, J.M., Eik-Nes, S., Torp, H., Kiss, G.: Automatic measurement of the fetal abdominal section on a portable ultrasound machine for use in low and middle income countries. In: 2016 IEEE International Ultrasonics Symposium (IUS). pp. 1–4. IEEE (2016)

15. Khan, N.H., Tegnander, E., Dreier, J.M., Eik-Nes, S., Torp, H., Kiss, G.: Automatic detection and measurement of fetal biparietal diameter and femur length—feasibility on a portable ultrasound device. Open Journal of Obstetrics and Gynecology **7**(3), 334–350 (2017)
16. Li, J., Wang, Y., Lei, B., Cheng, J.Z., Qin, J., Wang, T., Li, S., Ni, D.: Automatic fetal head circumference measurement in ultrasound using random forest and fast ellipse fitting. IEEE journal of biomedical and health informatics **22**(1), 215–223 (2017)
17. Lin, Z., Li, S., Ni, D., Liao, Y., Wen, H., Du, J., Chen, S., Wang, T., Lei, B.: Multi-task learning for quality assessment of fetal head ultrasound images. Medical image analysis **58**, 101548 (2019)
18. Ronneberger, O., Fischer, P., Brox, T.: U-net: Convolutional networks for biomedical image segmentation. In: International Conference on Medical image computing and computer-assisted intervention. pp. 234–241. Springer (2015)
19. Rueda, S., Fathima, S., Knight, C.L., Yaqub, M., Papageorghiou, A.T., Rahmatullah, B., Foi, A., Maggioni, M., Pepe, A., Tohka, J., et al.: Evaluation and comparison of current fetal ultrasound image segmentation methods for biometric measurements: a grand challenge. IEEE Transactions on medical imaging **33**(4), 797–813 (2013)
20. Salomon, L., Alfirevic, Z., Da Silva Costa, F., Deter, R., Figueras, F., Ghi, T., Glanc, P., Khalil, A., Lee, W., Napolitano, R., et al.: Isuog practice guidelines: ultrasound assessment of fetal biometry and growth. Ultrasound in Obstetrics & Gynecology **53**(6), 715–723 (2019)
21. Sandler, M., Howard, A., Zhu, M., Zhmoginov, A., Chen, L.C.: Mobilenetv2: Inverted residuals and linear bottlenecks. In: Proceedings of the IEEE conference on computer vision and pattern recognition. pp. 4510–4520 (2018)
22. Sarris, I., Ioannou, C., Chamberlain, P., Ohuma, E., Roseman, F., Hoch, L., Altman, D., Papageorghiou, A., Fetal, I., for the 21st Century (INTERGROWTH-21st), N.G.C.: Intra-and interobserver variability in fetal ultrasound measurements. Ultrasound in obstetrics & gynecology **39**(3), 266–273 (2012)
23. Sobhaninia, Z., Rafiei, S., Emami, A., Karimi, N., Najarian, K., Samavi, S., Soroushmehr, S.R.: Fetal ultrasound image segmentation for measuring biometric parameters using multi-task deep learning. In: 2019 41st Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC). pp. 6545–6548. IEEE (2019)
24. Zhang, L., Dudley, N.J., Lambrou, T., Allinson, N., Ye, X.: Automatic image quality assessment and measurement of fetal head in two-dimensional ultrasound image. Journal of Medical Imaging **4**(2), 024001 (2017)

AutoFB: Automating Fetal Biometry Estimation from Standard Ultrasound Planes (Supplementary Material)

Sophia Bano[1,2], Brian Dromey[1,3], Francisco Vasconcelos[1,2], Raffaele Napolitano[3], Anna L. David[3], Donald M Peebles[3], and Danail Stoyanov[1,2]

[1] Wellcome/EPSRC Centre for Interventional and Surgical Sciences(WEISS), University College London, London, UK
Department of Computer Science, University College London, London, UK
[2] Elizabeth Garrett Anderson Institute for Women's Health, University College London, London, UK

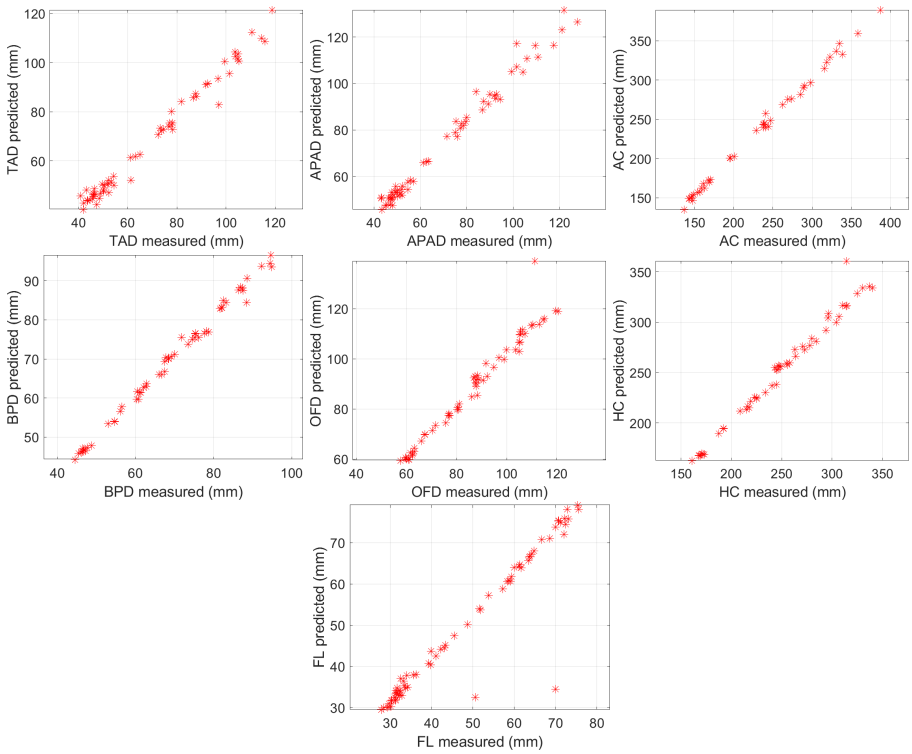


Fig. 1: Predicted fetal biometry from the best performing architecture (Deeplabv3+ with MobileNetv2) as reported in Table 2 and Fig. 3 versus clinically measured fetal biometry is shown plotted for TAD, APAD, AC, BPD, OFD, HC and FL. Observe that all measurements lie on a diagonal with a few outliers evident in OFD, HC and FL.