Physica Medica 88 (2021) 127–137

Contents lists available at ScienceDirect

Physica Medica

journal homepage: www.elsevier.com/locate/ejmp

Original paper

Automatic fetal biometry prediction using a novel deep convolutional network architecture

Mostafa Ghelich Oghli [a,b,*,1], Ali Shabanzadeh [a,*], Shakiba Moradi [a,1], Nasim Sirjani [a,1], Reza Gerami [c], Payam Ghaderi [a], Morteza Sanei Taheri [d], Isaac Shiri [e], Hossein Arabi [e], Habib Zaidi [e,f,g,h]

[a] Research and Development Department, Med Fanavarn Plus Co., Karaj, Iran
[b] Department of Cardiovascular Sciences, KU Leuven, Leuven, Belgium
[c] Radiation Sciences Research Center (RSRC), Aja University of Medical Sciences, Tehran, Iran
[d] R Department of Radiology, Shohada-e-Tajrish Hospital, Shahid Beheshti University of Medical Sciences, Tehran, Iran
[e] Division of Nuclear Medicine and Molecular Imaging, Geneva University Hospital, CH-1211 Geneva 4, Switzerland
[f] Geneva University Neurocenter, Geneva University, CH-1205 Geneva, Switzerland
[g] Department of Nuclear Medicine and Molecular Imaging, University of Groningen, University Medical Center Groningen, Groningen, Netherlands
[h] Department of Nuclear Medicine, University of Southern Denmark, Odense, Denmark

ARTICLE INFO

Keywords:
Fetal biometry
Ultrasound imaging
Deep learning
Convolutional neural network
Image classification

ABSTRACT

Purpose: Fetal biometric measurements face a number of challenges, including the presence of speckle, limited soft-tissue contrast and difficulties in the presence of low amniotic fluid. This work proposes a convolutional neural network for automatic segmentation and measurement of fetal biometric parameters, including biparietal diameter (BPD), head circumference (HC), abdominal circumference (AC), and femur length (FL) from ultrasound images that relies on the attention gates incorporated into the multi-feature pyramid Unet (MFP-Unet) network.

Methods: The proposed approach, referred to as Attention MFP-Unet, learns to extract/detect salient regions automatically to be treated as the object of interest via the attention gates. After determining the type of anatomical structure in the image using a convolutional neural network, Niblack's thresholding technique was applied as pre-processing algorithm for head and abdomen identification, whereas a novel algorithm was used for femur extraction. A publicly-available dataset (HC18 grand-challenge) and clinical data of 1334 subjects were utilized for training and evaluation of the Attention MFP-Unet algorithm.

Results: Dice similarity coefficient (DSC), hausdorff distance (HD), percentage of good contours, the conformity coefficient, and average perpendicular distance (APD) were employed for quantitative evaluation of fetal anatomy segmentation. In addition, correlation analysis, good contours, and conformity were employed to evaluate the accuracy of the biometry predictions. Attention MFP-Unet achieved 0.98, 1.14 mm, 100%, 0.95, and 0.2 mm for DSC, HD, good contours, conformity, and APD, respectively.

Conclusions: Quantitative evaluation demonstrated the superior performance of the Attention MFP-Unet compared to state-of-the-art approaches commonly employed for automatic measurement of fetal biometric parameters.

Introduction

Ultrasound is the modality of choice in prenatal diagnosis owing to its numerous advantages, including widespread availability, low cost, use of non-ionizing radiation and portability. It is the most commonly used method for two main purposes: fetal growth screening and assessment of pathologic and physiologic conditions. However, ultrasound has inherent limitations, like operator dependency, limited soft-tissue contrast, difficulty in the presence of low amniotic fluid, etc. [1]. Several conventional image processing and artificial intelligence-

* Corresponding authors at: 10th St. Shahid Babaee Blvd. Payam Special Economic Zone, Karaj, Iran.
E-mail addresses: m.g31_mesu@yahoo.com (M. Ghelich Oghli), shabanzadeh.ali@gmail.com (A. Shabanzadeh).
[1] Authors contributed equally to this manuscript.

https://doi.org/10.1016/j.ejmp.2021.06.020
Received 6 February 2021; Received in revised form 23 June 2021; Accepted 27 June 2021
Available online 6 July 2021

based approaches have been developed to overcome these limitations and reducing side effects in other medical imaging systems [2–4].

Monitoring of the fetal growth is performed using gestational age (GA) estimation, which is a function of fetal biometric parameters [5]. In this regard, measurement of fetal biometric parameters, including head circumference (HC), biparietal diameter (BPD), abdominal circumference (AC), and femur length (FL) is a prerequisite for this purpose. These standard biometric parameters, commonly reported on a routine second trimester scan, are defined based on fetus anatomy. For instance, BPD is defined as the diameter of the fetus skull from one parietal bone to the other and is measured on a transverse plane that contains the third ventricle and the thalami. The HC is measured on the same plane as BPD. The AC indicates the circumference of the fetal abdomen on an image acquired in the transverse section through the upper abdomen (containing fetal stomach, umbilical vein, and portal sinus) [6]. Finally, the FL denotes the distance from the head to the distal end of femur. Manual fetal biometric measurement is an error-prone and time-consuming procedure. Besides, it suffers from inter- and intra-sonographer variability. Therefore, there is an essential need for exploiting a robust and accurate method that measures the fetal biometric parameters automatically. This method improves the workflow and reduces user variability in measuring fetal biometric parameters. In this regard, a number of commerical software packages are available, including SonoBiometry [7], which exhibited accurate outcome, though they require manual invervention.

Machine learning (ML)-based techniques empowered novel potential clinical applications of medical imaging in recent years [8]. The outstanding capabilities of ML methods provide the potential to address the undeniable need for methods enabling the extraction of a complex hierarchy of features from images via their self-learning capacities [9]. Deep learning (DL) approaches that became popular in recent years can be trained to provide robust solutions to the variability in image quality/acquisition protocols, taking advantage of the processing power of graphics processing units [10]. These algorithms produce more generalizable and usually less interpretable features, as opposed to ML features that are designed in decomposable pipelines. Image segmentation and classification have been revolutionized by the introduction of DL algorithms [11].

U-net was proposed in 2015 for the segmentation of medical images with a limited dataset sample [12]. The network consisted of encoder (contraction) and decoder (expansion) paths and skip connections established between feature maps from the encoder section to the up-convolution layers at the same level in the decoder section. There are several extentions of U-net. Alom *et al.* [13] introduced RU-net and R2U-net, representing "recurrent convolutional neural network" and "recurrent residual convolutional neural network", respectively. In RU-net, there are recurrent convolutional layers [14] before the pooling layers and recurrent up-convolutional layers before up-convolution layers and before the output of the segmentation map. Conversely, in R2-U-net, the recurrent convolutional layers are replaced by residual recurrent convolutional layers. Oktay *et al.* [15] proposed another modification of the U-net architecture by adding AGs in the skip connection path and suggesting Attention U-net. They proposed grid-based gating that allows attention coefficients to be more specific to local regions. Furthermore, Lee *et al.* [16] combined the Attention U-net with R2U-net in an attempt to improve the overall performance of the network. In our previous work, we proposed a multi-feature pyramid U-net (MFP-Unet) [17], which takes the advantages of both U-net architecture and feature pyramid network (FPN) [18].

Most research studies in this field focused on fetal head segmentation owing to the availability of a general public dataset from fetal head circumference challenge [19], and the importance of biometric parameters related to the fetal head (i.e. HC and BPD). Heuvel *et al.* [19] proposed a pipeline composed of two main components summarized as pixel classifier and fetal skull detector. In the pixel classifier component, Haar-like features train a random forest classifier to locate the fetal skull. Then, the HC was extracted using Hough transform [20], dynamic programming and an ellipse fitting algorithm in the second component. The authors optimized three different systems that use one, two, and three pipelines to investigate the influence of gestational age in different trimesters on system performance. In another work, Sobhanina *et al.* [21] proposed a multi-task convolutional network based on Link-Net architecture [22] for the segmentation of fetal head and an optimization process to fit an ellipse over the segmented region. Other approaches attempted to segment fetal head and abdomen simultaneously. Sinclair *et al.* [23] trained a fully convolutional network (FCN) [24] over almost 2000 clinically annotated images and then optimized an ellipse to be fitted to the segmented region. They evaluated the performance of their method through comparison to intra- and inter-observer errors. Irene *et al.* [25] broke the problem of fetal head and abdomen segmentation into three steps. First, a region of interest is detected using YOLO algorithm [26]. Second, a Canny edge detector was applied to the resulting image and then a Hough transform [20] was utilized to detect the elliptic shape of the fetal head and abdomen. Finally, an efficient model, called the difference of Gaussian Revolved Along Elliptical Path (DoGell) was used to segment these regions [27]. The DoGell model is a fully automatic, image processing-based method aiming at segmenting the fetal head from original ultrasound images. Their method was based on minimizing a cost function between the observed image and a predefined surface. The surface revolves a difference of Gaussians along the elliptical path to model pixel values of the skull and surrounding areas.

A number of studies proposed a more general approach to predict additional fetal biometric parameters. For instance, Carneiro *et al.* [28] proposed a comprehensive system to detect and measure fetal anatomical structures including BPD, HC, AC, FL, humerus length (HL), and crown-rump length (CRL) automatically. They exploited atlas-based segmentation to train a constrained version of the probabilistic boosting tree [29]. Rahmatullah *et al.* [30] presented a method based on multilayer superpixel classification to segment the fetal head, femur, and humerus. They utilized a simple linear iterative clustering algorithm to generate square-shaped regions. Thereafter, three different features containing unary, shape, and image moments were extracted from each region. Finally, a random forest classifier was performed over a 5-fold cross-validation scheme.

This study proposes a comprehensive deep learning-based approach for prediction of BPD, HC, AC, and FL through automated segmentation of fetal head, abdomen, and femur. The proposed approach sets out to address the aforementioned challenges of ultrasound image segmentation while focusing on the following goals: (i) generalisability and versatility of the approach for the segmentation of all fetal anatomies with a wide range of variability, (ii) high accuracy of anatomy segmentation, and (iii) robustness to signal dropout and speckle noise.

To attain these goals, a novel and effective convolutional network architecture (multi-feature pyramid Unet: MFP-Unet), previously introduced for the segmentation of the left ventricle in echocardiography images [17], was upgraded and employed. In fetal ultrasound images, an object of interest representing the salient part of data can be normally defined. In this light, MFP-Unet was upgraded to automatically detect and focus on the object of interest without additional user intervention. To this end, an attention gate (AG) consisting of additional preceding object localization models to separate localization and subsequent segmentation steps [31] was incorporated into the MFP-Unet architecture. The AGs suppress feature activations in disjointed regions, thus increase the sensitivity and accuracy of the model with no additional computation burden. Overall, the contributions of this manuscript are threefold. First, we introduce a novel convolutional neural network architecture for the delineation of anatomical organs from ultrasound images. Second, we incorporate AGs into our previously introduced MFP-Unet network for the segmentation of fetal ultrasound images, which enables the network to focus on the object of interest within the images. Third, a preprocessing algorithm is proposed to remove irrelevant parts/structures in the fetal femur images to enforce

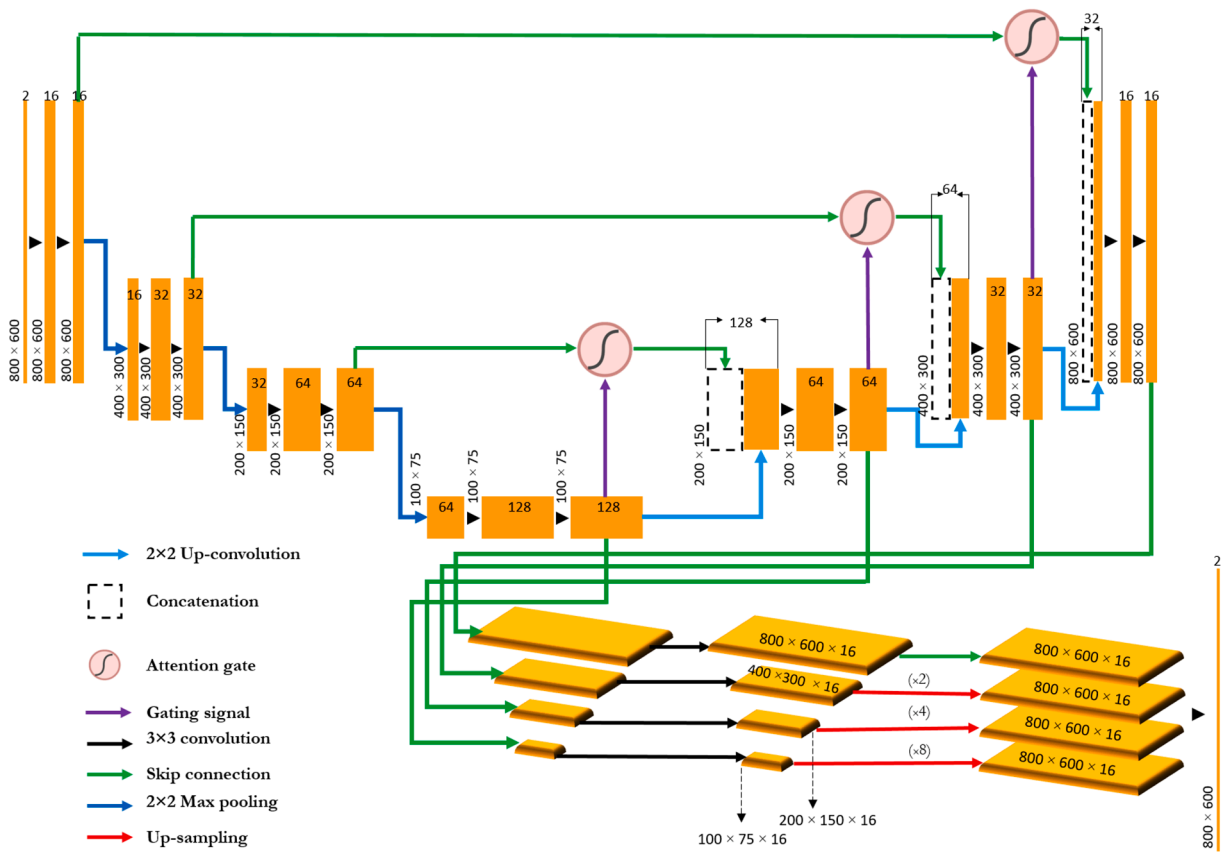

Fig. 1. Attention MFP-Unet architecture.

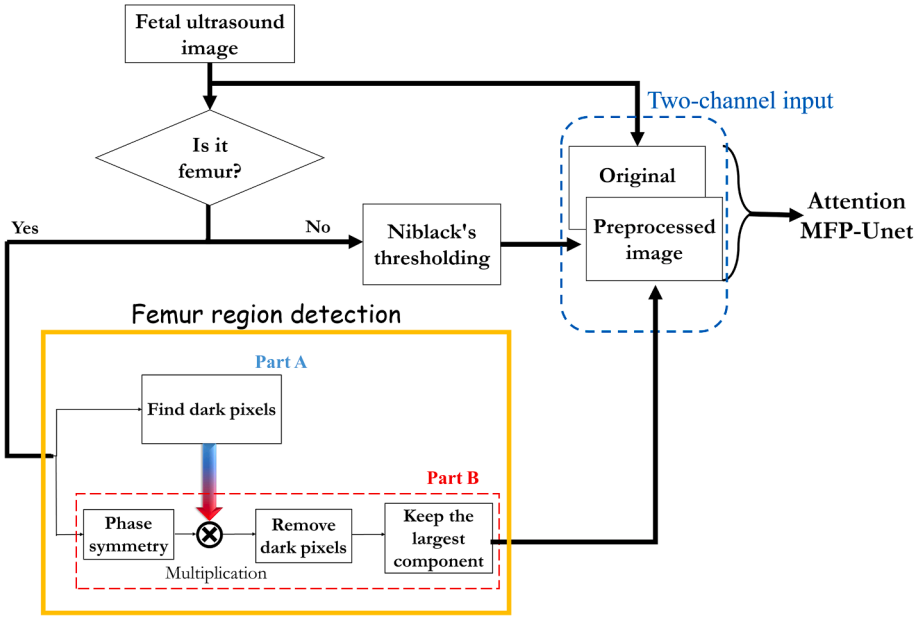

Fig. 2. The proposed algorithm for fetal image segmentation.

the Attention MFP-Unet to focus on the object of interest.

Materials and methods

Training and evaluation dataset

This work employed two distinct datasets, including a publicly-available and a local dataset. The first dataset is a large publicly

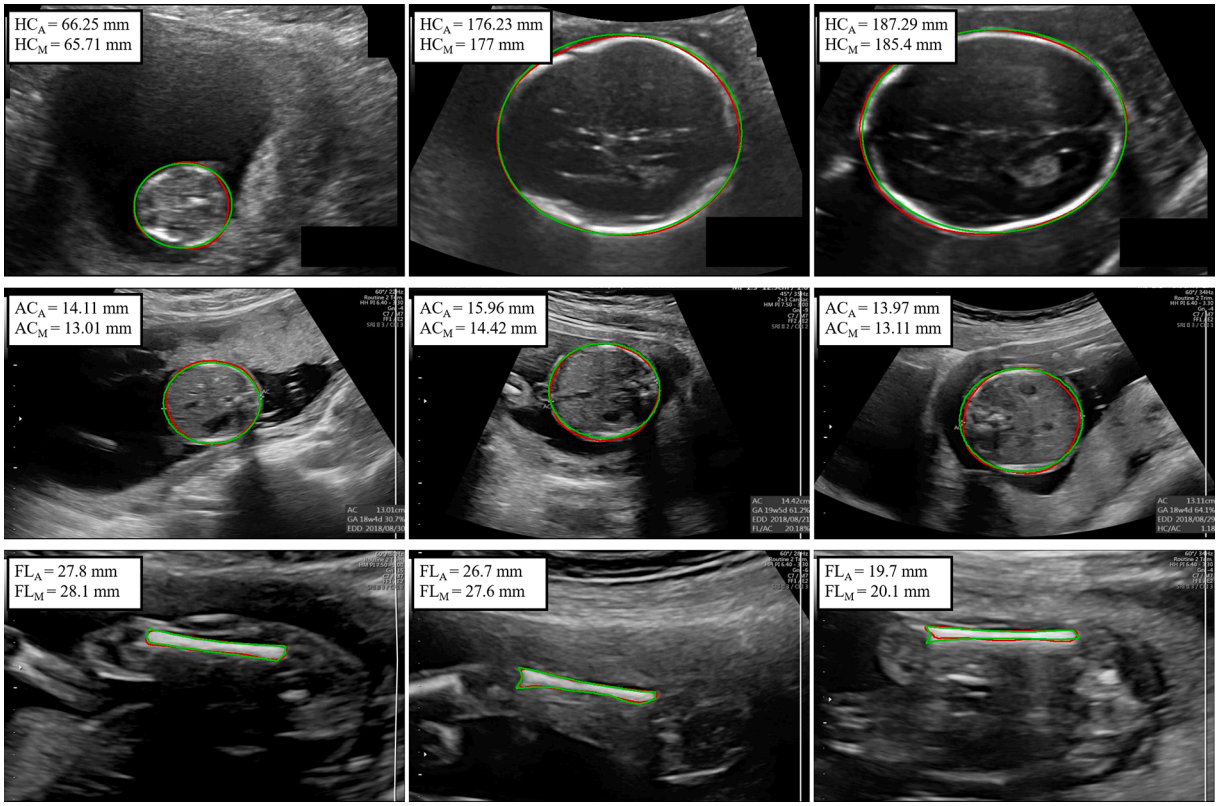

Fig. 3. Automatic (green) and manual (red) segmentations of fetal head (top row), abdomen (middle row) and femur (bottom row). The measured fetal biometry parameters using automatic and manual approaches are also shown. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

Table 1
Performance of the proposed network for fetal organ classification.

Data	Precision	Recall	F1 score
Train	1	1	1
Test	0.967	0.956	0.956

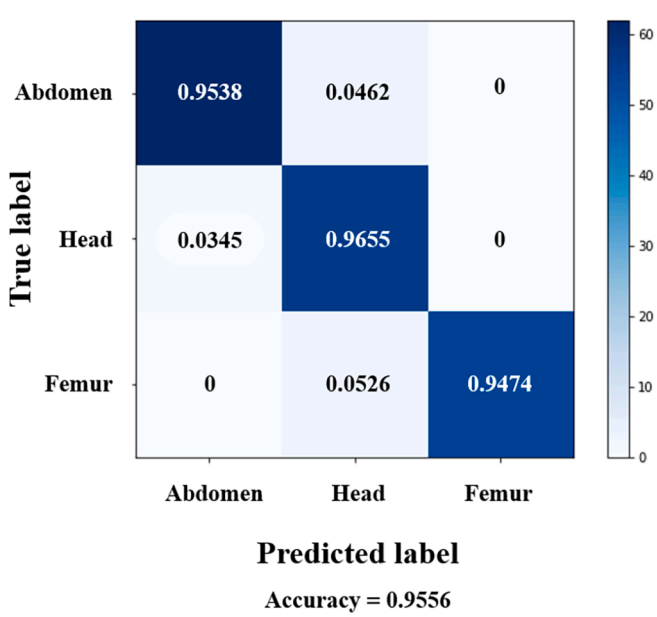

Fig. 4. Confusion matrix of classification network for test set.

available dataset for head circumference measurement from the Grand Challenge [19]. The second one is a local dataset consisting of the fetal abdomen, and femur images obtained from two different hospitals.

Description of the public dataset.

van den Heuvel et al. shared a dataset consisting of 1334 two-dimensional (2D) ultrasound images of the fetal head acquired at the Department of Obstetrics of the Radboud University Medical Center, Nijmegen, the Netherlands [19]. In total, ultrasound images of 551 pregnant women receiving a routine ultrasound-screening exam were included in this dataset. Expert sonographers using two high-end ultrasound machines, including Voluson E8 and Voluson 730 (General Electric, Austria) acquired the images between May 2014 and May 2015.

The whole dataset was divided into a training set of 999 images (75%) and a test set of 335 images (25%). Each 2D ultrasound image consisted of 800 × 540 pixels with a pixel size ranging from 0.052 to 0.326 mm. The annotated fetal head and measured circumference (in millimeters) were also provided. As the standard period for routine ultrasound screening for fetal biometry is the second trimester (i.e. between 14 and 26 weeks), most of the images were acquired during this period.

Description of the local dataset

A collection of 2D ultrasound images of the fetal abdomen and femur was prepared. To assure the sufficient image vriability in the training phase, different gestational ages were included in the image dataset. The elastic deformation method was used to augment the data by a factor of 10, since the dataset was not large enough for proper training of the network. The images were acquired from two distinct centers, including: (i) Alvand Medical Imaging Center, Tehran, Iran, and (ii) Laleh Hospital, Tehran, Iran. The ultrasound machines were Voluson E10 echocardiographic system (General Electric, Austria) with a C2-9-D XDclear probe.

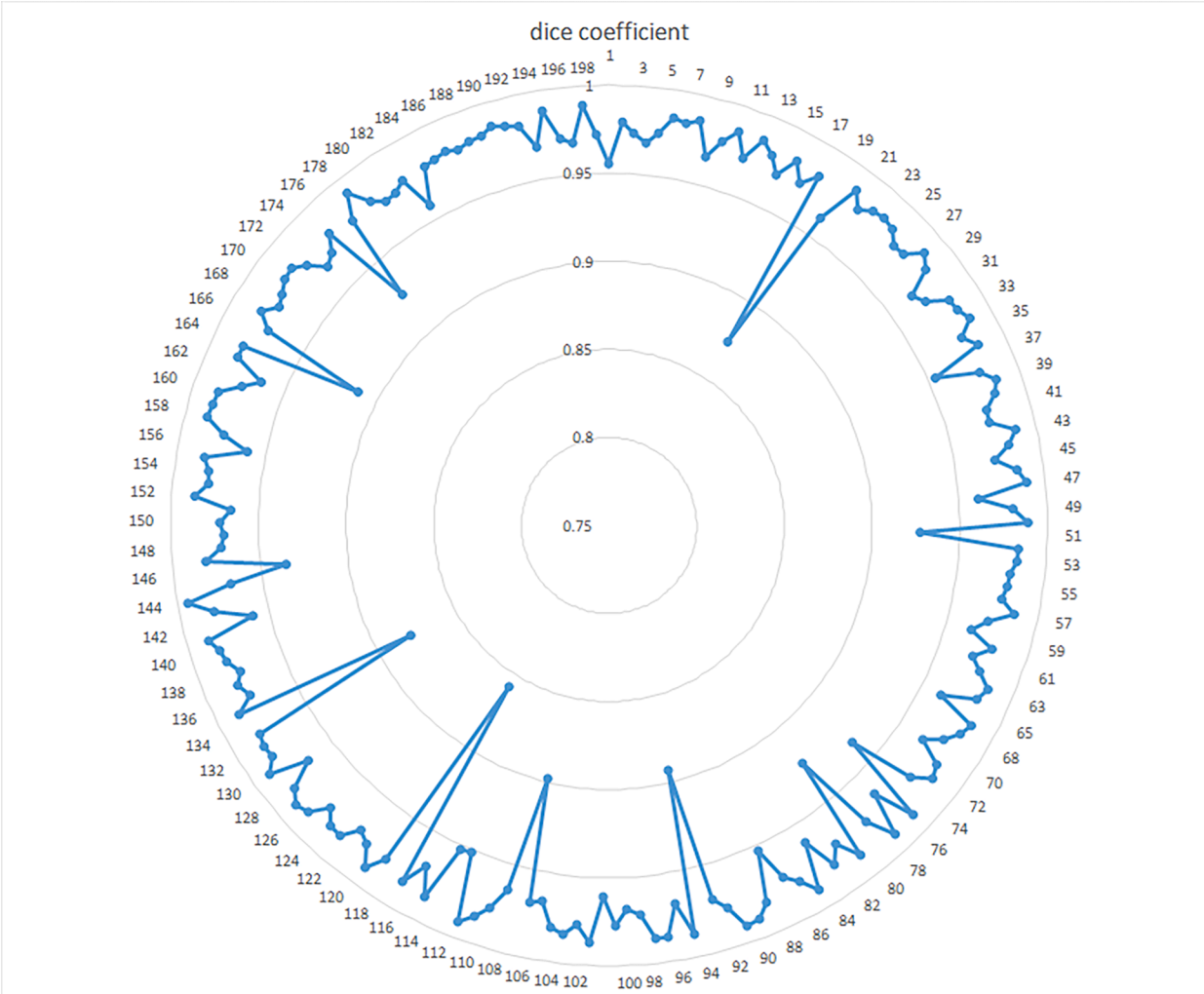

Fig. 5. Dice values obtained using the proposed approach for 198 subjects from the evaluation dataset.

Table 2
Performance of the proposed segmentation method compared with the different techniques using our dataset.

Method	Fetal organ	DSC[1]	HD[2] (mm)	Conformity	APD[3] (mm)	Good Contours (%)
Attention MFP-Unet	Abdomen	0.98	2.22	0.95	1.23	97.30
	Femur	0.91	1.14	0.80	0.20	100
MFP-Unet	Abdomen	0.95	4.50	0.86	1.58	94.87
	Femur	0.86	4.10	0.67	0.65	97.00
U-net	Abdomen	0.94	7.22	0.86	1.90	92.30
	Femur	0.84	3.50	0.62	0.27	100
Dilated U-net	Abdomen	0.94	4.08	0.86	1.46	94.87
	Femur	0.87	1.28	0.70	0.23	100
Attention U-net	Abdomen	0.95	3.87	0.88	1.28	**100**
	Femur	0.86	1.73	0.67	0.23	100
RU-net	Abdomen	0.98	3.84	0.95	1.50	100
	Femur	0.84	2.87	0.62	0.24	100
R2U-net	Abdomen	0.97	2.38	0.92	1.76	97.43
	Femur	0.85	2.98	0.65	0.27	100

[1] Dice Similarity Coefficient
[2] Hausdorff Distance
[3] Average Perpendicular Distance

Table 3
Segmentation performance of the proposed method compared with previously published works using the HC public dataset. Numbers format: mean value ± (standard deviation).

Method	DSC[1]	HD[2] (mm)	DF[3] (mm)	ADF[4] (mm)
Attention MFP-Unet	0.972 ± 0.12	2.67 ± 0.05	0.55 ± 4.72	2.35 ± 4.12
Heuvel *et al.* [19]	97.00 ± 2.80	2.00 ± 1.60	0.60 ± 4.30	2.80 ± 3.30
Sobhaninia *et al.* [21]	96.84 ± 2.89	1.72 ± 1.39	1.13 ± 2.69	2.12 ± 1.87
Ciurte *et al.* [44]	94.45 ± 1.57	4.60 ± 1.64	11.93 ± 5.32	–
Stebbing *et al.* [45]	97.23 ± 0.77	2.59 ± 1.14	−3.46 ± 4.06	–
Sun [46]	96.97 ± 1.07	3.02 ± 1.55	3.83 ± 5.66	–
Ponomarev *et al.* [47]	92.53 ± 10.22	6.87 ± 9.82	16.39 ± 24.88	–

[1] Dice Similarity Coefficient
[2] Hausdorff Distance
[3] Difference
[4] Absolute Difference

Table 4
Comparison the result of segmentation performance with and without femur detection algorithm.

Method	Fetal organ	DSC[1]	HD[2] (mm)	Conformity	APD[3] (mm)	Good Contours (%)
Attention MFP-Unet	Without femur detection algorithm	0.84	2.43	0.65	0.1	96
	With femur detection algorithm	0.91	1.14	0.80	0.20	100

[1] Dice Similarity Coefficient
[2] Hausdorff Distance
[3] Average Perpendicular Distance

The C2-9-D transducer has a bandwidth of 2.3–8.4 MHZ taking advantage of the XDclear technology, a single piezoelectric crystal (PMN-PT and PZN-PT) used to increase bandwidth compared to conventional PZT crystals. Two observers delineated the anatomical organs (abdomen and femur) using an in-house code developed in MATLAB 2018b (MathWorks, MA). The total dataset consisted of two parts: 158 2D ultrasound images of the fetal abdomen and 315 2D ultrasound images of the fetal femur. All images were split randomly into a training (75% of the dataset) and an evaluation (25% of the dataset) set. The pixel spacing parameter of the images is different in various datasets, ranging from 0.23 to 0.36 mm/pixel. The images were captured at a resolution of 800 × 600 pixels. The participants gave their written consent to participate in this study and the study was approved by the ethics committee of the institution.

MFP-Unet architecture

MFP-Unet is a hybrid architecture based on dilated U-net proposed for 2-D echocardiographic image segmentation [17]. The MFP stands for multi-feature pyramid combining the concept of feature pyramids with the dilated U-net network. Dilated kernels are sought due to their large receptive fields, which enable establishing relations between anatomical organs from medical images. Although convolutional kernels with higher dimensions could alternatively be chosen, dilated kernels are preferable because they keep the number of network parameters the same. In the U-net network, the input image is propagated through three different paths: encoder or contraction path, decoder or expansion path, and skip connections. At each level of abstraction in the contraction path, the resolution of feature maps decreases (by a factor of two), while the number of feature maps doubles after each level. Conversely, up-convolution layers in the expansion path up-samples the feature maps to enable precise localization. Finally, skip connections combine higher resolution features from the contraction path to the expansion path at the same‘ level in order to better localize and learn the prominent features.

The U-net architecture demonstrated superior performance compared to conventional models, such as FCN, for pixel-based image segmentation, which is also an effective solution with a limited dataset [12]. However, it overlooks the effect of feature maps in different scales “directly”. In other words, it lacks connections between the entire feature maps in the decoder path and the output segmentation map. This attribute of the U-net deprives the feature maps in different scales of the decoder path to share/propagate features, which may prevent the flow of information and result in unnecessary parameters. In MFP-Unet, we proposed to address this issue using the concept of FPN [18]. The intuition behind MFP-Unet was extracting feature maps from all levels of the decoder path and using them in the final pixel classification. To implement this idea, a 3 × 3 convolution layer was added after feature maps from all levels of the decoder path, followed by an up-sampling layer (with different factors for each layer) to fix the size of the feature maps. The last step was the concatenation of all the extracted feature maps and feeding them to a 1 × 1 convolution filter for the pixel-classification process.

The modification made on the MFP-Unet model enhanced notably its performance for anatomy segmentation from ultrasound images.

Attention MFP-Unet

As illustrated in Fig. 1, the main difference between Attention MFP-Unet and MFP-Unet is the integration of three AGs in the path of the skip connections. The AG proposed in [31] represented a soft attention module which does not add a large number of extra parameters. To be more specific, AG performs a series of operations on the feature maps of different layers, which can be summarized as follows:

1- The two input feature maps are convolved with 1 × 1 × 1 convolution kernels separately and the results passed through a ReLU activation function.
2- The resulting feature map of step 1 is convolved with 1 × 1 × 1 convolution kernel again followed by the application of a Sigmoid activation function.
3- A grid resampling of attention coefficients is performed using trilinear interpolation.
4- Next, concatenation is performed with the up-sampled feature maps at a lower level.
5- Finally, input feature-maps are multiplied by attention coefficients in an element-wise manner.

Attention coefficients determine the salient image regions to keep only the activations related to the object of interest. We utilized these AGs in the path of the skip connections at each level of contraction–expansion. The last feature maps of each level in both contraction and expansion paths are transmitted by a skip connection and treated as the input of AGs. Consequently, the result of the AG process is concatenated with the feature maps in the upper level of the expansion path. The AGs are integrated into the MFP-Unet architecture, which makes the network smarter by detecting the relevant (salient) parts of images automatically. This capability is extremely useful in fetal ultrasound segmentation because there is only one object of interest in these images (unlike echocardiography images of apical views that contain various heart chambers).

Proposed algorithm for fetal image segmentation

Fig. 2 shows an overview of the fetal image segmentation algorithm. The first stage is the identification of fetal anatomy. To this end, a simple but efficient convolutional neural network was trained on images of our dataset to delineate the fetal anatomical organs automatically. The details are described in Supplementary materials. Based on the selected anatomy, a specific pre-processing algorithm is performed on the original image and the outcome fed to the Attention MFP-Unet as the second channel of the input.

If the image contains a fetal head or abdomen, Niblack’s global thresholding is applied to make the borders on ultrasound images brighter [32]. Conversely, the proposed preprocessing algorithm described in the Supplementary material is performed if the input image contains a fetal femur. This algorithm is used due to the presence of deceptive structures, especially at the top of the image near the convex probe, that are strongly similar to the femur anatomy. These structures interfere with the saliency detection procedure performed by AGs, since they are also brighter than surrounding structures, like the femur. Our proposed algorithm relies on a hybrid approach for the detection of the region containing the femur in fetal images.

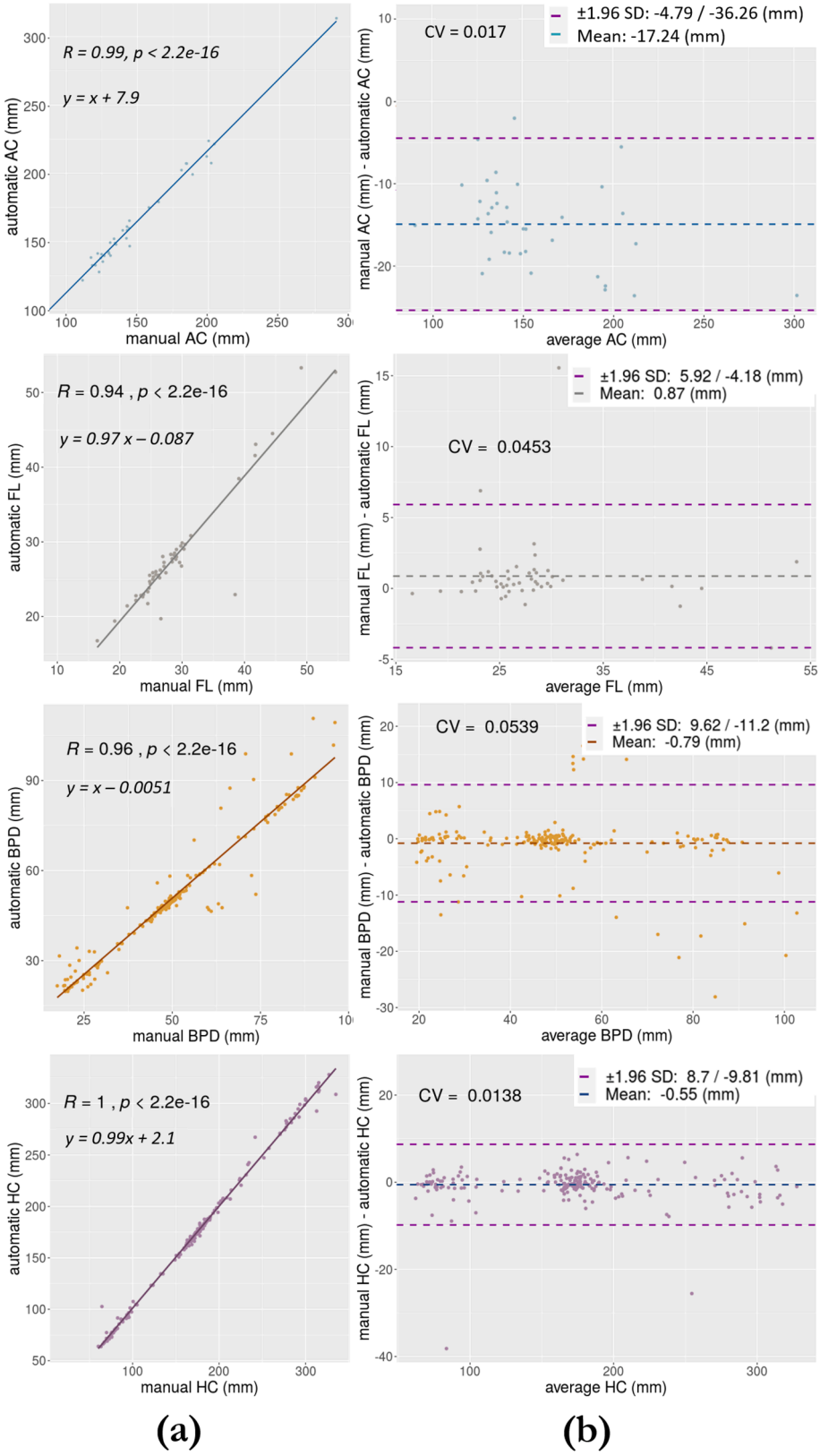

Fig. 6. Correlation analysis (a) and Bland-Altman plots (b) for fetal biometry parameters.

Fetal biometry

The final stage of our proposed algorithm is the extraction of fetal biometrical parameters, including BPD, HC, AC, and FL from the segmented regions. For fetal head and abdomen images, a direct method of ellipse fitting using the least square algorithm is used [33]. This method, an improved version of [34], reformulated the fitting task as a linear optimization problem with a quadratic constraint. Then, a

Table 5
Fetal biometry parameters obtained using the proposed method on the training and test sets of dataset 2. R is the correlation coefficient (ideal = 1) whereas a and b are coefficients of the linear regression fit (y = ax + b, ideally a = 1 and b = 0).

Dataset	BPD[1]	HC[2]	AC[3]	FL[4]
Training	R = 0.96 a = 0.99 b = 1	R = 1 a = 1 b = 0.91	R = 0.99 a = 1 b = 2.1	R = 0.98 a = 0.99 b = 1.3
Test	R = 0.96 a = 1 b = -0.0051	R = 0.96 a = 0.99 b = 2.1	R = 0.97 a = 1 b = 16	R = 0.94 a = 0.97 b = -0.09

[1] Biparietal Diameter
[2] Head Circumference
[3] Abdominal Circumference
[4] Femur Length

standard least-squares minimization solves the produced linear optimization problem directly. The main advantage of this algorithm [33] is the new formulation based on the block decomposition of matrices.

After fitting an optimized ellipse on the segmented region, five parameters of the fitted ellipse, including coordinates of the center (x, y), semi-major and semi-minor axes (a, b), and the ellipse angle (θ), are extracted. Consequently, the semi-minor axis of the fitted ellipse on the fetal head image would be BPD whereas the ellipse perimeter would be HC for the fetal head image and AC for the fetal abdomen. To extract FL, the following steps are followed:

1- The skeleton of the segmented region is extracted.
2- The contour of the skeleton is located.
3- The skeleton is fitted with a parabola curve/plane (since the most similar equation to the femur shape is a parabola).
4- The distance between femur line endpoints would indicate the FL size.

Implementation details

We used a system with 16 GB of RAM, a GPU based graphic card with 2176 CUDA cores (GeForce RTX 2060-A8G), and an Intel Xeon CPU. The network was implemented in the Python environment with Tensorflow r2 and Keras 2.2.4. To evaluate the performance of the model, an unbiased 5-fold cross-validation scheme was adopted through the following steps:

1- Stratification and partition of sonography images into five equally sized folds to ensure that each fold represents the whole dataset properly.
2- Performing five iterations of training and validation, such that within each iteration four folds of the data are used for the learning process, while one remaining fold is held out for validation. This procedure is performed on both "prepared" and "challenge" datasets.

Stochastic gradient descent algorithm with a weight decay of 0.0005 and a momentum of 0.9 is used in the training of Attention MFP-Unet. The learning rate was 0.001 whereas max epoch was 200.

Evaluation metrics

The evaluation of the proposed network was carried in two phases: first, the robustness and accuracy of the segmentation task were compared to the gold standard (manual annotations by experts). Second, the ability of the proposed network for predicting fetal biometric parameters was assessed in clinical routine. The proposed model was also compared with U-net, dilated U-net, Attention U-net, RU-net, R2U-net, MFP-Unet, and published results of other methods evaluated on the same dataset [19]. For segmentation task, the following measures were

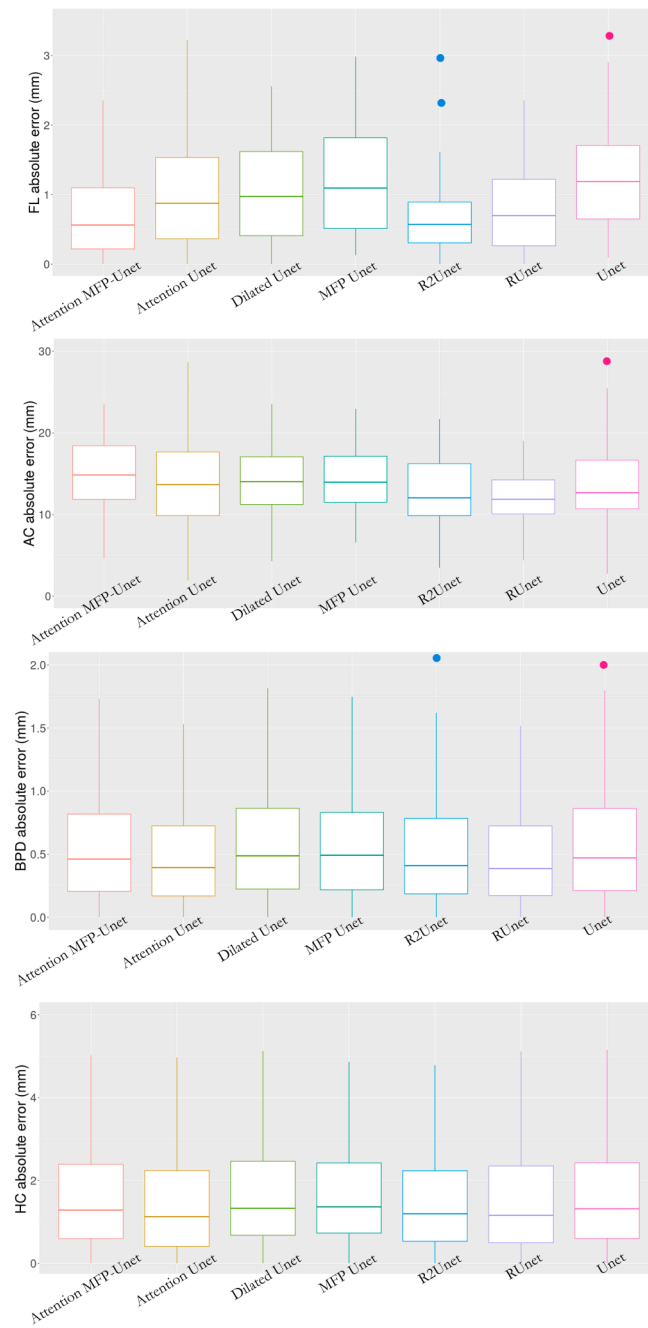

Fig. 7. Box plots of errors for fetal biometry parameters obtained using the different techniques.

calculated between manually drawn and predicted contours: Dice similarity coefficient (DSC) [35], Hausdorff distance (HD) [36], percentage of good contours, the conformity coefficient [37] and average perpendicular distance (APD) were calculated. The percentage of good contours is defined based on a recommendation of Radau et al. [38] that an APD less than 5 mm is an indicator of "good" segmentation. The DSC (Eq. (1)) measures the similarity between manually segmented region (AM), and automatically segmented region (AA).

$$DSC = \frac{2(A_A \cap A_M)}{A_A + A_M} \tag{1}$$

HD measures the maximum distance of contour A to the nearest point in contour B.

$$HD = max\left(\underset{a\in A}{max}\left(\underset{b\in B}{min}(a,b)\right), \underset{b\in B}{max}\left(\underset{a\in A}{min}(a,b)\right)\right) \quad (2)$$

where d(.,.) represents the Euclidean distance and A and B are manual and automatic defined contours.. The conformity coefficient (CC) is defined as the number of miss-segmented pixels divided by the number of correctly segmented pixels [37].

$$CC = \frac{3DSC - 2}{DSC} \quad (3)$$

APD measures the distance between the manually drawn contour and the automatically segmented contour, averaged over all contour points. Higher values of APD indicate that the two contours do not match closely.

For the second phase of the evaluation, four fetal biometric parameters (BPD, HC, AC, and FL) were computed on the automatic and manual segmentations. Bland-Altman analysis [39] and Pearsons' test were employed to perform the correlation analysis of the results. The coefficient of variation (CV) defined in Eq. (4) and the reproducibility coefficient (RPC), defined as standard deviation of the differences between manual and automatic results multiplied by 1.96, were also calculated.

$$CV = \frac{SD(auto - man)}{mean(auto) + mean(man)} \quad (4)$$

In Eq. (4), SD (auto – man) is the standard deviation of the differences between automatic and manual segmentation results, while mean (auto) and mean (man) are their mean values. In addition, the difference (DF) and the absolute difference (ADF) of predicted HC values in the public dataset were estimated to assess the performance of the proposed segmentation method and compare the results with other participants in the challenge [19]. DF is defined as:

$$DF = HC_M - HC_A \quad (5)$$

where HCM is the HC measured by an expert radiologist and HCA is the HC calculated automatically. ADF was defined as:

$$ADF = |HC_M - HC_A| \quad (6)$$

To evaluate the proposed network for the classification of fetal organs, we used Precision, Recall, F1 score metrics as well as confusion matrix [40].

Results

Qualitative results

Fig. 3 shows the automatic and manual (serving as standard of reference) segmentation results of fetal ultrasound images from both pubic and prepared datasets. At each row, the results of manual and automatic parameter prediction (i.e. BPD, HC, AC, and FL) are depicted.

Quantitative results

The results reflecting the evaluation of fetal organ classification using the convolutional neural network in terms of Precision, Recall, and F1 score are presented in table 1. The confusion matrix for the test set (20% of the whole dataset, i.e. 180 images) is depicted in Fig. 4.

DSC values obtained from Attention MFP-Unet model for 198 subjects in the evaluation of public dataset are illustrated in Fig. 5. A DSC of less than 0.86 was not observed in this dataset, which indicates that the performance of the proposed network does not depend on image variability. In other words, the proposed network presents good generalizability properties for the segmentation of subjects with large differences.

The average values and the standard deviation of the quantitative metrics are summarized in Tables 2 and 3 for the proposed algorithm and other approaches. Table 2 compares the performance of Attention MFP-Unet with U-net, dilated U-net, Attention U-net, RU-net, R2U-net, and MFP-Unet using the local dataset. Table 3 compares the results of the proposed approach withstate-of-the-art methods using the public dataset [19].

To investigate the impact of the preprocessing step, the training and evaluation of Attention MFP-Unet on femur images were repeated without applying the preprocessing algorithm. Table 4 shows the results of this comparison in terms of DSC, HD, conformity coefficient, APD, and good contours.

For clinical validation, fetal biometric parameters, including BPD, HC, AC and FL were computed and the Bland-Altman plots for these parameters are shown in Fig. 6. For BPD, HC, AC and FL parameters, correlation coefficients of 0.96, 1, 0.97, and 0.94 with the reference contours, respectively, were observed. The solid line in the middle of the Bland-Altman plots represents the mean value of the difference between manual and automatic segmentations, whereas the upper and lower dashed lines represent ± 1.96 standard deviation (SD) of the difference. The mean and confidence intervals of the difference between the automatic and proposed method BPD results were −0.79 mm and −11.2–9.62 mm, respectively. These values were −0.55 mm and −9.81–8.7 mm for HC, −17.24 mm and −36.26–1.79 mm for AC, and 0.87 mm and −4.18–5.92 mm for FL.

The three parameters obtained from the correlation analysis (R correlation coefficient and the two parameters representing linear regression analysis) are presented in Table 5. A strong correlation between the automatic and manual contours can be observed. The weakest performance was observed for FL parameters in the test set (R = 0.94).

Fig. 7 shows the box plots of absolute error for fetal biometric parameters obtained from the approaches listed in Table 2. The whiskers were plotted as a "circle" marker. Attention MFP-Unet led to the lowest HC and AC errors among the entire approaches.

Discussion

We proposed a novel network architecture for the segmentation of fetal ultrasound images and the estimation of fetal biometric parameters. The presented model combines MFP_Unet architecture [17] and AGs (utilized in the path of skip connections) [31]. This upgrade improves the accuracy of fetal organ segmentation due to the concentration of the network on the only "one" object in the image. Moreover, a preprocessing algorithm was proposed to identify the region containing the femur in fetal images to be treated as a second input channel in the proposed network. However, for images of fetal head and abdomen, Niblack's thresholding technique [32] was utilized as the second input channel.

The proposed network benefits from the capability of MFP-Unet in accurate segmentation of ultrasound images as well as the incorporation of the AGs for segmentation of fetal images. MFP-Unet takes advantage of U-net (robustness in image segmentation) and FPN (flexibility of the network that uses the features in all scales for prediction). Dilated convolution kernels [41] were employed in Attention MFP-Unet as well as MFP-Unet, which creates a more global receptive field by utilizing a wider field of view without increasing the size of the parameters.

To the best of our knowledge, most previous works in this field were devoted to fetal head segmentation and HC prediction. This is motivated by the existence of a publically-available dataset [19]. Hence, proposing a deep learning-based network outperforming existing methods for fetal ultrasound image segmentation and enabling to predict fetal biometric parameters in a robust way is highly demanded. The proposed preprocessing of the input image plays a significant role in this generality, which allows the use of this algorithm in clinical routine with minimum user interaction (i.e. the radiologist determines whether the input image is femur or not).

The performance of Attention MFP-Unet improved after the training data were augmented. The data augmentation strategy was adopted

since providing a large set of annotated fetal ultrasound images requires excessively long time. In addition, the proposed method for fetal ultrasound image segmentation exhibited promising performance with a relatively small training dataset. The runtime of Attention MFP-Unet is remarkably lower than classic U-net owing to the incorporated dilated kernels in the architecture.

Fig. 3 shows representative auto-segmentations of the head, abdomen and femur for three representative cases. According to the guidelines, after 32–34 menstrual weeks, the distal femoral epiphysis should be visualized, however, it should not be included in the measurement [42]. Our proposed network met well this important clinical criterion, while it has been overlooked by most previous works [28,43].

To properly evaluate the performance of the proposed method, two different datasets and several similar approaches were included in this study. The evaluation of the network using fetal head dataset [19] was compared with state-of-the-art approaches that used the same dataset (Table 3). Conversely, when the network was trained with our prepared dataset, it was compared with U-net and all upgraded networks based on U-net, including MFP-Unet, dilated U-net, Attention U-net, RU-net and R2U-net. Regarding Table 2, our approach led to promising results (except the good contours) in comparison with other approaches, while the poorest results were observed when using classic U-net. The superior performance of the proposed method is because of the combination of various blocks in the promoted U-net architectures. Table 3 also demonstrates the superiority of our approach in comparison with other approaches evaluated on the same database.

Fig. 6(b) shows the consistency between fetal biometric parameters resulting from automated and manual delineations. Most of the data points in the Bland-Altman plots are within the standard line and the mean line is close to zero. This indicates the consistency between the results of automatic and manual segmentation. In some cases of abdomen images, there is a slight difference between manual and automatic results, which stems from the fuzziness of fetal abdomen contour (vague margins). The low target to background contrast challenged our ellipse fitting algorithm to estimate the correct boundary of the fetal abdomen. According to Table 5, the HC obtained from manual and automated methods has the highest correlation among all parameters. This is due to the abundance of real fetal head images in the public dataset [19]. The main difference between our work and the van den Heuvel *et al.*'s work [19] in the measurement of HC is in dividing the training dataset based on the fetus age (the specified trimester) and consequently image morphology. Designing the deep model based on the specified trimester led to improved quantitative results. Conversely, van den Heuvel et al. adopted a semi-automatic approach, which requires user interaction in the test phase.

We used various evaluation metrics to assess the performance of Attention MFP-Unet, including DSC, HD, CC, APD, and percentage of good contours for the prepared dataset, and DSC, HD, DF and ADF for the public fetal head dataset, because we set out to compare the proposed method with other approaches using the same public dataset. Besides, we found many other quantitative parameters illustrative and meaningful for inclusion in this study.

Fig. 7 shows a boxplot representation of errors related to the calculated biometry parameters using the proposed segmentation method and other related approaches, where the whiskers of the U-net boxplot have the largest error range. The performance can be improved by increasing the number of training images.

Our proposed algorithm fits the desired ellipse that is needed in fetal head and abdomen images after the segmentation phase. While, Sobhaninia et. al's network learns the ellipse tuner parameters in the training phase of the network [21]. The independent approach of ellipse fitting, considering accurate segmentatio, results a more precise fitted ellipse.

The main limitations of the proposed model are deficiency of the training images and lack of publically available fetal ultrasound images. The first limitation indicates that it is important to have a fairly rich number of training images for the training of the Attention MFP-Unet, albeit a small training set with appropriate augmentation methods performed relatively well. The term "fairly rich" also means that the network's performance would be higher when it uses 1000 labeled images collected from different patients rather than 10 images from the 100 patients. The second limitation refers to the lack of annotated fetal ultrasound images in the abdomen and femur.

Conclusion

We presented a convolutional neural network architecture that automatically measures the BPD, HC, AC, and FL biometry parameters from fetal ultrasound images. Our network, as an extension of our previous work, incorporates AGs into the skip connections of the MFP-Unet. Utilizing a relatively large dataset along with an appropriate data augmentation algorithm resulted in satisfactory results with clinically tolerable errors. The results obtained using the proposed network exhibited a promising performance in the measurement of fetal biometric parameters. A comparison of the proposed method with other variants of U-net model demonstrated its superior performance. Future work will focus on the development of an efficient approach to automate the procedure of determining the type of fetal organ present in ultrasound images. Our algorithm could be extended to consider 3D images of fetal organs in the framework of a 3D attention MFP-Unet.

Acknowledgements

This work was supported by the Med Fanavaran Plus Co and the Swiss National Science Foundation under grant SNRF 320030_176052.

Appendix A. Supplementary data

Supplementary data to this article can be found online at https://doi.org/10.1016/j.ejmp.2021.06.020.

References

[1] Liu S, Wang Yi, Yang X, Lei B, Liu Li, Li SX, et al. Deep learning in medical ultrasound analysis: a review. Engineering 2019;5(2):261–75. https://doi.org/10.1016/j.eng.2018.11.020.
[2] Adineh-Vand A, Torabi M, Roshani GH, Taghipour M, Feghhi SAH, Rezaei M, et al. Application of adaptive neuro-fuzzy inference system for prediction of neutron yield of IR-IECF facility in high voltages. J Fusion Energy 2014;33(1):13–9.
[3] Hearn-Stebbins B. Normal fetal growth assessment: a review of literature and current practice. J Diagn Med Sonogr 1995;11(4):176–87. https://doi.org/10.1177/875647939501100403.
[4] Loughna P, Chitty L, Evans T, Chudleigh T. Fetal size and dating: charts recommended for clinical obstetric practice. Ultrasound 2009;17(3):160–6. https://doi.org/10.1179/174313409X448543.
[5] Bandeira Diniz P, Yin Y, Collins S. Deep learning strategies for ultrasound in pregnancy. Eur Med J Reprod Heal 2020;6:73–80.
[6] Rueda S, Fathima S, Knight CL, Yaqub M, Papageorghiou AT, Rahmatullah B, et al. Evaluation and comparison of current fetal ultrasound image segmentation methods for biometric measurements: a grand challenge. IEEE Trans Med Imaging 2014;33(4):797–813.
[7] GE Versana Club - SonoBiometry n.d. https://www.versanaclub.net/emea/sono-biometry (accessed August 31, 2020).
[8] Mohammadi R, Shokatian I, Salehi M, Arabi H, Shiri I, Zaidi H. Deep learning-based auto-segmentation of organs at risk in high-dose rate brachytherapy of cervical cancer. Radiother Oncol 2021;159:231–40.
[9] Arabi H, AkhavanAllaf A, Sanaat A, Shiri I, Zaidi H. The promise of artificial intelligence and deep learning in PET and SPECT imaging. Phys Med 2021;83:122–37.
[10] Shiri I, Arabi H, Salimi Y, Sanaat AH, Akhavanalaf A, Hajianfar G, et al. COLI-NET: Fully Automated COVID-19 Lung and Infection Pneumonia Lesion Detection and Segmentation from Chest CT Images. MedRxiv 2021.
[11] Shiri I, Arabi H, Sanaat A, Janebi E, Becker M, Zaidi H. Fully automated gross tumour volume delineation from PET in head and neck cancer using deep learning algorithms. Clin Nucl Med 2021.
[12] Ronneberger O, Fischer P, Brox T. U-Net: Convolutional Networks for Biomedical Image Segmentation. Int Conf Med image Comput Comput Interv, Springer 2015:234–41. https://doi.org/10.1007/978-3-319-24574-4_28.
[13] Alom MZ, Hasan M, Yakopcic C, Taha TM, Asari VK. Recurrent residual convolutional neural network based on u-net (r2u-net) for medical image segmentation. ArXiv Prepr ArXiv180206955 2018.

[14] Liang M, Hu X. Recurrent convolutional neural network for object recognition. Proc IEEE Conf Comput Vis pattern Recognit 2015:3367–75.
[15] Oktay O, Schlemper J, Le FL, Lee M, Heinrich M, Misawa K, et al. Attention u-net: Learning where to look for the pancreas. ArXiv Prepr ArXiv180403999 2018.
[16] Lee J. Image_Segmentation 2018. https://github.com/LeeJunHyun/Image_Segmentation.
[17] Moradi S, Oghli MG, Alizadehasl A, Shiri I, Oveisi N, Oveisi M, et al. MFP-Unet: A novel deep learning based approach for left ventricle segmentation in echocardiography. Phys Med 2019;67:58–69. https://doi.org/10.1016/j.ejmp.2019.10.001.
[18] Lin T-Y, Dollár P, Girshick R, He K, Hariharan B, Belongie S. Feature Pyramid Networks for Object Detection. CVPR 2017:2117–25.
[19] van den Heuvel TLA, de Bruijn D, de Korte CL, van Ginneken B. Automated measurement of fetal head circumference. PLoS One 2018;13:e0200412. https://doi.org/10.5281/ZENODO.1322001.
[20] Kerbyson D, Atherton T. Circle detection using Hough transform filters. Image Process its Appl 1995:370–4.
[21] Sobhaninia Z, Rafiei S, Emami A, Karimi N, Najarian K, Samavi S, et al. Fetal ultrasound image segmentation for measuring biometric parameters using multi-task deep learning. In: Annu Int Conf IEEE Eng Med Biol Soc. Institute of Electrical and Electronics Engineers (IEEE); 2019. p. 6545–8. https://doi.org/10.1109/embc.2019.8856981.
[22] Chaurasia A, Culurciello E. LinkNet: Exploiting encoder representations for efficient semantic segmentation. In 2017 IEEE Vis Commun Image Process VCIP 2017, vol. 2018- Janua, Institute of Electrical and Electronics Engineers Inc.; 2018, p. 1–4. doi:10.1109/VCIP.2017.8305148.
[23] Sinclair M, Baumgartner CF, Matthew J, Bai W, Martinez JC, Li Y, et al. Human-level Performance on Automatic Head Biometrics in Fetal Ultrasound Using Fully Convolutional Neural Networks. In Proc Annu Int Conf IEEE Eng Med Biol Soc EMBS, vol. 2018- July, Institute of Electrical and Electronics Engineers Inc.; 2018, p. 714–7. doi:10.1109/EMBC.2018.8512278.
[24] Long J, Shelhamer E, Darrell T. Fully convolutional networks for semantic segmentation. IEEE Conf Comput Vis pattern Recognit 2015:3431–40.
[25] Irene K, Yudha P. A, Haidi H, Faza N, Chandra W. Fetal Head and Abdomen Measurement Using Convolutional Neural Network, Hough Transform, and Difference of Gaussian Revolved along Elliptical Path (Dogell) Algorithm. ArXiv Prepr ArXiv191106298 2019.
[26] Redmon J, Divvala S, Girshick R, Farhadi A. You only look once: unified, real-time object detection. IEEE Conf Comput Vis Pattern Recognit 2016:779–88.
[27] Foi A, Maggioni M, Pepe A, Rueda S, Noble JA, Papageorghiou AT, et al. Difference of Gaussians revolved along elliptical paths for ultrasound fetal head segmentation. Comput Med Imaging Graph 2014;38(8):774–84. https://doi.org/10.1016/j.compmedimag.2014.09.006.
[28] Carneiro G, Georgescu B, Good S, Comaniciu D. Detection and measurement of fetal anatomies from ultrasound images using a constrained probabilistic boosting tree. IEEE Trans Med Imaging 2008;27(9):1342–55. https://doi.org/10.1109/TMI.2008.928917.
[29] Tu Z. Probabilistic boosting-tree: Learning discriminative models for classification, recognition, and clustering. Proc IEEE Int Conf Comput Vis 2005;II:1589–96. https://doi.org/10.1109/ICCV.2005.194.
[30] Rahmatullah R, Ma'Sum MA, Aprinaldi, Mursanto P, Wiweko B. Automatic fetal organs segmentation using multilayer super pixel and image moment feature. In Proc - ICACSIS 2014 2014 Int Conf Adv Comput Sci Inf Syst, Institute of Electrical and Electronics Engineers Inc.; 2014, p. 420–6. doi:10.1109/ICACSIS.2014.7065883.
[31] Schlemper Jo, Oktay O, Schaap M, Heinrich M, Kainz B, Glocker B, et al. Attention gated networks: learning to leverage salient regions in medical images. Med Image Anal 2019;53:197–207. https://doi.org/10.1016/j.media.2019.01.012.
[32] Niblack W. An introduction to digital image processing. Prentice-Hall Englewood Cliffs 1986.
[33] Halir R, Flusser J. Numerically stable direct least squares fitting of ellipses. In: 6th Int Conf Cent Eur Comput Graph Vis WSCG; 1998. p. 125–32.
[34] Fitzgibbon A, Pilu M, Fisher RB. Direct least square fitting of ellipses. IEEE Trans Pattern Anal Mach Intell 1999;21(5):476–80.
[35] Dice LR. Measures of the amount of ecologic association between species. Ecology 1945;26:297–302. https://doi.org/10.2307/1932409.
[36] Babalola KO, Patenaude B, Aljabar P, Schnabel J, Kennedy D, Crum W, et al. Comparison and Evaluation of Segmentation Techniques for Subcortical Structures in Brain MRI. Med Image Comput Comput Interv – MICCAI 2008, Springer, Berlin, Heidelberg; 2008, p. 409–16. doi:10.1007/978-3-540-85988-8_49.
[37] Chang H-H, Zhuang AH, Valentino DJ, Chu W-C. Performance measure characterization for evaluating neuroimage segmentation algorithms. Neuroimage 2009;47(1):122–35. https://doi.org/10.1016/j.neuroimage.2009.03.068.
[38] Radau P, Lu Y, Connelly K, Paul G, Dick A, Wright G. Evaluation framework for algorithms segmenting short axis cardiac MRI. MIDAS J-Cardiac MR Left Vent Segmentation Chall 2009;49.
[39] Bland JM, Altman DG. Statistical methods for assessing agreement between two methods of clinical measurement. Int J Nurs Stud 2010;47(8):931–6. https://doi.org/10.1016/j.ijnurstu.2009.10.001.
[40] Stehman SV. Selecting and interpreting measures of thematic classification accuracy. Remote Sens Environ 1997;62(1):77–89.
[41] Yu F, Koltun V. Multi-scale context aggregation by dilated convolutions. ArXiv Prepr ArXiv151107122 2016.
[42] Goldstein R, Filly R, Simpson G. Pitfalls in femur length measurements. J Ultrasound Med 1987;6:203–7.
[43] Shrimali V, Anand RS, Kumar V. Improved segmentation of ultrasound images for fetal biometry using morphological operators.. In: Proc 31st Annu Int Conf IEEE Eng Med Biol Soc Eng Futur Biomed EMBC 2009. IEEE Computer Society; 2009. p. 459–62. https://doi.org/10.1109/IEMBS.2009.5334470.
[44] Ciurte A, Bresson X, Cuadra MB. A semi-supervised patch-based approach for segmentation of fetal ultrasound imaging. Chall US Biometric Meas from Fetal Ultrasound Images, ISBI 2012, 2012, p. 5–7.
[45] Stebbing R V, McManigle, John E. A boundary fragment model for head segmentation in fetal ultrasound. Chall US Biometric Meas from Fetal Ultrasound Images, ISBI, 2012, p. 9–11.
[46] Sun C. Automatic fetal head measurements from ultrasound images using circular shortest paths. Chall US Biometric Meas from Fetal Ultrasound Images, ISBI 2012, 2012, p. 13–5.
[47] Ponomarev G V, Gelfand MS, Kazanov MD. A multilevel thresholding combined with edge detection and shape-based recognition for segmentation of fetal ultrasound images. Chall US biometric Meas from fetal ultrasound images, ISBI, 2012, p. 17–9.