ORIGINAL ARTICLE

Automated measurement network for accurate segmentation and parameter modification in fetal head ultrasound images

Peixuan Li[1,2,3,4,5] · Huaici Zhao[1,2,4,5] 🔟 · Pengfei Liu[1,2,3,4,5] · Feidao Cao[1,2,3,4,5]

Abstract

Measurement of anatomical structures from ultrasound images requires the expertise of experienced clinicians. Moreover, there are artificial factors that make an automatic measurement complicated. In this paper, we aim to present a novel end-to-end deep learning network to automatically measure the fetal head circumference (HC), biparietal diameter (BPD), and occipitofrontal diameter (OFD) length from 2D ultrasound images. Fully convolutional neural networks (FCNNs) have shown significant improvement in natural image segmentation. Therefore, to overcome the potential difficulties in automated segmentation, we present a novelty FCNN and add a regression branch for predicting OFD and BPD in parallel. In the segmentation branch, a feature pyramid inside our network is built from low-level feature layers for a variety of fetal head in ultrasound images, which is different from traditional feature pyramid building methods. In order to select the most useful scale and reduce scale noise, attention mechanism is taken for the feature's filter. In the regression branch, for the accurate estimation of OFD and BPD length, a new region of interest (ROI) pooling layer is proposed to extract the elliptic feature map. We also evaluate the performance of our method on large dataset: HC18. Our experimental results show that our method can achieve better performance than the existing fetal head measurement methods.

Keywords Fetal head measurement · Ultrasound image segmentation · Fully convolutional networks · Feature pyramid · ROI pooling

1 Introduction

Ultrasonic imaging is widely used in clinical examination since it does not use ionizing radiation and more low-costing compared with computed tomography (CT) and magnetic resonance imaging (MRI), which make it to be the first choice of prenatal care. A clear and accurate anatomical structure measurement is required in many clinical ultrasound diagnoses. In particular, the fetal head measurement can be used to estimate gestational age and monitor growth patterns [1]. In general, these measurements are performed by experienced clinical sonographers on account of ultrasound images, which are operator-dependent and machine-specific [2] leading to inter- and intra-observer variability. The automatic method of fetal biometric measurement can reduce the variability and doctors' workloads with no intra-observer variability [3]. Furthermore, there is still a severe shortage of well-trained sonographers in many countries, so an automated system may assist inexperienced clinicians to obtain an accurate measurement.

Typically, three standard fetal head biometric parameters were considered by using two-dimensional ultrasound measurements: head circumference (HC), biparietal diameter (BPD), occipitofrontal diameter (OFD). The guidelines state that BPD and HC are measured in the transaxial plane at the widest portion of the skull at the level of the thalami. When measuring BPD, the vernier is placed from the outer margin of the proximal skull to the inner margin of the distal

✉ Huaici Zhao
hczhao@sia.cn

1 Shenyang Institute of Automation, Chinese Academy of Sciences, Shenyang, 110016, China

2 Institutes for Robotics and Intelligent Manufacturing, Chinese Academy of Sciences, Shenyang, 110169, China

3 University of Chinese Academy of Sciences, Beijing, 100049, China

4 Key Laboratory of Opto-Electronic Information Processing, Chinese Academy of Sciences, Shenyang, 110016, China

5 Key Lab of Image Understanding and Computer Vision, Liaoning Province, Shenyang, 110016, China

skull and perpendicular to the cerebral line. OFD overlaps as much as possible with the middle cerebral. The HC parameter is calculated by drawing an ellipse around the outline of the skull [4]. It is beyond the scope of this paper to detail the measurement, and more details can be found in references [3, 4].

Due to the attenuation of ultrasonic transmission and acquisition characteristics of ultrasonic equipment, ultrasound images may contain speckle noise, discontinuous or ambiguous anatomical boundaries, and shadows, which make ultrasound image measurement become one of the most difficult medical imaging tasks [5]. Examples of some artifacts are shown in Fig. 1, where fetal head ultrasound images from the first to the third trimesters are shown in the top row, and the corresponding manually labeled images are shown in the bottom. It can be seen that the boundary of the fetal head structures is mostly incomplete or not obvious, and there are a lot of speckle noises, which makes it ambiguous with the surrounding tissue.

The previous methods can be mainly divided into two categories: one is to fit the ellipse equation by segmentation of the region of interest [6, 7], the other is to fit ellipse equation directly from the original image [8–11]. Whereas they all obtain the BPD and OFD by measuring the length of the minor axis and the major axis, these methods, which make no distinction between geometric and biological lengths, will lead to errors in the final measurement. Moreover, in spite of the great effort did in many fields, traditional machine learning approaches based on hand-crafted feature limit to develop more complex application scenarios. Recently, deep convolutional neural networks (CNNs) have become the dominant approach in different vision challenges by automatically extracting useful features. Wu et al. first brought the CNN to the fetal head' region segmentation by cascading three variants of the FCN [7]. However, using single scale information can cause

incomplete boundary prediction ambiguity because a fixed receptive field of a neural network can only infer boundary with a fixed region. As the red bounding box shown in Fig. 1, the small receptive fields obtain a good prediction, while the fixed receptive fields in the large incomplete areas result in fuzzy boundary prediction.

In this work, we propose a novel automated measurement system based on a deep neural network. First, we propose a fetal head segmentation network, which considers the very deep CNN with the large receptive field to extract the main areas of the head and then utilize the layers corresponding to different receptive fields to speculate the discontinuous areas in different sizes of a skull. Instead of simple average fusion, we propose a scale attention-based multi-scale module to fuse the different scales of information. After segmenting the region of the fetal head from the ultrasound image, the ellipse closest to the skull can be fitted by the least square error method. As mentioned earlier, most of these current methods treat geometric lengths as biological length, which results in measurement errors. After fitting the ellipse, we propose a regression network to correct this error by predicting the residual angle between the major and minor axis of the ellipse and the true OFD and BPD. We modify RoIAlign [12] to present an ellipse pooling module that directly obtains the elliptic feature and send it into the fully connected layers for angle prediction. The ellipse pooling module is efficient for the reason that it avoids re-extracting features from the original image.

In summary, the contributions of this paper are in four aspects:

1. The fetal head ultrasound images have speckle noise, boundary occlusion, and other artifacts, so the segmentation algorithm must be able to segment complete and occlusive region in a lot of noise. We propose that the

Fig. 1 Example of fetal head ultrasound images: without annotation and with annotation in blue from top to bottom. The first trimester, the second trimester, and the third trimester from left to right. Note that there are speckle noise, incomplete anatomical boundaries and other artifacts. The boundary in the red bounding box is relatively clear, while the boundary in the green box cannot be inferred from local brightness very well
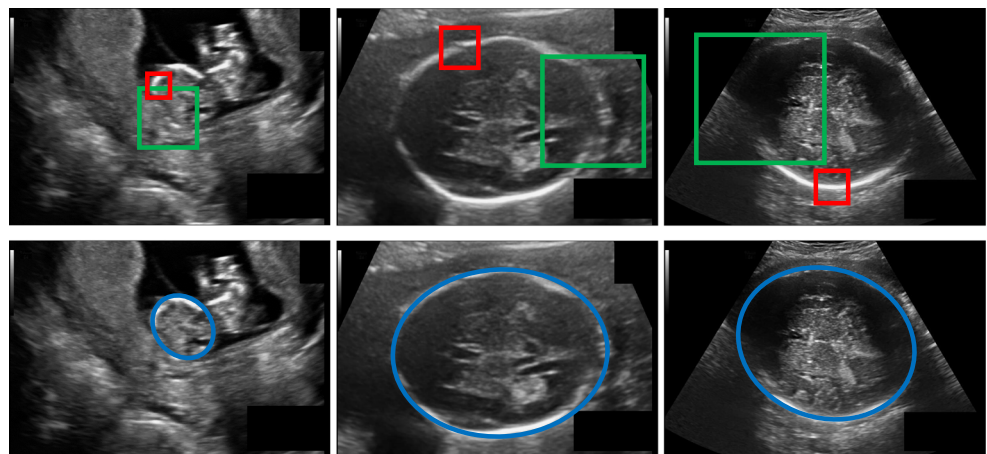
multi-scale method can be used for targeted boundary recognition and fitting.

2. We propose to build a scale attention-based feature pyramid to fuse the information of different receptive field layers.

3. We design a regression network for BPD and OFD prediction. In the proposed network, we design an ellipse pooling module to share the feature map which allows the regression network to reuse low feature layers in order to reduce time consumption.

4. We build a complete fetal head automatic measurement system that can be used to measure head circumference, OFD and BPD.

2 Related work

Fetal head measurement. Foi et al. presented a fully automatic method for fetal head measurement using signal processing and minimizing a cost function to directly fit an ellipse [8]. Ciurte et al. used a semi-supervised approach, which interpreted ultrasound segmentation to a graph cutting problem and solved it via min-cut and fast minimization algorithm [6]. Stebbing et al. first calculated the position and direction of the fetal head boundary, then used the random forest to find the inner and outer contour of the skull, and finally used it to fit the two ellipses [9]. Lu and Satwika et al. used different hough transform approaches to directly estimate the parameters of ellipse function [10, 11]. Ponomarev et al. distinguished skull from the background by multilevel threshold and recognized the segmented objects by using two introduced shape-based descriptor. [13]. Machine learning techniques are potentially included in fetal head measurement, such as [14–18]used Haar-Like features to detect the position of the skull, of which [18] first trained a random forest classifier with Haar-Like features to locate the fetal head and the ellipse parameters are fitted by dynamic programming and hough transform.

Deep learning-based segmentation network. Recent segmentation algorithms often convert an existing CNN structure to a FCNN such as AlexNet [19], VGGNet [20], and ResNet [21]. They use 1×1 convolution instead of fully connected layers to generate heatmap and perform some deconvolution layers for pixel-wise labeling [22]. Furthermore, intra-skipping connections are included to improve performance [23]. These networks have shown excellent performance in natural image understanding, maintaining the best records on many datasets. In the field of medical image, the FCNN is designed to a U shape structure as a state-of-the-art model [23]. These encoder-decoder architecture combined fine-grained and coarse-grained

features have been proved effective at a satisfactory level in CT or MRI images [24–26]. Wu et al. first introduce the CNN to fetal head segmentation in ultrasound images [7]. They show that FCNN can filter out most of the speckle noise and non-skull regions by cascading three variants of the FCN. They utilize the single scale information, which can cause incomplete boundary prediction ambiguity because a fixed receptive field of a neural network can only infer boundary with a fixed region; meanwhile, they also make no discrimination on biometric and geometric length.

3 Proposed methodology

The whole automatic measurement system can be seen in Fig. 2, which is mainly composed of three parts: (1) scale attention pyramid deep neural network (SAPNet) for head region segmentation; (2) regression network for OFD and BPD prediction; (3) the fusion module which takes advantage of the results of the two network to output the final result. Our approach achieves state-of-the-art performance in the available dataset HC18 [18] and estimates the total runtime of the system on NVIDIA 1080TI GPU to about 30 FPS. In the rest of this section, we will detail our system.

3.1 Dataset

The proposed method is based on deep learning which is very dependent on the amount and quality of data. Therefore, we will first detail the dataset. A total of 1334 two-dimensional fetal head ultrasound images were collected from the challenge of HC18, which were acquired from 551 pregnant women who received a routine ultrasound screening exam between May 2014 and May 2015. It should be noted that these fetuses are clinically healthy, which is very important to use this dataset to evaluate fetal development. These images were collected by experienced sonographers who went through the Voluson E8 and Vouson 730 ultrasound machine. The pixel size of each image is 800×540, and the distance between pixels corresponds to the real distance in the range of 0.052 to 0.326 mm. We can obtain 999 ground truth of 1334 images, and the remaining 335 results are submitted to the HC18 challenge website to evaluate algorithm rankings[1]. The ground truth is annotated manually by the sonographers, this was done by drawing an ellipse best fitting the circumference of the head skull, as shown in Fig. 3.
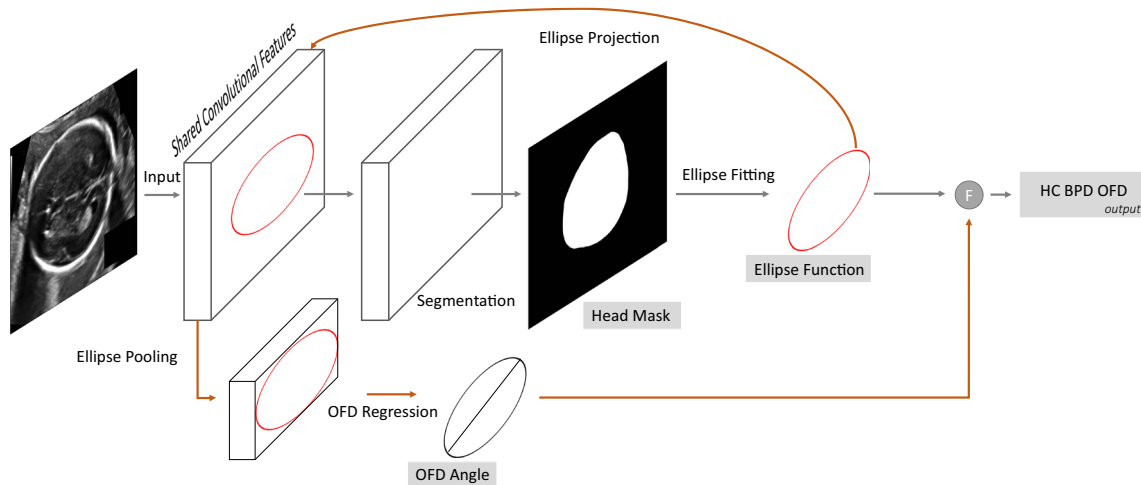
---

[1] https://hc18.grand-challenge.org/

Fig. 2 Complete automatic system for fetal head measurement. An ultrasound image is an input from the left side and a mask bitmap of the head region is provided by the segmentation network. This mask is subsequently fitted to an ellipse by the least square error method. The angle of occipitofrontal diameter is obtained by the input regression network after ellipse pooling. Finally, the ellipse equation and occipitofrontal diameter angle are fused to obtain the final output

We express the ellipse parametric equation as:

$$
\begin{aligned}
&Axy + By^2 + Cy + Dx + x^2 + E = 0, \\
&A = \frac{b^2 - a^2}{a^2 \sin^2\theta + b^2 \cos^2\theta}, \\
&B = \frac{a^2 \cos^2\theta + b^2 \sin^2\theta}{a^2 \sin^2\theta + b^2 \cos^2\theta}, \\
&C = -2c_y B - c_x A, \\
&D = -2c_x - c_y A, \\
&E = c_x^2 + c_y^2 B + c_x c_y A - \frac{a^2 b^2}{a^2 \sin^2\theta + b^2 \cos^2\theta}.
\end{aligned} \tag{1}
$$

Datasets are collected from fetuses of different gestation ages, as shown in Fig. 1. For these images of different ages varies greatly, our algorithm also tests the images in different gestation ages to better evaluate the quality. Table 1 shows the distribution of entire datasets from the first to the third trimester.

3.2 Data augmentation

The data augmentation scheme is a common approach to increase the amount of training data and tech the network desired invariance and robustness properties [23]. In the case of the fetal ultrasound image, we primarily need gray value, chrominance, contrast and sharpness invariance as well as the robustness of gaussian blurring in different window sizes. Especially, the random horizontal flipping of the training image seems to be the key augmentation to learn a segmentation network with very little ground truth. We randomly perturb an image within a specified range during each training session, detailed in Table 2.

3.3 SAPNet architecture

The SAPNet, illustrated in Fig. 4, is a segmentation model that can combines different scale information in feature level. Our network is based on the U-Net [23] structure which is a standard medical image segmentation network.
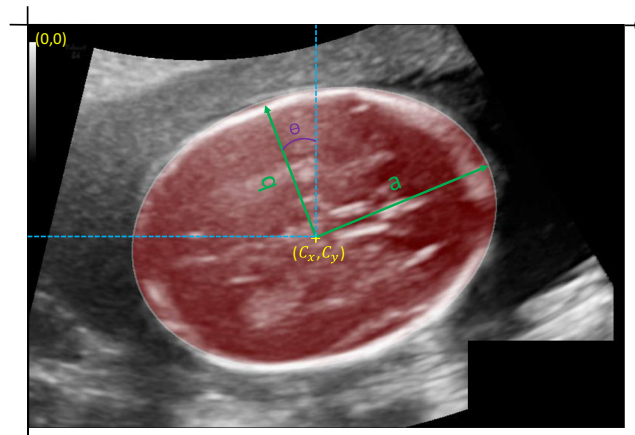

Fig. 3 The parameterized ground truth: **a** and **b** are the major and minor axes of the ellipse respectively. c_x, c_y are the center of the ellipse. θ is the rotation angle in degrees

Table 1 The distribution of dataset

Gestation period	Training set	Test set
First trimester	165	55
Second trimester	693	233
Third trimester	141	47
Total number	999	335

Table 2 Data enhancement methods and their range

Method	Brightness	Contrast	Sharpness	Gaussian blur
Range	0.4–1.8	0.3–1.8	0.3–2	2–9

In order to get a scale pyramid module with different receptivity, we modify the convolution layer of the encoding layer of U-Net [23] structure to dilated convolution. The dilated convolution introduces a new parameter called the expansion rate of the convolutional layer which defines the spacing of the values when the convolution kernel processes the data. This convolution method can discard the pooling layer to output the full-resolution feature map while still obtaining a large receptive field. Meanwhile, we use the output of the different encoder layers to form the feature pyramid layer. The high-level feature layer of the feature pyramid has a larger receptive field, while the lower-level feature layer has a smaller receptive field. Feature layers with small receptive fields perform well on the more continuous parts of the skull, while feature layers with larger receptive fields perform well on the skull that appeared to be discontinuous on the ultrasound images.

To generate pixel-wise segmentation, one can make use of attention mechanism to generate a mask on the feature map [27], which enables a scale-level weight matrix by convolution to indicate which scale should be noticed. Inspired by [28], we propose a scale attention module (SAM) that provides attention of a global context prior to select scale-wise features. Scale attention module fuses three different scale context information by providing a scale-level attention value. To build the scale attention module, we first use a dilated U-Net model to extract the feature pyramid. As shown in Fig. 4, similar to the U-Net, the encoding network of SAPNet also consists of five large blocks. To better extract context from different layers, we use the feature map of the last three large blocks after convoluting different dilated rates to build the feature pyramid. The size of the feature maps in the scale pyramid module is 1/4 of the input size.

As shown in Fig. 5, the bottleneck layer of dilated U-Net generates an attention feature layer after global average pooling convolution. This global pooling method provides a global context as guidance of the feature pyramid to select scale attention. We get attention feature from global average pooling after 1×1 convolution with batch normalization and sigmoid activation function. The final attention value is obtained by averaging the weight of feature maps according to the number of each layer in the feature pyramid. The attention value multiplies the attention
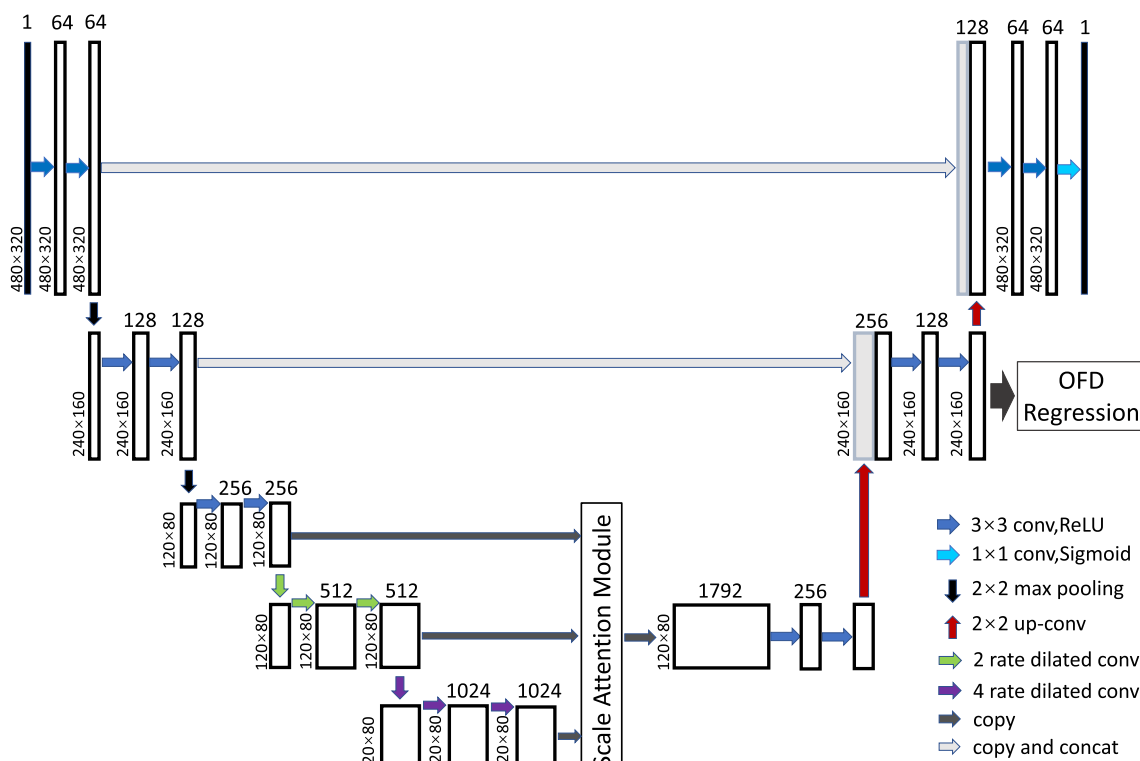

Fig. 4 Overview of the scale attention feature pyramid network. We use U-Net based on dilated convolution to extract features. Then, the scale attention module is applied to harvest different scale representations, followed by upsampling, concatenation, and 3×3 convolution

to form final module representation, which contains information about the different receptive fields. Finally, the pixel-wise prediction is got by connecting fine-grained layers

Fig. 5 Scale attention module: The multi-scale pyramid is achieved by connecting the feature maps of different receptive fields. The weight of each feature map can be obtained through the global average pooling and 1×1 convolution. The weight of a pyramid layer can be calculated by averaging the weight of corresponding feature maps
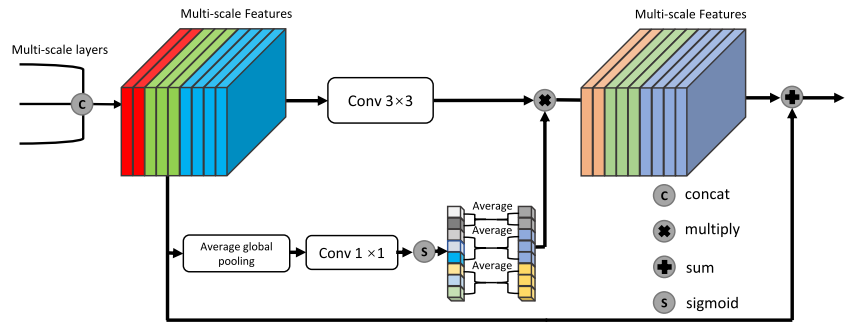
feature map and adds the original input to get the final output.

3.4 Ellipse fitting

We use the least square error method to fit the elliptical boundary of the segmentation according to Eq. 1. For the N points on the edge of the segmentation result, we can get the minimum target as:

$$Q = \sum_{i=1}^{N} (x_i^2 + B y_i^2 + A x_i y_i + D x_i + C y_i + E)^2 \qquad (2)$$

where (x_i, y_i) indicates the coordinates of the detected edge points. We can get the minimum value by:

$$\frac{\partial Q}{\partial A} = \frac{\partial Q}{\partial B} = \frac{\partial Q}{\partial C} = \frac{\partial Q}{\partial D} = \frac{\partial Q}{\partial E} = 0 \qquad (3)$$

This minimization problem can be converted into matrix equations following:

$$\begin{bmatrix} \widetilde{x^2 y^2} & \widetilde{xy^3} & \widetilde{x^2 y} & \widetilde{xy^2} & \widetilde{xy} \\ \widetilde{xy^3} & \widetilde{y^4} & \widetilde{xy^2} & \widetilde{y^3} & \widetilde{y^2} \\ \widetilde{xy^2} & \widetilde{y^3} & \widetilde{xy} & \widetilde{y^2} & \widetilde{y} \\ \widetilde{x^2 y} & \widetilde{xy^2} & \widetilde{x^2} & \widetilde{xy} & \widetilde{x} \\ \widetilde{xy} & \widetilde{y^2} & \widetilde{x} & \widetilde{y} & 1 \end{bmatrix} \begin{bmatrix} A \\ B \\ C \\ D \\ E \end{bmatrix} = \begin{bmatrix} -\widetilde{x^3 y} \\ -\widetilde{x^2 y^2} \\ -\widetilde{x^2 y} \\ -\widetilde{x^3} \\ -\widetilde{x^2} \end{bmatrix},$$

$$\widetilde{x} = \frac{1}{N} \sum_{i=1}^{N} x_i, \widetilde{y} = \frac{1}{N} \sum_{i=1}^{N} y_i, \widetilde{xy} = \frac{1}{N} \sum_{i=1}^{N} x_i y_i,$$

$$\widetilde{y^2} = \frac{1}{N} \sum_{i=1}^{N} y_i^2, \widetilde{x^3} = \frac{1}{N} \sum_{i=1}^{N} x_i^3, \widetilde{x^2 y} = \frac{1}{N} \sum_{i=1}^{N} x_i^2 y_i,$$

$$\widetilde{xy^2} = \frac{1}{N} \sum_{i=1}^{N} x_i y_i^2, \widetilde{y^3} = \frac{1}{N} \sum_{i=1}^{N} y_i^3, \widetilde{x^3 y} = \frac{1}{N} \sum_{i=1}^{N} x_i^3 y_i,$$

$$\widetilde{x^2 y^2} = \frac{1}{N} \sum_{i=1}^{N} x_i^2 y_i^2, \widetilde{xy^3} = \frac{1}{N} \sum_{i=1}^{N} x_i y_i^3, \widetilde{y^4} = \frac{1}{N} \sum_{i=1}^{N} y_i^4, \quad (4)$$

3.5 BPD and OFD prediction

The first step in the prediction of the BPD and OFD, as shown in the guideline, is to find the middle cerebral. In fact, the major and minor axis of the ellipse fitted by

the segmentation results is very close to OFD and BPD. Therefore, we obtain OFD in predicting the increment of the major axis by adding the branch of regression network in our SAPNet, as shown in Fig. 4. Our experiment shows that regression increment work is better than a direct prediction absolute angle. Regression networks only predict the angle of OFD because BPD is orthogonal to OFD. After knowing the angle of OFD, it is not difficult to obtain the angle of BPD. In order to eliminate the influence of areas outside the fetal head on the prediction of middle cerebral, we design an ellipse pooling layer to accurately locate the features inside the skull, as shown in Fig. 6. After founding the bounding box of the ellipse, it is projected onto the convolution layer of feature extraction by RoIAlign [12]. RoIAlign is an operation widely used in object detection tasks, which convert proposals of different shapes to fix shape as required by fully connected layers. The product of the feature map and the head mask can eliminate the impact of the area except for the head in the bounding box. OFD angle regression network consists of three fully connected layers whose final output passes through the activation function of $\sigma \times \tanh$. σ are the maximum values of clockwise or counterclockwise increment of the long axis and we set it as 5 for the experiment. As shown in Fig. 7, after rotating the OFD angle of the original image coordinate system, a new coordinate system with the X-axis parallel to the middle cerebral is obtained. At this point, the binary image obtained by the segmentation network is projected to the new coordinate system, in which the highest point corresponding to the X-axis is BPD, and the highest point corresponding to the Y-axis is OFD.

3.6 Network training

The network performance is optimized using the Adam [29]. We set base learning to 0.0001 and reduce by a factor of 0.8 at training error saltation. The momentum and weight decay are set to 0.9 and 0.0001. Due to the limitation of our computer hardware, we have adjusted the original 800×540 pixel image to 480×320 and set the batch size to 10 during training. In the validation set, 100 epochs are used to train all networks and after the final comparison in the
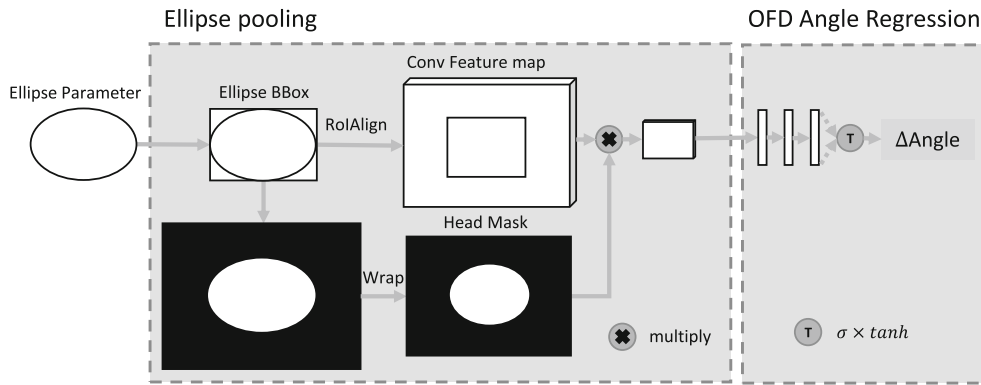
Fig. 6 An overview of the occipitofrontal diameter angle prediction module. The ellipse pooling layer uses max pooling to convert the feature inside any bounding box of the ellipse into a small feature map with a fixed spatial extent of $H \times W$ (e.g, 7×7), where H and W are layer hyper-parameters that are independent of any particular ROI. In order to obta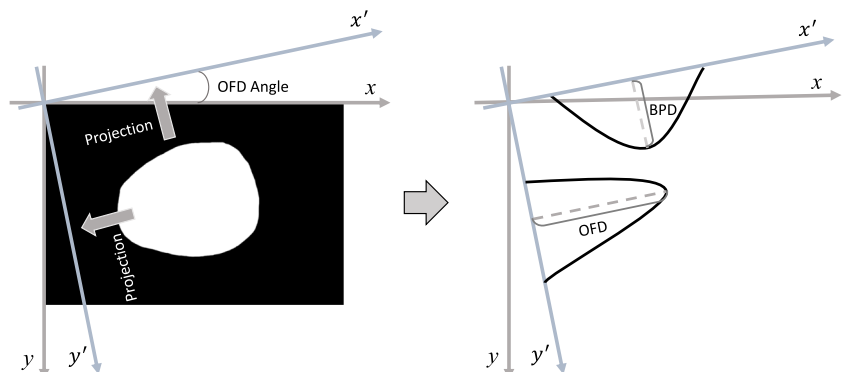in the elliptic region more accurately, we first multiplied the head mask before sampling the feature map. The occipitofrontal diameter angle regression network sends the output of the ellipse pooling layer into a regression network composed of three fully connected layers. The final output of the full connection layer finally obtains the increment value of the long axis of the ellipse through $\sigma \times tanh$ activation function

test set, we have trained 700 epochs. It is worth noting that our two network performances are improved by increasing the epoch number. We train our SAPNet by minimizing a cross-entropy loss:

$$
\begin{aligned}
L(\Gamma_{\boldsymbol{\theta}}(X), Y_{\text{truth}}) = \ & \frac{1}{m} \sum_{i=1}^{m} \Big(-Y_{\text{truth}}^{i} \log(P(\Gamma_{\boldsymbol{\theta}}^{i}(X) = 1 | \boldsymbol{\theta})) \\
& - (1 - Y_{\text{truth}}^{i}) \log(P(\Gamma_{\boldsymbol{\theta}}^{i}(X) = 0 | \boldsymbol{\theta})) \Big) + \lambda R(\boldsymbol{\theta}) \quad (5)
\end{aligned}
$$

where the $\Gamma(X)$ indicates the output network and the inputs are X and Y_{truth} represents the ground-truth image. The $\boldsymbol{\theta}$ denotes the network model parameters that need to be obtained through training. $R(\boldsymbol{\theta})$ is the regularization term where we use L_2 norm of the network weights. As in Mask R-CNN, the segmentation region is considered positive if it has IoU with the ground truth of at least 0.5 and negative otherwise. When one is the positive sample, the segmentation result is fitted to the ellipse and the regression network is trained. The L_1 loss function is adopted in the regression network.

4 Experimental results

We perform four quantitative experiments to evaluate the performance of our approach on the HC18 dataset. Firstly, we compare our system to U-Net baseline [23] in segmentation evaluation metrics and fit segmentation images boundary to ellipses in the least square method for HC, BPD, and OFD comparison. Secondly, we compare the effects of different components on automated measurement. Finally, we compare results with the best performers on the HC18 leader board. In the first three experiments, we divided annotated images into 80% training sets and 20% test sets as shown in Table 3 according to the number of images in the three pregnancy stages as shown in Table 1.

4.1 Evaluation metrics

The performance of the segmentation experiment is evaluated with three metrics in the mean pixel accuracy (mPA), the mean Intersection over Union (mIoU) and the Dice similarity coefficient (DSC).

Fig. 7 The binary mask image is projected to the coordinate system aligned with the middle cerebral, and the maximum projection values corresponding to each axis are biparietal diameter and the occipitofrontal diameter

Gestation period	Training set	Test set
First trimester	132	33
Second trimester	555	138
Third trimester	113	28
Total number	800	199

The mAP is used to evaluate the accuracy of an image that is correctly labeled with pixels.

$$mPA = \frac{N_{TP} + N_{TN}}{N_{TP} + N_{TN} + N_{FP} + N_{FN}}, \tag{6}$$

where N_{TP} is true positive which represents the number of pixels correctly classified by the fetal head, N_{TN} is true negative which represents the number of pixels correctly classified as background, and N_{FP} and N_{FN} are the number of the fetal head and background incorrectly annotated.

The mIoU is a common metric that calculates the ratio of intersection and union between two segmentation sets.

$$mIoU = \frac{IoU_{fh} + IoU_{bg}}{2},$$
$$IoU_{fh} = \frac{N_{TP}}{N_{TP} + N_{FN} + N_{FP}},$$
$$IoU_{bg} = \frac{N_{TN}}{N_{TN} + N_{FN} + N_{FP}} \tag{7}$$

where IoU_{fh} and IoU_{bg} represent the mean Intersection over Union of the fetal head and background annotated collection, respectively.

It is similar to the IoU metric the DSC gives an indication of overlapping area between our segmentation method and the ground truth.

$$DSC = \frac{|Area_M \cap Area_{GT}|}{|Area_M| + |Area_{GT}|}, \tag{8}$$

where $Area_M$ denotes the segmentation area in using our method and $Area_{GT}$ is the area of annotation of the ground truth.

The final result is the ellipse fitted by least squares after segmentation, and then we evaluate the metric of the Hausdorff distance (HD), the difference head circumference (DF), and the absolute difference head circumference (ADF).

The Hausdorff distance is a measure of the degree of similarity between two sets of points: Let P_{GT} and P_{OM} be the boundary points of the ground truth and proposed methods. p_{GT} denotes a point of P_{GT} and p_{OM} a point of P_{OM}. The minimum measurement distance of a point p to P_{GT} is defined as:

$$d_{min}(p, P_{GT}) = \min_{p_{GT} \in P_{GT}} ||p - p_{GT}||, \tag{9}$$

where $||.||$ denotes the Euclidean distance. The HD can then be expressed as:

$$HD(P_{GT}, P_{OM}) = \max\left(\max_{p_{GT} \in P_{GT}} d_{min}(p_{GT}, P_{OM}) \right.$$
$$\left. \max_{p_{OM} \in P_{OM}} d_{min}(p_{OM}, P_{GT}) \right). \tag{10}$$

The DF was defined as:

$$DF = HC_{GT} - HC_{OM}, \tag{11}$$

where HC_{OM} is the head circumference measured by ellipse proposed method and HC_{GT} by the ground truth.

The ADF was defined as:

$$DF = |HC_{GT} - HC_{OM}|. \tag{12}$$

In order to evaluate the performance of the regression network, RMSE (root mean square error) was used to calculate the error between the predicted value and the true OFD angle or length.

4.2 U-Net baseline comparison

In order to evaluate our proposed segmentation network, we conduct a series of experiments for segmentation performance comparison between the U-Net baseline and the proposed network with our best settings. We use qualitative and quantitative results to compare algorithms. The qualitative comparisons of results allow us to know where the algorithm has been improved, and the quantitative comparisons allow us to know how much the algorithm has improved. Since most of the baby fetal head segmentation is to predict fetal development, we need to know not only the average of the segmentation metric but also the worst and best conditions of our segmentation network. So we count all the prediction data of our segmentation algorithm to provide the performance of our system.

The qualitative comparisons in the proposed networks with the U-Net baseline can be seen in Fig. 8. The qualitative results show the superior ability of the proposed network to deal with the incomplete regions of the skull while producing a smooth segmentation in the complete regions of the skull. In the first trimester, the ultrasound image has many noises, uncertain areas, and unclear skull boundaries. U-Net can get lost in these areas, and even other contours may be identified as skulls. In the second pregnancy, there is a sudden saltus in the fetal skull border, which confuses the U-Net segmentation path. In the third trimester, the skull in the ultrasound image itself has large discontinuous areas, and the black template coverage of other information on the fetus in the dataset makes the discontinuous area increase and irregular, which results in the use of only a single scale U-Net often recognizes errors.
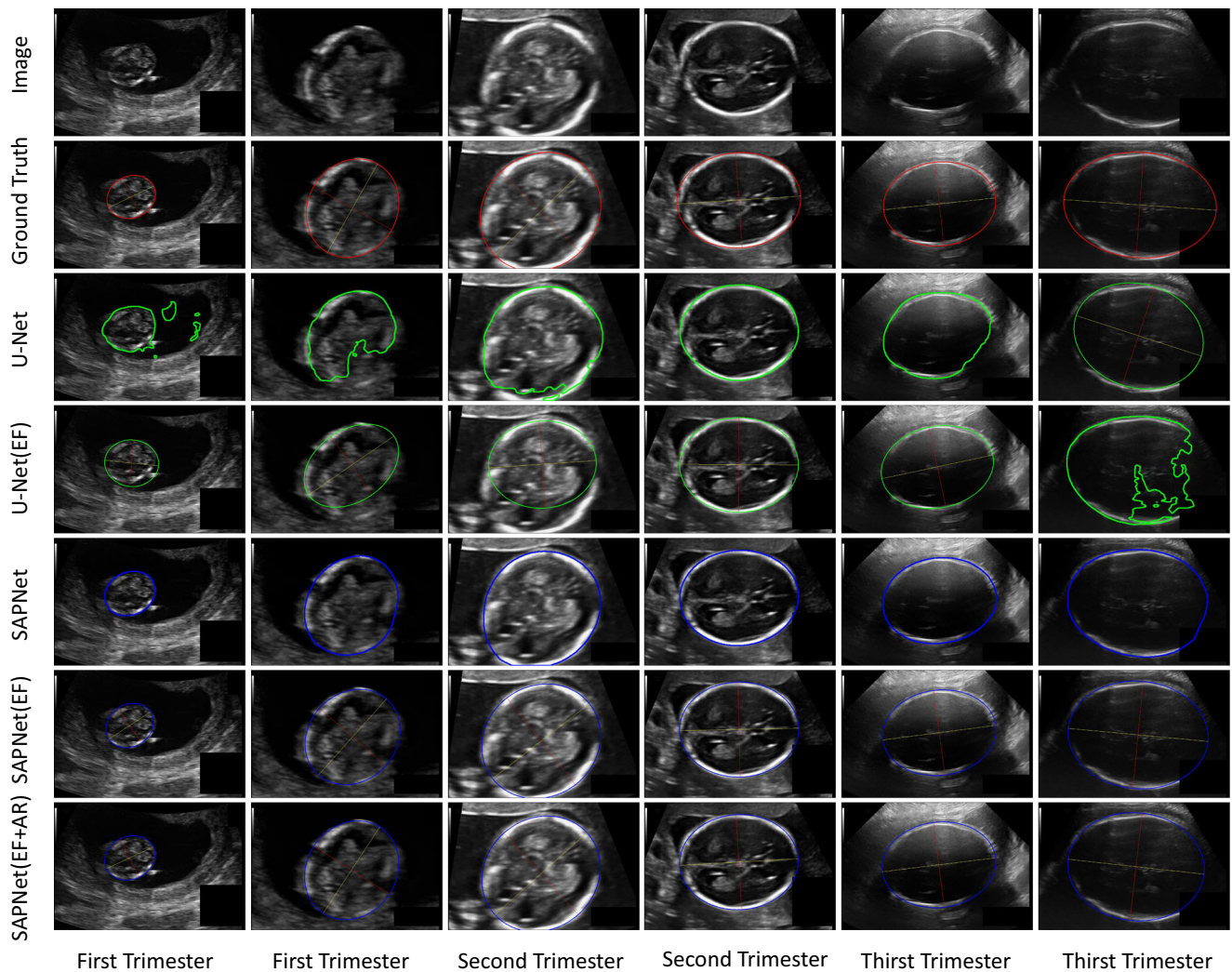
Fig. 8 Qualitative comparison of segmentation performance of networks on the fetal head ultrasound image. The original image and ground-truth region boundaries are shown in the first, second row, respectively. The outputs obtained using the U-Net baseline and proposed SAPNet are shown in the last row. The best results are shown by the SAPNet

To quantitatively demonstrate the performance of the U-Net and our proposed networks, we compare the results of two segmentation network without any post-processing. The output sizes of these three networks are adjusted to 320 × 480 and compared with the ground truth in the same size. Three assessment metrics of mIoU, mPA, and DSC are adopted for quantitative comparison. The performance between the ground truth and the results of the two networks is shown in Table 4. In our proposed SAPNet, we notice that a relatively greater improvement is performed by 3.05/1.83/2.57 and reaching 96.46%/98.02%/97.26%.

Table 4 Quantitative comparison of segmentation results for the U-Net baseline, proposed SAPNet from first trimester to third trimester

Method	U-Net			Our SAPNet		
Metric	mIoU (%) ± std	mAP (%) ± std	DSC (%) ± std	mIoU (%) ± std	mAP (%) ± std	DSC (%) ± std
First trimester	92.68 ± 7.98	94.40 ± 4.15	93.29 ± 10.03	95.06 ± 2.13	97.73 ± 1.44	95.57 ± 2.28
Second trimester	94.03 ± 2.64	97.62 ± 1.68	96.34 ± 1.99	96.82 ± 0.91	98.79 ± 0.52	97.57 ± 0.74
Third trimester	93.85 ± 6.53	95.20 ± 3.67	95.61 ± 3.81	96.06 ± 2.74	98.06 ± 1.48	97.39 ± 1.31
Average	93.81 ± 5.48	96.19 ± 2.96	95.69 ± 5.72	96.46 ± 1.77	98.02 ± 0.98	97.26 ± 1.50

Table 5 Comparison of the metric for ellipse after segmentation contours fitting

Metric	Period	U-Net	U-Net (EF)	Our SAPNet	Our SAPNet (EF)
DF (mm) ± std	First trimester	−1.85 ± 7.43	1.20 ± 7.33	0.77 ± 2.83	−0.16 ± 2.24
	Second trimester	−1.29 ± 2.26	−0.71 ± 2.61	0.29 ± 2.30	0.27 ± 2.31
	Third trimester	−1.16 ± 10.28	−1.03 ± 8.28	0.88 ± 5.98	0.30 ± 5.26
	Average	−1.30 ± 7.76	−1.28 ± 5.54	0.41 ± 3.17	−0.28 ± 2.95
ADF (mm) ± std	First trimester	5.23 ± 7.91	4.22 ± 6.06	2.22 ± 1.87	1.63 ± 1.52
	Second trimester	2.57 ± 6.91	2.00 ± 2.61	1.72 ± 1.55	1.78 ± 1.48
	Third trimester	7.65 ± 9.18	6.95 ± 8.88	4.69 ± 3.70	4.31 ± 2.91
	Average	3.28 ± 6.03	3.12 ± 4.75	2.23 ± 2.28	2.03 ± 2.13
DSC (%) ± std	First trimester	93.29 ± 10.03	93.60 ± 6.95	95.57 ± 2.28	95.51 ± 1.61
	Second trimester	96.34 ± 1.99	97.59 ± 1.28	97.57 ± 0.74	98.22 ± 0.75
	Third trimester	95.61 ± 3.81	95.62 ± 3.03	97.39 ± 1.31	97.46 ± 0.86
	Average	95.69 ± 5.72	96.23 ± 4.45	97.26 ± 1.50	97.72 ± 1.61
HD (mm) ± std	First trimester	1.95 ± 1.93	1.15 ± 1.47	0.59 ± 0.51	0.46 ± 0.36
	Second trimester	0.82 ± 3.18	0.55 ± 0.56	0.41 ± 0.38	0.40 ± 0.39
	Third trimester	2.43 ± 2.71	2.41 ± 2.71	0.85 ± 0.72	1.02 ± 0.78
	Average	1.28 ± 6.16	0.87 ± 1.38	0.50 ± 0.48	0.47 ± 0.46

"EF" denotes ellipse fitting for the segmentation network

In order to get the ellipse closest to the fetal head, we fit the segmentation results of the U-Net and SAPNet by least squares, illustrated in Section 3.4 in details. The fitted ellipses are compared by three assessment metrics: DF, ADF, DSC, and HD. All the evaluation metric values are listed in Table 5. In Table 6, we compare our OFD angle regression network to another network with ellipse fitting. It can be seen that the direct fitting ellipse using the segmentation network can only be similar in circumference, but its OFD and BPD have a large error. The performance of the regression network during the first and second trimester was absolutely superior to other methods. However, the performance of the regression network in the third trimester was almost the same as that of other methods, which we believe is related to the fact that the middle cerebral in this stage is not obvious in most ultrasound images.

Table 6 Comparison of OFD, BPD, and OFD angle prediction of different networks

Metric	Period	U-Net (EF)	SAPNet (EF)	SAPNet (AR)
OFD length (mm) ± std	First trimester	2.23 ± 4.09	1.24 ± 2.94	0.83 ± 2.93
	Second trimester	2.39 ± 3.22	0.87 ± 2.18	0.34 ± 1.76
	Third trimester	3.14 ± 3.82	2.12 ± 3.49	2.08 ± 3.43
	Average	2.51 ± 3.78	1.07 ± 3.43	0.52 ± 2.74
BPD length (mm) ± std	First trimester	1.33 ± 2.91	1.04 ± 2.45	0.46 ± 2.07
	Second trimester	1.07 ± 1.81	0.57 ± 1.49	0.41 ± 1.35
	Third trimester	1.83 ± 2.48	1.36 ± 2.28	1.40 ± 2.50
	Average	1.29 ± 2.45	0.82 ± 2.21	0.57 ± 1.83
Angle OFD (mm) ± std	First trimester	4.06 ± 4.23	3.99 ± 4.66	1.34 ± 2.91
	Second trimester	3.60 ± 4.23	3.34 ± 4.10	1.01 ± 2.23
	Third trimester	4.88 ± 4.91	4.71 ± 4.83	4.02 ± 3.91
	Average	3.86 ± 4.82	3.61 ± 4.34	1.67 ± 3.05

"EF" represents the measurement on the ellipse after fitting the output of the segmentation network. The length of OFD is the long axis of the ellipse, the length of BPD is the short axis, and the angle of OFD is the included angle between the long axis and the X-axis of the image. "AR" represents the measurement after the fusion angle of the regression network

Table 7 Ablation analysis of our proposed SAPNet with different settings

Network	mIoU (%)± std	mPA (%)± std	DSC(%)± std
SAPNet (DC)	95.16 ± 3.07	96.51 ± 1.89	95.58 ± 2.11
SAPNet (FP)	96.32 ± 2.14	97.67 ± 1.66	96.24 ± 2.08
SAPNet (SAM)	96.46 ± 1.77	98.02 ± 0.98	97.26 ± 1.50

"DC" denotes dilated convolution. "FP" denotes feature pyramid without scale attention. "SAM" means our scale attention module

4.3 Ablation experiments

To show the effectiveness of different components in our SAPNet, we present an ablation experiment to quantitative analysis of the following components: dilated convolution, feature pyramid, and scale attention module, as described before. As listed in Table 7, these experiments show that different factors have an effect on the final result.

Ablation study for segmentation network: As shown in Table 8, we first test the effect of different layers of feature pyramids on the final result, and the input feature map of each pyramid layer is adjusted to the same size. Finally, we find that the best results are achieved in three layers. In order to reduce the detail loss caused by upsampling on the small feature map of the feature pyramid layer, we replace the last two layers of the encoding network with dilated convolution. we notice that dilated convolution works are better than ordinary convolution, as shown in Table 7. Furthermore, when we replace the scale attention module to the feature pyramid, the performance of the network is further improved.

Ablation study for regression network: As shown in Table 9, we test the difference between the absolute and incremental OFD angle of regression network prediction. The absolute angle is the angle between OFD and the *X*-axis of the image, and the incremental angle is the intersection angle between the long axis of the ellipse after fitting and the real OFD. The activation function σ of the last layer of the regression network that predicts the absolute angle is set

Table 8 Detailed analysis of our proposed SAPNet with different layers of feature pyramids

Network	mIoU (%) ± std	mPA (%) ± std	DSC (%) ± std
Feature pyramid (2)	94.72 ± 3.98	95.67 ± 2.16	94.81 ± 3.22
Feature pyramid (3)	95.16 ± 3.07	96.51 ± 1.89	95.58 ± 2.11
Feature pyramid (4)	95.12 ± 3.14	96.46 ± 1.66	95.24 ± 2.18

Table 9 Comparison of OFD angle with different prediction methods

Network	OFD (mm) ± std	BPD (mm) ± std	Angle (°) ± std
Absolute angle	1.01 ± 3.01	0.66 ± 1.45	2.78 ± 2.82
Incremental angle	0.52 ± 2.74	0.57 ± 1.83	1.67 ± 3.05

All network parameters are the same except for the last layer activation function

to 180. Compared with absolute angle prediction, the incremental angle is equivalent to σ reduction, and the search space of deep neural network is also reduced.

To verify the performance of the ellipse pooling module, we add a new feature extraction layer with the same encoding structure as SAPNet, followed by the OFD Angle regression network. The feature extraction structure runs in parallel with SAPNet, relying on the back-propagation gradient of the regression network. The results are shown in Table 10.

4.4 Results in test set

Combining our best setting in the deep neural network, we experiment with the automated measurement system of fetal head on the HC18 test set. In evaluation, we use these best settings to train 700 epochs with the Adam optimizer, so the result would be better than our validation set. The final output of the entire system comes in the 1st place in the HC18 leader board (December 23, 2018, account name: shen-zexu). Without adding the regression network, the result of using only the output of the segmentation network with fitting ellipse ranks the fourth. In order to ensure the authority of the evaluation, we only selected published paper results for comparison, as shown in Table 11. Our best result with the SAPNet achieves a score 1.81±1.69/97.94±1.34/0.59± 2.41/1.22 ± 0.77 in terms of ADF, DSC, DF, and HD.

5 Discussion

The most important observation in our experiment is that we use multi-scale information to synthesize local and global context to identify the edge information of the skull, while

Table 10 Comparison of the results of both ellipse pooling and none ellipse pooling

Network	OFD (mm) ± std	BPD (mm) ± std	Angle (°) ± std
None ellipse pooling	2.34 ± 2.89	1.09 ± 2.08	3.38 ± 3.27
Ellipse pooling	0.52 ± 2.74	0.57 ± 1.83	1.67 ± 3.05

Table 11 Results of the HC18 challenge

Method	Rank	Period	ADF (mm) ± std	DSC (%) ± std	DF (mm) ± std	HD (mm) ± std
SAPNet (DC+SAM+AR)	1	First trimester	1.37 ± 1.25	96.96 ± 2.43	0.38 ± 1.82	0.80 ± 0.63
		Second trimester	1.67 ± 1.47	98.16 ± 0.92	0.49 ± 2.17	1.15 ± 0.68
		Third trimester	3.02 ± 2.47	98.03 ± 0.71	1.31 ± 3.69	2.06 ± 0.71
		Average	1.81 ± 1.69	97.94 ± 1.34	0.59 ± 2.41	1.22 ± 0.77
SAPNet	4	First trimester	1.63 ± 1.76	96.69 ± 3.00	0.69 ± 2.31	0.88 ± 0.66
		Second trimester	1.78 ± 1.61	98.10 ± 0.97	0.76 ± 2.28	1.19 ± 0.68
		Third trimester	3.06 ± 2.52	98.07 ± 0.71	1.47 ± 3.70	2.07 ± 0.78
		Average	1.93 ± 1.84	97.86 ± 1.57	0.85 ± 2.53	1.26 ± 0.77
Wu et al. [7]	9	First trimester	2.00 ± 2.45	93.65 ± 5.15	1.40 ± 2.84	1.54 ± 1.15
		Second trimester	2.07 ± 1.76	92.94 ± 6.55	1.22 ± 2.43	3.65 ± 3.18
		Third trimester	2.95 ± 2.48	93.52 ± 6.86	1.93 ± 3.35	5.74 ± 5.48
		Average	2.18 ± 2.01	93.14 ± 6.38	1.35 ± 2.64	3.60 ± 3.56
Thomas et al. [18]	25	First trimester	3.06 ± 4.83	94.65 ± 5.43	−0.32 ± 5.72	1.69 ± 2.28
		Second trimester	2.38 ± 2.38	97.63 ± 1.37	0.77 ± 3.28	1.62 ± 1.27
		Third trimester	4.79 ± 3.43	97.37 ± 1.17	0.56 ± 5.91	3.07 ± 1.60
		Average	2.83 ± 3.16	97.10 ± 2.73	0.56 ± 4.21	1.83 ± 1.60

The results of the published papers are listed. The final ranking is based on absolute head circumference. The results are evaluated on the testing set with no annotated information

the regression network can correct the elliptic geometric axes into biological OFD and BPD. A feature pyramid is established at the feature level to utilize local and global information corresponding to different sizes of receptive fields in the feature layer. Our network structure is quite different from the previous network that used a single scale to segment the head region of a fetus. Furthermore, we proposed a scale attention module for multi-scale information fusion, which yielded better performance in our experiments.

On the other hand, in different previous approaches that treat biological and geometric lengths equally, we add a regression network to obtain OFD and BPD by modifying the major and minor axes of the ellipse. In our regression network, the ellipse pooling module plays an important role, because it can combine ellipse parameters fitted by results of the segmentation network and visual feature to modify geometric lengths. Our experiments also demonstrate the effectiveness of our regression network.

There are also some problems with our proposed network and U-Net: as shown in Table 5, these networks perform worse in the first trimester compared with the second and third trimesters. This is because the fetus in the first gestation period has a softer skull tissue, which is very similar to the tissue inside the skull, so there is no obvious characteristic change between the skull and the inside of the fetal head in the ultrasound images. This can serve as an open question to further advance the measurement of the fetal head. One of the simplest treatments is to design a network structure for the first trimester.

6 Conclusion

In this work, we proposed a novel deep neural network that uses multi-scale information for fetal head segmentation and accurate BPD, OFD prediction in ultrasound images, and design an automatic measurement system based on the network structures. We design the SAPNet that establishes feature pyramids and uses attention mechanism to select feature layers. The SAPNet that uses scale information can fuse local and global information to infer skull boundaries that contain speckle noise or discontinuities. Based on the segmentation results of SAPNet, we obtain the head circumference by performing ellipse fitting in the least squares method. Ellipse pooling is used to project the ellipse parameters to the encoding feature layer of the segmentation network, and the elliptic geometric axes are modified by the regression network to obtain more accurate BPD and OFD. Our experimental results show that the proposed approach can achieve comparable performance with other models on the HC18 dataset. However, our results were only significant in ultrasound images of a single target. Future work should include multi-order data so that they be able to evaluate the performance on the fetal heads' regions of twins.

References

1. Loughna P, Chitty L, Evans T, Chudleigh T (2009) Fetal size and dating: charts recommended for clinical obstetric practice. Ultrasound 17(3):160–166
2. Rueda S, Fathima S, Knight CL, Yaqub M, Papageorghiou AT, Rahmatullah B, Foi A, Maggioni M, Pepe A, Tohka J (2014) Evaluation and comparison of current fetal ultrasound image segmentation methods for biometric measurements: a grand challenge. IEEE Trans Med Imaging 33(4):797
3. Sylvia R, Sana F, Knight CL, Mohammad Y, Papageorghiou AT, Bahbibi R, Alessandro F, Matteo M, Antonietta P, Jussi T (2014) Evaluation and comparison of current fetal ultrasound image segmentation methods for biometric measurements: a grand challenge. IEEE Trans Med Imaging 33(4):797–813
4. Verburg BO, Steegers EAP, Ridder D, Snijders MRJM, Smith E, Hofman A, Moll HA, Jaddoe VWV, Witteman JCM (2008) New charts for ultrasound dating of pregnancy and assessment of fetal growth: longitudinal data from a population-based cohort study. Ultrasound Obst Gyn 31(4):388–396
5. Noble JA (2010) Ultrasound image segmentation and tissue characterization. Proc Inst Mech Eng H 224(2):307–316
6. Ciurte A, bresson X, Cuadra MB (2012) A semi-supervised patch-based approach for segmentation of fetal ultrasound imaging. In: ISBI Challenge Us: Biometric Measurements from Fetal Ultrasound Images
7. Wu L, Xin Y, Li S, Wang T, Heng P-A, Ni D (2017) Cascaded fully convolutional networks for automatic prenatal ultrasound image segmentation. In: 2017 IEEE 14th International Symposium on Biomedical Imaging (ISBI 2017), IEEE, pp 663–666
8. Foi A, Maggioni M, Pepe A, Tohka J (2012) Head contour extraction from fetal ultrasound images by difference of Gaussians revolved along elliptical paths. In: Challenge Us: Biometric Measurements from Fetal Ultrasound Images Isbi
9. Stebbing RV, McManigle JE (2012) A boundary fragment model for head segmentation in fetal ultrasound. In: Proceedings of Challenge US: Biometric Measurements from Fetal Ultrasound Images, ISBI 2012, pp 9–11
10. Lu W, Tan J (2008) Detection of incomplete ellipse in images with strong noise by iterative randomized hough transform (irht). Pattern Recogn 41(4):1268–1279
11. Satwika IP, Habibie I, Ma'sum MA, Febrian A, Budianto E (2014) Particle swarm optimation based 2-dimensional randomized hough transform for fetal head biometry detection and approximation in ultrasound imaging. In: Advanced computer Science and Information Systems (ICACSIS), 2014 International Conference on, IEEE, pp 468–473
12. He K, Gkioxari G, Dollar P, Girshick R (2017) Mask r-cnn. IEEE Trans Pattern Anal Mach Intell PP(99):1–1
13. Ponomarev GV, Gelfand MS, Kazanov MD (2012) A multilevel thresholding combined with edge detection and shape-based recognition for segmentation of fetal ultrasound images. In: Proceedings of Challenge Us: Biometric Measurements from Fetal Ultrasound Images, ISBI 2012, Citeseer, vol 2012, pp 17–19
14. Carneiro G, Georgescu B, Good S, Comaniciu D (2008) Detection of fetal anatomies from ultrasound images using a constrained probabilistic boosting tree. IEEE Trans Medical Imaging 27(9):1342–1355
15. Zalud I, Good S, Carneiro G, Georgescu B, Aoki K, Green L, Shahrestani F, Okumura R (2009) Fetal biometry: a comparison between experienced sonographers and automated measurements. J Matern-Fetal Neo Med 22(1):43–50
16. Ni D, Yang Y, Li S, Qin J, Ouyang S, Wang T, Heng PA (2013) Learning based automatic head detection and measurement from fetal ultrasound images via prior knowledge and imaging parameters. In: Biomedical Imaging (ISBI), 2013 IEEE 10th International Symposium on, IEEE, pp 772–775
17. Jatmiko W, Habibie I, Ma'sum MA, Rahmatullah R, Satwika IP (2015) Automated telehealth system for fetal growth detection and approximation of ultrasound images. Int J Smart Sens Intell Syst 8(1)
18. van den Heuvel TL, de Bruijn D, de Korte CL, van Ginneken B (2018) Automated measurement of fetal head circumference using 2d ultrasound images. PloS One 13(8):e0200412
19. Krizhevsky A, Sutskever I, Hinton GE (2012) Imagenet classification with deep convolutional neural networks. In: International Conference on Neural Information Processing Systems, pp 1097–1105
20. Simonyan K, Zisserman A (2015) Very deep convolutional networks for large-scale image recognition. In: International Conference on Learning Representations
21. He K, Zhang X, Ren S, Sun J (2016) Deep residual learning for image recognition. In: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, pp 770–778
22. Long J, Shelhamer E, Darrell T (2015) Fully convolutional networks for semantic segmentation. In: IEEE Conference on Computer Vision and Pattern Recognition, pp 3431–3440
23. Ronneberger O, Fischer P, Brox T (2015) U-net: convolutional networks for biomedical image segmentation. In: International Conference on Medical Image Computing and Computer-Assisted Intervention. Springer, New York, pp 234–241
24. Alison NJ (2016) Reflections on ultrasound image analysis. Med Image Anal 33:33–37
25. Litjens G, Kooi T, Bejnordi BE, Aaa S, Ciompi F, Ghafoorian M, Jawm VDL, Van GB, Sánchez CI (2017) A survey on deep learning in medical image analysis. Med Image Anal 42(9):60–88
26. Shen D, Wu G, Suk HI (2017) Deep learning in medical image analysis. Annu Rev Biomed Eng 19(1):221–248
27. Chen L, Yang Y, Wang J, Xu W, Yuille AL (2016) Attention to scale: scale-aware semantic image segmentation. In: Computer vision and pattern recognition, pp 3640–3649
28. Li H, Xiong P, An J, Wang L (2018) Pyramid attention network for semantic segmentation, arXiv:1805.10180
29. Kingma D, Ba J (2014) Adam: A method for stochastic optimization, Comput Sci

Peixuan Li received the B.S. degree in automation from Tianjin University of Technology and Education in 2016. He is currently pursuing the Ph.D. degree with the Shenyang Institute of Automation, Chinese Academy of Sciences, Beijing, China. His current research interests include deep learning, medical image processing, SLAM and pattern recognition.

Huaici Zhao received the Ph.D. degree in Mechanical and Electronic Engineering from the Shenyang Institute of Automation, Chinese Academy of Sciences, in 2003. He is currently a Professor with the Shenyang Institute of Automation. His research interests cover deep learning, image deblurring, medical image processing, hyperspectral image processing.

Feidao Cao received the B.Eng degrees in automation from Liaoning Technical University, Liaoning, China, in 2015. He is currently pursuing the Ph.D. degree at Shenyang Institute of Automation, Chinese Academy of Sciences. His current research interests include image enhancement, machine learning for image processing.

Pengfei Liu is a PhD student at the Shenyang Institute of Automation, Chinese Academy of Sciences. He received his BS degree in Measurement and Control Technology in 2010 from the University of Harbin Institute of Technology and MS degree in engineering from the Civil Aviation University of China in 2015. His current research interests include deep learning, image deblurring, medical image processing, and hyperspectral image processing.