



RDHCformer: Fusing ResDCN and Transformers for Fetal Head Circumference Automatic Measurement in 2D Ultrasound Images

Chaoran Yang^{1,2}, Shanshan Liao³, Zeyu Yang⁴, Jiaqi Guo¹, Zhichao Zhang¹, Yingjian Yang^{1,2}, Yingwei Guo^{1,2}, Shaowei Yin³, Caixia Liu³ and Yan Kang^{1,2,5*}

¹ College of Medicine and Biological Information Engineering, Northeastern University, Shenyang, China, ² Medical Device Innovation Center, Shenzhen Technology University, Shenzhen, China, ³ Department of Obstetrics, Shengjing Hospital of China Medical University, Shenyang, China, ⁴ Department of Ultrasound, Shengjing Hospital of China Medical University, Shenyang, China, ⁵ Engineering Research Centre of Medical Imaging and Intelligent Analysis, Ministry of Education, Shenyang, China

OPEN ACCESS

Edited by:

Kuanquan Wang,
Harbin Institute of Technology, China

Reviewed by:

Yili Zhao,
University of California, San Francisco,
United States
Xueju Wang,
Peking University Third Hospital, China

Li Ke,
Shenyang University of
Technology, China

*Correspondence:

Yan Kang
kangyan@bmie.neu.edu.cn

Specialty section:

This article was submitted to
Precision Medicine,
a section of the journal
Frontiers in Medicine

Received: 05 January 2022

Accepted: 07 March 2022

Published: 29 March 2022

Citation:

Yang C, Liao S, Yang Z, Guo J, Zhang Z, Yang Y, Guo Y, Yin S, Liu C and Kang Y (2022) RDHCformer: Fusing ResDCN and Transformers for Fetal Head Circumference Automatic Measurement in 2D Ultrasound Images. *Front. Med.* 9:848904. doi: 10.3389/fmed.2022.848904

Fetal head circumference (HC) is an important biological parameter to monitor the healthy development of the fetus. Since there are some HC measurement errors that affected by the skill and experience of the sonographers, a rapid, accurate and automatic measurement for fetal HC in prenatal ultrasound is of great significance. We proposed a new one-stage network for rotating elliptic object detection based on anchor-free method, which is also an end-to-end network for fetal HC auto-measurement that no need for any post-processing. The network structure used simple transformer structure combined with convolutional neural network (CNN) for a lightweight design, meanwhile, made full use of powerful global feature extraction ability of transformer and local feature extraction ability of CNN to extract continuous and complete skull edge information. The two complement each other for promoting detection precision of fetal HC without significantly increasing the amount of computation. In order to reduce the large variation of intersection over union (IOU) in rotating elliptic object detection caused by slight angle deviation, we used soft stage-wise regression (SSR) strategy for angle regression and added KLD that is approximate to IOU loss into total loss function. The proposed method achieved good results on the HC18 dataset to prove its effectiveness. This study is expected to help less experienced sonographers, provide help for precision medicine, and relieve the shortage of sonographers for prenatal ultrasound in worldwide.

Keywords: prenatal ultrasound, fetal head circumference, rotating object detection, transformers, convolutional neural network

INTRODUCTION

Prenatal ultrasound is one of the most important examination methods during pregnancy due to its fast, low-risk and non-invasive characteristics. Fetal head circumference (HC) is one of the most essential biological indexes in accurate assessment of fetal development, which provides a method for monitoring fetal growth, estimating gestational age, and determining delivery mode. It is of

paramount importance to ensure the continued wellbeing of mothers and newborns both during and after pregnancy. In prenatal ultrasound screening, the fetal head circumference is measured on standard plane of thalamus according to the obstetric ultrasound guidelines (1, 2), and the circumference of ellipse can be identified as the fetal HC since the contour of skull is similar to an ellipse. During measurement, the contour of skull is marked by sonographers, and the HC can be calculated by ellipse parameters which is obtained through fitting on post-processing software embedded in ultrasound equipment. Some semi-automated HC measurement is available on newer OB ultrasound machines, like GE Voluson E10 (SonoBiometry). The measurement results of semi-automated methods are directly affected by the accuracy in performing segmentation.

However, it is challenging for AI models to measure HC due to blurred or incomplete skull edge in ultrasound images. Accurate measurement can provide an important reference for the evaluation of fetal growth and development. Therefore, in order to improve efficiency, reliability, and reduce the workload of doctors, clinical practice puts forward high requirements for automatic segmentation (3). It is of great significance to develop an efficient and accurate method for automatic measurement of fetal HC.

In this paper, a lightweight detection network that combined with Transformer and Convolutional neural network (CNN) is proposed to detect the position of the fetal head, regress the parameters of ellipse, and then solve the head circumference value through the parameters. For automatic measurement tasks of HC, it is a one-stage network of detection. The process does not require any post-processing, such as edge extraction or ellipse fitting, and the process comparison between our method and general detection method is shown in **Figure 1**. This work makes the following contributions:

- 1) To our knowledge, our method is the first to apply the rotating ellipse detection method to the skull edge detection task. This is a one-stage network based on anchor-free method;
- 2) Taking Res_DCN as baseline, Deformable Convolutional Networks (DCN) combined with ResNet can learn the features of irregular boundary better and promote capability of local feature extraction. Meanwhile, powerful global feature extraction ability of Transformer is used to obtain more abundant continuous features of boundary from the global view. The proposed approach combines simple Transformer structure with CNN to obtain complete and accurate elliptical information as much as possible without significantly increasing the amount of computation;
- 3) Soft Stagewise Regression (SSR) strategy is used to map angle regression problems into classification problems. Firstly, the angle is roughly classified, and then the dynamic range is introduced to make every bin can do translation and scaling for fine classification. Classify the angles from coarse to fine to make angle regression accuracy higher;
- 4) Kullback-Leibler Divergence (KLD) loss that is similar to IOU loss is added into total loss function to solve the problem that intersection over union (IOU) between ground truth (GT) and prediction changes greatly caused by small angle deviation or center point deviation of the rotating target, as the IOU of rotate target is difficult to calculate. KLD loss can further improve the regression accuracy of elliptic parameters;
- 5) The proposed method gets good results compared with other existing HC measurement methods in open data set of HC18. It is noteworthy that the method is simple and efficient without requiring any post-processing.

RELATED WORK AND MOTIVATION

Related Work

In the past research, many methods based on machine learning have been used to extract skull edge features, such as Haar-like features combined with different classifiers (4–9). There are also some methods based on gradient (10), threshold (11), active boundary model (12), contour fragment model (13), multi-groupfilters mixing (14) to extract features of skull region or boundary. After the skull features were extracted, different methods such as Hough transform (15) and ElliFit (16) were used to fit the elliptic skull boundary and further measure HC. Although some good results have been achieved by above methods, they all require prior knowledge or artificially designed features with poor robustness and large amount of calculation.

In recent years, CNN have been widely used in medical image segmentation (17, 18), Sinclair et al. (19) and Wu et al. (20) used the cascaded Fully Convolutional Network (FCN) to segment the skull region. U-net and its extended form have a symmetrical structure and extract rich features by using the fusion of different feature layers (21, 22). There are also some methods with different understanding of tasks, such as multi-organ segmentation (23), segmentation and regression multi-task methods (24), which are widely used in skull region or boundary segmentation. Skull boundary detection based on CNN segmentation method has excellent performance in regional segmentation, after predicting the skull region, a series of complex post-processing such as expansion, corrosion and edge extraction are carried out to obtain the skull boundary pixels, and then ellipse parameter fitting is carried out to solve HC, therefore, these methods have huge networks and cumbersome process. The measurement accuracy of head circumference depends on the segmentation result heavily, and the effect is not good for the ultrasonic image with unclear or incomplete boundary.

Object detection technology based on anchor method has good detection results for standard rectangular frame targets (25, 26), but there are no relevant studies on rotating elliptic object detection (i.e., skull edge detection task). The method based on anchor need to preset size of anchor according to IOU, and an appropriate number of anchors are selected with a certain threshold value (such as 0.5) as positive samples for regression distribution of objects. But this leads to two problems in rotating object detection: first, further aggravating the positive and negative sample imbalance. Angle prior should be added to the preset rotating anchor, doubling the number of preset anchors. In addition, rotating anchor angle slightly deviated from GT will lead to sharp decline of IOU. Second, classification is inconsistent with regression. Many studies have discussed this

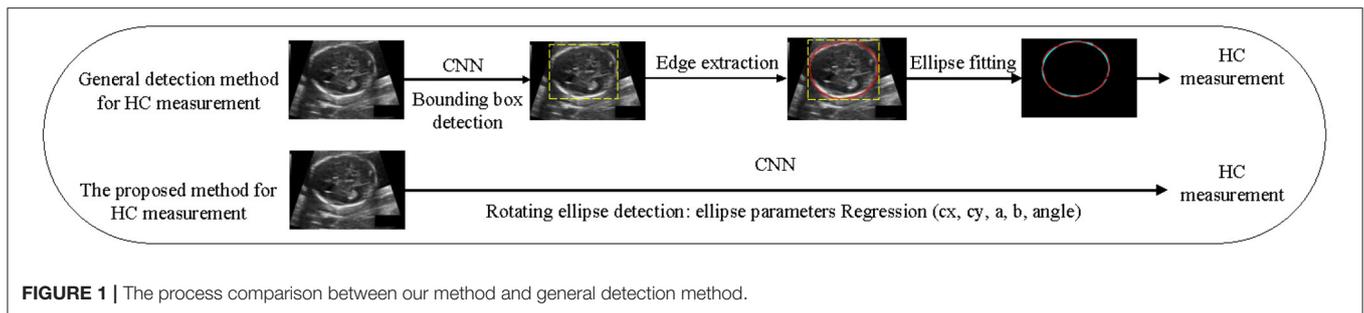


FIGURE 1 | The process comparison between our method and general detection method.

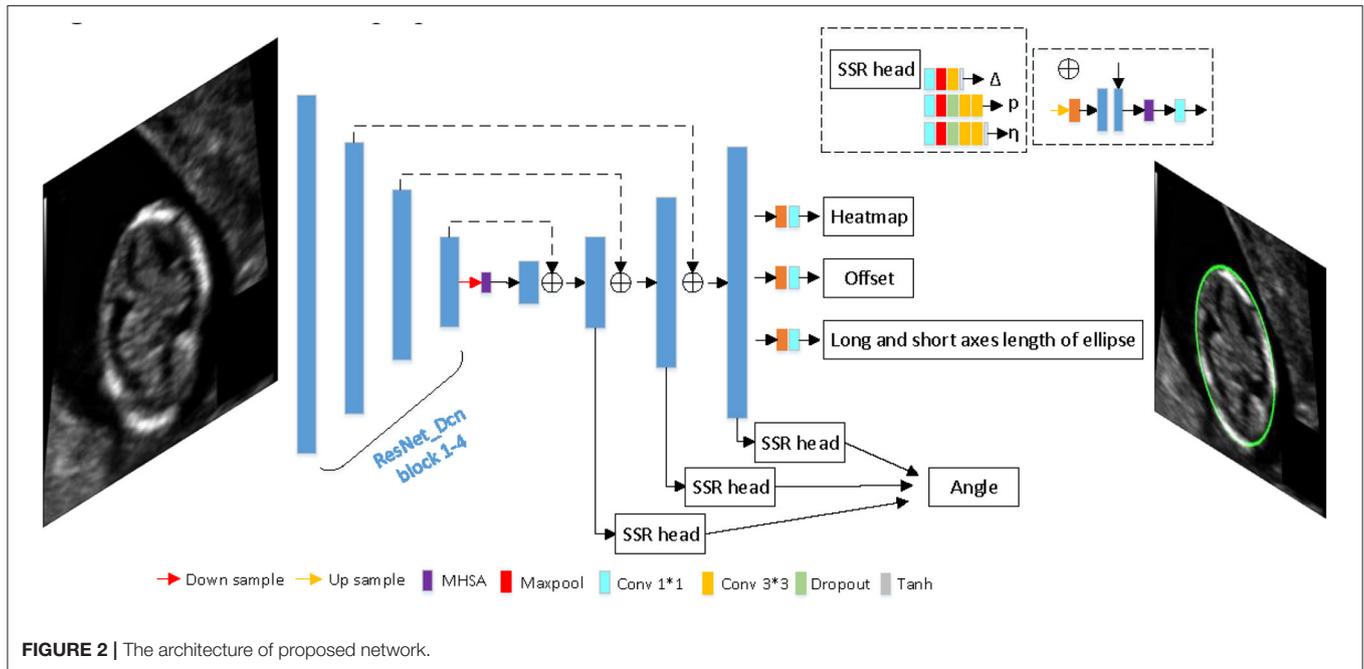


FIGURE 2 | The architecture of proposed network.

problem (27), that is, the classification score of predicted results is inconsistent with the positioning accuracy, so inaccurate positioning may be selected when passing the NMS stage or selecting detection results according to the classification score, while the well-positioned anchor is omitted or suppressed.

Motivation

For skull edge detection task, due to factors such as fetuses at different gestational ages and different positions, the skull edge presents elliptic shapes of different sizes. The detection method based on Anchor needs to design sizes of different proportions according to prior, which is a very complicated process. In addition, IOU between GT and prediction changes greatly because of small angle deviation or center point deviation of the rotating target. Recently, the object detection method based on anchor-free has been greatly developed. CenterNet (28) detects the center point of the object first, and then directly regress the width and height of the object. Of course, we can directly regress a rotation angle to expand CenterNet to rotating object detection. However, the size and angle actually depend on different rotating coordinate systems, so it is difficult to directly

regress parameters. To sum up, we are committed to studying a lightweight and high-precision rotating ellipse detection network for skull edge detection. The proposed method is a one-stage method based on anchor-free to solve the above problems.

METHODS

In this section, we first describe the overall architecture of the proposed method, and then explain the Gaussian distribution of GT, output maps, and KLD loss function in detail. The output maps are used to generate the oriented ellipse of the objects.

Architecture

Since our goal is to build a lightweight network, we didn't choose backbone which is too complicated. The proposed network is based on an asymmetric U-shaped architecture (see **Figure 2**). We use the block 1–5 of ResNet_DCN as the backbone, simple Multi-head-self-attention [MHSA, see **Figure 3**, details in reference (29)] is used in encoder's last bottleneck module and the whole up-sampling process. Deformable convolution and self-attention mechanism are used to improve the access

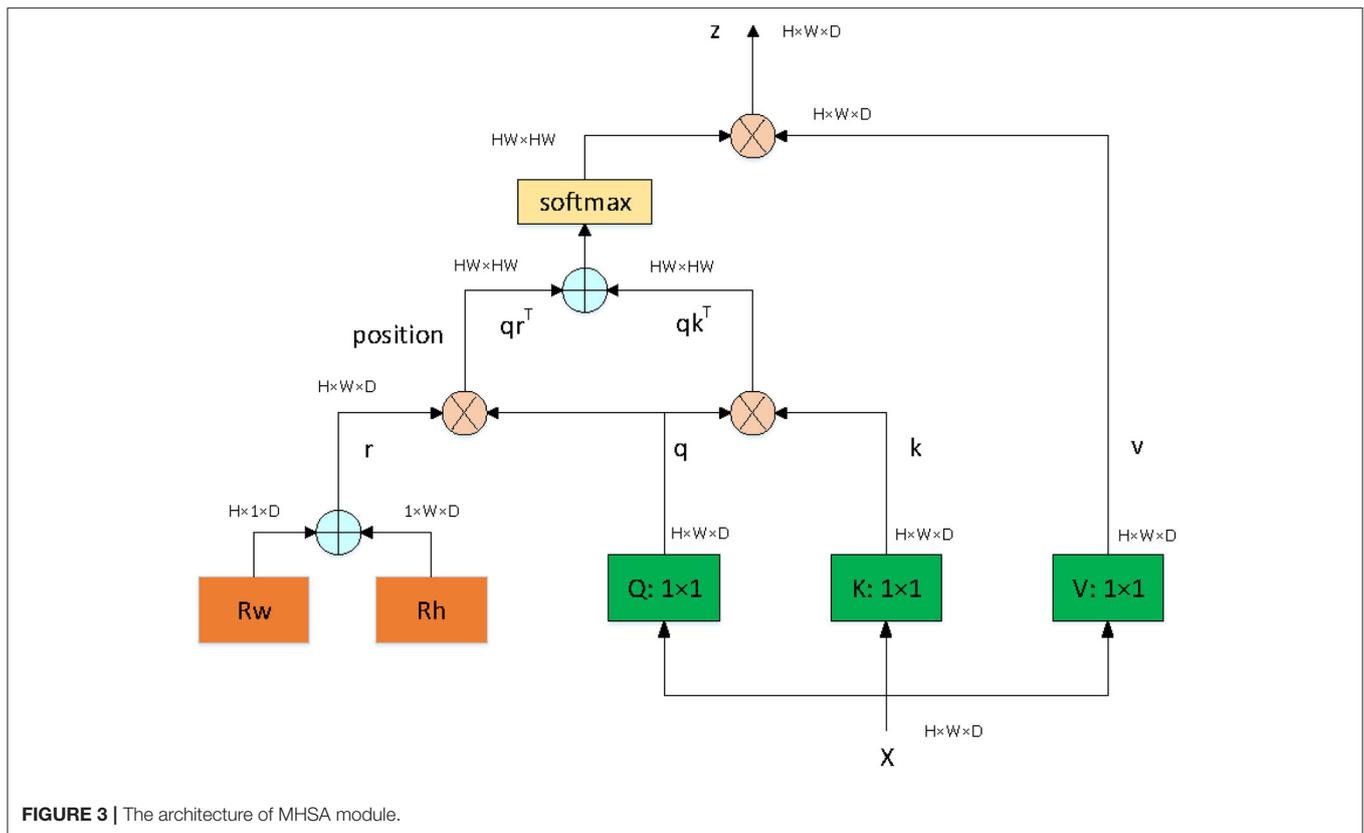


FIGURE 3 | The architecture of MHA module.

to local information and the continuity of irregular boundary. In decoder, output features of encoder are up-sampled to 1/4 of the input image (scale $s = 4$), we combine a deep layer with a shallow layer through skip connections to share both the high-level semantic information and low-level finer details. In particular, we first up-sample a deep layer to the same size of the shallow layer through bilinear interpolation. The up-sampled features map is refined through a 3×3 convolutional layer. The refined feature map is then concatenated with the shallow layer, followed by a 1×1 convolutional layer to refine the channel-wise features. In the end, four detection heads are used for ellipse parameters regression (heatmap, center offset, long and short axes of ellipse, angle).

Gaussian Distribution of GT

The proposed method locates the target based on free-anchor, we need to map GT of keypoints to a 2D Gaussian distribution on the heatmap. The mapping method in reference (28) is not friendly to targets with a large aspect ratio, especially for ellipse, so we modified the mapping method. The GT of an ellipse is $(c_x, c_y, a, b, \text{angle})$, where (c_x, c_y) is the center point of an oriented ellipse, a and b are long and short axes of ellipse, respectively, angle is the angle between the short axis and the vertical direction. We generate the smallest horizontal enclosing rectangular box of the ellipse (b_x, b_y, w, h) , (b_x, b_y) is the center point of smallest horizontal enclosing rectangular box of the ellipse, w and h are width and height of rectangular box, respectively. We map GT

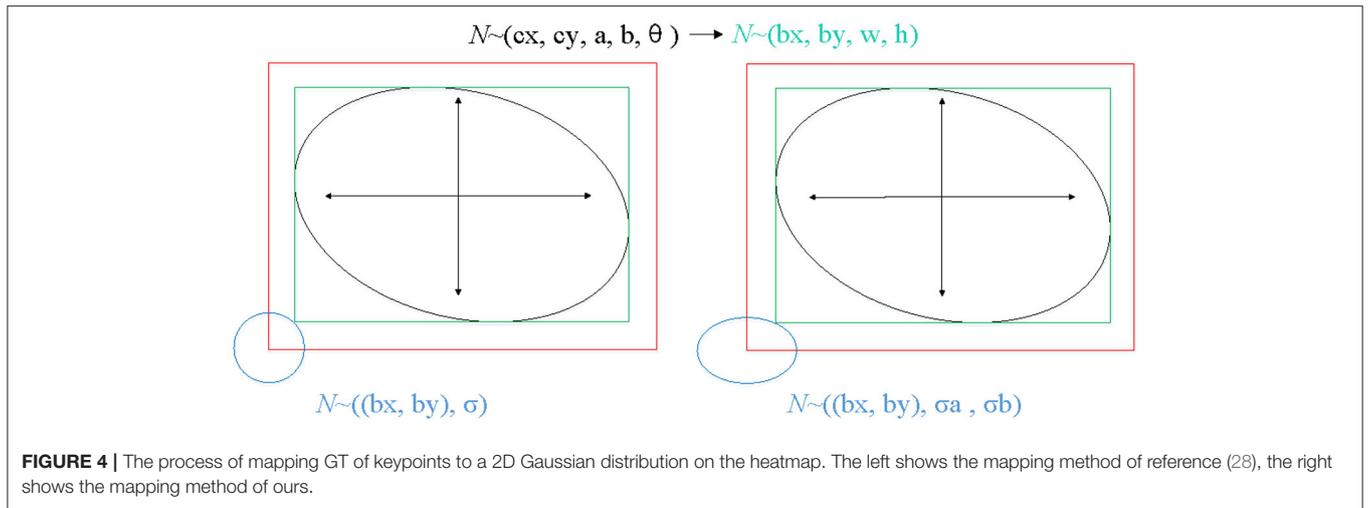
(b_x, b_y, w, h) which can be predicted as a positive sample to 2D Gaussian distribution $\exp(-(\frac{(p_x-b_x)^2}{2\sigma_a^2} + \frac{(p_y-b_y)^2}{2\sigma_b^2}))$, where σ is a box size-adaptive standard deviation, (see Figure 4). The heatmap $P \in R^{c \times \frac{H}{s} \times \frac{W}{s}}$, H and W are height and width of input image respectively, c is set to 1 in this work.

Output Maps Heatmap

In this work, we use the heatmap $\rho \in R^{c \times \frac{H}{s} \times \frac{W}{s}}$ to detect the center points of arbitrarily oriented objects, where c is corresponding to one object category. The predicted heatmap value at a particular center point is regarded as the confidence of the object detection. We use the variant focal loss to train the heatmap:

$$L_h = -\frac{1}{N} \sum_i \begin{cases} (1 - \rho_i)^\alpha \log(\rho_i) & \text{if } P_i = 1 \\ (1 - P_i)^\beta \rho_i^\alpha \log(1 - \rho_i) & \text{otherwise} \end{cases} \quad (1)$$

where P and ρ refer to the ground-truth and the predicted heatmap values, i indexes the pixel locations on the feature, N is the number of objects. α and β are the hyper-parameters that control the contribution of each point. α is set to 2 and β is set to 4.



Center Offset

We transform GT of center point by down-sample from the input image, which is a floating-point type $C = (\frac{c_x}{s}, \frac{c_y}{s})$. However, the predicted center point is an integer. To compensate for the discretization error between the floating center point and the integer center point caused by the output stride, we predict an offset map $O \in R^{2 \times \frac{H}{s} \times \frac{W}{s}}$ can be defined as:

$$O = \left(\frac{c_x}{s} - \left\lfloor \frac{c_x}{s} \right\rfloor, \frac{c_y}{s} - \left\lfloor \frac{c_y}{s} \right\rfloor \right) \quad (2)$$

The offset is optimized with a smooth L_1 loss:

$$L_O = \frac{1}{N} \sum_{k=1}^N \text{Smooth } L_1(O_k - o_k) \quad (3)$$

where N is the total number of objects, o refers to the ground-truth offsets, k indexes the objects.

The smooth L_1 loss can be expressed as:

$$\text{Smooth } L_1(x) = \begin{cases} 0.5x^2 & \text{if } |x| < 1 \\ |x| - 0.5 & \text{otherwise} \end{cases} \quad (4)$$

Long and Short Axes of Ellipse

We regress to long and short axes of ellipse for each object, $B = (a, b) \in R^{2 \times \frac{H}{s} \times \frac{W}{s}}$, where a is long axes length, b is short axes length. It can be optimized with a smooth L_1 loss:

$$L_B = \frac{1}{N} \sum_{k=1}^N \text{Smooth } L_1(B_k - b_k) \quad (5)$$

where B and b are the ground-truth and the predicted ellipse parameters, respectively.

Angle

Accurate angle regression is very important for rotating object detection, a small angle variation has marginal influence on the total loss in training, but it may induce a large IOU difference

between the predicted ellipse and the ground-truth ellipse. Because of the symmetry of the ellipse, the rotation angles $\theta \in [0, 180)$. Soft-stagewise regression strategy is adopted for angle regression, which takes angle regression as a multi-classification task. We set it as a three-stage classification task ($S_1 = 18, S_2 = 10, S_3 = 10$). In the first stage, the angle $\theta \in [0, 180)$ is divided into S_1 parts with a span of $180 / S_1$. In the second stage, $[0, 180 / S_1]$ is divided into S_2 parts with a span of $180 / S_1 / S_2$. The third stage is similar, as shown in the **Figure 5**. In each stage, it is a multi-classification task, the sum of the probability of each class and the representative angle of the current class is taken as the final prediction value. The angle is predicted by the following formula for soft-stagewise regression:

$$\theta = \sum_{k=1}^K \sum_{i=0}^{S_k-1} p_i^{(k)} i \left(\frac{V}{\prod_{j=1}^k s_j} \right) \quad (6)$$

where $V \in [0, 180)$, $p_i^{(k)}$ refers to the probability of each class for each stage, The last term in the above equation is the bin width $\omega_k = \frac{V}{\prod_{j=1}^k s_j}$ for the k -th stage and i is the bin index. Reference (30) introduced a dynamic range for each bin, that is, it allowed each bin to be shifted and scaled according to the input image. For adjusting the bin width ω_k at the k -th stage, SSR introduce a term Δk to modify s_k into s_k^* as follows:

$$s_k^* = s_k(1 + \Delta k) \quad (7)$$

where Δk is the output of a regression network given the input image. For shifting bins, we add an offset term η to each bin index i . The bin index i is modified as follows:

$$i = i + \eta_i^{(k)} \quad (8)$$

Thus, the output of SSR head are $p_i^{(k)}$, Δk , and $\eta_i^{(k)}$, the angle is regressed from coarse to fine by introducing dynamic range

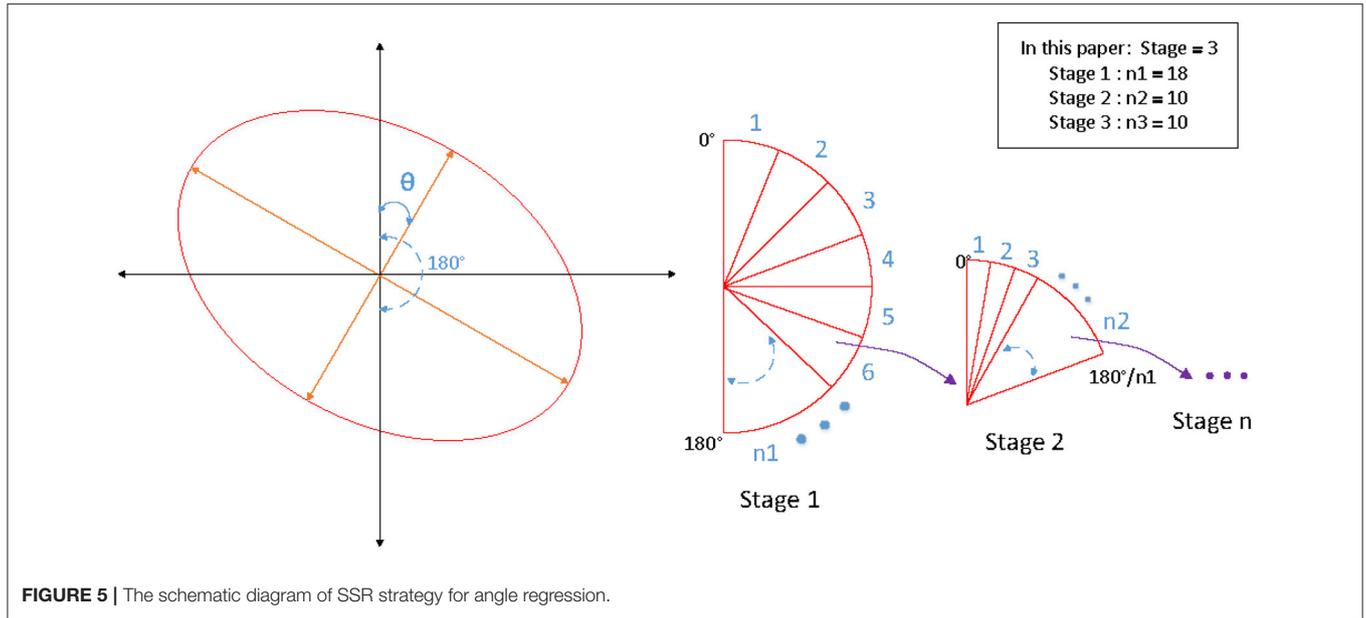


FIGURE 5 | The schematic diagram of SSR strategy for angle regression.

that each bin can do translation and scaling so as to improve the precision of angle regression and reduce the error as much as possible. Angle regression can be optimized with a smooth L_1 loss:

$$L_{\theta} = \frac{1}{N} \sum_{k=1}^N \text{Smooth } L_1(\theta_k - \theta_k^*) \quad (9)$$

where θ_k and θ_k^* are the ground-truth and the predicted ellipse parameters, respectively.

Kullback-Leibler Divergence Loss

The IOU of rotate object is difficult to be calculated, we use KLD to measure the similarity of the two distributions to approximate the IOU. KLD loss also has some advantages when optimizing parameters. When one of the parameters is optimized, the other parameters will be used as its weight to dynamically adjust the optimization rate. In other words, the optimization of parameters is no longer independent, that is, optimizing one parameter will also promote the optimization of other parameters. The optimization of this virtuous circle is the key to KLD as an excellent rotation regression loss. Reference (31) proved its derivability and advantages. Convert GT of the ellipse (c_x, c_y, a, b, θ) into a 2-D Gaussian $N(\mu, \varepsilon)$, (see **Figure 6**). Specifically, the conversion is:

$$\begin{aligned} \mu &= (c_x, c_y)^T \\ \varepsilon^{1/2} &= \begin{pmatrix} \frac{b}{2} \cos^2 \theta + \frac{a}{2} \sin^2 \theta & \frac{b-a}{2} \cos \theta \sin \theta \\ \frac{b-a}{2} \cos \theta \sin \theta & \frac{b}{2} \sin^2 \theta + \frac{a}{2} \cos^2 \theta \end{pmatrix} \end{aligned} \quad (10)$$

$X_p \sim N_p(\mu_p, \varepsilon_p)$ and $X_t \sim N_t(\mu_t, \varepsilon_t)$, the KLD between two 2-D Gaussian is:

$$\begin{aligned} D_{kl}(N_p \| N_t) &= \frac{1}{2} (\mu_p - \mu_t)^T \varepsilon_t^{-1} (\mu_p - \mu_t) + \frac{1}{2} \text{Tr}(\varepsilon_t^{-1} \varepsilon_p) \\ &+ \frac{1}{2} \ln \frac{|\varepsilon_t|}{|\varepsilon_p|} - 1 \end{aligned} \quad (11)$$

The KLD loss is:

$$L_{reg} = 1 - \frac{1}{1 + \log(D_{kl}(N_p \| N_t) + 1)} \quad (12)$$

The final regression loss is:

$$L_{total} = L_h + L_O + L_{\theta} + L_{reg} \quad (13)$$

EXPERIMENTS

Datasets and Implementation Details

Dataset is from the HC18 grand-challenge¹ which provided 1334 2D ultrasound images from standard planes, a training set with 999 images and a test set with 335 images. Manual annotations of HC were made by senior experts. Since the data set only provides standard planes, that is to say, each image has a target, and the target accounts for a large proportion of the image. In order to balance the positive and negative samples, we used two ways to generate negative samples, one way is to remove the target in the image and fill it with surrounding information, the other way is to randomly crop the image into patches, and then resize them to the size of the network input. If the IOU with GT is < 0.3 , it will be considered as a negative sample. The size of each 2D ultrasound image is approximately $540 * 800$ with the pixel size ranging from 0.052 to 0.6 mm. Data augmentation is

¹Available online at: <https://hc18.grand-challenge.org>.

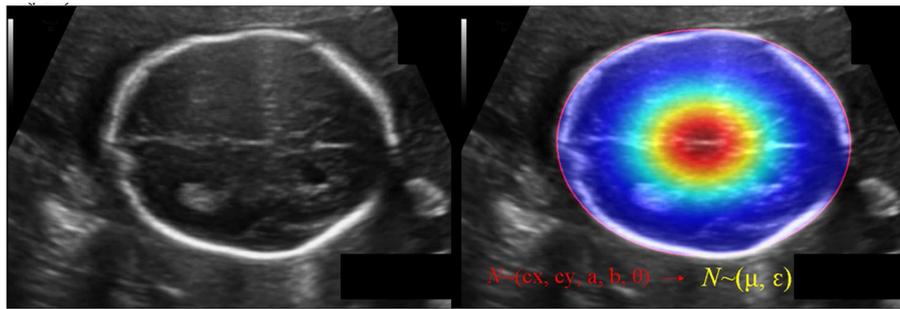


FIGURE 6 | The schematic diagram of converting GT of the ellipse (c_x, c_y, a, b, θ) into a 2-D Gaussian $N(\mu, \epsilon)$.

TABLE 1 | The comparison results between our method and other common methods base on segmentation algorithm.

Model	MAE \pm std (mm)	ME \pm std (mm)	Params (M)	AP
U-Net (33)	2.36 \pm 5.60	0.41 \pm 2.91	31.042	-
U-Net++ (34)	2.29 \pm 2.33	0.27 \pm 2.73	9.163	-
CE-Net (35)	2.24 \pm 2.28	0.16 \pm 2.12	29.003	-
SE-Unet (36)	2.27 \pm 3.61	0.09 \pm 2.33	-	-
HC18 challenge best	1.72 \pm 1.60	0.04 \pm 2.35	-	-
Ours	1.97 \pm 1.89	0.11 \pm 2.71	24.31	84.45

MAE, Mean Absolute Error; ME, Mean Error; std, standard deviation. Param, the size of model parameters; M, Mbyte; AP, Average precision. Bold values represent the best value of each indicator.

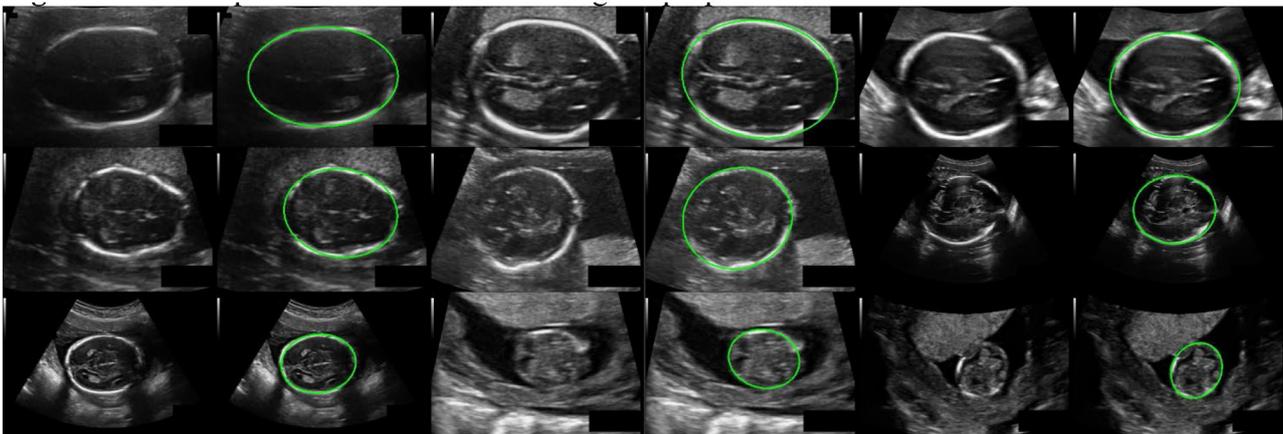


FIGURE 7 | Some examples of detection results using the proposed method.

essential to make model more robust. The data augmentation strategy was as follows: Rotation: rotation angle is $[-30^\circ, 30^\circ]$, and the interval is 10° . Scale transformation: the scaling ratio is $[0.85, 1.15]$, and the interval is 0.05. Gamma transformation: gamma factor is $[0.5, 1.5]$, and the interval is 0.1. Flip: the input image is flipped randomly. After data augmentation, training set is expanded from 999 to 12,999, of which 200 are used as validation set and the rest are used as a new training set. The Stochastic Gradient Descent (SDG) optimizer is selected, the initial learning rate is set to 0.005, the momentum is 0.9, the dropout rate is 0.1, and the batchsize is set to 16. The training

procedure is completed on two NVIDIA GeForce RTX 2080TI graphics cards.

Evaluation Metrics

In order to comprehensively evaluate the performance of the model and conduct comparative analysis, regression Average Precision (AP), Mean Absolute Error (MAE), Mean Error (ME) of head circumference are adopted as the evaluation metrics of the model in this paper. HC can be calculated as follows (32):

$$HC = \pi \left[3(a + b) - \sqrt{(3a + b)(a + 3b)} \right] \quad (14)$$

where a and b are parameters of semi-long axis and semi-short axis of the ellipse. Mean Absolute Error of fetal HC is defined as:

$$MAE_{hc} = \frac{1}{N} \sum_{i=1}^N |\widehat{HC}_i - HC_i| \tag{15}$$

Mean Error of fetal HC is defined as:

$$ME_{hc} = \frac{1}{N} \sum_{i=1}^N (\widehat{HC}_i - HC_i) \tag{16}$$

Where \widehat{HC} and HC denote the HC measured by the proposed method and the real value of fetal HC, respectively.

Average Precision is a commonly used evaluation metric in object detection that obtained by calculating the area of Precision-Recall curve.

TABLE 2 | Ablation experiments results.

Backbone	Method	AP (with MHSA/ not)
Res_DCN-50	Smooth L1 (center, a, b, angle)	81.83/77.33
	Smooth L1 (center, a, b) + SSR (angle)	83.25/79.97
	Smooth L1 (center, a, b) + SSR (angle) + KLD	84.45/81.71

Comparison of Res_DCN-50 with or without MHSA module, SSR, and KLD loss. Bold values represent the best value of each indicator.

RESULTS

Our method has achieved good results on HC18 dataset, Average precision (AP) is 84.45%. MAE \pm std (mm) is 1.97 ± 1.89 , ME \pm std (mm) is 0.11 ± 2.71 , the parameter size of the proposed model is 24.31M. **Table 1** shows the comparison results between our method and other common methods base on segmentation algorithm. It can be seen that our method has achieved good skull edge detection results without significantly increasing the amount of model parameters, and it can be comparable to the state-of-the-art method. It is worth noting that our method is simple and efficient. Unlike methods based on segmentation algorithm, our method do not need any complicated post-processing, which is an end-to-end network strictly for head circumference detection task. There are some examples of detection results in **Figure 7**.

Ablation Analysis

Some ablation experiments were conducted to prove the effectiveness of each module design in our algorithm. Taking Res_DCN-50 as the backbone as an example, the experimental results are as shown in **Table 2**. It can be seen that using normal Smooth L1 function and without MHSA module achieved the AP: 77.33%, while adding the MHSA module AP: 81.83%, increased 4.5%, it indicated that the MHSA module has a significant improvement for the task. With the addition of MHSA module, AP was increased by 1.42-83.25% by using SSR detection head for angle, then after adding KLD Loss, AP was increased

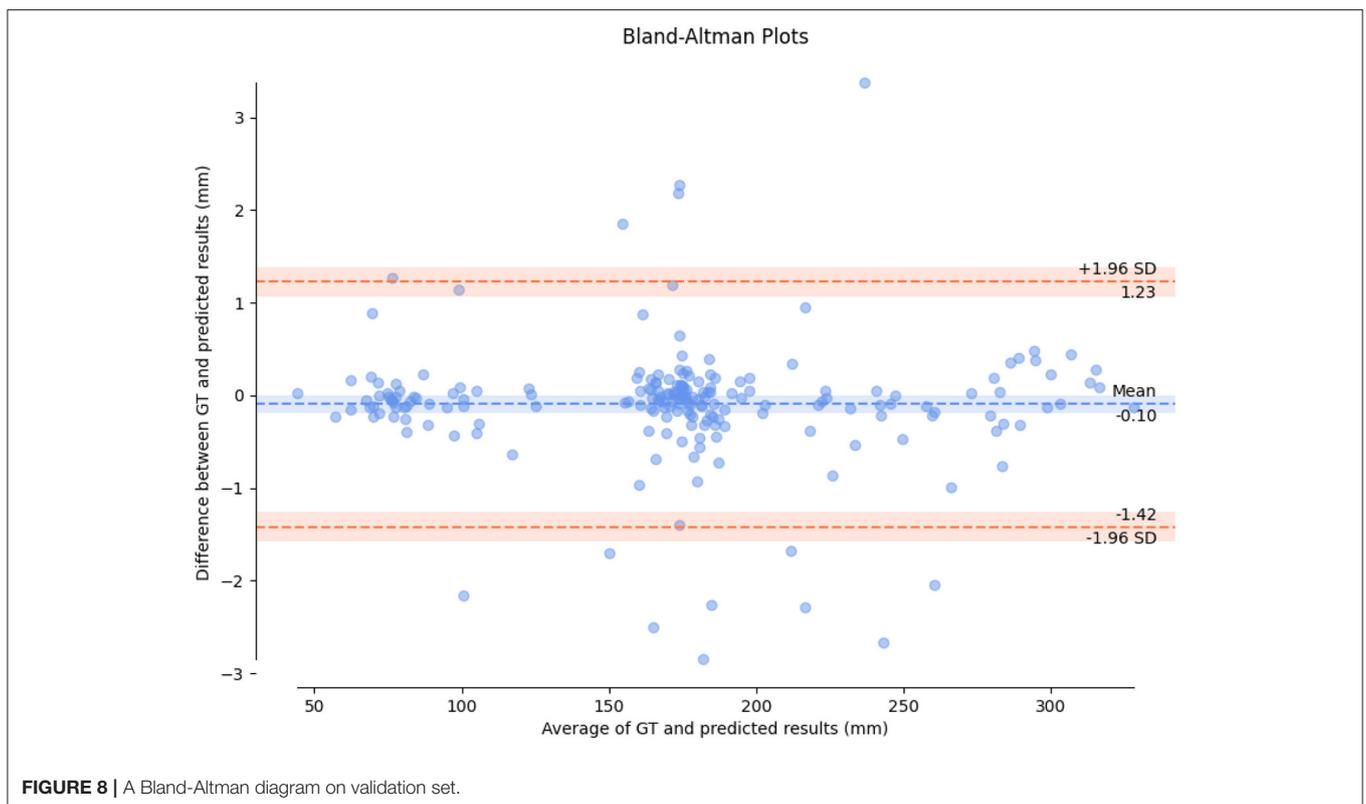


FIGURE 8 | A Bland-Altman diagram on validation set.

by 1.2–84.45% further, at this time, compared with no MHSA module AP: 81.71%, it is an increase of 2.74%. This indicated that the excellent global feature extraction ability of MHSA module improves the model's ability to extract skull edge continuity features, it is helpful for each module of the network.

Consistency Analysis

In order to evaluate the consistency between HC measured by the proposed method and real value of HC, we draw a Bland-Altman diagram on validation set, as shown in **Figure 8**. Compared with the real value of HC, the Mean Difference of HC measurement is -0.10 mm with a 95% confidence interval and the error ranges from -1.42 – 1.23 mm. It indicated the HC measured by the proposed method has a good consistency with the real value.

CONCLUSION

A new fetal head circumference auto-measurement method based on rotating ellipse detection has been proposed in this paper, which is a strictly end-to-end detection method without any post-processing for the task. As far as we know, this is the first application of end-to-end detection network to measure fetal head circumference directly. We combine transformer and CNN because convolution operations can extract rich context features in local area and transformer (MHSA) module can capture long-distance feature relationship benefitting from its ability of global and dynamic receptive fields. The two complement each other for promoting detection precision of fetal HC without significantly increasing the amount of computation. For the task of rotating elliptical object detection, the precision of angle regression is very important. Slight angle deviation will bring large changes in IOU. Therefore, we used SSR strategy for angle regression and added KLD that is approximate to IOU loss into total loss function. These methods significantly improve the detection precision. This study is expected to help less experienced sonographers, provide help for precision medicine, and relieve the shortage of sonographers for prenatal ultrasound in worldwide. There are also some shortcomings in our work, a little deviation can be allowed in predicting the location of target

center point in the inference stage (that is, positive sample can be determined if the IOU is greater than a certain threshold), therefore, in order to facilitate calculation, we conducted pre-processing operation in the process of mapping the center point of ellipse to 2D Gaussian distribution on the heatmap. We generated the smallest horizontal enclosing rectangle of the ellipse, and used center point of rectangle as the new center point for mapping. There is a slight error with the center point of the ellipse, which may affect the precision of the detection results. This is also the study direction that we need to improve in the future.

DATA AVAILABILITY STATEMENT

Publicly available datasets were analyzed in this study. This data can be found at: <https://hc18.grand-challenge.org>.

AUTHOR CONTRIBUTIONS

CY: method and experiment design and manuscript writing. SL, ZY, and SY: data analysis and interpretation. JG, ZZ, YY, and YG: method investigation and experiment analysis. YK and CL: supervision. All authors contributed to the article and approved the submitted version.

FUNDING

This study was supported by funding from the National Natural Science Foundation of China (grant numbers 81401143), National Key Research and Development Program of China (grant numbers 2018YFC1002900), the Stable Support Plan for Colleges and Universities in Shenzhen, China (grant numbers SZWD2021010), and the National Natural Science Foundation of China (grant numbers 62071311).

ACKNOWLEDGMENTS

The authors thank HC18 challenge in grand-challenge for providing an open source data.

REFERENCES

- Chudleigh T. The 18+020+6 weeks fetal anomaly scan national standards. *Ultrasound*. (2010) 18:92–8. doi: 10.1258/ult.2010.010014
- Sanders RC, James AE. The principles and practice of ultrasonography in obstetrics and gynecology. *JAMA*. (1981) 245:80. doi: 10.1001/jama.1981.03310260058043
- Sarris I, Ioannou C, Chamberlain P, Ohuma E, Roseman F, Hoch L. Intra- and interobserver variability in fetal ultrasound measurements. *Ultrasound Obstet Gynecol*. (2012) 39:266–73. doi: 10.1002/uog.10082
- Nadiyah P, Rofiqah N, Firdaus Q, Sigit R, Yuniarti H. Automatic detection of fetal head using haar cascade and fit ellipse. *International Seminar on Intelligent Technology and Its Applications*. Surabaya: IEEE. (2019).
- Jatmiko W, Habibie I, Ma'sum MA, Rahmatullah R, Satwika PI. Automated telehealth system for fetal growth detection and approximation of ultrasound images. *Int J Smart Sens Intell Syst*. (2015) 8:697–719. doi: 10.21307/ijssis-2017-779
- Namburete AIL, Noble JA. Fetal cranial segmentation in 2D ultrasound images using shape properties of pixel clusters. In: *IEEE International Symposium on Biomedical Imaging* (2013).
- van den Heuvel TLA, de Bruijn D, de Korte CL, Ginneken, BV. Automated measurement of fetal head circumference using 2D ultrasound images. *PLoS ONE*. (2018) 13:e0200412. doi: 10.1371/journal.pone.0200412
- Yaqub M, Kelly B, Papageorghiou AT, Noble JA. Guided random forests for identification of key fetal anatomy and image categorization in ultrasound scans. In: *International Conference on Medical Image Computing and Computer-Assisted Intervention*, Munich. (2015).
- 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 Imag*. (2008) 27:1342–55. doi: 10.1109/TMI.2008.928917
- Irene K, 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 (2019) arXiv:1911.06298.

11. Ponomarev GV, Gelfand MS, Kazanov MD. *A Multilevel Thresholding Combined With Edge Detection and Shape-Based Recognition for Segmentation of Fetal Ultrasound Images*. Proc. Chall. US biometric Meas. from fetal ultrasound images, ISBI (2012).
 12. Rahayu KD, Sigit R, Agata D, Pambudi A, Istiqomah N. Automatic gestational age estimation by femur length using integral projection from fetal ultrasonography. *International Seminar on Application for Technology of Information and Communication*. Semarang: IEEE. (2018).
 13. Stebbing RV, McManigle JE. *A Boundary Fragment Model for Head Segmentation in Fetal Ultrasound*. Proc. Chall. US Biometric Meas. from Fetal Ultrasound Images, ISBI (2012).
 14. Zhang L, Dudley NJ, Lambrou T, Allinson N, Ye X. Automatic image quality assessment and measurement of fetal head in two-dimensional ultrasound image. *J Med Imaging*. (2017) 4:024001. doi: 10.1117/1.JMI.4.2.024001
 15. Hough PVC. *Method and Means for Recognizing Complex Patterns*. U.S. Patent. Washington, DC.; Patent and Trademark Office. (1962).
 16. Prasad DK, Leung MKH, Quek C. ElliFit: an unconstrained, non-iterative, least squares based geometric ellipse fitting method. *Pattern Recognit*. (2013) 46:1449–65. doi: 10.1016/j.patcog.2012.11.007
 17. Altaf F, Islam SM, Akhtar N, Janjua NK. Going deep in medical image analysis: concepts, methods, challenges and future directions. *IEEE Access*. (2019) 7:99540–72. doi: 10.1109/ACCESS.2019.2929365
 18. Girshick R, Donahue J, Darrell T, Malik J. Rich feature hierarchies for accurate object detection and semantic segmentation. *Proc. IEEE Conf. Comput. Vis. Pattern Recognit*. (2014) 580–587. doi: 10.1109/CVPR.2014.81
 19. Sinclair M, Baumgartner CF, Matthew J, Bai WJ, Martinez JC, Li YW, et al. Human-level Performance On Automatic Head Biometrics In Fetal Ultrasound Using Fully Convolutional Neural Networks. *Conference proceedings: Annual International Conference of the IEEE Engineering in Medicine and Biology Society*. (2018) 714–717. doi: 10.1109/EMBC.2018.8512278
 20. Wu L, Xin Y, Li S, Wang T, Heng PA, Ni D. Cascaded fully convolutional networks for automatic prenatal ultrasound image segmentation. In: *International Symposium on Biomedical Imaging (ISBI 2017)*. IEEE, Melbourne. (2017).
 21. Sobhaninia Z, Emami A, Karimi N, Samavi S. Localization of fetal head in ultrasound images by multiscale view and deep neural networks. In: *International Computer Conference, Computer Society of Iran (CSICCI)*. IEEE, Tehran. (2020).
 22. Skeika EL, Da Luz MR, Fernandes BJT, Siqueira HV, De Andrade MLSC. Convolutional neural network to detect and measure fetal skull circumference in ultrasound imaging. *IEEE Access*. (2020) 8:191519–29. doi: 10.1109/ACCESS.2020.3032376
 23. Aji CP, Fatoni MH, Sardjono TA. Automatic measurement of fetal head circumference from 2-dimensional ultrasound. In: *International Conference on Computer Engineering, Network, and Intelligent Multimedia (CENIM)*. IEEE, Surabaya. (2019).
 24. 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: *International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC)*. IEEE, Berlin. (2019).
 25. Redmon J, Divvala S, Girshick R, Farhadi A. You only look once: unified, real-time object detection. In: *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, Las Vegas, NV. (2016).
 26. Liu W, Anguelov D, Erhan D, Szegedy C, Reed S, Fu CY, et al. Ssd: single shot multibox detector. In: *European Conference on Computer Vision*. Springer. Cham. (2016).
 27. Ming Q, Zhou Z, Miao L, Zhang H, Li L. *Dynamic Anchor Learning for Arbitrary-Oriented Object Detection*. New York: AAAI. (2020).
 28. Zhou X, Wang D, Krähenbühl P. Objects as Points. *arXiv*. (2019) arXiv:1904.07850.
 29. Srinivas A, Lin TY, Parmar N, Shlens J, Abbeel P, Vaswani A. Bottleneck transformers for visual recognition. In: *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition* (2021).
 30. Yang TY, Huang YH, Lin YY, Hsiu PC, Chuang YY. SSR-Net: a compact soft stagewise regression network for age estimation. *IJCAI*. (2018) 5:7. doi: 10.24963/ijcai.2018/150
 31. Yang X, Yang X, Yang J, Ming Q, Wang W, Tian Q, et al. *Learning High-Precision Bounding Box for Rotated Object Detection Via Kullback-Leibler Divergence*. *Advances in Neural Information Processing Systems*. NeurIPS. (2021).
 32. Ma'Sum MA, Jatmiko W, Tawakal MI, Al Afif F. Automatic fetal organs detection and approximation in ultrasound image using boosting classifier and hough transform. In: *International Conference on Advanced Computer Science and Information System*. IEEE (2014).
 33. Ronneberger O, Fischer P, Brox T. U-Net: convolutional networks for biomedical image segmentation. In: *International Conference on Medical Image Computing and Computer-Assisted Intervention (MICCAI)*, Munich. (2015).
 34. Xiao X, Lian S, Luo Z, Li S. Weighted res-unet for high-quality retina vessel segmentation. In: *International Conference on Information Technology in Medicine and Education (ITME)*. IEEE, Hangzhou. (2018).
 35. Gu Z, Cheng J, Fu H, Zhou K, Hao H, Zhao Y, et al. CE-Net: context encoder network for 2D medical image segmentation. *IEEE Trans Med Imag*. (2019) 39:2281–92. doi: 10.1109/TMI.2019.2903562
 36. Qiao D, Zulkernine F. Dilated squeeze-and-excitation U-net for fetal ultrasound image segmentation. In: *IEEE Conference on Computational Intelligence in Bioinformatics and Computational Biology (CIBCB)*, Chiang Mai. (2020).
- Conflict of Interest:** The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.
- Publisher's Note:** All claims expressed in this article are solely those of the authors and do not necessarily represent those of their affiliated organizations, or those of the publisher, the editors and the reviewers. Any product that may be evaluated in this article, or claim that may be made by its manufacturer, is not guaranteed or endorsed by the publisher.
- Copyright © 2022 Yang, Liao, Yang, Guo, Zhang, Yang, Guo, Yin, Liu and Kang. This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY). The use, distribution or reproduction in other forums is permitted, provided the original author(s) and the copyright owner(s) are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.