Ellipse Guided Multi-task Network for Fetal Head Circumference Measurement

Jinting Wang[a,b,c], Zhiwen Fang[a,b,c,*], Sheng Yao[a,b,c,*], Feng Yang[a,b,c,*]

[a]*School of Biomedical Engineering, Southern Medical University, Guangzhou, 510515, China*
[b]*Guangdong Provincial Key Laboratory of Medical Image Processing, Southern Medical University, Guangzhou, 510515, China*
[c]*Guangdong Province Engineering Laboratory for Medical Imaging and Diagnostic Technology , Southern Medical University, Guangzhou, 510515, China*

Abstract

Precise fetal head circumference measurement by ultrasound imaging is of great significance for prenatal examination. However, missing or blurring boundaries caused by artifacts and noises challenge measurement accuracy. The inconsistency between the segmentation pseudo-label and the ellipse contours also generates measurement errors. To improve the measurement performances of fetal head circumference, in this study, we propose an ellipse-guided multi-task network that measures the fetal head circumference according to detected ellipse boundary pixels. In the proposed network, an region segmentation branch is designed to learn region features of the fetal head, and a feature fusion module is applied to combine region features with boundary features, which contribute to exploring more context information of fetal head and locating boundary pixels in boundary missing or blurring regions. A loss function is also designed in the network to ensure the boundary estimation in an ellipse shape. Experiments are conducted on both the public fetal head circumference measurement dataset HC-18 and a self-built ultrasonic phantom dataset. The experimental results demonstrate that the proposed method achieves excellent performances to compete with other state-of-the-art methods in fetal head circumference measurement, which can handle the conditions of boundary missing or region blurring.

Keywords: Ultrasound imaging, Fetal head circumference measurement, Multi-task learning, Ellipse shape constrain

*Corresponding author.
E-mail address: fzw310@gmail.com (Zhiwen Fang)
yaosheng@smu.edu.cn(ShengYao)
yangf@smu.edu.cn (Feng Yang).

1. Introduction

Ultrasound imaging is widely used for monitoring fetal growth and development during the prenatal examination due to the advantages of non-invasive, non-radioactive, painless, and cost-effective [1]. In prenatal examinations, fetal head circumference (HC) is an important physiological indicator, which is required to be precisely measured for gestational age prediction [2]. Nevertheless, currently, clinical HC measurement based on ultrasound imaging requires well-trained and experienced sonographers, since ultrasound images are not intuitive and HC measurement results are operator-dependent and machine-specific that leading to inter-and intra-observer variability [3] [4]. Therefore, an automatic measurement system of fetal ultrasound images is needed, which not only can reduce the variability but also reduce the workload of clinicians.

One of the commonly used approaches in automatic HC measurement is ellipse estimation [5, 6, 7, 8, 9], as shown in Fig. 1(b). Specifically, deep-learning-based measurement methods [5, 6, 8] have been presented to regress the ellipse parameters directly from the original ultrasound images. However, these methods with a regression branch suffer from high measurement errors in the ellipse parameter estimation since ultrasound imaging contains various artifacts, such as motion blurring, missing boundaries, acoustic shadows, and speckle noises, which may cause boundary missing (the region in the red bounding box in Fig. 1(a)) or blurring (the region in the orange bounding box in Fig. 1(a))of the fetal head. To solve this problem, HC measurement approaches based on the segmentation contour of fetal head area are proposed [7, 9]. With the abundant context information of the segmentation results, these segmentation-based methods can improve the performance of evaluating the parameters of the fetal head region to a certain extent. Nevertheless, the segmentation results are unreliable, because the annotation provided by the clinician is fetal head elliptical contour (shown in Fig. 1(b)) rather than the ground truth of the segmentation of the real fetal head area. To train segmentation models [7, 9], the whole ellipse is treated as the pseudo-label of the segmentation (displayed in Fig. 1(c)). As shown in Fig. 1, the shape of the fetal head region is not a standard el-

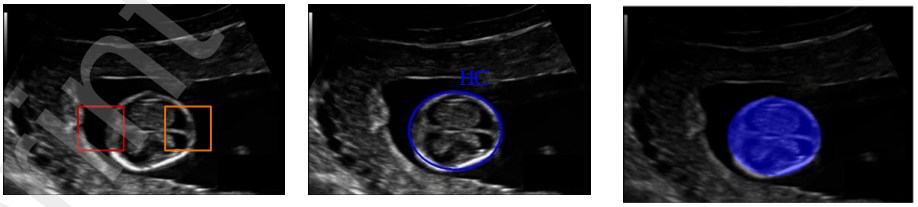

Figure 1: (a) is an original ultrasound image and (b) is an ellipse annotation produced by the sonographer, (c) is a segmentation pseudo-label generated by filling the ellipse annotation showed in (b). The boundary missing region is illustrated in red bounding box and the boundary in orange bounding box is blurring.

lipse, which will lead to the model errors between the fetal head region and the pseudo-label of the segmentation in the training phase. This problem will weaken the performance of segmentation-based measurement methods. Obviously, for the irregular fetal head shape in ultrasound images, it is much easier to detect the fetal head boundary pixels and fit an ellipse that is as consistent as possible with the boundary than to obtain an accurate ellipse segmentation. Since the elliptical contours are provided in this task, we argue that an elliptical boundary, which can mostly cover the fetal head region, would achieve the reliable measurement of fetal head circumference.

In this paper, we propose an ellipse-guided multi-task network to boost the performance of HC measurement in ultrasound images. Specifically, a boundary detection branch is developed to locate fetal head boundary pixels considering that HC measurement is based on the contours composed of these boundary pixels. Nevertheless, the existence of artifacts and maternal uterus which texture structure is similar to fetal head boundary cause missing or blurring boundaries in ultrasound images would weak the accuracy of boundary detection. Therefore, the region features of the head area are extracted to assist boundary detection, which can contribute to locating the boundary of the fetal head and effectively remove false boundary detection. Moreover, to improve the robustness of the boundary detection, a feature fusion module is designed to integrate the boundary features and region features. More useful context information can be extracted from the fused features which improve the capability of boundary detection in low contrast ultrasound images and even eliminates boundary leakages in the boundary missing or blurring regions. As the shape prior of the fetal head is an ellipse, an ellipse-shaped loss function is proposed to constraint the results of boundary detection. Since the annotations are standard ellipses, the ellipse-shaped loss can effectively guide the detected boundary contours close to the elliptical annotations as much as possible. This is particularly important for the improvement of performance for HC measurement. The proposed method is evaluated on the public dataset (HC-18 [3]) and a self-built simulated phantom ultrasound dataset. The experimental results show that the proposed method outperforms the state-of-the-art methods. The main contributions of this work are summarized as follows:

•An ellipse-guided network based on multi-task learning is proposed for fetal head circumference measurement.

•The fetal head boundaries are obtained by the boundary detection branch, without having to resort to segmentation, which reduces errors caused by the inconsistency between the segmentation pseudo-label and the fetal head area.

•The localization information of the fetal head boundaries in the region features from the region segmentation branch and the ellipse information provided by an ellipse-shape loss are applied to improve the measurement performance.

The remainder of this paper is organized as follows: In Section 2, we describe the published fetal head measurement methods based on a single model and multi-task learning. Then the details of our method are illustrated and introduced in Section 3. The experimental results and discussions on HC-18 [3] dataset are conducted in Section 4. The experimental results and discussions on

our self-built dataset are conducted in Section 5. Finally, Section 6 concludes the whole paper.

2. Deep learning-based method for fetal head circumference measurement

Recently, deep learning methods have won great success in HC measurement. The work in HC [9] develops two probabilistic CNN methods: Monte Carlo Dropout during inference and probabilistic UNet. The generated set of segmentation masks is used to reject acquired images that produce sub-optimal HC measurements. In [7], a cascaded FCN is applied for fetal head segmentation. Approaches based on a single segmentation task such as [7, 9] may ignore the potential information in related tasks that can improve measurement performance. As well-known that multi-task learning has been used successfully across all applications of machine learning [10, 11, 12, 13]. The multi-task learning strategy also has been used in HC measurement [6, 8, 14]. [8] combines the classification information of the fetal pregnancy cycle to improve the segmentation performance. For multi-task learning, the higher the correlation between tasks, the greater the promotion effect on the main task [15]. Based on the simplification of fetal head circumference into an ellipse circumference as mentioned in Section 1, ellipse information is much useful to improve the measurement performance. Regressing ellipse parameters from the segmentation features is employed as an auxiliary task in a multi-task neural network [14], which can be trained by minimizing a compound cost function composed of segmentation dice score and mean square error (MSE). The localization information of the fetal head is also considered for HC measurement performance improvement. A combined fetal head localization and fetal head segmentation method based on Mask R-CNN is proposed in [16]. A regression branch is added in an FCNN, and the localization information from the segmentation mask is applied to improve the[6].

Most of these above approaches take segmenting the whole fetal head as a preliminary step for HC measurement. However, some works [17, 18] have been proposed which aim at directly extracting biomarkers from medical images. The goal is to avoid intermediate steps, such as segmentation, that may be computationally expensive and prone to errors. A similar approach is proposed for HC measurement in [5], where a regression CNN is applied to estimate the HC from fetal head ultrasound images, without having to resort to segmentation. Considering that the localization of the fetal head is beneficial to improve HC measurement performance, a region proposal CNN for head localization and centering is used in [19]. The experimental results demonstrate that it can improve the HC delineation performance of the regression CNN.

According to these considerations, in this study, we design a direct approach to estimate the HC by fetal head boundary detection with a multi-task learning network. For improvement of detection performance, the fetal head region features are used as localization information of fetal head boundary. And the el-

lipse information is introduced into the network by an ellipse-shaped loss which guides the detected boundary close to the ellipse shape.

3. Improving ellipse boundary detection with region features fusion for fetal head circumference measurement

Although the fetal head boundaries are incomplete due to artifacts and noise interference in ultrasound images, a good fetal head circumference method should be able to provide an accurate ellipse. It means that the model should not only avoid the interference caused by artifacts and noise, but also obtain the ellipse contours from these incomplete boundary pixels. The proposed method in this work extracts the region features of the fetal head area for localization, and guides the output of the network close to the ellipse shape by an ellipse-shape loss. These make our method has superior performance under the interference of artifacts and noise.

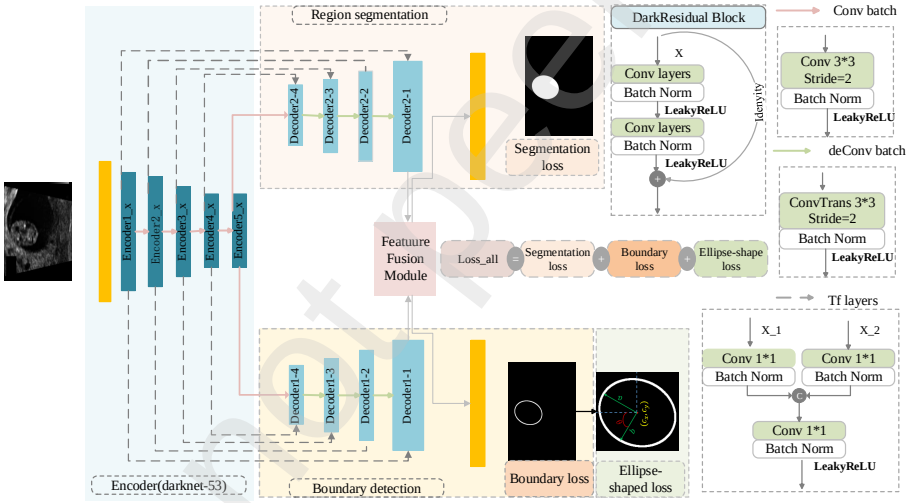

Figure 2: The architecture of our network. The network consists of a Darknet-53 encoder path and two decoder paths. One decoder path is the head region segmentation branch, the other is the ellipse detection branch. The features from the segmentation branch and ellipse detection branch are fed into a feature fusion module, which can help the ellipse detection branch to accurately locate the head boundary. And an ellipse-shaped loss is designed to do shape constraints on boundary detection results.

Fig.2 shows the architecture of our proposed network, the proposed multi-task ellipse-guided network includes a framework for region segmentation and an ellipse detection. Inspired by [20, 21, 22], which show that coarse segmentation results can infer the context information for performance improvement, We introduce segmentation features into the boundary detection branch to dynamically extract the context information contained in the segmentation, which

helps the boundary detection branch to distinguish boundary pixels from background pixels. The region features from the segmentation branch and boundary features from the detection branch are integrated and further selected to explore more context information for boundary detection. The shape prior information is also incorporated into the ellipse detection branch, which can guide the feature extraction in the model. Fit the output of the boundary detection with an ellipse, and an ellipse-shaped loss is designed to supervise the ellipse parameters. From the fitted ellipse, the HC length is computed. More details will be introduced in this section.

3.1. Ellipse detection and region segmentation framework

Generally, the context information is useful for improving the performance in many visual applications [23, 24, 25]. As mentioned above, the information in region segmentation features is also useful for distinguishing boundary pixels and noise. For enhancing the ability of ellipse detection of the fetal head, the context information in fetal head region features is introduced into boundary detection for an accurate boundary. A feature fusion module is used to integrate information between the region segmentation features and the ellipse detection features, which contributes to accurate ellipse detection due to the provision of boundary and region information of the fetal head.

In this study, we modify the Darknet-53 [26] as an encoder path to extract detailed context features, the two decoder paths are the region segmentation branch and the ellipse detection branch, respectively. As shown in Fig. 2, each the i_th Encoder block contains x_i DarknetResidual block (where $x_i \in \{1, 2, 8, 8, 4\}$, and i$\in \{1, 2, 3, 4, 5\}$). We apply Conv batch instead of Maxpooling layer to do downsampling since some significant spatial information is lost by Maxpooling operations [27], which is important for boundary detection. The Conv batch block contains a convolutional layer (kernel size = 3, stride = 2). The convolution is followed by batch normalization and a LeakyRelu with a constant multiplier α, equal to 0.1 to control the slope of the activation function for negative values. Similarly, the Tf layerS presented in Fig. 2 are also applied both on the two decoders to concatenate the correspondingly feature map from the encoder path. Each decoder block contains 1 DarknetResidual Block. A deConv batch block is used to do upsampling.

To alleviate the influence of boundary missing and blurring regions for the ellipse detection, the framework in our method simultaneously learns the region information in the segmentation features and boundary information in the detection features. And the region information can be treated as boundary localization in boundary missing or blurring regions. For exploring more context information of fetal head boundary, a feature fusion module is designed to achieve features integration between region features and boundary features. The details of the feature fusion module will be presented in the next section.

3.2. Feature fusion module

For the information integration between the region and the boundary, a feature fusion module is designed to fuse context features in the two decoding paths.

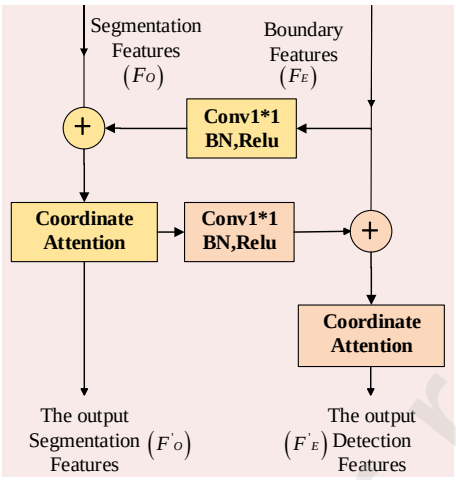

Figure 3: Feature fusion module

The fusion module can assist the ellipse detection branch in accurately locating boundaries according to the region information. The feature fusion module is shown in Fig. 3. The region segmentation features F_O contain rich context information, while the boundary features F_E contain rich shape information. The fusion block can be formulated as follows:

$$F'_E = \alpha(f(F_O) + F_E) + F_E \tag{1}$$

$$F'_O = \alpha(f(F'_E) + F_O) + F_O \tag{2}$$

where F'_E denotes the output detection features and F'_O denotes the output segmentation features, f means 1×1 convolution with BN and Relu. α means coordinate attention module [28]. Some visualization examples are given in Fig. 4. From the comparison, we can see that the processed features by feature fusion module (illustrated as Fig. 4(d) and Fig. 4(i)) contain more accurate boundary information in boundary missing or blurring regions than original features (presented as Fig. 4(b) and Fig. 4(g)), which improves the boundary detection performance, as shown in Fig. 4(c) and Fig. 4(e), Fig. 4(h) and Fig. 4(j). Obviously, the boundary detection capability of the ellipse detection framework is get improved so that weak boundaries in boundary missing or blurring regions can be correctly identified. It means that the feature fusion of the segmentation context information is necessary.

3.3. Ellipse-shaped loss

To keep the boundary close to the fetal head ellipse shape and achieve higher performance, ellipse fitting is performed on the output of the boundary detection branch, and an ellipse-shaped loss is designed (as shown in Fig. 2) considering the shape constraint [29, 30], which can be defined as

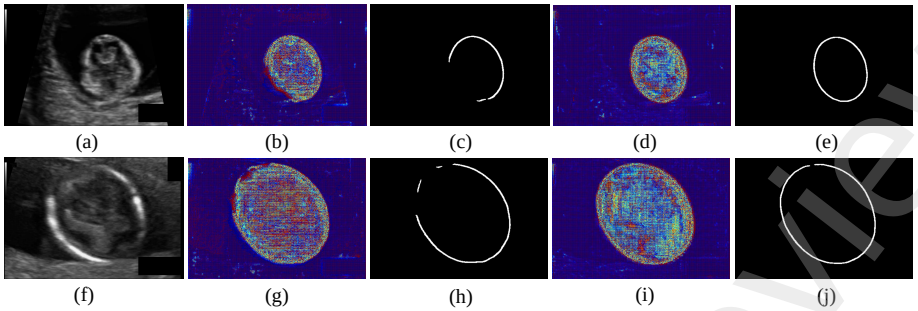

Figure 4: Comparison between with and without feature fusion module. (a) is an ultrasound image with a blurring boundary. (b) is the output feature without the feature fusion module in the model. (c) is the detected head boundary without the feature fusion module. (d) is the output feature with the feature fusion module in the model. (e) is the detected head boundary with the feature fusion module. (f) is an ultrasound image with boundary missing. (g) is the output feature without the feature fusion module in the model. (h) is the detected head boundary without the feature fusion module. (i) is the output feature with the feature fusion module in the model. (j) is the detected head boundary with the feature fusion module. It can be seen that more details about head boundary can be found by the context information from the segmentation branch.

$$L_S = MSE(\hat{y}(a,b,x,y,\theta), y(a,b,x,y,\theta)) \tag{3}$$

where MSE is the Mean Square Error function. a, b, x, y, θ are the ellipse parameters,a and b are major and minor semi-axes of ellipse, x and y are center coordinates of ellipse, θ is rotation angle of ellipse, as shown in Fig. 2. $\hat{y}(a,b,x,y,\theta)$ is the prediction of ellipse fitted by detected boundary pixels, $y(a,b,x,y,\theta)$ is the provided label. The least-square fitting is used to do ellipse fitting on the detected boundary pixels' coordinates. The ellipse-shaped loss L_S can guide the training of our model in an end-to-end fashion.

With the obtained ellipse-shaped loss in Eq. 3, the final loss function of our multi-task network can be defined as:

$$L = L_{dice}(\hat{y_S}, y_S) + L_{dice}(\hat{y_E}, y_E) + L_{MSE}(\hat{y}(a,b,x,y,\theta), y(a,b,x,y,\theta)) \tag{4}$$

where the segmentation loss $L_{dice}(\hat{y_S}, y_S)$, the boundary detection loss $L_{dice}(\hat{y_E}, y_E)$ and the ellipse-shaped loss $L_{MSE}(\hat{y}(a,b,x,y,\theta), y(a,b,x,y,\theta))$. $\hat{y_S}$ is segmentation prediction, y_S is the segmentation pseudo-label. And $\hat{y}_E$ is the boundary detection results, y_E is the fetal head ellipse contour. With the final loss, the the multi-task model can eventually learn the features of fetal head boundary and region under the constraint of ellipse shape.

4. Experiments on HC-18 dataset

4.1. Experimental setup

In experiments, we evaluated the performance of our method on a publicly available dataset HC-18. Table 1 shows the distribution of HC-18 datasets from

the first to the third trimester. Since the ground truth was not available in the original HC-18 test set, we divided 90% annotated images of the HC-18 training set into our training set and the rest images into our new test set. The ablation experiments were studied on the new test set, as listed in Table 2. Annotated images of HC-18 had the resolution of 800×540, with the pixel size from 0.0520 mm to 0.326 mm. In our experiments, all images were padded to 800×800 pixels with zero, then resized to 416×416 pixels. Online data augmentation such as rotation, gray transformation, and horizontal flip was randomly applied at each training iteration. The batch size was set as 8 and the epochs were 200. In the training step, Adam optimizer with an initial learning rate of 0.0001 was used to minimize the multi-task loss in Eq. 4. All experiments were implemented using the deep learning framework Pytorch on a computer with NVIDIA GTX 1080Ti GPUs.

Table 1: The distribution of HC-18 dataset

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

Table 2: The distribution of experimental dataset

Period	Training set	Test set
First trimester	150	15
Second trimester	623	70
Third trimester	125	16
Total	898	101

To evaluate the HC measurement performance on HC-18, we chose three metrics including Dice similarity coefficient (DSC) [%], Absolute Difference (ADF) [mm], and Hausdorff Distance (HD) [mm]. They are defined as:

$$DSC = \frac{2(Y_p \cap Y_g)}{\mid Y_p \mid + \mid Y_g \mid} \tag{5}$$

$$ADF = \mid HC_p - HC_g \mid \tag{6}$$

$$\begin{aligned} HD &= max(h(Y_g, Y_p), h(Y_p, Y_g)) \\ h(Y_p, Y_g) &= max_{Y_g \in Y_p} min_{Y_p \in Y_g} \parallel Y_p - Y_g \parallel \\ h(Y_g, Y_p) &= max_{Y_p \in Y_g} min_{Y_g \in Y_p} \parallel Y_g - Y_p \parallel \end{aligned} \tag{7}$$

where Y_g denotes the fetal head area delimited by the ellipse contours delineated by clinician, Y_p is the fetal head area delimited by the ellipse obtained with our proposed method. HC_p represents the circumference measured from the ellipse detection branch and HC_g is the clinician manual annotation.

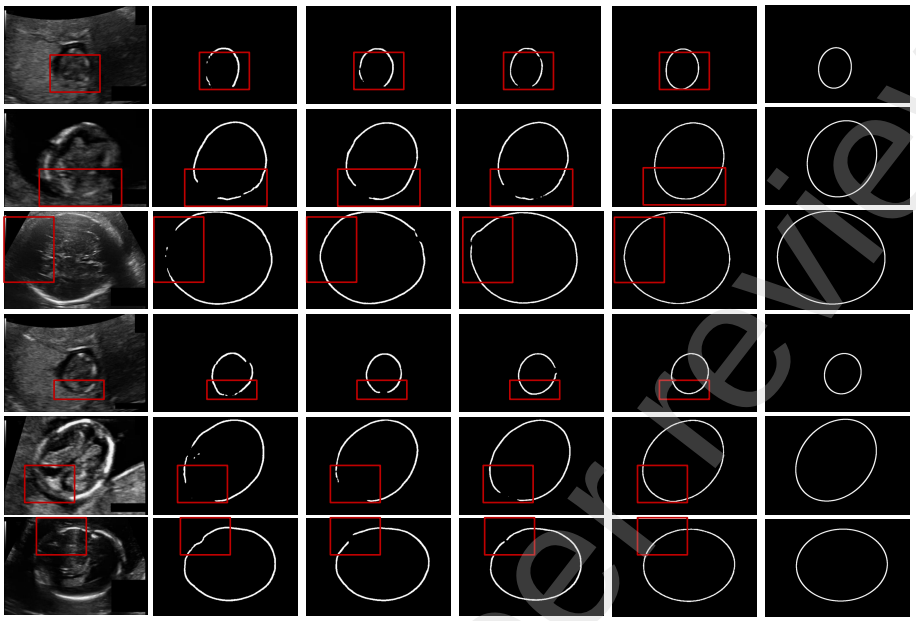

Figure 5: Qualitative evaluation of ablation study on the new test dataset. The first and the fourth row show examples with poor fetal head boundary contrast to the background. The images in the second and the fifth row are examples of fetal head boundary-blurring regions. The third and the sixth row presents examples with boundary missing regions. Each column shows the original ultrasound images (first column), the corresponding ellipse annotations (sixth column), the ellipse detection results obtained by Base (second column), Base+Seg (third column), Base+Seg+ESFF (fourth column), Base+Seg+ESFF+Shape (fifth column), respectively.

4.2. Ablation experiment on the new test set

To show the effectiveness of different components in our model, we present an ablation experiment on the new test set to quantitative and qualitative analysis of the following components in Table 3 and Fig. 5: boundary detection branch (Base), region segmentation branch (Seg), feature fusion module (ESFF) and ellipse-shaped loss (Shape). The lowest mean AD (2.06 mm) is obtained with the Base (boundary detection only), which shows the challenge of boundary missing or blurring regions for boundary detection. However, it can be observed that adding the Seg yields improvements in terms of ADF (0.13 mm). We can see from the second column and the third column in Fig. 5, after adding the region segmentation branch, more accurate boundary pixels can be detected in boundary missing or blurring regions noted in the red bounding box. This achievement can be explained by learning region features which encourage the network to explore more information to distinguish the fetal head area from the background. Benefiting from the ESFF, the performance of DSC, HD, and ADF in the Table 3 has been further improved, increasing by 0.25%, 0.08 mm, and 0.06 mm respectively. As shown in the fourth column in the figure, the wrong

boundary pixels are corrected, and the detected boundaries of the fetal head are more complete. This means that the designed feature fusion module can select better context information in the boundary features and region features, thereby facilitating the boundary detection branch to locate the head boundary. With the Shape added in ellipse detection branch, the detected boundary is closer to the elliptical shape, and the detected boundary is more complete. And according to the Table 3, DSC value, HD value and ADF value increased by 0.03%, 0.04 mm, 0.05 mm respectively. Obviously, the ellipse-shaped loss does well in shape constraint and guiding the detected boundary contours much closer to the ellipses covering the fetal head area, even for examples of irregular head shape such as the second row and the fourth row in Fig. 5.

Table 3: Ablation results on the new test dataset. Base is the Darknet-53 encoder with the ellipse detection branch network. Seg denotes the fetal hand segmentation branch. ESFF means the feature fusion module between the segmentation branch and the ellipse detection branch. Shape means the ellipse-shaped loss.

Methods	DSC(%)	HD(mm)	ADF(mm)
Base	98.05±1.74	1.18±0.74	2.06±2.03
Base + Seg	97.78±1.36	1.21±1.11	1.93±1.41
Base + Seg + ESFF	98.03±1.07	1.13±0.68	1.87±1.59
Base + Seg + ESFF + Shape	**98.06±1.37**	**1.09±0.63**	**1.82±1.35**

4.3. Ablation experiment on HC-18 test set

Combined with our best settings in the deep neural network, we experimented with our proposed HC measurement method on the HC-18 test set. It can be seen from the measurement results listed in the Table 4 that all the components of our model have improved the performance on the HC-18 test set. Compared with Base, Base + Seg shows better performance with DSC and ADF increased by about 0.23% and 0.11 mm, respectively. The ESFF has further improved performance, with DSC, HD, ADF increased by 0.16%, 0.13 mm, 0.09 mm, respectively. This reflects that the context information contained in the region features is beneficial to the boundary detection in boundary missing or blurring regions. Under the supervision of Shape, DSC, AD, HD have been improved by 0.07%, 0.01 mm, 0.06 mm, respectively. This achievement can be explained by limiting the detected boundary pixels to the contour of the ellipse as much as possible by the ellipse loss function.

4.4. Comparison with previous studies on HC-18 test set

Performance comparisons are conducted with different state-of-the-art methods, where the results are listed in Table 5. All these methods were developed and tested using the HC-18 [3] dataset. The relative performance metrics, exhibited in Table 5, are extracted from the corresponding published papers. The fetal head measurement method based on a single model [31] obtained the worst performance. Similarly, our proposed method and the works [14], [6] all employ

Table 4: Ablation studies on the HC-18 test dataset. The bold number indicates the best performance in each column.

Methods	DSC(%)	HD(mm)	ADF(mm)
Base	97.97 ± 1.09	1.25 ± 0.71	2.11 ± 1.97
Base + Seg	97.74 ± 1.24	1.36 ± 1.94	2.00 ± 1.94
Base + Seg + ESFF	97.90 ± 1.26	1.23 ± 0.69	1.91 ± 1.79
Base + Seg + ESFF + Shape	**97.97 ± 1.15**	**1.22 ± 0.77**	**1.85 ± 1.96**

Table 5: Comparison with other methods on HC-18 test dataset. The bold number indicates the best performance in each column.

Methods	DSC(%)	HD(mm)	ADF(mm)
[31]	95.49 ± 4.11	2.44 ± 1.96	2.45 ± 2.55
[14]	96.84 ± 2.89	1.72 ± 1.39	2.12 ± 1.87
[3]	97.00 ± 2.80	2.00 ± 1.60	2.80 ± 3.30
[19]	97.76 ± 1.32	1.32 ± 0.73	1.90 ± 1.77
Ours	**97.97 ±1.15**	**1.22 ±0.77**	**1.85 ±1.96**

multi-task learning strategy, and the significant improvements are obtained than the results in [31] with DSC improved by 1.35%, 1.27%, 2.48%, HD increased by 1.28 mm, 1.11 mm, 1.78 mm and ADF increased by 0.33 mm, 0.55 mm, 0.60 mm, respectively. It demonstrates that more information about the fetal head can be excavated by applying the multi-task strategy, which contributes to improving the performance of the model. As listed in Table 5, the results of our proposed method and [6] are both better than [14] in terms of DSC improved by 1.13% and 0.92%, HD improved by 0.5 mm and 0.4 mm, and ADF improved by 0.27 mm and 0.12 mm. Since direct approaches are designed to estimate the HC from ultrasound images in our two work, without having to resort to segmentation like [14]. Apparently, the inconsistency between the segmentation pseudo-label and the ellipse annotation worse the accuracy of segmentation results as mentioned in Section 1. The proposed approach achieves the best performance listed in Table 5, and the results of our method get improvement than [6] by 0.21% (DSC), 0.1 mm (HD), 0.05 mm (ADF). Comparison with the localization information in the bounding box of the fetal head is obtained by a bounding box regression CNN, the boundary localization information learned from ellipse region features of the fetal head in our method improves the performance of HC measurement more effectively.

5. Experiments on our self-built dataset

5.1. Experimental setup

As mentioned, the main idea of our method is doing the measurement on the detected boundary pixels which is helpful to reduce errors caused by the

inconsistency between the segmentation pseudo-label and the ellipse annotations. And the boundary pixels in missing or blurring boundary regions can be more easily localized by the proposed method. To validate the effectiveness and robustness of the proposed method, experiments were conducted on a self-built dataset.

We used hand-held ultrasound equipment to collect phantom data which is regarded as the simulation of fetal head ultrasound images. The boundary in the ultrasound phantom image was randomly occluded to simulate boundary-missing cases (For each image, we used cut image blocks to randomly occlude boundary of $\frac{1}{10}$ circumference). And measuring the circumference of the circular area in the ultrasound image was used to simulate the HC measurement. Here we designed four groups of experiments on our self-built dataset to verify the effectiveness and robustness of our method for cases with boundary missing or blurring cases: ultrasound images with 5 MHz center frequency (data 1, high-resolution images with clear boundaries), ultrasound images with 3.5 MHz frequency center frequency (data 2, low-resolution images with blurring boundaries), ultrasound images with 5 MHz center frequency and boundary occlusion (data 3), ultrasound images with 3.5 MHz and boundary occlusion (data 4). Data 1 and data 2 were conducted as a comparative experiment to study the boundary detection ability of our method with different blurriness images. Data 3 and data 4 were used to study the measurement accuracy of our method with the boundary-missing images. We compared the experiment results by UNet++ [32] and Zaral et al. [14], one is a measurement method based on a single segmentation model, the other is a measurement method based on multi-task learning. The experimental parameters and software and hardware environment were the same as those experiments in Section 4, except the epochs were set 30 for our self-built dataset.

5.2. Ultrasonic data acquisition and processing

Table 6: Equipment parameters for ultrasound image acquisition.

Parameters	Machine Index	Gain	Dynamic Range	Center Frequency
data1	0.7	80dB	80dB	5MHz
data2	0.6	88dB	90dB	3.5MHz

The handheld ultrasound equipment we used was Sonoster UProbe (a digital ultrasonic imaging diagnostic instrument with 128 arrays convex probe shown in Fig. 6(a)) and the ultrasound phantom is CIRS 040GSE(Fig. 6(b)). And ultrasonic images with different resolutions were obtained with the center frequencies of 5 MHz and 3.5 MHz respectively (an example as Fig. 6(c), the equipment parameters of ultrasonic image acquisition are listed in Table 6.) Placed the probe directly on the surface of the phantom and changed the parameters of the probe to obtain ultrasound images of the phantom with different resolutions. Images were stored in MP4 format, and each MP4 file contained 100 images.

After obtaining the phantom ultrasound image, the image was clipped according to the position of the circular area , and the clipped sizes were 395×278 pixels to 222×166 pixels. The pixel size of these imgaes is 0.0036mm. Then we labeled the clipped image with LabelMe (an online open annotation tool). The data was augmented by rotation(90°, 180°, 270°) and gamma transformation (gamma factor $\gamma \in \{0.6, 0.7, 0.8, 0.9, 1.1, 1.2, 1.3, 1.4\}$). The private simulated phantom ultrasound dataset included 700 images for training and 100 images for testing. The size of images in this dataset were from 395×278 pixels to 222×166 pixels, resized to 256×256 pixels with zero padding.

(a) (b) (c)

Figure 6: The equipment of our experiments. (a) is the handheld ultrasound equipment,(b) is the ultrasound phantom, (c) is an ultrasound image with 5MHz center frequency of the probe.

5.3. Results on our self-built dataset

We now compare our method with two published studies [14], [32] on our self-built dataset. For the fairness of the results, the same experimental parameters and experimental data were used for these three methods. With results shown in Fig. 7, it can be observed that: our proposed method consistently outperforms these two methods on our self-built dataset in terms of DSC, HD, and ADF metrics for all four experiments. For ultrasound images with 5 MHz or 3.5 MHz center frequency, which means these images with different blurriness boundaries, the results by UNet++ [32] drops 0.0003 mm, the results by [14] drops 0.0007 mm, and the results of our method drops 0.0001 mm. Our method achieves the least accuracy reduction compared with UNet++ [32] and [14] in ADF values. These results further validate the effectiveness of our proposed method for boosting the measurement performance in boundary-blurring cases. Our proposed method and the work [14] both apply the multi-task learning strategy, but the results of our method are better. Obviously, the intermediate steps such as segmentation in HC measurement are prone to errors.

For ultrasound images with boundary occlusion, which means these images with boundary missing regions. The results show that our proposed method achieves the best performance in terms of DSC, HD, and ADF metrics. The measurement method based on UNet++ [32] is most affected and the accuracy of the three metrics decreased sharply. [14] applies a multi-task strategy to improve the accuracy of boundary-blurring cases, but their method does not

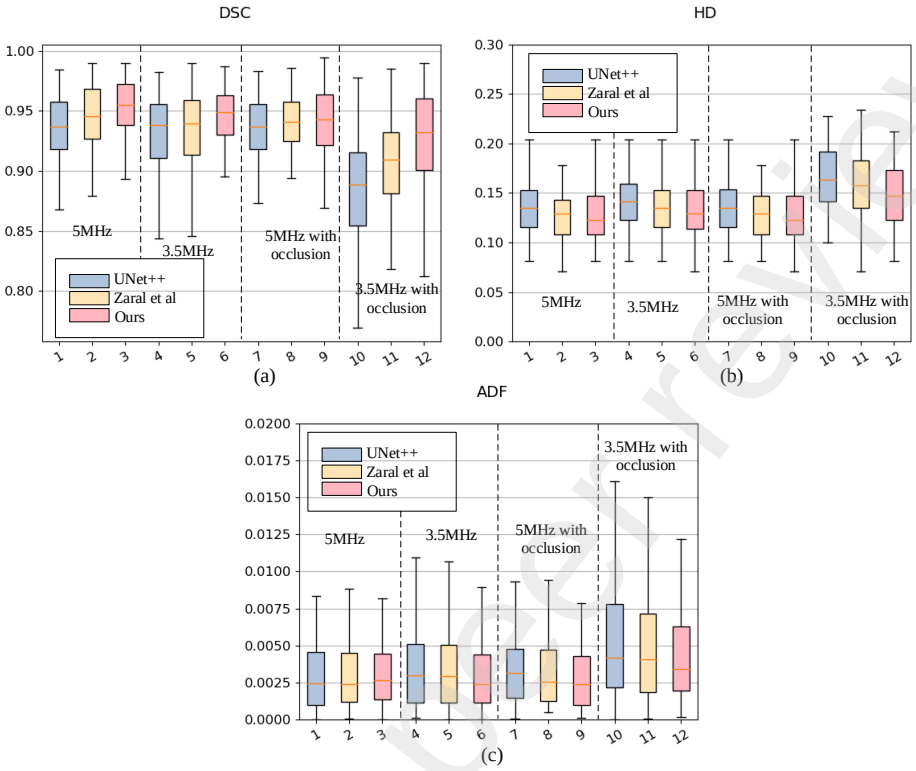

Figure 7: Results on our self-built dataset. (a)-(c) are DSC(%), HD(mm), ADF(mm) respectively. We compared our method with other two (UNet++ [32], Zaral et al. [14]) on the four experiments (5 MHz denotes ultrasound images with 5 MHz center frequency (data 1), 3.5 MHz denotes ultrasound images with 3.5 MHz center frequency (data 2), 5 MHz with occlusion denotes ultrasound images with 5 MHz center frequency and boundary occlusion (data 3), 3.5 MHz with occlusion denotes ultrasound images with 3.5 MHz center frequency and boundary occlusion (data 4).

work in boundary-missing cases, with reduction by 0.6% (DSC), 0.0039 mm (HD), 0.0054 mm (ADF) on data 3, and reduction by 4.38% (DSC), 0.0276 mm (HD), 0.0035 mm (ADF) on data 4, respectively. They can not effectively guide the segmentation in these boundary-missing pixels by regressing elliptic parameters on the segmentation features. Differently, we add ellipse-shaped loss on boundary detection results, which avoids the difficulty of network training caused by the fluctuation of shape constraints, and can guide the prediction results closer to the ellipse more effectively. The results of the proposed method slightly drop 0.1% (DSC), 0.0006mm (HD), 0.0054mm (ADF) on data 3, and 0.19% (DSC), 0.0014 mm (HD), 0.0016 mm (ADF) on data 4, respectively.

The results presented above show that the proposed method achieves the most superior performance in terms of ADF and its average value and variance for the designed four experiments. That convincingly demonstrates the effective-

ness and robustness of our proposed method in handling cases with boundary missing or blurring regions for fetal head circumference measurement.

6. CONCLUSION

In this work, we propose an ellipse-guided network for fetal head circumference measurement in ultrasound imaging, which focuses on pixel prediction in boundary missing or blurring regions. The core idea is to introduce region features by the feature fusion module and detect the fetal head boundary under the ellipse shape supervision for improving the measurement performances. Both the public dataset HC-18 and a self-built dataset are used to validate the effectiveness of our proposed method. The experimental results show that the proposed method achieves DSC 97.97%, HD 1.22 mm, and ADF 1.85 mm, which are better than the performances of the other state-of-the-art approaches. It is also worth mentioning that this work is among the first attempts to directly detect the fetal head boundaries for fetal head circumference measurement, without resorting to segmentation edges, which provides great support to sonographers in the clinical practice of prenatal examination.

Acknowledgments

This work was supported in part by the National Science Foundation of China (NSFC) (61771233).

References

[1] J. Jang, Y. Park, B. Kim, S. M. Lee, J.-Y. Kwon, J. K. Seo, Automatic estimation of fetal abdominal circumference from ultrasound images, IEEE journal of biomedical and health informatics 22 (5) (2017) 1512–1520.

[2] U. Schmidt, D. Temerinac, K. Bildstein, B. Tuschy, J. Mayer, M. Sütterlin, J. Siemer, S. Kehl, Finding the most accurate method to measure head circumference for fetal weight estimation, European Journal of Obstetrics & Gynecology and Reproductive Biology 178 (2014) 153–156.

[3] T. L. van den Heuvel, D. de Bruijn, C. L. de Korte, B. v. Ginneken, Automated measurement of fetal head circumference using 2d ultrasound images, PloS one 13 (8) (2018) e0200412.

[4] I. Sarris, C. Ioannou, P. Chamberlain, E. Ohuma, F. Roseman, L. Hoch, D. Altman, A. Papageorghiou, I. Fetal, N. G. C. for the 21st Century (INTERGROWTH-21st), Intra-and interobserver variability in fetal ultrasound measurements, Ultrasound in obstetrics & gynecology 39 (3) (2012) 266–273.

[5] J. Zhang, C. Petitjean, P. Lopez, S. Ainouz, Direct estimation of fetal head circumference from ultrasound images based on regression cnn, in: Medical Imaging with Deep Learning, PMLR, 2020, pp. 914–922.

[6] P. Li, H. Zhao, P. Liu, F. Cao, Automated measurement network for accurate segmentation and parameter modification in fetal head ultrasound images, Medical & Biological Engineering & Computing 58 (11) (2020) 2879–2892.

[7] L. Wu, Y. Xin, S. Li, T. Wang, P.-A. Heng, D. Ni, Cascaded fully convolutional networks for automatic prenatal ultrasound image segmentation, in: 2017 IEEE 14th international symposium on biomedical imaging (ISBI 2017), IEEE, 2017, pp. 663–666.

[8] P. Liu, H. Zhao, P. Li, F. Cao, Automated classification and measurement of fetal ultrasound images with attention feature pyramid network, in: Second Target Recognition and Artificial Intelligence Summit Forum, Vol. 11427, International Society for Optics and Photonics, 2020, p. 114272R.

[9] S. Budd, M. Sinclair, B. Khanal, J. Matthew, D. Lloyd, A. Gomez, N. Toussaint, E. C. Robinson, B. Kainz, Confident head circumference measurement from ultrasound with real-time feedback for sonographers, in: International Conference on Medical Image Computing and Computer-Assisted Intervention, Springer, 2019, pp. 683–691.

[10] X. Li, Y. Guo, F. Jiang, L. Xu, F. Shen, Z. Jin, Y. Wang, Multi-task refined boundary-supervision u-net (mrbsu-net) for gastrointestinal stromal tumor segmentation in endoscopic ultrasound (eus) images, IEEE Access 8 (2020) 5805–5816.

[11] Y. Ruan, D. Li, H. Marshall, T. Miao, T. Cossetto, I. Chan, O. Daher, F. Accorsi, A. Goela, S. Li, Mb-fsgan: Joint segmentation and quantification of kidney tumor on ct by the multi-branch feature sharing generative adversarial network, Medical image analysis 64 (2020) 101721.

[12] E. Z. Chen, X. Dong, X. Li, H. Jiang, R. Rong, J. Wu, Lesion attributes segmentation for melanoma detection with multi-task u-net, in: 2019 IEEE 16th International Symposium on Biomedical Imaging (ISBI 2019), IEEE, 2019, pp. 485–488.

[13] X. Liu, H. Wang, Advnet: Multi-task fusion of object detection and semantic segmentation, in: 2019 Chinese Automation Congress (CAC), IEEE, 2019, pp. 3359–3362.

[14] Z. Sobhaninia, S. Rafiei, A. Emami, N. Karimi, K. Najarian, S. Samavi, S. R. Soroushmehr, Fetal ultrasound image segmentation for measuring biometric parameters using multi-task deep learning, in: 2019 41st annual international conference of the IEEE engineering in medicine and biology society (EMBC), IEEE, 2019, pp. 6545–6548.

[15] S. Ruder, An overview of multi-task learning in deep neural networks, arXiv preprint arXiv:1706.05098 (2017).

[16] B. Al-Bander, T. Alzahrani, S. Alzahrani, B. M. Williams, Y. Zheng, Improving fetal head contour detection by object localisation with deep learning, in: Annual Conference on Medical Image Understanding and Analysis, Springer, 2019, pp. 142–150.

[17] X. Zhen, S. Li, Towards direct medical image analysis without segmentation, arXiv preprint arXiv:1510.06375 (2015).

[18] X. Zhen, Z. Wang, A. Islam, M. Bhaduri, I. Chan, S. Li, Direct volume estimation without segmentation, in: Medical Imaging 2015: Image Processing, Vol. 9413, International Society for Optics and Photonics, 2015, p. 94132G.

[19] M. C. Fiorentino, S. Moccia, M. Capparuccini, S. Giamberini, E. Frontoni, A regression framework to head-circumference delineation from us fetal images, Computer Methods and Programs in Biomedicine 198 (2021) 105771.

[20] L. Yu, H. Chen, Q. Dou, J. Qin, P.-A. Heng, Automated melanoma recognition in dermoscopy images via very deep residual networks, IEEE transactions on medical imaging 36 (4) (2016) 994–1004.

[21] I. G. Díaz, Incorporating the knowledge of dermatologists to convolutional neural networks for the diagnosis of skin lesions, arXiv preprint arXiv:1703.01976 (2017).

[22] Y. Xie, J. Zhang, Y. Xia, C. Shen, A mutual bootstrapping model for automated skin lesion segmentation and classification, IEEE transactions on medical imaging 39 (7) (2020) 2482–2493.

[23] G. Wang, W. Li, S. Ourselin, T. Vercauteren, Automatic brain tumor segmentation using cascaded anisotropic convolutional neural networks, in: International MICCAI brainlesion workshop, Springer, 2017, pp. 178–190.

[24] P. Mlynarski, H. Delingette, A. Criminisi, N. Ayache, 3d convolutional neural networks for tumor segmentation using long-range 2d context, Computerized Medical Imaging and Graphics 73 (2019) 60–72.

[25] D. Guo, L. Wang, T. Song, G. Wang, Cascaded global context convolutional neural network for brain tumor segmentation, in: International MICCAI Brainlesion Workshop, Springer, 2019, pp. 315–326.

[26] J. Redmon, A. Farhadi, Yolov3: An incremental improvement, arXiv preprint arXiv:1804.02767 (2018).

[27] O. Ronneberger, P. Fischer, T. Brox, U-net: Convolutional networks for biomedical image segmentation, in: International Conference on Medical image computing and computer-assisted intervention, Springer, 2015, pp. 234–241.

[28] Q. Hou, D. Zhou, J. Feng, Coordinate attention for efficient mobile network design, in: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, 2021, pp. 13713–13722.

[29] A. Biswas, P. Bhattacharya, S. Maity, 3d segmentation of liver and its lesions using optimized geometric contours, Procedia computer science 133 (2018) 240–247.

[30] A. Saito, S. Nawano, A. Shimizu, Joint optimization of segmentation and shape prior from level-set-based statistical shape model, and its application to the automated segmentation of abdominal organs, Medical image analysis 28 (2016) 46–65.

[31] Y. Rong, D. Xiang, W. Zhu, F. Shi, E. Gao, Z. Fan, X. Chen, Deriving external forces via convolutional neural networks for biomedical image segmentation, Biomedical optics express 10 (8) (2019) 3800–3814.

[32] Z. Zhou, M. M. R. Siddiquee, N. Tajbakhsh, J. Liang, Unet++: A nested u-net architecture for medical image segmentation, in: Deep learning in medical image analysis and multimodal learning for clinical decision support, Springer, 2018, pp. 3–11.