You Can Detect It: Fetal Biometric Estimation Using Ellipse Detection

Hongyuan Zhang[1,2], Haoyu Xie[2], Tingting Ye[3(✉)], and Songxiong Wu[4(✉)]

[1] CAIR, HKISI-CAS, Hong Kong SAR, China
[2] Medical AI Lab, Shenzhen University Medical School, Shenzhen, China
[3] Department of Ultrasound, Shenzhen University General Hospital and Shenzhen University Clinical Medical Academy, Shenzhen, China
ytty126@163.com
[4] Department of Radiology, Shenzhen University General Hospital and Shenzhen University Clinical Medical Academy, Shenzhen, China
wusongxiong@email.szu.edu.cn

Abstract. The cardiothoracic diameter ratio (CTR) biometric in four-chamber ultrasound plane is often measured for diagnosing congenital heart disease. However, due to the commonly existing artifacts like acoustic shadowing, manual measurement can be time-consuming and labor-intensive task, and may results in high measurements variability. Presently, one of the most popular approaches is segmentation-based methods, which utilize deep learning networks to segment the cardiac and thoracic regions. Then, the metric is calculated through an ellipse fitting scheme. This is inefficient, and requires additional post-processing. To tackle the above problems, in this paper, we therefore present an one-stage ellipse detection network, namely EllipseDet, which detects the cardiac and thoracic regions in ellipse, and then automatically calculates the CTR biometric in four-chamber view. In particular, we formulate the network that detects the center of each object as points and regresses the ellipses' parameters simultaneously. Besides, we propose a novel ellipse feature alignment module and Ellipse-IoU loss to further regulate the regression procedure. We have evaluated EllipseDet on a clinical echocardiogram dataset and the experimental results show that our proposed framework outperforms several state-of-the-art methods. As an open science, source code, images dataset and pre-trained weights are available at https://github.com/szuboy/FOCUS-dataset.

Keywords: Biometric Estimation · Ellipse Detection · Four-chamber Ultrasound View · Cardiothoracic Diameter Ratio

1 Introduction

Fetal congenital heart disease is one of the most common forms of birth defects worldwide, with an approximate incidence rate of 8% among newborn babies [20]. Four-chamber view of the fetal echocardiography is an important yet unique tool for assessing the fetus for the presence of congenital heart disease [5]. In the

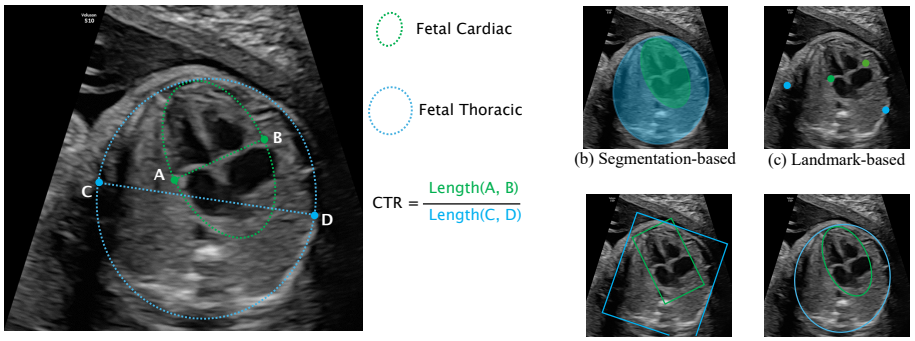


Fig. 1: (a) The definition of the CTR biometric in the four-chamber view and (b-e) the comparisons of object representation in different estimation pipelines.

four-chamber view, assessment and diagnosis are based on the cardiothoracic diameter ratio (CTR) [19] biometric, as shown in Fig. 1(a). Specifically, the CTR assesses heart size by comparing the maximum cardiac diameter to the thoracic diameter. However, due to the commonly existing acoustic shadowing artifact in fetal echocardiogram, the manual measurements is a labor-intensive and time-consuming process, and suffer from large inter-operator variance. Therefore, automatic and reliable fetal cardiac measurements are in high demand.

Recently, many deep learning methods have been proposed for fetal biometrics estimation [18,3,1], but they still cannot achieve satisfactory results due to the low quality of four-chamber view. Existing deep learning-based methods can be classified into three categories: segmentation-based methods, landmark-based methods and box-based methods, as shown in Fig. 1. Segmentation-based methods [15,18,12] are currently the most common approach, utilizing segmentation results as an intermediate processing step. However, these approaches make biometrics estimation heavily dependent on these results, leading to error accumulation across both stages [26]. Landmark-based methods [1], on the other hand, detect the landmarks directly on the image without relying on segmentation masks. Unfortunately, ambiguous boundaries caused by massive speckle noise and artifacts significantly hinder the effectiveness of landmark-based approaches. Besides, rotation object detection methods [3,23] can also be used as candidate solutions. However, traditional bounding box may perform poorly in representing such ball-shaped biomedical objects [16]. Thus, how to design a suitable method to represent the cardiac and thoracic regions is the key procedure for achieving precise biometric estimation.

In this paper, we present a one-stage detection framework, called EllipseDet, for fetal cardiac biometrics measurement in four-chamber scans. Unlike previous paradigms [15,18,1], our proposed method uses ellipses to represent the cardiac and thoracic regions and directly regresses the ellipses's parameters in one-stage without requiring additional post-processing, as shown in Fig. 1(e). We also in-

troduce a novel feature alignment module and loss function tailored for detecting ellipse-shaped fetal cardiac and thoracic regions, making it suitable for practical medical object detection. **Overall, our contributions are listed as follows:**

- **Ellipse Representation:** We propose a simple ellipse representation for fetal biometrics estimation that better depicts fetal structures. In addition, we introduce a new ellipse feature alignment module and ellipse intersection over union (Ellipse-IoU) loss to further regulate the training procedure.
- **Optimized Biomedical Object Detection:** To the best of our knowledge, our EllipseDet is the first one-stage approach with ellipse representation and optimized Ellipse-IoU loss for fetal biometric estimation.
- **Superior Detection and Clinical Benefits:** The extensive experimental results demonstrate that our proposed method, EllipseDet, achieves superior detection performance and offers greater clinical benefits compared to other state-of-the-art techniques.
- **Public Fetal Cardiac Dataset:** We contribute the first publicly available fetal cardiac biometrics dataset, providing a foundation for future research.

2 EllipseDet for Fetal Biometrics Estimation

2.1 Architecture Overview

The overview of our EllipseDet is shown in Fig. 2. EllipseDet is a simple, unified network consisting of a backbone network [9], a feature pyramid network [13], and three task-specific heads. The backbone and feature pyramid network are designed to generate multi-scale feature maps for the subsequent head networks. And, the heatmap and offset heads are used to determine the central location, while the the ellipse regression head determines the parameters of the ellipse.

Ellipse Representation. As shown in Fig. 1(e), given an ellipse object, we represent it as $E_i = (x_i, y_i, a_i, b_i, \theta_i)$ for the i-th object with class c_i, where (x_i, y_i) denotes the center point, a_i and b_i are the lengths of the semi-major and semi-minor axes, respectively, and θ_i corresponds to the tilt angle. The parameters $x_i, y_i, a_i, b_i, \theta_i \in \mathcal{R}^1$, where $\theta_i \in [-\frac{\pi}{2}, \frac{\pi}{2})$. Thus, we model the object as one center and three parameters.

Backbone Network. Following prior works [25,3], the ResNet50 [9] with Feature Pyramid Network (FPN) is adopted as the backbone network to produce multi-scale feature maps from input ultrasound image. We construct the pyramid with levels P_2 through P_4, where l indicates pyramid level and P_l has resolution 2^l lower than the input [13]. And, all pyramid levels have 256 channels.

Regression Head. As illustrated in Fig. 2, an ellipse object is decoupled into location and size, corresponding to the two branches in the regression head. Each branch has four convolutional layers and interacts with each other through a feature alignment module. However, since the output feature size is smaller than that of the input image, we use heatmap-offset aggregation [28,3] to correct the discretization error from the output stride. Finally, there are three outputs

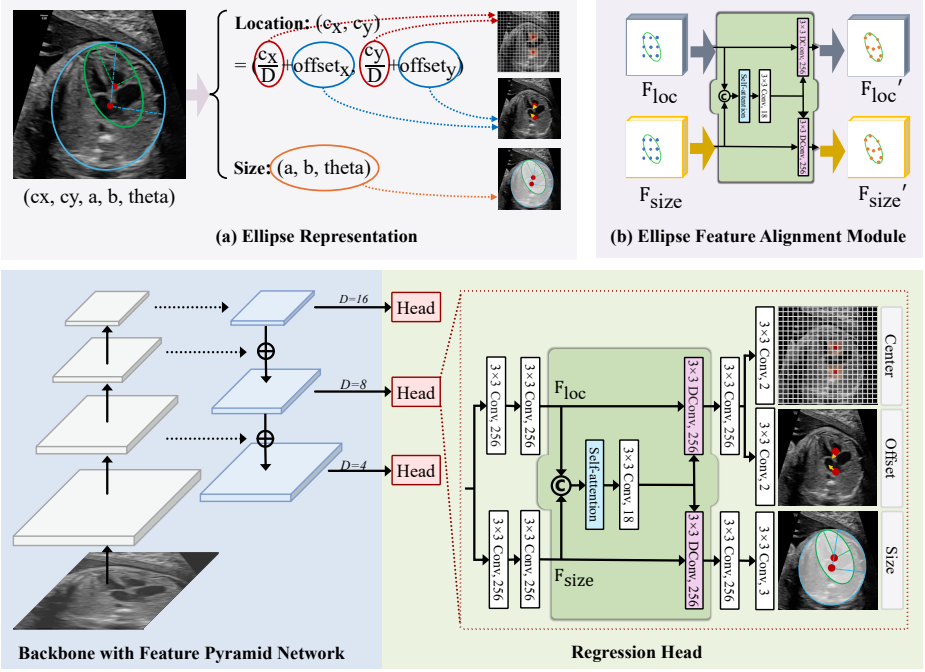


Fig. 2: The network structure of EllipseDet and the explanation of the modules: (a) Ellipse representation for thoracic and cardiac instances, (b) Visualization of the alignment between thoracic and cardiac features using the proposed ellipse feature alignment module.

in each head. Specifically, we predict the heatmap $\hat{Y} \in \mathcal{R}^{\frac{W}{D} \times \frac{H}{D} \times C}$, local offset $\hat{O} \in \mathcal{R}^{\frac{W}{D} \times \frac{H}{D} \times 2}$, and a size $\hat{S} \in \mathcal{R}^{\frac{W}{D} \times \frac{H}{D} \times 3}$, where D is the output stride, representing the down-sampling factor of the prediction, C is the number of object categories, and the three channels in size prediction represent the semi-major axis, semi-minor axis and angle, respectively. Following standard practice [10,28], the ground truth of the heatmap is modeled as a 2D Gaussian kernel:

$$Y = \exp\left(-\frac{(x-\tilde{x}_i)^2 + (y-\tilde{y}_i)^2}{2\sigma_p^2}\right) \tag{1}$$

where the $\tilde{x}_i = \lfloor \frac{x_i}{D} \rfloor$ and $\tilde{y}_i = \lfloor \frac{y_i}{D} \rfloor$ are the down-sampled target center point location and σ_p is the size-adaptive standard deviation [10]. The predicted heatmap is optimized using pixel-wise regression loss $\mathcal{L}_{heatmap}$ with focal loss [14]. Besides, the offset is calculated as $o = (\frac{x_i}{D} - \tilde{x}_i, \frac{y_i}{D} - \tilde{y}_i)$. The offset loss function $\mathcal{L}_{offset}$ is optimized with $\mathcal{L}_1$ loss. However, the training of the size regression branch is non-trivial. For an ellipse instance, its center and size are relevant and should be trained as a whole, rather than as independent regression tasks. To address this challenge, we introduce the Ellipse-IoU Loss, which is further discussed in the Section 2.3.

Ellipse Assembling and Biometric Estimation. During inference, we only assemble ellipse from predictions with the highest confidence scores across all FPN levels without any post-processing. Specifically, we first extract the peak in the heatmap for each category independently, and its location $(\hat{x}_i, \hat{y}_i)$ is taken as the center position of the cardiac or thoracic object. Meanwhile, the offset $(\delta x_i, \delta y_i)$ is obtained from $\hat{O}$. Then, the bounding ellipse is formed with center point $\hat{p}$ and size $\hat{s}$ as follows:

$$\hat{p} = ((\hat{x}_i + \delta x_i) * D, (\hat{y}_i + \delta y_i) * D) \qquad \hat{s} = \hat{S}_{\hat{x}_i, \hat{y}_i} = (\hat{a}_i, \hat{b}_i, \hat{\theta}_i) \tag{2}$$

Once we obtain the cardiac and thoracic parameters, we can calculate and estimate the CTR biometric as: $\text{CTR} = \hat{a}_{cardiac}/\hat{a}_{thoracic}$.

2.2 Ellipse Feature Alignment Module

Since conventional networks for rotated object detection have limited generalization to rotation variations, better feature alignment generally means better classification and regression in object detection [11,24]. Thus, as shown in Fig. 2, we introduce a simple yet effective ellipse feature alignment module (EFAM) in the regression head, enhancing the collaboration between location and size features during detection. It reintegrates the decoupled location and size features to provide shared offsets, which keeps task-specific features and mitigates inconsistent predictions. The detailed architecture of the module is presented in Fig. 2(b). We first use and concatenate the features $F_{\text{loc}} \in \mathcal{R}^{\frac{W}{D} \times \frac{H}{D} \times 256}$ and $F_{\text{size}} \in \mathcal{R}^{\frac{W}{D} \times \frac{H}{D} \times 256}$ of the two branches to form the task-interaction feature $F_{\text{loc-size}} \in \mathcal{R}^{\frac{W}{D} \times \frac{H}{D} \times 512}$, which contains location and size information. Then, we incorporate a self-attention module [21] into the EFAM to further improve ellipse feature learning. This operation allows for dynamic weighting of global rotated ellipse object features. Given the task-interaction feature $F_{\text{loc-size}}$, we compute the attention feature as follows:

$$F^{'}_{\text{loc-size}} = \text{Attention}(F_{\text{loc-size}}) = \text{softmax}\left(\frac{QK^T}{\sqrt{d_k}}\right)V \tag{3}$$

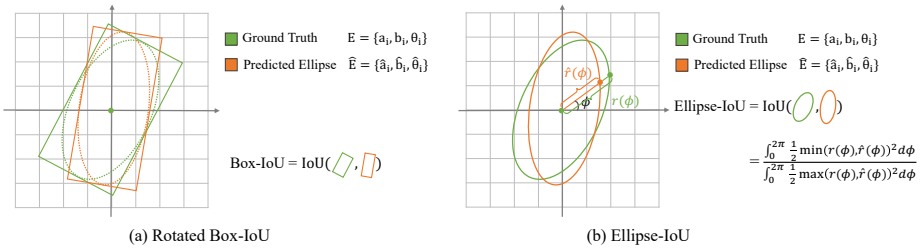
where Q, K, V represent the queries, keys, and values of $F_{\text{loc-size}}$, respectively, and d_k is the dimension of the keys. After obtaining the feature $F^{'}_{\text{loc-size}}$, we apply a 3×3 convolutional layer to output the offsets $F_{\text{offset}} \in \mathcal{R}^{\frac{W}{D} \times \frac{H}{D} \times 18}$. Using F_{offset}, we can adaptively align the features for location and size tasks:

$$F^{'}_{\text{loc}} = f^{\text{loc}}_{\text{dcn}}(F_{\text{loc}}, F_{\text{offset}}) \qquad F^{'}_{\text{size}} = f^{\text{size}}_{\text{dcn}}(F_{\text{size}}, F_{\text{offset}}) \tag{4}$$

Here, $f^{\text{loc}}_{\text{dcn}}(\cdot)$ and $f^{\text{size}}_{\text{dcn}}(\cdot)$ both the standard deformable convolutional operation [6]. Then, we can use the aligned features $F^{'}_{\text{loc}}$ and $F^{'}_{\text{size}}$ for regression.

2.3 Ellipse-IoU Loss

Compared with smooth-l_1 loss, the IoU loss considers the optimization as a whole. However, existing IoU losses [27] are box-based and not tailored for elliptical biomedical objects. In this work, we derive an differential algorithm to compute IoU for ellipse representation and achieve competitive performance.



(a) Rotated Box-IoU (b) Ellipse-IoU

Fig. 3: Comparison of different Rotated Box-IoU [27] and Ellipse-IoU metrics.

As shown in Fig. 3, we first introduce the Ellipse-IoU loss based on the definition of IoU, instead of relying on the compromise method of the bounding rectangle box. This approach is more effective for accurately representing elliptical objects. Based on the ellipse representation [22], the Ellipse-IoU for a given instance is calculated as follows:

$$\text{Ellipse-IoU} = \frac{\int_0^{2\pi} \frac{1}{2}\min(r(\phi), \hat{r}(\phi))^2 d\phi}{\int_0^{2\pi} \frac{1}{2}\max(r(\phi), \hat{r}(\phi))^2 d\phi} \approx \lim_{n\to\infty} \frac{\sum_{i=1}^{N} \frac{1}{2}\min(r(\phi), \hat{r}(\phi))^2 \Delta\phi}{\sum_{i=1}^{N} \frac{1}{2}\max(r(\phi), \hat{r}(\phi))^2 \Delta\phi} \tag{5}$$

where ϕ is the angle with the x-axis, and $r(\phi)$ and $\hat{r}(\phi)$ are the axis lengths corresponding to the angle ϕ.[5] When N approaches infinity, the discrete form converges to the continuous form. We assume that the sampling is uniform, leading to $\phi = \frac{2\pi}{N}$, which further simplifies the expression:

$$\text{Ellipse-IoU} = \frac{\sum_{i=1}^{N} \min(r(\phi), \hat{r}(\phi))^2}{\sum_{i=1}^{N} \max(r(\phi), \hat{r}(\phi))^2} \tag{6}$$

Since the optimal Ellipse-IoU is always 1, the Ellipse-IoU loss can be defined as:

$$\text{Ellipse-IoU Loss} = 1 - \text{Ellipse-IoU} = 1 - \frac{\sum_{i=1}^{N} \min(r(\phi), \hat{r}(\phi))^2}{\sum_{i=1}^{N} \max(r(\phi), \hat{r}(\phi))^2} \tag{7}$$

2.4 Final Training Loss Function

In this study, we do not normalize the scale and instead use the raw pixel coordinates directly, scaling the loss by a constant λ_{size}. The final loss is:

$$\mathcal{L}_{final} = \mathcal{L}_{heatmap} + \lambda_{offset}\mathcal{L}_{offset} + \lambda_{size}\mathcal{L}_{size} \tag{8}$$

where λ_{offset} and λ_{size} weights for the respective loss terms. To balance the Ellipse-IoU loss function and other loss functions, we set $\lambda_{size} = 10$ and $\lambda_{offset} = 1$ in all our experiments unless specified otherwise. In addition, we set the number of samples $N = 36$ for the Ellipse-IoU loss function by default.

[5] For any rotated ellipse instance (x, y, a, b, θ), the corresponding axis length r in ϕ can be expressed as: $r(\phi) = \sqrt{(a\cos(\phi - \theta))^2 + (b\sin(\phi - \theta))^2}$.

Table 1: Experimental results between different methods across different metrics.

Methods	Cardiac Region		Thoracic Region		Biometric
	DSC (%) ↑	HD95 (pixel) ↓	DSC (%) ↑	HD95 (pixel) ↓	P_{CTR}(%) ↑
U-Net[17] [MICCAI2015]	89.12 ± 1.62	25.63 ± 4.24	93.96 ± 0.67	32.74 ± 4.68	90.23 ± 1.37
Rotated Faster RCNN[24] [ICRSA2020]	86.60 ± 0.72	29.79 ± 1.28	91.00 ± 0.94	39.82 ± 4.60	88.30 ± 1.63
EllipseNet[3] [MICCAI2022]	77.87 ± 2.48	48.37 ± 6.89	81.36 ± 3.24	70.46 ± 8.79	45.15 ± 15.12
Oriented RepPoints[11] [CVPR2022]	60.00 ± 3.10	109.96 ± 10.06	69.42 ± 5.25	140.57 ± 27.50	61.94 ± 11.83
BiometryNet[1] [MICCAI2022]	/	/	/	/	35.98 ± 12.61
E2EBM-Net[8] [CBM2023]	/	/	/	/	31.45 ± 39.96
TransUNet[4] [MedIA2024]	92.48 ± 0.36	16.76 ± 0.72	**96.48 ± 0.18**	16.26 ± 0.96	93.32 ± 1.38
EllipseDet(N=36) [OURS]	**93.18 ± 0.24**	**14.47 ± 0.96**	94.87 ± 0.84	**16.19 ± 0.86**	**94.41 ± 1.33**

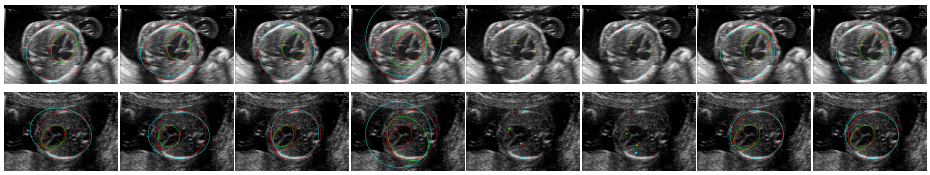
3 Experiment and Results

3.1 Experiment Protocol

Fetal Cardiac Ultrasound Dataset: The dataset comprises 300 images, and is divided into 200 training sets, 50 validation sets and 50 test sets based on the patient-level. The annotation work involved the participation of 2 experts with more than 5-year clinical experience. It includes images of 217 subjects from two medical centers within the public FETAL_PLANE database [2] and incorporates ellipse annotations for both cardiac and thoracic across all cases.
Implementation Details: All the ultrasound images are resized 512×512 to before feeding into the network. Random left-right flipping, scaling, shifting and random Gaussian noise are added for data augmentation. Following previous studies [4], the segmentation model is trained using a combination of loss functions: Dice loss function and binary cross-entropy (BCE) loss function, with a BCE weight of 0.2. We trained all model with batch size 8 and initial learning rate 1.25×10^{-4} for 120 epochs. All experiments were conducted on a Linux Ubuntu 16.04 system, which was equipped with 2 NVIDIA TITAN RTX GPUs.
Evaluation Metrics: All experiments were evaluated using the Dice similarity coefficient (DSC) and 95th percentile Hausdorff distance (HD95). In addition, we presented the measurement precision of the CTR biometric. The CTR is formulated as $R = b_C/b_T$, where b_C represents the length of minor axis of the cardiac ellipse, b_T represents that of the thoracic ellipse [3]. Then, the CTR precision can be defined as: $P_{CTR} = (1 - \frac{|R_{\text{true}} - R_{\text{pred}}|}{R_{\text{true}}}) \times 100\%$, where R_{pred} and R_{true} denote the predicted CTR and the ground truth CTR, respectively.

3.2 Comparisons with State-of-the-arts

To demonstrate the effectiveness of EllipseDet, we conducted experiments across three aspects: 1) Segmentation-based method: U-Net [17] and TransUNet [4]; 2) Landmark-based method: E2EBM-Net [8] and BiometryNet [1]; and 3) Detection-based method: EllipseNet [3], Oriented RepPoints [11] and Rotated Faster RCNN [24]. For U-Net and TransUNet, we employed the ellipse-fit post-processing method [7] to obtain final ellipse prediction results. Landmark-based methods use landmarks to directly compute CTR biometric, while detection-based methods utilize the inscribed ellipse of the rotated bounding box for evaluation metrics.



(a) U-Net (b) Rotated Faster RCNN (c) EllipseNet (d) Oriented RepPoints (e) BiometryNet (f) E2EBM-Net (g) TransUNet (h) **EllipseDet (OURS)**

Fig. 4: Visualizations of different methods are shown, with red indicating true labels, cyan for predicted thoracic labels, and green for predicted cardiac labels.

Table 2: Ablation studies of different components of our method on the dataset.

FPN	EFAM	Box-IoU	Ellipse-IoU		Cardiac Region		Thoracic Region		Biometric
			N=36	N=100	DSC (%) ↑	HD95 (pixel) ↓	DSC (%) ↑	HD95 (pixel) ↓	P_{CTR}(%) ↑
					84.45 ± 1.62	31.11 ± 5.23	86.96 ± 3.78	63.11 ± 8.64	84.53 ± 6.46
✓					89.67 ± 0.50	24.17 ± 3.56	90.18 ± 1.95	24.80 ± 5.02	89.17 ± 1.73
✓	✓				90.19 ± 0.69	23.23 ± 1.47	90.94 ± 1.63	23.36 ± 4.55	92.10 ± 1.82
✓	✓	✓			91.24 ± 0.33	17.35 ± 1.20	92.02 ± 1.27	18.26 ± 2.61	93.32 ± 1.38
✓	✓		✓		93.18 ± 0.24	**14.47 ± 0.96**	94.87 ± 0.84	16.09 ± 0.86	94.41 ± 1.33
✓	✓			✓	**93.46 ± 0.22**	14.55 ± 0.97	**95.08 ± 0.77**	**15.60 ± 0.88**	**95.06 ± 1.45**

Quantitative Results: We listed the detection performance in Table 1. As expected, due to the lack of semantic information, the landmark-based E2EBM-Net [8] and BiometryNet [1] show poor biometric estimation performance ($P_{CTR} < 40\%$). Compared with the U-Net [17] and TransUNet [4] methods, our method does not require additional ellipse fitting post-processing methods and achieves state-of-the-arts results in almost all metrics(DSC $> 90\%, P_{CTR} > 90\%$). Besides, thanks to the optimization of the ellipse objects, EllipseDet also outperforms the other box-based methods in region segmentation and CTR estimation.

Qualitative Results: Fig. 4 shows two cases illustrating the performance of different methods. As we can see, our EllipseDet aligns closely with true labels, effectively capturing thoracic and cardiac regions, while other methods struggle with boundaries. It also demonstrates superior accuracy in challenging scenarios.

3.3 Ablation Studies

Table 2 shows ablation studies on the key components of our method with the fetal cardiac ultrasound dataset. The multi-scale features of the FPN can bring about a 5% performance improvement. Besides, using our proposed EFAM, the performance significantly improves by 3% in CTR estimation. Compared with the smooth-l_1 and Box-IoU loss, proposed Ellipse-IoU loss function brings continuous performance improvements in both segmentation and biometric estimation. Besides, the Ellipse-IoU loss function is not sensitive to the number of samples N. When $N = 36$ and $N = 100$, all metric performances are close (<1%).

4 Conclusion

This paper proposes an one-stage ellipse detection network for the cardiac and thoracic regions detection. In order to better represent the fetal cardiac and thoracic, two innovative modules (EFAM and Ellipse-IoU) are proposed for the optimization of elliptical objects. Extensive experimental results also show the effectiveness of our proposed EllipseDet method.

Disclosure of Interests. The authors have no competing interests to declare that are relevant to the content of this article.

References

1. Avisdris, N., Joskowicz, L., Dromey, B., David, A.L., Peebles, D.M., Stoyanov, D., Ben Bashat, D., Bano, S.: Biometrynet: landmark-based fetal biometry estimation from standard ultrasound planes. In: International Conference on Medical Image Computing and Computer-Assisted Intervention. pp. 279–289. Springer (2022)
2. Burgos-Artizzu, X.P., Coronado-Gutiérrez, D., Valenzuela-Alcaraz, B., Bonet-Carne, E., Eixarch, E., Crispi, F., Gratacós, E.: Evaluation of deep convolutional neural networks for automatic classification of common maternal fetal ultrasound planes. Scientific Reports **10**(1), 10200 (2020)
3. Chen, J., Zhang, Y., Wang, J., Zhou, X., He, Y., Zhang, T.: Ellipsenet: anchor-free ellipse detection for automatic cardiac biometrics in fetal echocardiography. In: Medical Image Computing and Computer Assisted Intervention–MICCAI 2021: 24th International Conference, Strasbourg, France, September 27–October 1, 2021, Proceedings, Part VII 24. pp. 218–227. Springer (2021)
4. Chen, J., Mei, J., Li, X., Lu, Y., Yu, Q., Wei, Q., Luo, X., Xie, Y., Adeli, E., Wang, Y., et al.: Transunet: Rethinking the u-net architecture design for medical image segmentation through the lens of transformers. Medical Image Analysis **97**, 103280 (2024)
5. Copel, J.A., Pilu, G., Green, J., Hobbins, J.C., Kleinman, C.S.: Fetal echocardiographic screening for congenital heart disease: the importance of the four-chamber view. American journal of obstetrics and gynecology **157**(3), 648–655 (1987)
6. Dai, J., Qi, H., Xiong, Y., Li, Y., Zhang, G., Hu, H., Wei, Y.: Deformable convolutional networks. In: Proceedings of the IEEE international conference on computer vision. pp. 764–773 (2017)
7. Fitzgibbon, A., Pilu, M., Fisher, R.B.: Direct least square fitting of ellipses. IEEE Transactions on pattern analysis and machine intelligence **21**(5), 476–480 (1999)
8. Gao, Z., Tian, Z., Pu, B., Li, S., Li, K.: Deep endpoints focusing network under geometric constraints for end-to-end biometric measurement in fetal ultrasound images. Computers in Biology and Medicine **165**, 107399 (2023)
9. He, K., Zhang, X., Ren, S., Sun, J.: Deep residual learning for image recognition. In: Proceedings of the IEEE conference on computer vision and pattern recognition. pp. 770–778 (2016)
10. Law, H., Deng, J.: Cornernet: Detecting objects as paired keypoints. In: Proceedings of the European conference on computer vision (ECCV). pp. 734–750 (2018)
11. Li, W., Chen, Y., Hu, K., Zhu, J.: Oriented reppoints for aerial object detection. In: Proceedings of the IEEE/CVF conference on computer vision and pattern recognition. pp. 1829–1838 (2022)

12. Liang, B., Peng, F., Luo, D., Zeng, Q., Wen, H., Zheng, B., Zou, Z., An, L., Wen, H., Wen, X., et al.: Automatic segmentation of 15 critical anatomical labels and measurements of cardiac axis and cardiothoracic ratio in fetal four chambers using nnu-netv2. BMC Medical Informatics and Decision Making **24**(1), 128 (2024)
13. Lin, T.Y., Dollár, P., Girshick, R., He, K., Hariharan, B., Belongie, S.: Feature pyramid networks for object detection. In: Proceedings of the IEEE conference on computer vision and pattern recognition. pp. 2117–2125 (2017)
14. Lin, T.Y., Goyal, P., Girshick, R., He, K., Dollár, P.: Focal loss for dense object detection. In: Proceedings of the IEEE international conference on computer vision. pp. 2980–2988 (2017)
15. Liu, L., Tang, D., Li, X., Ouyang, Y.: Automatic fetal ultrasound image segmentation of first trimester for measuring biometric parameters based on deep learning. Multimedia Tools and Applications **83**(9), 27283–27304 (2024)
16. Nguyen, E.H., Yang, H., Deng, R., Lu, Y., Zhu, Z., Roland, J.T., Lu, L., Landman, B.A., Fogo, A.B., Huo, Y.: Circle representation for medical object detection. IEEE transactions on medical imaging **41**(3), 746–754 (2021)
17. Ronneberger, O., Fischer, P., Brox, T.: U-net: Convolutional networks for biomedical image segmentation. In: Medical image computing and computer-assisted intervention–MICCAI 2015: 18th international conference, Munich, Germany, October 5-9, 2015, proceedings, part III 18. pp. 234–241. Springer (2015)
18. Sinclair, M., Baumgartner, C.F., Matthew, J., Bai, W., Martinez, J.C., Li, Y., Smith, S., Knight, C.L., Kainz, B., Hajnal, J., et al.: Human-level performance on automatic head biometrics in fetal ultrasound using fully convolutional neural networks. In: 2018 40th annual international conference of the IEEE engineering in medicine and biology society (EMBC). pp. 714–717. IEEE (2018)
19. Sylwestrzak, O., Respondek-Liberska, M.: Echocardiographic metods of fetal heart size assessment-heart to chest area ratio and transversal heart diameter. Prenatal Cardiology **8**(1) (2018)
20. Van Der Linde, D., Konings, E.E., Slager, M.A., Witsenburg, M., Helbing, W.A., Takkenberg, J.J., Roos-Hesselink, J.W.: Birth prevalence of congenital heart disease worldwide: a systematic review and meta-analysis. Journal of the American College of Cardiology **58**(21), 2241–2247 (2011)
21. Vaswani, A., Shazeer, N., Parmar, N., Uszkoreit, J., Jones, L., Gomez, A.N., Kaiser, Ł., Polosukhin, I.: Attention is all you need. Advances in neural information processing systems **30** (2017)
22. Xie, E., Sun, P., Song, X., Wang, W., Liu, X., Liang, D., Shen, C., Luo, P.: Polarmask: Single shot instance segmentation with polar representation. In: Proceedings of the IEEE/CVF conference on computer vision and pattern recognition. pp. 12193–12202 (2020)
23. Xie, X., Cheng, G., Wang, J., Yao, X., Han, J.: Oriented r-cnn for object detection. In: Proceedings of the IEEE/CVF international conference on computer vision. pp. 3520–3529 (2021)
24. Yang, S., Pei, Z., Zhou, F., Wang, G.: Rotated faster r-cnn for oriented object detection in aerial images. In: Proceedings of the 2020 3rd International Conference on Robot Systems and Applications. pp. 35–39 (2020)
25. Yang, X., Yan, J., Ming, Q., Wang, W., Zhang, X., Tian, Q.: Rethinking rotated object detection with gaussian wasserstein distance loss. In: International conference on machine learning. pp. 11830–11841. PMLR (2021)
26. Zhao, L., Li, N., Tan, G., Chen, J., Li, S., Duan, M.: The end-to-end fetal head circumference detection and estimation in ultrasound images. IEEE/ACM Transactions on Computational Biology and Bioinformatics (2022)

27. Zhou, D., Fang, J., Song, X., Guan, C., Yin, J., Dai, Y., Yang, R.: Iou loss for 2d/3d object detection. In: 2019 international conference on 3D vision (3DV). pp. 85–94. IEEE (2019)
28. Zhou, X., Wang, D., Krähenbühl, P.: Objects as points. arXiv preprint arXiv:1904.07850 (2019)