Progressive Semi-supervised Landmark Detection Algorithm For Intrapartum Ultrasound Measurement

Zelan Li, Hansen Zhang, Zhengyang Zhang, Yan Cheng, Siqi Wang, and Jianning Chi✉

Northeastern University, Shenyang, China
chijianning@mail.neu.edu.cn

Abstract. The angle of progression (AoP) in intrapartum ultrasound is critical for evaluating fetal head descent and rotation during labor, and the angle formed by these three points (PS1, PS2, and FH1). Manual AoP measurement is time-consuming, labor-intensive, and lacks standardization—limitations, automated methods with incomplete labels can improve the efficiency. However, automated methods encounter two significant challenges: firstly, the scarcity of landmark annotations provided by experienced obstetricians may lead to network overfitting and poor generalization; secondly, the anatomical landmarks in ultrasound images are often too small, resulting in insufficient information and feature learning for algorithms. To address these challenges, inspired by the clinical workflow of manual AoP assessment, we propose a progressive semi-supervised landmark detection algorithm, which first locates and identifies the pubic symphysis (PS) and the fetal head (FH) region, and then detects the landmarks of three keypoints to calculate the AoP. Specifically, in the first stage, we utilize the spatial information of landmarks to generate scribbles of the foreground and background of the PS and the FH region. These scribbles are fed to a frozen segmentation foundation model named ScribblePrompt to get coarse segmentation and detection results as pseudo labels, which can help the network concentrate on PS and FH regions. After the first stage of pseudo-label pre-training, the following fine-tuning utilizes pre-trained models to learn landmarks with confidence-guided weight loss to train on labeled and unlabeled data, improving the robustness and generalization of the algorithm. The experimental results show that our algorithm achieved good landmark detection results.

Keywords: Landmark detection · Semi-supervised learning · Intrapartum ultrasound.

1 Introduction

Labor is a dynamic process that requires continuous monitoring to ensure the safety of both mother and fetus [9,4]. A fundamental component of ultrasound-based intrapartum assessment lies in the accurate identification of anatomical

landmarks within intrapartum ultrasound images, as these landmarks serve as the basis for calculating critical clinical parameters such as the angle of progression (AoP) [23,7]. The AoP offers pivotal insights into fetal head descent and rotation during labor, and its measurement directly informs clinical decision-making regarding obstetric interventions. However, current clinical workflows rely on time-intensive manual landmark identification and AoP calculation by experienced obstetricians, and inherent intra- and inter-observer variability in these manual processes undermines the reliability and consistency of measurement results. In contrast, automated landmark detection algorithms for ultrasound images enable the rapid acquisition of standardized assessment outcomes, significantly reducing the time burden on clinical practitioners and thereby facilitating the efficient management of labor processes [28,13,3]. Consequently, an efficient and accurate automated ultrasound landmark detection algorithm holds substantial clinical application value [2,29].

Existing landmark detection approaches can be categorized into two paradigms: 1) Regression methods based on directly comparing landmark coordinates of prediction and landmarks. 2) Methods that treated landmark detection as a classification task based on landmark heatmaps, which typically transform the prediction of landmark locations into a heatmap classification problem. For instance, Sofka et al. [19] proposed a Fully Convolutional Neural Network (FCN) for the accurate automatic detection of measurement points in ultrasound video sequences, utilizing Long Short-Term Memory cells to ensure temporal consistency. YOLOv11 Pose [11] extends the original YOLO framework by integrating a keypoint detection module that enables simultaneous real-time object detection and human pose estimation, leveraging a single neural network to predict both bounding boxes and pose keypoints effectively. Alternatively, certain approaches [16,26] generate a Gaussian heatmap for each landmark and take the location with the maximum value as the ultimate prediction. However, these automated methods face two key challenges in intrapartum Ultrasound Measurement: 1) Landmark annotation relies on experienced obstetricians to manually label ultrasound images, a process that is both tedious and time-consuming. The scarcity of landmark annotations provided by experienced obstetricians can lead to network overfitting and poor generalization; 2) the anatomical landmarks in ultrasound images are often too small, providing limited information per point, resulting in insufficient information and feature learning for algorithms.

In this study, inspired by the clinical workflow of manual AoP assessment [14,12,1], we propose a progressive semi-supervised landmark detection algorithm based on the state-of-the-art YOLOv11 Pose framework for intrapartum ultrasound measurement. Adopting a coarse-to-fine strategy, the algorithm first locates and identifies the pubic symphysis (PS) and the fetal head (FH) region, and then detects the landmarks of three keypoints to calculate the AoP. In the first stage, spatial information of landmarks is leveraged to generate scribble annotations for the foreground and background regions corresponding to the pubic symphysis (PS) and fetal head (FH). These scribbles are input to a frozen segmentation foundation model, ScribblePrompt [24], to obtain coarse segmentation results

and detection outputs as pseudo-labels, thereby guiding the network to focus on these critical anatomical regions [30,10]. Following the first-stage pseudo-label pre-training, the network incorporates anatomical priors, after which the second stage employs a confidence-weighted strategy for labeled and unlabeled data training to fine-tune the pre-trained model, yielding precise landmark detection results. Experimental findings demonstrate that the proposed algorithm achieves promising performance in landmark detection tasks. In summary, the contributions of our work are as follows:

1. We proposed a progressive learning strategy for intrapartum ultrasound measurement learning landmark from coarse to fine.
2. We proposed a scribble generation strategy based on prior anatomical structure, which extracts anatomical prior information of ultrasound images from landmarks, expanding prior knowledge from point range to region range.
3. We proposed a confidence-weighted learning strategy to utilize unlabeled images to enhance the generalization and robustness of the algorithm.

2 Related work

2.1 Self-supervised learning

2.2 Semi-supervised learning

2.3 Landmark detection

3 Method

Motivation The main process for measuring the Angle of Progression (AoP) involves obstetricians first identifying the two farthest points (PS1 and PS2) along the contour of the pubic symphysis (PS) [17,5,6]. Then, a tangent line is drawn from the rightmost point (PS1) such that it just touches the fetal head (FH), with the intersection point marked as the third point (FH1). The AoP is defined as the angle formed by these three points (PS1, PS2, and FH1).

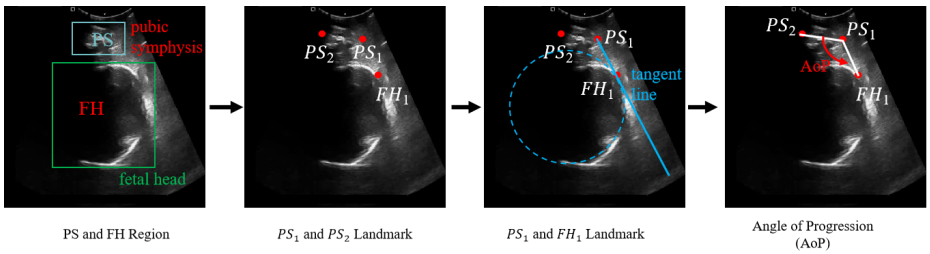

Fig. 1: The main process for measuring the Angle of Progression (AoP)

We observe that these landmarks are derived from the spatial relationships between distinct meaningful regions in ultrasound images, rather than from the direct features of individual points. However, in the task of the IUGC2025 challenge, the limited labeled data only provides the coordinates of PS1, PS2, and FH1. Since these three points are respectively located in two separate regions of interest (ROIs)—the pubic symphysis (PS) and the fetal head (FH)—it is not feasible to directly obtain the prior anatomical information of the PS and FH regions using only these landmarks. To address this limitation, we propose leveraging foundation models with high generalization ability to generate coarse region indications, thereby providing the network with the necessary prior anatomical information.

3.1 Overall Framework

As shown in Figure 6, the proposed network progressively learns to detect coarse-to-fine landmarks in fetal ultrasound images for aortic pressure (Aop) measurement. The training process consists of two stages: In the first stage, we generate high-confidence scribbles based on anatomical spatial relationships as prompts for ScribblePrompt [24] to produce coarse segmentation results and the region of interest (ROI) pseudo-labels. In the second stage, the pre-training backbone network YOLOv11 is utilized to learn landmarks from labeled images and then generate pseudo labels for unlabeled images. The unlabeled images are used to help improve the generalization and robustness of the network through a confidence-weighted strategy.

3.2 Coarse Prior Anatomical Area Learning

Generation of Scribble Prompts The ScribblePrompt [24] proposed by Hallee E. Wong et al. is a flexible neural network-based interactive segmentation tool for biomedical imaging. It allows human annotators to segment previously unseen structures using scribbles [15]. Therefore, we propose a method of spatial information computation: by extracting the spatial relationships between labeled coordinates and ROI regions, we generate scribbles that serve as prompts for input into ScribblePrompt. For the pubic symphysis (PS) region: since PS1 and PS2 are the two farthest points along the PS contour, the line connecting them must lie within the PS contour and thus can be used as a foreground scribble. Mathematically, let the pixel coordinates of PS1 and PS2 be denoted as $\text{PS1} = (x_{\text{ps1}}, y_{\text{ps1}})$ and $\text{PS2} = (x_{\text{ps2}}, y_{\text{ps2}})$ respectively. The line segment $L_{\text{PS1-PS2}}$ (serving as the foreground scribble) can be expressed in two-point form as:

$$\frac{y - y_{\text{ps1}}}{y_{\text{ps2}} - y_{\text{ps1}}} = \frac{x - x_{\text{ps1}}}{x_{\text{ps2}} - x_{\text{ps1}}}, \tag{1}$$

In contrast, the extended parts at both ends of this line can be used as background scribbles. The rays extending $L_{\text{PS1-PS2}}$ beyond PS2 (denoted as $L^{+}_{\text{PS1-PS2}}$)

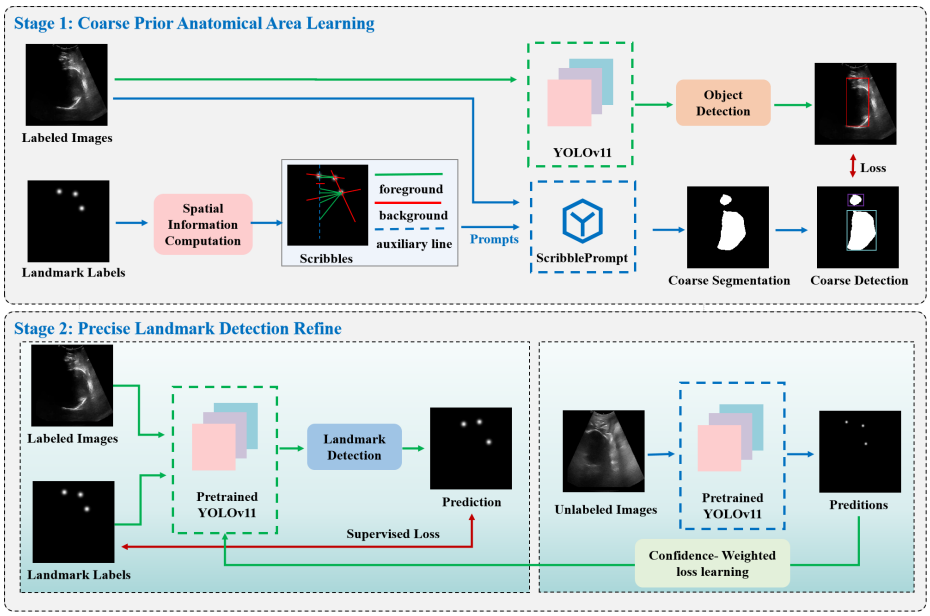

Fig. 2: Overall framework of proposed methods. In stage one, the network learns the coarse prior anatomical area through ScribblePrompt results, which prompted spatial information. In stage two, the network learns landmarks from labeled images and is further refined by unlabeled images with a confidence-weighted loss strategy.

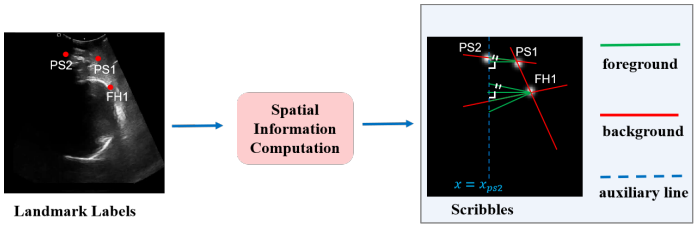

Fig. 3: The main process for measuring the Angle of Progression (AoP)

and beyond PS1 (denoted as $L^{-}_{\text{PS1-PS2}}$) are given by:

$$L^{+}_{\text{PS1-PS2}} : \frac{y - y_{\text{ps2}}}{y_{\text{ps2}} - y_{\text{ps1}}} = \frac{x - x_{\text{ps2}}}{x_{\text{ps2}} - x_{\text{ps1}}}, \tag{2}$$

$$L^{-}_{\text{PS1-PS2}} : \frac{y - y_{\text{ps1}}}{y_{\text{ps2}} - y_{\text{ps1}}} = \frac{x - x_{\text{ps1}}}{x_{\text{ps2}} - x_{\text{ps1}}}, \tag{3}$$

Additionally, in most cases, the contour of a normal pubic symphysis appears convex rather than concave in radiographic images [15]. Therefore, we first draw an auxiliary line perpendicular to the x-axis through PS2, which is expressed as $x = x_{\text{ps2}}$. Then, drawing a line perpendicular to this auxiliary line (one above and one below the auxiliary line) as foreground scribbles will introduce little or no noise, which can enhance the foreground segmentation result of ScribblePrompt.

Fetal Head (FH) Area For the fetal head (FH) area, the labeled data only includes the third point (FH1), which is located at the tangent position of the FH area. Let the pixel coordinate of FH1 be FH1 $= (x_{\text{fh1}}, y_{\text{fh1}})$. After connecting PS1 and FH1, the line segment $L_{\text{PS1-FH1}}$ (and its extended parts) all lie within the background region, so this line can be designated as a background scribble. The equation of $L_{\text{PS1-FH1}}$ in two-point form is:

$$\frac{y - y_{\text{ps1}}}{y_{\text{fh1}} - y_{\text{ps1}}} = \frac{x - x_{\text{ps1}}}{x_{\text{fh1}} - x_{\text{ps1}}}, \tag{4}$$

Moreover, the FH area occupies a relatively large proportion of ultrasound images. When a perpendicular line is drawn from FH1 to the aforementioned auxiliary line $x = x_{\text{ps2}}$, the segment of this perpendicular line (within the FH area) will lie entirely within the FH area with minimal error. Additionally, the lines parallel to the line connecting PS2 and PS1 also lie within the FH area. To make the segmentation result more consistent with the ground truth, we further adopt lines with included angles of 10° and 30° (relative to the reference line) as scribbles for the foreground region of the FH area. Let the perpendicular line from FH1 to $x = x_{\text{ps2}}$ have a foot of perpendicular $H = (x_{\text{ps2}}, y_{\text{fh1}})$. The lines forming 10° and 30° with the perpendicular line FH1 $- H$ (serving as foreground scribbles) have slopes $\tan(10°) \approx 0.1763$, $\tan(170°) \approx -0.1763$, $\tan(30°) \approx 0.577$, $\tan(150°) \approx -0.577$ respectively. Their equations are:

$$y - y_{\text{fh1}} = \pm 0.1763(x - x_{\text{fh1}}) \quad \text{(for } 10° \text{ included angle)} \tag{5}$$

$$y - y_{\text{fh1}} = \pm 0.577 \cdot (x - x_{\text{fh1}}) \quad \text{(for } 30° \text{ included angle)} \tag{6}$$

We take the segments of these lines within the FH area as the foreground scribbles.

Generation of Pseudo Labels After generating scribble prompts from landmarks, we input these prompts into the parameter-frozen ScribblePrompt model to obtain segmentation results for the two target regions. Although the foundation model produces coarse segmentation outputs, direct utilization of these

results would introduce substantial noise due to the dense classification nature of segmentation tasks. However, since our algorithm's core goal is to obtain point labels, only the anatomical positional relationships of each region in ultrasound images are required. Thus, we extract roughly constrained bounding boxes from the segmentation results to train the YOLOv11 backbone network, facilitating regional feature extraction.

Prior Information Learning To extract prior object detection information, we optimize YOLOv11 during the model learning process by integrating multiple loss components. These components jointly address the core tasks of object classification, bounding box regression, and object confidence estimation. The total loss function for object detection training is a weighted combination of three key loss terms: classification loss, bounding box regression loss, and object confidence loss, as defined in Eq. 7.

$$\mathcal{L}_{\text{detect}} = \lambda_{\text{cls}} \cdot \mathcal{L}_{\text{cls}} + \lambda_{\text{reg}} \cdot \mathcal{L}_{\text{reg}} + \lambda_{\text{conf}} \cdot \mathcal{L}_{\text{conf}} \tag{7}$$

where λ_{cls}, λ_{reg}, and λ_{conf} are cross-validated hyperparameters used to balance the contributions of each loss component. $\mathcal{L}_{\text{cls}}$ is the classification loss, which employs Cross-Entropy (CE) loss to ensure accurate classification of different types of landmarks:

$$\mathcal{L}_{\text{cls}}(y, \hat{y}) = -\frac{1}{N} \sum_{i=1}^{N} \sum_{c=1}^{C} y_{i,c} \log(\hat{y}_{i,c}) \tag{8}$$

$\mathcal{L}_{\text{conf}}$ denotes the object existence confidence loss, which uses Binary Cross-Entropy (BCE) loss to distinguish between valid bounding boxes and background regions:

$$\mathcal{L}_{\text{conf}} = -\frac{1}{N_{\text{obj}} + N_{\text{noobj}}} \left(\sum_{i \in \text{obj}} C_i \log(\hat{C}_i) + \sum_{i \in \text{noobj}} (1 - C_i) \log(1 - \hat{C}_i) \right) \tag{9}$$

$\mathcal{L}_{\text{reg}}$ represents the bounding box regression loss, which utilizes Mean Squared Error (MSE) loss to optimize the spatial alignment between predicted and ground-truth bounding boxes:

$$\mathcal{L}_{\text{reg}} = \sum_{i=1}^{N_{\text{obj}}} \mathbb{1}_{\text{obj}} \left[(x_i - \hat{x}_i)^2 + (y_i - \hat{y}_i)^2 + (w_i - \hat{w}_i)^2 + (h_i - \hat{h}_i)^2 \right] \tag{10}$$

In the above formulas: N denotes the total number of bounding boxes; $C = 2$ represents the total number of box categories; $y_i = (x_i, y_i, w_i, h_i)$ and $\hat{y}_i = (\hat{x}_i, \hat{y}_i, \hat{w}_i, \hat{h}_i)$ are the ground-truth and predicted coordinates (center (x, y) and dimensions (w, h)) of the i-th bounding box, respectively; $y_{i,c}$ and $\hat{y}_{i,c}$ denote the one-hot pseudo labels and predicted probability that the i-th box belongs to the c-th category, respectively; N_{obj} and N_{noobj} are the counts of positive

(object-containing) and negative (background) boxes; C_i and $\hat{C}_i$ represent the ground-truth (1 for positive, 0 for negative) and predicted confidence scores of the i-th box; $\mathbb{1}_{\text{obj}}$ is an indicator function (1 for positive boxes, 0 otherwise);

3.3 Precise Landmark Detection Refine

Labeled Images Learning After generating ROI detection pseudo labels, the backbone network YOLOv11 can extract anatomical prior information from ultrasound images, which makes it possible that during the refinement process, implicit associations between key points and anatomical structures are added, accelerating the training of the network.

In the landmark detection phase, we use a multi-task joint loss function, incorporating an auxiliary ROI object detection loss into the specialized regression loss for keypoint detection, which enables the network to simultaneously learn both tasks effectively, and the ROI detection can help the network concentrate on the spatial and semantic relationship of landmarks. For the landmarks l and the predicted landmarks $\hat{l}$, the specific formula is as follows:

$$\mathcal{L}_{\text{label}} = \mathcal{L}_{\text{reg}}(l, \hat{l}) + \alpha \cdot \mathcal{L}_{\text{detect}}(y, \hat{y}) \tag{11}$$

where α are the weight coefficients used to balance the loss components, $\mathcal{L}_{\text{reg}}$ represents the regression loss of landmarks, which can be calculated using the Mean Squared Error (MSE) to measure the spatial deviation between the predicted and ground-truth key points:

$$\mathcal{L}_{\text{loc}}(l, \hat{l}) = \frac{1}{N} \sum_{i=1}^{N} \left\| l_i - \hat{l}_i \right\|^2 \tag{12}$$

where N represents the number of key points.

Unlabeled Image Learning After training the network on labeled data, the trained model was used to generate landmark pseudo-labels for the unlabeled dataset. By setting a threshold for the output confidence score, we excluded images where the network failed to make valid predictions, leaving only high-confidence pseudo-labeled data for subsequent training.

To further enhance the model's learning efficiency on pseudo-labeled data, we introduce information entropy to quantify the confidence of the model's predictions and incorporate this confidence into the loss function as a dynamic weight. During training, the confidence is incorporated into the loss function as a weight coefficient:

$$\mathcal{L}_{\text{unlabel}} = \mathcal{L}_{\text{reg}}(l, \hat{l}) + \mathcal{C}(p, \hat{p}) \cdot \mathcal{L}_{\text{reg}}(p, \hat{p}) \tag{13}$$

For the prediction result p output by the pre-trained network and the network prediction $\hat{p}$ obtained after introducing noise, the confidence $\mathcal{C}(p, \hat{p})$ is defined as:

$$\mathcal{C}(p, \hat{p}) = \frac{1}{4}\left(\frac{p \cdot \hat{p}}{\|p\| \cdot \|\hat{p}\|} + 1\right) \tag{14}$$

where $p \cdot \hat{p}$ denotes the dot product of the two prediction vectors, and $\|p\|$, $\|\hat{p}\|$ represent the L2 norms of p and $\hat{p}$, respectively. The confidence $\mathcal{C}(p, \hat{p})$ ranges from $[0, 0.5]$, with values closer to 0.5 indicating higher consistency between the predictions before and after noise injection, thus reflecting a higher confidence in the predictions of the model on unlabeled data. This weighted loss mechanism allows the model to adaptively allocate learning resources during training using unlabeled data, thereby improving the robustness of landmark detection.

4 Experiments

4.1 Experimental Materials

Dataset The experiment utilizes a transperineal ultrasound dataset provided by the Intrapartum Ultrasound Grand Challenge 2025 (IUGC 2025) [8], which contains 31,421 images with 300 labeled images. During training, the labeled data was randomly split into training and testing sets at a 9:1 ratio. Due to the test phase was already closed, we use metrics from the validation phase online as ablation and comparison experiment results.

Evaluation Metrics We conducted a quantitative comparison using two standard landmark localization and parameter estimation evaluation metrics: Mean Radial Error (MRE) and Absolute Parameter Difference (APD) [22]. The former assesses the spatial localization accuracy of predicted landmark points relative to ground truth, while the latter evaluates the deviation between the estimated Angle of Polarization (AoP) (derived from predicted landmarks) and the ground truth AoP. The definitions of each metric are as follows:

$$\text{MRE} = \frac{1}{N}\sum_{i=1}^{N}\sqrt{(x_i^{pred} - x_i^{gt})^2 + (y_i^{pred} - y_i^{gt})^2} \tag{15}$$

where (x_i^{pred}, y_i^{pred}) and (x_i^{gt}, y_i^{gt}) are the predicted and ground truth coordinates of the i-th landmark, and N is the total number of landmarks.

$$\text{APD} = |\text{AoP}^{pred} - \text{AoP}^{gt}| \tag{16}$$

where AoP^{pred} and AoP^{gt} are the predicted and ground truth Angle of Pose, respectively.

Parameter Setting and Implementation In the first training stage, we set a learning rate of 0.001, a batch size of 8, and trained for 100 epochs with images resized to 640×640. In the second training stage, we set a learning rate of 0.001, a total batch size of 8, a labeled batch size of 4, a unlabeled batch size of 4,

and trained for 100 epochs with images resized to 640×640. The network was optimized using the SGD optimizer. For comparative experiments, we adhered to the same parameters. All experiments were carried out using PyTorch on an Nvidia 3090 GPU equipped with 24GB of memory.

4.2 Ablation Study

Ablation studies were conducted to verify the effectiveness of each component in the proposed framework. Table 1 summarizes the performance of models with different design configurations, where Model A (Baseline) uses only YOLOv11, trained exclusively on labeled images for landmark detection; Model B incorporates a pre-training stage for coarse anatomical region prior learning, followed by fine-tuning the pre-trained YOLOv11 on labeled images for landmark detection; Model C builds on Model B by adding a confidence-weighted loss term for unlabeled images during the training process.

Table 1: Quantitative ablation results of models with different network learning strategies. The baseline used is YOLOv11 with only the labeled images for training.

Model	Strategy			Metrics	
	Baseline	+ Coarse Prior	+ Unlabeled Learning	MRE (pixel) ↓	APD (degree) ↓
A	✓			31.8969	13.5713
B	✓	✓		21.2317	9.4166
C	✓	✓	✓	17.7447	7.6243

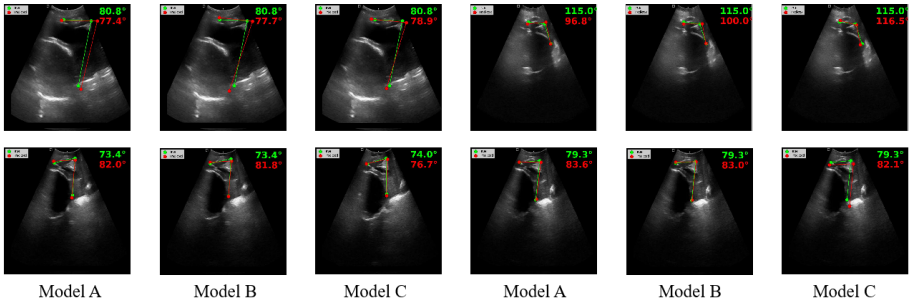

Fig. 4: Landmark detection results of different ablation models, green represents ground truth, red represents prediction results.

Effectiveness of Coarse Prior Anatomical Area Learning As shown in Table 1, by comparing Model A and Model B, it can be observed that the integration of the Coarse Prior Anatomical Area Learning module enables the network to capture the correlations between different regions in ultrasound images, leading to a significant improvement in the detection accuracy of key landmarks. Specifically, the MRE decreases substantially from 31.8969 to 21.2317, and the APD also reduces from 13.5713 to 9.4166.

The visualization results in Fig. 4 further confirm that after incorporating anatomical prior learning, each landmark predicted by the network is closer to the ground truth. In particular, the detection performance for the FH1 landmark has been remarkably enhanced.

Effectiveness of Precise Landmark Detection Refine As shown in Table 1, by comparing Model B and Model C, it can be seen that after incorporating unlabeled images into the network training via the confidence-weighted loss during the refinement stage, the network's key landmark detection performance is further improved. Specifically, the MRE decreases from 21.2317 to 17.7447, and the APD reduces from 9.4166 to 7.6243. These results verify the effectiveness of the semi-supervised strategy proposed in this study.

The visualization results in Fig. 4 demonstrate that after introducing the confidence-weighted loss, the AoP predicted by the network is more consistent with the ground truth.

4.3 Comparison Study

We evaluated the AoP measurement performance of our proposed framework and compared it with several fully-supervised methods and semi-supervised methods: 1) Unet [18]; 2) Cenet [20]; 3) YOLOv11 [11]; 4) Mean Teacher [21] ; 5) Adversarial Network [27]; 6)DFGC [25].

Table 2: Quantitative results of different Comparison networks.

Supervision	Model	Task type	MRE (pixel) ↓	APD (degree) ↓
Fully-supervised	UNet	Regression	26.7731	10.4903
	UNet	Classification	24.2149	10.1297
	CENet	Regression	27.6230	9.8838
	CENet	Classification	20.7143	8.4383
	YOLOv11	Regression	31.8969	13.5713
Semi-supervised	Mean Teacher	Classification	27.7819	10.3221
	Adversarial Network	Classification	32.6527	14.2298
	DFGC	Classification	24.7469	9.1388
	ours	Regression	**17.7447**	**7.6243**

As presented in Table 2, different algorithms exhibit varying landmark detection performances when adopting distinct learning strategies. For the Unet

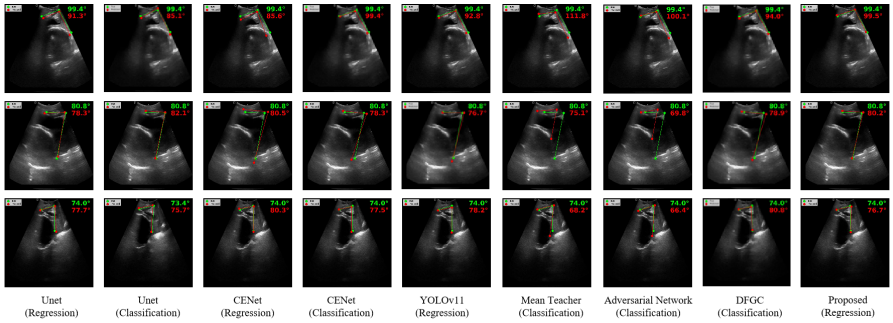

Fig. 5: Landmark detection results of different comparison methods, green represents ground truth, red represents prediction results.

and CENet algorithms, compared with direct coordinate regression tasks (which directly obtain landmark coordinates), using heatmaps for image classification tasks achieves a reduction in MRE by 2.5582 and 6.9087, respectively, with the APD also decreasing by more than 0.3. The nnUNet algorithm, which employs heatmaps for classification, yields better metrics than CENet. When used independently, YOLOv11 shows slightly inferior performance compared to other fully supervised algorithms. For the Mean Teacher algorithm and Adversarial Network applied to landmark detection with UNet as its backbone, the extremely limited information in heatmaps generated from key points prevents the algorithm from learning effective features, resulting in excessively large MRE and APD values that render it unable to detect target points effectively. The DFGC algorithm, specifically designed for ultrasound images, can learn target-related features effectively (achieving an MRE of 24.7469 and an APD of 9.1388) but still performs less favorably than the algorithm proposed in this study.

Fig. 5 further confirms that the proposed algorithm achieves significant improvements in detecting the PS1 and PS2 landmarks compared to other algorithms. For FH1 detection, the algorithm substantially enhances positional accuracy, ensuring that FH1 is located as close as possible on the fetal head contour—avoiding mislocalization within the FH region—and thus enabling more accurate calculation of the AoP.

5 Discussion

During the competition, our experiments revealed notable findings: combining a semi-supervised framework with the standard UNet drastically degrades landmark detection performance, as evidenced by substantial worsening of metrics like MRE and angular offset mean absolute error (APD) (exceeding 100). This underscores the limitations of generic encoder-decoder architectures for semi-supervised intrapartum ultrasound measurements. The core issue likely stems from the sparsity and ambiguity of ultrasound keypoints—semi-supervised

pseudo-labels generated from unlabeled data tend to propagate noise when learning from isolated, context-poor points. While UNet excels at dense pixel-wise segmentation, it fails to capture sparse, high-resolution spatial cues critical for keypoint localization, instead overfitting irrelevant background textures or blurred boundaries and amplifying distance/angle estimation errors.

In contrast, integrating a regional background (e.g., the PS1-PS2-FH1 triangle) defined by three key landmarks with a conventional encoder-decoder architecture significantly improves detection accuracy. This method incorporates spatial priors, framing the task as regional feature extraction rather than isolated point regression. This approach provides richer contextual information, such as relative positional relationships and boundary transitions between the pubic symphysis and fetal head. The proposed encoder-decoder architecture is optimized for hierarchical feature aggregation, enabling more effective regional pattern learning, reducing keypoint ambiguity, and enhancing the robustness of distance and angle estimation.

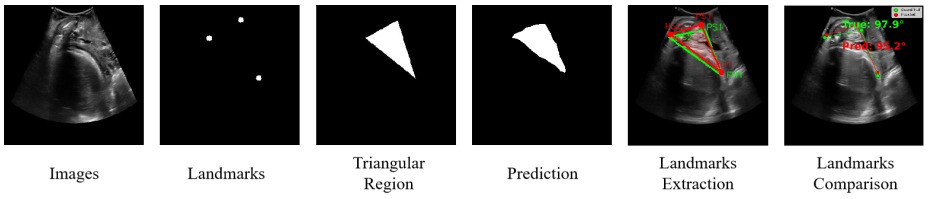

Fig. 6: Schematic diagram of learning landmark detection from the triangular region composed of landmarks.

Based on the advantages of introducing regional background information, future work can further optimize models based on regional feature extraction. Explore how to combine multi-level spatial prior knowledge, utilize geometric relationships and semantic information between key points, not limited to the connection of three landmark points, but also consider more complex regional constraints, thereby further improving detection accuracy and model robustness.

6 Conclusion

This study focuses on addressing the limitations of manual Angle of Progression measurement in intrapartum ultrasound and tackles the core challenges faced by automated AoP landmark measurement methods: the reliance on massive manual landmark annotations and the difficulty of detecting minute fetal head anatomical landmarks. To resolve these issues, a progressive semi-supervised landmark detection algorithm for automated AoP measurement is proposed, which aligns with the clinical workflow of manual AoP assessment. In the first stage, spatial information of key landmarks is leveraged to generate foreground/background scribbles for the pubic symphysis and fetal head. These

scribbles are input to the frozen ScribblePrompt segmentation foundation model to obtain coarse segmentation results, which serve as pseudo-labels to guide the network in focusing on key anatomical regions. In the second stage, the network extracts anatomical priors from ultrasound images. It undergoes fine-tuning via weight-allocated learning on both labeled and unlabeled data—effectively reducing the demand for annotated data while enhancing the model's ability to learn discriminative features of minute landmarks. Experimental results confirm that the proposed algorithm achieves favorable performance in key anatomical landmark detection, laying a solid foundation for accurate automated AoP measurement. By integrating anatomical prior learning and semi-supervised learning, the algorithm not only mitigates the dependency on manual annotation but also improves the robustness of landmark detection against tissue confusion, thereby providing a clinically valuable tool for objective, standardized, and efficient AoP assessment during labor.

References

1. Bai, J., Lekadir, K., Ni, D., Slimani, S., Campello, V.M., Ohene-Botwe, B., Lu, Y., Chen, G., Hou, H., Qiu, D., Zhou, Z.: Intrapartum ultrasound grand challenge 2024. In: 27th International Conference on Medical Image Computing and Computer Assisted Intervention (MICCAI 2024). Zenodo (2024). https://doi.org/10.5281/zenodo.10979813, https://doi.org/10.5281/zenodo.10979813
2. Bai, J., Ou, Z., Lu, Y., Ni, D., Chen, G.: Pubic symphysis-fetal head segmentation from transperineal ultrasound images. In: International Conference on Medical Image Computing and Computer Assisted Intervention (MICCAI) (2023)
3. Bai, J., Sun, Z., Yu, S., Lu, Y., Long, S., Wang, H., Qiu, R., Ou, Z., Zhou, M., Zhi, D., et al.: A framework for computing angle of progression from transperineal ultrasound images for evaluating fetal head descent using a novel double branch network. Frontiers in physiology **13**, 940150 (2022)
4. Bai, J., Zhou, Z., Ou, Z., Koehler, G., Stock, R., Maier-Hein, K., Elbatel, M., Martí, R., Li, X., Qiu, Y., et al.: Psfhs challenge report: pubic symphysis and fetal head segmentation from intrapartum ultrasound images. Medical Image Analysis **99**, 103353 (2025)
5. Chen, G., Bai, J., Ou, Z., Lu, Y., Wang, H.: Psfhs: intrapartum ultrasound image dataset for ai-based segmentation of pubic symphysis and fetal head. Scientific Data **11**(1), 436 (2024)
6. Chen, Z., Lu, Y., Long, S., Campello, V.M., Bai, J., Lekadir, K.: Fetal head and pubic symphysis segmentation in intrapartum ultrasound image using a dual-path boundary-guided residual network. IEEE journal of biomedical and health informatics **28**(8), 4648–4659 (2024)
7. Chen, Z., Ou, Z., Lu, Y., Bai, J.: Direction-guided and multi-scale feature screening for fetal head–pubic symphysis segmentation and angle of progression calculation. Expert Systems with Applications **245**, 123096 (2024)
8. CodaBench: Iugc2025: International unsupervised geospatial challenge. https://www.codabench.org/competitions/7105/ (2025), accessed: 2025-08-30
9. Haws, R.A., Yakoob, M.Y., Soomro, T., Menezes, E.V., Darmstadt, G.L., Bhutta, Z.A.: Reducing stillbirths: screening and monitoring during pregnancy and labour. BMC pregnancy and childbirth **9**(Suppl 1), S5 (2009)

10. Jiang, J., Wang, H., Bai, J., Long, S., Chen, S., Campello, V.M., Lekadir, K.: Intrapartum ultrasound image segmentation of pubic symphysis and fetal head using dual student-teacher framework with cnn-vit collaborative learning. In: International conference on medical image computing and computer-assisted intervention. pp. 448–458. Springer (2024)
11. Khanam, R., Hussain, M.: Yolov11: An overview of the key architectural enhancements. arXiv preprint arXiv:2410.17725 (2024)
12. Lu, Y., Zhi, D., Zhou, M., Lai, F., Chen, G., Ou, Z., Zeng, R., Long, S., Qiu, R., Zhou, M., et al.: Multitask deep neural network for the fully automatic measurement of the angle of progression. Computational and mathematical methods in medicine **2022**(1), 5192338 (2022)
13. Lu, Y., Zhou, M., Zhi, D., Zhou, M., Jiang, X., Qiu, R., Ou, Z., Wang, H., Qiu, D., Zhong, M., et al.: The jnu-ifm dataset for segmenting pubic symphysis-fetal head. Data in brief **41**, 107904 (2022)
14. Ou, Z., Bai, J., Chen, Z., Lu, Y., Wang, H., Long, S., Chen, G.: Rtseg-net: a lightweight network for real-time segmentation of fetal head and pubic symphysis from intrapartum ultrasound images. Computers in biology and medicine **175**, 108501 (2024)
15. van Ovost, A., Hanff, D.F., Serner, A., van Klij, P., Agricola, R., Weir, A.: Radiographic assessment of the pubic symphysis in elite male adolescent football players: development and reliability of the maturing adolescent pubic symphysis (maps) classification. European journal of radiology **167**, 111068 (2023)
16. Payer, C., Štern, D., Bischof, H., Urschler, M.: Regressing heatmaps for multiple landmark localization using cnns. In: International conference on medical image computing and computer-assisted intervention. pp. 230–238. Springer (2016)
17. Qiu, R., Zhou, M., Bai, J., Lu, Y., Wang, H.: Psfhsp-net: an efficient lightweight network for identifying pubic symphysis-fetal head standard plane from intrapartum ultrasound images. Medical & Biological Engineering & Computing **62**(10), 2975–2986 (2024)
18. 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)
19. Sofka, M., Milletari, F., Jia, J., Rothberg, A.: Fully convolutional regression network for accurate detection of measurement points. In: International Workshop on Deep Learning in Medical Image Analysis. pp. 258–266. Springer (2017)
20. Tao, H., Xie, C., Wang, J., Xin, Z.: Cenet: A channel-enhanced spatiotemporal network with sufficient supervision information for recognizing industrial smoke emissions. IEEE internet of things journal **9**(19), 18749–18759 (2022)
21. Tarvainen, A., Valpola, H.: Mean teachers are better role models: Weight-averaged consistency targets improve semi-supervised deep learning results. Advances in neural information processing systems **30** (2017)
22. Wang, C.W., Huang, C.T., Hsieh, M.C., Li, C.H., Chang, S.W., Li, W.C., Vandaele, R., Marée, R., Jodogne, S., Geurts, P., et al.: Evaluation and comparison of anatomical landmark detection methods for cephalometric x-ray images: a grand challenge. IEEE transactions on medical imaging **34**(9), 1890–1900 (2015)
23. Wang, X., Xiao, F., Li, J.: Intrapartum ultrasound monitoring in second-stage labor: Impact on delivery outcomes. Journal of Radiation Research and Applied Sciences **18**(4), 101846 (2025)

24. Wong, H.E., Rakic, M., Guttag, J., Dalca, A.V.: Scribbleprompt: fast and flexible interactive segmentation for any biomedical image. In: European Conference on Computer Vision. pp. 207–229. Springer (2024)
25. Yao, Y., Duan, X., Qu, A., Chen, M., Chen, J., Chen, L.: Dfcg: A dual-frequency cascade graph model for semi-supervised ultrasound image segmentation with diffusion model. Knowledge-Based Systems **300**, 112261 (2024)
26. Yin, S., Wang, S., Chen, X., Chen, E., Liang, C.: Attentive one-dimensional heatmap regression for facial landmark detection and tracking. In: Proceedings of the 28th ACM International Conference on Multimedia. pp. 538–546 (2020)
27. Zhang, Y., Yang, L., Chen, J., Fredericksen, M., Hughes, D.P., Chen, D.Z.: Deep adversarial networks for biomedical image segmentation utilizing unannotated images. In: International conference on medical image computing and computer-assisted intervention. pp. 408–416. Springer (2017)
28. Zhou, M., Yuan, C., Chen, Z., Wang, C., Lu, Y.: Automatic angle of progress measurement of intrapartum transperineal ultrasound image with deep learning. In: International conference on medical image computing and computer-assisted intervention. pp. 406–414. Springer (2020)
29. Zhou, Z., Lu, Y.: Baseline method at the intrapartum ultrasound grand challenge 2024. In: Intrapartum Ultrasound Grand Challenge, pp. 1–10. Springer (2024)
30. Zhou, Z., Lu, Y., Bai, J., Campello, V.M., Feng, F., Lekadir, K.: Segment anything model for fetal head-pubic symphysis segmentation in intrapartum ultrasound image analysis. Expert Systems with Applications **263**, 125699 (2025)