DSNT-DeepUNet: A Coordinate Prediction Method for Intrapartum Ultrasound

Yang Zi[1][0009−0004−4629−1883], Liu Qingchen[1][0009−0003−2453−1770], Hu Yuchen[1][0009−0003−3564−5079], Kuang Jingfan[1][0009−0008−2907−2530], Song Shanglin[1][0009−0003−9015−9663], and Wang Jianlin[1]✉[0009−0004−8500−7670]

1.The First Hospital of Lanzhou University, Lanzhou Gansu 730000, China
448706606@qq.com

Abstract. Intrapartum ultrasound monitoring is critical for maternal-fetal safety, yet traditional manual annotation of key anatomical landmarks (PS1, PS2, FH1) faces bottlenecks such as significant inter-observer variability and time-intensive processes, hindering standardized implementation of the WHO Labor Care Guide (LCG). This study proposes DSNT-DeepUNet, a deep learning-based ultrasound coordinate prediction model. By integrating a U-Net backbone with a Differentiable Spatial to Numerical Transform (DSNT) layer, it achieves end-to-end mapping from raw ultrasound images to keypoint coordinates. The model employs a multi-task loss function to simultaneously optimize coordinate accuracy and heatmap distribution, while an 8-fold cross-validation strategy and dynamic data augmentation techniques significantly enhance generalization capability. On an independent test set, the model achieved an angle of progression prediction error of 4.7005 pixels and an average point distance error of 14.7712 pixels, with PS1 and PS2 localization errors at 9.0600 and 11.5661 pixels respectively, ranking sixth in a public challenge. This solution successfully eliminates subjective variations in manual annotation, demonstrating effective and precise ultrasound coordinate prediction.

Keywords: Intrapartum ultrasound monitoring · Anatomical landmark localization · DSNT network.

1 Introduction

Intrapartum ultrasound examination, as a non-invasive and real-time fetal monitoring method, is widely used to assess fetal position, predict delivery methods, and aid clinical decision-making [9, 11]. Studies have demonstrated its superiority over traditional digital vaginal examination in terms of objectivity, reproducibility, and patient compliance, significantly reducing discomfort and infection risks [6, 11]. During the procedure, clinicians must manually annotate key anatomical landmarks—such as the symphysis pubis, the midpoint of the fetal cranium, and the umbilical cord insertion site—on ultrasound images to measure parameters like head-perineum distance (HPD) and angle of progression

(AOP), which are critical for evaluating fetal descent and predicting delivery outcomes [5, 10]. However, due to subjective variations in operator experience, this annotation process exhibits significant inconsistency: multiple studies report an average localization deviation of 10–15 pixels for the same landmark across different clinicians, with a single annotation typically requiring 3–5 minutes [4]. Such limitations hinder rapid, precise monitoring in time-sensitive clinical settings, particularly during the second stage of labor where timely decision-making is crucial for avoiding adverse maternal and neonatal outcomes [5, 9]. With the advancement of deep learning in medical imaging, automated keypoint detection via convolutional neural networks (CNNs) has emerged as a research focus, primarily following two approaches. First, direct coordinate regression uses fully connected layers at the network's terminus to predict coordinates [1, 7]. While structurally simple, this method struggles to leverage spatial contextual information, limiting sub-pixel localization accuracy, and is highly sensitive to the spatial distribution of training data, which can hinder generalization. Second, heatmap-based methods generate probability distribution maps for anatomical points, with coordinates determined via non-differentiable argmax operations. Although preserving multi-scale features, this framework prevents end-to-end optimization (due to argmax's non-differentiability) and introduces quantization errors from limited heatmap resolution. To address these constraints, the Differentiable Spatial to Numerical Transform (DSNT) was proposed [7]. By computing spatial expectations over heatmaps, DSNT maintains spatial probability modeling while mapping discrete pixels to continuous coordinates, enabling gradient propagation and reducing quantization errors, all without introducing additional parameters [7]. Although DSNT has demonstrated superior performance in fields such as human pose estimation and facial keypoint detection [7], its application in medical ultrasound imaging remains in its nascent stages. Previous studies in medical imaging have largely relied on segmentation-based approaches, such as U-Net [8], for structure localization, which require pixel-level annotations and are time-consuming to produce [1, 3]. Alternatively, regression-based methods that directly output coordinates have been explored to reduce annotation burden [1, 3], yet they often lack the ability to provide spatial interpretability. Recent work has also shown that combining regression with implicit localization, as in biomarker regression networks [3], can yield both accurate measurements and spatial maps without segmentation labels, though such methods are still underexplored in ultrasound. This study utilized 300 cases of data provided by the competition organizers [2] and engaged three physicians with three years of experience in ultrasound diagnostics to annotate an additional 169 cases, resulting in a total of 469 intrapartum ultrasound datasets. For the first time, the DSNT module was seamlessly integrated into the heatmap branch of a deep U-Net, establishing an end-to-end differentiable localization framework. Building on this, the study designed a composite loss function that combines pixel-level Euclidean distance with heatmap distribution regularization, automatically optimizing their weights through grid search to balance coordinate accuracy and probability distribution quality. Furthermore, the study employed stratified 8-

fold cross-validation and rigorous statistical testing to comprehensively evaluate model performance, with additional testing on a separate test set to validate the practical improvements in annotation efficiency and accuracy. The results demonstrate that the proposed method not only significantly reduces the average localization error of keypoints but also substantially shortens physicians' annotation time, offering a reliable and feasible technical solution for intelligent assisted analysis of intrapartum ultrasound.

2 Method Design

2.1 Methodology

The model architecture in this study adopts a deep symmetric dual-branch collaborative optimization framework, as illustrated in Fig. 1, to fully integrate multi-scale feature extraction with precise coordinate decoding capabilities. The heatmap generation branch employs an improved U-Net as its backbone network, featuring a seven-level progressive downsampling and symmetric upsampling design that enables hierarchical analysis from local textures to global anatomical structures. Specifically, the encoding phase utilizes 3×3 convolutional kernels with a stride of 2 for feature mapping during each downsampling step, followed by batch normalization and ReLU activation to effectively mitigate gradient vanishing and enhance the network's nonlinear representation capacity. As the network deepens, the number of channels progressively doubles from an initial 64 to 512, expanding the receptive field to capture anatomical information at varying scales. Concurrently, each encoder stage incorporates a 2×2 max-pooling operation at its endpoint to rapidly reduce feature map dimensions while preserving critical spatial information, ensuring computational efficiency.

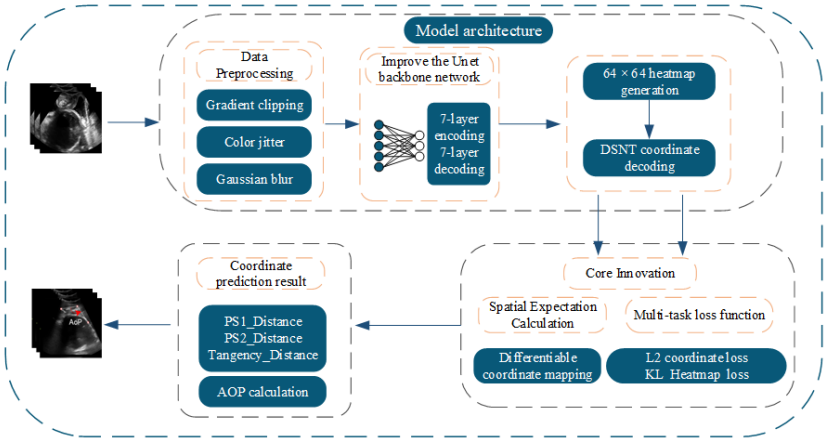

Fig. 1. Overall workflow

2.2 Network Architecture

The model architecture in this study adopts a deep symmetric dual-branch collaborative optimization framework, as shown in Fig. 2, to fully integrate multi-scale feature extraction with precise coordinate decoding capabilities. The heatmap generation branch employs an improved U-Net as its backbone network, featuring a seven-level progressive downsampling and symmetric upsampling design that enables hierarchical analysis from local textures to global anatomical structures. Specifically, the encoding phase utilizes 3×3 convolutional kernels with a stride of 2 for feature mapping during each downsampling step, followed by batch normalization and ReLU activation to effectively mitigate gradient vanishing and enhance the network's nonlinear representation capacity. As the network deepens, the number of channels progressively doubles from an initial 64 to 512, expanding the receptive field to capture anatomical information at varying scales. Concurrently, each encoder stage incorporates a 2×2 max-pooling operation at its endpoint to rapidly reduce feature map dimensions while preserving critical spatial information, ensuring computational efficiency.

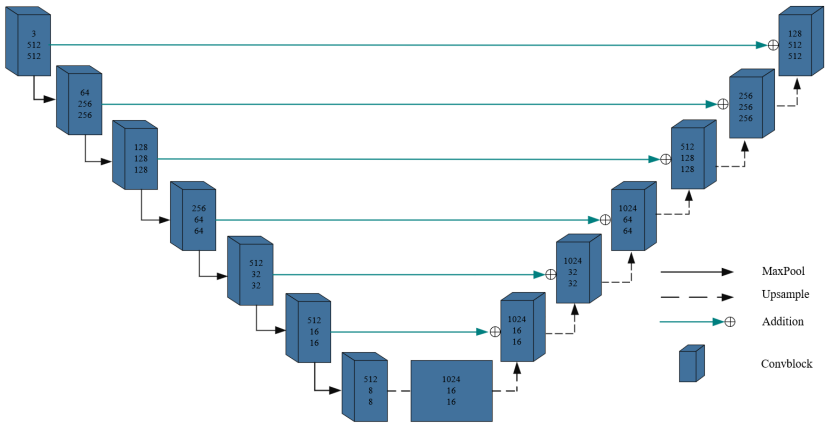

Fig. 2. Model Architecture Diagram

During the decoding phase, the model performs layer-wise upsampling via transposed convolution and smoothly restores spatial resolution through bilinear interpolation. Consistent with the traditional U-Net architecture, the decoder employs skip connections to fuse high-resolution features from the encoder at corresponding scales with the upsampled results, thereby mitigating the loss of deep semantic information. On this basis, the concatenated high-dimensional tensor undergoes further refinement through two consecutive 3×3 convolutional modules, effectively balancing spatial details with semantic integrity.

At the fourth decoding level, the 256-dimensional features extracted from the encoder's third level are concatenated with the upsampled 512-dimensional

features, forming a 768-dimensional feature descriptor. This descriptor is then refined through successive convolutions to ensure that fused features at all levels adequately represent anatomical structural variations. Finally, the decoder output employs a 1×1 convolution to reduce the channel dimension to the number of keypoints (three channels in this study), generating a 64×64 heatmap where each channel corresponds to the probability distribution cloud of an anatomical landmark. This heatmap resolution, approximately one-eighth of the original image size, maintains sufficient spatial detail while keeping computational costs within a reasonable range.

The probability distribution maps generated by the heatmap branch are further fed into the DSNT (Differentiable Spatial to Numerical Transform) layer, enabling differentiable conversion from discrete probabilities to continuous coordinates. During model initialization, the DSNT module constructs and registers a Cartesian grid coordinate buffer with equidistant values in the range [-1, 1], which is efficiently reused via tensor operations. In the forward pass, a pixel-wise Softmax is first applied to each heatmap channel to ensure the probability distribution adheres to normalization axioms. Subsequently, the expected coordinates of the keypoints are computed by performing an element-wise Hadamard product between the normalized probabilities and grid coordinates, followed by summation along the spatial dimensions. This mechanism fundamentally eliminates the quantization error inherent in traditional argmax operations and allows gradients to backpropagate directly from the coordinate loss layer to the heatmap generation branch, enabling fully end-to-end training.

In terms of model hyperparameter tuning, empirical optimization determined an optimal balance between a seven-level depth and 3×3 convolutional kernel size. The seven-level depth ensures sufficient receptive field coverage for complete anatomical structures while avoiding gradient vanishing issues associated with excessively deep networks. Fixed-size medium kernels enhance local feature representation while keeping parameter counts manageable. The channel doubling strategy draws inspiration from biological visual systems' multiscale processing, enabling the network to efficiently capture spatial details across different levels. The 64×64 heatmap resolution provides sufficiently fine-grained probability distributions for the DSNT layer while maintaining controllable GPU memory usage.

Overall, this dual-branch architecture forms a closed-loop optimization pipeline from 512×512-pixel ultrasound image inputs to three sets of keypoint pixel coordinates. The spatial features distilled by the heatmap branch not only encapsulate rich local and global information but also establish a differentiable coordinate learning pathway with the DSNT module. During training, the network optimizes both heatmap distribution quality and coordinate precision through a composite loss function, enabling the heatmap branch and DSNT layer to coevolve and ultimately achieve sub-pixel localization accuracy. This framework provides an efficient, reliable, and easily deployable solution for automated annotation of intrapartum ultrasound images.

3 Experiments

3.1 Data and Preprocessing

This study constructed a dataset sourced from multiple Grade A tertiary hospitals and maternal and child health centers across China, including the First Affiliated Hospital of Jinan University, Zhujiang Hospital of Southern Medical University, Nanfang Hospital of Southern Medical University, the Third Affiliated Hospital of Sun Yat-sen University, Guangzhou Women and Children's Medical Center, and over ten other medical institutions, ensuring broad representativeness and clinical authenticity. The image data were acquired via transperineal ultrasound examinations using devices from various manufacturers, such as Philips CS50, Toshiba Aplio300, Voluson P8, Esaote MyLab, Mindray Resona series, and Youkey Q7, thereby ensuring diversity in imaging equipment.

Image acquisition was performed by an experienced specialized team, with all operators possessing more than seven years of expertise in ultrasound diagnostics. A standardized image acquisition protocol was followed, which included probe preparation, the application of coupling gel, and fine adjustments in positioning to ensure clear visualization of key pelvic and fetal anatomical landmarks while minimizing artifacts. Each case corresponds to a single ultrasound image. The training set consists of 300 images, and the validation set contains 100 images, making it one of the largest publicly available labeled intrapartum ultrasound datasets to date. Furthermore, three obstetricians from the First Hospital of Lanzhou University, each with over seven years of experience, independently annotated 169 ultrasound cases from an unlabeled dataset provided by the competition organizers. These data were also incorporated into the training set to enhance model performance.

To evaluate the model's robustness and generalization ability, an 8-fold stratified cross-validation approach was employed for dataset partitioning. Stratification was performed based on two dimensions—fetal presentation (cephalic, breech, transverse) and scanning laterality (left, right)—to maintain proportional distribution of different categories across all folds. Each fold was then alternately used as the validation set while the remaining seven folds served as the training set. This strategy not only maximized the utilization of limited clinical data but also ensured that each model evaluation covered diverse anatomical and scanning scenarios.

During preprocessing, to mitigate variations caused by different ultrasound devices and scanning parameters, all DICOM pixel values were first linearly normalized to the [0,1] range. Images with original resolutions ranging from 512×512 to 768×768 were resized proportionally using bilinear interpolation and center-cropped to a uniform 512×512 resolution to eliminate edge noise interference in model training. A PyTorch-based augmentation pipeline was applied to each image, including random color jitter (brightness, contrast, and saturation adjustments with an intensity of 0.4 each, and hue perturbation of 0.1), Gaussian blur , and a 50% probability of horizontal or vertical flipping. Additionally, random rotation (within ±15°) and translation (up to ±10 pixels) were applied. All

geometric and pixel-level transformations were synchronized with the original annotations via affine transformation matrices to ensure consistency between model inputs and target outputs.

To generate the target heatmaps for training, each ground truth keypoint was treated as the center of a 2D Gaussian distribution in the 64×64 low-resolution space. A continuous probability density map was then constructed within the [0,1] range, serving as the Gaussian target. These heatmaps, corresponding one-to-one with the network outputs, were used for subsequent regularization loss calculations, enabling the model to learn not only coordinate regression accuracy but also spatial consistency in probability distribution.

3.2 Evaluation Metrics

This experiment selects the following 4 evaluation metrics: Mean Squared Error (MSE), Mean Absolute Error (MAE), Average Point Distance (APD), and AOP_MAE.

Mean Squared Error (MSE): Quantifies the average squared difference between predicted and true values, emphasizing penalties for larger errors.

$$\mathrm{MSE} = \frac{1}{n}\sum_{i=1}^{n}(y_i - \hat{y}_i)^2$$

Mean Absolute Error (MAE): Measures the average absolute deviation between predicted and true values, more robust than MSE.

$$\mathrm{MAE} = \frac{1}{n}\sum_{i=1}^{n}|y_i - \hat{y}_i|$$

Average Point Distance (APD): Evaluates the average Euclidean distance between corresponding points in trajectory or spatial point prediction.

$$\mathrm{APD} = \frac{1}{n}\sum_{i=1}^{n}\sqrt{(x_i - \hat{x}_i)^2 + (y_i - \hat{y}_i)^2}$$

Error Metric for Anchor Offset Probability (AOP_MAE): Measures statistical error of bounding box offset in object detection or tracking.

$$\mathrm{AOP\_MAE} = \frac{1}{K}\sum_{k=1}^{K}\left|\mathrm{AOP}_{\mathrm{true}}^{(k)} - \mathrm{AOP}_{\mathrm{pred}}^{(k)}\right|$$

Additionally, we employed the Optuna hyperparameter optimization framework to systematically tune critical model parameters, including the learning rate,

regularization factor λ, Gaussian heatmap standard deviation σ, and gradient clipping norm `clip_grad_norm`. The optimization objective was to minimize the combined validation loss $\ell = \ell_{euc} + \lambda \cdot \ell_{reg}$. After 50 trials, the optimal hyperparameter set was determined as: learning rate = $5.352137134504593 \times 10^{-5}$, $\lambda = 3.732135829310275$, $\sigma = 2.3022166650613793$, and `clip_grad_norm` = 0.7221588712934679. This automated tuning process significantly enhanced model stability and convergence efficiency.

3.3 Experimental Environment and Configuration

This experiment was conducted on a Windows 11 Professional operating system. The hardware configuration consists of:

- NVIDIA GeForce RTX 4090 GPU with 24GB VRAM
- Intel Core i9-13900K processor (24 cores, 32 threads, base frequency 3.00GHz)
- 64GB DDR5 RAM
- 2TB NVMe SSD

The GPU's Tensor Core architecture and 24GB VRAM provide powerful parallel computing capabilities for model training, while the processor's 24 physical cores efficiently support data preprocessing workflows.

The experiment is based on a PyTorch framework implementing the DSNT keypoint detection model. Parameter updates were performed using the Adam optimizer with an initial learning rate of 5.35×10^{-5}. During training, a dynamic learning rate scheduling strategy was employed: when the validation loss showed no improvement for 10 consecutive epochs, the learning rate decayed to 50% of its current value. This approach effectively balances convergence speed with training stability.

3.4 Loss Function Analysis

In deep learning keypoint detection tasks, the DSNT loss function achieves end-to-end coordinate regression through a dual-path supervision mechanism. Its core consists of Euclidean distance loss and probability distribution regularization loss, mathematically expressed as:

$$\ell = \ell_{\text{euc}} + \lambda \cdot \ell_{\text{reg}}$$

where λ is the tunable regularization factor (set to 3.73 in the code). The Euclidean loss directly constrains the normalized coordinate space: first, the true pixel coordinates are linearly transformed to the interval $[-1, 1]$, then the mean squared error between predicted coordinates and transformed true coordinates is calculated:

$$\ell_{\text{euc}} = \frac{1}{N} \sum_{i=1}^{N} \|\mathbf{C}_i - \mathbf{C}_{\text{gt},i}^{\text{norm}}\|^2$$

The regularization loss aligns the heatmap distribution via Kullback-Leibler (KL) divergence: based on true coordinates, a Gaussian target heatmap is generated where the heat value for each keypoint channel is determined by a 2D Gaussian function:

$$H_{\text{gt}}(x, y) = \exp\left(-\frac{(x - x_c)^2 + (y - y_c)^2}{2(\sigma/\hbar)^2}\right)$$

This loss computes the logarithmic probability difference between the predicted heatmap $\hat{H}$ and the target heatmap:

$$\ell_{\text{reg}} = \text{KL}(H_{\text{gt}} \| \hat{H}) = \frac{1}{NK} \sum_{i=1}^{N} \sum_{k=1}^{K} H_{\text{gt}}^{(i,k)} \left(\log H_{\text{gt}}^{(i,k)} - \log \hat{H}^{(i,k)}\right)$$

This dual-path design enables the model to simultaneously learn precise coordinate localization and heatmap representations conforming to spatial probability distributions, significantly enhancing regression stability.

4 Results and Analysis

4.1 Analysis of Loss and Pixel Error

In the eight-fold training, the seventh fold achieved the best performance. The analysis of loss and pixel error for the seventh fold is shown in the figure below (Fig. 3). Over the course of 66 training epochs, both the training loss and validation loss exhibited a consistent declining trend. The training loss decreased steadily from an initial value of approximately 1840 to around 1302, while the validation loss showed a similar downward trend, starting from about 1556 and eventually converging near 1364. This synchronous reduction in both losses indicates effective learning without signs of overfitting. The validation pixel error demonstrated significant improvement, dropping from nearly 20 pixels to below 10 pixels, reflecting enhanced localization accuracy of the model. It is worth noting that the learning rate was reduced twice during training (at epochs 44 and 58), which effectively contributed to the stabilization of the loss curves.

4.2 Analysis of Experimental Results

This study systematically evaluated the performance of three models on both validation and test sets, with the results shown in Table 1. The baseline U-Net model, an improved model with multi-scale fusion (Deep-UNet), and a model incorporating both multi-scale fusion and DSNT modules (DSNT-DeepUNet). The evaluation considered four key performance metrics: Mean Squared Error (MSE), Mean Absolute Error (MAE), Average Point Distance (APD), and Average Angular Deviation (AOP).

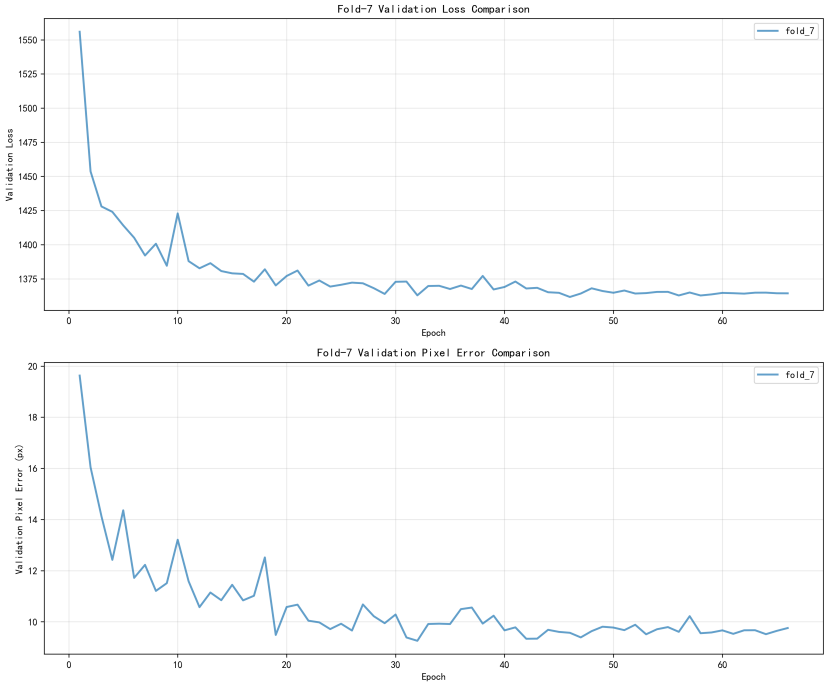

Fig. 3. Training and validation loss trends

Experimental results demonstrate that the DSNT-DeepUNet model, integrating multi-scale features and DSNT modules, achieved the best overall performance on both datasets. Specifically, on the validation set, DSNT-DeepUNet attained the lowest MSE (178.0195) and AOP (4.7906), with its MAE (9.9812) and APD (15.5995) outperforming the baseline U-Net (MSE: 632.6917, MAE: 13.0450, APD: 20.2612, AOP: 8.0364) while approaching or slightly trailing Deep-UNet (MSE: 219.8383, MAE: 9.9350, APD: 15.5393, AOP: 5.5850).

On the more challenging test set, DSNT-DeepUNet's advantages were more pronounced, achieving the best values across all four metrics: MSE (180.9396), MAE (9.4630), APD (14.7712), and AOP (4.7005). In comparison, Deep-UNet's performance on the test set (MSE: 273.8074, MAE: 9.8526, APD: 15.3954, AOP: 5.9237) remained significantly better than baseline U-Net (MSE: 887.5399, MAE: 14.0043, APD: 21.8273, AOP: 8.3727) but showed a comprehensive gap compared to DSNT-DeepUNet.

Overall, the multi-scale fusion strategy alone (Deep-UNet) effectively enhanced model performance, while the additional integration of DSNT modules (DSNT-DeepUNet) brought more substantial accuracy improvements. This was particularly evident in position (APD) and angular (AOP) prediction accuracy, ultimately establishing DSNT-DeepUNet as the top-performing model architecture in this study.

Table 1. Experimental Results

Dataset	Model	MSE	MAE	APD	AOP
Validation	Unet	632.6917	13.0450	20.2612	8.0364
Validation	Deep-UNet	219.8383	**9.9350**	**15.5393**	5.5850
Validation	DSNT-DeepUNet	**178.0195**	9.9812	15.5995	**4.7906**
Test	Unet	887.5399	14.0043	21.8273	8.3727
Test	Deep-UNet	273.8074	9.8526	15.3954	5.9237
Test	DSNT-DeepUNet	**180.9396**	**9.4630**	**14.7712**	**4.7005**

5 Model Tooling

Our team has operationalized the model described in the paper and developed a real-time automated tool for predicting the Angle of Progression (AoP) from intrapartum ultrasound images. When a single ultrasound frame is uploaded, the system automatically detects and precisely annotates three key anatomical landmarks—the two endpoints of the pubic symphysis (PS1, PS2) and the fetal head point (FH1)—within a standardized 512×512-pixel image coordinate system, and computes the AoP from these spatial coordinates. The system also generates a visualization showing the annotated landmarks, the connecting lines, and the angle markers, and returns this visual output together with the numerical AoP result instantly. This end-to-end, fully automated

pipeline eliminates manual intervention, substantially shortens processing time, and reduces annotation variability, thereby providing an efficient and consistent method for clinical assessment of fetal head progression. The web link is: http://61.178.78.27:50210/aop_prediction/, as illustrated in Fig. 4.

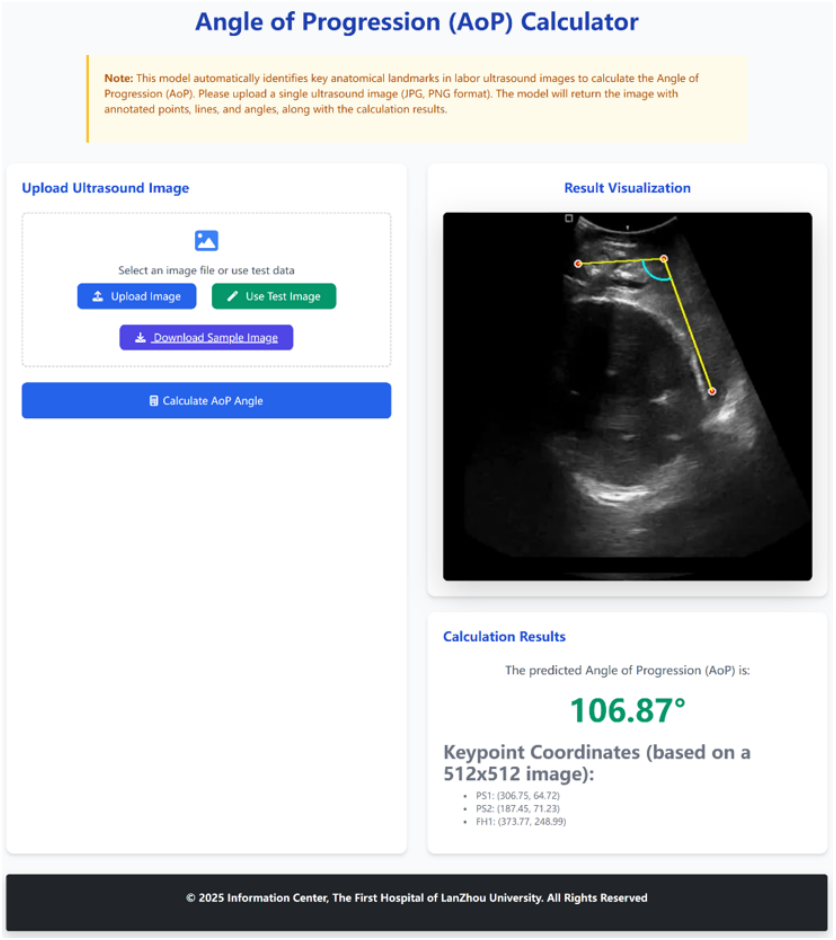

Fig. 4. Model Tooling

6 Conclusion

This study addresses the challenges of significant inter-observer variability and time-consuming manual annotation of key anatomical landmarks in intrapartum ultrasound monitoring by proposing a deep learning-based coordinate prediction

model named DSNT-DeepUNet. The model integrates a deep U-Net backbone with a Differentiable Spatial to Numerical Transform (DSNT) module, achieving end-to-end mapping from raw ultrasound images to keypoint coordinates. This approach effectively mitigates quantization errors and non-differentiability issues inherent in traditional methods, significantly improving localization accuracy and inference efficiency.

By incorporating a multi-task loss function, the model optimizes coordinate prediction accuracy while enhancing regularization constraints on the heatmap probability distribution, thereby striking a balance between spatial consistency and numerical regression stability. An eight-fold stratified cross-validation strategy and dynamic data augmentation methods were employed, substantially improving the model's generalization capability and robustness. Experimental results demonstrate that DSNT-DeepUNet outperforms both the baseline U-Net and the Deep-UNet model with only multi-scale fusion modules on an independent test set, particularly in reducing keypoint distance errors and improving the prediction accuracy of the angle of progression (AoP).

Furthermore, the model has been operationalized into a web-based system capable of real-time ultrasound image processing and automatic AoP calculation, providing a consistent, efficient, and non-invasive auxiliary tool for clinical practice. The system is publicly accessible for testing and demonstrates considerable potential for clinical application.

Although the proposed model achieved sixth place in the open challenge with no missed detections, indicating high reliability, it still exhibits certain errors when handling complex anatomical structures such as tangent points. Future work will focus on optimizing the model architecture through the incorporation of attention mechanisms and multi-modal information to improve recognition capability for challenging cases, while expanding the dataset to enhance generalization across diverse devices and populations.

Acknowledgements The authors of this paper declare that the method implemented for participation in the Landmark Detection Challenge for Intrapartum Ultrasound Measurement (Intrapartum Ultrasound Grand Challenge 2025) did not employ any pre-trained models or external datasets beyond those provided by the organizers. We extend our gratitude to all data contributors for making the ultrasound images publicly accessible.

This work was supported by the Education Development Foundation of Lanzhou University.

References

1. Aghasizade, M., Kiyoumarsioskouei, A., Hashemi, S., Torabinia, M., Caprio, A., Rashid, M., Xiang, Y., Rangwala, H., Ma, T., Lee, B., et al.: A coordinate-regression-based deep learning model for catheter detection during structural heart interventions. Appl. Sci. **13**, 7778 (2023). https://doi.org/10.3390/app13137778

2. Bai, J., Khobo, I., Lu, Y., Ni, D., Yaqub, M., Lekadir, K., Ma, J., Li, S.: Landmark detection challenge for intrapartum ultrasound measurement meeting the actual clinical assessment of labor progress. In: Medical Image Computing and Computer Assisted Intervention 2025 (MICCAI) (2025). https://doi.org/10.5281/zenodo.15081529
3. Cano-Espinosa, C., González, G., Washko, G.R., Cazorla, M., Estépar, R.S.J.: Biomarker localization from deep learning regression networks. IEEE Trans. Med. Imaging **39**(6), 2121–2132 (2020). https://doi.org/10.1109/TMI.2020.2965486
4. Hu, N.N., He, Y.F.: Application value of ultrasound in labor. Chinese Journal of Clinical Research **35**(05), 721–725 (2022). https://doi.org/10.13429/j.cnki.cjcr.2022.05.027
5. Huo, G.G., Chang, Y., Chen, X.: Value of transperineal ultrasound measurement of angle of progression and head-perineum distance in predicting delivery mode and duration in the second stage of labor. Chinese Journal of Practical Gynecology and Obstetrics **37**(03), 373–377 (2021). https://doi.org/10.19538/j.fk2021030123
6. Li, P.M., Wu, Z.M., Yao, L.M.: Effects of intrapartum ultrasound monitoring of fetal heart rate and fetal position combined with new labor stage time limit management on labor process and pregnancy outcome in advanced age parturients. Medical Innovation of China **20**(21), 147–152 (2023)
7. Nibali, A., He, Z., Morgan, S., Prendergast, L.: Numerical coordinate regression with convolutional neural networks. arXiv preprint (2018)
8. Ronneberger, O., Fischer, P., Brox, T.: U-net: Convolutional networks for biomedical image segmentation. In: MICCAI (2015)
9. Yang, J.: Clinical study on the evaluation of delivery mode by intrapartum ultrasound combined with vaginal examination. Guide of China Medicine **22**(09), 55–58 (2024). https://doi.org/10.15912/j.issn.1671-8194.2024.09.016
10. Yue, Z.Z., Wang, J.Y., Ni, Y., et al.: Predictive value of transperineal ultrasound measurement of angle of progression and head-perineum distance in the second stage of labor for delivery mode and duration. Journal of Clinical and Experimental Medicine **21**(11), 1196–1200 (2022)
11. Zhang, Q.J., Yan, J.Y.: Application of intrapartum ultrasound in labor management. Chinese Journal of Practical Gynecology and Obstetrics **40**(02), 142–147 (2024). https://doi.org/10.19538/j.fk2024020104