Anatomy-Aware Standard Plane Localization in 3D Ultrasound with Geometric Constraints

Jianlong Nie[1,2], Xi Ouyang[2], Yuyu Zhou[4], Qi Wan[5], Zhong Xue[2], Xiaohuan Cao[2(✉)], and Dinggang Shen[1,2,3(✉)]

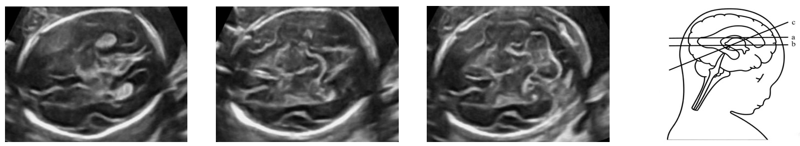
[1] School of Biomedical Engineering & State Key Laboratory of Advanced Medical Materials and Devices, ShanghaiTech University, Shanghai 201210, China
Dinggang.Shen@gmail.com
[2] Shanghai United Imaging Intelligence Co., Ltd., Shanghai 200030, China
xiaohuan.cao@uii-ai.com
[3] Shanghai Clinical Research and Trial Center, Shanghai 201210, China
[4] Shanghai United Imaging Healthcare Co., Ltd., Shanghai 201807, China
[5] Wuhan United Imaging Healthcare Co., Ltd., Wuhan 430206, China

Abstract. Standard plane localization is a critical prerequisite in ultrasound scanning for disease diagnosis. Although three-dimensional (3D) ultrasound acquires volumetric data superior to two-dimensional (2D) imaging, manually localizing multiple standard planes simultaneously within 3D volume remains labor-intensive and prone to inter-observer variability. To address this, we propose an anatomy-aware multi-task learning framework with structural prior integration and inter-plane geometric constraints to jointly localize multiple standard planes. To validate the effectiveness and generalizability of our method, we curated a multi-center dataset comprising 800 fetal brain volumes. Our approach outperforms state-of-the-art methods, achieving localization errors of 1.09 mm / 3.85°, 1.11 mm / 3.93° and 1.20 mm / 4.03° on three key standard planes of the fetal brain (i.e., the transventricular plane, the transthalamic plane and the transcerebellar plane, respectively). These findings underscore the method's potential to enhance clinical ultrasound scanning workflows.

Keywords: 3D ultrasound · Fetal brain · Standard plane localization · Multi-task learning · Geometric constraints

1 Introduction

Fetal brain ultrasound, as a non-invasive imaging modality, plays a pivotal role for assessing neurodevelopment and detecting anomalies [3]. Clinical guidelines require acquisition of several standard planes (SPs), including transventricular (TV), transthalamic (TT), and transcerebellar (TC) planes [6] (see Fig. 1). Each of these planes contains key anatomical structures [19]. Although 3D ultrasound enables full brain volume capture and retrospective SP extraction, automated SP localization remains challenging due to poor image quality, large 3D searching space, and high inter-subject anatomical variability. Robust and precise SP



(a) TV plane (b) TT plane (c) TC plane (d) Fetal brain SPs

Fig. 1: Examples of standard planes in 3D fetal brain ultrasound.

localization is essential to boost diagnostic efficiency and to reduce operator dependency.

Existing approaches for standard plane localization in 3D ultrasound can be broadly categorized into two paradigms. The first paradigm treats the task as 2D slice classification, where densely sampled planes are classified using handcrafted features [13] or deep convolutional neural networks [1, 2]. To improve the performance, recent studies have incorporated attention mechanisms [9, 15] and multi-task learning [4, 5]. Additionally, anatomical prior knowledge has been leveraged to enrich feature representations [12, 23] or to constrain the searching space [14, 18]. However, 2D-based approaches ignore 3D geometric consistency and are sensitive to sampling strategies.

The second paradigm directly estimates continuous plane parameters from the volume [10]. These approaches often employ anatomical landmarks or reference planes to initially restrict the solution space firstly [16, 21]. Reinforcement learning [7, 22, 25] and diffusion models [8] have also been explored to iteratively refine plane estimates. Such multi-stage strategies or iterative optimization are struggling with error propagation, initialization bias and computational cost.

Despite progress, current methods still struggle to achieve robust and precise SP localization that fully exploits the rich anatomical and spatial context embedded within 3D ultrasound volumes. Critically, each SP is defined by specific anatomical structures, thus integrating detailed structural priors holds considerable promise for improving localization accuracy. Moreover, multiple SPs derived from the same fetal organ exhibit well-defined geometric relationships (as shown in Fig. 1(d)). However, Current approaches typically predict them independently, thereby ignoring valuable inter-plane constraints and forgoing opportunities for shared representation learning.

To address these limitations, we propose a unified framework to jointly estimate the parameters of multiple SPs by explicitly integrating anatomical structure information. The main contributions are summarized as follows:

1. We formulate multi-plane localization as a multi-task learning problem, enabling shared representation learning across SPs and thereby enhancing both computational efficiency and inter-plane consistency.
2. We incorporate segmentation mask of key anatomical structures not only to guide image feature extraction but also to serve as direct geometric cues for plane parameter prediction.

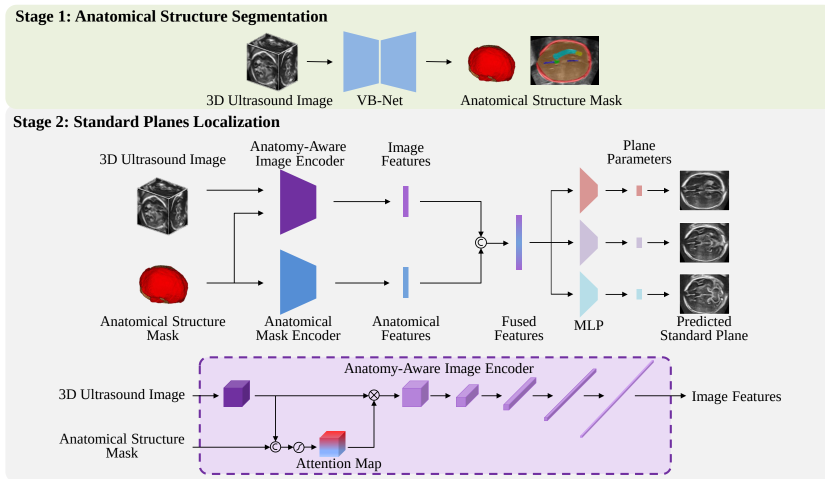


Fig. 2: Overview of our proposed anatomy-aware multi-plane localization framework. The segmentation network (Stage 1) produces an anatomical mask, which is fused with the ultrasound volume for joint multiple plane localization (Stage 2).

3. We introduce a novel inter-plane geometric constraint loss to explicitly encodes the known spatial relationships among SPs to regularize the joint prediction process.

2 Method

Our method comprises two sequential stages, as illustrated in Fig. 2. In the first stage, we use a 3D segmentation network to delineate key anatomical structures within the fetal brain ultrasound volume, producing a voxel-wise anatomical mask. In the second stage, we introduce an anatomy-aware localization module for joint localization of the three standard planes of the fetal brain: TV, TT, and TC. Specifically, both the 3D ultrasound volume and the predicted anatomical mask are processed by a multi-task regression network, which simultaneously predicts the parameters of all three planes. This unified design explicitly integrates structural priors to guide accurate and geometrically consistent plane localization.

2.1 Anatomical Structure Segmentation

Standard planes in 3D ultrasound are defined by the presence and spatial arrangement of specific anatomical structures. To encode this prior knowledge, we employ a VB-Net [17] to segment the input 3D ultrasound volume and generate voxel-wise labels of key anatomical structures in the fetal brain. We adopt the original network architecture while optimizing the training configuration for our specific dataset and task. To accommodate our ultrasound dataset, all input volumes are resampled to isotropic spacing of $1.0 \times 1.0 \times 1.0$ mm^3 using linear interpolation. All other training settings follow the coarse model configuration described in [17]. The resulting segmentation mask provides structured anatomical priors for informing the subsequent plane localization stage.

2.2 Anatomy-Aware Standard Plane Localization

To focus computation on anatomically relevant regions, we first compute a bounding box to enclose all foreground voxels in the anatomical mask and apply this cropping jointly to both the original ultrasound volume and the mask. Then, the cropped inputs are processed by two parallel encoders: (1) an anatomy-aware image encoder and (2) an anatomical mask encoder. Their outputs are fused to enable multi-task regression of the three standard plane parameters.

Anatomy-Aware Image Encoder This encoder processes the cropped ultrasound volume in conjunction with the anatomical mask via a spatial attention mechanism. Initial image features $\mathbf{f}_\mathrm{i}$ are extracted using a 3D $3 \times 3 \times 3$ convolution with batch normalization (BN) and rectified linear unit (ReLU). The initial image features and mask are concatenated and passed through a convolutional subnetwork ($3 \times 3 \times 3$ conv, ReLU, $1 \times 1 \times 1$ conv, and sigmoid) to generate an attention map $\mathbf{A}$. Then, this attention map is element-wise multiplied with the initial image features $\mathbf{f}_\mathrm{i}$. Subsequently, four downsampling steps are applied, each using $2 \times 2 \times 2$ max-pooling (stride 2) and a basic block ($3 \times 3 \times 3$ conv, BN, and ReLU) that doubles feature channels. The final image features $\hat{\mathbf{f}}_\mathrm{i}$ are flattened. This design captures increasingly abstract and context-rich image representations.

Anatomical Mask Encoder The segmentation mask encodes rich spatial relationships among anatomical structures essential for defining standard planes. To capture inter-structure spatial relationships, we employ a mask encoder with the same architecture as the image encoder but without the attention module, extracting anatomical features $\hat{\mathbf{f}}_\mathrm{m}$ directly from the segmentation mask. The final anatomical features $\hat{\mathbf{f}}_\mathrm{m}$ are flattened.

Feature Fusion and Plane Regression The encoded image features $\hat{\mathbf{f}}_\mathrm{i}$ and anatomical features $\hat{\mathbf{f}}_\mathrm{m}$ are concatenated and fed into a two-layer fully-connected MLP (256 units, ReLU, and 256 units). For standard plane $k \in \{\mathrm{TV}, \mathrm{TT}, \mathrm{TC}\}$, the network outputs a 4-dimensional vector: its first three components are L2-normalized to yield the unit plane normal $\boldsymbol{n}^k$, and the last component gives the signed distance d^k from the origin. Here the origin is set as the center of an ultrasound volume. The multi-task formulation enables shared representation learning across all three planes.

2.3 Loss Function

The total loss function consists of two components: (1) three per-plane regression losses ($\mathcal{L}_{\text{TV}}$, $\mathcal{L}_{\text{TT}}$, $\mathcal{L}_{\text{TC}}$) that penalize errors in normal orientation and distance offset; and (2) two inter-plane geometric constraint losses ($\mathcal{L}_{\text{parallel}}$, $\mathcal{L}_{\text{coplanar}}$) that encode consistent anatomical relationships among the standard planes.

Single Plane Loss Each standard plane in 3D space is represented by $ax + by + cz = d$. When the coefficients are normalized such that $a^2 + b^2 + c^2 = 1$, the plane's unit normal is given by $\boldsymbol{n} = (a, b, c)$, and the signed distance from the origin to the plane is d.

The single plane loss combines an angular term and a distance term: (1) to emphasize small-angle errors, we use the squared magnitude of the cross product between predicted and ground-truth normals; (2) for the distance offset, we adopt the L_1 loss. The loss for a single plane is formulated as:

$$\mathcal{L}_k = \alpha \cdot \|\boldsymbol{n}_{\text{E}}^k \times \boldsymbol{n}_{\text{G}}^k\|^2 + \beta \cdot |d_{\text{E}}^k - d_{\text{G}}^k|, \tag{1}$$

where $\boldsymbol{n}_{\text{E}}$ and $\boldsymbol{n}_{\text{G}}$ are the predicted and ground-truth unit normals, d_{E} and d_{G} are the corresponding signed distances, $k \in \{\text{TV}, \text{TT}, \text{TC}\}$, and α, β are balancing weights.

Inter-Plane Geometric Constraint Loss Despite inter-subject anatomical variability, the spatial configuration of fetal brain structures imposes consistent geometric relationships among the standard diagnostic planes. Specifically: (1) the TV and TT planes are clinically required to be nearly parallel; and (2) the TC plane is acquired with a slight caudal tilt relative to TV/TT, but its orientation remains anatomically constrained such that the normals of all three planes approximately lie in a common anatomical plane. To enforce these priors, we introduce two geometric constraint losses.

$$\mathcal{L}_{\text{parallel}} = \max\left(0, \cos\left(\tau_{\text{p}}\right) - \boldsymbol{n}_{\text{E}}^{\text{TV}} \cdot \boldsymbol{n}_{\text{E}}^{\text{TT}}\right), \tag{2}$$

where $\boldsymbol{n}_{\text{E}}^{\text{TV}}$ and $\boldsymbol{n}_{\text{E}}^{\text{TT}}$ are the normalized predicted plane normals for the TV and TT planes, respectively, and τ_{p} is the maximum allowable angle deviation in radian.

$$\mathcal{L}_{\text{coplanar}} = \max\left(0, |\boldsymbol{n}_{\text{E}}^{\text{TV}} \cdot \left(\boldsymbol{n}_{\text{E}}^{\text{TT}} \times \boldsymbol{n}_{\text{E}}^{\text{TC}}\right)| - \tau_{\text{c}}\right), \tag{3}$$

where $\boldsymbol{n}_{\text{E}}^{\text{TV}}$, $\boldsymbol{n}_{\text{E}}^{\text{TT}}$ and $\boldsymbol{n}_{\text{E}}^{\text{TC}}$ denote the normalized predicted plane normals for the TV, TT, and TC planes respectively, and τ_{c} is a coplanarity error threshold.

The total loss is the sum of three per-plane losses and two geometric regularizers:

$$\mathcal{L}_{\text{total}} = \lambda_1 \cdot (\mathcal{L}_{\text{TV}} + \mathcal{L}_{\text{TT}} + \mathcal{L}_{\text{TC}}) + \lambda_2 \cdot (\mathcal{L}_{\text{coplanar}} + \mathcal{L}_{\text{parallel}}), \tag{4}$$

where λ_1 and λ_2 balance between three per-plane losses and two inter-plane geometric constraint losses.

3 Experimental Results

3.1 Data Collection and Cohort Characteristics

We constructed a multi-center 3D fetal brain ultrasound dataset comprising 800 volumes from 780 pregnant women in the second trimester. All raw volumes were anonymized, and those with severe artifacts, poor image quality, or incomplete coverage were excluded by experienced sonographers. Data were acquired across four clinical sites. Maternal age ranged from 18 to 46 years.

Annotation was performed using 3D Slicer, where both anatomical structure segmentation mask and standard planes were manually delineated by two trained annotators under sonographer supervision. A two-stage quality control procedure was applied, with ambiguous cases reviewed by senior clinicians to yield consensus ground truth. The dataset was partitioned by patient into training (580), validation (70), and testing (150) sets to prevent data leakage.

3.2 Implementation Details

Our method was implemented in PyTorch and trained on an NVIDIA H20 GPU using the Adam optimizer with a learning rate of 1e-4 and batch size of 96. All volumes were resampled to a uniform spacing of $1.0\times1.0\times1.0$ mm^3 and normalized to [0, 1] using min-max normalization. Data augmentation included random translation, rotation, and scaling with a probability of 0.5, applied to the resampled volumes. Subsequently, volumes were cropped to $128\times128\times128$ voxels centered at the image center. The best model for testing was selected in terms of validation performance. Loss weights in Eq. (1) and Eq. (4) were $\alpha = 100.0$, $\beta = 1.0$, $\lambda_1 = 1.0$, $\lambda_2 = 100.0$. Thresholds $\tau_\mathrm{p} = 0.002$ and $\tau_\mathrm{c} = 0.003$ were determined based on statistical analysis of plane orientations in the training cohort and validated on the validation set to ensure generalizability.

3.3 Result Analysis

We evaluate localization accuracy using three metrics: angular deviation (*Ang*), distance error (*Dis*), and Structural Similarity (SSIM) [20]. *Ang* and *Dis* are defined as:

$$Ang = \arccos\left(\frac{\boldsymbol{n}_\mathrm{E} \cdot \boldsymbol{n}_\mathrm{G}}{\|\boldsymbol{n}_\mathrm{E}\| \cdot \|\boldsymbol{n}_\mathrm{G}\|}\right), \tag{5}$$

$$Dis = |d_\mathrm{E} - d_\mathrm{G}|, \tag{6}$$

where $\boldsymbol{n}_\mathrm{E}$ and $\boldsymbol{n}_\mathrm{G}$ are the predicted and ground-truth unit normals, and d_E, d_G are distances from the volume origin.

Inter- and intra-observer variability analysis on 60 random cases validated the reliability of anotations of SPs. High consistency was observed across all three SPs (inter-observer: 2.41°–2.43° angular deviation, 0.58–0.79 mm distance error; intra-observer: 2.29°–2.36° in angle, 0.61–1.12 mm in distance), confirming a robust, reliable benchmark for all localization experiments.

Table 1: Comparison of different methods on SPs localization. ‡ indicates $p < 0.001$, † indicates $p < 0.01$, * indicates $p < 0.05$.

Plane	Metric	Seg	Regress	UDF	RL-US	ITN	Ours
TV	Ang(°)	$5.04\pm3.68^{\ddagger}$	$5.67\pm7.40^{\dagger}$	$7.05\pm7.41^{\ddagger}$	$8.29\pm4.03^{\ddagger}$	$10.49\pm6.26^{\ddagger}$	**3.85 ± 2.15**
	Dis(mm)	1.42 ± 6.43	$1.46\pm1.35^{\ddagger}$	$1.57\pm1.26^{\dagger}$	$2.12\pm1.81^{\ddagger}$	$2.56\pm2.45^{\ddagger}$	**1.09 ± 0.85**
	SSIM	0.80 ± 0.12	$0.79\pm0.12^{\ddagger}$	$0.77\pm0.12^{\ddagger}$	$0.74\pm0.14^{\ddagger}$	$0.70\pm0.15^{\ddagger}$	**0.82 ± 0.11**
TT	Ang(°)	$5.03\pm3.67^{\ddagger}$	$5.85\pm7.44^{\ddagger}$	$7.37\pm7.92^{\ddagger}$	$7.05\pm3.89^{\ddagger}$	$8.62\pm5.95^{\ddagger}$	**3.93 ± 2.56**
	Dis(mm)	1.57 ± 6.40	1.18 ± 1.08	$1.48\pm1.37^{\ddagger}$	$3.31\pm3.57^{\ddagger}$	$2.79\pm2.81^{\ddagger}$	**1.11 ± 0.93**
	SSIM	$0.81\pm0.11^{*}$	$0.80\pm0.12^{\ddagger}$	$0.78\pm0.12^{\ddagger}$	$0.74\pm0.14^{\ddagger}$	$0.72\pm0.15^{\ddagger}$	**0.82 ± 0.11**
TC	Ang(°)	4.10 ± 2.36	$6.75\pm7.58^{\ddagger}$	$8.73\pm7.46^{\ddagger}$	$7.95\pm3.65^{\ddagger}$	$8.00\pm4.87^{\ddagger}$	**4.03 ± 2.79**
	Dis(mm)	$2.24\pm6.20^{*}$	1.30 ± 1.07	$1.81\pm1.45^{\ddagger}$	$2.19\pm1.74^{\ddagger}$	$2.65\pm3.00^{\ddagger}$	**1.20 ± 1.00**
	SSIM	$0.75\pm0.12^{\ddagger}$	$0.75\pm0.12^{\ddagger}$	$0.72\pm0.13^{\ddagger}$	$0.65\pm0.15^{\ddagger}$	$0.66\pm0.17^{\ddagger}$	**0.78 ± 0.12**

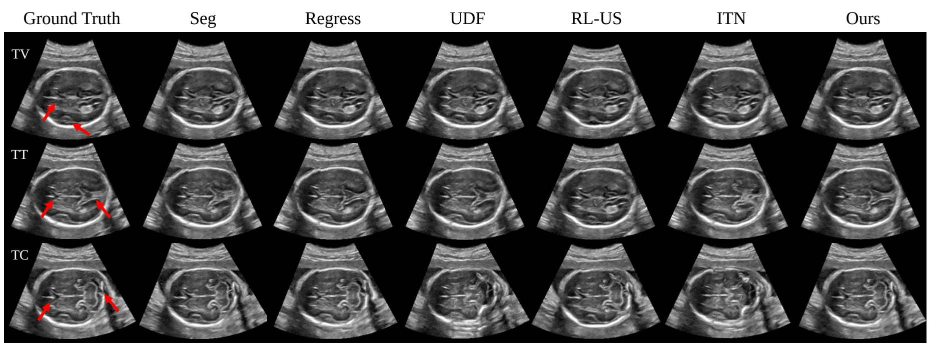


Fig. 3: Visual comparison of SPs localization results of one subject across different methods. Red arrows in the ground truth mark regions with clear inter-method localization discrepancies.

We compare our method with five baselines: (1) segmentation-based (Seg), using anatomical structure centroids to predict planes; (2) regression-based (Regress), employing a 3D CNN trained with MSE loss for explicit plane parameters regression; (3) unsigned distance field-based (UDF) [24], extracting planes from predicted distance fields generated by VB-Net; (4) RL-US [21], a reinforcement learning-based ultrasound standard plane localization method; and (5) ITN [11], an iterative transformation network for standard plane detection.

Table 1 reports quantitative results on TV, TT, and TC planes. Paired t-tests confirm our method achieves state-of-the-art performance across most of the metrics, with statistically significant improvement over all competitors. Segmentation-based method shows competitive performance but larger variance, as segmentation failures under poor visualization propagate to localization outliers. Our method can capture inter-structure relationships for consistent predictions. UDF, RL-US, and ITN show higher errors, suggesting implicit representations or reinforcement strategies struggling with clinical precision. Lower variance indicates robustness of our method.

Table 2: Ablation study for analyzing different components. * means that each column indicates whether the corresponding component is included (✓) or excluded (✗): anatomy-aware attention, anatomical mask encoder, inter-plane geometric constraint loss.

Plane	Metric	Strategies*			
		(✗ ✓ ✓)	(✓ ✗ ✓)	(✓ ✓ ✗)	(✓ ✓ ✓)
TV	Ang(°)	4.652±2.993	4.623±3.121	4.113±2.352	**3.850±2.153**
	Dis(mm)	1.149±0.872	1.136±0.967	1.148±0.867	**1.085±0.849**
	SSIM	0.805±0.116	0.807±0.118	0.814±0.114	**0.816±0.110**
TT	Ang(°)	4.617±3.004	4.636±3.111	4.250±2.918	**3.929±2.561**
	Dis(mm)	1.112±1.017	1.099±1.012	1.171±1.009	**1.108±0.926**
	SSIM	0.814±0.117	0.815±0.117	0.817±0.114	**0.821±0.109**
TC	Ang(°)	5.273±4.136	5.521±4.479	4.646±3.434	**4.031±2.789**
	Dis(mm)	1.296±1.015	1.368±1.000	1.316±1.018	**1.198±0.997**
	SSIM	0.760±0.123	0.758±0.126	0.768±0.123	**0.784±0.120**

The qualitative visualization in Fig. 3 is fully consistent with quantitative findings. The planes generated by Seg and Regress show higher anatomical consistency with the ground truth than those from UDF, RL-US, and ITN. Specifically, UDF, RL-US and ITN exhibit obvious misalignment of key anatomical landmarks, and detail loss in the TT and TC planes. In contrast, our method achieves near-perfect alignment with the ground truth across all planes.

3.4 Ablation Study

We evaluate component contributions by removing the anatomy-aware attention module, anatomical mask encoder, and inter-plane geometric constraint loss (Table 2). Full model performs best across all planes. Ablation results confirm that both the anatomy-aware attention and mask encoder are critical for integrating anatomical structure information, driving the core performance gain of our framework. Inter-plane geometric constraint loss has smaller but consistent impact across all planes.

3.5 Clinical Biometric Evaluation from SPs

Biometric analysis on testing cases shows that measurements from predicted planes are clinically equivalent to ground truth. Absolute errors are minimal: 0.39±0.38 mm for atrial width of the lateral ventricle in the TV plane, 0.43±0.42 mm for biparietal diameter in the TT plane, and 0.83±0.86 mm for cisterna magna depth in the TC plane. Paired t-tests suggests no significant differences across all planes ($p > 0.05$), validating diagnostic reliability of our method.

4 Conclusion

We have presented a two-stage framework for multiple standard planes localization in 3D fetal brain ultrasound: (1) deep learning segmentation of key anatomical structures, and (2) multi-task regression of standard planes. We explicitly leverage anatomical priors via anatomy-aware image encoding, anatomical structure encoding, and inter-plane geometric constraints. Experiments show our approach outperforms state-of-the-art methods, achieving errors of 1.09–1.20 mm and 3.85°–4.03° across three planes. Limitations include dependency on segmentation quality and potential challenges with rare anatomical anomalies. Future work includes end-to-end training, extending to additional standard planes within the fetal brain, generalizing the framework to standard views of other anatomical structures (e.g., fetal heart and abdomen).

Acknowledgments. This work was supported in part by National Natural Science Foundation of China (grant numbers 82441023, 62131015, U23A20295, 82394432), the China Ministry of Science and Technology (STI2030-Major Projects-2022ZD0209000, STI2030-Major Projects-2022ZD0213100), The Key R&D Program of Guangdong Province, China (grant number 2023B0303040001), HPC Platform of ShanghaiTech University, and National Key R&D Program of China (2024YFF1207103).

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

References

1. Abdillahi, M.A., Baker, M.R., Doğru, A., Buyrukoğlu, S.: Enhanced fetal plane classification in ultrasound imaging via prototypical networks and few-shot learning. Journal of Imaging Informatics in Medicine pp. 1–18 (2025)
2. Baumgartner, C.F., Kamnitsas, K., Matthew, J., Smith, S., Kainz, B., Rueckert, D.: Real-time standard scan plane detection and localisation in fetal ultrasound using fully convolutional neural networks. In: International Conference on Medical Image Computing and Computer-assisted Intervention. vol. LNCS 9901, pp. 203–211. Springer (2016)
3. Beyer, T., Bidaut, L., Dickson, J., Kachelriess, M., Kiessling, F., Leitgeb, R., Ma, J., Shiyam Sundar, L.K., Theek, B., Mawlawi, O.: What scans we will read: imaging instrumentation trends in clinical oncology. Cancer Imaging **20**(1), 38 (2020)
4. Cai, Y., Sharma, H., Chatelain, P., Noble, J.A.: Multi-task SonoEyeNet: Detection of fetal standardized planes assisted by generated sonographer attention maps. In: International Conference on Medical Image Computing and Computer-assisted Intervention. vol. LNCS 11070, pp. 871–879. Springer (2018)
5. Cai, Y., Sharma, H., Chatelain, P., Noble, J.A.: SonoEyeNet: Standardized fetal ultrasound plane detection informed by eye tracking. In: 2018 IEEE 15th International Symposium on Biomedical Imaging (ISBI 2018). pp. 1475–1478. IEEE (2018)
6. Fiorentino, M.C., Villani, F.P., Di Cosmo, M., Frontoni, E., Moccia, S.: A review on deep-learning algorithms for fetal ultrasound-image analysis. Medical Image Analysis **83**, 102629 (2023)

7. He, G., Li, Z., Zhu, Z., Han, T., Cao, Y., Chen, C., Huang, Y., Dou, H., Liang, L., Zhang, F., et al.: CLP-Net: An advanced artificial intelligence technique for localizing standard planes of cleft lip and palate by three-dimensional ultrasound in the first trimester. BMC Pregnancy and Childbirth **25**(1), 10 (2025)
8. Huang, Y., Xu, Y., Dou, H., Deng, J., Yang, X., Zheng, H., Ni, D.: Uncertainty-aware diffusion and reinforcement learning for joint plane localization and anomaly diagnosis in 3D ultrasound. In: International Conference on Medical Image Computing and Computer-Assisted Intervention. vol. LNCS 15963, pp. 650–660. Springer (2025)
9. Krishna, T.B., Kokil, P.: Automatic identification of fetal biometry planes from ultrasound images: An assistive tool for healthcare professionals. IEEE Sensors Journal **24**(24), 42365–42372 (2024)
10. Li, Y., Cerrolaza, J.J., Sinclair, M., Hou, B., Alansary, A., Khanal, B., Matthew, J., Kainz, B., Rueckert, D.: Standard plane localisation in 3D fetal ultrasound using network with geometric and image loss. In: Medical Imaging with Deep Learning (2018)
11. Li, Y., Khanal, B., Hou, B., Alansary, A., Cerrolaza, J.J., Sinclair, M., Matthew, J., Gupta, C., Knight, C., Kainz, B., et al.: Standard plane detection in 3D fetal ultrasound using an iterative transformation network. In: International Conference on Medical Image Computing and Computer-Assisted Intervention. vol. LNCS 11080, pp. 392–400. Springer (2018)
12. Ni, D., Li, T., Yang, X., Qin, J., Li, S., Chin, C.T., Ouyang, S., Wang, T., Chen, S.: Selective search and sequential detection for standard plane localization in ultrasound. In: International MICCAI Workshop on Computational and Clinical Challenges in Abdominal Imaging. pp. 203–211. Springer (2013)
13. Rahmatullah, B., Papageorghiou, A., Noble, J.A.: Automated selection of standardized planes from ultrasound volume. In: International Workshop on Machine Learning in Medical Imaging. pp. 35–42. Springer (2011)
14. Ryou, H., Yaqub, M., Cavallaro, A., Roseman, F., Papageorghiou, A., Noble, J.A.: Automated 3D ultrasound biometry planes extraction for first trimester fetal assessment. In: International Workshop on Machine Learning in Medical Imaging. pp. 196–204. Springer (2016)
15. Schlemper, J., Oktay, O., Chen, L., Matthew, J., Knight, C., Kainz, B., Glocker, B., Rueckert, D.: Attention-gated networks for improving ultrasound scan plane detection. arXiv preprint arXiv:1804.05338 (2018)
16. Schmidt-Richberg, A., Schadewaldt, N., Klinder, T., Lenga, M., Trahms, R., Canfield, E., Roundhill, D., Lorenz, C.: Offset regression networks for view plane estimation in 3D fetal ultrasound. In: Medical Imaging 2019: Image Processing. vol. 10949, pp. 907–912. SPIE (2019)
17. Shi, F., Hu, W., Wu, J., Han, M., Wang, J., Zhang, W., Zhou, Q., Zhou, J., Wei, Y., Shao, Y., et al.: Deep learning empowered volume delineation of whole-body organs-at-risk for accelerated radiotherapy. Nature Communications **13**(1), 6566 (2022)
18. Thomas, S., Kausch, L., Kunze, H., Privalov, M., Klein, A., Barbari, J., Martin Vicario, C., Franke, J., Maier-Hein, K.: Computer-assisted contralateral side comparison of the ankle joint using flat panel technology. International Journal of Computer Assisted Radiology and Surgery **16**(5), 767–777 (2021)
19. Wang, Y., Ge, X., Ma, H., Qi, S., Zhang, G., Yao, Y.: Deep learning in medical ultrasound image analysis: A review. IEEE Access **9**, 54310–54324 (2021)

20. Wang, Z., Bovik, A.C., Sheikh, H.R., Simoncelli, E.P.: Image quality assessment: From error visibility to structural similarity. IEEE Transactions on Image Processing **13**(4), 600–612 (2004)
21. Yang, X., Dou, H., Huang, R., Xue, W., Huang, Y., Qian, J., Zhang, Y., Luo, H., Guo, H., Wang, T., et al.: Agent with warm start and adaptive dynamic termination for plane localization in 3D ultrasound. IEEE Transactions on Medical Imaging **40**(7), 1950–1961 (2021)
22. Yang, X., Huang, Y., Huang, R., Dou, H., Li, R., Qian, J., Huang, X., Shi, W., Chen, C., Zhang, Y., et al.: Searching collaborative agents for multi-plane localization in 3D ultrasound. Medical Image Analysis **72**, 102119 (2021)
23. Yang, X., Ni, D., Qin, J., Li, S., Wang, T., Chen, S., Heng, P.A.: Standard plane localization in ultrasound by radial component. In: 2014 IEEE 11th International Symposium on Biomedical Imaging (ISBI). pp. 1180–1183. IEEE (2014)
24. Zhang, C., Lin, G., Yang, L., Li, X., Komura, T., Schaefer, S., Keyser, J., Wang, W.: Surface extraction from neural unsigned distance fields. In: 2023 IEEE/CVF International Conference on Computer Vision (ICCV). pp. 22474–22483. IEEE (2023)
25. Zou, Y., Dou, H., Huang, Y., Yang, X., Qian, J., Zhen, C., Ji, X., Ravikumar, N., Chen, G., Huang, W., et al.: Agent with tangent-based formulation and anatomical perception for standard plane localization in 3D ultrasound. In: International Conference on Medical Image Computing and Computer-assisted Intervention. vol. LNCS 13434, pp. 300–309