This MICCAI paper is the Open Access version, provided by the MICCAI Society. It is identical to the accepted version, except for the format and this watermark; the final published version is available on SpringerLink.

USPose: Improving 3D Fetal Brain Plane Localization for Volumetric Reconstruction from 2D Ultrasound with SO(3)-Equivariance

Jayroop Ramesh*[,1], Mark C. Eid*[,1,2], Pak-Hei Yeung[1,3], João F. Henriques[2], Christian Rupprecht[2], and Ana I.L. Namburete[1]

[1] Oxford Machine Learning in NeuroImaging Lab, University of Oxford, U.K.
[2] Visual Geometry Group, University of Oxford, U.K.
[3] Quantitative Healthcare Analysis (qurAI) Group, University of Amsterdam, The Netherlands
jayroop.ramesh@cs.ox.ac.uk

Abstract. Accurate localization of 2D ultrasound planes within the 3D fetal brain is paramount for automated fetal assessment. However, free-hand ultrasound acquisition is highly operator-dependent, and mapping 2D cross-sections to 3D anatomy remains a major challenge. Traditional pose estimation methods relying on Euler angles or quaternions suffer from discontinuities and singularities, leading to ambiguous localization and poor fetal brain recovery. We propose USPose, a novel SO(3)–equivariant deep learning framework that directly maps ultrasound features to the rotation manifold, capturing pose distributions rather than single-point estimates. By integrating a spatial-parameter-based regression network, USPose enables precise, real-time plane localization and reduces rotational ambiguity. Our approach achieves state-of-the-art localization accuracy and competitive resampled image quality compared with task-specific and foundation-model approaches. Furthermore, USPose yields better poses than prior work for downstream automated 2D-to-3D reconstruction on real clinical ultrasound data, thereby lowering the barrier to interpretation for quantitative volumetric analysis.

Keywords: Fetal neurosonography · Plane localization · Reconstruction

1 Introduction

Fetal brain development follows a precisely timed and highly structured trajectory [19], with disruptions in early pregnancy associated with congenital abnormalities or neonatal mortality [8]. Two-dimensional (2D) ultrasound (US) is the most widely used imaging modality for fetal monitoring due to its real-time acquisition, cost-effectiveness, and portability. However, 2D US is inherently operator-dependent, requiring skilled sonographers to manually acquire standard planes. This dependence poses a challenge, particularly in low- and middle-income countries, where trained personnel may be scarce.

*These authors contributed equally.

During freehand scanning, acquiring diagnostically relevant imaging planes is challenging due to subjective acquisition protocols [24], fetal movement, and acoustic shadowing [25, 26], often leading to inconsistent and ambiguous plane localization [27, 28]. Novice sonographers, in particular, struggle with precise probe positioning, leading to pose misestimations that degrade the reliability of fetal brain assessments [27]. Such errors can lead to incorrect anatomical interpretations, necessitating the reduction of pose ambiguity for plane localization imperative for improving clinical reliability.

Deep learning has advanced automated US plane detection and 2D-to-3D pose estimation, but existing methods lack quantification of pose uncertainty, leaving sonographers unaware of when predictions are unreliable [26]. To address this, we propose USPose, a framework that integrates equivariance and feature mapping from 2D images to the SO(3) rotation manifold to enhance pose estimation accuracy while accounting for pose ambiguity.

Our contributions are as follows:

- USPose incorporates SO(3)-equivariance with a spatial parameter regressor to achieve state-of-the-art performance on localization and image quality metrics.
- USPose addresses pose ambiguity in the downstream task of 2D-to-3D volumetric reconstruction from freehand 2D US videos, supporting sonographers by enabling improved reconstruction fidelity in the absence of ground-truth poses.

2 Related Work

Ultrasound plane localisation. Remarkable strides have been made in standard plane detection, sensorless localization of 2D planes in volumetric coordinate space [1], and assisted navigation to standard planes [5, 18]. Approaches such as SonoNet [1] and ITN [15] perform standard plane detection. PoseGuideNet [18] uses convolutional neural networks (CNNs) to estimate 2D-to-3D spatial correspondences. However, these methods rely on predefined standard planes, making them ill-suited for freehand US scans, where anatomical structures are arbitrarily positioned.

Pose Estimation in Ultrasound. Pose estimation is traditionally treated as a point regression problem, where networks minimize errors relative to ground-truth 3D rotations [23]. The work in [9] performs pose estimation from arbitrarily oriented 2D images in an MRI context. Many methods use Euler angles, which suffer from singularities and discontinuities [30]. Alternative representations, such as quaternions, provide continuous mappings but introduce double coverage, leading to ambiguity. Recent 6D and 9D pose representations mitigate these issues by enforcing orthogonality constraints [7], but maintaining validity is challenging.

Pose Ambiguity and Equivariance. Existing methods focus on pose accuracy but neglect accounting for ambiguous and uncertain frames, which is important for robustness and generalizability. The work outlined in [22] has shown

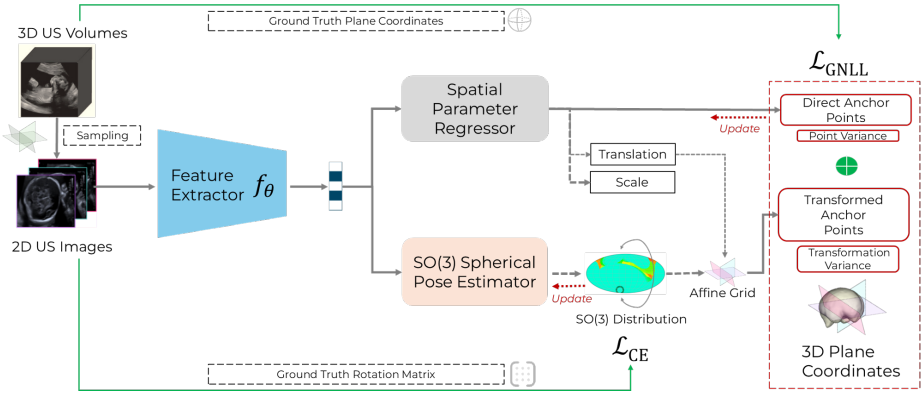

Fig. 1: **Pipeline.** During training, 2D slices sampled from aligned 3D volumes are augmented and used as input. Features extracted by an equivariant backbone and a smaller pretrained network are used to predict plane locations and rotational distributions on the SO(3) manifold.

this limitation within US, where uncertainty quantification through common approaches such as ensembling or Monte Carlo dropout, such as in MRI [9], does not provide meaningful results. Recent work suggests that high-entropy distributions and geodesic distances in SO(3) can address pose ambiguity [12], yet US-specific implementations are lacking.

Our method addresses this by integrating equivariant learning to improve pose reliability in freehand US scanning.

3 Methods

In this section, we will describe our method and problem formulation in detail. The pipeline of our proposed work is presented in Fig. 1.

Problem Definition. Given a 2D ultrasound image, our model estimates the 3D plane in a reference brain atlas to which the image corresponds. During training, given a pose p we sample an image I ($I \in \mathbb{R}^{H \times W}$) from 3D volumes aligned to a 3D anatomical atlas [19] at pose p, The final outputs of the models are i) the predicted 3D plane pose $\hat{p}$, ii) the predicted non-parametric categorical distribution of SO(3) rotations $P(\hat{R}|I)$ and iii) a predicted variance $\hat{\sigma}^2$. We also predict translation t and scaling α parameters in an intermediary step. We denote a ground truth 3D plane pose p with three anchor points defining the plane, which is a set of Cartesian coordinates $p_i \in \mathbb{R}^{3\times3}$ corresponding to the center, bottom right, and bottom left corners [28]. The ground truth rotation in relation to the origin associated with the pose p is denoted by $R \in \mathbb{R}^{3\times3}$. The final 3D predicted rotation matrix $\hat{R}$ is determined by taking the argmax of the predicted distribution $P(\hat{R}|I)$.

3.1 Pose Estimation

The backbone f_θ extracts the feature embedding $\mathcal{Z} = f_\theta(I)$, which is then projected onto S^2 via an orthographic projection ψ, followed by S^2 and SO(3) group equivariant convolutions, respectively. This results in a signal over the Fourier basis of SO(3), which is then queried in the spatial domain with Q discrete points on an equivolumetric grid to create a categorical distribution over rotations.

Mapping from image to sphere. An orthographic projection is generated using a HEALPix grid [3] over a hemisphere to map the 2D feature map $\mathcal{Z}$ to a spherical signal $\psi \in \mathbb{R}^{Q_{2D}}$ where Q_{2D} denotes the number of points distributed across the 2-sphere S^2 (2D surface embedded in 3D space) [12]. The signal ψ now contains spatial information from the original image that maps the 2D pixels to 3D coordinates, supported over one hemisphere due to the fixed perspective of the image plane. When positions in the S^2 grid do not have corresponding pixels in the image, interpolation is applied to smooth transitions with projection dimension D^ψ. The signal ψ is moved to the frequency domain using fast Fourier transforms to express ψ as a linear combination of spherical harmonics. This allows for convolving continuous signals over the S^2 with the SO(3) group. The Fourier series is truncated at frequency L and corrections [12] are used to mitigate sampling errors if the spatial signal has a high frequency [14]. Symmetries in US data can be preserved using equivariant neural networks to improve performance. Convolutions inherently satisfy translation equivalence, but rotation equivariance requires integrating additional components. To account for in-plane rotations (i.e., rotation about the plane normal) [28], $\mathcal{Z}$ undergoes the following step for preserving the 3D symmetry of the representation.

Parameterization with SO(3) Distributions. The Fourier signal ψ is processed with group convolutions that preserve 3D symmetry using filters with global support. Nonlinearities are used between convolution layers by mapping the signal to the spatial domain to apply ReLU, then mapping back to the Fourier domain [4]. A single SO(3)-equivariant group convolution is then applied to produce a linear combination of Wigner-D matrices [14]. These are irreducible representations of the SO(3) group, such that any rotation in 3D space can be expressed using these matrices [12]. The resulting signal ψ can be normalized with a softmax operation to produce a categorical distribution and queried using an equivolumetric HEALPix grid over SO(3) and across Q_{3D} discrete rotations.

Spatial Parameter Regressor. The extracted features $\mathcal{Z}$ are also passed through dedicated fully connected layers, from which scale (α), translation (t), and an SO(3) variance (σ_1^2) are predicted. The top-scoring rotation from the predicted SO(3) distribution is combined with the spatial parameters to construct an affine grid $[\alpha \cdot \hat{R}|\hat{t}]$ with 6DoF, from which a first set of transformed anchor points p_1 is derived. A second set of anchor points $\hat{p}_2$ and variance σ_2^2 is directly regressed independently via another shallow MLP, to provide an inductive bias to the transformed points. The two sets of anchor point estimates are then averaged to produce $\hat{p}$ and σ^2 using the law of total variance [13]. This has the added

benefit of capturing heteroscedasticity [22] for the sets of anchor points, as the predicted variances are sample dependent, allowing for adaptive weighing.

3.2 Loss Formulation

The overall objective of our training is to optimize the predicted anchor points $\hat{p}$ and SO(3) equivariant distribution $P(\hat{R}|I)$, given by Eq. (1). This allows USPose to accurately localize plane locations, predict high probabilities for equivalent rotations, and sample 2D image frames similar to the input from corresponding volumes using the predicted parameters.

$$\mathcal{L}_{\text{total}} = \lambda_{\text{GNLL}}\mathcal{L}_{\text{GNLL}}(\hat{p}, p, \hat{\sigma^2}) + \lambda_{\text{CE}}\mathcal{L}_{\text{CE}}(\hat{R}, R), \tag{1}$$

where $\mathcal{L}_{\text{GNLL}}$ is the Gaussian Negative Log-Likelihood loss, and $\mathcal{L}_{\text{CE}}$ is the cross-entropy loss between $P(\hat{R}|I)$ and the ground truth R. λ_{GNLL} and λ_{CE} are their respective weights at values of 1.0 and 0.5, respectively. $\mathcal{L}_{\text{GNLL}}$ is preferred over $\mathcal{L}_{\text{MSE}}$ based on the findings that explicitly regressing variance terms (to capture uncertainty) alongside point predictions to learn parametric distributions, is more robust [13, 22].

4 Experiments

Implementation details. We input a 2D grayscale image I into the f_θ VGG-19 [22, 28] backbone to obtain features $\mathcal{Z}$. The Spherical Pose Estimator maps features onto S^2 with a HEALPix grid of $Q_{2D} = 20$ points. The dimensionality of the S^2 projection is $D^\psi = 512$, and the maximum frequency level is $L = 6$. We employ a SO(3) equivolumetric HEALPix grid with $Q_{3D} = 2.4M$ points. During training, we randomly resample 64 2D slices from each 3D volume per epoch over a total of 300 epochs, ensuring sampling diversity. Training used an AdamW optimizer with learning rate 1×10^{-4} and a cosine learning rate scheduler. The code to replicate the results of our method and baselines is available at XXXX.

Dataset. The 3D US fetal brain volumes ($160 \times 160 \times 160$ voxels at an isotropic resolution of 0.6 mm^3) were obtained as part of the INTERGROWTH-21st study [21]. The volumes were pre-aligned to a common reference atlas space using [20], and additional details are available in [19].

Training and Testing. 49 3D volumes acquired at 21 gestational weeks were divided into training (35), validation (4), and testing (10) splits. For each 3D volume in each training epoch, a set of 64 2D images (128 during testing) was randomly sampled with respect to the midplane positioned at the central axis of the brain, using sampling parameters and augmentation values used in previous work [22]. This was performed to simulate actual free-hand acquisition by operators when locating target view planes.

Evaluation Metrics. We evaluate the localization accuracy with the Euclidean distance (**ED**, in voxels), Plane angle (**PA**, in radians), and mean squared error (**MSE**, in voxels^2) between predicted and ground truth 3D plane poses

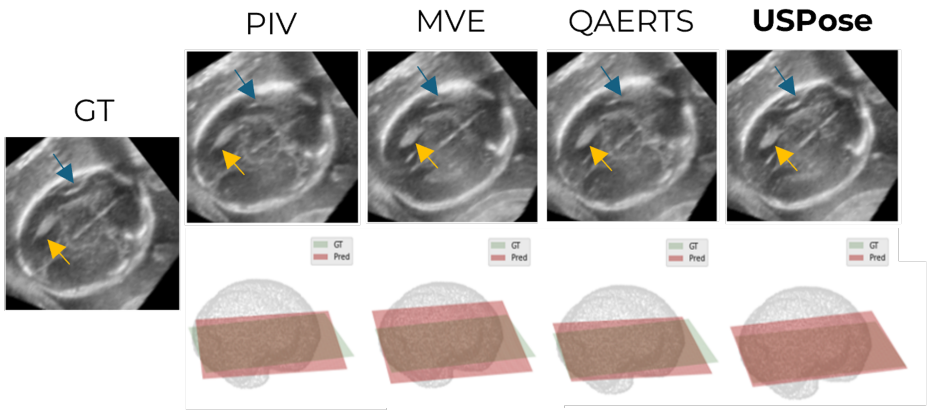

Fig. 2: **Qualitative Comparison.** Given an input frame, we show the predicted planes of different models and images sampled from the predicted plane location. The second row shows a 3D visualization of the planes in the brain atlas. USPose shows relatively better pose recovery for the anatomical structures of the choroid plexus and Sylvian fissure [24], indicated by the blue and yellow arrows respectively.

p [28]. We evaluated the quality of images sampled with predicted and ground truth 3D plane poses p with Normalized Cross-Correlation (**NCC**) and Structural Similarity (**SSIM**) [22]. We measure median rotation error (i.e geodesic distance) (**Rot Err.**) as well as accuracy between predicted and ground truth rotations within the rotation error thresholds $\leq 15°$ (**Acc@15°**) and $\leq 30°$ (**Acc@30°**) [12].

5 Results and Discussion

We compared USPose with (i) recent approaches from the literature, namely PlaneInVol [27, 28], MVE [22] and QAERTS [22], and (ii) relevant foundation models: EchoCare [29], USFM [11], UltraFed [10], FetalClip [17] and MedSAM [16]. We employ linear probing with frozen features for the foundation models to directly regress pose $\hat{p}$ and variance σ^2 for supervision with $\mathcal{L}_{\text{GNLL}}$ loss.

Quantitative results. As shown in Table 1, our proposed USPose consistently outperforms existing approaches across both localization and resampled image fidelity. USPose achieves the lowest ED of 0.110, PA of 0.304, and MSE of 0.100, indicating better plane localization than other methods. We also report the highest NCC and SSIM values of 0.708 and 0.675, respectively, for image quality. Although the median rotation error is the lowest, accuracies within the rotation thresholds are *marginally* lower. This hints that the model did not perform as well on a few outliers. Foundational models are designed for various downstream tasks, but it appears that localization is still too specialized a problem, and ex-

Table 1: **Comparisons.** Bold indicates best. ↑ indicates higher is better and ↓ indicates lower is better. Standard deviations ranged between 0.09-0.20, with all methods showing similar average variability per metric.

Model	ED↓	PA↓	MSE↓	NCC↑	SSIM↑	Acc@15↑	Acc@30↑	Rot Err.
PlaneInVol [27, 28]	0.119	0.325	0.127	0.686	0.606	94.6	**100**	7.67°
MVE [22]	0.111	0.321	0.102	0.688	0.648	**94.9**	**100**	7.50°
QAERTS [22]	0.122	0.328	0.115	0.673	0.629	92.8	99.8	7.75°
EchoCare [29]	0.166	0.482	0.183	0.470	0.500	6.3	36.2	39.53°
USFM [11]	0.204	0.487	0.270	0.394	0.483	5.4	26.0	47.69°
UltraFed [10]	0.199	0.524	0.263	0.430	0.522	4.5	27.4	46.10°
FetalClip [17]	0.183	0.434	0.282	0.426	0.516	4.7	23.1	65.15°
MedSAM [16]	0.196	0.404	0.249	0.426	0.511	4.8	27.7	44.94°
USPose	**0.110**	**0.304**	**0.100**	**0.708**	**0.675**	94.1	99.8	**7.17°**

isting US and medical general-purpose frozen features are insufficient to predict accurate poses. Among them, EchoCare performs relatively better. Additional fine-tuning or distillation-based approaches [2] for foundational models may lead to better results, but are beyond the scope of this work. From Fig. 2, we observe that USPose yields closer overlap between the predicted planes (red) and ground truth planes (green). USPose also attains improved similarities for tissue boundaries and anatomical structures of interest [18, 20, 24].

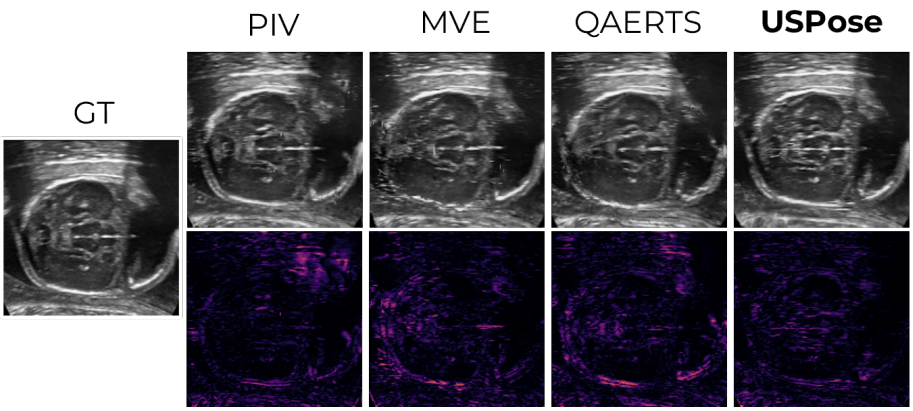

Fig. 3: **Reconstruction Results.** Native 2D freehand US videos without annotations are reconstructed by a SotA Gaussian Splatting model [6] using predicted poses from each of the methods. Top: Novel Image, Bottom: Difference between novel image and the ground truth, displayed using the `inferno` colormap.

Table 2: Quantitative results of reconstruction performance on native freehand videos. The SSIM (↑) scores shown are the average over all frames held out for testing after training for 2000 epochs. Bold indicates best.

Method	Video 1	Video 2	Video 3	Video 4	Avg.	Std.
PIV	0.557	0.623	0.426	0.615	0.555	0.091
MVE	0.651	0.610	0.501	0.624	0.596	0.066
QAERTS	0.677	0.583	0.440	0.578	0.569	0.098
USPose	**0.677**	**0.693**	**0.567**	**0.661**	**0.650**	**0.057**

Table 3: **Ablations.** Mean results for quantitative metrics during ablation tests. Bold indicates best. ↑ indicates higher is better and ↓ indicates lower is better.

Model	ED↓	PA↓	MSE↓	NCC↑	SSIM↑	Acc@15°	Acc@30°	Rot Err.
USPose	**0.110**	**0.304**	0.100	**0.708**	**0.675**	94.1	99.8	**7.17°**
w/o SPR	0.138	0.336	0.152	0.644	0.584	92.3	**100**	7.85°
w/o $\mathcal{L}_{\text{CE}}$	0.117	0.306	0.113	0.625	0.619	92.7	99.9	7.79°
w/ $D^{\psi} = 256$	0.113	0.287	0.105	0.693	0.643	**95.4**	99.8	7.34°
w/ $D^{\psi} = 768$	0.115	0.355	**0.094**	0.703	0.662	94.5	99.9	7.27°
w/ $D^{\psi} = 1024$	0.122	0.317	0.118	0.680	0.620	94.8	**100**	7.29°

Reconstruction assessment. Four freehand US videos (gestational ages between 20 and 21 weeks) had their frames partitioned into a train (80%) and test (20%) set. The pose estimation models were then applied and the predicted poses given to UltraGauss [6], an ultrasound-specific reconstruction model. 300k gaussians were used, and the reconstruction process was run for 2000 epochs. Once a complete 3D reconstruction of the brain was obtained, views of the held-out frames at their estimated poses were sampled. A subset is shown in Fig. 3, with quantitative results for all held-out frames reported in Table 2. Note that these metrics reflect the *combined* error from both pose estimation and reconstruction. Fig. 3 visualizes the use of predicted locations to sample the corresponding slices from the 2D videos, with the difference maps of USPose showing the most similarity with the ground truth.

Ablations. We investigated the effect of changing the proposed components in Table 3. Removing the Spatial Parameter Regressor (w/o SPR) leads to the largest degradation across nearly all metrics. The additional loss appears necessary as it finds high probabilities for the set of equivalently likely orientation rotations if training images have unseen symmetries and pose ambiguity (w/o $\mathcal{L}_{\text{CE}}$ hurts NCC and SSIM). We also examine the orthographic projection dimension D^{ψ} used to parameterise S^2: a smaller dimension ($D^{\psi} = 256$) achieves the highest Acc@15° but sacrifices reconstruction quality, while a larger dimension ($D^{\psi} = 768$) yields the best MSE and competitive NCC/SSIM at the cost of higher planar error.

6 Conclusion

Automated approaches to 3D pose prediction in 2D fetal brain images are affected by variability in image quality, which limits practical deployment. We propose mapping equivariant features from the 2D image domain to the 3D rotation manifold to improve localization by constraining movements in a continuous SO(3) space. Our model USPose achieves superior accuracy compared to both prior task-specific and foundation-model approaches, enabling better pose recovery and 2D-to-3D reconstruction. This has implications for obstetric utility for enhanced fetal growth monitoring by providing sonographers with a reliable avenue for quantitative volumetric analysis.

Acknowledgments. M.C.E is kindly supported by the EPSRC CDT in Autonomous Intelligent Machines and Systems (EP/S024050/1) and Amazon Web Services; J.F.H by the Royal Academy of Engineering (RF/201819/ 18/163); A.I.L.N. by the Bill and Melinda Gates Foundation.

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

References

1. Baumgartner, C.F., Kamnitsas, K., Matthew, J., Fletcher, T.P., Smith, S., Koch, L.M., Kainz, B., Rueckert, D.: Sononet: real-time detection and localisation of fetal standard scan planes in freehand ultrasound. IEEE transactions on medical imaging **36**(11), 2204–2215 (2017)
2. Busbridge, D., Shidani, A., Weers, F., Ramapuram, J., Littwin, E., Webb, R.: Distillation scaling laws. In: Singh, A., Fazel, M., Hsu, D., Lacoste-Julien, S., Berkenkamp, F., Maharaj, T., Wagstaff, K., Zhu, J. (eds.) Proceedings of the 42nd International Conference on Machine Learning. Proceedings of Machine Learning Research, vol. 267, pp. 5977–6045. PMLR (13–19 Jul 2025)
3. Calabretta, M.R., Roukema, B.F.: Mapping on the healpix grid. Monthly Notices of the Royal Astronomical Society **381**(2), 865–872 (2007)
4. Cohen, T., Welling, M.: Group equivariant convolutional networks. In: ICML. pp. 2990–2999. PMLR (2016)
5. Di Vece, C., Cirigliano, A., Lous, M.L., Napolitano, R., David, A.L., Peebles, D., Jannin, P., Vasconcelos, F., Stoyanov, D.: Measuring proximity to standard planes during fetal brain ultrasound scanning. arXiv preprint arXiv:2404.07124 (2024)
6. Eid, M.C., Namburete, A.I., Henriques, J.F.: UltraGauss: Ultrafast gaussian reconstruction of 3D ultrasound volumes. In: ICLR (2026)
7. Geist, A.R., Frey, J., Zobro, M., Levina, A., Martius, G.: Learning with 3d rotations, a hitchhiker's guide to so (3). arXiv preprint arXiv:2404.11735 (2024)
8. Goergen, S.K., Alibrahim, E., Govender, N., Stanislavsky, A., Abel, C., Prystupa, S., Collett, J., Shelmerdine, S.C., Arthurs, O.J.: Diagnostic assessment of foetal brain malformations with intra-uterine mri versus perinatal post-mortem mri. Neuroradiology **61**, 921–934 (2019)
9. Hou, B., Khanal, B., Alansary, A., McDonagh, S., Davidson, A., Rutherford, M., Hajnal, J.V., Rueckert, D., Glocker, B., Kainz, B.: 3-d reconstruction in canonical co-ordinate space from arbitrarily oriented 2-d images. IEEE transactions on medical imaging **37**(8), 1737–1750 (2018)
10. Jiang, Y., Feng, C.M., Ren, J., Wei, J., Zhang, Z., Hu, Y., Liu, Y., Sun, R., Tang, X., Du, J., et al.: From pretraining to privacy: federated ultrasound foundation model with self-supervised learning. npj Digital Medicine **8**(1), 714 (2025)
11. Jiao, J., Zhou, J., Li, X., Xia, M., Huang, Y., Huang, L., Wang, N., Zhang, X., Zhou, S., Wang, Y., et al.: Usfm: A universal ultrasound foundation model generalized to tasks and organs towards label efficient image analysis. Medical image analysis **96**, 103202 (2024)
12. Klee, D.M., Biza, O., Platt, R., Walters, R.: Image to sphere: Learning equivariant features for efficient pose prediction. arXiv preprint arXiv:2302.13926 (2023)
13. Lakshminarayanan, B., Pritzel, A., Blundell, C.: Simple and scalable predictive uncertainty estimation using deep ensembles. NeurIPS **30** (2017)
14. Lee, J., Cho, M.: 3d equivariant pose regression via direct wigner-d harmonics prediction. NeurIPS **37**, 127765–127798 (2025)
15. 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. pp. 392–400. Springer (2018)
16. Ma, J., He, Y., Li, F., Han, L., You, C., Wang, B.: Segment anything in medical images. Nature communications **15**(1), 654 (2024)

17. Maani, F., Saeed, N., Saleem, T., Farooq, Z., Alasmawi, H., Diehl, W., Mohammad, A., Waring, G., Valappi, S., Bricker, L., et al.: Fetalclip: A visual-language foundation model for fetal ultrasound image analysis. arXiv preprint arXiv:2502.14807 (2025)
18. Men, Q., Guo, X., Papageorghiou, A.T., Noble, J.A.: Pose-guidenet: Automatic scanning guidance for fetal head ultrasound from pose estimation. In: MICCAI. pp. 700–710. Springer (2024)
19. Namburete, A.I., Papież, B.W., Fernandes, M., Wyburd, M.K., Hesse, L.S., Moser, F.A., Ismail, L.C., Gunier, R.B., Squier, W., Ohuma, E.O., et al.: Normative spatiotemporal fetal brain maturation with satisfactory development at 2 years. Nature **623**(7985), 106–114 (2023)
20. Namburete, A.I., Xie, W., Yaqub, M., Zisserman, A., Noble, J.A.: Fully-automated alignment of 3d fetal brain ultrasound to a canonical reference space using multitask learning. Medical image analysis **46**, 1–14 (2018)
21. Papageorghiou, A.T., Kennedy, S.H., Salomon, L.J., Altman, D.G., Ohuma, E.O., Stones, W., Gravett, M.G., Barros, F.C., Victora, C., Purwar, M., et al.: The intergrowth-21st fetal growth standards: toward the global integration of pregnancy and pediatric care. American journal of obstetrics and gynecology **218**(2), S630–S640 (2018)
22. Ramesh, J., Dinsdale, N., Yeung, P.H., Namburete, A.I.: Geometric transformation uncertainty for improving 3d fetal brain pose prediction from freehand 2d ultrasound videos. In: MICCAI. pp. 419–429. Springer (2024)
23. Salehi, S.S.M., Khan, S., Erdogmus, D., Gholipour, A.: Real-time deep pose estimation with geodesic loss for image-to-template rigid registration. IEEE transactions on medical imaging **38**(2), 470–481 (2018)
24. Salomon, L., Alfirevic, Z., Berghella, V., Bilardo, C., Chalouhi, G., Costa, F.D.S., Hernandez-Andrade, E., Malinger, G., Munoz, H., Paladini, D., et al.: Isuog practice guidelines (updated): performance of the routine mid-trimester fetal ultrasound scan. Ultrasound in Obstetrics and Gynecology **59**(6), 840–856 (2022)
25. Singla, R., Ringstrom, C., Hu, R., Lessoway, V., Reid, J., Rohling, R., Nguan, C.: Speckle and shadows: ultrasound-specific physics-based data augmentation for kidney segmentation. In: MIDL. pp. 1139–1148. PMLR (2022)
26. Tirindelli, M., Eilers, C., Simson, W., Paschali, M., Azampour, M.F., Navab, N.: Rethinking ultrasound augmentation: A physics-inspired approach. In: MICCAI. pp. 690–700. Springer (2021)
27. Yeung, P.H., Aliasi, M., Haak, M., 21st Consortium, I., Xie, W., Namburete, A.I.: Adaptive 3d localization of 2d freehand ultrasound brain images. In: MICCAI. pp. 207–217. Springer (2022)
28. Yeung, P.H., Aliasi, M., Papageorghiou, A.T., Haak, M., Xie, W., Namburete, A.I.: Learning to map 2d ultrasound images into 3d space with minimal human annotation. Medical Image Analysis **70**, 101998 (2021)
29. Zhang, H., Wu, Y., Zhao, M., Chen, Z., Li, R., Zhu, F., Zhao, H., Yuan, X., Yang, M., Qiu, C., Cong, X., Chen, H., Luan, L., Wong, R.H.L., Liao, H., Graham, C.A., Chang, S., Tao, G., Yi, D., Lei, Z., Navab, N., Ourselin, S., Luo, J., Liu, H., Meng, G.: A fully open and generalizable foundation model for ultrasound clinical applications (2025), `https://arxiv.org/abs/2509.11752`
30. Zhou, Y., Barnes, C., Lu, J., Yang, J., Li, H.: On the continuity of rotation representations in neural networks. In: CVPR. pp. 5745–5753 (2019)