Pose-GuideNet: Automatic Scanning Guidance for Fetal Head Ultrasound from Pose Estimation

Qianhui Men[1], Xiaoqing Guo[1], Aris T. Papageorghiou[2], and J. Alison Noble[1]

[1] Institute of Biomedical Engineering, University of Oxford, Oxford, UK
qianhui.men@eng.ox.ac.uk
[2] Nuffield Department of Women's & Reproductive Health, University of Oxford, Oxford, UK

Abstract. 3D pose estimation from a 2D cross-sectional view enables healthcare professionals to navigate through the 3D space, and such techniques initiate automatic guidance in many image-guided radiology applications. In this work, we investigate how estimating 3D fetal pose from freehand 2D ultrasound scanning can guide a sonographer to locate a head standard plane. Fetal head pose is estimated by the proposed *Pose-GuideNet*, a novel 2D/3D registration approach to align freehand 2D ultrasound to a 3D anatomical atlas without the acquisition of 3D ultrasound. To facilitate the 2D to 3D cross-dimensional projection, we exploit the prior knowledge in the atlas to align the standard plane frame in a freehand scan. A semantic-aware contrastive-based approach is further proposed to align the frames that are off standard planes based on their anatomical similarity. In the experiment, we enhance the existing assessment of freehand image localization by comparing the transformation of its estimated pose towards standard plane with the corresponding probe motion, which reflects the actual view change in 3D anatomy. Extensive results on two clinical head biometry tasks show that Pose-GuideNet not only accurately predicts pose but also successfully predicts the direction of the fetal head. Evaluations with probe motions further demonstrate the feasibility of adopting Pose-GuideNet for freehand ultrasound-assisted navigation in a sensor-free environment.

Keywords: Fetal Pose Estimation · Ultrasound Navigation · Multimodal Image Registration · Probe Guidance

1 Introduction

Ultrasonography requires an operator to be proficient in 3D spatial perception to interpret the entire region of interest from the perspective of a single acoustic window. Clinical sonography workflow usually involves coarsely localizing anatomy of interest, and then ultrasound probe refinement to reach a high-quality anatomical standard plane (SP) for accurate examination. The task is extremely challenging for obstetric clinicians when examining multiple anatomies of a moving fetus [20]. An automatic system combining the two processes - fetal pose localization and navigation to SP (as described in Fig. 1), would greatly
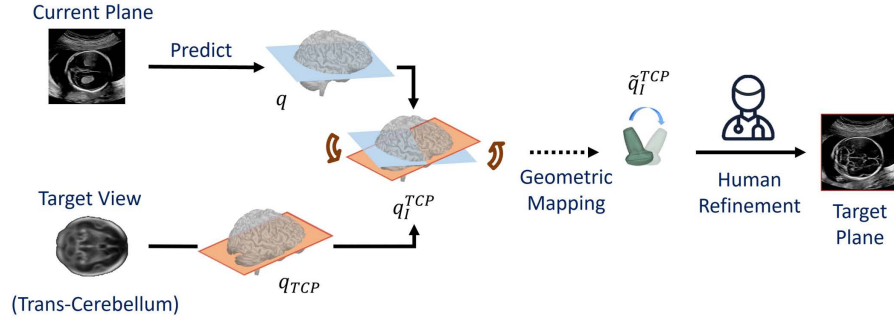
Fig. 1. Principle of *Pose-GuideNet*. The system provides guidance on how to move to a SP of 3D anatomy, and the corresponding probe movement would be inferred through geometric mapping from 3D anatomy to the probe coordinate.

reduce scanning workload as well as provide trainee sonographers and occasional users with heuristic guidance of probe manipulation to support SP acquisition.

Existing algorithms in the literature typically regard fetal SP detection as a classification problem of still ultrasound (US) images [2, 11, 21] or video clips [3, 19]. Recently, guidance-based frameworks have been developed [5, 14, 15] to provide probe movement instructions for SP acquisition with the help of an external motion tracker. Droste et al. [5] proposed a behavioral cloning system learned from the action of expert operators to estimate the next probe movement, and it is further extended [14, 15] with visual focus that simulates sonographer hand-eye coordination for more accurate guidance. However, these guidance methods are supervised by motion signals and rely on the previous probe motion to infer how to move the probe towards SP. For the detection or guidance methods mentioned, a key overlooked factor is the natural correlation between a freehand scanning 2D frame and its position in the 3D reference frame of the viewed anatomy, which is crucial for navigating within a 3D space during SP searching.

Methods for localizing a 2D US plane in a volumetric data space mainly refer to pose estimation under 3D US [16]. With the known 3D spatial information, an accurate 2D to 3D mapping can be achieved for pose estimations [4, 18, 24]. Yeung et al. [26] attempted to infer the pose of a freehand 2D fetal head image from a 3D US-based network. The estimation is less satisfactory due to the inherent differences between the two imaging techniques. For example, 3D US is normally anatomy-centered, while the structure in a freehand 2D scan can be arbitrary position. To increase cross-modality performance, cycle consistency is later adopted in [25] to link between 3D US slices and freehand images with a minimum manual registration. However, these methods are limited by an effective evaluation metric without the actual 3D location of a 2D US frame, and they are not targeted for SP acquisition guidance.

In this work, we fill the literature gap by combining the two fields: registering the freehand US image to a volumetric atlas and automatically guiding towards

SP in 3D anatomy. A uniform model, *Pose-GuideNet*, is proposed to first localize 2D US image without the need for manual registration, and then provide scanning guidance to reach a SP image. The clinical target is head biometry acquisitions, where a sonographer typically searches for the trans-ventricular plane (TVP) and trans-cerebellar plane (TCP) to assess fetal brain development and estimate gestational age. Pose-GuideNet first performs *in-plane alignment* by registering the acquired SP frame in a real-world freehand scan to the biometry view in 3D fetal head atlas. The rest of the scanning frames are then registered through *out-of-plane alignment* with their geometry affinity informed by their anatomical similarity. The method gives a relatively precise measurement by comparing with probe motion data that reflects the true 3D position of the current plane, which also facilitates the new evaluation methods of freehand image localization techniques. Experimental results with both motion-based and image-based evaluations demonstrate that Pose-GuideNet performs accurate pose estimation by preserving the direction and anatomical details of the query 2D image. The estimated transformation towards SP may also drive the real-world deployment of probe displacement in SP acquisition.

2   Methods

Figure 1 outlines the process of scanning guidance with *Pose-GuideNet*. The system anticipates the transformation of the current US plane to the pre-defined target viewpoint in 3D anatomy (*i.e.* TVP and TCP in fetal head scan). The transformation can be further converted to probe movement for a human sonographer to manipulate the probe to reach the target plane in freehand 2D scanning. Figure 2 depicts the training workflow of the proposed pose localization method. Since the position of 2D US image is unknown, a two-stage learning strategy is considered for Pose-GuideNet: A pose encoder is initially trained on sampled atlas 2D slices to capture the generic mapping between the anatomical feature and its corresponding position in 3D space; It is then fine-tuned on freehand 2D US scans by 1) the *Geometric-guided In-plane Alignment* - supervised alignment of the head standard planes and their directions based on the prior geometric knowledge on the atlas, 2) the *Semantic-aware Out-of-plane Alignment* - unsupervised alignment of the off-plane frames to atlas w.r.t. their anatomical similarity learned by a semantic encoder.

Problem Definition. The principle of the method involves establishing a correspondence between the reference 2D US image and its cross-sectional view in an open-sourced 3D atlas volume [17]. Given any 2D US image frame I in a freehand fetal head scanning stream, Pose-GuideNet registers the US image by predicting its location $\boldsymbol{\theta} = \{\boldsymbol{q}, \boldsymbol{\delta}\}$ including orientation (in quaternion) $\boldsymbol{q} = [w, x, y, z]$ and translation $\boldsymbol{\delta} = [\Delta_x, \Delta_y, \Delta_z]$ in 3D brain anatomy. Then the transformation of the US image towards SP is defined as $\boldsymbol{q}_I^{sp} = \boldsymbol{q}^* \boldsymbol{q}_{sp}$ (quaternion multiplication) and $\boldsymbol{\delta}_I^{sp} = \boldsymbol{\delta}_{sp} - \boldsymbol{R}(\boldsymbol{q}_I^{sp}) \cdot \boldsymbol{\delta}$, with $\boldsymbol{q}_{sp}$ and $\boldsymbol{\delta}_{sp}$ representing the orientation and the translation of SP under the same coordinate of the 3D atlas.
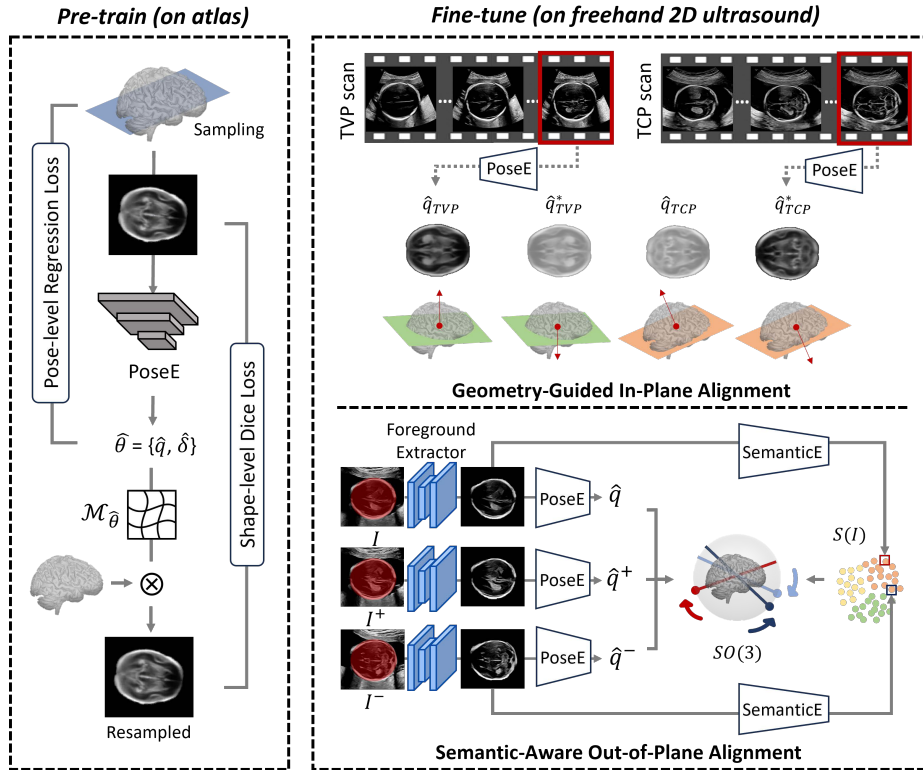
Fig. 2. Overview of *Pose-GuideNet* for pose localization of fetal head.

Here, q^* denotes the quaternion conjugate and $R(q)$ denotes the 3×3 rotation matrix of q, and $sp \in \{\text{TVP, TCP}\}$.

2.1   Pose Localization on 3D Atlas

Unlike existing methods [18, 25] that require the alignment and registration of 3D US volumes to atlas, we directly sample from the atlas volume to learn its positional mapping without introducing the 3D US modality. The backbone of Pose-GuideNet is a pose encoder (PoseE) based on ResNeXt50 [23]. Different from [25, 26] that solely constrain the 3D position, we also incorporate the shape-level similarity loss on the resampled 2D slice generated through a differentiable spatial transformer [10] to stabilize the training process. This is done by sampling the transformed grid $\mathcal{M}_{\hat{\theta}}$ of the estimated parameters $\hat{\theta}$ on the atlas volume (denoted as V). Since the intensity-based image constraint will easily overwhelm the adaption to US image fine-tuning, dice similarity coefficient (DSC) is used here to preserve the region of interest in the resampled slice. Specifically, the training objective on 3D atlas combines the shape-level dice loss and the pose-

level regression loss:

$$\mathcal{L}_{atlas}(I^{atl}) = 1 - \frac{2\sum_k (V \circ \mathcal{M}_{\hat{\boldsymbol{\theta}}}(k)) \cdot y(I^{atl}(k))}{\sum_k (V \circ \mathcal{M}_{\hat{\boldsymbol{\theta}}}(k)) + \sum_i y(I^{atl}(k))} + ||\boldsymbol{\theta} - \hat{\boldsymbol{\theta}}||_2, \qquad (1)$$

where I^{atl} is the input atlas slice at pose $\boldsymbol{\theta}$, $\hat{\boldsymbol{\theta}}$ is its predicted pose, and $y(I^{atl}(k))$ is the binary label for I^{atl} at pixel k. $\circ$ denotes the grid sampling operation.

2.2   Cross-Dimension Cross-Modality Alignment on 2D US

The dimensional gap between 2D US and 3D atlas is one of the main challenges that affects the 2D/3D registration accuracy. To overcome this obstacle, we propose to 1) establish an explicit mapping based on the cross-dimensional shared SP frame, and 2) for the frames with unknown mapping, learn the closeness of pose in 3D from their image-level conformity. The whole process includes the standard plane alignment in a head acquisition scan with its geometric prior and the alignment of its searching planes based on their anatomical features. As a primary concern for head standard plane refinement, we focus on understanding the change in orientation of the head plane without considering its translation. The rotation group defined on the 3D anatomy is isomorphic to *SO(3)* [7] on a Riemannian manifold structure, which allows us to establish a geodesic distance representing the most direct path between the orientations of two planes. Here, geodesic distance d_G is adopted on the two types of alignments to measure the closeness of 2D US images in the 3D atlas.

Another challenge is the cross-modality difference, and we reduce it by automatic recognition and removal of the unnecessary image background in 2D US through cascading an image localization net [10] with MedSAM [12]. The extracted foreground of fetal head is further processed by bilinear filtering [1] to produce a smoothed, atlas-like image before pose localization.

Geometry-Guided In-Plane Alignment. To ensure the geometry correctness, the target SP of a scan is aligned to one of the standard viewpoints (TVP or TCP) in atlas, each of which contains two opposite directions with the predefined geometry positions (as shown in the top right of Figure 2). Specifically, given a US SP image I_{sp}, the geodesic loss is given by

$$\mathcal{L}_{us}^{in}(I_{sp}) = \alpha d_G(\boldsymbol{q}_{sp}, \hat{\boldsymbol{q}}_{sp}) = \arccos(|<\boldsymbol{q}_{sp}, \hat{\boldsymbol{q}}_{sp}>|) \qquad (2)$$

that is computed between the SP positional prior $\boldsymbol{q}_{sp}$ and its prediction $\hat{\boldsymbol{q}}_{sp}$, where $|<\cdot,\cdot>|$ is the absolute inner product between two normalized quaternions [13]. Here, α is 0.5 given that the angle on *SO(3)* is half of the angle between their corresponding 3D orientations [7]. This in-plane alignment step allows Pose-GuideNet to identify the standard viewpoint, facilitating the geometric error corrections during registration.

Semantic-Aware Out-of-Plane Alignment. In a US scan, the searching frames before SP are of diverse angles in the 3D head anatomy, which are essentially spanned across the rotation manifold. Inspired by the unsupervised

manifold learning [9], we preserve the consistency in US images by contrasting their mappings on the rotation manifold. The manifold nearest neighbors (*i.e.* positive pairs) are selected from consecutive frames in the scan, which are geometrically closest to each other in the scanning sequence. The negative items are randomly selected frames within the same scan. However, given that two timely distinct frames may be visually similar, we also incorporate semantic similarity in negative items to advise their geometric closeness. The semantic similarity is to compare their anatomical characteristics from a semantic encoder (SemanticE) formed by SonoNet [2]. In particular, the alignment of an out-of-plane frame I can be regularized by a semantic-aware contrastive loss of the quaternion triplet:

$$\mathcal{L}_{us}^{out}(I) = -\log \frac{\exp(\cos(\alpha d_G(\hat{\boldsymbol{q}}, \hat{\boldsymbol{q}}^+))/\tau)}{\sum_{n=1}^{N} < S(I), S(I_n^-) > \exp(\cos(\alpha d_G(\hat{\boldsymbol{q}}, \hat{\boldsymbol{q}}_n^-))/\tau)}, \qquad (3)$$

where I_n^- is the nth negative sample of I with the total number N set to 5. $S(I)$ is the encoded semantic feature, and the temperature τ is set to 0.8 which controls the uniformity of the embedding distribution [22].

The overall objective for 2D US fine-tuning is to jointly optimize $\mathcal{L}_{us}^{in}$ and $\mathcal{L}_{us}^{out}$ on a batch of US images with in- and out-of-plane collections.

3   Experiments

3.1   Data and Experimental Settings

In this study, the 2D US scans were collected as part of the PULSE (Perception Ultrasound by Learning Sonographic Experience) [6]. The clinical fetal US scans were conducted on a GE Voluson E8 scanner. The US sequence is selected 10s before the *cine-buffer-corrected* SP and pre-processed to 6 Hz to reduce redundancy. This study received approval from UK Research Ethics Committee. There are 192 scanning sequences for the 2nd trimester fetal head biometry acquisition, resulting in 10810 valid image planes. The training and test split is 7428/3382. Among all test planes, 2081 are for TVP scans and 1301 for TCP. At each plane, the synchronized probe orientation $\tilde{\boldsymbol{q}}$ is recorded by an IMU motion sensor, which is used for evaluation purpose. The 3D fetal head atlas used is open-sourced [17]. 1059 US brain volumes collected from 899 fetuses were used to generate the fetal head atlases for different gestational ages. Here, we use the atlas at 20 week pregnancy as the median description of the fetal head in the 2nd trimester. To better align with the atlas slice, the US video frame was cropped to 224×288 and resized to 160×160 while keeping the aspect ratio. The experiments were run with PyTorch 1.10.1 on a 32GB NVIDIA Tesla V100 GPU. Pose-GuideNet is pre-trained and fine-tuned for 800 and 100 epochs respectively, using Adam optimizer with a batch size of 8 and a learning rate of 1e-4. To prevent overfitting, the last linear layer in PoseE is frozen during fine-tuning.

Table 1. Quantitative results of the estimated plane in terms of motion transformation towards SP and image-level similarity.

	Architecture	TVP			
		Motion	Image		
		KL Divergence ↓	Dice(%) ↑	NCC ↑	MS-SSIM ↓
Sensorless	Random plane	19.89	43.52	0.169±0.03	0.614±0.07
	Yeung et al.	18.89	55.03	0.271±0.07	0.556±0.08
	Ours (*pre-train*)	18.17	65.47	0.277±0.09	0.537±0.10
	Ours (*ft. Geo.*)	18.37	65.59	0.306±0.08	0.530±0.07
	Ours (*ft. Geo. Sem.*)	**17.33**	**66.91**	**0.315±0.09**	**0.529±0.09**
Sensored	US-GuideNet	15.01	-	-	-
	Multimodal-GuideNet	**14.79**	-	-	-
	Architecture	TCP			
		Motion	Image		
		KL Divergence ↓	Dice(%) ↑	NCC ↑	MS-SSIM ↓
Sensorless	Random plane	20.32	44.89	0.183±0.03	0.619±0.06
	Yeung et al.	20.03	57.11	0.308±0.10	0.536±0.08
	Ours (*pre-train*)	19.13	69.11	0.312±0.05	0.532±0.05
	Ours (*ft. Geo.*)	16.42	67.19	0.341±0.08	0.521±0.07
	Ours (*ft. Geo. Sem.*)	**15.47**	**70.12**	**0.353±0.08**	**0.511±0.07**
Sensored	US-GuideNet	15.27	-	-	-
	Multimodal-GuideNet	**14.62**	-	-	-

3.2 Evaluations with Motion-level Correspondence

A clinical goal for 2D US image localization is to find the biometry plane for fetal measurements. Because of the inherent correspondence between the probe manipulation and the field of view change in the scanned anatomy [8], we utilize the probe rotation towards its reached SP, *i.e.* $\tilde{q}_I^{sp} = \tilde{q}^* \tilde{q}_{sp}$, for evaluation of the predicted transformation between the current plane and its expected standard plane, *i.e.* $\hat{q}_I^{sp} = \hat{q}^* q_{sp}$. Since the probe and the 3D fetal head positioning are operated under different coordinate systems, their angular or positional distances are not directly comparable. We thus evaluate on the distributions of these two rotation dynamics in a scan using the statistical metric of KL Divergence. Other than the variations of our method, we compared with random plane assignment and registration-based approach [26] with atlas pre-training, and US guidance methods *US-GuideNet* [5] and *Multimodal-GuideNet* [15] based on probe motion prediction towards SP. These probe motion-based methods are considered *sensored* as they predict under a behavioral cloning strategy given the previous step from a motion sensor.

The results of different architectures are given in the *Motion* column of Table 1. On these two types of head biometry scans, Pose-GuideNet consistently improves under the *sensorless* method by gradually incorporating in-plane geometric (*ft. Geo.*) and out-of-plane semantic (*ft. Geo. Sem.*) alignments, which reduce the motion error close to the baselines using extra sensor(s). Practically, it is more challenging to locate TCP than TVP with the anatomical structure of cerebellum much smaller in the brain. TCP is thus less likely to be sampled
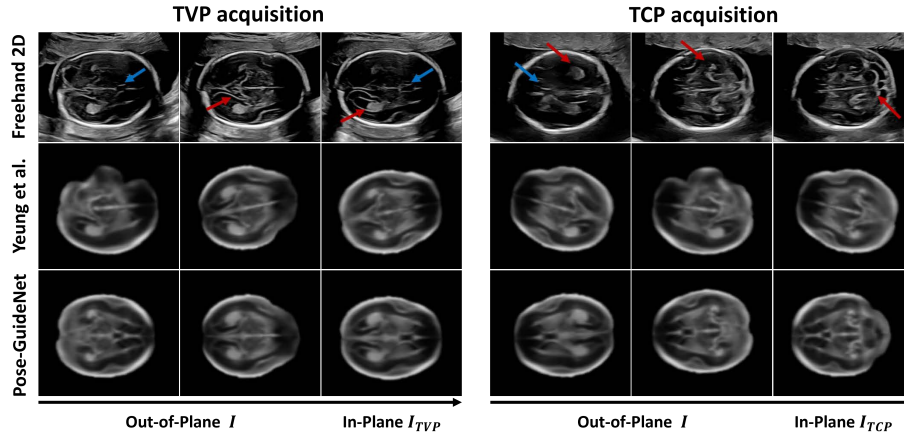
Fig. 3. The retrieved atlas planes of two example 2D head biometry acquisitions.

during pre-training with a larger localization error than TVP (20.03>18.89 and 19.13>18.17). However, by fine-tuning with geometric-guided alignment, Pose-GuideNet can greatly improve cross-dimensional registration by reducing the motion difference from 19.13 to 16.42. This is because the key plane serves as a common baseline that establishes a spatial correspondence between the 2D US and 3D anatomy. This correspondence is necessary for accurate mapping especially for few-shot planes such as TCP.

3.3   Evaluations with Image-level Correspondence

The visual similarity between the queried 2D US image and the retrieved plane from 3D anatomy is also compared in Table 1. Dice score, Normalized Correlation Coefficient (NCC), and Multi-Scale Structural Similarity Index Measure (MS-SSIM) are used as image similarity metrics, where MS-SSIM is more comprehensive than SSIM to tolerate scale variations of US images. From the results, Pose-GuideNet generally shows consistent performance across different metrics for both planes. A drop of dice score for TCP is observed when including the geometric constraint (69.11% vs. 67.19%). This may be due to that in atlas, the occipital bone ending around the cerebellum can be greatly deformed within a small angle change [17] (as in supplementary material), which makes it sensitive to the segmentation-based measurement. Figure 3 shows the result of retrieved atlas planes from in- and out-of-plane frames in two head scanning sequences. The anatomical structures (as highlighted by colored arrows) and head directions are both aligned well with the input 2D image under Pose-GuideNet. Some key structures, such as cavum septum pellucidum (CSP) highlighted in blue arrows, are challenging to observe due to the acoustic shadow caused by ultrasound artifacts. However, Pose-GuideNet can better estimate them by learning from the appearance of a neighbor frame with a different luminance condition.

4 Conclusion

We propose Pose-GuideNet which estimates the 3D pose from freehand 2D US image and its transformation towards SP in 3D anatomy, for the guidance of fetal head biometry acquisition. The approach registers US image to a 3D atlas volume using its prior knowledge of SP geometry as a baseline mapping, and using the anatomical similarity to consistently align the rest of off-plane frames in the scan. Extensive experiments show that the predicted plane is highly aligned with the input US image in terms of both anatomical structures and direction. Note that the Pose-GuideNet is a sensor-free method that can stand alone for fetal head pose estimation. As a future step, geometric correspondence can be incorporated that converts the transformation in 3D anatomy to probe motion in real-world deployment. This will enable a multi-functional guidance system for trainee sonographers, by providing not only a reference plane in 3D but also the decisive motion in hand towards the target plane.

Disclosure of Interests. The authors have no competing interests in the paper as required by the publisher.

References

1. Banterle, F., Corsini, M., Cignoni, P., Scopigno, R.: A low-memory, straightforward and fast bilateral filter through subsampling in spatial domain. In: Computer Graphics Forum. vol. 31, pp. 19–32 (2012)
2. 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)
3. Chen, H., Wu, L., Dou, Q., Qin, J., Li, S., Cheng, J.Z., Ni, D., Heng, P.A.: Ultrasound standard plane detection using a composite neural network framework. IEEE transactions on cybernetics **47**(6), 1576–1586 (2017)
4. Dou, H., Yang, X., Qian, J., Xue, W., Qin, H., Wang, X., Yu, L., Wang, S., Xiong, Y., Heng, P.A., et al.: Agent with warm start and active termination for plane localization in 3d ultrasound. In: International Conference on Medical Image Computing and Computer-Assisted Intervention. pp. 290–298 (2019)
5. Droste, R., Drukker, L., Papageorghiou, A.T., Noble, J.A.: Automatic probe movement guidance for freehand obstetric ultrasound. In: Medical Image Computing and Computer Assisted Intervention–MICCAI 2020: 23rd International Conference, Lima, Peru, October 4–8, 2020, Proceedings, Part III 23. pp. 583–592 (2020)
6. Drukker, L., Sharma, H., Droste, R., Alsharid, M., Chatelain, P., Noble, J.A., Papageorghiou, A.T.: Transforming obstetric ultrasound into data science using eye tracking, voice recording, transducer motion and ultrasound video. Sci. Rep. **11**(1), 1–12 (2021)
7. Huynh, D.Q.: Metrics for 3d rotations: Comparison and analysis. Journal of Mathematical Imaging and Vision **35**, 155–164 (2009)

8. Ihnatsenka, B., Boezaart, A.P.: Ultrasound: Basic understanding and learning the language. International journal of shoulder surgery **4**(3), 55 (2010)
9. Iscen, A., Tolias, G., Avrithis, Y., Chum, O.: Mining on manifolds: Metric learning without labels. In: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. pp. 7642–7651 (2018)
10. Jaderberg, M., Simonyan, K., Zisserman, A., et al.: Spatial transformer networks. Advances in neural information processing systems **28** (2015)
11. Lee, L.H., Gao, Y., Noble, J.A.: Principled ultrasound data augmentation for classification of standard planes. In: International Conference on Information Processing in Medical Imaging. pp. 729–741 (2021)
12. Ma, J., He, Y., Li, F., Han, L., You, C., Wang, B.: Segment anything in medical images. Nature Communications **15**(1), 654 (2024)
13. Mahendran, S., Ali, H., Vidal, R.: 3d pose regression using convolutional neural networks. In: Proceedings of the IEEE International Conference on Computer Vision Workshops. pp. 2174–2182 (2017)
14. Men, Q., Teng, C., Drukker, L., Papageorghiou, A.T., Noble, J.A.: Multimodal-guidenet: Gaze-probe bidirectional guidance in obstetric ultrasound scanning. In: International Conference on Medical Image Computing and Computer-Assisted Intervention. pp. 94–103 (2022)
15. Men, Q., Teng, C., Drukker, L., Papageorghiou, A.T., Noble, J.A.: Gaze-probe joint guidance with multi-task learning in obstetric ultrasound scanning. Medical image analysis **90**, 102981 (2023)
16. Mohamed, F., Siang, C.V.: A survey on 3d ultrasound reconstruction techniques. Artificial Intelligence—Applications in Medicine and Biology pp. 73–92 (2019)
17. 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)
18. 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 multi-task learning. Medical image analysis **46**, 1–14 (2018)
19. Pu, B., Li, K., Li, S., Zhu, N.: Automatic fetal ultrasound standard plane recognition based on deep learning and iiot. IEEE Transactions on Industrial Informatics **17**(11), 7771–7780 (2021)
20. Salomon, L.J., Alfirevic, Z., Berghella, V., Bilardo, C., Hernandez-Andrade, E., Johnsen, S., Kalache, K., Leung, K.Y., Malinger, G., Munoz, H., et al.: Practice guidelines for performance of the routine mid-trimester fetal ultrasound scan. Ultrasound in Obstetrics & Gynecology **37**(1), 116–126 (2011)
21. Sundaresan, V., Bridge, C.P., Ioannou, C., Noble, J.A.: Automated characterization of the fetal heart in ultrasound images using fully convolutional neural networks. In: 2017 IEEE 14th international symposium on biomedical imaging (ISBI 2017). pp. 671–674 (2017)
22. Wang, F., Liu, H.: Understanding the behaviour of contrastive loss. In: Proceedings of the IEEE/CVF conference on computer vision and pattern recognition. pp. 2495–2504 (2021)
23. Xie, S., Girshick, R., Dollár, P., Tu, Z., He, K.: Aggregated residual transformations for deep neural networks. In: Proceedings of the IEEE conference on computer vision and pattern recognition. pp. 1492–1500 (2017)
24. Yang, X., Shi, W., Dou, H., Qian, J., Wang, Y., Xue, W., Li, S., Ni, D., Heng, P.A.: Fetusmap: fetal pose estimation in 3d ultrasound. In: Medical Image Computing

and Computer Assisted Intervention–MICCAI 2019: 22nd International Conference, Shenzhen, China, October 13–17, 2019, Proceedings, Part V 22. pp. 281–289 (2019)

25. 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: International Conference on Medical Image Computing and Computer-Assisted Intervention. pp. 207–217 (2022)

26. 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)