Robust Fetal Pose Estimation across Gestational Ages via Cross-Population Augmentation

Sebastian Diaz[1,2], Benjamin Billot[3], Neel Dey[1,4,5], Molin Zhang[2], Esra Abaci Turk[4,6], P. Ellen Grant[4,6], Polina Golland[1], and Elfar Adalsteinsson[2]

[1] Computer Science & Artificial Intelligence Laboratory, MIT, Cambridge, USA
[2] Research Laboratory of Electronics, MIT, Cambridge, USA
[3] Inria, Epione team, Sophia-Antipolis, France
[4] Harvard Medical School, Boston, USA
[5] Massachusetts General Hospital, Boston, USA
[6] Boston Children's Hospital, Boston, USA
sebodiaz@csail.mit.edu

Abstract. Fetal motion is a critical indicator of neurological development and intrauterine health, yet its quantification remains challenging, particularly at earlier gestational ages (GA). Current methods track fetal motion by predicting the location of annotated landmarks on 3D echo planar imaging (EPI) time-series, primarily in third-trimester fetuses. The predicted landmarks enable simplification of the fetal body for downstream analysis. While these methods perform well within their training age distribution, they consistently fail to generalize to early GAs due to significant anatomical changes in both mother and fetus across gestation, as well as the difficulty of obtaining annotated early GA EPI data. In this work, we develop a cross-population data augmentation framework that enables pose estimation models to robustly generalize to younger GA clinical cohorts using only annotated images from older GA cohorts. Specifically, we introduce a fetal-specific augmentation strategy that simulates the distinct intrauterine environment and fetal positioning of early GAs. Our experiments find that cross-population augmentation yields reduced variability and significant improvements across both older GA and challenging early GA cases. By enabling more reliable pose estimation across gestation, our work potentially facilitates early clinical detection and intervention in challenging 4D fetal imaging settings. Code is available at https://github.com/sebodiaz/cross-population-pose.

Keywords: fetal · neurodevelopment · augmentation

1 Introduction

Fetal motion is a critical biomarker for neurological development, with diminished movement patterns linked to neurodevelopmental disorders [13] and intrauterine complications such as hypoxia, infection, and growth restriction [2]. Accurate quantification of these movements is essential for effective monitoring and timely clinical interventions. Moreover, distinguishing patterns in fetal
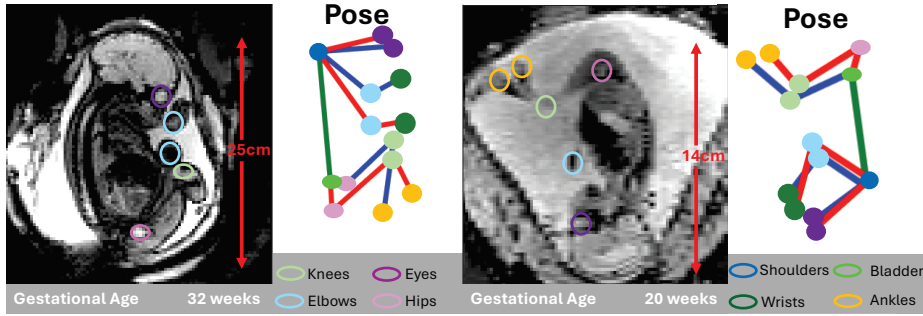
Fig. 1. A comparison of 3D MRI of fetuses at different gestational ages (GA). Annotations mark anatomical keypoints. (**Left**) A 32-week GA fetus scanned at 3mm isotropic resolution, similar to subjects used for training [22]. (**Right**) A 20-week GA fetus scanned at 2mm isotropic resolution. Pose estimation networks trained on higher GA fetuses (**Left**) do not generalize well to younger GA fetuses (**Right**).

motion—such as frequency, rhythm, or complexity—may help differentiate between healthy and at-risk populations, offering deeper insight into developmental trajectories. Traditionally, fetal motion has been assessed through subjective maternal perception, with quantitative automated fetal tracking research still in its infancy.

Recent methods have used ultrasound for movement quantification by tracking anatomical landmarks [5], but its associated low tissue contrast makes these landmarks difficult to detect. Fetal Magnetic Resonance Imaging (MRI) is a promising alternative as it enables rapid acquisitions with large field-of-view (FOV) and sufficient contrast to reliably identify anatomical landmarks and enable motion studies.

Recent research has demonstrated the potential of convolutional neural networks (CNNs) for fetal pose estimation and motion analysis [20,22]. Inspired by human pose estimation [19,27], these models predict Gaussian-distributed heatmaps centered on annotated keypoints. The methods achieve impressive performance on third-trimester populations (27-37 weeks GA), but unfortunately this cohort is not fully representative of clinical practice, where scans cover a broader GA range and may have anisotropic resolution (Fig. 1). Pose estimation in younger fetuses (18-22 weeks GA) presents further challenges. Compared to third-trimester fetuses, they exhibit less defined joint articulation [4], have four times less mass [11], and display greater intrauterine mobility, leading to a wider range of poses and frame-to-frame motion. These factors complicate annotation, even for experts, make labeling more time-consuming than their older counterparts, where anatomical features are more distinct and movement is reduced.

Such domain gaps [14] are common in medical imaging due to variations in acquisition protocols, scanner hardware, image resolution, and sites [16,21,25]. Data augmentation, attempts to bridge this gap by perturbing the source domain

to encourage robust feature learning [23,26]. However, even with the extensive use of existing augmentations, these models fail to generalize to early GAs as they do not fully account for the drastic morphological differences across gestation.

Clinical fetal MRI presents a severe case of domain shift. Fetal growth introduces a fundamental mismatch – smaller, younger fetuses differ significantly in size, musculoskeletal development, and movement patterns. Without explicitly addressing this discrepancy, models trained on readily available older GA data struggle to generalize to clinical cases at earlier GAs, where accurate pose estimation is most needed.

To address the scarcity of diverse training data for fetal pose estimation, we developed a novel fetal-specific augmentation strategy that captures the unique spatial arrangement of the uterus earlier in development. By leveraging an annotated dataset with segmented regions for the fetus and uterus, we randomly scale and warp fetal bodies and inpaint them onto uteruses from different subjects. This process simulates the wide variation of spatial configurations seen at younger gestational ages (GAs). Importantly, drawing inspiration from methods like SynthMorph [8] and SynthSeg [3], our approach acknowledges that generating realistic data isn't always necessary to achieve significant performance improvements. Through training with our proposed augmentation, we observe substantial gains in fetal pose estimation performance and robustness across all anatomical regions and gestational ages, particularly in clinical cohorts where younger GAs are more common.

2   Methods

2.1   Fetal motion estimation by supervised keypoint detection

Problem formulation: An overview of our method is seen in Fig 2A. Given an EPI time-series $\mathcal{S} \in \mathbb{R}^{H \times W \times D \times T}$, where T is the number of volumes in the sequence, we define a single 3D volume $V_t \in \mathbb{R}^{H \times W \times D}$ as a sample at time t from $\mathcal{S}$. The goal of fetal pose estimation is to localize anatomical landmarks by estimating a set of keypoints $k_1, k_2, \ldots, k_{15}$ where each keypoint k_i corresponds to an anatomical landmark in the fetus represented by its 3D voxel coordinates $k_i = (x_i, y_i, z_i) \in \mathbb{R}^3$ within the volume V_t. Instead of directly predicting keypoint coordinates, we represent them with heatmaps $H_i \in \mathbb{R}^{H \times W \times D}$, where each voxel (x, y, z) encodes the likelihood of the keypoint location. Specifically, we adopt a probabilistic approach, where the ground truth heatmap for keypoint k_i is modeled as a 3D isotropic Gaussian with standard deviation σ centered on (x_i, y_i, z_i).

During training, we optimize our network to predict a set of heatmaps $\hat{H} = \{\hat{H}_1, \hat{H}_2, \ldots, \hat{H}_{15}\}$ from an input volume V where $\hat{H}_i$ approximates H_i. The training objective is to minimize the Mean Squared Error (MSE) between the predicted and the ground truth heatmaps. At inference time, the keypoint coordinates, $\hat{k}_i = (\hat{x}_i, \hat{y}_i, \hat{z}_i)$ are extracted from each predicted heatmap $\hat{H}_i$ by first
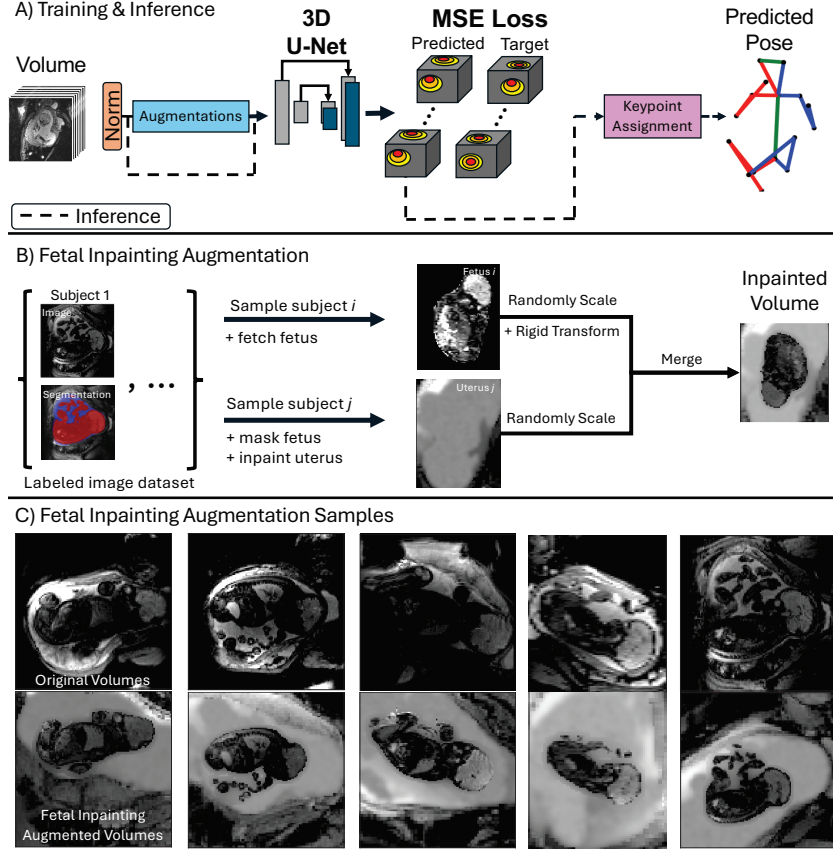
Fig. 2. Overview of the proposed method. (A) A network is trained to output Gaussian heatmaps, from which keypoints are extracted by taking the center-of-mass around the heatmap's `argmax`. (B) Proposed fetal inpainting augmentation. (C) Views of a training volume augmented with the proposed synthetic amniotic fluid and inpainting transform. Top and bottom rows are the raw EPI volumes and associated fetal body embedded into a uterus, respectively. Gamma correction was applied for visualization.

taking the highest activation:

$$u^* = (x_i^*, y_i^*, z_i^*) = \arg\max_{(x,y,z)} \hat{H}_i(x, y, z),$$

followed by a weighted local refinement around the neighborhood, $\mathcal{N} = \{u^* + v : v \in \{-1, 0, 1\}^3\}$, with a small constant, $\epsilon = 10^{-10}$:

$$\hat{k}_i = \frac{\sum_{u \in \mathcal{N}} u \cdot \hat{H}_i(u)}{\epsilon + \sum_{u \in \mathcal{N}} \hat{H}_i(u)}$$

2.2   Fetal-specific augmentations

MRI-specific augmentations Conventional data augmentation methods – such as simple rotations, scalings, flips, or intensity adjustments – benefit generalization by increasing the diversity of training examples [9]. Although these were employed in prior work [22,24,25], they fail to capture the nature of clinical data. Building off the augmentations included by the prior methods, we first incorporate a range of transformations to further enhance model robustness. These include additive noise, anisotropy, bias fields, gamma, and spikes typically encountered in MRI acquisitions. In our work, we extend these transformations by increasing their severity over commonly used settings [3,17].

Fetal inpainting To capture the fact that early GA fetuses are smaller, in highly variable poses, and exhibit a substantially higher amniotic fluid-to-body ratio (Fig. 1), we propose a novel augmentation. Specifically, we extract segmentation masks for the fetal body B and the surrounding amniotic fluid A, where the uterus is defined as $U = B \cup A$ and I is original image. We then generate a semi-realistic amniotic fluid intensity distribution in the absence of the fetal body. We achieve this by removing the body and synthesizing the fluid intensities in the vacated space as: $I_s(x) = \tilde{A} + \epsilon(x), x \in B$, where $\tilde{A}$ is the median of the amniotic fluid intensity and $\epsilon(x) \sim \mathcal{N}(0, \sigma_\epsilon^2)$ represents additive noise. To ensure smooth transitions between the synthesized fluid and the original fluid, we apply a Gaussian kernel $\mathcal{G}(x; \sigma_u)$ over U:

$$I^{'}(x) = [\mathbb{1}_{x \notin B} I(x) + \mathbb{1}_{x \in B} I_s(x)] * \mathcal{G}(x; \sigma_u), \quad x \in U,$$

and scale $I^{'}$ by a factor $\gamma \sim p(\gamma)$ resulting in I_U. We add the newly created image pair (U, I_U) and body pair (B, I_B) to the training collection (Fig 2B). To generate the augmentation, a fetal body B with its intensity image I_B and uterus image I_U with its mask U are randomly sampled from the training collection. We scale B and by a factor $\alpha \sim p(\alpha)$ and apply a random rigid transformation $T \sim p(T)$ to obtain a transformed body $I_B(T(\alpha B))$ that satisfies $T(\alpha B) \subset U$. The augmented sample consists of a fetal body $I_B(T(\alpha B))$ placed in a randomly sampled synthetic volume I_U, which is then added to the training dataset.

2.3   Implementation details

We perform all augmentations online which increases the variability of the training data and improves robustness of the network. We use TorchIO [17] functions for the conventional augmentation implementations. These augmentations include random noise, random K-space spike artifacts, random bias field corruptions, random rotations, and random gamma adjustments. In the event of a scaling augmentation, we multiply the σ of the ground truth heatmap by the chosen factor to discourage overlapping heatmaps as the subject becomes smaller. We further randomly sample anisotropic image resolution transformations with downsampling factors between 1.5 and 2. Augmentations are not applied to the
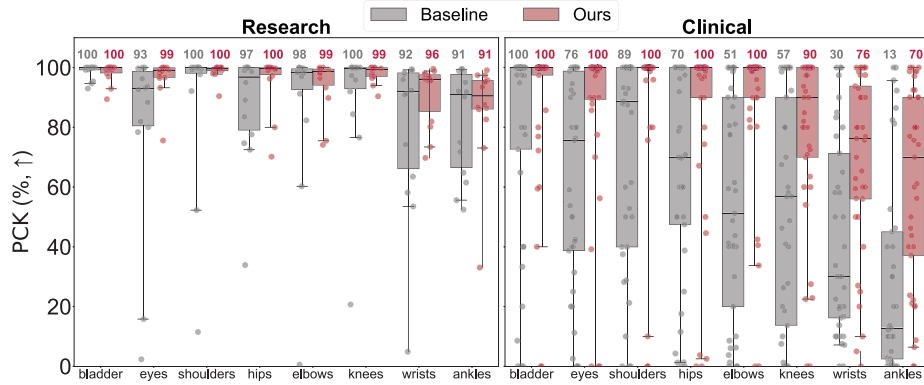
Fig. 3. PCK (threshold = 10 mm) statistics. Left and right keypoints are binned together. Median values are reported above each column. Our method demonstrates improvement in the research cohort (left) and significantly improves estimation in the clinical cohort (right).

fetal inpainted volumes. For fetal inpainting subjects, we use an in-house segmentation network, trained on ground truth segmentation masks, to obtain body and amniotic fluid masks.

For keypoint detection, we use a lightweight 3D UNet [6,18] variation which has demonstrated superior performance in per-voxel tasks compared to larger models [7]. Our UNet uses ReLU activations, with 4 pooling operations, an initial embedding dimension of 16, and a channel multiplier of 4, with two convolutional blocks per level. We train our UNet for 1,000 epochs with a batch size of 16, crop size of 64, using Adam optimizer [10] with a linear decay scheduler and an initial learning rate of 0.0002.

3   Experiments and results

3.1   Data

Our study uses two distinct manually labeled datasets. The **research** dataset was acquired with a single-shot gradient-echo (GRE) EPI acquisitions, 3 mm isotropic, TR=2.5-4s, TE=32–38ms, FA=90°. This dataset was collected as part of a study that included predominately higher GAs (27-37 weeks; avg. = 31.74). It includes 19,816 volumes from 77 acquisitions and 53 unique patients. The **clinical** dataset is acquired with the same GRE EPI mentioned above with TR=2.2-5s, TE=37ms, and FA=90°. In-plane resolution ranged from 1.8-2 mm and slice thickness ranged from 2-3 mm. It includes 989 volumes from 37 acquisitions and 28 subjects with an average GA of 21.75 weeks. Thirty of the acquisitions had isotropic 2 mm resolution. This data was collected during routine clinical acquisitions, making it representative of cases encountered in everyday practice.
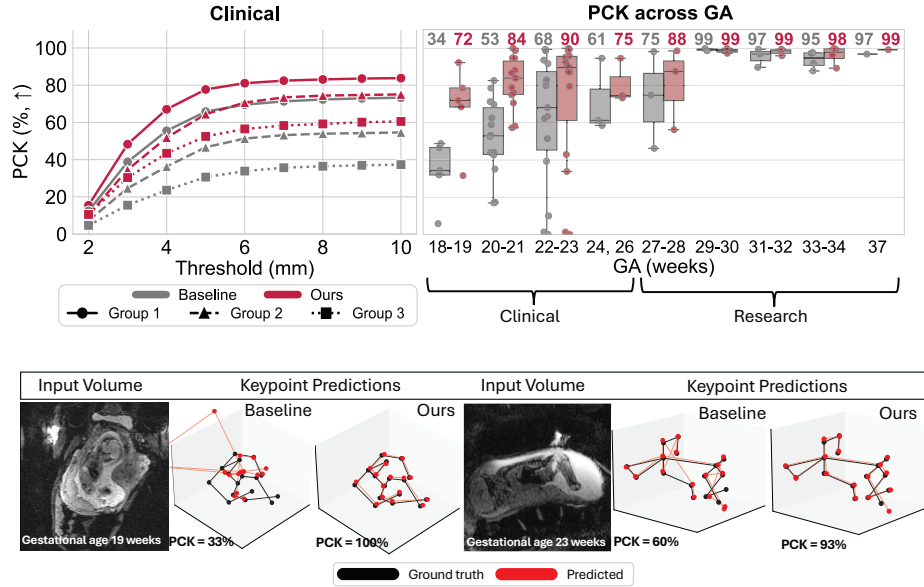
Fig. 4. (**Top Left**) **PCK vs. Threshold**. Our method versus the baseline's performance on the **_clinical_** dataset across threshold values binned into three groups. Group 1 (○): bladder, eyes, shoulders, and hips. Group 2 (△): elbows and knees. Group 3 (□): ankles and wrists. (**Top Right**) **PCK vs. GA**. PCK statistics as a function of GA. Median values are reported above each box. (**Bottom**) Qualitative predictions. Far right time-series animation is shown in the supplementary.

3.2   Experimental setup

Training: Our networks are exclusively trained on the **_research_** dataset, dividing (without subject overlap) the 77 times-series into 50 (N=11,618 scans), 13 (N=5,186), and 14 (N=3,462) for training, validation and testing, respectively. 15% of the total training volumes are fetal inpainting volumes, when applicable. For training, we sampled 8,000 volumes from the training pool.

Baseline: We adapt the original method proposed for volumetric fetal keypoint detection [22] by converting the open-source TensorFlow 1.x [1] code (`https://github.com/daviddmc/fetal-pose`) into PyTorch [15]. This method has demonstrated high performance in research-grade acquisitions [22] and influenced observation motion analysis research [20], making it the most relevant and meaningful baseline for comparison. We maintain the same pre-processing, augmentation, architecture, and training parameters, training for 400 epochs using a batch size of 8, a crop size of 64, the AdamW optimizer [12] (weight decay = 0.0001), a cosine warmup scheduler, an initial learning rate of 0.001. Data aug-

Method	Bladder	Eyes	Shoulders	Hips	Elbows	Knees	Wrists	Ankles
			Research data					
Ours	98 ± 3	**97 ± 6**	**99 ± 2**	**96 ± 9**	**95 ± 9**	**98 ± 3**	**91 ± 10**	**86 ± 16**
Ours−FI	98 ± 2	91 ± 23	95 ± 13	93 ± 13	89 ± 23	92 ± 19	83 ± 19	82 ± 21
Ours−FI−Aniso.	**99 ± 2**	88 ± 22	93 ± 19	90 ± 19	90 ± 21	93 ± 19	81 ± 25	85 ± 14
Ours−FI−Int.	**99 ± 2**	80 ± 32	91 ± 24	87 ± 26	85 ± 29	92 ± 18	80 ± 26	78 ± 21
Ours−FI−Noise	97 ± 6	87 ± 27	92 ± 24	94 ± 10	86 ± 27	90 ± 21	83 ± 26	79 ± 24
Ours−FI−Spike	99 ± 3	83 ± 30	92 ± 25	93 ± 12	87 ± 26	91 ± 21	85 ± 25	81 ± 19
Ours−FI−Scale	98 ± 3	87 ± 25	89 ± 27	86 ± 26	84 ± 27	88 ± 26	77 ± 27	79 ± 25
			Clinical data					
Ours	**88 ± 25**	**86 ± 29**	**90 ± 27**	**84 ± 32**	**87 ± 27**	**77 ± 29**	**69 ± 29**	**61 ± 33**
Ours−FI	84 ± 28	84 ± 28	87 ± 26	81 ± 31	82 ± 27	74 ± 35	66 ± 32	57 ± 34
Ours−FI−Aniso.	77 ± 35	78 ± 32	80 ± 31	64 ± 38	68 ± 33	50 ± 36	53 ± 29	43 ± 32
Ours−FI−Int.	71 ± 38	54 ± 39	71 ± 35	54 ± 38	71 ± 30	55 ± 37	50 ± 34	35 ± 30
Ours−FI−Noise	76 ± 36	85 ± 31	86 ± 30	69 ± 38	81 ± 30	70 ± 36	68 ± 28	49 ± 37
Ours−FI−Spike	84 ± 30	76 ± 34	83 ± 31	73 ± 35	78 ± 32	62 ± 35	59 ± 35	39 ± 32
Ours−FI−Scale	72 ± 38	77 ± 32	83 ± 29	66 ± 35	72 ± 33	59 ± 38	62 ± 32	42 ± 35

Table 1. Ablation: PCK (Mean ± standard deviation, threshold = 10 mm) per acquisition on both datasets. **Bolded** numbers are highest performance in each dataset. "FI" denotes Fetal Inpainting. "Int." is short for "intensity" and refers to gamma and bias field transforms. "Aniso." denotes anisotropy.

mentation settings were replicated from [22] and included intensity, rotations, and scaling transformations.

3.3   Results

We evaluate keypoint localization using percentage of correct keypoints (PCK), which is defined as a frequency the predicted keypoint falls within a predefined threshold distance from the ground truth. Figure 3 reports PCK statistics across different acquisitions in the research (n=14) and clinical (n=37) datasets. Our method consistently outperforms the baseline, achieving higher median PCK scores with notably lower variability across acquisitions, highlighting its robustness. In particular, keypoints corresponding to the wrists and ankles – areas indicative of significant gestational motion – exhibit median performance increases of 46% and 57%, respectively, in the clinical cohort.

We further stratified the keypoints into three groups based on detection difficulty: Group 1 (eyes, shoulders, hips, bladder), Group 2 (elbows, knees), and Group 3 (ankles, wrists). As illustrated in Figure 4, our method delivers substantial and consistent gains across all groups. The most dramatic improvements are observed in Group 3, which comprises the smallest and most challenging features.

To investigate the influence of GA on detection accuracy, we merged the datasets and performed a PCK analysis within discrete GA bins (Fig. 4). As expected, performance saturates for both methods on older subjects, while our method demonstrates a significant advantage on younger populations. Furthermore, the narrower spread of errors across subjects underscores the robustness and consistency of our approach.

Ablations: To understand the effect of inpainting augmentation in conjunction with other standard transformations, we conduct ablations on each augmentation while maintaining the same training hyperparameters and report results in Table 1. We find that all added augmentation contribute to the robustness of the model across all anatomical landmarks, with fetal inpainting combined with all remaining augmentations yielding the best model overall.

4   Conclusion

This study addresses the clinical need for a more reliable fetal pose estimation method. We demonstrate how our augmentation approach, including a fetal inpainting method, can significantly improve performance, particularly in the early GA range. These advancements make the model clinically applicable, enabling clinicians to answer critical questions regarding fetal motion and neurological development with greater accuracy.

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

References

1. Abadi, M., Agarwal, A., Barham, P., Brevdo, E., Chen, Z., Citro, C., Corrado, G.S., Davis, A., Dean, J., Devin, M., Ghemawat, S., Goodfellow, I., Harp, A., Irving, G., Isard, M., Jia, Y., Jozefowicz, R., Kaiser, L., Kudlur, M., Levenberg, J., Mané, D., Monga, R., Moore, S., Murray, D., Olah, C., Schuster, M., Shlens, J., Steiner, B., Sutskever, I., Talwar, K., Tucker, P., Vanhoucke, V., Vasudevan, V., Viégas, F., Vinyals, O., Warden, P., Wattenberg, M., Wicke, M., Yu, Y., Zheng, X.: TensorFlow: Large-scale machine learning on heterogeneous systems (2015), http://tensorflow.org/, software available from tensorflow.org
2. Ayala, K., Huynh, C., Voegtline, K., Rutherford, H.J.: Made to move: A review of measurement strategies to characterize heterogeneity in normal fetal movement. Infant Behavior and Development **75**, 101949 (2024)
3. Billot, B., Greve, D.N., Puonti, O., Thielscher, A., Van Leemput, K., Fischl, B., Dalca, A.V., Iglesias, J.E.: Synthseg: Segmentation of brain mri scans of any contrast and resolution without retraining. Medical Image Analysis **86**, 102789 (2023)
4. Chauvin, N.A., Victoria, T., Khwaja, A., Dahmoush, H., Jaramillo, D.: Magnetic resonance imaging of the fetal musculoskeletal system. Pediatric Radiology **50**(13), 2009–2027 (2020)
5. Chen, C., Yang, X., Huang, Y., Shi, W., Cao, Y., Luo, M., Hu, X., Zhu, L., Yu, L., Yue, K., Zhang, Y., Xiong, Y., Ni, D., Huang, W.: Fetusmapv2: Enhanced fetal pose estimation in 3d ultrasound. Medical Image Analysis **91**, 103013 (2024)
6. Çiçek, Ö., Abdulkadir, A., Lienkamp, S.S., Brox, T., Ronneberger, O.: 3d u-net: Learning dense volumetric segmentation from sparse annotation. Med. Image Comput. Comput. Assist. Interv. pp. 424–432 (2016)

7. Dey, N., Billot, B., Wong, H.E., Wang, C.J., Ren, M., Grant, P.E., Dalca, A.V., Golland, P.: Learning general-purpose biomedical volume representations using randomized synthesis. International Conference on Learning Representations (2025)

8. Hoffmann, M., Billot, B., Greve, D.N., Iglesias, J.E., Fischl, B., Dalca, A.V.: Synthmorph: Learning contrast-invariant registration without acquired images. IEEE Transactions on Medical Imaging **41**(3), 543–558 (Mar 2022)

9. Isensee, F., Jaeger, P.F., Kohl, S.A.A., Petersen, J., Maier-Hein, K.H.: nnu-net: a self-configuring method for deep learning-based biomedical image segmentation. Nature Methods **18**(2), 203–211 (2020)

10. Kingma, D.P., Ba, J.: Adam: A method for stochastic optimization. International Conference on Learning Representations (2015)

11. Kiserud, T., Piaggio, G., Carroli, G., Widmer, M., Carvalho, J., Neerup Jensen, L., Giordano, D., Cecatti, J.G., Abdel Aleem, H., Talegawkar, S.A., Benachi, A., Diemert, A., Tshefu Kitoto, A., Thinkhamrop, J., Lumbiganon, P., Tabor, A., Kriplani, A., Gonzalez Perez, R., Hecher, K., Hanson, M.A., Gülmezoglu, A.M., Platt, L.D.: The world health organization fetal growth charts: A multinational longitudinal study of ultrasound biometric measurements and estimated fetal weight. PLOS Medicine **14**(1), e1002220 (2017)

12. Loshchilov, I., Hutter, F.: Decoupled weight decay regularization. International Conference on Learning Representations (2017)

13. Lüchinger, A.B., Hadders-Algra, M., van Kan, C.M., de Vries, J.I.P.: Fetal onset of general movements. Pediatric Research **63**(2), 191–195 (2008)

14. Pan, S.J., Yang, Q.: A survey on transfer learning. IEEE Transactions on Knowledge and Data Engineering **22**(10), 1345–1359 (2010)

15. Paszke, A., Gross, S., Massa, F., Lerer, A., Bradbury, J., Chanan, G., Killeen, T., Lin, Z., Gimelshein, N., Antiga, L., Desmaison, A., Köpf, A., Yang, E.Z., DeVito, Z., Raison, M., Tejani, A., Chilamkurthy, S., Steiner, B., Fang, L., Bai, J., Chintala, S.: Pytorch: An imperative style, high-performance deep learning library. In: Wallach, H.M., Larochelle, H., Beygelzimer, A., d'Alché Buc, F., Fox, E.B., Garnett, R. (eds.) NeurIPS. pp. 8024–8035 (2019)

16. Perone, C.S., Ballester, P., Barros, R.C., Cohen-Adad, J.: Unsupervised domain adaptation for medical imaging segmentation with self-ensembling. NeuroImage **194**, 1–11 (2019)

17. Pérez-García, F., Sparks, R., Ourselin, S.: Torchio: A python library for efficient loading, preprocessing, augmentation and patch-based sampling of medical images in deep learning. Computer Methods and Programs in Biomedicine **208**, 106236 (2021)

18. Ronneberger, O., Fischer, P., Brox, T.: U-net: Convolutional networks for biomedical image segmentation. Med. Image Comput. Comput. Assist. Interv. p. 234–241 (2015)

19. Toshev, A., Szegedy, C.: Deeppose: Human pose estimation via deep neural networks. IEEE Conference on Computer Vision and Pattern Recognition p. 1653–1660 (2014)

20. Vasung, L., Xu, J., Abaci-Turk, E., Zhou, C., Holland, E., Barth, W.H., Barnewolt, C., Connolly, S., Estroff, J., Golland, P., Feldman, H.A., Adalsteinsson, E., Grant, P.E.: Cross-sectional observational study of typical in utero fetal movements using machine learning. Developmental Neuroscience **45**(3), 105–114 (2022)

21. Xu, C., Li, J., Wang, Y., Wang, L., Wang, Y., Zhang, X., Liu, W., Chen, J., Vatian, A., Gusarova, N., Ye, C., Zheng, Z.: Simix: A domain generalization method for cross-site brain mri harmonization via site mixing. NeuroImage **299**, 120812 (2024)

22. Xu, J., Zhang, M., Turk, E.A., Zhang, L., Grant, E., Ying, K., Golland, P., Adalsteinsson, E.: Fetal pose estimation in volumetric MRI using a 3D convolution neural network. Med. Image Comput. Comput. Assist. Interv. **11767**, 403–410 (2019)
23. Yuan, Z., Fang, Z., Huang, Z., Wu, F., Yao, Y.F., Li, Y.: Adapting pre-trained generative model to medical image for data augmentation p. 79–89 (2024)
24. Zhang, M., Golland, P., Grant, P.E., Adalsteinsson, E.: Fetaldiffusion: Pose-controllable 3d fetal mri synthesis with conditional diffusion model (2024)
25. Zhang, M., Xu, J., Abaci Turk, E., Grant, P.E., Golland, P., Adalsteinsson, E.: Enhanced detection of fetal pose in 3d mri by deep reinforcement learning with physical structure priors on anatomy. In: Medical Image Computing and Computer Assisted Intervention–MICCAI 2020: 23rd International Conference, Lima, Peru, October 4–8, 2020, Proceedings, Part VI 23. pp. 396–405. Springer (2020)
26. Zhang, Y., Yang, L., Chen, J., Fredericksen, M., Hughes, D.P., Chen, D.Z.: Deep adversarial networks for biomedical image segmentation utilizing unannotated images. In: Descoteaux, M., Maier-Hein, L., Franz, A., Jannin, P., Collins, D.L., Duchesne, S. (eds.) Medical Image Computing and Computer Assisted Intervention MICCAI 2017. pp. 408–416. Springer International Publishing, Cham (2017)
27. Zheng, C., Wu, W., Chen, C., Yang, T., Zhu, S., Shen, J., Kehtarnavaz, N., Shah, M.: Deep learning-based human pose estimation: A survey. ACM Computing Surveys **56**(1), 1–37 (2023)