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.

A3F: Anatomy-Aware Actionable Feedback for Fetal Ultrasound

Angela Feixue Wang[1 [0009−0009−9798−4205]], Hala Lamdouar[1 [0000−0002−8091−0367]], Jayne Lander[2 [0000−0002−3643−4893]], Aris T. Papageorghiou[2 [0000−0001−8143−2232]], and J. Alison Noble[1 [0000−0002−3060−3772]]

[1] Department of Engineering Science, University of Oxford, Oxford, UK
feixue.wang@eng.ox.ac.uk
[2] Nuffield Department of Women's & Reproductive Health, University of Oxford, Oxford, UK

Abstract. The training of fetal ultrasound sonographers is limited by a shortage of experienced trainers. This questions whether an AI system can be developed to provide scalable ultrasound training support. However, existing AI-assisted fetal ultrasound models typically focus on image interpretation or quality assessment, but do not translate assessment scoring into acquisition guidance. We introduce A3F, a multi-task Anatomy-Aware Actionable Feedback framework that directly guides a trainee on image magnification, gain, anatomy centering, and acoustic shadow without an explicit intermediate image interpretation stage. We also establish a design principle for actionable ultrasound feedback: clinically deployable guidance should be succinct and action-focused to minimise cognitive load on a sonographer. Results show that decoupling actionable feedback categories into independent prediction heads enables logically consistent prediction of co-occurring instructions, without compromising real-time inference speed. Ablation study demonstrates that anatomical context injection and anatomy-specific augmentations further strengthen model performance. These findings establish the feasibility of directly learning clinically meaningful actionable feedback from fetal ultrasound images, and motivate future user studies to evaluate its impact on sonographer training. We make the code publicly available, alongside a lightweight annotation tool tailored for fast annotation used in the research. Both are available at: https://github.com/angela-feixue-wang/A3F.

Keywords: Fetal ultrasound · Actionable feedback · Human-AI collaboration · AI-assisted human skill assessment.

1 Introduction

Many studies have reported a global shortage of fetal ultrasound sonographers [25,26] which limits its application as a diagnostic imaging modality worldwide. For instance, a national workforce survey reported the average sonographer vacancy rate nearly doubled from 12.6% in 2019 [21] to 24.2% in 2025 [23], and

professional bodies are calling for increased training capacity [22]. However, fetal ultrasound sonography is highly complex and operator-dependent, usually requiring years of training to build proficiency. Furthermore, training capacity is limited by the number of experienced sonographers available to provide training. This motivates the development of AI systems to support training.

Fetal ultrasound AI-assistive models have advanced considerably in recent years, but typically focus on describing image or video content [2,10,14] and quality assessment [4,6,7,24]. These approaches do not instruct a trainee on what they need to do to improve a suboptimal image.

Drawing inspiration from actionable feedback in classical computer vision [3], we define actionable feedback as specific, coach-like instruction that communicates both what is wrong and how to fix it. Despite its importance, actionable feedback remains underexplored in the ultrasound AI literature. Li et al. [15] treated actionable feedback as a two-stage process, first generating image descriptions and then deriving feedback from them. But the generated feedback is limited to generic comments like "continue scanning". Ultrasound probe guidance is another important related field [12,17,18,19,27]. However, to the best of our knowledge, actionable feedback for other clinically important ultrasound acquisition parameters (e.g., magnification level and overall gain) has only been addressed in one prior study [13].

To motivate our approach, we analysed expert sonographer voiceover reviews of retrospective second trimester scan videos from the PULSE dataset [9]. This revealed four recurring feedback categories: (1) magnification level, (2) overall gain, (3) whether the region of interest is centered on the screen, and (4) whether acoustic shadow is preventing a clear view of the region of interest. We also observed that feedback is inherently tied to the anatomical view being examined, leading us to propose to incorporate anatomical context to learn anatomy-aware feedback representations.

Unlike existing methods that formulate actionable feedback as a two-stage process [15], we propose that actionable feedback can be directly learned from image data without an explicit intermediate image interpretation stage. Building on prior work [13], we extend feedback prediction from a single-task to a unified multi-task framework with dedicated prediction heads for each feedback category, thus preventing logical inconsistencies when multiple feedback categories are predicted simultaneously. We further demonstrate that anatomy-awareness by injecting anatomical context during training is a powerful mechanism to capture domain-specific characteristics in fetal ultrasound.

We make the following technical contributions: (1) We introduce a multi-task anatomy-aware actionable feedback framework that predicts actionable feedback directly from ultrasound images, eliminating the need for an explicit image interpretation stage as well as preventing logical inconsistencies when multiple corrective actions are predicted simultaneously; (2) We establish a design principle for actionable ultrasound feedback: clinically grounded feedback should be succinct and action-focused to minimise sonographer cognitive load; (3) We demonstrate that simple anatomy-specific augmentations efficiently mitigates class imbalance

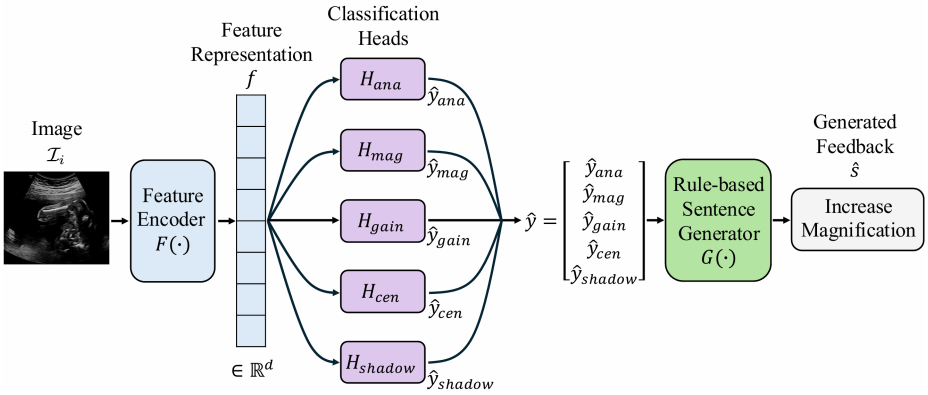

Fig. 1. An overview of the proposed A3F architecture.

by synthesising realistic suboptimal variants from good-quality images, enabling the model to learn from both positive and negative examples.

2 Methods

2.1 Multi-Task Anatomy-Aware Actionable Feedback

Figure 1 shows the proposed A3F architecture. Given an image dataset $\mathcal{D} = \{(\mathcal{I}_i, \mathbf{y}_i)\}_{i=1}^{N}$, a shared feature encoder $F(\cdot)$ maps an image $\mathcal{I}_i$ to a feature representation f_i:

$$f_i = F(\mathcal{I}_i) \in \mathbb{R}^d. \tag{1}$$

Five task-specific classification heads map f_i to logits z_{ana}, z_{mag}, z_{gain}, z_{cen}, and z_{shadow}, which are subsequently classified into labels $\hat{y}_{\text{ana}}$, $\hat{y}_{\text{mag}}$, $\hat{y}_{\text{gain}}$, $\hat{y}_{\text{cen}}$, and $\hat{y}_{\text{shadow}}$ using a softmax function for multi-class classification tasks (anatomy, magnification, gain, and centering) and a sigmoid function for the binary classification task (shadow). We adopt focal loss [16] to mitigate class imbalance:

$$\mathcal{L}_{\text{focal}} = -\alpha_t (1 - p_t)^{\gamma} \log(p_t) \tag{2}$$

where α_t is the class-balancing factor, p_t is the predicted probability of the ground-truth class, and γ is the focusing parameter. For multi-class classification, p_t is obtained from the softmax output corresponding to the true class, whereas for binary classification $p_t = p$ when $y = 1$ and $p_t = 1 - p$ when $y = 0$.

The overall training objective is to optimise the weighted sum of the task-specific losses:

$$\mathcal{L}_{\text{total}} = \lambda_{\text{ana}}\mathcal{L}_{\text{ana}} + \lambda_{\text{mag}}\mathcal{L}_{\text{mag}} + \lambda_{\text{gain}}\mathcal{L}_{\text{gain}} + \lambda_{\text{cen}}\mathcal{L}_{\text{cen}} + \lambda_{\text{shadow}}\mathcal{L}_{\text{shadow}} \tag{3}$$

where $\mathcal{L}_{\text{ana}}$, $\mathcal{L}_{\text{mag}}$, $\mathcal{L}_{\text{gain}}$, $\mathcal{L}_{\text{cen}}$, and $\mathcal{L}_{\text{shadow}}$ denote the focal loss computed using Eq. (2) for the anatomy, magnification, gain, centering, and shadow classification tasks, respectively. The weights λ_{ana}, λ_{mag}, λ_{gain}, λ_{cen}, and λ_{shadow} balance the contributions of individual tasks during joint optimisation.

2.2 Clinically Grounded Feedback

During inference, the predicted labels $\hat{y}_{\text{ana}}$, $\hat{y}_{\text{mag}}$, $\hat{y}_{\text{gain}}$, $\hat{y}_{\text{cen}}$, and $\hat{y}_{\text{shadow}}$ are passed into a rule-based sentence generator $G(\cdot)$ to map labels to a human-readable feedback sentence $\hat{s}$, given $\hat{\mathbf{y}} = [\hat{y}_{\text{ana}}, \hat{y}_{\text{mag}}, \hat{y}_{\text{gain}}, \hat{y}_{\text{cen}}, \hat{y}_{\text{shadow}}]$.

$$\hat{s} = G(\hat{\mathbf{y}}) \tag{4}$$

Following the proposed clinically grounded fetal ultrasound feedback design principle, $G(\cdot)$ prioritises brevity by focusing on information that directly supports actionable guidance. This design choice was informed by consultation with two expert sonographers, both of whom preferred concise instructions such as *"increase magnification"* over more verbose sentences like *"femur, good gain and good centering, need to increase magnification"* or *"femur, need to increase magnification"*. In addition, $G(\cdot)$ suppresses feedback output for images whose predicted anatomical views fall outside NHS FASP target anatomical views [20], as these views are not target acquisition planes.

2.3 Anatomy-Specific Augmentation

We observed that dataset class imbalance is particularly strong in the magnification, gain, and centering tasks, where some suboptimal classes are severely underrepresented. Take sagittal spine views as an example, *"good_mag"* has 1038 samples but *"reduce_mag"* only has six. To generate additional training samples, we introduce an anatomy-specific augmentation strategy during training and validation to generate controlled suboptimal variants for good images. Examples are displayed in Fig. 2. Given an ultrasound image, one or more of the following operations were randomly applied to good images: (1) scaling the field of view to simulate over- and under-magnification; (2) scaling and off-setting image intensity to mimic excessive or insufficient gain; and (3) shifting the region of interest to mimic off-centered views.

3 Experiments and Results

3.1 Dataset

11,045 ultrasound images were extracted from 577 full-length routine second-trimester ultrasound scans performed by both newly-qualified and experienced sonographers from the PULSE dataset [9]. All ultrasound scans were performed in the same ultrasound examination room on General Electric (GE) Healthcare Voluson E8 or E10 ultrasound machines with standard curvilinear (C1-5-D,

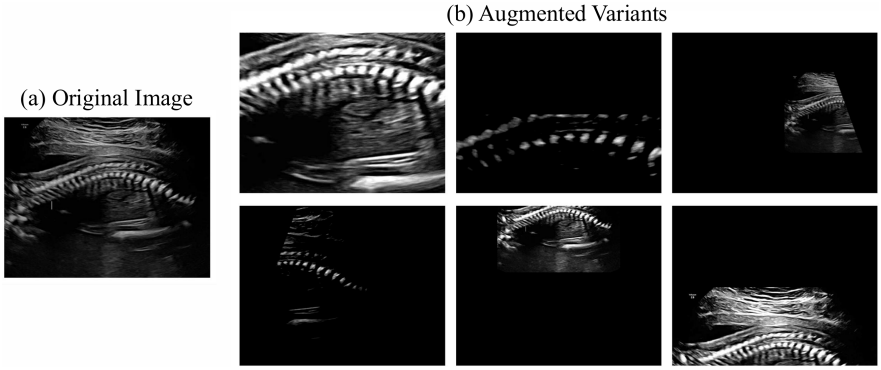

Fig. 2. Example anatomy-specific augmentations for a sagittal spine view.

C1-6-D, C2-9-D) and 3D/4D transducers (RAB6-D, RC6M). Scan videos were recorded at 30 fps.

Images were manually labelled using a custom developed keystroke-based annotation tool with fully configurable key-to-label mappings. Compared with mouse-driven tools such as CVAT, our keystroke annotator reduces unnecessary hand movements and accelerates annotation workflow. Annotation was performed by a researcher with fetal ultrasound research experience. 1% of images were randomly selected and reviewed by an experienced fetal ultrasound sonographer, whose feedback was incorporated across all annotations. To enable anatomical context injection during training, we assigned anatomical labels according to the NHS FASP target anatomical views [20], together with an *"Other"* category for non-FASP views. Training/validation (80%) and test (20%) data were split chronologically by scan date. The test set was set aside for evaluation. Five-fold cross-validation was performed for the training/validation set where folds were split by scan date. Anatomy-specific augmentations, as described in section 2.3, were applied to the training/validation sets. To prevent data leakage, each original image and all its augmentations were assigned to the same fold.

3.2 Implementation Details

We evaluated three feature extractor backbones: ResNet-50 [11], SonoNet [5], and ViT/B-16 [8]. Each model was trained for 100 epochs with a batch size of 16, on a NVIDIA RTX Pro 4500 GPU. We used the Adam optimiser and performed hyperparameter optimisation using Optuna [1]. Inference time was measured using a batch size of 1, with timings averaged over $2,000$ forward passes after a warm-up period.

3.3 Results and Discussion

During inference, the model predicts all five tasks directly from input image. Model evaluation results on the test set (2209 images) are presented in Table 1.

Task	Metric	Backbone		
		ViT/B-16 [8]	ResNet-50 [11]	**SonoNet [5]**
Anatomy	Accuracy (%)	67.9 ± 2.4	84.7 ± 3.9	**88.5 ± 4.4**
	Precision (%)	63.0 ± 3.0	83.0 ± 4.6	**87.5 ± 5.6**
	Recall (%)	67.0 ± 1.5	88.8 ± 2.0	**90.3 ± 2.2**
Mag.	Accuracy (%)	86.6 ± 1.1	88.2 ± 1.5	**90.3 ± 1.6**
	Precision (%)	58.6 ± 3.1	70.4 ± 6.9	**75.5 ± 9.1**
	Recall (%)	43.7 ± 2.0	56.5 ± 6.9	**57.1 ± 5.9**
Gain	Accuracy (%)	91.2 ± 0.7	91.0 ± 1.7	**92.5 ± 1.3**
	Precision (%)	63.9 ± 2.3	63.6 ± 4.7	**68.9 ± 5.4**
	Recall (%)	74.2 ± 7.6	68.8 ± 4.7	**72.9 ± 4.7**
Centering	Accuracy (%)	95.4 ± 0.7	95.8 ± 1.9	**96.6 ± 0.9**
	Precision (%)	32.2 ± 5.8	52.9 ± 5.8	**64.7 ± 15.2**
	Recall (%)	37.3 ± 3.9	62.1 ± 5.8	**58.4 ± 5.0**
Shadow	Accuracy (%)	89.0 ± 2.0	90.3 ± 2.2	**91.2 ± 2.4**
	Precision (%)	75.0 ± 2.0	78.6 ± 3.2	**78.4 ± 4.2**
	Recall (%)	74.7 ± 2.9	81.1 ± 7.1	**81.4 ± 5.1**

Table 1. Model evaluation results on test set (2209 images). Values are reported as task-wise macro-average ± standard deviation across models trained and validated on the five cross-validation folds.

Task	Metric	Ablation Study			
		Aug(×) AH(×)	Aug(✓) AH(×)	Aug(×) AH(✓)	**Aug(✓) AH(✓)**
Anatomy	Accuracy (%)	n/a	n/a	81.9	**89.6**
	Precision (%)	n/a	n/a	88.3	**92.5**
	Recall (%)	n/a	n/a	81.0	**86.9**
Mag.	Accuracy (%)	88.5	82.7	90.3	**91.5**
	Precision (%)	45.2	65.1	74.5	**88.5**
	Recall (%)	41.2	56.1	43.4	**61.3**
Gain	Accuracy (%)	92.6	91.4	92.3	**93.5**
	Precision (%)	61.2	64.9	93.2	**77.0**
	Recall (%)	57.6	65.5	52.7	**77.6**
Centering	Accuracy (%)	93.7	96.2	97.5	**97.8**
	Precision (%)	42.0	35.9	31.0	**83.1**
	Recall (%)	37.2	50.0	32.0	**64.9**
Shadow	Accuracy (%)	92.8	92.8	92.8	**93.2**
	Precision (%)	86.9	78.6	83.0	**82.3**
	Recall (%)	64.6	79.5	73.0	**75.9**

Table 2. Ablation study. We take the best model (SonoNet trained on fold No.1) and gradually remove anatomy-specific augmentation (Aug) and the anatomy-aware head (AH). × indicates the component is removed. ✓ indicates the component is included.

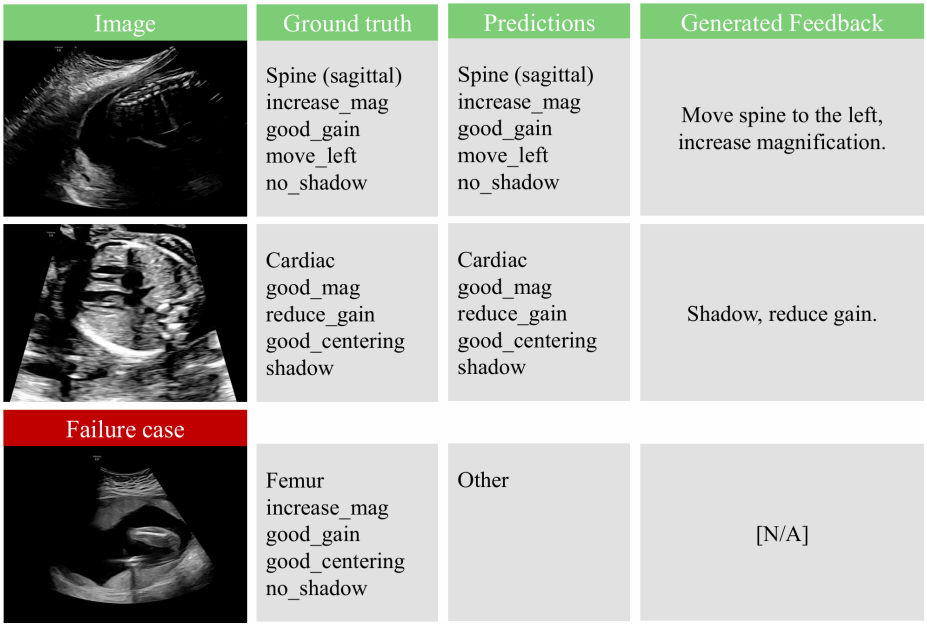

Fig. 3. Qualitative examples from the best-performing model, showing the input image, ground-truth labels, predicted labels, and generated feedback.

Model inference time per image was 1.76 ± 0.11 ms for ResNet-50, 1.12 ± 0.07 ms for SonoNet, and 4.93 ± 0.08 ms for ViT/B-16. $G(\cdot)$ sentence generation time was 0.04 ± 0.003 ms per image irrespective of backbone. This indicates that A3F is capable of real-time feedback. SonoNet achieved the highest overall performance and was therefore adopted for subsequent experiments.

Ablation study in Table 2 shows model performance degradation when gradually removing anatomy-specific augmentations and the anatomy-aware head, validating the contribution of each component to overall model performance. Fig. 3 and Fig. 4 display qualitative examples including failure cases and task-wise confusion matrices, respectively, both for the best model. We find that the model performs well on cases where the anatomy of interest is appropriately magnified, but tends to fall short when the anatomy of interest is not enlarged and become misclassified as *"Other"*. The confusion matrices also reflect the inherent subjectivity of fetal ultrasound image evaluation. For borderline cases, subtle differences in expert judgement may lead to different but equally reasonable feedback labels (e.g., *"good_centering"* and *"move_left"*). This is particularly evident for feedback classes represented by only a small number of test samples, and highlights the need for multi-sonographer validation in future work.

The expert sonographers' preference for brief, action-focused feedback suggests that clinical effectiveness is not achieved by maximising the amount of information presented to the user. During a real-time fetal ultrasound exam-

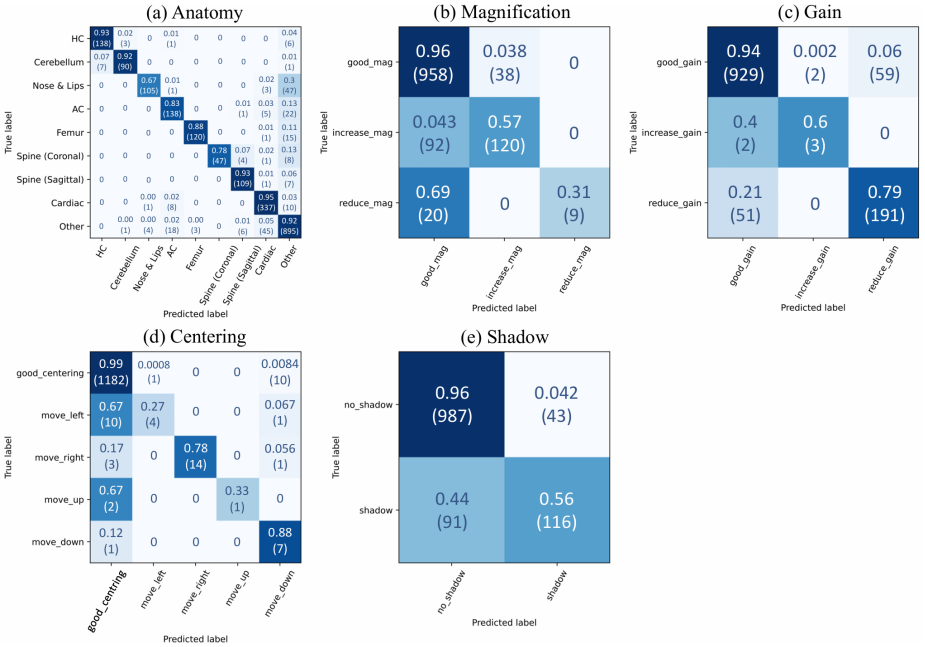

Fig. 4. Confusion matrices from the best-performing model.

ination, a sonographer needs to simultaneously manipulate the probe, adjust imaging parameters, interpret the image, and communicate with the patient. Consequently, concise guidance that minimises cognitive load is better suited to support real-time scanning.

Compared with the dual-head approach in prior work [13] where only one classification head is used for mutually exclusive feedback prediction, A3F's four independent feedback classification heads support concurrent prediction of more than one feedback (e.g., low magnification and shadow). A3F improves anatomy classification by 15.1, 18.4, and 15.5 percentage points in accuracy, precision, and recall, respectively. Although direct comparison of feedback prediction is complicated by differences in formulation, our gain and shadow heads exceed the previous study's overall instruction prediction performance, whereas the magnification and centering heads show slightly lower recall (3.5–7.1 percentage points).

4 Conclusion

This study demonstrates the feasibility of learning clinically meaningful actionable feedback directly from fetal ultrasound images. By combining a unified anatomy-aware multi-task actionable feedback framework with anatomy-specific augmentation, A3F improves anatomy, gain, and shadow feedback by up to 18.4 percentage points in precision compared to a previous dual-head approach. A3F

supports real-time deployment at an average inference time of 1.12 ms per image for the best-performing model. These results establish technical feasibility and motivate future user studies to quantify A3F's impact on trainee performance.

Acknowledgments. The authors acknowledge the University of Oxford Department of Engineering Science and the EPSRC Turing AI Fellowship for Ultra Sound Multimodal video-based human-machine collaboration (EP/X040186/1).

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

References

1. Akiba, T., Sano, S., Yanase, T., Ohta, T., Koyama, M.: Optuna: A next-generation hyperparameter optimization framework. In: Proceedings of the 25th ACM SIGKDD International Conference on Knowledge Discovery & Data Mining. pp. 2623–2631. Association for Computing Machinery, New York, NY, USA (2019). https://doi.org/10.1145/3292500.3330701
2. Alsharid, M., Sharma, H., Drukker, L., Chatelain, P., Papageorghiou, A.T., Noble, J.A.: Captioning ultrasound images automatically. In: Shen, D., Liu, T., Peters, T.M., Staib, L.H., Essert, C., Zhou, S., Yap, P.T., Khan, A. (eds.) MICCAI 2019. LNCS, vol. 11767, pp. 338–346. Springer, Cham (2019). https://doi.org/10.1007/978-3-030-32251-9_37
3. Ashutosh, K., Nagarajan, T., Pavlakos, G., Kitani, K., Grauman, K.: ExpertAF: Expert actionable feedback from video. In: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR). pp. 13582–13594 (2025)
4. Bashir, Z., Lin, M., Feragen, A., Mikolaj, K., Taksøe-Vester, C., Christensen, A.N., Svendsen, M.B.S., Fabricius, M.H., Andreasen, L., Nielsen, M., et al.: Clinical validation of explainable AI for fetal growth scans through multi-level, cross-institutional prospective end-user evaluation. Scientific Reports **15**(1), 2074 (2025). https://doi.org/10.1038/s41598-025-86536-4
5. 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). https://doi.org/10.1109/TMI.2017.2712367
6. Chen, C., Huang, Y., Yang, X., Hu, X., Zhang, Y., Tan, T., Xue, W., Ni, D.: Enhancing fetal ultrasound image quality assessment with multi-scale fusion and clustering-based optimization. Biomedical Signal Processing and Control **102**, 107249 (2025). https://doi.org/10.1016/j.bspc.2024.107249
7. Chen, L., He, F., Lei, T., Wang, S., Wang, Y., Wang, N., Yan, Y., Xie, H., Huang, Y.: Intelligent quality control for first trimester ultrasound scans in Liaoning Province of China. npj Digital Medicine **9**(1), 35 (2025). https://doi.org/10.1038/s41746-025-02207-8
8. Dosovitskiy, A., Beyer, L., Kolesnikov, A., Weissenborn, D., Zhai, X., Unterthiner, T., Dehghani, M., Minderer, M., Heigold, G., Gelly, S., et al.: An image is worth 16x16 words: Transformers for image recognition at scale. In: 9th International Conference on Learning Representations (ICLR) (2020)

9. 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. Scientific Reports **11**(14109), 2045–2322 (2021). https://doi.org/10.1038/s41598-021-92829-1
10. Guo, X., Men, Q., Noble, J.A.: MMSummary: multimodal summary generation for fetal ultrasound video. In: Linguraru, M.G., Dou, Q., Feragen, A., Giannarou, S., Glocker, B., Lekadir, K., Schnabel, J.A. (eds.) MICCAI 2024. LNCS, vol. 15004, pp. 678–688. Springer, Cham (2024). https://doi.org/10.1007/978-3-031-72083-3_63
11. He, K., Zhang, X., Ren, S., Sun, J.: Deep residual learning for image recognition. In: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR). pp. 770–778 (2016). https://doi.org/10.1109/CVPR.2016.90
12. Jiang, Z., Bi, Y., Zhou, M., Hu, Y., Burke, M., Navab, N.: Intelligent robotic sonographer: Mutual information-based disentangled reward learning from few demonstrations. The International Journal of Robotics Research **43**(7), 981–1002 (2024). https://doi.org/10.1177/02783649231223547
13. Lamdouar, H., Wang, A.F., Guo, X., Men, Q., Lander, J., Papageorghiou, A.T., Noble, J.A.: FOCUS: Towards fetal obstetric corrective ultrasound guidance. In: MICCAI 2026. LNCS (2026), to be published.
14. Li, J., Su, T., Zhao, B., Lv, F., Wang, Q., Navab, N., Hu, Y., Jiang, Z.: Ultrasound report generation with cross-modality feature alignment via unsupervised guidance. IEEE Transactions on Medical Imaging **44**(1), 19–30 (2025). https://doi.org/10.1109/TMI.2024.3424978
15. Li, X., Huang, D., Zhang, Y., Navab, N., Jiang, Z.: Semantic scene graph for ultrasound image explanation and scanning guidance. In: Gee, J.C., Alexander, D.C., Hong, J., Iglesias, J.E., Sudre, C.H., Venkataraman, A., Golland, P., Kim, J.H., Park, J. (eds.) MICCAI 2025. LNCS, vol. 15968, pp. 500–510. Springer, Cham (2025). https://doi.org/10.1007/978-3-032-05114-1_48
16. Lin, T.Y., Goyal, P., Girshick, R., He, K., Dollár, P.: Focal loss for dense object detection. IEEE Transactions on Pattern Analysis and Machine Intelligence **42**(2), 318–327 (2020). https://doi.org/10.1109/TPAMI.2018.2858826
17. Men, Q., Guo, X., Papageorghiou, A.T., Noble, J.A.: Pose-GuideNet: Automatic scanning guidance for fetal head ultrasound from pose estimation. In: Linguraru, M.G., Dou, Q., Feragen, A., Giannarou, S., Glocker, B., Lekadir, K., Schnabel, J.A. (eds.) MICCAI 2024. LNCS, vol. 15004, pp. 700–710. Springer, Cham (2024). https://doi.org/10.1007/978-3-031-72083-3_65
18. Men, Q., Teng, C., Drukker, L., Papageorghiou, A.T., Noble, J.A.: MultimodalGuideNet: Gaze-probe bidirectional guidance in obstetric ultrasound scanning. In: Wang, L., Dou, Q., Fletcher, P.T., Speidel, S., Li, S. (eds.) MICCAI 2022. LNCS, vol. 13437, pp. 94–103. Springer, Cham (2022). https://doi.org/10.1007/978-3-031-16449-1_10
19. Men, Q., Zhao, H., Drukker, L., Papageorghiou, A.T., Noble, J.A.: ScanAhead: Simplifying standard plane acquisition of fetal head ultrasound. Medical Image Analysis **104**, 103614 (2025). https://doi.org/10.1016/j.media.2025.103614
20. NHS: 20-week screening scan. https://www.gov.uk/government/publications/fetal-anomaly-screening-programme-handbook/20-week-screening-scan (2025), last accessed 2 June 2026
21. The Society of Radiographers: Ultrasound Workforce UK Census 2019. https://www.sor.org/getmedia/cb5f34dd-15b2-4595-a37b-7ebdc4bb8d4f/ultrasound_workforce_uk_census_2019.pdf_2 (2019), last accessed 2 June 2026

22. The Society of Radiographers: SoR publishes case studies on sonographer training capacity. https://www.sor.org/news/ultrasound/sor-publishes-case-studies-on-sonographer-training (2025), last accessed 2 June 2026
23. The Society of Radiographers: Sonography vacancy rates have increased dramatically, SoR ultrasound census reveals. https://www.sor.org/news/ultrasound/sonography-vacancy-rates-have-increased-dramatical (2026), last accessed 2 June 2026
24. Wang, Y., Yang, Q., Drukker, L., Papageorghiou, A., Hu, Y., Noble, J.A.: Task model-specific operator skill assessment in routine fetal ultrasound scanning. International Journal of Computer Assisted Radiology and Surgery **17**(8), 1437–1444 (2022). https://doi.org/10.1007/s11548-022-02642-y
25. Waring, L., Miller, P.K., Sloane, C., Bolton, G.: Charting the practical dimensions of understaffing from a managerial perspective: The everyday shape of the UK's sonographer shortage. Ultrasound **26**(4), 206–213 (2018). https://doi.org/10.1177/1742271X18772606
26. Won, D., Walker, J., Horowitz, R., Bharadwaj, S., Carlton, E., Gabriel, H.: Sound the alarm: The sonographer shortage is echoing across healthcare. Journal of Ultrasound in Medicine **43**(7), 1289–1301 (2024). https://doi.org/10.1002/jum.16453
27. Xu, H., Wu, J., Cao, G., Chen, Z., Lei, Z., Liu, H.: Transforming surgical interventions with embodied intelligence for ultrasound robotics. In: Linguraru, M.G., Dou, Q., Feragen, A., Giannarou, S., Glocker, B., Lekadir, K., Schnabel, J.A. (eds.) MICCAI 2024. LNCS, vol. 15006, pp. 703–713. Springer, Cham (2024). https://doi.org/10.1007/978-3-031-72089-5_66