Supplementary Materials for Task Model-Specific Operator Skill Assessment in Routine Fetal Ultrasound Scanning

Yipei Wang[1*], Qianye Yang[2], Lior Drukker[3], Aris Papageorghiou[3], Yipeng Hu[1,2] and J. Alison Noble[1]

[1*]Institute of Biomedical Engineering, Department of Engineering Science, University of Oxford, Oxford, UK.
[2]Department of Medical Physics and Biomedical Engineering, University College London, London, UK.
[3]Nuffield Department of Women's & Reproductive Health, University of Oxford, Oxford, UK.

*Corresponding author(s). E-mail(s): yipei.wang@eng.ox.ac.uk;

Table 1 Summary of data contribution from the 11 sonographers (S1-S11) used in this study. HC=head circumference, AC=abdominal circumference.

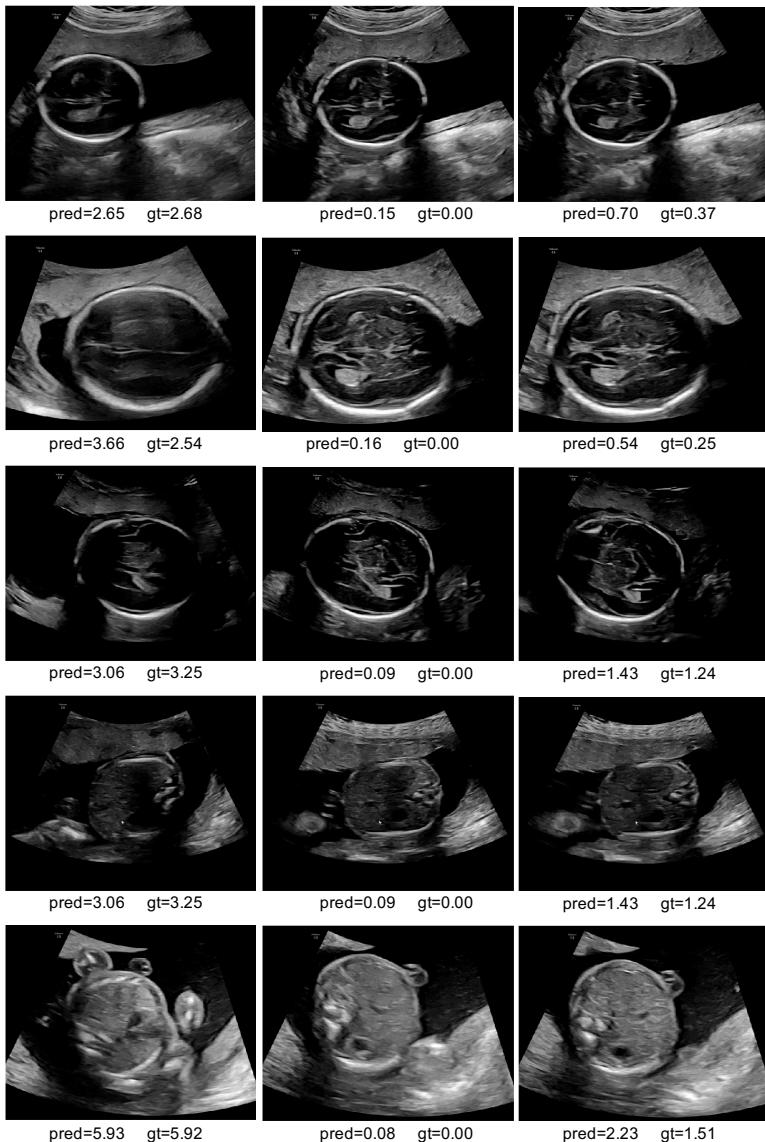| Sonographer		S1	S2	S3	S4	S5	S6	S7	S8	S9	S10	S11	Sum
HC	#Clips	83	11	5	2	1	3	3	1	8	28	1	146
	#Frames	26211	3237	1529	568	212	990	1080	336	2376	9430	360	46329
AC	#Clips	84	7	2	3	0	4	3	0	14	28	4	149
	#Frames	27401	2022	561	930	0	1148	932	0	4394	9021	610	47019

Table 2 Dataset partitioning used in the experiments. HC=head circumference, AC=abdominal circumference, S1-S11 are the 11 sonographers.
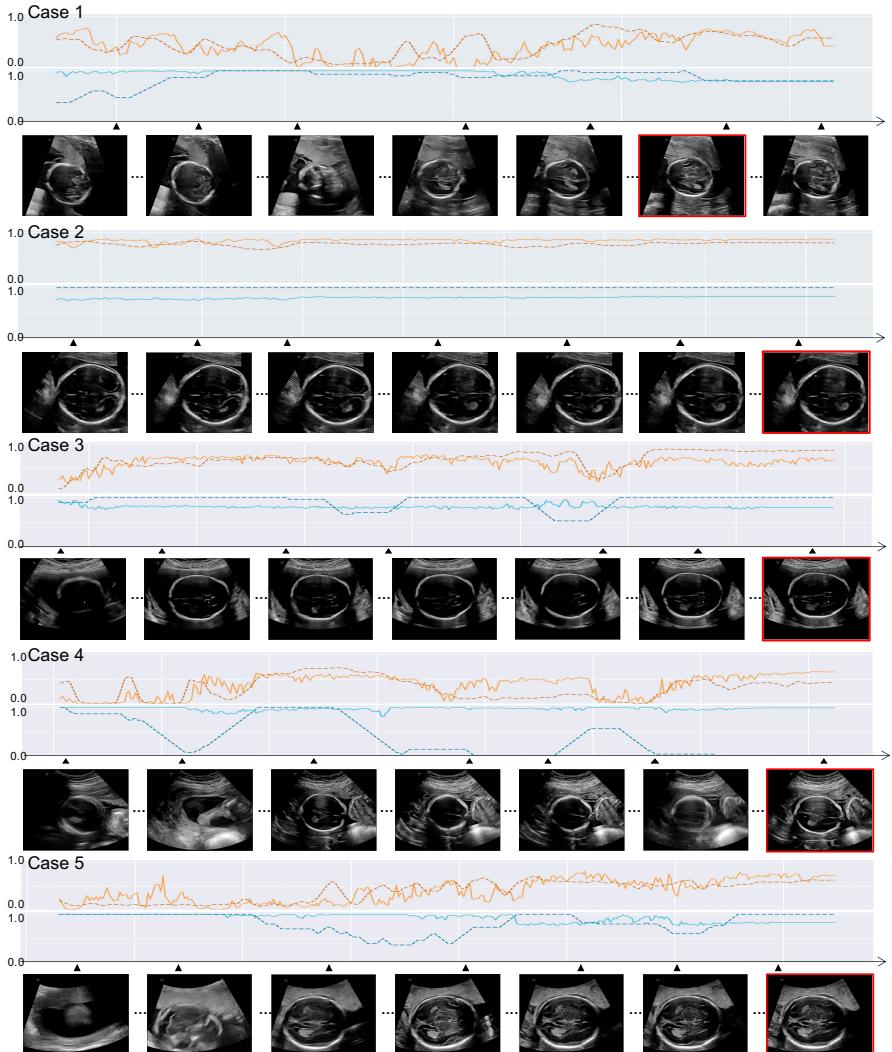
Task	Dataset	Development set					Test set							
		Operator				#Clips	Operator							#Clips
HC	A	S1	S2	S9	S10	130	S3	S4	S5	S6	S7	S8	S11	16
	B	S2	S6	S9	S10	50	S1	S3	S4	S5	S7	S8	S11	96
AC	C	S1	S3	S9	S10	128	S2	S4	S6	S7	S11	-	-	21
	D	S2	S3	S9	S10	51	S1	S3	S4	S6	S7	-	-	98

Table 3 Ablation study results for different network architectures at $\delta_{min}=0.04$.

Criterion	δ_{min}	Model	RMSE	PCC	Criterion	δ_{min}	Model	RMSE	PCC
f_{skill}^{acc}	0.01	VGG16	0.380±0.132	0.711±0.366	f_{skill}^{pos}	0.01	VGG16	0.246±0.065	0.759±0.188
		ResNet18	0.391±0.132	0.498±0.276			ResNet18	0.167±0.036	0.886±0.202
	0.02	VGG16	0.328±0.159	0.459±0.397		0.02	VGG16	0.229±0.090	0.510±0.183
		ResNet18	0.305±0.169	0.328±0.447			ResNet18	0.247±0.073	0.611±0.231
	0.03	VGG16	0.341±0.239	-0.033±0.247		0.03	VGG16	0.206±0.063	0.630±0.326
		ResNet18	0.335±0.249	0.141±0.229			ResNet18	0.245±0.076	0.513±0.316
	0.04	VGG16	0.164±0.189	-0.176±0.200		0.04	VGG16	0.171±0.116	0.212±0.257
		ResNet18	0.164±0.196	-0.450±0.246			ResNet18	0.119±0.063	0.773±0.304

pred=2.65   gt=2.68   pred=0.15   gt=0.00   pred=0.70   gt=0.37

pred=3.66   gt=2.54   pred=0.16   gt=0.00   pred=0.54   gt=0.25

pred=3.06   gt=3.25   pred=0.09   gt=0.00   pred=1.43   gt=1.24

pred=3.06   gt=3.25   pred=0.09   gt=0.00   pred=1.43   gt=1.24

pred=5.93   gt=5.92   pred=0.08   gt=0.00   pred=2.23   gt=1.51

Fig. 1 Visual results of the mixed label generator f_{linear} performance. All images in the same row are from the same scanning sequence and, for reference, the middle images are the diagnostic planes selected by sonographers. Both the predicted percentage relative error $\hat{\delta}$ and the ground-truth percentage error δ are presented.

4 *Supplementary Material*

Fig. 2 Additional cases complementing the Section 4.4 and Figure 2 in the main paper. f_{skill}^{pos} and f_{skill}^{acc} are plotted in orange and blue, respectively, with both task model-generated scores and the predicted scores in dotted and solid lines, respectively. The red boxed frames were the manually identified standard planes.