Task-Conditioned Parameter-Efficient Adaptation for Robust Ultrasound Landmark Localization and Biometric Measurement

Pei-Qiu Huang[1][0000−0001−6278−4566], Yu Zhao*,[2][0009−0006−0800−8447], and Jieyu Liu*,[2][0000−0003−4203−9985]

[1] School of Automation, Central South University, Changsha 410083, China
pqhuang@csu.edu.cn, artrainzy@126.com
[2] Department of Ultrasound Diagnostic, The Second Xiangya Hospital of Central South University, Changsha 410011, Hunan, China
liujieyu99@csu.edu.cn

Abstract. Robust ultrasound biometry requires accurate anatomical landmark localization across different anatomies, scanners, and acquisition protocols. This paper presents a task-conditioned heatmap localization framework for landmark localization and biometric measurement in prenatal, intrapartum, pediatric cardiac, and adult cardiac ultrasound images. The core of the method is a task-aware architecture combined with semi-supervised learning. The framework adapts a frozen Ultrasound Representation Foundation Model (URFM) with a Vision Transformer (ViT)-Huge backbone using Low-Rank Adaptation (LoRA), fuses low-, middle-, and last-layer features using task and group conditioning, and predicts landmarks through a group-aware expert adaptor and a multi-readout heatmap head. To exploit unlabeled data, we further use semi-supervised pseudo-labeling: 5,000 images with confidence scores no lower than 0.75 are merged with the manually labeled data for training. On 619 samples from nine tasks, the framework achieves a task-averaged mean radial error (MRE) of 24.26 and reports mean absolute error (MAE) values for 29 biometric endpoints.

Keywords: Ultrasound biometry · Task-conditioned adaptation · Semi-supervised learning · LoRA.

1 Introduction

Quantitative ultrasound biometry is central to perinatal and cardiovascular assessment, including fetal growth and intrapartum guidelines from the International Society of Ultrasound in Obstetrics and Gynecology (ISUOG) [16, 6] and echocardiographic chamber quantification [11]. Although these applications use different biometric parameters, most measurements are derived from predefined

* Corresponding authors.

anatomical landmarks. These measurements may be reported as distances, angles, circumferences, or chamber dimensions, but they share the same computational basis: anatomical points or curves must first be localized in the ultrasound image. Reliable landmark localization is therefore an important step for reproducible ultrasound measurement.

Automatic ultrasound landmark localization remains difficult because speckle noise, acoustic shadowing, view-dependent appearance, and operator-dependent acquisition cause large variation in image appearance. In a multi-scenario setting, the difficulty is not only caused by image quality, but also by the differences among tasks. Some tasks contain only a few stable points, whereas others involve long contours, cardiac structures, or landmarks whose locations are defined by task-specific measurement rules. The amount of manual annotation is also uneven across anatomies, which makes it hard to train a separate robust model for every measurement. These factors require a model that can share useful ultrasound representations while still adapting to different landmark layouts and measurement geometry.

Deep learning methods have shown strong potential in medical image analysis [12] and fetal ultrasound analysis [5], including standard-plane localization [1] and fetal head circumference measurement [8]. Representative segmentation and localization networks, such as U-Net and nnU-Net [15, 10], as well as heatmap-based pose and anatomical landmark models [13, 14], provide useful design references for dense prediction and point localization. However, many existing systems are designed for one anatomy or one measurement setting. A fully task-specific design reduces cross-task sharing and increases training cost, while a fully shared predictor may ignore task-specific landmark definitions, geometry, and scale shifts [2, 7]. A useful framework should therefore combine a strong shared encoder, task-conditioned adaptation, and an effective way to use unlabeled ultrasound images, especially because medical annotations are often scarce or imperfect [3, 18]. In this paper, the method is built around two core ideas: task-aware, i.e., task-conditioned, feature adaptation and semi-supervised pseudo-label learning. The first idea lets the network change its behavior according to the target task, while the second idea increases training diversity without requiring extra manual annotation, following the general principle of teacher-student and high-confidence pseudo-label learning [19, 17, 20].

We propose a task-conditioned multitask heatmap localization framework based on a pre-trained Ultrasound Representation Foundation Model (URFM) with a Vision Transformer (ViT)-Huge backbone. The encoder follows the ViT design [4] and is adapted with Low-Rank Adaptation (LoRA) [9], and low-, middle-, and last-layer Transformer features are fused under task and task-group conditioning. A group-aware expert adaptor models shared, group-specific, and task-specific feature transformations, and a multi-readout head combines heatmaps from multiple feature levels. The adaptor is applied to a single fused representation from one encoder pass, so the method is a task-conditioned feature adaptation framework rather than multi-model fusion. The training strategy further uses semi-supervised pseudo-labeling with 5,000 high-confidence unlabeled

images and selective fine-tuning for difficult tasks, which enlarges the training distribution while keeping the manually labeled validation and testing protocol unchanged.

The main contributions are centered on two core parts. First, we design a task-conditioned adaptation architecture with hierarchical feature fusion, group-aware expert adaptation, and multi-readout heatmap prediction. Second, we introduce a semi-supervised parameter-efficient training strategy that combines LoRA adaptation, pseudo-labeled data, and selective fine-tuning. We also provide a unified landmark-based formulation for multi-scenario ultrasound biometry.

2 Task Description

Let $\mathcal{T} = \{t_1, \ldots, t_M\}$ be the task set. For task t, the labeled data are $\mathcal{D}_t = \{(I_i^t, P_i^t, S_i^t)\}_{i=1}^{N_t}$, where I_i^t is the original image, $P_i^t = \{(x_{i,k}^t, y_{i,k}^t)\}_{k=1}^{K_t}$ contains K_t landmarks, and $S_i^t = (H_i^t, W_i^t)$ is the original image size. After preprocessing, I_i^t becomes X_i^t, and P_i^t is converted to a Gaussian heatmap $M_i^t \in \mathbb{R}^{K_t \times h \times w}$. For unlabeled data $\mathcal{U} = \{(I_i^u, t_i^u, S_i^u)\}_{i=1}^{N_u}$, a teacher model assigns pseudo heatmaps and confidence scores c_i^u. Predictions with $c_i^u \geq \kappa$ are retained, with $\kappa = 0.75$, and the top $N_p = 5{,}000$ images form $\mathcal{D}_p$. The augmented training set for task t is $\mathcal{A}_t = \mathcal{D}_t \cup \mathcal{D}_{p,t}$; pseudo-labeled samples are used only for training. This formulation separates the localization problem from the measurement problem: the network learns task-specific landmark heatmaps, landmark coordinates are recovered using Distribution-Aware Coordinate Representation of Keypoints (DARK) decoding, and biometric values are computed by deterministic measurement functions. This separation allows different clinical measurements to share the same heatmap learning objective, and it also makes pseudo-labeled images easy to add because both manual labels and pseudo labels are represented in the same heatmap form.

The goal is to learn f_Θ that predicts $\widehat{M}_i^t = f_\Theta(X_i^t, t)$. Landmarks and biometric measurements are decoded by

$$\widehat{P}_i^t = \mathrm{DARK}(\widehat{M}_i^t, S_i^t), \qquad \widehat{B}_i^t = \phi_t(\widehat{P}_i^t, S_i^t), \tag{1}$$

where $\phi_t(\cdot)$ is the task-specific measurement function. The frozen encoder has parameters θ_0, and the trainable set is $\Theta = \{\theta_L, \theta_N, \theta_E, \theta_R\}$ for LoRA, the fusion neck, the expert adaptor, and task-specific readout heads. Each task has an embedding e_t, a weight w_t, and a group $\gamma(t) \in \mathcal{G}$ with group embedding $q_{\gamma(t)}$. The selected tasks and parameters for the second training stage are denoted by $\mathcal{T}_s \subseteq \mathcal{T}$ and $\Omega_s \subseteq \Theta$.

3 Method

3.1 Overall Framework

The framework follows a semi-supervised heatmap localization pipeline. It is built around two core parts: a task-conditioned adaptation path and a semi-

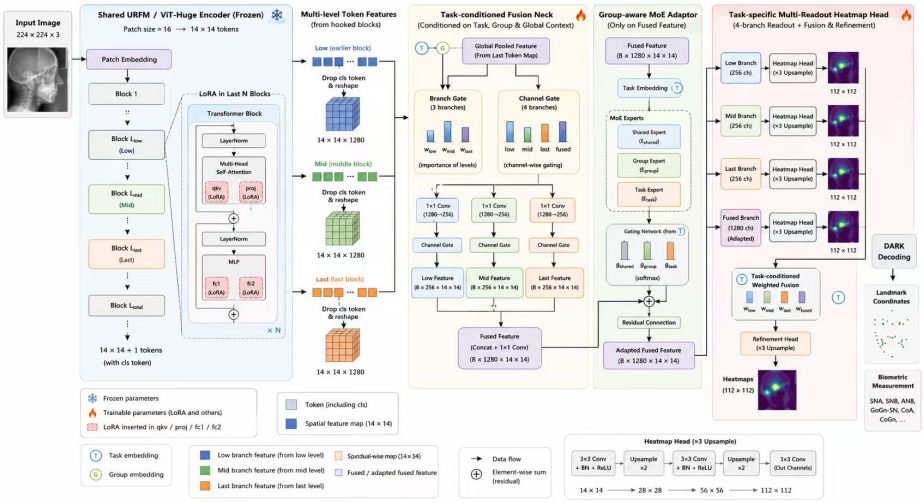

Fig. 1. Overall workflow of the proposed method.

supervised training path. First, an ultrasound image is processed by a frozen URFM/ViT-Huge encoder, where LoRA parameters are inserted into selected attention and multi-layer perceptron (MLP) layers in the last Transformer blocks. For a frozen linear layer W, LoRA computes $y = Wx + \frac{\alpha_L}{r_L} CAx$, where A and C are trainable low-rank matrices and r_L is the rank. Then, low-, middle-, and last-layer token features are reshaped into spatial maps and fused by a task-conditioned neck. Next, a group-aware expert adaptor refines the fused feature, and a task-specific multi-readout head predicts landmark heatmaps. During training, the model uses both manually labeled samples and 5,000 high-confidence pseudo-labeled images selected with $\kappa = 0.75$. Finally, DARK decoding converts heatmaps to landmark coordinates, and task-specific measurement functions compute biometric values. Fig. 1 gives the workflow, Algorithm 1 gives the training/inference procedure, and Sections 3.2–3.5 describe the key operations. The feature fusion in Sec. 3.2 explains how multi-level tokens are selected for each task, Sec. 3.3 describes how the fused feature is adapted by shared, group, and task experts, Sec. 3.4 defines the heatmap readout, and Sec. 3.5 details the semi-supervised two-stage training process.

The overall design is driven by two requirements. First, a unified model should avoid retraining a full encoder for every anatomy, because the amount of labeled data differs across tasks and full fine-tuning can easily overfit small tasks. LoRA therefore provides a compact way to adapt the frozen encoder while preserving the general ultrasound representation. Second, the network must keep task-specific geometry visible to the decoder. Task and group embeddings are used in the fusion, expert, and readout modules so that the same backbone can produce different localization behavior for different measurement settings.

Algorithm 1 Task-conditioned semi-supervised adaptation

Require: $\mathcal{T}$, $\{\mathcal{D}_t\}$, $\mathcal{U}$, $\gamma(t)$, $f_{\Theta^{\mathrm{T}}}$, $\kappa = 0.75$, $N_p = 5{,}000$, $\mathcal{T}_s$, θ_0, $\Theta = \{\theta_L, \theta_N, \theta_E, \theta_R\}$, E_1, E_2.

Ensure: Θ, predicted landmarks $\widehat{P}$, and measurements $\widehat{B}$.

Freeze θ_0, insert θ_L, and initialize θ_N, θ_E, and θ_R.
Forward modules: task-conditioned fusion (Sec. 3.2), group-aware adaptation (Sec. 3.3), and multi-readout prediction (Sec. 3.4).
Predict $\{(\widetilde{M}_i^u, c_i^u)\}_{i=1}^{N_u}$ on $\mathcal{U}$ with $f_{\Theta^{\mathrm{T}}}$.
Select the top N_p predictions satisfying $c_i^u \geq \kappa$ to form $\mathcal{D}_p$; set $\mathcal{A}_t = \mathcal{D}_t \cup \mathcal{D}_{p,t}$.
Stage 1: semi-supervised joint adaptation (Sec. 3.5)
for $e = 1, \ldots, E_1$ and $t \in \mathcal{T}$ **do**
    Sample $\mathcal{B} \subset \mathcal{A}_t$, optionally apply mixup-style heatmap interpolation (MixUp) to obtain $(\widetilde{X}_j^t, \widetilde{M}_j^t)$.
    Predict $\widehat{M}_j^t = f_\Theta(\widetilde{X}_j^t, t)$ and update $\Omega_1 = \Theta$ by minimizing $\mathcal{L}_t$.
end for
Stage 2: selective fine-tuning (Sec. 3.5)
for $e = 1, \ldots, E_2$ and $t \in \mathcal{T}_s$ **do**
    Sample $\mathcal{B} \subset \mathcal{A}_t$, predict $\widehat{M}_j^t = f_\Theta(X_j^t, t)$, and update Ω_s by minimizing $\mathcal{L}_t$.
end for
Inference (Sec. 3.5)
Predict $\widehat{M}^* = f_\Theta(X^*, t^*)$, decode $\widehat{P}^* = \mathrm{DARK}(\widehat{M}^*, S^*)$, and compute $\widehat{B}^* = \phi_{t^*}(\widehat{P}^*, S^*)$.

3.2 Task-conditioned Hierarchical Feature Fusion

A single Transformer level is often insufficient for ultrasound biometry because different landmarks rely on different visual evidence. Boundary-related points benefit from local texture and edge responses, while cardiac and contour-based measurements need broader structural context. Given an input $\widetilde{X}_j^t$, the adapted encoder therefore returns low-, middle-, and last-level token features. After removing the class token, they are reshaped into spatial maps $(F_{j,l}^t, F_{j,m}^t, F_{j,h}^t)$. The fusion neck projects them to branch features, computes a global average pooling (GAP) descriptor, and generates task/group-conditioned gates:

$$\begin{aligned} G_{j,s}^t &= \psi_s(F_{j,s}^t), \quad r_j^t = \xi([\mathrm{GAP}(G_{j,l}^t), \mathrm{GAP}(G_{j,m}^t), \mathrm{GAP}(G_{j,h}^t)]), \\ a_j^t &= 0.75 + \sigma(g_a(r_j^t) + \tau g_t(e_t) + g_g(q_{\gamma(t)})), \quad c_j^t = 0.5 + \sigma(g_c([r_j^t, e_t])). \end{aligned} \tag{2}$$

For $s \in \{l, m, h\}$, the branch output is $U_{j,s}^t = a_{j,s}^t(c_{j,s}^t \odot G_{j,s}^t)$, and the fused feature is

$$U_{j,f}^t = c_{j,f}^t \odot \psi_f([U_{j,l}^t, U_{j,m}^t, U_{j,h}^t]). \tag{3}$$

The branch gate a_j^t adjusts the contribution of each feature level, and the channel gate c_j^t selects task-relevant channels inside each level. This design lets each task select the feature levels and channels that match its anatomical structure without creating a separate encoder. It is especially useful when tasks have different scales, because small local measurements can emphasize lower-level details, whereas large contours can draw more information from middle and high-level features.

3.3 Group-aware Expert Adaptation

The group-aware adaptor is applied only to the fused feature $U^t_{j,f}$, so it is not a fusion of multiple complete models. It contains a shared expert E_{sh}, a group expert $E_{\gamma(t)}$, and a task expert E_t:

$$\pi^t = \mathrm{softmax}(g_E(e_t) + b^t_E). \tag{4}$$

With $E^j_{\mathrm{sh}} = E_{\mathrm{sh}}(U^t_{j,f})$, $E^j_{\mathrm{g}} = E_{\gamma(t)}(U^t_{j,f})$, and $E^j_{\mathrm{t}} = E_t(U^t_{j,f})$, the adapted feature is

$$V^t_j = \pi^t_{\mathrm{sh}} E^j_{\mathrm{sh}} + \lambda_{\mathrm{g}} \pi^t_{\mathrm{g}} E^j_{\mathrm{g}} + \lambda_{\mathrm{t}} \pi^t_{\mathrm{t}} E^j_{\mathrm{t}} + \rho U^t_{j,f}. \tag{5}$$

The residual term preserves stable shared information, while the experts provide shared, group-specific, and task-specific corrections for different landmark layouts. The shared expert learns transformations common to ultrasound landmark localization, the group expert captures similarities among related views or anatomical structures, and the task expert provides a small correction for the final measurement geometry. This structure increases adaptation capacity after feature fusion, but it still uses one backbone and one forward path.

3.4 Multi-readout Heatmap Prediction

The task-specific head reads four branches: low, middle, last, and fused. This is useful because the best readout level can vary with the endpoint. For example, short distance measurements may depend on sharp local responses, while chamber and circumference measurements need more global shape information. For $s \in \{l, m, h\}$, $H^t_{j,s} = R^t_s(U^t_{j,s})$, and $H^t_{j,f} = R^t_f(V^t_j)$. The task embedding produces readout weights $\beta^t = \mathrm{softmax}(g_R(e_t))$, and the final heatmap is

$$\bar{H}^t_j = \sum_{s \in \{l,m,h,f\}} \beta^t_s H^t_{j,s}, \qquad \widehat{M}^t_j = \bar{H}^t_j + \eta R^t_\Delta(\bar{H}^t_j). \tag{6}$$

The refinement branch R^t_Δ learns a residual correction on the weighted heatmap rather than replacing the main prediction. This multi-readout design balances fine boundary cues, structural cues, semantic context, and fused task-conditioned information.

3.5 Semi-supervised Two-stage Training and Inference

Training minimizes the heatmap loss

$$\mathcal{L}_t = \frac{w_t}{B} \sum_{j=1}^{B} \left\| \widehat{M}^t_j - \widetilde{M}^t_j \right\|_2^2, \tag{7}$$

where $\widetilde{M}^t_j$ is either a manual target heatmap or a pseudo target heatmap. The pseudo-labeling step is performed before the final training stages. The teacher

model predicts heatmaps on the unlabeled pool, and only samples with confidence scores no lower than $\kappa = 0.75$ are considered; the top 5,000 samples are merged with the manually labeled data. This filtering is important because pseudo labels are helpful only when they add appearance diversity without introducing too much label noise.

Stage 1 is the semi-supervised joint adaptation stage: it updates all trainable modules on $\{\mathcal{A}_t\}$ so that the LoRA parameters, fusion neck, expert adaptor, and readout heads learn from both manual and pseudo-labeled heatmaps. This stage builds a common task-conditioned representation for all measurement settings. Stage 2 is selective fine-tuning: it updates only Ω_s on selected tasks $\mathcal{T}_s$, focusing capacity on harder tasks while keeping the shared representation stable. At inference, no pseudo label is used; the trained model predicts $\widehat{M}^*$, decodes $\widehat{P}^*$, and computes $\widehat{B}^*$ for the requested task.

4 Experiments and Results

We evaluate the proposed framework on nine ultrasound landmark localization and biometric measurement tasks. All experiments are based on the dataset released for the Foundation-Model-for-Ultrasound Biometry Challenge at the Medical Image Computing and Computer Assisted Intervention (MICCAI) 2026 conference, including the labeled task data and the unlabeled images used for pseudo-label generation. The evaluation set contains 619 samples in total and covers fetal abdominal circumference (FA), inferior vena cava (IVC), intrapartum angle of progression (AOP), parasternal short-axis echocardiography (PSAX), parasternal long-axis echocardiography (PLAX), fetal ultrasound gland/cervix measurement (FUGC), fetal femur measurement, fetal head circumference (HC), and apical four-chamber echocardiography (A4C). The main localization metric is mean radial error (MRE), and measurement errors are reported as mean absolute error (MAE).

Table 1. Task-level evaluation results.

Task	Samples	Avg. MRE ↓	Avg. MAE ↓
FA	188	18.84	77.63
IVC	10	30.29	14.41
AOP	60	13.59	9.40
PSAX	18	36.59	20.40
PLAX	26	17.14	8.60
FUGC	20	9.51	8.79
Fetal femur	62	22.65	20.68
HC	215	45.87	74.50
A4C	20	23.82	20.68
Overall	619	**24.26**	–

Table 1 shows a task-averaged MRE of 24.26 across the nine evaluation tasks. FUGC, AOP, and PLAX have the lowest MREs, while HC and PSAX are harder because of larger contours and stronger view variation. This pattern is consistent with the task design: sparse or locally constrained measurements are easier to

localize, whereas contour-based and cardiac view-variable tasks require more stable global structure modeling.

Table 2. Endpoint-level biometric measurement errors. Pixel-based measurements are reported in px, and the AOP angle is reported in degrees.

Task	Endpoint	MAE ↓	Task	Endpoint	MAE ↓
FA	FA	77.63 px	PLAX	STJ	7.41 px
IVC	IVC	14.41 px	PLAX	VAO	8.89 px
AOP	AOP	8.48°	FUGC	CL	8.79 px
AOP	HSD	10.31 px	Fetal femur	FL	20.68 px
PSAX	PA	23.68 px	HC	BPD	16.47 px
PSAX	RVOT	17.12 px	HC	HC	132.53 px
PLAX	AAO	9.74 px	A4C	LA horiz.	15.89 px
PLAX	AV	7.67 px	A4C	LA vert.	16.52 px
PLAX	IVS	4.98 px	A4C	LV horiz.	34.46 px
PLAX	LA	10.74 px	A4C	LV vert.	12.31 px
PLAX	LV	13.10 px	A4C	RA horiz.	30.01 px
PLAX	LVOT	7.18 px	A4C	RA vert.	16.54 px
PLAX	LVPW	5.52 px	A4C	RV horiz.	24.60 px
PLAX	RV	8.74 px	A4C	RV vert.	15.15 px
PLAX	RVOT	10.61 px			

The endpoint-level results in Table 2 provide a more detailed view of the measurement behavior. For PLAX, most endpoint errors remain within 4.98-13.10 px, suggesting that the model can handle several cardiac structures within the same view. For AOP, the angle error is 8.48° and the head-symphysis distance error is 10.31 px, showing that the localized landmarks can support both angular and distance-based measurements. The largest error appears on HC, where circumference estimation amplifies local landmark errors along a large contour. These observations show that the pipeline supports multiple measurement settings, with large-contour and view-variable cases remaining the main difficulties.

5 Conclusion

This paper addresses robust ultrasound landmark localization and biometric measurement across multiple clinical scenarios. The proposed framework combines parameter-efficient adaptation of a frozen URFM/ViT-Huge encoder, task-conditioned hierarchical feature fusion, group-aware expert adaptation, multi-readout heatmap prediction, and semi-supervised pseudo-label expansion. Experiments on 619 samples from nine tasks show a task-averaged MRE of 24.26 and report measurement errors for 29 biometric endpoints. The results suggest that a unified task-conditioned model can support diverse ultrasound measurement settings, while the higher errors on HC and PSAX indicate that large-contour measurements and view-variable cardiac cases remain challenging. Future work will explore contour-aware objectives and uncertainty-aware pseudo-label selection to further improve these difficult cases.

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). https://doi.org/10.1109/TMI.2017.2712367
2. Caruana, R.: Multitask learning. Machine Learning **28**(1), 41–75 (1997). https://doi.org/10.1023/A:1007379606734
3. Cheplygina, V., de Bruijne, M., Pluim, J.P.W.: Not-so-supervised: A survey of semi-supervised, multi-instance, and transfer learning in medical image analysis. Medical Image Analysis **54**, 280–296 (2019). https://doi.org/10.1016/j.media.2019.03.009
4. Dosovitskiy, A., Beyer, L., Kolesnikov, A., Weissenborn, D., Zhai, X., Unterthiner, T., Dehghani, M., Minderer, M., Heigold, G., Gelly, S., Uszkoreit, J., Houlsby, N.: An image is worth 16x16 words: Transformers for image recognition at scale. In: International Conference on Learning Representations (2021), https://openreview.net/forum?id=YicbFdNTTy
5. Fiorentino, M.C., Villani, F.P., Di Cosmo, M., Frontoni, E., Moccia, S.: A review on deep-learning algorithms for fetal ultrasound-image analysis. Medical Image Analysis **83**, 102629 (2023). https://doi.org/10.1016/j.media.2022.102629
6. Ghi, T., Eggebø, T., Lees, C., Kalache, K., Rozenberg, P., Youssef, A., Salomon, L.J., Tutschek, B.: ISUOG practice guidelines: intrapartum ultrasound. Ultrasound in Obstetrics & Gynecology **52**(1), 128–139 (2018). https://doi.org/10.1002/uog.19072
7. Guan, H., Liu, M.: Domain adaptation for medical image analysis: A survey. IEEE Transactions on Biomedical Engineering **69**(3), 1173–1185 (2022). https://doi.org/10.1109/TBME.2021.3117407
8. van den Heuvel, T.L.A., de Bruijn, D., de Korte, C.L., Ginneken, B.v.: Automated measurement of fetal head circumference using 2d ultrasound images. PLOS ONE **13**(8), e0200412 (2018). https://doi.org/10.1371/journal.pone.0200412
9. Hu, E.J., Shen, Y., Wallis, P., Allen-Zhu, Z., Li, Y., Wang, S., Wang, L., Chen, W.: LoRA: Low-rank adaptation of large language models. In: International Conference on Learning Representations (2022), https://openreview.net/forum?id=nZeVKeeFYf9
10. 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 (2021). https://doi.org/10.1038/s41592-020-01008-z
11. Lang, R.M., Badano, L.P., Mor-Avi, V., Afilalo, J., Armstrong, A., Ernande, L., Flachskampf, F.A., Foster, E., Goldstein, S.A., Kuznetsova, T., Lancellotti, P., Muraru, D., Picard, M.H., Rietzschel, E.R., Rudski, L., Spencer, K.T., Tsang, W., Voigt, J.U.: Recommendations for cardiac chamber quantification by echocardiography in adults: an update from the american society of echocardiography and the european association of cardiovascular imaging. European Heart Journal – Cardiovascular Imaging **16**(3), 233–270 (2015). https://doi.org/10.1093/ehjci/jev014
12. Litjens, G., Kooi, T., Bejnordi, B.E., Setio, A.A.A., Ciompi, F., Ghafoorian, M., van der Laak, J.A.W.M., van Ginneken, B., Sánchez, C.I.: A survey on deep learning in medical image analysis. Medical Image Analysis **42**, 60–88 (2017). https://doi.org/10.1016/j.media.2017.07.005
13. Newell, A., Yang, K., Deng, J.: Stacked hourglass networks for human pose estimation. In: Computer Vision – ECCV 2016. pp. 483–499. Springer (2016). https://doi.org/10.1007/978-3-319-46484-8_29
14. Payer, C., Stern, D., Bischof, H., Urschler, M.: Integrating spatial configuration into heatmap regression based CNNs for landmark localization. Medical Image Analysis **54**, 207–219 (2019). https://doi.org/10.1016/j.media.2019.03.007

15. Ronneberger, O., Fischer, P., Brox, T.: U-Net: Convolutional networks for biomedical image segmentation. In: Medical Image Computing and Computer-Assisted Intervention – MICCAI 2015. pp. 234–241. Springer (2015). https://doi.org/10.1007/978-3-319-24574-4_28
16. Salomon, L.J., Alfirevic, Z., Da Silva Costa, F., Deter, R.L., Figueras, F., Ghi, T., Glanc, P., Khalil, A., Lee, W., Napolitano, R., Papageorghiou, A., Sotiriadis, A., Stirnemann, J., Toi, A., Yeo, G.: ISUOG practice guidelines: ultrasound assessment of fetal biometry and growth. Ultrasound in Obstetrics & Gynecology **53**(6), 715–723 (2019). https://doi.org/10.1002/uog.20272
17. Sohn, K., Berthelot, D., Carlini, N., Zhang, Z., Zhang, H., Raffel, C.A., Cubuk, E.D., Kurakin, A., Li, C.L.: FixMatch: Simplifying semi-supervised learning with consistency and confidence. In: Advances in Neural Information Processing Systems. vol. 33, pp. 596–608 (2020), https://arxiv.org/abs/2001.07685
18. Tajbakhsh, N., Jeyaseelan, L., Li, Q., Chiang, J.N., Wu, Z., Ding, X.: Embracing imperfect datasets: A review of deep learning solutions for medical image segmentation. Medical Image Analysis **63**, 101693 (2020). https://doi.org/10.1016/j.media.2020.101693
19. Tarvainen, A., Valpola, H.: Mean teachers are better role models: Weight-averaged consistency targets improve semi-supervised deep learning results. In: Advances in Neural Information Processing Systems. vol. 30, pp. 1195–1204 (2017), https://arxiv.org/abs/1703.01780
20. Xie, Q., Luong, M.T., Hovy, E., Le, Q.V.: Self-training with noisy student improves ImageNet classification. In: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. pp. 10687–10698 (2020). https://doi.org/10.1109/CVPR42600.2020.01070