MABA: Measurement-Aware Biometry Adaptation for Landmark-Based Ultrasound

Zhaorui Fan, Liwen Wang, Xingbo Dong*, and Zhe Jin

School of Artificial Intelligence, Anhui University, Hefei, China
3222855695@qq.com, liwenwang@stu.ahu.edu.cn
xingbo.dong@ahu.edu.cn, jinzhe@ahu.edu.cn

Abstract. This technical report describes our system for the FU_Biometry challenge, which evaluates landmark localization and biometric-parameter error across prenatal, intrapartum, and cardiac ultrasound. MABA combines a self-supervised encoder, task-routed heatmap heads, differentiable decoding, measurement-aware objectives, and gated pseudo labels. The standalone student scored 30.32 on visible validation; separate prediction-level Donor Fusion audits scored 24.2509 and 24.0358 MRE (30.3981 MAE) on the refreshed 619-row manifest, versus 32.2257 for the organizer baseline. These audits are not new trained models or hidden-test results.

Keywords: Ultrasound biometry · Landmark detection · Measurement-aware learning · Foundation models · Challenge report

1 Introduction

Ultrasound biometry supports fetal growth assessment, intrapartum monitoring, and cardiac structural measurement. It includes fetal head, abdomen, and femur measurements, labor-progress parameters, and cardiac chamber or vessel diameters. In routine practice, these measurements are affected by operator variability, scanner differences, speckle noise, acoustic shadowing, fetal or cardiac motion, and off-plane acquisition [16,6,2,5]. Landmark-based automation keeps an interpretable intermediate representation: anatomical points are localized first, and task rules convert those points into distances, circumferences, angles, or other biometric parameters [2,4,7,11].

The FU_Biometry challenge is a multi-domain landmark benchmark [10,9]: one system must cover prenatal, intrapartum, pediatric, and adult cardiac tasks with heterogeneous measurement definitions and long-tailed labels. AOP has 4000 local rows, whereas IVC, PSAX, PLAX, and A4C have only 38, 49, 87, and 108.

The public organizer repository defines the baseline code, unified-model rule, and output format [9,8]. Its released package scores 32.2257 Average MRE and 34.00 mean MAE on the refreshed 619-row evaluation, anchoring comparison because visible labels are unavailable.

* Corresponding author.

MABA adapts a self-supervised ViT encoder through routed heatmap heads and differentiable biometric readouts [14]. Deterministic heads, measurement readouts, and task-aware supervision turn the encoder into a challenge system. We call this trained single-encoder package the *standalone MABA student*. Donor Fusion is instead a post-hoc replacement of selected prediction rows for an online audit, not a second feature extractor, a weight-level fusion module, or the deployed model.

Contributions. First, we formulate a measurement-aware adaptation framework that jointly optimizes heatmap localization, decoded coordinates, anatomical geometry, and biometric measurements under deterministic task routing. Second, we develop a long-tail training strategy that combines task-balanced resampling with confidence-gated pseudo labels. Third, we provide a transparent evaluation protocol that distinguishes the standalone MABA model from post-inference prediction audits and reports task-level official feedback.

2 Related Work and Positioning

Generic landmark detection often starts from heatmap regression, followed by discrete or continuous coordinate decoding. Stacked hourglass and HRNet-style networks are common high-resolution baselines because they preserve spatial detail through repeated refinement or parallel multi-resolution branches [13,17]. Medical landmark methods further add spatial-configuration priors so that predictions respect plausible anatomy [15]. These ideas motivate our use of dense heatmaps, soft-argmax decoding, and geometry-aware loss terms.

Ultrasound biometry requires handling speckle, shadowing, foreshortening, and anatomical ambiguity [6,2,5,3]. FU_Biometry extends plane-specific settings to one submission spanning prenatal, intrapartum, pediatric, and adult cardiac images with different landmark counts and measurement rules.

The organizer baseline is a multi-head EfficientNet-B4 with deterministic task routing [8]. We retain the unified-submission constraint but replace coordinate-only supervision with heatmap decoding, measurement-coupled losses, and pseudo labels [18,1]. DINOv2 provides transferable visual features, while task heads and measurement readouts provide the challenge-specific interface [14].

3 Task and Data

FU_Biometry evaluates landmark prediction and derived ultrasound measurement estimation [10,8]. The official score equally weights MRE and parameter error, and the challenge emphasizes one unified solution across subgroups and imaging domains. We therefore separate method description, official validation, and local diagnostics.

Our local labeled data contain 6768 rows across nine task IDs: A4C, AOP, FA, FUGC, HC, IVC, PLAX, PSAX, and fetal_femur. The visible-validation evaluator contains 619 rows; its task counts are reported in Table 1. Because visible

Table 1. Local labeled training and current visible-validation distribution. "Scarce" marks tasks with fewer than 120 labeled rows, which require explicit attention in sampling, pseudo-labeling, and result interpretation.

Group	Task ID	Labeled train rows	Current visible rows
Scarce	A4C	108	20
Rich	AOP	4000	60
Medium	FA	500	188
Medium	FUGC	260	20
Rich	HC	999	215
Scarce	IVC	38	10
Scarce	PLAX	87	26
Scarce	PSAX	49	18
Medium	fetal_femur	727	62
–	Total	6768	619

Note. Current visible rows are taken from the 2026-07-02 CodaBench evaluator log for the refreshed 619-sample manifest. Official-score rows are reported from archived scoring logs.

CSV files lack ground-truth landmarks, `val_data` metrics are not substitutes for CodaBench feedback.

Scarce tasks have limited labeled support, while measurements span different scales and definitions. Scarce/medium/rich are internal data-volume bins, not clinical categories or difficulty ranks; we therefore report both MRE and MAE per task.

3.1 Terminology and Evidence Protocol

Local proxy metrics are development metrics, not official scores; a *manifest* is the evaluator's expected sample-ID/task list, not training data. *Scarce* denotes fewer than 120 labeled rows; *medium* and *rich* are internal data-volume bins. *Donor Fusion* is a prediction-level JSON audit after coverage, duplicate, and manifest checks, not feature fusion, a backbone change, or a multi-model hidden-test claim.

3.2 Evaluation Setting and Reporting Protocol

The archived student used 5414 training and 1354 labeled-only validation rows for model selection and fixed-protocol diagnosis, not visible evaluation. CodaBench supplies a 619-row raw-image/task manifest without local landmark labels: its score is the visible-validation outcome, while local proxies diagnose components. The archived 23.49 calibration and DINOv2-L/reg4 probe used a 649-row manifest, so they are not direct before-and-after comparisons.

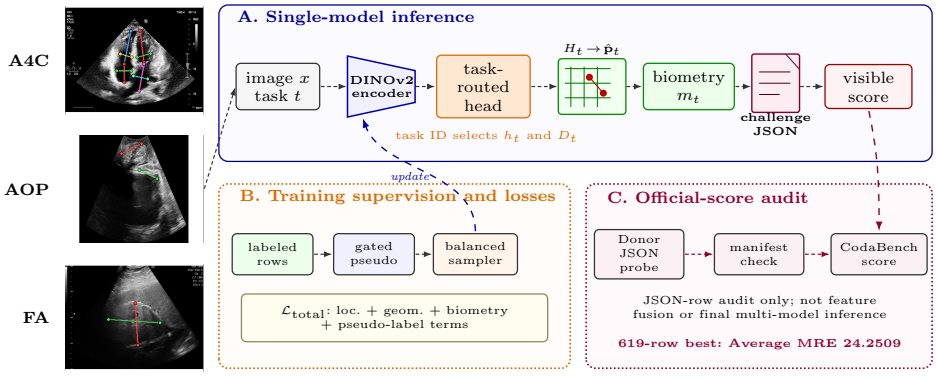

Fig. 1. MABA overview. A single DINOv2 encoder uses task-index routing to produce landmarks and biometric measurements, while training supervision and JSON-level official-score auditing are kept separate.

4 Method

4.1 MABA Architecture and Task Routing

For image x and task identifier t, MABA computes $\mathbf{F} = f_\theta(x)$ and the routed head predicts

$$H_t = h_t(\mathbf{F}) \in \mathbb{R}^{K_t \times 64 \times 64}, \tag{1}$$

where K_t is the task-specific landmark count. The task ID selects the heatmap head and measurement decoder after the shared encoder, retaining one feature extractor with task-specific landmark cardinalities.

Model provenance. The standalone record is the archived 2026-06-06 DINOv2-B ("vit_base_patch14_dinov2") cloud student, submission 775826; later unscored development does not replace it.

4.2 Landmark Decoding and Measurement Consistency

During training, heatmap logits are passed through a sigmoid and decoded by soft-argmax. For landmark k, the decoder is

$$\hat{\mathbf{p}}_k = \sum_{u,v} \operatorname{softmax}(\tau H_{t,k,u,v}) \begin{bmatrix} u/W \\ v/H \end{bmatrix}, \qquad \tau = 45. \tag{2}$$

The decoder collects $\hat{\mathbf{p}} = D_t(H_t)$, normalizes coordinates to $[0, 1]$, and maps them to original pixels. Inference selects task-wise argmax, temperature soft-argmax, or local-window soft-argmax before measurement $m_t(\hat{\mathbf{p}})$ [12,15].

4.3 Measurement-Aware Losses

For task t, ground-truth landmarks $\mathbf{p}$, predictions $\hat{\mathbf{p}}$, heatmap targets G_t, and available measurement targets $\mathbf{y}_t$, MABA jointly supervises landmark localization and derived biometry. The objective extends the coordinate-MSE baseline

[8] with differentiable measurement terms:

$$\begin{aligned}\mathcal{L} = \mathcal{L}_{hm} + \lambda_{coord}\mathcal{L}_{coord} + \lambda_{geom}\mathcal{L}_{geom} + \lambda_{num}\mathcal{L}_{num} + \lambda_{rel}\mathcal{L}_{rel} \\ + \lambda_{perm}\mathcal{L}_{perm} + \lambda_{dir}\mathcal{L}_{dir} + \lambda_{pseudo}\mathcal{L}_{pseudo}.\end{aligned} \tag{3}$$

In the cloud student run, $\lambda_{coord} = 0.008$, $\lambda_{geom} = 0.006$, $\lambda_{num} = 0.008$, $\lambda_{rel} = 0.001$, $\lambda_{perm} = 0.004$, $\lambda_{dir} = 0.004$, $\lambda_{pseudo} = 0.045$.

The heatmap and coordinate terms supervise dense landmark maps and decoded continuous points:

$$\begin{aligned}\mathcal{L}_{hm} &= \frac{1}{K_t HW} \sum_{k,u,v} \left(\sigma(H_{t,k,u,v}) - G_{t,k,u,v}\right)^2, \\ \mathcal{L}_{coord} &= \frac{1}{K_t} \sum_{k=1}^{K_t} \mathrm{SmoothL1}(\hat{\mathbf{p}}_k - \mathbf{p}_k).\end{aligned} \tag{4}$$

Here $\sigma(\cdot)$ is sigmoid and $H = W = 64$. Let $m_t(\cdot)$ be the evaluator's measurement function and $c_{t,r}(\cdot)$ be a geometric constraint. The measurement and geometry terms are

$$\begin{aligned}\mathcal{L}_{num} &= \frac{1}{Q_t} \sum_{q=1}^{Q_t} |m_{t,q}(\hat{\mathbf{p}}) - y_{t,q}|, \\ \mathcal{L}_{rel} &= \frac{1}{Q_t} \sum_{q=1}^{Q_t} \frac{|m_{t,q}(\hat{\mathbf{p}}) - y_{t,q}|}{|y_{t,q}| + \epsilon}, \\ \mathcal{L}_{geom} &= \frac{1}{R_t} \sum_{r=1}^{R_t} \mathrm{SmoothL1}(c_{t,r}(\hat{\mathbf{p}}) - c_{t,r}(\mathbf{p})).\end{aligned} \tag{5}$$

$\mathcal{L}_{num}$ is computed in pixels or degrees; $\mathcal{L}_{rel}$ controls scale imbalance; $\mathcal{L}_{geom}$ includes ellipse-style circumference for FA/HC, angle and HSD for AOP, paired cardiac/vessel diameters for IVC/PLAX/PSAX/A4C, and segment lengths for FUGC/fetal_femur.

For endpoint-ambiguous pairs $(a, b) \in \mathcal{P}_t$, permutation and direction terms are

$$\begin{aligned}\mathcal{L}_{perm} &= \frac{1}{|\mathcal{P}_t|} \sum_{(a,b)} \min\{d_{ab}, d_{ba}\}, \\ \mathcal{L}_{dir} &= \frac{1}{|\mathcal{P}_t|} \sum_{(a,b)} \left(1 - \hat{\mathbf{v}}_{ab}^{\top} \mathbf{v}_{ab}\right),\end{aligned} \tag{6}$$

where d_{ab} is SmoothL1 endpoint distance and d_{ba} swaps the targets; $\hat{\mathbf{v}}_{ab}$ and $\mathbf{v}_{ab}$ are unit predicted and labeled segment directions. For pseudo-labeled rows,

$$\mathcal{L}_{pseudo} = z_i(\mathcal{L}_{hm}^{pseudo} + \mathcal{L}_{coord}^{pseudo} + \mathcal{L}_{geom}^{pseudo}). \tag{7}$$

The gate z_i is one only when confidence and measurement-plausibility checks pass. For the later structured IVC pool, the recorded thresholds were teacher

confidence ≥ 0.88, TTA disagreement ≤ 0.08, selector score ≥ 0.85, and measurement $|z| \leq 2.0$. These thresholds belong to that structured pool rather than retroactively defining every earlier pseudo row. The 473-row cloud run used the same family of confidence, consistency, and measurement checks, but its archived record does not expose a complete numeric threshold vector. Medium-consistency rows are used only as weak or consistency candidates in later audits, while rejected rows are excluded.

4.4 Pseudo-Label Training

In the 2026-06-06 cloud run, score-selected teachers produced 473 pseudo rows for AOP, FA, HC, and IVC (weight 0.75) on the 5414/1354 labeled split. Pseudo rows pass confidence and measurement-plausibility gates; task-balanced resampling, not loss reweighting, uses multipliers 22, 16, 14, 22, 3, 2, and 1 for AOP, FA, HC, IVC, PSAX, PLAX, and other tasks. The structured-IVC thresholds above are not cutoffs for every earlier pseudo row. Training initializes the DINO student from the prior checkpoint, selects by labeled-validation MRE, decodes task-wise, and packages both outputs.

4.5 Standalone MABA and Donor Fusion

Standalone MABA is one encoder with routed heads, measurement-aware losses, and pseudo-label training; Donor Fusion is a separate post-inference audit, not an architectural or deployment claim.

4.6 Submission Verification

Submission verification checks that both required outputs match the active 619-row CodaBench manifest; scores are taken from archived CodaBench logs.

Table 2. Implementation details.

Item	Setting
Environment	RTX 4090 D 24 GB, AMP, PyTorch, organizer baseline code
Model/input	Standalone MABA: `vit_base_patch14_dinov2`; 518×518 input; 64×64 heatmap; $\sigma = 1.8$
Latest development	B5/B7: DINOv2-S, 128×128 heatmaps, ROI/geometry/uncertainty/task adapter; local Docker only
Optimization	AdamW cosine, batch 4, 3 epochs; LR base 2×10^{-7}, heads 2×10^{-6}
Data/loss	5414 labeled + 473 pseudo / 1354 val; coord .008, geom .006, num .008, pseudo .045
Structured gate	IVC pool: confidence $\geq$.88; TTA disagreement $\leq$.08; selector $\geq$.85; $\lvert z\rvert \leq 2.0$
Sampling/decode	AOP 22, FA 16, HC 14, IVC 22, PSAX 3, PLAX 2, others 1; soft/local-window/argmax

5 Experiments and Results

5.1 Official Website Scores

The official CodaBench website is the visible-validation record. Submission 882801 reports 24.035778 Average MRE and 30.398111 Average MAE for `fused_best_v7.zip` on the refreshed 619-row manifest. This five-row IVC overlay is an audit, not a standalone student: standalone MABA is 775826 (30.32), the earlier Donor Fusion audit is 823914 (24.2509), and the overlay is 882801 (24.0358). The rejected FUGC-only selector scored 24.3607; 23.49 used an older 649-row manifest.

Table 3. Visible official validation subgroup summary for the archived Donor Fusion audit submission 823914. Group metrics are unweighted task-level averages. Lower is better.

Group	Tasks	Avg. MRE	Avg. MAE
Scarce	A4C, IVC, PLAX, PSAX	27.31	18.65
Medium	FA, FUGC, fetal_femur	17.14	39.40
Rich	AOP, HC	28.81	42.67
All tasks	9 challenge tasks	**24.2509**	**30.90**

Note. Pixel-scale MRE depends on anatomical scale and measurement definition; these internal data-volume groups are not clinical categories.

Table 4. Per-task official visible-validation results for the current audit submission 882801. Lower is better.

Group	Task	Val. rows	Official MRE	Official MAE	DINOv2-L/reg4 MRE†
Scarce	A4C	20	26.136	19.431	32.3220
Rich	AOP	60	14.262	10.853	81.0673
Medium	FA	188	20.444	87.883	28.9835
Medium	fetal_femur	62	20.399	21.109	26.7827
Medium	FUGC	20	10.572	9.199	5.0502
Rich	HC	215	43.294	73.745	54.0571
Scarce	IVC	10	28.839	21.867	46.6721
Scarce	PLAX	26	16.778	9.209	20.9062
Scarce	PSAX	18	35.598	20.287	48.0876
All tasks	–	619	**24.035778**	**30.398111**	38.2142†

Note. Official columns are from submission 882801; † DINOv2-L/reg4 uses an older 649-row manifest and is diagnostic only.

5.2 Development Diagnostics

Local proxy metrics are diagnostic. DINOv2-L/reg4 scores 11.9273 locally but 38.2142 after upload. FUGC remains relatively strong (5.0502 MRE), whereas AOP (81.0673), HC (54.0571), IVC (46.6721), PSAX (48.0876), and A4C (32.3220) are poor in the older 649-row evaluation. The non-uniform pattern is consistent with low-sample forgetting, task imbalance, decoder or calibration mismatch, and overfitting to the local proxy, rather than with a simple capacity effect; the task-wise values are included in Table 4.

5.3 Ablation Study on Measurement-Aware Losses

Table 5. Local loss ablation. Lower is better.

Variant	Active losses	MRE	Param. error	Composite
Base heatmap only	$\mathcal{L}_{hm}$	20.42	21.17	20.80
+ Geometry consistency	$\mathcal{L}_{hm} + \mathcal{L}_{geom}$	19.94	20.04	19.99
+ Numeric measurement	$\mathcal{L}_{hm} + \mathcal{L}_{geom} + \mathcal{L}_{num}$	**19.88**	**19.91**	**19.89**

Geometry reduces local parameter error from 21.17 to 20.04; the numeric term gives composite 19.89. This is a local component diagnostic, not leaderboard evidence: variants fix the labeled split, encoder, heads, preprocessing, optimizer, sampling schedule, and decoder, changing only active loss terms.

5.4 Qualitative Prediction Checks

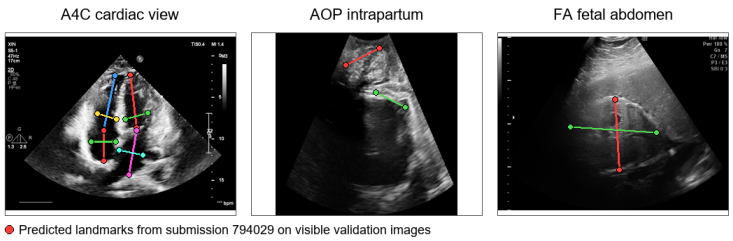

Fig. 2. Qualitative visible-validation predictions on A4C, AOP, and FA. Ground truth is unavailable locally, so overlays show predictions only.

Without local visible labels, these overlays are sanity checks, not quantitative evidence; they illustrate risks from shadowing, off-plane cardiac views, and scale-dependent fetal landmarks.

6 Discussion and Conclusion

The standalone MABA student (30.32) is the paper's runnable model: one DINOv2-B encoder with deterministic task routing, measurement-aware supervision, task-balanced sampling, and gated pseudo-label training. The controlled loss ablation holds the split, encoder, heads, preprocessing, optimizer, sampler, and decoder fixed while changing loss terms. It diagnoses components locally, but does not establish leaderboard improvement or complete official ablations.

The 24.2509 Donor Fusion audit and five-row IVC overlay (24.0358 MRE, 30.3981 MAE) are prediction-level audits, not MABA-network changes; only standalone MABA is a single runnable model. They use the refreshed 619-row manifest, whereas the 23.49 record and DINOv2-L/reg4 probe use the 649-row manifest. The task-level table therefore identifies where the shared student needs calibration and long-tail supervision, rather than supporting a clinical ranking across views with different scales, definitions, and sample counts.

Official ablations for gating, backbone choice, and decoding remain incomplete; the numerical gate vector applies only to the later structured IVC pool, not the archived 473-row run. A new single-container student should isolate these choices on one fixed manifest before device/site validation, calibration, and reader-agreement studies.

References

1. Arazo, E., Ortego, D., Albert, P., O'Connor, N.E., McGuinness, K.: Pseudo-labeling and confirmation bias in deep semi-supervised learning. In: 2020 International Joint Conference on Neural Networks (IJCNN). pp. 1–8 (2020). https://doi.org/10.1109/IJCNN48605.2020.9207304
2. Bai, J., Tang, Y., Liu, X., Hu, J., Li, Y., Chen, X., Wang, Y., Ma, C., et al.: IUGC: A benchmark of landmark detection in end-to-end intrapartum ultrasound biometry. Medical Image Analysis **110**, 103960 (2026). https://doi.org/10.1016/j.media.2026.103960
3. Bai, J., Tang, Y., Zhou, Z., Islam, M., Tabassum, M., Almar-Munoz, E., Liu, H., Meng, H., et al.: FUGC: Benchmarking semi-supervised learning methods for cervical segmentation. IEEE Transactions on Medical Imaging **45**(6), 2950–2960 (2026). https://doi.org/10.1109/TMI.2026.3666364
4. Bai, J., Zhou, Z., Ou, Z., Koehler, G., Stock, R., Maier-Hein, K., Elbatel, M., Marti, R., et al.: PSFHS challenge report: Pubic symphysis and fetal head segmentation from intrapartum ultrasound images. Medical Image Analysis **99**, 103353 (2025). https://doi.org/10.1016/j.media.2024.103353
5. Bai, J., Zhou, Z., Tang, Y., Gan, J., Liang, Z., Fan, J., Mcguire, L.B., Clarke, J.L., et al.: Beyond benchmarks of IUGC: Rethinking requirements of deep learning method for intrapartum ultrasound biometry from fetal ultrasound videos. Medical Image Analysis **111**, 104043 (2026). https://doi.org/10.1016/j.media.2026.104043
6. 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
7. Chen, G., Bai, J., Ou, Z., Lu, Y., Wang, H.: PSFHS: Intrapartum ultrasound image dataset for AI-based segmentation of pubic symphysis and fetal head. Scientific Data **11**(1), 436 (2024). https://doi.org/10.1038/s41597-024-03266-4
8. Deng, B., Tang, Y., Li, J., Huang, Y., Wang, L., Zhang, Y., Zhan, Y., Lu, H., Zhang, X., Bai, J.: Baseline method of the foundation model challenge for ultrasound image analysis. arXiv preprint arXiv:2602.01055 (2026), https://arxiv.org/abs/2602.01055
9. FU_Biometry Challenge Organizers: Foundation-Model-for-Ultrasound-Biometry: Baseline code. https://github.com/lijiake2408/Foundation-Model-for-Ultrasound-Biometry (2026), gitHub repository. Accessed 2026-08-08
10. FU_Biometry Challenge Organizers: FU_Biometry: Foundation model for ultrasound biometry. https://www.codabench.org/competitions/15590/ (2026), codaBench competition page. Accessed 2026-08-08
11. Jiang, J., Wang, H., Bai, J., Long, S., Chen, S., Campello, V.M., Lekadir, K.: Intrapartum ultrasound image segmentation of pubic symphysis and fetal head using dual student-teacher framework with CNN-ViT collaborative learning. In: Medical Image Computing and Computer Assisted Intervention – MICCAI 2024. pp. 448–458. Springer (2024). https://doi.org/10.1007/978-3-031-72378-0_42
12. Li, Y., Yang, S., Liu, P., Zhang, S., Wang, Y., Wang, Z., Yang, W., Xia, S.T.: SimCC: A simple coordinate classification perspective for human pose estimation. In: Computer Vision – ECCV 2022. pp. 89–106. Springer (2022). https://doi.org/10.1007/978-3-031-20068-7_6
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. Oquab, M., Darcet, T., Moutakanni, T., Vo, H., Szafraniec, M., Khalidov, V., Fernandez, P., Haziza, D., Massa, F., El-Nouby, A., Assran, M., Ballas, N., Galuba, W., Howes, R., Huang, P.Y., Li, S.W., Misra, I., Rabbat, M., Sharma, V., Synnaeve, G., Xu, H., Jegou, H., Mairal, J., Labatut, P., Joulin, A., Bojanowski, P.: DINOv2: Learning robust visual features without supervision. arXiv preprint arXiv:2304.07193 (2023), https://arxiv.org/abs/2304.07193
15. 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
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. Sun, K., Xiao, B., Liu, D., Wang, J.: Deep high-resolution representation learning for human pose estimation. In: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. pp. 5686–5696 (2019). https://doi.org/10.1109/CVPR.2019.00584
18. 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 (2017)