Towards Universal Ultrasound Biometry with Task-aware Foundation Models

Taiyu Han [1] Yuan Bi[1] Nassir Navab[1]

[1] Computer Aided Medical Procedures (CAMP),
TU Munich, Germany

yuan.bi@tum.de

Abstract. Ultrasound landmark detection is a fundamental component of intelligent ultrasound biometry, playing a critical role in improving the efficiency and consistency of clinical measurements. However, substantial appearance variations, anatomical diversity, and heterogeneous spatial distributions across different anatomical structures make it difficult for conventional single-task models to achieve unified landmark localization. To address this challenge, we propose a task-aware foundation model for universal ultrasound biometry. The proposed framework employs a frozen DINOv2 backbone as a unified encoder and introduces parameter-efficient task adaptation through Visual Prompt Tuning (VPT) and shared/task-specific Low-Rank Adaptation (LoRA). Furthermore, a task-adaptive heatmap decoder and a multi-level supervision strategy are developed, where diffusion-derived structural anchors generated from unlabeled ultrasound images are exploited to provide auxiliary supervision for learning more robust anatomical representations, thereby improving landmark localization across diverse anatomical structures. Extensive experiments on the official Foundation Model for Ultrasound Biometry Challenge Hidden Validation Set demonstrate that the proposed method enables automatic estimation of more than 20 clinically relevant biometric measurements across prenatal, intrapartum, pediatric, and adult cardiac ultrasound examinations, achieving an average Mean Radial Error (MRE) of 31.14 and an average Mean Absolute Error (MAE) of 33.07. These results demonstrate that parameter-efficient task adaptation effectively unlocks the potential of vision foundation models for universal ultrasound landmark detection and provides a promising paradigm for building unified medical world models for intelligent ultrasound analysis.

Keywords: Ultrasound Biometry, Foundation Model, Landmark Detection.

1 Introduction

Ultrasound imaging is widely used in prenatal screening, cardiovascular diagnosis, and abdominal examination because of its real-time, non-invasive, and low-cost characteristics [1]. Many clinical decisions rely on anatomical measurements, such as fetal head circumference, femur length, cardiac chamber dimensions, and vascular

diameters [2]. These measurements rely on accurate localization of anatomical landmarks [3]. Therefore, ultrasound landmark detection is fundamental to automatic ultrasound biometry and computer-assisted diagnosis. Although deep learning has significantly improved ultrasound landmark detection [4], most existing methods are designed for a single anatomical structure or clinical task. Building a unified model that generalizes across diverse anatomical structures remains challenging.

Compared with natural images, universal ultrasound landmark detection is considerably more difficult. Ultrasound images often suffer from speckle noise, acoustic shadowing, low signal-to-noise ratio, and ambiguous tissue boundaries. Multi-center ultrasound images also exhibit significant appearance variations caused by different hospitals, ultrasound devices, and sonographers, resulting in severe domain shifts [5]. Anatomical structures also vary dramatically in morphology, scale, and spatial configuration. Consequently, a unified model should learn robust visual representations while adapting efficiently to diverse anatomical tasks.

Deep learning-based ultrasound landmark detection has evolved from convolutional neural networks (CNNs) [6] to Vision Transformers (ViTs) [7]. Early approaches mainly adopted CNN-based architectures, such as Hourglass [8] and U-Net [9], to regress landmark heatmaps. Although these methods are effective in capturing local image patterns, their limited receptive fields restrict long-range anatomical reasoning. ViTs alleviate this limitation through global self-attention and have demonstrated superior capability in modeling global contextual information. More recently, vision foundation models, represented by DINOv2 [10] and SAM [11], have achieved remarkable success by leveraging large-scale self-supervised pretraining. Their powerful and transferable visual representations open new opportunities for medical image analysis. Nevertheless, the application of foundation models to ultrasound landmark detection remains limited. Existing studies generally focus on a single anatomical task and rely on full or partial fine-tuning. Such strategies require updating a large number of parameters and fail to balance shared anatomical knowledge with task-specific characteristics [12]. Therefore, how to efficiently adapt a unified foundation model to heterogeneous ultrasound landmark detection tasks remains an open research question.

To advance research on foundation models for ultrasound biometry, the Foundation Model for Ultrasound Biometry Challenge was organized as part of the MICCAI 2026 Medical World Model (MWM) Workshop. The challenge provides a large-scale multi-center ultrasound dataset covering prenatal, intrapartum, pediatric, and adult cardiac examinations. To address these challenges, we propose a Unified Multi-level Task-aware Parameter-efficient Adaptation Framework for universal ultrasound landmark detection. A frozen vision foundation model serves as the shared feature encoder. Task adaptation is achieved collaboratively at the input, encoder, and decoder levels through Visual Prompt Tuning (VPT) and shared/task-specific Low-Rank Adaptation (LoRA), enabling efficient transfer across heterogeneous anatomical structures while introducing only a small number of trainable parameters. Furthermore, we design a multi-level supervision strategy that combines heatmap supervision, coordinate supervision, anatomical geometry constraints, and diffusion-derived structural anchor supervision. The main contributions of this work are summarized as follows:

- We propose a unified task-aware parameter-efficient adaptation framework for universal ultrasound landmark detection. The framework introduces collaborative task adaptation at the input, encoder, and decoder levels, allowing a frozen vision foundation model to effectively capture both shared anatomical knowledge and task-specific representations with minimal trainable parameters.
- We propose a multi-level supervision strategy for robust anatomical landmark localization. The proposed optimization combines heatmap regression, coordinate regression, anatomical geometry constraints, and diffusion-derived structural anchor supervision obtained from unlabeled ultrasound images. This strategy enhances feature learning at multiple representation levels and significantly improves localization accuracy and robustness in challenging ultrasound scenarios.

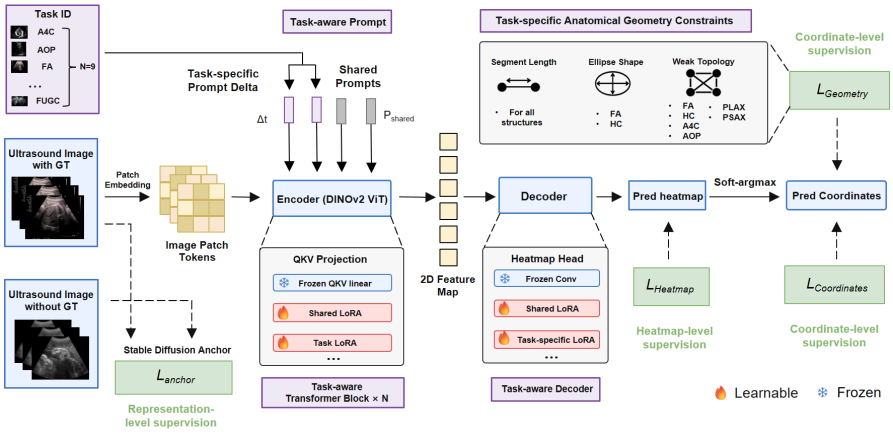

Fig. 1. Overall framework of the proposed multi-level task-aware parameter-efficient adaptation framework with multi-level supervision.

2 Methods

2.1 Task-aware Landmark Detection Network

To enable unified ultrasound landmark detection across diverse anatomical structures, we propose a unified multi-level task-aware parameter-efficient adaptation framework (Fig. 1). A frozen DINOv2 serves as the shared encoder, while task-aware parameter-efficient adaptation is performed at the input, Transformer, and prediction head levels. Multi-level supervision further improves representation learning and landmark localization. The network consists of a shared encoder and a task-adaptive decoder. Given an ultrasound image I and its corresponding task identifier t, the model predicts the landmark heatmaps and landmark coordinates for the target anatomical structure. The input image is first divided into fixed-size patches through the Patch Embedding layer and converted into a sequence of patch tokens.

At the input level, VPT [13] injects task information into the frozen encoder. Instead of learning independent prompts, each prompt is decomposed into a shared component and a task-specific component,

$$\mathrm{P}_t = \mathrm{P}_{shared} + \Delta\mathrm{P}_t \quad (1)$$

where P_{shared} captures visual priors shared across different ultrasound tasks, while ΔP_t models task-specific anatomical characteristics. The prompt tokens are concatenated with the patch tokens before entering the Transformer encoder.

At the Transformer parameter level, we further introduce shared LoRA and task-specific LoRA into the QKV projections of every self-attention layer. Instead of fine-tuning the entire backbone, only low-rank adaptation modules are optimized, while all pretrained DINOv2 parameters remain frozen. The projection matrix can be formulated as:

$$W = W_0 + \Delta W_{shared} + \Delta W_t \quad (2)$$

where W_0 denotes the frozen pretrained weight, ΔW_{shared} represents the shared low-rank adapter that captures common visual patterns across ultrasound tasks, and ΔW_t denotes the task-specific adapter responsible for modeling anatomical differences. By optimizing only these lightweight adaptation modules, the proposed framework preserves the strong representation capability of the pretrained foundation model while achieving parameter-efficient adaptation across heterogeneous anatomical structures.

At the prediction head level, a dedicated prediction head is assigned to each anatomical task to accommodate different numbers and spatial distributions of landmarks. To maintain a consistent parameter-efficient adaptation strategy, the first convolutional layer of each decoder is further equipped with shared and task-specific LoRA branches. The convolution operation can be expressed as

$$Y = f(F) + \Delta f_{shared}(F) + \Delta f_t(F) \quad (3)$$

where $f(\cdot)$ denotes the frozen convolution, and Δf_{shared} and Δf_t denote the shared and task-specific low-rank convolutional adapters, respectively. The decoder outputs task-specific heatmaps, which are converted into landmark coordinates using Soft-argmax.

2.2 Multi-level Supervision

To further improve both feature representation and landmark localization accuracy, we propose a multi-level supervision strategy that jointly optimizes the network from four complementary aspects, namely representation-level supervision, heatmap-level supervision, coordinate-level supervision, and geometry-level supervision. The overall supervision framework is illustrated in Fig. 1.

Inspired by diffusion-based unsupervised learning [14], we introduce an auxiliary anchor prediction head after the encoder to exploit the abundant structural information contained in unlabeled ultrasound images. Rather than directly predicting anatomical landmarks, this branch encourages the encoder to recover diffusion-derived structural anchors, providing additional anatomical supervision on the learned feature representations. Specifically, approximately 60K unlabeled ultrasound images are first processed using a frozen Stable Diffusion v1.5 model. A set of learnable text tokens is optimized, and the corresponding cross-attention maps are used to extract anatomical

attention peaks as pseudo structural anchors. After confidence filtering, these anchors serve as supervisory signals for the auxiliary prediction branch. Given the encoder feature map F, the Anchor Prediction Head predicts: $\widehat{\mathrm{A}} = G(\mathrm{F})$. $G(\cdot)$ denotes the Anchor Prediction Head. The representation-level supervision is defined as

$$\mathcal{L}_{anchor} = MSE(\hat{A}, A) \tag{4}$$

where A represents the diffusion-generated structural anchors. Unlike landmark supervision, the anchor loss is applied only to the encoder representations. It therefore guides the encoder to learn anatomically meaningful features without directly constraining the final landmark predictions, leading to improved generalization across heterogeneous ultrasound datasets.

Heatmap supervision serves as the primary optimization objective of the proposed framework. Given the predicted heatmap $\widehat{\mathrm{H}}$ and the ground-truth heatmap H, the loss is defined as

$$\mathcal{L}_{heatmap} = \mathrm{MSE}(\sigma(\widehat{\mathrm{H}}), \mathrm{H}) \tag{5}$$

where $\sigma(\cdot)$ denotes the Sigmoid activation function.

A Smooth L1 loss $\mathcal{L}_{\text{coord}}$ is applied to directly supervise the predicted landmark coordinates. Although heatmap supervision effectively captures local responses, it does not explicitly enforce anatomical consistency among landmarks. We therefore introduce task-specific anatomical geometry constraints to regularize the predicted landmark configuration according to the corresponding clinical measurement protocol. Specifically, the geometry loss consists of three complementary priors: segment length consistency, ellipse shape preservation, and weak topology regularization for dense landmark configurations. Weak Topology Constraint softly regularizes the distances between predefined anatomically related landmark pairs, encouraging globally consistent landmark configurations without imposing rigid geometric assumptions. These constraints jointly preserve clinically meaningful measurements while maintaining the global anatomical structure. The overall geometry loss is formulated as

$$\mathcal{L}_{\text{geom}} = \mathcal{L}_{\text{segment}} + 0.5\mathcal{L}_{\text{ellipse}} + 0.15\mathcal{L}_{\text{topology}} \tag{6}$$

Finally, the network is optimized using the weighted combination of all supervision terms,

$$\mathcal{L} = \mathcal{L}_{\text{heatmap}} + \lambda_{\text{anchor}}\mathcal{L}_{\text{anchor}} + \mathcal{L}_{\text{coord}} + \lambda_{\text{geom}}\mathcal{L}_{\text{geom}} \tag{7}$$

where λ_{anchor} and λ_{geom} denote the corresponding loss weights. To stabilize optimization, the geometry loss is gradually activated using a linear warm-up strategy, where λ_{geom} increases linearly from zero to its predefined value during the first five training epochs. In addition, the anchor loss weight λ_{anchor} follows a cosine annealing schedule, enabling the auxiliary representation supervision to adapt smoothly throughout the training process.

3 Experiments and Results

3.1 Experimental Settings

The proposed method was evaluated on the multi-center ultrasound landmark detection dataset released by the MWM Workshop. The dataset covers nine anatomical structures

from prenatal, intrapartum, pediatric, and adult cardiac ultrasound, including Apical Four Chamber (A4C), Angle of Progression (AOP), Fetal Abdomen (FA), Fetal Umbilical Gastric (FUGC), Head Circumference (HC), Inferior Vena Cava (IVC), Parasternal Long-Axis (PLAX), Parasternal Short-Axis (PSAX), and fetal femur. It contains more than 190K ultrasound images, among which approximately 6.8K images are manually annotated for clinical biometry, while the remaining unlabeled images are exploited to generate diffusion-derived structural anchors for representation-level supervision. The number of landmarks varies from 2 to 22 across different tasks, reflecting substantial anatomical diversity. During training, the official training set was randomly split into training and validation subsets with a ratio of 8:2, and the same split was used for all comparison and ablation experiments to ensure a fair evaluation. Generalization across imaging centers was further evaluated on the official Hidden Validation Set. Following the challenge protocol, performance was evaluated using Mean Radial Error (MRE) and Mean Absolute Error (MAE), which measure landmark localization accuracy and clinical biometry error, respectively. All experiments were implemented in PyTorch and trained on a single NVIDIA A100 GPU.

3.2 Quantitative Results

To evaluate the effectiveness of the proposed framework, we first compare it with four representative landmark detection methods, including the official Baseline, U-Net, YOLO11-Pose, and UOD-DATR [15]. All methods are trained using the same data split and evaluated under the official metrics. As shown in Table 1, our method achieves the best overall performance, with an average MRE and MAE of 34.35 and 21.76, representing improvements of 9.3% and 7.5% over the official baseline, respectively. In particular, our framework consistently performs well on anatomically challenging structures, including A4C, Femur, FUGC, HC, IVC, and PSAX, demonstrating the

Table 1. Comparison with state-of-the-art methods on the internal validation set.

Evaluation (MRE/MAE)	Baseline	UNet	YOLO	UOD-DATR	Ours
Avg	37.87/23.53	45.54/28.50	123.34/73.37	38.20/23.91	34.35/21.76
A4CE	37.78/22.56	56.63/33.68	42.85/25.26	52.70/30.74	36.32/21.52
AOP	8.84/5.61	7.52/4.78	23.08/14.21	5.12/3.25	6.47/4.11
FA	23.08/14.56	18.03/11.27	113.07/58.57	10.70/6.72	23.52/14.82
Femur	46.92/28.64	47.54/29.25	175.37/99.75	51.46/31.66	41.09/25.30
FUGC	8.39/5.45	11.27/7.27	66.27/45.37	7.72/4.93	7.52/4.93
HC	36.33/22.55	59.32/37.34	122.38/69.07	49.04/30.50	32.55/20.34
IVC	78.16/46.98	68.78/43.93	402.28/245.76	56.75/35.43	49.29/31.56
PLAX	37.90/23.24	45.40/28.45	55.34/32.75	28.54/17.74	57.71/37.10
PSAX	63.38/42.14	95.36/60.49	109.41/69.57	81.77/54.18	54.64/36.19

effectiveness of task-aware parameter-efficient adaptation in learning shared representations across diverse anatomical structures. Although our method is slightly inferior to single-task approaches on AOP, FA, and PLAX, it achieves the best overall accuracy, indicating a favorable trade-off between task-specific optimization and multi-task generalization.

To further investigate the contribution of each component, ablation experiments are conducted on the official Hidden Validation Set by removing VPT, Task-specific LoRA, and Multi-level Supervision individually. The quantitative results are summarized in Table 2. The complete model achieves the best overall performance, reducing the average MRE from 85.49 to 31.14 compared with the official baseline. Removing either VPT, LoRA, or Multi-level Supervision consistently degrades performance, while removing multiple components causes further deterioration. This confirms that task-aware adaptation and multi-level supervision provide complementary benefits for learning task-specific representations and accurate landmark localization. Although our overall MRE is higher than several top-ranked challenge entries, the comparison is not fully equivalent because different solutions may adopt task-specific optimization or ensemble strategies. In contrast, our framework targets unified and parameter-efficient adaptation with a frozen backbone and a small number of trainable parameters. Despite this constraint, it achieves competitive performance across diverse anatomical structures.

Table 2. Ablation study of different components on the official hidden validation set.

Evaluation (MRE/MAE)	Base-line	w/o VPT	w/o LoRA	w/o VPT/LoRA	w/o Multi-level Loss	Ours
Avg	85.49/41.41	41.65/39.25	46.09/44.44	46.07/38.62	40.56/37.18	31.14/33.07
A4C	141.27/55.23	25.64/19.18	31.54/23.25	34.48/28.53	27.07/22.27	24.8/22.31
AOP	19.55/13.13	17.99/10.85	20.05/10.87	19.02/9.79	18.11/11.6	17.68/11.71
FA	226.16/94.94	22.2/92.4	27.3/101.29	29.7/101.76	20.63/82.12	21.52/85.93
Femur	36.84/46.25	29.43/24.44	24.13/25.87	28.35/29.9	28.11/21.34	24.17/23.87
FUGC	16.01/13.75	16.71/15.55	12.27/14.03	14.49/16.56	11.39/9.37	15.37/18.3
HC	42.56/55.57	55.83/81.07	52.38/69.15	55.69/75.05	55.69/80.73	54.5/73.6
IVC	46.84/23.36	55.03/55.13	71.81/79.34	73.29/20.27	50.81/49.45	37.24/23.69
PLAX	123.7/33.57	113.28/34.03	114.87/35.48	114.71/36.23	112.75/33.09	44.95/13.26
PSAX	116.48/36.86	38.75/20.62	60.42/40.69	44.92/29.53	40.47/24.68	40.0/24.94

3.3 Case Study and Failure Analysis

To better understand the strengths and limitations of the proposed method, qualitative analysis was performed on the official Hidden Validation Set. As domain metadata were unavailable, we focused on representative anatomical failure patterns. The largest performance gaps compared with the top-ranked solution were observed on HC, PLAX,

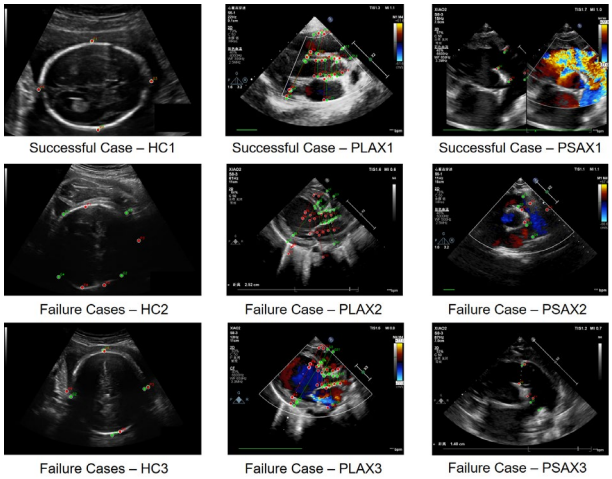

Fig. 2. Case study and anatomical analysis of representative cases on HC, PLAX, and PSAX. Green markers denote ground truth, and red markers denote predictions.

and PSAX. One successful case and two representative failure cases were selected for each structure (Fig. 2).

For HC, acoustic shadows, attenuation, and image noise often obscure the elliptical contour, making the landmark positions difficult to determine. In addition, all four landmarks lie on the ellipse boundary and have no fixed semantic ordering. This permutation ambiguity increases the difficulty of learning stable landmark correspondences. For PLAX, the predicted measurement segments generally have reasonable lengths, but their orientations are occasionally incorrect, resulting in global landmark displacement. This suggests that the model captures the overall anatomical scale but is less effective at modeling long-range spatial dependencies across multiple cardiac regions. In contrast, the main challenge of PSAX arises from imaging modality variation. When color Doppler ultrasound images are encountered, blood-flow signals partially obscure grayscale anatomical structures, leading to consistent localization errors and reduced cross-modality robustness.

4 Discussion and Conclusion

We presented a task-aware vision foundation model for universal ultrasound biometry. Built upon a frozen DINOv2 encoder, the proposed framework combines parameter-efficient task adaptation with multi-level supervision to achieve unified landmark detection across multiple anatomical structures. Experimental results demonstrate improved localization accuracy and cross-center generalization with minimal trainable parameters, highlighting the potential of vision foundation models for unified multi-organ ultrasound analysis. Remaining challenges include ambiguous anatomical boundaries, complex spatial relationships, and cross-modality appearance variations. Future work will explore medical world models that integrate anatomical priors,

multimodal foundation models, and temporal ultrasound sequences for more robust and generalizable ultrasound understanding.

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

References

1. Han, T., Ning, G., Liang, H., Li, Z., Jiang, Z., Chen, F., Kang, Y., Luo, J., Liao, H.: Toward clinical applications of intelligent robotic ultrasound systems. IEEE Reviews in Biomedical Engineering. 1–21 (2025). https://doi.org/10.1109/RBME.2025.3610605.
2. 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, 715–723 (2019). https://doi.org/10.1002/uog.20272.
3. Noble, J.A., Boukerroui, D.: Ultrasound image segmentation: a survey. IEEE Transactions on Medical Imaging. 25, 987–1010 (2006). https://doi.org/10.1109/TMI.2006.877092.
4. 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, 2204–2215 (2017). https://doi.org/10.1109/TMI.2017.2712367.
5. 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.
6. Payer, C., Štern, 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.
7. Li, Y., Zhang, S., Wang, Z., Yang, S., Yang, W., Xia, S.-T., Zhou, E.: TokenPose: Learning Keypoint Tokens for Human Pose Estimation. In: 2021 IEEE/CVF International Conference on Computer Vision (ICCV). pp. 11293–11302 (2021). https://doi.org/10.1109/ICCV48922.2021.01112.
8. Newell, A., Yang, K., Deng, J.: Stacked Hourglass Networks for Human Pose Estimation. In: Leibe, B., Matas, J., Sebe, N., and Welling, M. (eds.) Computer Vision – ECCV 2016. pp. 483–499. Springer International Publishing, Cham (2016). https://doi.org/10.1007/978-3-319-46484-8_29.
9. Ronneberger, O., Fischer, P., Brox, T.: U-Net: Convolutional Networks for Biomedical Image Segmentation. In: Navab, N., Hornegger, J., Wells, W.M., and Frangi, A.F. (eds.) Medical Image Computing and Computer-Assisted Intervention – MICCAI 2015. pp. 234–241. Springer International Publishing, Cham (2015). https://doi.org/10.1007/978-3-319-24574-4_28.

10. 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. Transactions on Machine Learning Research Journal. (2024). https://doi.org/10.48550/arxiv.2304.07193.
11. Ravi, N., Gabeur, V., Hu, Y.-T., Hu, R., Ryali, C., Ma, T., Khedr, H., Rädle, R., Rolland, C., Gustafson, L., Mintun, E., Pan, J., Alwala, K.V., Carion, N., Wu, C.-Y., Girshick, R., Dollar, P., Feichtenhofer, C.: SAM 2: Segment Anything in Images and Videos. International Conference on Learning Representations. 2025, 28085–28128 (2025).
12. Zhang, S., Metaxas, D.: On the challenges and perspectives of foundation models for medical image analysis. Medical Image Analysis. 91, 102996 (2024). https://doi.org/10.1016/j.media.2023.102996.
13. Jia, M., Tang, L., Chen, B.-C., Cardie, C., Belongie, S., Hariharan, B., Lim, S.-N.: Visual Prompt Tuning. In: Avidan, S., Brostow, G., Cissé, M., Farinella, G.M., and Hassner, T. (eds.) Computer Vision – ECCV 2022. pp. 709–727. Springer Nature Switzerland, Cham (2022). https://doi.org/10.1007/978-3-031-19827-4_41.
14. Hedlin, E., Sharma, G., Mahajan, S., He, X., Isack, H., Kar, A., Rhodin, H., Tagliasacchi, A., Yi, K.M.: Unsupervised Keypoints from Pretrained Diffusion Models. Presented at the Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (2024).
15. Zhu, H., Quan, Q., Yao, Q., Liu, Z., Zhou, S.K.: UOD: Universal One-Shot Detection of Anatomical Landmarks. In: Greenspan, H., Madabhushi, A., Mousavi, P., Salcudean, S., Duncan, J., Syeda-Mahmood, T., and Taylor, R. (eds.) Medical Image Computing and Computer Assisted Intervention – MICCAI 2023. pp. 24–34. Springer Nature Switzerland, Cham (2023). https://doi.org/10.1007/978-3-031-43907-0_3.