

Prototype Memory-Guided Training-Free Anomaly Classification and Localization in Prenatal Ultrasound

Huanwen Liang^{1,2*}, Yuhao Huang^{1,5*}, Xiliang Zhu⁴, Yuanji Zhang⁶, Xuedong Deng⁷, Xinru Gao⁸, Guowei Tao⁹, Yuhao Zhang^{1,2}✉, and Dong Ni^{1,3,4}

¹ Medical Ultrasound Image Computing (MUSIC) Lab, Shenzhen University, Shenzhen, China

² School of Biomedical Engineering, Medical School, Shenzhen University, Shenzhen, China

³ School of Artificial Intelligence, Shenzhen University, Shenzhen, China

⁴ School of Biomedical Engineering and Informatics, Nanjing Medical University, Nanjing, China

⁵ Centre for Artificial Intelligence and Robotics, Hong Kong Institute of Science & Innovation, Chinese Academy of Sciences, Hong Kong, China

⁶ Shenzhen Luohu People's Hospital (The Third Affiliated Hospital of Shenzhen University), Shenzhen, China.

⁷ Center for Medical Ultrasound, The Affiliated Suzhou Hospital of Nanjing Medical University, Suzhou Municipal Hospital, Nanjing Medical University, Suzhou, China

⁸ Northwest Women's and Children's Hospital, Xian, China

⁹ Qilu Hospital of Shandong University, Jinan, China

Abstract. Prenatal anomaly classification and localization is of critical importance for fetal health and pregnancy management. Although ultrasound (US) is the primary modality for prenatal screening, accurate diagnosis remains challenging due to the low prevalence and high heterogeneity of anomalies. Existing deep learning methods for prenatal tasks rely on large-scale annotated datasets, which are difficult to obtain in practice. Although few-shot learning alleviates data scarcity, it typically requires fine-tuning for new categories, limiting its practicality in resource-limited clinical settings. To address these challenges, we propose a training-free framework for multi-class prenatal US anomaly classification and localization that operates with only a few reference images per class, representing the first exploration of this setting. Our framework comprises three key components: (1) a memory bank with multi-granular prototypes that explicitly models both class-level semantics and anomaly characteristics; (2) a prototype-driven soft merging mechanism that aggregates discriminative features to detect the anomaly region; and (3) a class-aware refinement strategy that leverages prototype consistency to improve category prediction. Extensively validated on a multi-center prenatal US dataset containing 1,149 cases, with a total of 2,357 images and 9 categories, our proposed method outperforms the competitors.

Keywords: Anomaly Localization · Training-free · Prenatal Ultrasound

* Huanwen Liang and Yuhao Huang contributed equally.

1 Introduction

Prenatal anomalies are functional or structural abnormalities that occur during fetal growth [7,16]. These anomalies can seriously affect fetal survival and postnatal quality of life [21]. Therefore, accurate diagnosis is critical for pregnancy management. Ultrasound (US) is widely used for prenatal anomaly screening and diagnosis because it is non-invasive, radiation-free, and enables real-time imaging [19,2]. However, the low incidence and high categorical diversity of prenatal anomalies lead to heterogeneous US appearances, making accurate identification difficult even for experienced sonographers. Consequently, prenatal anomaly screening becomes highly operator-dependent and susceptible to substantial inter-operator variability, underscoring the clinical value of automated methods for assisting screening and reducing diagnostic variability.

Learning-based method in prenatal US focuses on various tasks including plane detection [22,5], biomarker measurement [26], and disease diagnosis [6,25]. Zhang et al. [24] proposed a deep learning-based method to identify the standard fetal midsagittal plane in 3D US volumes for crown-rump length measurement. Lin et al. [9] developed a YOLOv3-based system to detect fetal brain standard planes and to classify abnormal images. Liang et al. [8] introduced medical knowledge to enhance anomaly recognition and achieved prenatal abdominal anomaly classification without relying on standard planes. These studies have shown strong performance in prenatal US tasks. However, most methods rely on large amounts of annotated data. Since prenatal anomalies have a low incidence and data access is constrained by patient privacy, constructing large-scale annotated datasets is costly. This hinders the further development of these methods.

Few-shot learning, e.g., few-shot object detection (FSOD), is considered an effective solution to alleviate data scarcity [17,13,14,12]. SMILe [14] introduces submodular mutual information as the learning objective to encourage well-separated feature clusters between base and novel classes, thereby reducing class confusion and catastrophic forgetting in FSOD. Zhu et al. [28] propose TRR-CCM, an US FSOD method that leverages circular channel Mamba and topological relationship reasoning to improve detection and localization accuracy. However, existing methods rely on large-scale annotated data for pretraining from the same modality. They also require fine-tuning on novel categories, which limits their practicality in resource-limited clinical settings.

Foundation vision models (FVMs), such as Contrastive Language-Image Pre-training (CLIP) [18] and Self-Distillation with No Labels (DINO) [3,15,20], have recently achieved strong representation learning ability through pretraining on large-scale datasets. Recent studies have explored applying these foundation models to training-free medical image segmentation [27,11,23], showing their potential in low-annotation or annotation-free settings. As shown in Fig. 1, although DINO is not trained on US data, it can still capture fetal structural features, demonstrating strong fine-grained representation ability.

In this study, we propose a training-free framework for prenatal anomaly classification and localization based on vision foundation encoder. Our framework does not require fine-tuning and does not rely on large amounts of annotated

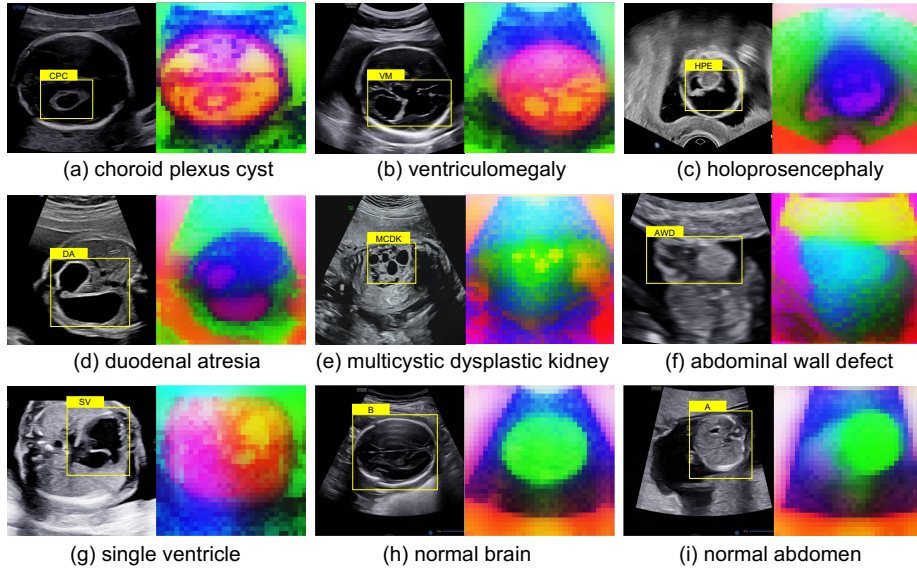


Fig. 1. Visualization of DINO features. Each image (right) displays the first three principal components of the feature representation, mapped to RGB color channels. yellow boxes indicate region of interest.

data, providing a practical solution for prenatal anomaly classification and localization. Our contributions are threefold. (1) To the best of our knowledge, this is the first study to perform multi-disease prenatal anomaly classification and localization in a training-free setting without relying on extensive annotated data. (2) We develop a unified framework that jointly incorporates memory bank construction, anomaly feature aggregation, and class-aware refinement to facilitate accurate and efficient localization of abnormal regions. (3) We validated our approach on a multi-center prenatal anomaly US dataset covering 9 categories, with 1,149 cases and 2,357 images. Experimental results demonstrated the effectiveness of our method.

2 Methodology

Fig. 2 illustrates the overall framework of the proposed training-free method, which relies on only a few annotated reference images for accurate anomaly classification and localization. Our model consists of three stages. In **Stage 1**, a memory bank was constructed to store class-wise and anomaly-related prototype representations. Then, **Stage 2** introduced an anomaly feature aggregation strategy that selects candidate anomaly-related prototypes based on class-wise feature similarity and computes anomaly scores through anomaly-feature matching with anomaly-aware soft merging. Last, a class-aware refinement strategy was equipped to improve the anomaly category discrimination in **Stage 3**.

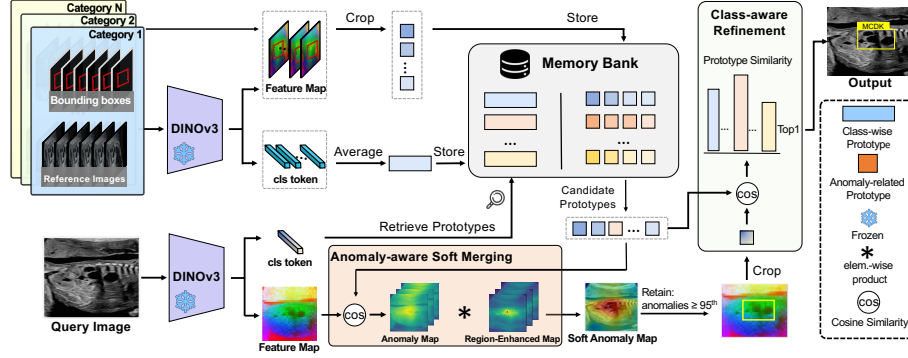


Fig. 2. Overview of our proposed framework.

2.1 Multi-Granular Prototype Memory Bank Construction

A good vision foundation encoder E plays a vital role in extracting features for memory bank construction. Most recently, DINOv3 has shown powerful performance to provide visual features for multiple medical imaging tasks [10]. Inspired by this study, we employ DINOv3 [20] as the encoder E in our work.

Given a set of reference images $\{I_i\}_{i=1}^N$ with anomaly bounding box annotations $\{B_i = \{(x_1, y_1), (x_2, y_2)\}\}_{i=1}^N$, where $N = C \times K$, C denotes the number of categories, K denotes the number of reference images per category, (x_1, y_1) and (x_2, y_2) denote the top-left and bottom-right coordinates of the bounding box, respectively. DINOv3 first partitions I_i into patches and processes them through L transformer layers, producing patch embeddings $z^l, l \in L$ and a corresponding $[\text{CLS}]^l$ token at each layer.

$$E = \{\text{layer}^1, \text{layer}^2, \dots, \text{layer}^L\}, \quad (1)$$

$$\{[\text{CLS}]^l, z^l\} = \text{layer}^l(z^{l-1}), z^0 = I_i. \quad (2)$$

Class-wise Prototype Construction for Coarse-grained Recognition.

For each reference image I_i , we pass it through the encoder E and extract the $[\text{CLS}]^L$ token from the last layer. After ℓ_2 normalization, the token is stored as a global class feature $g_i \in \mathbb{R}^D$, where D denotes the feature dimension. Then, for each category $c \in C$, class features from reference images belonging to the same category are averaged to form a class-wise prototype:

$$\mathbf{p}_c = \frac{1}{K} \sum_{i \in \mathcal{I}_c} g_i, \quad (3)$$

where $\mathcal{I}_c$ denotes the index set of category c . These class-wise prototypes are stored in the memory bank as class-wise representations.

Anomaly-Related Prototype Construction for Fine-grained Identification. For each reference image I_i with anomaly bounding box annotation

B_i , we pass I_i through the encoder E and extract ℓ_2 -normalized patch embeddings z^L from the last layer as dense feature map $\mathbf{F}_i \in \mathbb{R}^{d \times h \times w}$. Due to the low resolution of the original feature maps, a high-resolution upsampling method, JAFAR [4], is adopted to enhance spatial details to obtain high-resolution feature map $\mathbf{F}_{hr_i}$. For each image, we aggregate features within the annotated bounding box to form an anomaly-related prototype:

$$\mathbf{q}_i = \frac{1}{|B_i|} \sum_{(x,y) \in B_i} \mathbf{F}_{hr_i}(x,y). \quad (4)$$

These anomaly-related prototypes capture localized anomaly features and complement the class-wise prototypes stored in the memory bank, thereby forming a multi-granular prototype representation and improving anomaly localization.

2.2 Prototype-Guided Anomaly Feature Aggregation

During inference, given a query image I_q , its class feature is extracted using the encoder E and matched with the class-wise prototypes to select top- M candidate categories $\mathbf{p}_c$. The corresponding anomaly-related prototypes are then retrieved from the memory bank for anomaly aggregation.

Anomaly-Feature Matching. First, we retrieve candidate anomaly-related prototypes $\{\mathbf{q}_{c,k}\}$. Then, we extract dense feature map $\mathbf{F}_q$ from the query image and upsample to obtain high-resolution feature map $\mathbf{F}_{hr}$. We compute the cosine similarity between each candidate prototype and $\mathbf{F}_{hr}$ (see Equ. 5), resulting in one anomaly map per retrieved prototype.

$$\mathbf{A}_{c,k} = \cos(\mathbf{q}_{c,k}, \mathbf{F}_{hr}). \quad (5)$$

Anomaly-Aware Soft Merging. In US images, distinct anatomical structures can exhibit visually similar feature patterns, resulting in feature-level ambiguity that introduces noise and leads to false detections when features are matched independently. Meanwhile, we noticed that a local patch feature is often highly similar to its neighboring patches, which can be leveraged to strengthen region-level attention. Hence, for each anomaly map, our anomaly-aware soft merging strategy first selects the top- S high-similarity locations and computes their centroid $\mathbf{f}_{c,k}$. Then, the cosine similarity between the centroid feature $\mathbf{f}_{c,k}$ and $\mathbf{F}_{hr}$ was evaluated to obtain the region-enhanced map:

$$\mathbf{R}_{c,k} = \cos(\mathbf{f}_{c,k}, \mathbf{F}_{hr}), \quad (6)$$

which can be used to softly merge the K anomaly maps:

$$\mathbf{A}_c = \frac{1}{K} \sum_{k=1}^K \mathbf{R}_{c,k} \mathbf{A}_{c,k} \quad (7)$$

Anomaly Localization. Patches with anomaly values above the 95th percentile threshold in $\mathbf{A}_c$ are retained to generate a binary mask. Then, the minimum enclosing rectangle is computed as the final anomaly regions Ω .

Table 1. Quantitative comparison under the 32-shot setting across Splits 1-5 in terms of mAP@0.5 ($p < 0.05$). **Bold** indicates the best performance. All results are reported as (mean(std)) (%). TF: Training-free.

		CPC	VM	HPE	SV	DA	MCDK	AWD	B	A	mAP
	DeFRCN	52.60 (1.78)	39.23 (2.94)	55.63 (1.29)	50.18 (6.19)	52.97 (4.91)	35.18 (9.38)	42.34 (2.51)	36.16 (2.59)	54.90 (2.15)	46.58 (2.44)
FSOD	DiGeo	48.81 (4.75)	38.77 (7.46)	40.00 (4.42)	69.03 (10.21)	54.47 (6.73)	34.09 (10.39)	28.50 (6.73)	38.52 (3.17)	57.12 (0.82)	45.48 (4.46)
	SMILe-FSOD	50.5 (1.29)	44.11 (2.13)	41.40 (1.19)	62.89 (7.21)	53.91 (5.62)	39.22 (7.03)	32.18 (4.23)	39.45 (1.94)	54.64 (2.36)	46.48 (2.58)
	TRR-CCM	51.74 (4.20)	36.79 (4.56)	53.12 (4.49)	71.13 (4.89)	53.8 (4.42)	38.36 (6.70)	40.75 (1.20)	43.18 (2.62)	54.52 (3.56)	49.26 (2.79)
	ProtoSAM	22.57 (3.39)	38.13 (3.33)	26.66 (3.71)	28.85 (5.97)	55.96 (6.44)	55.8 (6.27)	10.23 (7.78)	64.65 (7.67)	65.71 (5.62)	40.95 (2.86)
TF	MAUP	25.46 (4.32)	37.61 (7.01)	15.89 (7.21)	38.17 (8.86)	35.02 (4.22)	39.72 (8.58)	17.01 (5.43)	66.93 (6.95)	55.22 (3.47)	36.78 (2.34)
	Ours	50.66 (6.35)	71.00 (3.56)	58.19 (5.12)	71.29 (5.59)	76.77 (1.44)	64.67 (2.16)	36.17 (2.43)	46.39 (2.70)	70.88 (4.91)	60.67 (1.82)

2.3 Class-Aware Refinement with Enhanced Memory Alignment

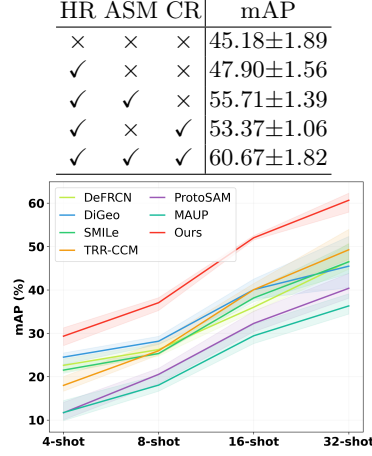
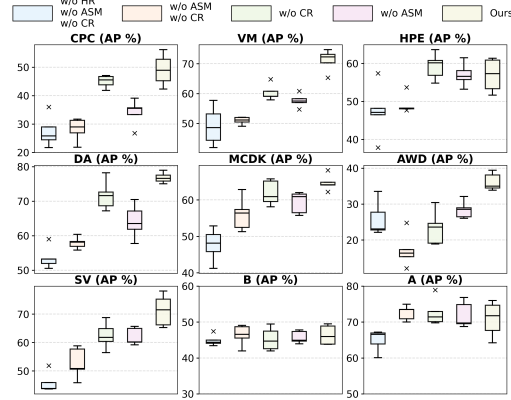
Based on the anomaly regions obtained in the previous step, this stage further refines their categories via a class-aware refinement strategy with the constructed memory bank. For each predicted anomaly region, we extract its region features from $\mathbf{F}_{hr}$ and compute an average feature embedding:

$$\mathbf{e}_i = \frac{1}{|\Omega_i|} \sum_{(x,y) \in \Omega_i} \mathbf{F}_{hr}(x,y), \quad (8)$$

where Ω_i denotes the bounding box of the i -th anomaly region. We then compute the cosine similarity between $\mathbf{e}_i$ and the retrieved anomaly-related prototypes, and select the class with the highest similarity as the predicted label, with the similarity value serving as the confidence score.

3 Experiment Results

Materials and Implementation Details. With local Institutional Review Board approval (*No.* K-2023-067-H01), we collected a multi-center prenatal US dataset comprising 1,149 cases and 2,357 images, encompassing 9 categories spanning 3 anatomical regions (brain, heart, and abdomen). All images were annotated by sonographers with classification labels and bounding boxes: the anomaly class is annotated with the anomaly region, and the normal class is annotated with the region of interest. Specifically, the fetal brain images contain four types, including choroid plexus cyst (CPC, 122 cases / 226 images), ventriculomegaly (VM, 164 cases / 312 images), holoprosencephaly (HPE, 104 cases

Table 2. Ablation study results.**Fig. 3.** Performance visualization.**Fig. 4.** Ablation study on different categories.

/ 238 images), and normal brain (B, 183 cases / 329 images). The heart-related category contain single ventricle (SV, 29 cases / 197 images). The abdomen-related categories contain duodenal atresia (DA, 178 cases / 255 images), multicystic dysplastic kidney (MCDK, 77 cases / 261 images), abdominal wall defect (AWD, 51 cases / 181 images), and normal abdomen (A, 241 cases / 358 images). The dataset was split at the case level by sampling cases without replacement. Once the accumulated number of images exceeded 32, 32 images were randomly selected to form a fold, and the corresponding cases were removed from the pool. The process was repeated to construct five reference folds (Splits 1-5), each containing 32 images per category (1,440 images in total), while the remaining cases were used for testing (917 images in total). Within each reference set, we randomly selected 4, 8, 16, or all 32 images per category to construct the 4-shot, 8-shot, 16-shot, and 32-shot settings, respectively.

Our approach was implemented in PyTorch using an NVIDIA RTX 4090 GPU (24GB). Images were resized to 512×512 . We employed a pretrained DINOv3-Small backbone, producing 64×64 feature map. The JAFAR upsampling method was applied to upsample features to 256×256 without additional pre-training. For hyperparameters, we set $M = 2$ and $S = 16$. Performance was evaluated using mAP@0.5 with TorchMetrics. Codes can be found in GitHub.

Quantitative Analysis. We compare our approach with official implementations of FSOD methods, including DeFRCN [17], DiGeo [13], SMILe-FSOD [14], and TRR-CCM [28], which are fine-tuned on our reference sets, as well as training-free methods, including ProtoSAM [1] and MAUP [27]. Table 1 shows the results. Our method outperforms the comparative methods across most categories. Compared with the second-best method, our approach achieves an 11.41% higher mAP. Fig. 3 shows the mAP performance with error estimation

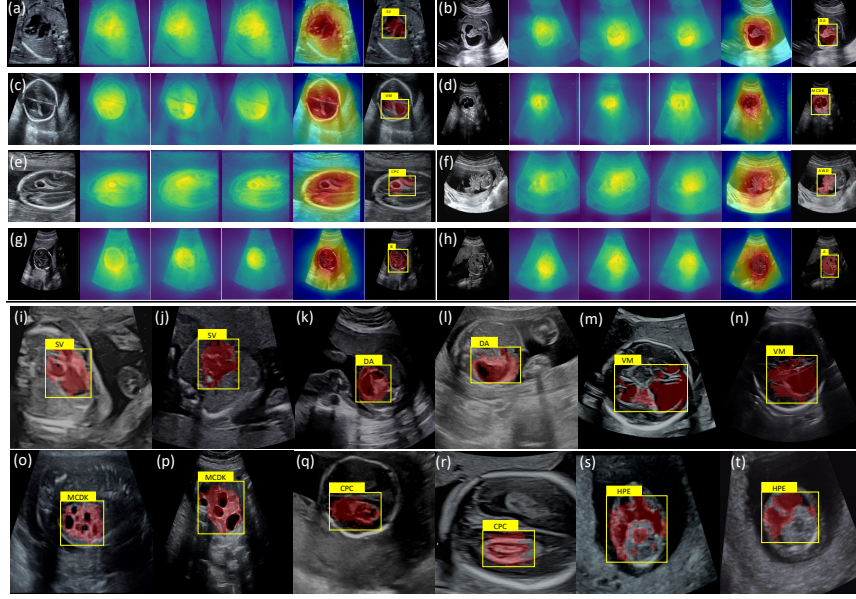


Fig. 5. Visualization of typical cases. (a)-(h) show representative examples. From left to right: the query image, similarity maps of three prototypes, the anomaly map, and the predicted anomaly region with category label. (i)-(t) shows additional cases.

across splits. As the number of reference images increases, our method exhibits steady performance gains, demonstrating strong scalability and outperforming all compared methods overall. Table 2 shows the ablation results. High-resolution (HR) upsampling improves mAP by 2.72%. Anomaly-aware soft merging (ASM) improves mAP by 7.81%. Class-aware refinement (CR) improves mAP by 5.47%. Using all components improves mAP by 15.49%. Fig. 4 shows the ablation results on different categories. It can be observed that each proposed strategy provides systematic performance improvements across all categories.

Qualitative Analysis. We visualize several typical cases in Fig. 5. For each query image (a)-(h), we show three candidate prototypes. We observe that, after aggregating the similarity maps from these candidate prototypes, the resulting anomaly map is well concentrated on the fetal region. The mask and bounding box obtained using the 95th-percentile threshold accurately localize the anomaly region. These results indicate that, even without training and using only a few reference images, our approach can accurately localize anomaly regions and correctly classify their categories, highlighting its effectiveness.

4 Conclusion

In this paper, we proposed a novel training-free framework for multi-class prenatal anomaly classification and localization. Our method requires only a few

reference images, addressing the scarcity of abnormal data and eliminating the need for retraining. The proposed framework consists of three stages: (1) memory bank construction to store class-level and anomaly-related representations from reference images. (2) anomaly feature aggregation to localize abnormal regions via anomaly-aware soft merging. (3) class-aware refinement to improve anomaly category discrimination. We evaluated our method on a multi-center dataset using only a few reference images per class. The results demonstrated the effectiveness of our approach. In future work, we will extend our approach to additional anomaly types and perform comprehensive clinical validation.

Acknowledgments. This work is supported by the Frontier Technology Development Program of Jiangsu Province (No. BF2024078), National Natural Science Foundation of China (Nos. 12326619, 62572324), Science and Technology Planning Project of Guangdong Province (No. 2023A0505020002).

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

References

1. Ayzenberg, L., Giryas, R., Greenspan, H.: Protosam: One-shot medical image segmentation with foundational models. arXiv preprint arXiv:2407.07042 (2024)
2. Bilardo, C., Chaoui, R., Hyett, J., Kagan, K., Karim, J., Papageorgiou, A., Poon, L., Salomon, L., et al.: Isuog practice guidelines (updated): performance of 11–14-week ultrasound scan. *Ultrasound in Obstetrics and Gynecology* **61**(1) (2023)
3. Caron, M., Touvron, H., Misra, I., Jégou, H., Mairal, J., Bojanowski, P., Joulin, A.: Emerging properties in self-supervised vision transformers. In: *Proceedings of the IEEE/CVF international conference on computer vision*. pp. 9650–9660 (2021)
4. Couairon, P., Chambon, L., Serrano, L., Haugeard, J.E., Cord, M., Thome, N.: Jafar: Jack up any feature at any resolution. arXiv preprint arXiv:2506.11136 (2025)
5. Dou, H., Huang, Y., Huang, Y., Yang, X., Zhen, C., Zhang, Y., Xiong, Y., Huang, W., Ni, D.: Standard plane localization using denoising diffusion model with multi-scale guidance. *Computer Methods and Programs in Biomedicine* p. 108619 (2025)
6. Huang, Y., Xu, Y., Dou, H., Deng, J., Yang, X., Zheng, H., Ni, D.: Uncertainty-aware diffusion and reinforcement learning for joint plane localization and anomaly diagnosis in 3d ultrasound. In: *International Conference on Medical Image Computing and Computer-Assisted Intervention*. pp. 650–660. Springer (2025)
7. Laurichesse Delmas, H., Kohler, M., Doray, B., Lémery, D., Francannet, C., Quistebert, J., Marie, C., Perthuis, I.: Congenital unilateral renal agenesis: prevalence, prenatal diagnosis, associated anomalies. data from two birth-defect registries. *Birth defects research* **109**(15), 1204–1211 (2017)
8. Liang, H., Xu, J., Zhang, Y., Huang, Y., Zhang, Y., Yang, X., Li, R., Deng, X., Liu, Y., Tao, G., et al.: Medical-knowledge driven multiple instance learning for classifying severe abdominal anomalies on prenatal ultrasound. In: *International Conference on Medical Image Computing and Computer-Assisted Intervention*. pp. 344–354. Springer (2025)
9. Lin, M., He, X., Guo, H., He, M., Zhang, L., Xian, J., Lei, T., Xu, Q., Zheng, J., Feng, J., et al.: Use of real-time artificial intelligence in detection of abnormal image

- patterns in standard sonographic reference planes in screening for fetal intracranial malformations. *Ultrasound in Obstetrics & Gynecology* **59**(3), 304–316 (2022)
10. Liu, C., Chen, Y., Shi, H., Lu, J., Jian, B., Pan, J., Cai, L., Wang, J., et al.: Does dinov3 set a new medical vision standard? benchmarking 2d and 3d classification, segmentation, and registration. *arXiv preprint arXiv:2509.06467* (2025)
 11. Liu, Y., Xiao, H., Chai, J., Zhang, Y., Wang, R., et al.: Synpo: Boosting training-free few-shot medical segmentation via high-quality negative prompts. In: *International Conference on Medical Image Computing and Computer-Assisted Intervention*. pp. 594–603. Springer (2025)
 12. Lu, X., Diao, W., Mao, Y., Li, J., et al.: Breaking immutable: Information-coupled prototype elaboration for few-shot object detection. In: *Proceedings of the AAAI conference on artificial intelligence*. vol. 37, pp. 1844–1852 (2023)
 13. Ma, J., Niu, Y., Xu, J., et al.: Digeo: Discriminative geometry-aware learning for generalized few-shot object detection. In: *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*. pp. 3208–3218 (2023)
 14. Majee, A., Sharp, R., Iyer, R.: Smile: Leveraging submodular mutual information for robust few-shot object detection. In: *European Conference on Computer Vision*. pp. 350–366. Springer (2024)
 15. Oquab, M., Darcet, T., Moutakanni, T., Vo, H., Szafraniec, M., Khalidov, V., Fernandez, P., Haziza, D., Massa, F., El-Nouby, A., et al.: Dinov2: Learning robust visual features without supervision. *arXiv preprint arXiv:2304.07193* (2023)
 16. Pechriggl, E., Blumer, M., Tubbs, R.S., Olewnik, L., Konschake, M., Fortélny, R., Stofferin, H., et al.: Embryology of the abdominal wall and associated malformations—a review. *Frontiers in Surgery* **9**, 891896 (2022)
 17. Qiao, L., Zhao, Y., Li, Z., Qiu, X., Wu, J., Zhang, C.: Defrcn: Decoupled faster r-cnn for few-shot object detection. In: *Proceedings of the IEEE/CVF international conference on computer vision*. pp. 8681–8690 (2021)
 18. Radford, A., Kim, J.W., Hallacy, C., Ramesh, A., Goh, G., Agarwal, S., Sastry, G., Askell, A., Mishkin, P., Clark, J., et al.: Learning transferable visual models from natural language supervision. In: *International conference on machine learning*. pp. 8748–8763. PmLR (2021)
 19. Salomon, L., Alfirevic, Z., Berghella, V., Bilardo, C., Chalouhi, G., Costa, F.D.S., Hernandez-Andrade, E., Malinger, G., Munoz, H., Paladini, D., et al.: Isuog practice guidelines (updated): performance of the routine mid-trimester fetal ultrasound scan. *Ultrasound in Obstetrics and Gynecology* **59**(6), 840–856 (2022)
 20. Siméoni, O., Vo, H.V., Seitzer, M., Baldassarre, F., Oquab, M., Jose, C., Khalidov, V., Szafraniec, M., Yi, S., Ramamonjisoa, M., et al.: Dinov3. *arXiv preprint arXiv:2508.10104* (2025)
 21. Xie, X., Pei, J., et al.: Global birth prevalence of major congenital anomalies: a systematic review and meta-analysis. *BMC Public Health* **25**(1), 449 (2025)
 22. Yang, X., Huang, Y., Huang, R., Dou, H., Li, R., Qian, J., Huang, X., Shi, W., Chen, C., Zhang, Y., et al.: Searching collaborative agents for multi-plane localization in 3d ultrasound. *Medical Image Analysis* **72**, 102119 (2021)
 23. Yang, Y., Su, G., Hu, J., Sammarco, F., Geiping, J., Wolfers, T.: Medsamix: A training-free model merging approach for medical image segmentation. *arXiv preprint arXiv:2508.11032* (2025)
 24. Zhang, Y., Huang, Y., Chen, C., Hu, X., Pan, W., Luo, H., Huang, Y., Wang, H., Cao, Y., Yi, Y., et al.: Comparative study of 2d vs. 3d ai-enhanced ultrasound for fetal crown-rump length evaluation in the first trimester. *BMC Pregnancy and Childbirth* **25**(1), 766 (2025)

25. Zhang, Y., Huang, Y., Dou, H., Zhu, X., Ling, C., Yang, Z., Liang, L., Li, J., Liang, S., Li, R., et al.: Artificial intelligence for detecting fetal orofacial clefts and advancing medical education. *Nature Communications* (2026)
26. Zhang, Y., Yang, X., Ji, C., Hu, X., Cao, Y., Chen, C., Sui, H., Li, B., Zhen, C., Huang, W., et al.: Deep learning model for real-time nuchal translucency assessment at prenatal us. *Radiology: Artificial Intelligence* **7**(4), e240498 (2025)
27. Zhu, Y., Zhang, H.: Maup: Training-free multi-center adaptive uncertainty-aware prompting for cross-domain few-shot medical image segmentation. In: *International Conference on Medical Image Computing and Computer-Assisted Intervention*. pp. 326–336. Springer (2025)
28. Zhu, Y., Liang, B., Li, N., Zhao, L., Li, X., Li, H., Yang, F., Pu, B.: Anatomical structure few-shot detection utilizing enhanced human anatomy knowledge in ultrasound images. In: *International Conference on Medical Image Computing and Computer-Assisted Intervention*. pp. 35–45. Springer (2025)