This MICCAI paper is the Open Access version, provided by the MICCAI Society. It is identical to the accepted version, except for the format and this watermark; the final published version is available on SpringerLink.

Anatomical Structure Few-Shot Detection Utilizing Enhanced Human Anatomy Knowledge in Ultrasound Images

Ying Zhu[1], Bocheng Liang[2], Ningshu Li[3], Lei Zhao[4], Xi Li[1], Hao Li[1(✉)], Fengwei Yang[5], and Bin Pu[3(✉)]

[1] School of Information Science and Engineering, Yunnan University, China
lihao707@ynu.edu.cn
[2] Shenzhen Maternity and Child Healthcare Hospital, Women and Children's Medical Center, Southern Medical University, China
[3] Electronic and Computer Engineering, The Hong Kong University of Science and Technology, China
eebinpu@ust.hk
[4] College of Computer Science and Electronic Engineering, Hunan University, China
[5] Department of Mathematics, University of British Columbia, Canada

Abstract. Deep learning-based models have significantly advanced clinical ultrasound tasks by detecting anatomical structures within vast ultrasound image datasets. However, their remarkable performance inherently requires extensive training of annotated medical datasets. Few-shot learning addresses the challenge of limited labeled data for model training. Currently, few-shot learning in the field of medical image analysis mainly focuses on classification and semantic segmentation, with relatively fewer studies on object detection. In this paper, we propose a novel few-shot anatomical structure detection method in ultrasound images called TRR-CCM, which consists of Circular Channel Mamba (CCM) and Topological Relationship Reasoning (TRR) based on human anatomy knowledge. CCM, as a new Mamba variant, performs contextual modeling of anatomical structures and captures long- and short-term dependencies. TRR learns spatial topological relationships between human anatomical structures to further improve the accuracy of detection and localization. Experimental results on two fetal ultrasound datasets demonstrate that TRR-CCM outperforms 9 state-of-the-art baseline methods.

Keywords: Few-shot anatomical structure detection · Circular channel Mamba · Topological relationship Reasoning.

1 Introduction

Deep learning (DL)-based anatomical structure detection methods have been very successful on several ultrasound tasks such as standard view localization

Ying Zhu and Bocheng Liang contributed equally to this work.
The code is available at https://github.com/yuyizhilian/TRR-CCM.

[4,14], quality control [8], and diagnosis of structural screening [16,18]. However, the performance of DL-based models relies on large amounts of annotated data. In some cases, the acquisition of large datasets is restricted on account of ethical and privacy regulations. Meanwhile, labeling massive data requires specialized medical experts, entailing huge labor and effort. As a viable solution to address data scarcity [12], few-shot learning has shown great potential in the field of medical imaging and has achieved significant success in tasks such as classification [6,9,26] and semantic segmentation [20,22]. However, previous studies have ignored the multi-medical object detection few-shot learning in ultrasound images.

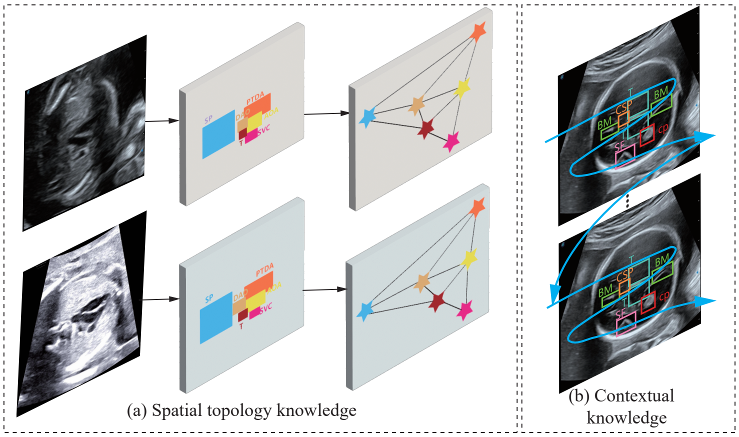


Fig. 1. Consistency of two fixed knowledge from human anatomy. (a) Spatial topology knowledge. (b) Contextual knowledge.

Ultrasound image analysis faces significant challenges due to domain shifts caused by heterogeneous device parameters (e.g., frequency, gain), operator-dependent acquisition techniques (probe angle/pressure), and intrinsic noise artifacts (low contrast, speckle noise) [24], which degrade cross-domain generalization of few-shot learning. Despite the deployment of numerous methods [15,17,27] for detecting ultrasound structures, the challenges continue to exist. Nevertheless, the ***invariant spatial-topological relationships*** (Fig. 1(a)) and ***anatomical context relationships*** (Fig. 1(b))derived from human physiology establish a unified prior knowledge framework to compensate for such variations and improve model robustness. *(1)* For extracting spatial topological knowledge, we employ graph reasoning to obtain the spatial topological relationships of the anatomical structures. Due to the intrinsic properties of human anatomy, there is significant consistency in the topological representations derived from anatomical structures in ultrasound images, as shown in Fig. 1(a). The highly consistent topology graph can help the model better understand the image from a global perspective, providing guidance and constraints for the detection of anatomical structures. Especially for few-shot object detection, the spatial topology graph can help the model predict novel classes. *(2)* We utilize a Mamba-based approach [3] to extract contextual semantic information (i.e., long- and short-term contextual dependencies) about anatomical structures. As shown in Fig. 1(b), the selective scanning mechanism of Mamba focuses on extracting core semantics from long sequences, enhances global context understanding, and boosts semantic modeling efficiency and accuracy. Furthermore, Mamba reduces sequence modeling complexity to linear via selective state-space models, dynamically filtering input information to cut parameters while preserving high efficiency. However, the previous Mamba-based methods focus on capturing spatial correlations and often neglected channel information [1]. The lack of channel mixing in Mamba architecture causes stability issues in larger networks [13] and limits its ability to model global information.

Based on the above analysis, we propose a new few-shot object detection method (TRR-CCM) in ultrasound images that integrates Circular Channel Mamba (CCM) and Topological Relationship Reasoning (TRR) for anatomical structure detection. CCM effectively captures long-range dependencies in the image with linear computational complexity, ensuring an understanding of the global context while retaining channel-specific features. In TRR, we propose to deploy a spatially-aware graph convolutional network to learn the spatial topological relationships between anatomical structures efficiently and adaptively. In summary, our contributions can be summarized as follows:

1. We design Circular Channel Mamba to capture both long-range and short-range dependencies of multiple anatomical structures while retaining crucial channel information.
2. We propose a Topological Relationship Reasoning that encodes human anatomy knowledge as graph relations utilizing graph convolution learning the spatial topological relationships, thereby enhancing the detection performance and robustness of the model.
3. Extensive experiments were conducted on two fetal ultrasound image datasets, and results demonstrate that our method outperforms state-of-the-art 9 baseline methods, showing its potential for clinical application.

2 Method

Fig. 2 shows the overview of our proposed method. TRR-CCM consists of Circular Channel Mamba (CCM) and Topological Relationship Reasoning (TRR), which divides the base class training phase and fine-tuning phase into two subphases. (1) The features extracted by the backbone network are encoded with positional information and then fed into CCM to capture long-term contextual dependencies. These visual features are subsequently enhanced by attention mechanisms and subjected to a global semantic pool. (2) In TRR, an interpretable sparse adjacency matrix is first learned from the visual features, retaining only

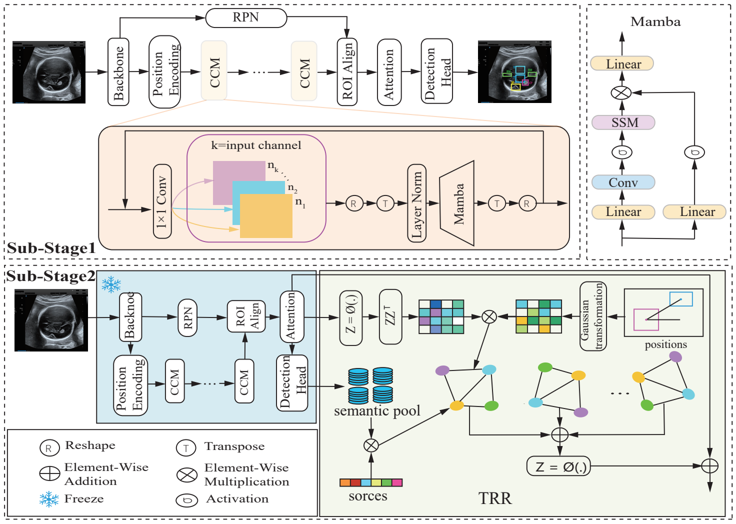


Fig. 2. The overall pipeline of the proposed TRR-CCM framework.

the most relevant connections for object recognition. Subsequently, the semantic representations from the global semantic pool are mapped to each region. A Graph Convolutional Network (GCN) is employed to learn the topological relationships between anatomical structures. (3) Finally, these fusion features are concatenated from the long- and short-term dependence level and topological relation level for further prediction.

2.1 Circular Channel Mamba

We design a new Mamba variant called Circular Channel Mamba to extract short- and long-term contextual dependencies of the anatomical structure. The former Mamba is a deep learning architecture based on Structured State Space Models (SSMs). SSMs function as a linear time-invariant system to map a one-dimensional sequence $x(t) \in \mathbb{R}^L$ to $y(t) \in \mathbb{R}^L$ through a hidden state $h(t) \in \mathbb{R}^L$, denoted as:

$$h_t = \overline{A}h_{t-1} + \overline{B}x_t, \quad y_t = Ch_t, \tag{1}$$

where $\overline{A} = exp(\varDelta A)$ and $\overline{B} = exp(\varDelta A)^{-1}(exp(\varDelta A) - I) \cdot \varDelta B$. Here, A and B are continuous-time parameters, while $\overline{A}$ and $\overline{B}$ are discrete-time parameters. Additionally, $\varDelta$ is a timescale parameter.

CCM extends the selective SSM mechanism to the channel dimension to capture inter-channel feature dependencies, and to extract long-term dependencies and channel information. Specifically, given input tensor $X \in \mathbb{R}^{B \times C \times H \times W}$, we initially employ 1×1 convolutions for cross-channel information aggregation and feature transformation. Here, B represents the batch size, and C represents the number of channels. Additionally, H and W represent the height and width of the feature map, respectively. Then for the multi-channel feature maps output from the previous layer, each channel is first separated, and convolution operations are performed on the individual channel feature maps, which are then recombined, expressed as

$$\overline{X}(c,m,n)=\sum_{i=0}^{K-1}\sum_{j=0}^{K-1}X_{c,m+i,n+j}\cdot W_{c,i,j}, \tag{2}$$

where $\overline{X}(c,m,n)$ is the value at position (m,n) in the c-th channel of the output feature map, $X_{c,m+i,n+j}$ is the value at position (m+i,n+j) in the c-th channel of the input feature map and $W_{c,i,j}$ is the weight at position (i,j) of the convolution kernel corresponding to the input channel c. Then reshape $\overline{X}$ to $X' \in \mathbb{R}^{B\times C\times T}$, where $T = H \times W$, and transpose to $X^T \in \mathbb{R}^{B\times T\times C}$. We encode the global context of X^T through:

$$\hat{X} = Mamba(LN(X^T)), \tag{3}$$

where LN stands for Layer Normalization. The final feature $\hat{X}$ is transposed and reshaped back to $\hat{X}' \in \mathbb{R}^{B\times C\times H\times W}$. Then, the output features are added to the original features and re-input into the CCM:

$$X_{out} = CCM(X + \hat{X}'). \tag{4}$$

In our experiment, we employ a series of stacked CCM blocks to enhance the short- and long-term feature extraction capability.

2.2 Topological Relationship Reasoning

Topological relationship reasoning assists the model in conducting structured reasoning based on spatial topological knowledge. First, learn sparse matrices from visual features. We model an undirected region-to-region graph G, represented as $G = \langle \mathcal{N}, \mathcal{E}\rangle$. Here, each node in $\mathcal{N}$ corresponds to a region proposal, and each edge $e_{i,j} \in \mathcal{E}$ represents the relationship between two nodes. We transform visual features f = $\{f_i\}_{i=1}^N$, $\{f_i\} \in \mathbb{R}^D$ into the latent space Z by non-linear transformation denoted by

$$z_i = \phi(f), i = 1, 2, ..., N, \tag{5}$$

where $z_i \in \mathbb{R}^L$, L is the dimension of the latent space and $\phi(.)$ is a non-linear function. Let $\mathbf{Z} \in \mathbb{R}^{N\times L}$ be the collection of normalized $\{z_i\}_{i=1}^N$, the adjacency matrix can be calculated as $\mathcal{E} = \mathbf{Z}\mathbf{Z}^T$. To alleviate the computational burden, for each node, only the K most relevant nodes are selected as neighboring nodes: $Neighbour(i) = TOP_k(\mathcal{E}_{i1}, \mathcal{E}_{i2}, ..., \mathcal{E}_{iN})$.

Then, the semantic information from the global semantic pool is assigned to every region. The global semantic pool $W \in \mathbb{R}^{C\times(D+1)}$ is obtained from the preceding classification layer. Meanwhile, the classifier and softmax function yield classification scores for the region proposals, denoted as $S \in \mathbb{R}^{N\times C}$. Here,

C denotes the number of classes, and D the feature dimension of the classifier input. Thus, the regional representations of the nodes $X \in \mathbb{R}^{N\times(D+1)}$ can be computed as a matrix multiplication: $X = SW$.

To enable the GCN to model and interact with spatial topological information, we employ a polar coordinate function $P(i,j) = (d,\theta)$ to construct the topological information. This function returns a two-dimensional vector that calculates the angle and distance between the centers of two region proposals (x_i, y_i) and (x_j, y_j), e.g., $d = \sqrt{(x_i - x_j)^2 + (y_i - y_j)^2}$ and angle $\theta = \arctan\left(\frac{y_j - y_i}{x_j - x_i}\right)$. Then pass this topological information into a Gaussian kernel to generate the relationship weights between adjacent nodes. Formally, given a graph node i, the information propagation and aggregation operation can be expressed as follows:

$$f_m^{'}(i) = \sum_{j\in Neighbour(i)} \omega_m(P(i,j))x_j e_{ij}, \tag{6}$$

$$\omega_m(P(i,j)) = \exp(-\frac{1}{2}(P(i,j) - u_m)^T \varSigma_m^{-1}(P(i,j) - u_m)), \tag{7}$$

where $Neighbour(i)$ denotes the neighborhood of node i and ω_m is the *m-th* Gaussian kernel, u_m and $\varSigma_m$ are learnable 2×1 mean vector and 2×2 covariance matrix. For each node i, $f_m^{'}(i)$ is computed as a weighted sum of the neighboring semantic representations X, where the Gaussian kernel $\omega_m(.)$ encodes the spatial information of the regions. Then $f_m^{'}(i)$ is concatenated over K kernels.

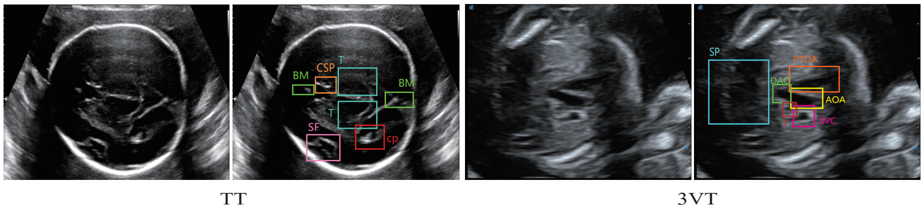
3 Experiments

3.1 Datasets

We employ two fetal ultrasound image datasets, transthalami (TT) [7] and three vessels and trachea (3VT) [18], for model evaluation. The TT dataset was sourced from Shenzhen Maternal and Child Health Hospital through various manufacturers (including Samsung, SonoScape, and Philips). The range of gestational weeks covered in these datasets spans from 18 to 32 weeks. TT contains 726 fetal brain US images with seven anatomical structure categories. 3VT contains 913 fetal heart US images with six anatomical structure categories. Moreover, both datasets were divided into training, validation, and test sets following a 7:1:2 ratio. Fig. 3 shows examples of original and labeled data in two datasets.

3.2 Implementation Details

We use DeFRCN [19] as the base network and Faster R-CNN as the base detector. ResNet101 is taken as the backbone and we use the weights pre-trained on ImageNet in initialization. We set the number of layers of the GNN to 2, the number of neighboring nodes for each node to 32, and the number of Gaussian kernels to 10. During training, our model is trained using SGD optimizer with a momentum of 0.9 and a weight decay of 0.00005. The learning rate is set to



TT

3VT

Fig. 3. Examples of two datasets.

0.01, and the batch size is set to 8. The metrics for evaluating the overall network performance are the mean average precision at IoU=0.5 (mAP@50). We take nine state-of-the-art FSOD methods as baseline methods to evaluate the effectiveness of the proposed method, including TFA [23], FSCE [21], DeFRCN [19], DCFS [2], MFDC [25], VFA [5], ICPE [10], DiGeo [11], and TKR [7].

3.3 Results

Table 1. Detection results for the TT dataset under the three settings. Bold and underlined numbers denote the 1st and 2nd scores.

Models	Split 1					Split 2					Split 3				
	1	2	3	5	10	1	2	3	5	10	1	2	3	5	10
TFA	63.4	64.4	63.9	64.2	65.3	66.3	66.0	63.7	62.9	63.8	67.2	70.2	71.5	69.3	70.6
FSCE	64.8	64.1	49.0	53.0	60.0	57.8	64.0	47.7	54.6	65.7	64.5	71.2	44.4	50.1	59.9
DeFRCN	67.4	68.6	69.7	69.6	72.8	73.9	72.5	73.9	77.1	78.6	70.8	72.0	71.4	72.9	76.6
DCFS	69.1	70.7	71.1	73.2	74.8	76.8	78.1	80.4	80.2	82.0	70.6	73.6	75.0	75.9	75.4
MFDC	67.7	69.7	73.3	77.7	80.8	76.3	76.7	76.2	78.0	83.1	72.2	74.8	73.8	75.3	81.1
VFA	53.7	58.7	64.1	51.6	51.8	64.4	48.4	65.6	49.8	41.2	51.9	67.4	52.9	52.7	52.9
ICPE	71.1	68.4	65.6	66.7	71.1	65.7	65.1	65.8	66.4	65.3	70.0	67.2	68.3	68.3	66.9
DiGeo	64.6	63.5	68.5	70.5	69.5	69.6	67.9	71.0	70.7	75.4	70.8	67.9	70.1	72.3	76.1
TKR	69.6	70.0	73.1	76.6	76.7	81.0	82.4	82.0	83.2	87.8	76.0	75.6	77.5	79.5	82.3
Ours	**71.3**	**71.8**	**75.8**	**79.8**	**81.0**	**82.8**	**84.5**	**86.3**	**86.5**	**91.0**	**79.4**	**78.3**	**77.9**	**83.3**	**86.8**

As shown in Table 1, our work significantly outperforms the existing competitive baseline methods in each few-shot setting on the TT dataset. In the 5-shot case, our method surpasses the second-best method by 2.1%, 3.3%, and 3.8% on data splits 1, 2, and 3, respectively. In the 3-shot case, our method exceeds the second-best method by 2.5% and 4.3% on data splits 1 and 2, respectively. Notably, in the 10-shot case of data split 3, our method achieves a substantial improvement of 4.5% compared to the second-best method. It can be observed that our method improves on this dataset in most cases with the number of shots increasing.

Table 2 presents the results of the 3VT, demonstrating that our method outperforms the baseline in most cases, with the best performance on data split

3. In the 5-shot case, our method is superior to the second-best method by 5.2%, 2.4% and 1.3% on data split 1, 2 and 3, respectively. In the 10-shot case, our method outperforms the second-best method by 3.4% and 3.1% on data split 1 and 3, respectively. As the number of shots increases, our method also improves performance in most cases.

Table 2. Detection results for the 3VT dataset under the three settings. Bold and underlined numbers denote the 1st and 2nd scores.

Models	Split 1					Split 2					Split 3				
	1	2	3	5	10	1	2	3	5	10	1	2	3	5	10
TFA	56.3	61.7	61.8	62.2	62.7	60.4	62.4	62.3	60.2	61.7	62.7	62.7	63.3	64.7	67.4
FSCE	43.3	54.8	33.0	35.3	42.8	53.6	54.6	38.3	35.5	42.7	47.2	53.8	37.7	36.9	42.7
DeFRCN	57.8	58.4	59.2	57.4	60.5	58.5	59.2	59.5	63.2	66.2	62.3	63.5	65.2	66.2	67.7
DCFS	58.9	58.0	61.2	61.1	63.9	61.3	62.2	64.3	68.2	68.8	62.9	64.0	65.0	66.3	68.7
MFDC	61.3	62.9	65.3	66.3	70.2	62.5	62.9	62.1	66.2	68.9	63.5	64.8	67.7	70.2	73.8
VFA	46.5	52.0	52.1	56.7	65.4	46.4	48.9	49.3	53.7	60.1	33.4	48.5	50.8	55.7	60.6
ICPE	61.1	61.7	61.8	60.3	61.3	57.8	57.7	59.6	60.7	60.8	58.3	58.6	58.4	58.0	57.3
DiGeo	61.0	60.5	62.6	66.3	67.1	61.7	60.9	62.8	61.5	62.3	61.9	63.1	63.6	63.6	65.8
TKR	63.9	66.0	68.8	65.9	72.2	**63.1**	64.0	**64.8**	70.0	73.0	64.4	64.6	70.1	72.0	74.5
Ours	**64.1**	**66.1**	**69.5**	**71.5**	**75.6**	62.5	**65.5**	64.7	**72.4**	**73.2**	**64.8**	**66.5**	**71.2**	**73.3**	**77.6**

3.4 Ablation

Table 3 shows the ablation studies of the CCM and TRR modules on data split 1 of the TT and 3VT datasets. As shown in Table 3, TRR outperformed CCM on the 3VT dataset, while CCM outperformed TRR on the TT dataset. Both components improve our model's performance, validating the effectiveness of TRR-CCM. Specifically, the addition of CCM only achieves a significant improvement over the baseline method by 1.4% to 9.4% on TT dataset, and 5.9% to 12.1% on 3VT dataset, respectively. Only by adding the TRR, it outperforms the baseline on most of the shots of TT, while the 3VT dataset sees an improvement of over 5.7% in all cases. When both CCM and TRR are included, our method boosts by 3.9%, 3.2%, 6.1%, 10.2%, and 8.2% in the cases of 1, 2, 3, 5, and 10 shot on data split 1 of TT, respectively.

Table 3. Ablation of each component of split 1.

Method	CCM	TRR	TT					3VT				
			1	2	3	5	10	1	2	3	5	10
Baseline	-	-	67.4	68.6	69.7	69.6	72.8	57.8	58.4	59.2	57.4	60.5
Ours	✓	✗	68.8	**72.3**	75.1	79.0	79.4	63.7	64.3	66.2	67.6	72.6
	✗	✓	68.5	68.7	70.2	69.9	70.1	63.5	65.7	67.2	68.0	71.6
	✓	✓	**71.3**	71.8	**75.8**	**79.8**	**81.0**	**64.1**	**66.1**	**69.5**	**71.5**	**75.6**

4 Conclusion

In this work, we propose a novel few-shot medical object detection method in ultrasound images called TRR-CCM. CCM excels in capturing long- and short-term dependencies while simultaneously retaining critical channel-specific information, thereby enriching the holistic comprehension of the anatomical structure's contextual nuances. TRR learns topological knowledge of human anatomy with a high degree of consistency to assist models in more accurate localization and classification. Experimental results on two datasets demonstrate the superiority of TRR-CCM, which shows the potential of our method for structure detection in clinical applications.

Acknowledgments. This study was funded by the Special Project of Central Government for Local Science and Technology Development (No. 202407AC110005), Yunnan Province Chen Jie Expert Workstation (No. 202305AF150065), Xingdian Talent Project of Yunnan Province (XYYC-CYCX-2022-0005), and the Natural Science Foundation of Hunan Province (No. 2025JJ60408).

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

References

1. Behrouz, A., Santacatterina, M., Zabih, R.: Mambamixer: Efficient selective state space models with dual token and channel selection. ArXiv **abs/2403.19888** (2024)
2. Gao, B.B., Chen, X., Huang, Z., Nie, C., Liu, J., Lai, J., Jiang, G., Wang, X., Wang, C.: Decoupling classifier for boosting few-shot object detection and instance segmentation. In: Neural Information Processing Systems (2022)
3. Gu, A., Dao, T.: Mamba: Linear-time sequence modeling with selective state spaces. arXiv preprint arXiv:2312.00752 (2023)
4. Guo, J., Tan, G., Lin, J., Pu, B., Wen, X., Wang, C., Li, S., Li, K.: Anatomical structures detection using topological constraint knowledge in fetal ultrasound. Neurocomputing **619**, 129143 (2025)
5. Han, J., Ren, Y., Ding, J., Yan, K., Xia, G.: Few-shot object detection via variational feature aggregation. In: AAAI Conference on Artificial Intelligence (2023)
6. Jiang, H., Gao, M., Li, H., Jin, R., Miao, H., Liu, J.: Multi-learner based deep meta-learning for few-shot medical image classification. IEEE Journal of Biomedical and Health Informatics **27**(1), 17–28 (2023)
7. Li, X., Tan, Y., Liang, B., Pu, B., Yang, J., Zhao, L., Kong, Y., Yang, L., Zhang, R., Li, H., et al.: Tkr-fsod: Fetal anatomical structure few-shot detection utilizing topological knowledge reasoning. IEEE Journal of Biomedical and Health Informatics (2024)
8. Lin, Z., Li, S., Ni, D., Liao, Y., Wen, H., Du, J., Chen, S., Wang, T., Lei, B.: Multi-task learning for quality assessment of fetal head ultrasound images. Medical image analysis **58**, 101548 (2019)
9. Lu, L., Cui, X., Tan, Z., Wu, Y.: Medoptnet: Meta-learning framework for few-shot medical image classification. IEEE/ACM Transactions on Computational Biology and Bioinformatics **21**(4), 725–736 (2024)

10. Lu, X., Diao, W., Mao, Y., Li, J., Wang, P., Sun, X., Fu, K.: Breaking immutable: Information-coupled prototype elaboration for few-shot object detection. In: AAAI Conference on Artificial Intelligence (2022)
11. Ma, J., Niu, Y., Xu, J., Huang, S., Han, G., Chang, S.F.: Digeo: Discriminative geometry-aware learning for generalized few-shot object detection. 2023 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR) pp. 3208–3218 (2023)
12. Nayem, J., Hasan, S.S., Amina, N., Das, B., Ali, M.S., Ahsan, M.M., Raman, S.: Few shot learning for medical imaging: A comparative analysis of methodologies and formal mathematical framework. ArXiv **abs/2305.04401** (2023)
13. Patro, B.N., Agneeswaran, V.S.: Simba: Simplified mamba-based architecture for vision and multivariate time series. ArXiv **abs/2403.15360** (2024)
14. Pu, B., Li, K., Li, S., Zhu, N.: Automatic fetal ultrasound standard plane recognition based on deep learning and iiot. IEEE Transactions on Industrial Informatics **17**(11), 7771–7780 (2021)
15. Pu, B., Lv, X., Yang, J., Dong, X., Lin, Y., Li, S., Li, K., Li, X.: Leveraging anatomical consistency for multi-object detection in ultrasound images via source-free unsupervised domain adaptation. In: Proceedings of the AAAI Conference on Artificial Intelligence. vol. 39, pp. 6532–6540 (2025)
16. Pu, B., Lv, X., Yang, J., He, G., Dong, X., Lin, Y., Li, S., Ying, T., Liu, F., Chen, M., Jin, Z., Li, K., Li, X.: Unsupervised domain adaptation for anatomical structure detection in ultrasound images. In: International Conference on Machine Learning (2024)
17. Pu, B., Wang, L., Yang, J., Dong, X., Ma, B., Chen, Z., Zhao, L., Li, S., Li, K.: Anatomical knowledge mining and matching for semi-supervised medical multi-structure detection. In: Proceedings of the AAAI Conference on Artificial Intelligence. vol. 39, pp. 6523–6531 (2025)
18. Pu, B., Wang, L., Yang, J., He, G., Dong, X., Li, S., Tan, Y., Chen, M., Jin, Z., Li, K., Li, X.: M3-uda: A new benchmark for unsupervised domain adaptive fetal cardiac structure detection. In: 2024 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR). pp. 11621–11630 (2024)
19. Qiao, L., Zhao, Y., Li, Z., Qiu, X., Wu, J., Zhang, C.: Defrcn: Decoupled faster r-cnn for few-shot object detection. 2021 IEEE/CVF International Conference on Computer Vision (ICCV) pp. 8661–8670 (2021)
20. Shen, Y., Fan, W., Han, Z., Zhou, D.: Multi-level feature-guided network for few-shot medical image segmentation. In: 2024 27th International Conference on Computer Supported Cooperative Work in Design (CSCWD). pp. 1346–1351 (2024)
21. Sun, B., Li, B., Cai, S., Yuan, Y., Zhang, C.: Fsce: Few-shot object detection via contrastive proposal encoding. 2021 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR) pp. 7348–7358 (2021)
22. Tang, K., Wang, S., Chen, Y.: Cross modulation and region contrast learning network for few-shot medical image segmentation. IEEE Signal Processing Letters **31**, 1670–1674 (2024)
23. Wang, X., Huang, T.E., Darrell, T., Gonzalez, J.E., Yu, F.: Frustratingly simple few-shot object detection. ArXiv **abs/2003.06957** (2020)
24. Wells, P.N.: Ultrasound imaging. Physics in medicine & biology **51**(13), R83 (2006)
25. Wu, S., Pei, W., Mei, D., Chen, F., Tian, J., Lu, G.: Multi-faceted distillation of base-novel commonality for few-shot object detection. In: European Conference on Computer Vision (2022)

26. Yan, J., Feng, K., Zhao, H., Sheng, K.: Siamese-prototypical network with data augmentation pre-training for few-shot medical image classification. In: 2022 2nd International Conference on Frontiers of Electronics, Information and Computation Technologies (ICFEICT). pp. 387–391 (2022)
27. Zhao, L., Li, K., Pu, B., Chen, J., Li, S., Liao, X.: An ultrasound standard plane detection model of fetal head based on multi-task learning and hybrid knowledge graph. Future Generation Computer Systems **135**, 234–243 (2022)