Submission and Formatting Instructions for International Conference on Machine Learning (ICML 2026)

Anonymous Authors[1]

Abstract

This document provides a basic paper template and submission guidelines. Abstracts must be a single paragraph, ideally between 4–6 sentences long. Gross violations will trigger corrections at the camera-ready phase.

1. Introduction

2. Related Work

This section reviews related work from three perspectives: efficient network architecture design, small object detection, and frequency domain analysis in object detection.

2.1. Efficient Network Architecture Design

Efficient network architectures are crucial for real-time object detection. The YOLO series has been a cornerstone, with recent iterations like YOLOv10 (Wang et al., 2024) and YOLOv11 (Khanam & Hussain, 2024) achieving improvements in accuracy and speed. FBRT-YOLO (Xiao et al., 2025) further enhances YOLO for small object detection through feature complementary mapping modules and multi-kernel perception units.

Beyond YOLO, DETR (Detection Transformer) (Carion et al., 2020) eliminated anchor boxes and introduced end-to-end detection, though its quadratic complexity limits efficiency. RT-DETR (Zhao et al., 2024) and DINO (Zhang et al., 2022a) address this through efficient attention mechanisms. ELAN (Zhang et al., 2022b) and LSNet (Wang et al., 2025) demonstrate effective layer aggregation strategies, with LSNet employing large kernels for global perception and small kernels for local aggregation. MobileU-ViT (Tang et al., 2025) combines CNNs and Transformers in a lightweight framework using large-kernel depthwise separable convolutions.

[1]Anonymous Institution, Anonymous City, Anonymous Region, Anonymous Country. Correspondence to: Anonymous Author <anon.email@domain.com>.

Preliminary work. Under review by the International Conference on Machine Learning (ICML). Do not distribute.

2.2. Small Object Detection

Small object detection faces challenges due to limited spatial information and weak feature responses. Traditional frameworks like Faster R-CNN (Ren et al., 2015) and SSD (Liu et al., 2016) lose critical information during downsampling. HS-FPN (Shi et al., 2025) addresses this by leveraging discrete cosine transform (DCT) to extract high-frequency components through a High-frequency Perception (HFP) module and capturing pixel-level spatial dependencies via a Spatial Dependency Perception (SDP) module.

Multi-scale feature fusion approaches build upon Feature Pyramid Networks (FPN) (Lin et al., 2017), with recent enhancements incorporating attention mechanisms and specialized modules for small object detection (Shi et al., 2025). Context information modeling through large receptive fields also proves effective. LSNet (Wang et al., 2025) combines large and small kernels for global and local context, while FBRT-YOLO (Xiao et al., 2025) employs multi-kernel perception units to capture cross-scale relationships efficiently.

2.3. Frequency Domain Analysis in Object Detection

Frequency domain analysis offers complementary perspectives for feature extraction. Global Filter Networks (GFNet) (Rao et al., 2023) replace self-attention layers with global filter layers using 2D FFT/IFFT, achieving linear complexity while maintaining global receptive fields. Frequency Dynamic Convolution (FDConv) (Chen et al., 2025) constructs frequency-disentangled convolution kernels by dividing the spectrum into low, medium, and high-frequency bands, with Kernel Spectrum Modulation (KSM) and Frequency Band Modulation (FBM) for adaptive weighting.

Wavelet-based methods also demonstrate effectiveness. Haar wavelet downsampling (Xu et al., 2023) preserves edge and texture information crucial for small object detection. Wavelet Convolutions (WTConv) (Finder et al., 2025) expand receptive fields with logarithmic complexity growth. Adaptive Complex Wavelet Informed Transformer Operators (Li et al., 2025) integrate complex wavelet transforms into Transformers for multi-resolution analysis, showing that frequency domain techniques can be seamlessly integrated into modern architectures.

References

Carion, N., Massa, F., Synnaeve, G., Usunier, N., Kirillov, A., and Zagoruyko, S. End-to-end object detection with transformers. *arXiv preprint arXiv:2005.12872*, 2020.

Chen, L., Gu, L., Li, L., Yan, C., and Fu, Y. Frequency dynamic convolution for dense image prediction. *arXiv preprint arXiv:2503.18783*, 2025.

Finder, S. E., Amoyal, R., Treister, E., and Freifeld, O. Wavelet convolutions for large receptive fields. In Leonardis, A., Ricci, E., Roth, S., Russakovsky, O., Sattler, T., and Varol, G. (eds.), *Computer Vision – ECCV 2024*, pp. 363–380, Cham, 2025. Springer Nature Switzerland. ISBN 978-3-031-72949-2.

Khanam, R. and Hussain, M. YOLOv11: An overview of the key architectural enhancements. *arXiv preprint arXiv:2410.17725*, 2024.

Li, X., Jiao, L., Liu, F., Yang, S., Zhu, H., Liu, X., Li, L., and Ma, W. Adaptive complex wavelet informed transformer operator. *IEEE Transactions on Multimedia*, 27:3513–3526, 2025. doi: 10.1109/TMM.2025.3535392.

Lin, T.-Y., Dollár, P., Girshick, R., He, K., Hariharan, B., and Belongie, S. Feature pyramid networks for object detection. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, pp. 2117–2125, 2017.

Liu, W., Anguelov, D., Erhan, D., Szegedy, C., Reed, S., Fu, C.-Y., and Berg, A. C. SSD: Single shot multibox detector. In *Computer Vision–ECCV 2016: 14th European Conference, Amsterdam, The Netherlands, October 11-14, 2016, Proceedings, Part I*, pp. 21–37. Springer, 2016.

Rao, Y., Zhao, W., Zhu, Z., Zhou, J., and Lu, J. GFNet: Global filter networks for visual recognition. *IEEE Transactions on Pattern Analysis and Machine Intelligence (TPAMI)*, 45(9):10960–10973, September 2023. doi: 10.1109/TPAMI.2023.3263824.

Ren, S., He, K., Girshick, R., and Sun, J. Faster R-CNN: Towards real-time object detection with region proposal networks. In *Advances in Neural Information Processing Systems (NeurIPS)*, volume 28, 2015.

Shi, Z., Hu, J., Ren, J., Ye, H., Yuan, X., Ouyang, Y., He, J., Ji, B., and Guo, J. HS-FPN: High frequency and spatial perception fpn for tiny object detection. *arXiv preprint arXiv:2412.10116*, 2025.

Tang, F., Nian, B., Ding, J., Ma, W., Quan, Q., Dong, C., Yang, J., Liu, W., and Zhou, S. K. Mobile U-ViT: Revisiting large kernel and U-shaped vit for efficient medical image segmentation. *arXiv preprint arXiv:2508.01064*, 2025.

Wang, A., Chen, H., Liu, L., Chen, K., Lin, Z., Han, J., and Ding, G. YOLOv10: Real-time end-to-end object detection. *arXiv preprint arXiv:2405.14458*, 2024.

Wang, A., Chen, H., Lin, Z., Han, J., and Ding, G. LSNet: See large, focus small. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, 2025.

Xiao, Y., Xu, T., Xin, Y., and Li, J. FBRT-YOLO: Faster and better for real-time aerial image detection. *arXiv preprint arXiv:2504.20670*, 2025.

Xu, G., Liao, W., Zhang, X., Li, C., He, X., and Wu, X. Haar wavelet downsampling: A simple but effective downsampling module for semantic segmentation. *Pattern Recognition*, 143:109819, 2023. ISSN 0031-3203. doi: https://doi.org/10.1016/j.patcog.2023.109819. URL https://www.sciencedirect.com/science/article/pii/S0031320323005174.

Zhang, H., Li, F., Liu, S., Zhang, L., Su, H., Zhu, J., Ni, L. M., and Shum, H.-Y. DINO: DETR with improved denoising anchor boxes for end-to-end object detection. *arXiv preprint arXiv:2203.03605*, 2022a.

Zhang, X., Zeng, H., Guo, S., and Zhang, L. Efficient long-range attention network for image super-resolution. In *European Conference on Computer Vision*, pp. 649–667. Springer, 2022b.

Zhao, Y., Lv, W., Xu, S., Wei, J., Wang, G., Dang, Q., Liu, Y., and Chen, J. Detrs beat yolos on real-time object detection. *arXiv preprint arXiv:2304.08069*, 2024.

A. You *can* have an appendix here.

You can have as much text here as you want. The main body must be at most 8 pages long. For the final version, one more page can be added. If you want, you can use an appendix like this one.

The `\onecolumn` command above can be kept in place if you prefer a one-column appendix, or can be removed if you prefer a two-column appendix. Apart from this possible change, the style (font size, spacing, margins, page numbering, etc.) should be kept the same as the main body.