

Date: December 30, 2025

Subject: Submission of "HSP-DETR: Hierarchical Sparse Perception Transformer for Real-Time Aerial Object Detection"

Dear Editor,

Real-time detection of small objects in UAV imagery faces a critical dilemma: heavy models offer accuracy but fail on resource-constrained edge devices, while lightweight models often sacrifice the fine-grained details essential for recognizing tiny targets. To resolve this conflict, we submit our original research, "**HSP-DETR**" for your consideration.

We propose a lightweight Transformer architecture that achieves an optimal Pareto frontier between detection accuracy and inference efficiency. Our approach introduces an **Asymmetric Hierarchical Synergy (ASH)** strategy to RT-DETR, incorporating three key innovations:

1. **Detail Preservation:** A **DLEP** module utilizing SPDConv to prevent feature loss during downsampling, ensuring tiny targets remain distinguishable.
2. **Efficient Fusion:** A **LHCGF** mechanism that enables cross-scale feature interaction with minimal computational overhead.
3. **Scale-Adaptive Supervision:** A **SAQL** loss function that dynamically re-weights supervision signals to focus on hard-to-detect small objects.

Key Results:

On the challenging **VisDrone2019 dataset**, HSP-DETR significantly outperforms the baseline RT-DETR. It achieves a **2.9% increase in small object accuracy (APs)** and a **4.1% increase in mAP50:95**, while simultaneously **reducing computational complexity (GFLOPs) by 29.52%**.

This work offers a practical solution for deploying high-performance vision transformers on aerial platforms, fitting well within the scope of Computer Vision and Image Understanding.

We declare that this manuscript is original, has not been published, and is not under consideration elsewhere.

Sincerely,

Prof. Lei Liao (Corresponding Author)