

Hierarchical Post-Hoc Verification of Low-Confidence Detections for Robust Vehicle Tracking

Miká Müller

Abstract—The intrinsic trade-off between recall and precision limits the dependability of object detectors in safety-critical applications. This is addressed by presenting a post-hoc verification framework that adds lightweight, hierarchical auxiliary heads to a YOLOv11 detector. These heads use a rule-based "Waterfall Rescue" logic to validate low-confidence proposals at coarser semantic levels. Two significant engineering contributions make this method possible: a decoupled head architecture that reduces inter-head task conflicts and a "Unified-Compact-ROI" feature extraction pipeline that resolves I/O bottlenecks during training. The system achieves a +18.7% absolute increase in MOTA when tested on a zero-shot tracking benchmark. This gain is driven by a 27.6% relative reduction in false negatives, albeit with an increase in identity switches, demonstrating a pragmatic and effective method for augmenting computationally constrained detectors to significantly improve their robustness for high-recall applications.

Index Terms—Auxiliary Heads, False Positive Reduction, Hierarchical Object Detection, Vehicle Tracking, YOLOv11

I. INTRODUCTION

THE dependability of object detection systems is crucial in high-stakes situations like autonomous driving and public surveillance. Modern one-stage detectors, such as YOLOv11 [1], offer remarkable speed and accuracy, but they are essentially constrained by a difficult trade-off between recall and precision that is controlled by a single confidence threshold. False negatives (FNs) are inevitably increased when a high threshold is set to reduce false positives (FPs). Even a brief inability to identify an object in Multi-Object Tracking (MOT) can cause a lost track and a serious lapse in situational awareness.

Standard detectors are further limited by being "hierarchically blind," treating classes as a flat, unstructured list. They lack the intrinsic understanding that a *Bus* is a type of *heavy_vehicle*, which is in turn a *vehicle*. This semantic context is a powerful prior for reasoning about detections that are assigned low confidence at a fine-grained level but are unambiguously objects when viewed through a coarser lens.

A modular, post-hoc hierarchical verification framework is suggested as a solution to these drawbacks. The system uses lightweight auxiliary heads that work on shared features from the detector's neck to enhance a pre-trained, L3 (leaf-level) detector model, denoted as M_{L3} . As shown in Fig. 1, these auxiliary heads for L2 (group) and L1 (super) classification, H_{L2} and H_{L1} respectively, are then used by a "Waterfall Rescue" inference logic to either validate (rescue) or reject low-confidence L3 proposals. By using this method, the base



Fig. 1. Illustration of the Waterfall Rescue mechanism. (Left) The baseline YOLOv11 detector at a standard confidence threshold (> 0.5) misses several distant vehicles (false negatives). (Right) The hierarchical model, operating at a low candidate threshold, successfully recovers these detections. The L1/L2 auxiliary heads validate the recovered objects (orange and red boxes) and assign the appropriate hierarchical labels. Original image stems from the MIO-TCD validation dataset.

detector can be run at a much lower, high-recall threshold, giving the specialized verification heads the responsibility of confirming borderline cases and filtering false positives.

The main contributions of this work are:

- A modular post-hoc framework for hierarchical verification that improves a pre-trained detector without costly retraining.
- An efficient "Unified-Compact-ROI" feature extraction pipeline, which makes the training of data-intensive auxiliary heads computationally feasible.
- A decoupled auxiliary head architecture that resolves inter-head task conflict, enabling robust learning for both coarse and fine-grained validation tasks.
- A comprehensive MOT evaluation that quantifies the core trade-off of the system: a massive reduction in false negatives at the cost of increased identity switches, leading to a substantial overall improvement in tracking accuracy.

II. RELATED WORK

This work lies at the crossroad of contemporary one-stage object detection, post-hoc detection refinement, and hierarchical classification.

A. Hierarchical and Cascade-Based Detection

In computer vision, the idea of employing cascades for effective detection is fundamental. For instance, the Viola-Jones face detector used a series of basic classifiers to quickly eliminate non-face areas, allocating processing power only to those that showed promise [2]. This coarse-to-fine filtering principle remains a cornerstone of efficient detection systems.

End-to-end models that incorporate class taxonomies directly into their architecture are a modern adaptation of this

Miká Müller is with the University of Göttingen, Germany
e-mail: mika.mueller@stud.uni-goettingen.de.

Thank you for your interest.

Due to the unpublished nature of this research,
the full report is available upon request.

Request Full Report