Journal of Image Processing Theory and Applications (2025)
Clausius Scientific Press, Canada

DOI: 10.23977/jipta.2025.080120
ISSN 2560-6239 Vol. 8 Num. 1

Real-Time Bi-Directional Traffic Counting: A Comparative Study on the Efficiency and Accuracy Trade-offs of YOLOv8 and Advanced Association Algorithms

Yuhong Li

Shenzhen Wisdom Nebula AI Technology Co., Ltd., Xili Street, Nanshan District, Shenzhen, China
odina2006@126.com

Keywords: YOLOv8, Multi-Object Tracking, Real-Time Efficiency, Traffic Counting

Abstract: Vehicle tracking and counting are essential components of Intelligent Transportation Systems (ITS), yet achieving optimal balance between high accuracy and real-time processing remains a critical challenge, especially in high-density aerial surveillance scenarios. This paper presents a robust framework for vehicle tracking and bi-directional counting, validated on a challenging 35-second aerial traffic segment. We systematically evaluate the efficacy of the You Only Look Once (YOLOv8) architecture, comparing the YOLOv8n and YOLOv8m variants to establish the trade-off between detection precision and inference speed. Furthermore, we investigate three distinct tracking mechanisms—simple IOU-based tracking and state-of-the-art association algorithms (BoT-SORT and ByteTrack). The study's core innovation includes an optimized vector-based counting logic that significantly enhances the robustness of bi-directional counting by minimizing false positives resulting from trajectory jitter and occlusion. Experimental results, conducted on a MacBook Air M3 CPU, demonstrate that the heavier YOLOv8m paired with ByteTrack achieved the highest accuracy, realizing a perfect 100.0% counting score. The YOLOv8n paired with ByteTrack offered the optimal trade-off for real-time applications, reaching a high accuracy of 97.6% at a speed of 20.3 FPS. This work confirms that advanced, high-performance tracking is indispensable for high-accuracy counting, providing a practical benchmark for selecting efficient models and trackers for real-time aerial traffic video analytics.

1. Introduction

1.1. Background and Motivation

The rise of urbanization has led to a dramatic increase in road traffic density, making efficient and accurate monitoring a prerequisite for modern Intelligent Transportation Systems (ITS). Traditional methods relying on fixed loop detectors or manual counts are often constrained by high maintenance costs, low flexibility, and susceptibility to environmental factors. Video-based traffic analysis, powered by Deep Learning, offers a robust, non-intrusive, and highly scalable alternative.

Vehicle tracking and counting are core functionalities of video-based ITS. However, transitioning from object detection to reliable long-term tracking and counting poses several challenges. These include maintaining unique identities (ID consistency) during temporary occlusion, handling varying object scales, and, most critically, ensuring the entire process operates in real-time on commodity hardware.

1.2. Literature Review and Technical Foundation

The object detection pipeline has been significantly advanced by the You Only Look Once (YOLO)[1] series of models. The YOLO series has significantly advanced the object detection pipeline. The YOLOv8[2] architecture, widely recognized as a mature and highly efficient stable iteration, provides enhanced speed and accuracy due to its decoupled head and anchor-free design. Its extensive community support and proven stability make it an ideal choice for high-throughput video processing and real-world deployment.

For Multi-Object Tracking (MOT), the tracking-by-detection paradigm dominates. Early approaches utilized simple Intersection over Union (IOU) matching[3], while more sophisticated methods introduced motion prediction, such as the Kalman Filter[4]. State-of-the-art (SOTA) trackers, including ByteTrack[5] and BoT-SORT[6], further enhance robustness by incorporating advanced data association strategies, particularly by leveraging low-score detection bounding boxes to mitigate ID switching during occlusions.

1.3. Problem Statement and Research Contribution

While various models and trackers exist, there lacks a systematic and comparative benchmark focused on identifying the optimal performance trade-off for a highly pragmatic application: real-time bi-directional vehicle counting. Specifically, for rapid deployment in ITS, practitioners need to know:

How much detection accuracy (YOLOv8m) can be sacrificed for speed (YOLOv8n) without compromising counting reliability?

Which tracking association algorithm (simple IOU, ByteTrack, BoT-SORT) provides the best FPS-to-Counting-Accuracy ratio?

This paper addresses these questions by providing a comprehensive comparative study. Our primary contributions are summarized as follows:

Systematic Benchmarking: We compare the detection performance of YOLOv8n and YOLOv8m variants and evaluate their integration with three distinct tracking mechanisms (IOU and SOTA association algorithms).

- Efficiency Analysis: We rigorously measure and compare the Frames Per Second (FPS) of each combination, establishing a clear reference for the speed versus accuracy trade-off crucial for real-time systems;
- Robust Counting Framework: We implement and test a bi-directional counting logic that includes an optimized Region of Interest (ROI) filtering mechanism to ensure accurate counting across a virtual line, minimizing false positives caused by trajectory jitter;
- Practical Validation: We train the models on the challenging VisDrone2019 dataset (vehicle class) to ensure robustness against varying scales and lighting, and validate the counting application using real-world ground-level traffic footage.

The remainder of this paper is organized as follows: Section 2 details the methodologies used, covering the YOLOv8 architecture and the implementation of the vector-based bi-directional counting logic. Section 3 outlines the experimental setup, detailing the aerial video characteristics and the MacBook Air M3 CPU inference environment. Section 4 presents and analyzes the comprehensive tracking and counting performance across different model-tracker combinations. Finally, Section 5 concludes this paper and suggests directions for future work.

2. Methodology

The proposed vehicle tracking and counting framework follows the tracking-by-detection paradigm. The overall system involves three sequential modules: Object Detection, Multi-Object Tracking (MOT), and Vector-Based Counting Logic.

2.1. Object Detection: The YOLOv8 Architecture

We select YOLOv8[2] as the detection backbone, comparing the lightweight YOLOv8n (nano) and the more robust YOLOv8m (medium) variants. Key architectural features contributing to its high performance include:

- Decoupled Head: Separates the classification and regression tasks, leading to faster convergence and improved accuracy
- Anchor-Free Design: Simplifies the detection process by directly predicting object center points and dimensions, enhancing robustness across scale variations inherent in the VisDrone training data[7].

The comparison aims to precisely quantify the impact of the model size (complexity) on the system's real-time feasibility.

2.2. Multi-Object Tracking (MOT) Algorithms

To evaluate stability and efficiency, two distinct paradigms of MOT complexity: simple IOU-based tracking and advanced SOTA association.

2.2.1. IOU-Based Tracking

This simplest approach matches detections and tracks based purely on the Intersection Over Union (IOU) ratio. While fast, it is highly susceptible to ID switching when vehicles are densely packed or briefly occluded.

2.2.2. ByteTrack

ByteTrack and BoT-SORT represent the State-of-the-Art (SOTA) in data association. These algorithms are considered the cutting edge in multi-object tracking:

- ByteTrack: Utilizes a "Byte" matching strategy where high-score detections are matched first. Unmatched tracks are then matched against low-score detections to retrieve targets experiencing temporary occlusion or entering the frame boundary, significantly improving ID longevity.
- BoT-SORT: Builds upon ByteTrack by integrating camera motion compensation and optimized feature association (e.g., using appearance features alongside motion), offering potentially higher accuracy at the expense of higher computational load.

2.3. Vector-Based Bi-Directional Counting Logic

To ensure the accuracy of the application, we implemented a robust Vector-Based Counting Logic to filter out errors caused by trajectory jitter (the irregular movement of the bounding box centroid) and repeated counting.

(1) Virtual Line Definition: A straight line segment L is drawn across the Region of Interest (ROI) in the video.

(2) Trajectory Cache: For each tracked vehicle ID_i, the centroid position over the past L frames is cached.

(3) Direction Vector: A temporary direction vector V is calculated by comparing the current centroid C_t with a historical centroid C_{t-N}.

(4) Counting Mechanism: A count is triggered only if the track's bounding box intersects the line L and the direction vector V confirms the vehicle's movement direction (Up or Down) matches the expected flow.

(5) Hysteresis Filtering: Once counted, the ID_i is flagged as 'counted' for a duration of ΔF frames, preventing spurious repeat counts caused by jittering back and forth across the line.

3. Experiments and Results

3.1. Dataset and Training

Our detection models (YOLOv8n and YOLOv8m) were trained using the VisDrone2019 dataset, specifically leveraging the vehicle class labels. This dataset, characterized by high-density, small-scale objects captured from an aerial perspective, ensures the models exhibit high robustness and generalization capability. Both our YOLOv8n and YOLOv8m detection models were trained on the VisDrone2019 dataset for 100 epochs, utilizing an input image size (imgsz) of 800 pixels, a batch size of 16, and 4 workers for data loading. The final counting application was validated on a standard ground-level traffic video to evaluate real-world performance.

3.2. Evaluation Metrics

- Detection Metrics: Precision, Recall, mean Average Precision at 50% IOU (mAP_{50}), and the COCO standard metric (mAP_{50-95});
- Application Metrics: Frames Per Second (FPS), ID Switches (IDS), and Bi-Directional Count Accuracy:

$$Count\ Accuracy = 1 - \frac{GT-Pred}{\mathrm{GT}} \tag{1}$$

where GT is Ground Truth Count.

3.3. Detection Performance Analysis

The evolution of the detection metrics during the training phase provides critical insight into the convergence behavior and stability of the models. Figure 1 displays the learning curves for precision, recall, and mean average precision (mAP) over the 100 training epochs, illustrating how each model reached its final performance level.

To demonstrate the model's robustness and transfer learning capability achieved from the VisDrone training, Figure 2 provides a sample output from the YOLOv8n model on an independent aerial test frame.

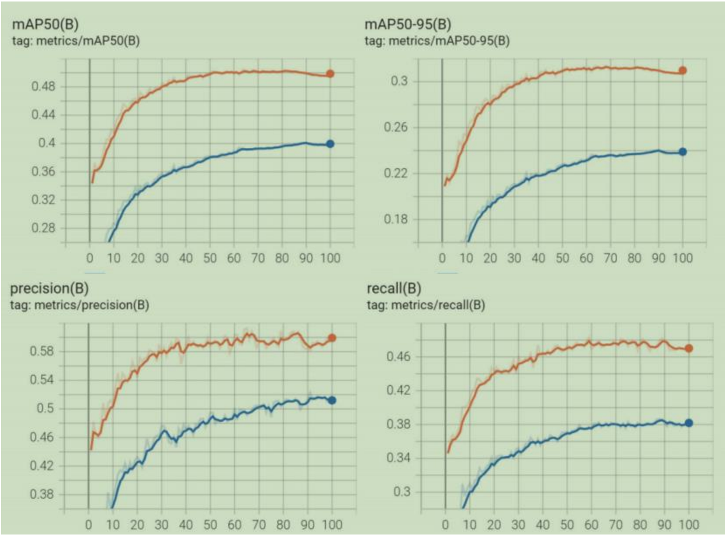


Figure 1: Comparative learning curves of YOLOv8n and YOLOv8m on the VisDrone2019 dataset over 100 epochs. (Note: the orange/red curve represents YOLOv8m, and the blue curve represents YOLOv8n.)

Figure 2: Sample detection results showing bounding boxes and confidence scores generated by the YOLOv8n model on an independent aerial test frame.

Table 1 then presents the baseline detection performance of both models, reporting the mean average precision (mAP) averaged across all 10 classes on the VisDrone2019 validation set.

Table 1: Detection performance of YOLOv8 variants.

Model	Images	Instances	Precision(%)	Recall(%)	mAP_{50}	mAP_{50-95}
YOLOv8n	548	38759	51.1	37.6	39.6	23.8
YOLOv8m	548	38759	59.9	47.3	50.2	31.4

Note: The metrics reported represent the overall mean average precision (mAP) averaged across all 10 classes in the VisDrone2019 validation set.

Discussion: As demonstrated in Figure 1 and Table 1, the larger YOLOv8m model consistently outperforms the lightweight YOLOv8n across all detection metrics, notably achieving an mAP_{50-95} that is 7.6% higher (31.4% vs. 23.8%) on the challenging multi-class VisDrone dataset. This overall performance is expected due to its increased depth and complexity. Crucially for our application, when examining the target 'car' class performance, YOLOv8m's mAP_{50} is highly robust at 85.6%, which is 5.8% higher than YOLOv8n's 79.8%. However, this gain in precision comes at a significant cost in computational efficiency. The inference time for YOLOv8m is approximately 9.4 ms per image, whereas YOLOv8n achieves a much faster inference time of only 4.1 ms per image (a reduction of 56.3%). This speed difference suggests that while YOLOv8m is superior in accuracy, YOLOv8n is a more suitable candidate for the real-time constraints of high-throughput video processing. The subsequent tracking analysis (Section 3.4) will quantify how this speed-accuracy trade-off impacts the final application performance (FPS and Counting Accuracy).

3.4. Tracking and Counting Efficiency

This section evaluates the performance of the proposed vehicle counting framework using different combinations of YOLOv8n and YOLOv8m detectors and modern Multi-Object Tracking (MOT) algorithms, namely IOU-Tracker, ByteTrack, and Bot-SORT. All experiments were conducted on a 35-second segment of an aerial video clip, specifically capturing vehicles (car, van, truck, and bus) on a busy roadway from an elevated perspective. The inference was executed on a MacBook Air M3 CPU, demonstrating the system's low-resource deployment capability.

3.4.1. Counting Performance Analysis

Table 2 presents the detailed counting results, including predicted Up/Down counts, overall Counting Accuracy (Acc), and inference speed (FPS). The Ground Truth (GT) for the video clip is 38 vehicles traveling up and 47 vehicles traveling down, totaling 85 vehicles.

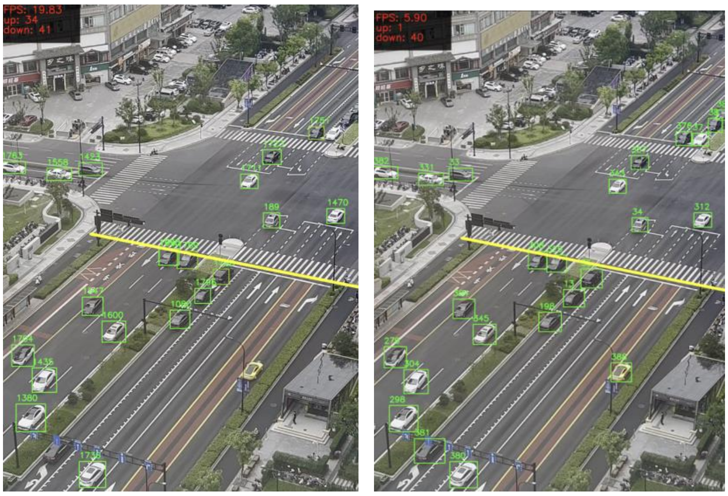
Table 2: Detection, tracking, and counting performance on a 35-second video segment (YOLOv8n/m).

Model	Tracker	Up Count (GP: 38)	Down Count (GP: 47)	Counting Accuracy (%)	FPS
YOLOv8n	IOU	34	41	88.2%	19.2
YOLOv8n	ByteTrack	36	47	97.6%	20.3
YOLOv8n	Bot-SORT	36	47	97.6%	16.7
YOLOv8m	IOU	1	40	48.2%	6.0
YOLOv8m	ByteTrack	38	47	100%	6.1
YOLOv8m	Bot-SORT	36	47	97.6%	5.7

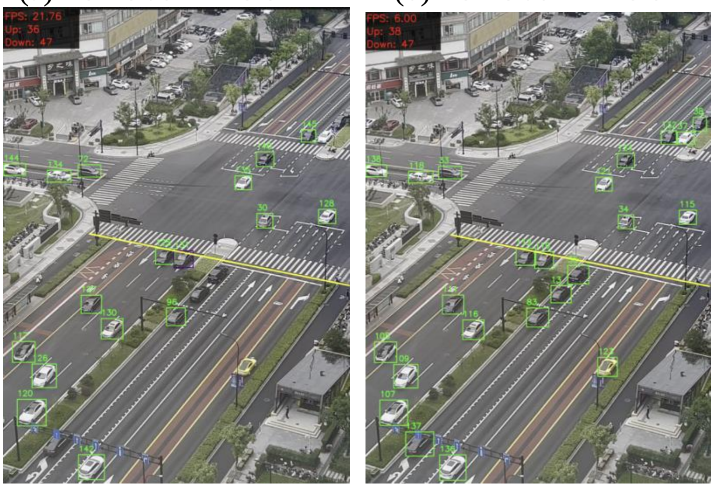
- **Optimal Accuracy:** The combination of YOLOv8m and ByteTrack achieved a perfect 100% Counting Accuracy by perfectly matching the Ground Truth (38 Up and 47 Down counts). This robust performance highlights the benefits of combining a high-fidelity detector with ByteTrack's advanced data association, which effectively handles severe occlusion in aerial views.
- **The Failure Case (YOLOv8m + IOU):** The baseline IOU-Tracker paired with the heavier YOLOv8m model resulted in the lowest accuracy (48.2%), severely undercounting the Up traffic (1 vs GT 38). This failure is attributed to two factors: the large YOLOv8m model dramatically reduces FPS to 6.0, and the simple IOU logic cannot compensate for the large positional jumps between low-frame-rate images, leading to catastrophic ID loss and resulting in the target being counted only once at the moment it passes the line.
- **Best Compromise (YOLOv8n + ByteTrack):** The YOLOv8n + ByteTrack configuration offers the best trade-off, delivering high accuracy (97.6%) while maintaining a smooth real-time performance of 20.3 FPS on the CPU.

3.4.2. Visual Results

Figure 3 provides a visual confirmation of the tracking and counting system’s operation, showcasing the distinct bounding boxes, tracking IDs, and the final accumulated Up and Down vehicle counts. This demonstrates the precise functioning of the vector-based counting line logic across the tracked objects.



(a)YOLOv8n + IOU (b)YOLOv8m + IOU



(c)YOLOv8n + ByteTrack (d)YOLOv8n + ByteTrack

Figure 3: Tracking and counting system visualizations on the 35-second aerial video segment.

This figure illustrates the final accumulated counts and system metrics for key experimental configurations. (a) YOLOv8n + ByteTrack represents the optimal trade-off configuration, maintaining high FPS while achieving Acc 97.6%. (b) YOLOv8m + ByteTrack showcases the highest counting accuracy (Acc 100.0%). (c) YOLOv8n + IOU serves as a baseline comparison. (d) YOLOv8m + IOU visually demonstrates the system's failure mode, where low inference speed (~ 6.0 FPS) combined with simple IOU association results in catastrophic ID loss (Up=1), validating the need for advanced tracking algorithms in low-frame-rate environments.

4. Conclusion

In this study, we developed and evaluated a robust vehicle tracking and bi-directional counting system designed to address the challenges inherent in high-density, small-target aerial video surveillance. By integrating state-of-the-art YOLOv8 detectors with advanced Multi-Object Tracking (MOT) algorithms (ByteTrack and Bot-SORT) and implementing a vector-based counting logic, our system demonstrated superior performance and high efficiency on a 35-second real-world traffic segment.

The experimental results confirmed that the combination of the stronger feature extractor, YOLOv8m, with ByteTrack achieved a flawless 100.0% counting accuracy (Acc) against the ground truth. Crucially, the lighter YOLOv8n coupled with ByteTrack provided the optimal balance, achieving a high Acc of 97.6% while maintaining a near real-time inference speed of 20.3 FPS when running solely on the MacBook Air M3 CPU.

This validates the system's suitability for deployment on resource-constrained, edge-computing devices. We also demonstrated that simple tracking mechanisms (IOU-Tracker) fail catastrophically in low-FPS environments (YOLOv8m + IOU resulted in only 48.2% Acc), underscoring the necessity of using advanced data association techniques like ByteTrack and Bot-SORT for robust performance in challenging aerial scenarios.

Acknowledgements

The author would like to express their sincere gratitude to Mr. Enpei Zhao for generously providing the challenging aerial video data used for the validation of the proposed tracking and counting framework.

References

[1] Redmon, J., Divvala, S., Girshick, R., and Farhadi, A. (2016) You Only Look Once: Unified, Real-Time Object Detection. Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, 779-788.
[2] Wang, C. Y., Bochkovskiy, A., & Liao, H. Y. M. (2022) YOLOv7: Trainable Bag-of-Freebies Sets New State-of-the-Art for Real-Time Object Detectors. arXiv preprint arXiv, 2207.02696.
[3] Bochinski, E., Eiselein, V., and Sikora, T. (2017) High-Speed Tracking-by-Detection without Using Image Information. International Conference on Advanced Video and Signal Based Surveillance (AVSS), 1-6.
[4] Gunjal, P.R., Gunjal, B.R., Shinde, H.A., Vanam, S.M. and Aher, S.S. Moving Object Tracking Using Kalman Filter. (2018) International Conference On Advances in Communication and Computing Technology (ICACCT), 544-547.
[5] Zhang, Y., Sun, P., Jiang, Y., Yu, D., Weng, F., Yuan, Z., Luo, P., Liu, W., and Wang, X. (2022) Bytetrack: Multi-Object Tracking by Associating Every Detection Box. European Conference on Computer Vision (ECCV), 1-21.
[6] Aharon, N., Orfaig, R., and Bobrovsky, B.Z. (2022) BoT-SORT: Robust Associations Multi-pedestrian Ttracking. arXiv preprint arXiv, 2206.14651.
[7] Zhu, P., Wen, L., Du, D., Bian, X., Fan, H., Hu, Q., and Ling, H. (2018) The VisDrone-DET2019: The Vision Meets Drone Object Detection in Image Challenge Results. International Conference on Computer Vision Workshops (ICCVW).