The Impact of Weather, Lighting, and Camera Viewpoint on the Traffic Monitoring Performance of YOLO-SORT

Enas Elshebli
Szechenyi Istvan University, Gyor, Hungary
elshebli.enas.abd.almajeed@hallagto.sze.hu (corresponding author)

Erdos Ferenc
Szechenyi Istvan University, Gyor, Hungary
erdosf@sze.hu

Anwar Khan
Department of Electronics, University of Peshawar, Peshawar, KPK, Pakistan
arkhan@uop.edu.pk

Vijayakumar Varadarajan
Swiss School of Business and Management, Geneva, Switzerland
vijayakumar.varadarajan@uts.edu.au

Received: 24 February 2026 | Revised: 31 March 2026 | Accepted: 17 April 2026

Licensed under a CC-BY 4.0 license | Copyright (c) by the authors | *DOI: https://doi.org/10.48084/etasr.18369*

ABSTRACT

This study investigates the condition sensitivity of a camera-based traffic monitoring pipeline that combines YOLOv8s detection with SORT tracking for real-time vehicle counting. Video data were collected from a pedestrian bridge above a two-lane highway in Győr, Hungary, under multiple weather conditions (clear, rainy, foggy), lighting (day, night) conditions, and viewpoint (central, moderate shift, maximum available shift) configurations. Pixels-per-meter calibration, normalized trajectories, and bounding-box dimensions across angles, and a trajectory-based count line produced event-level ground truth. Performance was evaluated using minute-level precision, recall, and F1-score, with condition-wise and angle-wise differences assessed via Wilcoxon signed-rank tests with Holm's correction and associated effect sizes. Across all angles, the pipeline maintained high F1-scores, with nighttime conditions yielding the highest and most stable performance, supported by consistent street lighting and lower traffic flow. Fog produced the most persistent degradation, while rain mainly caused brief precision dips during onset. Aggregated across conditions, no statistically significant differences were observed between viewpoints, indicating that within the studied range of angles, environmental factors dominate over camera orientation. Overall, the results suggest that YOLOv8s and SORT can support accurate real-time vehicle counting in CCTV-like deployments while remaining sensitive to reduced visibility and peak-traffic occlusions.

Keywords-traffic monitoring; vehicle detection and tracking; YOLO; SORT; camera viewpoint analysis; real-time video analytics

I. INTRODUCTION

Traffic data collection is fundamental to the development of efficient and sustainable transportation systems. Detailed information—such as vehicle volumes, classifications, and speeds—supports effective planning and management. As urban congestion and mobility demands grow, the need for accurate, timely traffic data has become increasingly critical [1]. Consequently, the rise of smart cities and intelligent transportation systems has stimulated substantial research, highlighting a growing need for accurate real-time traffic monitoring.

Conventionally, traffic data has been collected through manual methods or simple automated sensors. Manual counting, conducted either by observers in the field or through analysis of video recordings, has been regarded as the standard for accuracy, but remains labor-intensive, costly, and cannot

reasonably provide continuous 24/7 monitoring due to practical and economic constraints [2]. Automatic Traffic Counting and Classification (ATCC) systems, such as inductive loop detectors and pneumatic road tubes, have offered faster and more continuous data collection capabilities. However, they typically require substantial installation effort and ongoing maintenance [3], limiting their practicality in settings where deployment flexibility or low-cost scalability is required.

To overcome these limitations, recent advances in Artificial Intelligence (AI) and computer vision have begun to transform traffic data collection. Deep learning techniques now enable automated detection and classification of road users from video recordings with high accuracy and speed [4]. Models like YOLO have offered real-time object detection capabilities, identifying vehicles, pedestrians, and other objects in each video frame [1]. When combined with tracking algorithms such as the Simple Online and Realtime Tracker (SORT), these models can not only count vehicles but also follow their trajectories through a scene, estimating speeds, travel paths, and interaction events [5]. These AI-powered systems extend beyond simple detection, enabling richer traffic descriptors—such as lane-specific volumes, turning movements, and safety-related interactions—captured over longer periods and broader spatial extents than conventional methods.

AI-based systems such as YOLO-SORT represent a promising alternative to conventional infrastructure-based monitoring. Despite progress in automation, many existing systems still rely on fixed roadside infrastructure [6], and most evaluation studies isolate either detection accuracy under specific adverse conditions or tracking stability under comparatively favorable environments. The combined sensitivity of a lightweight detection–tracking pipeline to weather, lighting, and camera viewpoint remains poorly characterized in the literature. Addressing this gap is essential for establishing how such systems perform under the full range of variability present in real-world CCTV deployments. This study evaluates YOLO-SORT for real-time vehicle counting under jointly varying environmental and geometric conditions, providing empirical evidence on its adaptability and robustness in CCTV-like deployments.

YOLO detects objects in real time by predicting bounding boxes and class scores in a single network pass, with non-maximum suppression producing the final detections [7]. Later iterations improved the accuracy-speed trade-off with techniques such as multi-scale training, batch normalization, and anchor boxes [8]. This unified design avoids region-proposal staging and minimizes per-frame latency, which is critical for streaming traffic videos.

Related architectures include one-stage Single Shot Detectors (SSD) [9], anchor-free Fully Convolutional One-Stage (FCOS) object detection [10], and two-stage faster Region-based Convolutional Neural Networks (R-CNN) [11]. RetinaNet adds focal loss to stabilize one-stage training under foreground-background imbalance [12], while EfficientDet achieves strong accuracy via compound scaling [13]. Large comparative studies show that YOLO-style one-stage models typically offer the best balance of speed and accuracy for real-time settings, whereas two-stage methods trade higher accuracy for higher latency [14]. For traffic surveillance, where throughput, end-to-end delay, and robustness to scene dynamics matter, this balance makes YOLO a pragmatic choice: it is fast enough for live CCTV-like streams yet accurate enough to support downstream analytics such as lane-level counting and flow estimation [7, 14].

Tracking methods such as SORT [15], DeepSORT [16], and ByteTrack [17] extend these detectors to multi-object tracking by linking per-frame detections into continuous trajectories. SORT is particularly attractive for lightweight traffic pipelines due to its low computational overhead and minimal tuning, while DeepSORT and ByteTrack introduce appearance cues or low-score associations to improve identity persistence under occlusion. SORT is intentionally adopted in this study so that environmental and geometric effects can be evaluated without confounding from heavier re-identification modules.

Prior studies report reliable vehicle detection across day, night, and adverse weather when YOLO models are trained appropriately [18, 19]. Although these results establish the suitability of YOLO for controlled scenarios, evaluations in the literature are typically conducted under constrained conditions—either analyzing detection performance in specific adverse environments or examining tracking stability under relatively favorable settings. Most YOLO-based traffic counters assume stable lighting, a single camera viewpoint, and limited weather variation. Importantly, to the best of our knowledge, no published work quantifies how a lightweight detection-tracking pipeline behaves when multiple environmental and geometric factors vary simultaneously. This creates a gap concerning the joint sensitivity of YOLO-SORT to fog, rain, day-night illumination shifts, and camera-orientation changes.

Recent deployments have paired YOLO detectors with lightweight Multi-Object Tracking (MOT) frameworks to build practical traffic counters from video streams. For example, in [20], YOLO was integrated with DeepSORT for low-cost vehicle monitoring, reporting strong real-time performance, although without a systematic analysis of viewing geometry or weather interactions. More challenging settings have also been explored: In [21], daytime and nighttime counting-classification was performed with YOLOv3 under low illumination. Newer systems migrate to YOLOv8 and incorporate MOT for continuous identities and higher-level indicators; for instance, in [22], YOLOv8 was combined with DeepSORT for real-time tracking and congestion analysis, and in [23], TrackNCount (YOLOv8+DeepSORT) was introduced for joint tracking, counting, and speed estimation [23]. These works collectively show increasing maturity in YOLO-based traffic analytics, but they do not systematically evaluate multi-factor condition sensitivity.

Although recent studies have improved YOLO-based detectors for adverse weather [24-26] and low-light environments [27-29], these efforts focus primarily on detection and do not evaluate the full detector-tracker pipeline. Real-time YOLO-based vehicle detection systems have also been deployed on embedded and edge platforms for intelligent transportation and ADAS applications [30, 31], demonstrating

practical feasibility under resource-constrained deployment settings. As a result, the interaction between degraded visibility and downstream data association remains underexplored.

Recent low-light and degraded-visibility YOLO studies provide useful context for interpreting the performance changes observed under reduced visibility. Lightweight detector variants such as 3L-YOLO [32] show that low-light performance can be improved through targeted architectural changes without relying on a separate image-enhancement stage, and their evaluation on dedicated low-light benchmarks such as ExDark, ExDark+, and DARK FACE highlights the role of weak contrast, blurred boundaries, and unstable confidence in nighttime detection. Vehicle- and road-scene-oriented extensions such as LLD-YOLO [27] and DRF-YOLO [33] further indicate that recent designs increasingly combine illumination-aware modules, attention-enhanced feature extraction, and multi-scale feature fusion, with validation on datasets such as ExDark and the nighttime subset of BDD100K. Together, this literature helps explain why degraded illumination and visibility primarily affect contrast-sensitive localization and detection stability in traffic scenes.

The broader YOLO family has also continued to evolve in ways relevant to robustness and real-time deployment. YOLOv10 introduces end-to-end NMS-free training through consistent dual assignments together with a holistic efficiency-accuracy-driven redesign of detector components, while Ultralytics YOLOv11 emphasizes improved feature extraction, multi-task support, and deployment-oriented efficiency within a practical software framework. YOLOv12 moves further toward an attention-centric real-time architecture, aiming to strengthen global feature modeling while preserving competitive inference speed. Taken together, these developments provide a useful detector-side context for interpreting performance under reduced visibility, low contrast, and deployment-oriented real-time constraints [34-36].

Real-time YOLO studies in harsh operational settings further broaden the deployment context of recent detection pipelines. In environments such as underground mining and offshore oil platforms, recent models have been evaluated under complex lighting, cluttered backgrounds, occlusion, and real-time operational constraints rather than only under controlled laboratory conditions [37, 38]. These studies show that lightweight or deployment-oriented YOLO variants are increasingly being assessed in safety-critical scenes where visibility degradation, background interference, and computational efficiency must be handled simultaneously, making them relevant for interpreting robustness in practical monitoring applications.

These studies provide useful context, but they remain centered mainly on detection quality or detector redesign, whereas this study examines the behavior of a fixed YOLOv8s+SORT pipeline at the level of detection, tracking, and event-based counting under jointly varying weather and lighting conditions and viewpoint configurations.

It is important to distinguish the nature of these stressors: adverse weather primarily degrades image quality (contrast loss, scattering, sensor noise), whereas camera-viewpoint changes alter projective geometry, occlusion frequency, and apparent object size. Although both forms of variation occur simultaneously in roadside CCTV systems, prior studies have examined them almost exclusively in isolation. Currently, a unified evaluation that treats environmental and geometric variation jointly is missing.

This study addresses these gaps by evaluating the YOLO-SORT model across three key dimensions: weather (clear, foggy, rainy), lighting (day, night), and camera orientation (central, moderate shift, maximum available shift). Two primary configurations are tested:

- Daylight scenarios with varying weather and angles, and
- Nighttime scenarios under clear and rainy weather conditions with varying angles.

This partial factorial design enables the analysis of key two-way interactions while preserving interpretability. The novelty lies in its systematic multi-condition sensitivity analysis, which offers quantitative insights into the joint effects of environmental and geometric variability on detection and tracking performance.

II. RESEARCH METHODOLOGY

This section describes the experimental framework used to evaluate the YOLO-SORT pipeline under jointly varying environmental and geometric conditions. The methodology standardizes all model settings, camera parameters, Region of Interest (RoI) definitions, and evaluation procedures so that any observed performance differences arise solely from scene-level factors rather than from algorithmic or sensor variation.

This study evaluates a modular, camera-based traffic pipeline that performs real-time vehicle detection, tracking, and counting from video recordings. The pipeline operates frame-by-frame on recorded CCTV-like footage and consists of four main stages: (i) video ingestion and RoI definition, (ii) object detection with YOLO, (iii) multi-object tracking with SORT, and (iv) trajectory-based event counting and performance evaluation. All experiments use fixed, globally applied model settings across every condition to ensure comparability, and no task-specific retraining or fine-tuning is performed. This ensures that all differences observed in performance arise from environmental and geometric factors rather than model adaptation.

A. Video Ingestion and Region of Interest (RoI) Definition

Videos are processed at their native resolution and frame rate and read sequentially frame-by-frame. For each viewpoint, a roadway RoI is defined in the image to exclude other non-traffic areas. In this study, the RoI is represented by two horizontal line segments—an upper and a lower boundary—that visually bracket the carriageway. These boundaries serve as a simple and reproducible geometric mask, ensuring that subsequent detection, tracking, and counting operate strictly within the active traffic zone. A line-based ROI is intentionally used instead of polygonal masks to avoid viewpoint-dependent variability and to maintain perfect consistency across all environmental conditions.

Figure 1 illustrates a typical frame with ROI boundaries (blue) and YOLO detections (green). Only vehicles traveling within the delineated roadway corridor are processed, while areas outside the RoI remain excluded from analysis. This consistent geometric constraint isolates the measurement zone across all environmental and geometric conditions.

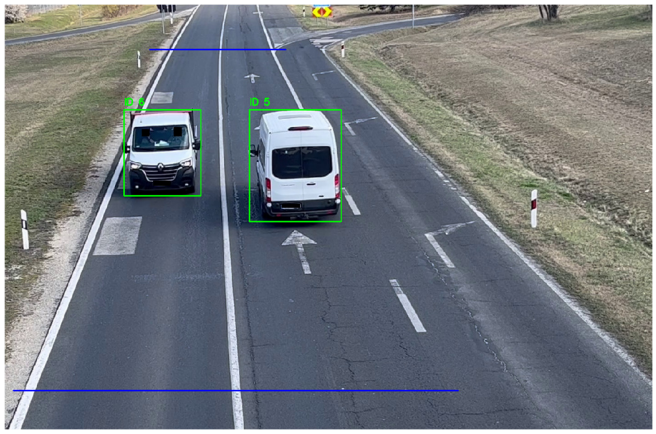

Fig. 1. Example video frame illustrating the roadway RoI used for analysis.

B. Object Detection with YOLO

Vehicle detection is performed using the YOLOv8 small configuration (YOLOv8s), initialized with MS-COCO pretrained weights. Detector hyperparameters were held constant across all conditions: 640×640 px input resolution (letterboxed), 0.5 confidence threshold, 0.5 NMS Intersection over Union (IoU) threshold, 100 (fixed) maximum detections per frame, and identical preprocessing settings. Inference was executed in streaming mode (batch = 1) with FP16 precision to maintain real-time throughput. Standard post-processing is applied:

- Non-Maximum Suppression (NMS) is used to remove duplicate or heavily overlapping boxes for the same object.
- Only road-user classes relevant to traffic analysis are retained, which are cars, buses, trucks, and motorcycles.
- A confidence threshold is applied so that low-score proposals are discarded before tracking.
- A maximum number of detections per frame and a minimum bounding-box area gate are used to suppress spurious tiny detections outside the roadway.

These post-processing choices follow standard practice and are kept fixed across all conditions; no per-condition tuning is performed so that differences in performance reflect environmental and geometric effects rather than parameter optimization. The remaining detections, in the form $[x_1, y_1, x_2, y_2, score]$, are then passed to the SORT tracker after RoI filtering.

C. Multi-Object Tracking with SORT

Multi-object tracking is handled by the SORT algorithm, which combines a linear Kalman filter with an IoU-based assignment step. SORT is used with its standard configuration: maximum track age at 30 frames, minimum track confirmation hit at 3, and IoU association threshold at 0.3, and all Kalman filter covariance matrices are kept at published defaults. Using the published defaults ensures that tracking behavior reflects canonical SORT performance.

- Each tracked object is represented by a 7-dimensional state vector describing the bounding box center, scale (area), aspect ratio, and their temporal derivatives.
- At each frame, every active tracker propagates its state forward using a constant-velocity motion model. This yields a predicted bounding box for the current frame.
- Predicted boxes are matched to current detections using an IoU-based cost matrix.
- Matched trackers are updated with their corresponding detection via the Kalman filter measurement update.
- Unmatched detections initialize new trackers.
- Trackers without an associated detection increment an internal "time since update" counter.

This fixed configuration ensures that cross-condition differences originate from scene variation rather than tracker tuning.

D. Trajectory-Based Counting

Vehicle counting is performed on track trajectories rather than on raw detections. A single virtual count line is defined in image coordinates, common to all angles and conditions, and placed approximately orthogonal to the direction of traffic flow within the RoI. For each confirmed track:

- The sequence of track centers over time is monitored.
- A track contributes exactly one count when it crosses the count line with a consistent direction (signed crossing).
- A small spatial hysteresis band is applied around the line to avoid multiple counts due to jitter or small oscillations near the boundary.
- A minimum track age before eligibility for counting, so very short-lived, unreliable tracks are ignored.
- Tolerance for short detection gaps, handled via Kalman prediction, to avoid splitting a single vehicle into multiple tracks.
- ROI and border gating to ensure only events occurring fully inside the analysis zone are counted.

Ground-truth counts are obtained by manual frame-by-frame inspection of the same videos, using the identical RoI and count-line definitions. The predicted events (from YOLO-SORT) are compared against these manual counts to derive true positives, false positives, and false negatives at the event level. Since the study also examines short-term temporal stability, precision and recall are computed at a per-minute resolution across each one-hour session.

E. Environmental and Geometric Configurations

This study evaluates how environmental visibility, illumination, and camera placement jointly influence performance. All recordings were acquired using fixed camera parameters (exposure, gain, ISO, and focal length) to avoid sensor-induced variation across conditions. Locking these parameters ensures that environmental effects (fog, rain, night) influence only the scene, not the sensor response. All recordings were collected under natural conditions at the study site, without artificial augmentation or laboratory staging. The environmental dimension includes three daytime weather states—clear, rainy, and foggy—and two nighttime states—clear and rainy. Fog was present only during daytime hours and could not be replicated at night, resulting in an asymmetry that reflects real-world availability rather than experimental omission. Each condition was recorded for one continuous hour, allowing the evaluation of both overall performance and short-term temporal variability.

Weather labels were determined through direct on-site observation and verified retrospectively by inspecting the recorded footage for visibility range, contrast reduction, road-surface reflections, and precipitation intensity. Clear conditions refer to sessions with no precipitation and stable visibility. Rain covers both light and moderate rainfall observed at the site, with the wet pavement clearly visible in the recordings. Fog refers to sessions with reduced contrast, softened edges, and intermittent partial occlusions caused by suspended water droplets, which mainly affected early segments of the recording as fog density varied throughout the hour. Night sessions were recorded under fixed roadway illumination. The camera's exposure settings remained constant across night recordings to ensure comparability.

To examine geometric effects, recordings were captured from three distinct camera viewpoints positioned along a pedestrian bridge overlooking the highway. These viewpoints represent commonly encountered CCTV placements in roadway monitoring. Angle 1 corresponds to a near-frontal, slightly elevated perspective that provides a clear view of approaching traffic with minimal lateral distortion. Angle 2 is a moderately shifted oblique placement that introduces perspective compression, altering apparent vehicle width and orientation. Angle 3 is positioned farther along the bridge, yielding a sharper lateral view with increased foreshortening and more pronounced perspective effects. These three placements create controlled geometric diversity while maintaining a consistent elevation and distance from the traffic stream. The three camera angles were selected to reflect common roadside deployments: a central angle representing ideal mounting, a moderately shifted angle reflecting realistic constraints, and a more oblique angle corresponding to non-ideal roadside installations. This variation enables a practical assessment of YOLO-SORT's tolerance to viewpoint deviations commonly encountered in real-world systems.

All recordings were captured with the camera mounted rigidly to prevent changes in roll, pitch, or yaw during acquisition. The viewpoints differ only in their lateral placement, not in height or camera orientation. Because perspective and magnification vary across angles, each viewpoint received an independent Pixels-Per-Meter (PPM) calibration, derived from the length of road markings situated near the count line. This per-view calibration ensures that bounding-box dimensions and track displacements are interpreted in a consistent metric scale across all angles, mitigating geometric distortions that would otherwise confound comparisons.

A partial factorial design was used for environmental and geometric conditions. For daytime weather states, recordings were collected at all three viewpoints, establishing a complete three-angle×three-weather configuration. For nighttime conditions, recordings were available for clear and rainy weather, also across all three angles. This results in a structured set of condition-angle combinations that supports comparison across environments while preserving interpretability. The design reflects realistic constraints: environmental conditions occurred naturally, and nighttime fog—frequently unpredictable—did not occur during the study period.

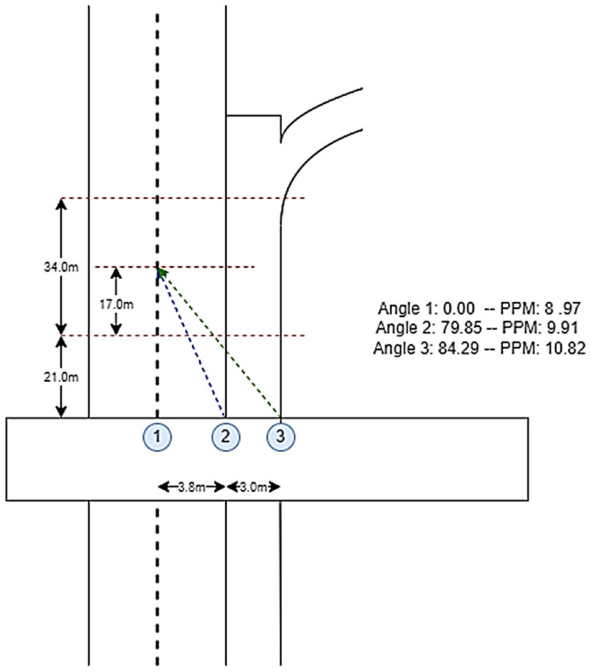

Fig. 2. Camera-viewpoint geometry, illustrating the three lateral positions on the bridge, their viewing angles toward the roadway, and the PPM calibration derived from road-marking measurements.

Figure 2 depicts the camera placement at three distinct viewpoints—Angle 1: 0.00° (PPM = 8.97), Angle 2: 79.85° (PPM = 9.91), and Angle 3: 84.29° (PPM = 10.82)—used to acquire video for the YOLO-SORT analysis. The elevated, laterally spaced mounting positions emulate common traffic-surveillance geometries. The higher PPM values observed for Angles 2 and 3 arise from field-of-view compression at oblique orientations, where the same pixel grid covers a smaller physical footprint, improving per-pixel object distinguishability. PPM was estimated from on-site measurements and verified using an independent check segment; discrepancies beyond a small tolerance triggered re-

measurement. The final PPM values were then used to normalize bounding-box dimensions and track displacements, ensuring a consistent metric scale across viewpoints and eliminating perspective-induced size bias. This normalization also stabilizes SORT's motion predictions by allowing all trajectories to be interpreted in a uniform physical space.

Traffic flow at the site followed its natural daily cyclical pattern. Morning recordings, used for daytime sessions, coincide with higher vehicle density, whereas evening recordings correspond to lower density. These variations were not controlled experimentally but are relevant for interpreting performance dynamics, as traffic density influences the degree of occlusion and the complexity of track association. Nevertheless, each condition-angle pair spans a full hour, allowing performance metrics to be evaluated over homogeneous and heterogeneous traffic segments within the same session.

Overall, this environmental and geometric configuration provides a controlled yet realistic basis for assessing the sensitivity of YOLO-SORT to visibility, illumination, and camera-viewpoint changes. The configuration ensures that both environmental and geometric influences can be examined independently and jointly, without relying on synthetic weather effects or simulated camera geometries.

III. EXPERIMENTATION AND ANALYSIS

Video data were collected above a highway segment in Győr, Hungary, using a mobile phone camera (iPhone 13) as a practical surrogate for fixed CCTV to evaluate the algorithms rather than the hardware. The recordings were captured from a pedestrian bridge above the carriageway, providing an unobstructed elevated viewpoint comparable to typical roadside surveillance installations. The device was rigidly mounted and operated with locked focus and exposure, fixed focal length, and constant 1080p/30fps. These settings closely match those of widely deployed traffic-surveillance units such as the Hikvision DS-2CD2T47G2-L [39], which also operates at fixed 1080p/30fps with a fixed-focus, fixed-focal-length lens. Consequently, the resulting imagery is representative of a standard CCTV stream for detection and tracking. Although a smartphone lacks 24/7 ruggedization, this limitation is irrelevant to the study's objective, which focuses on algorithmic performance under controlled, CCTV-like imagery rather than hardware evaluation.

All recordings were conducted on workdays. The daytime sessions were recorded during the 07:00–08:00 peak hour, while the nighttime sessions were recorded at 19:00–20:00. The environmental conditions included clear, rainy, and foggy daytime scenarios and clear and rainy nighttime scenarios, with one hour per condition and per angle to obtain stable minute-level profiles. The study focuses exclusively on vehicle detection and tracking using the YOLO-SORT pipeline; therefore, no human subjects were involved.

Performance is reported per condition and viewpoint at the event level, based on count-line crossings. Ground truth is obtained by manually counting vehicles in the recorded videos using the same ROI and virtual count line as the model. The YOLO-SORT pipeline produces predicted crossing events, compared against the manual counts to derive true positives, false positives, and false negatives at the event level.

- Recall quantifies the model's ability to retrieve relevant instances, computed as the proportion of true positives among all actual positives (TP/[TP+FN]); higher recall corresponds to fewer missed vehicles.
- Precision measures the reliability of positive predictions, computed as the proportion of true positives among all predicted positives (TP/[TP+FP]); higher precision corresponds to fewer spurious detections.
- F1-score is the harmonic mean of precision and recall, providing a single summary metric when both false positives and false negatives are important.

Sensitivity analyses were conducted on precision, recall, and F1-score.

- Within-angle comparisons: Condition effects were evaluated using the Wilcoxon signed-rank test with Holm's step-down correction for multiple comparisons [40, 41]. Effect sizes were quantified using the Wilcoxon r statistic. Wilcoxon was selected because per-minute F1 values were non-normally distributed and may exhibit temporal dependence, making nonparametric testing the appropriate choice.
- Across-angle comparisons: Condition-aggregated F1-scores were compared pairwise across the three viewpoints using the same Wilcoxon–Holm procedure and effect-size measure.

Experiments were run on a laptop workstation with Windows 11 64-bit, Intel Core i7-14650HX (24 logical CPUs, base ~2.2 GHz), 16 GB RAM. GPU acceleration was enabled at the NVIDIA GeForce RTX 4050 Laptop GPU (≈5.9 GB VRAM, ~13.9 GB total graphics memory, DirectX 12). All videos were processed locally on this machine without cloud resources.

To evaluate real-time feasibility, the YOLOv8s+SORT pipeline was benchmarked on the same workstation for all experiments. End-to-end processing was performed at 640×640 resolution with FP16 precision and a batch size of 1, matching the settings applied during evaluation. Across all conditions, the pipeline achieved a mean throughput of 55.1 fps (range: 52–58 fps), corresponding to an average per-frame latency of approximately 18 ms. GPU utilization during processing ranged between 42% and 55%, indicating that the system operates well within the computational capacity of the RTX 4050 Laptop GPU. These measurements confirm that the pipeline can process 30 FPS CCTV-style video streams in real time with substantial performance headroom.

A. Angle 1: Condition-Wise Performance Comparison

Condition-wise performance for Angle 1 in Table I reports ten paired contrasts; nine are significant after Holm correction ($p \leq 0.016$), with only Clear Night vs. Rainy Night non-significant ($p = 0.175$). Absolute Wilcoxon r spans moderate to very large effects, peaking for Clear Night vs. Foggy Day and for Foggy Day vs. Rainy Night ($|r| \approx 0.85$), and is the smallest

among the significant contrasts for Clear Day vs. Rainy Day ($|r| \approx 0.31$).

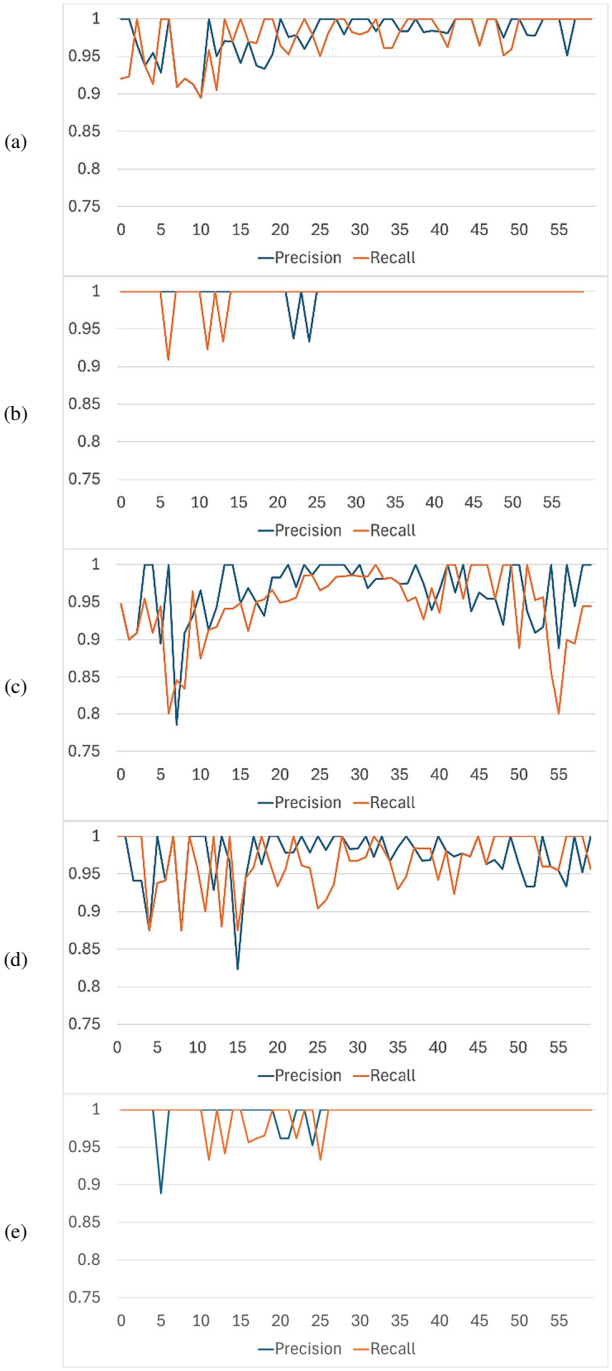

Fig. 3. Per-minute Precision and Recall for Angle 1 (a) Day Clear; (b) Night Clear; (c) Day Fog; (d) Day Rain; (e) Night Rain.

Hodges-Lehmann (HL) (first-second) median shifts range from 0.000 to about 0.033, with the largest magnitude for Clear Night vs. Foggy Day (≈+0.033) and the smallest non-zero for Foggy Day vs. Rainy Day (≈−0.009). These results indicate a clear ordering: night conditions (clear or rainy) outperform daytime weather, fog depresses F1 the most, and rain day produces a modest yet reliable drop relative to Clear Day; Clear vs. Rainy Night is statistically indistinguishable.

TABLE I. ANGLE 1 PAIRWISE CONDITION COMPARISONS ON F1 (WILCOXON SIGNED-RANK), REPORTING R-EFFECT SIZE, HODGES–LEHMANN MEDIAN SHIFT, AND HOLM-ADJUSTED P-VALUES

Condition - Condition	Effect size (r)	HL shift	Significance
Clear Day – Clear Night	-0.715	-0.0138	<0.001
Clear Day – Foggy Day	-0.559	0.0176	<0.001
Clear Day – Rainy Day	-0.314	0.0145	0.016
Clear Day – Rainy Night	-0.743	-0.0102	<0.001
Clear Night – Foggy Day	-0.853	0.0334	<0.001
Clear Night – Rainy Day	-0.846	0.0242	<0.001
Clear Night – Rainy Night	-0.408	0.0	0.175
Foggy Day – Rainy Day	-0.365	-0.0088	0.005
Foggy Day – Rainy Night	-0.851	-0.0303	<0.001
Rainy Day – Rainy Night	-0.827	-0.0222	<0.001

At Angle 1 (Figure 3), precision-recall traces are uniformly high with condition-specific quirks. Clear Night (b) is the most stable: recall stays pinned at 1, and precision only shows brief, shallow dips (~0.95–0.98) in the first third of the hour. Clear Day (a) is similarly strong—both curves hover near 0.97–1 with minor oscillations. Rainy Day (d) preserves high recall after minute ~15 but exhibits a couple of short precision dents (~0.85–0.9) during early showers, then recovers to the ~0.97 band. Foggy Day (c) is the most variable: an early precision trough (~0.78) and softened recall (~0.90–0.95) precede a steady climb toward ~0.97 as visibility improves. Rainy Night (e) remains robust overall; recall is essentially flat at 1, and precision is near-ceiling with only brief dips around minutes ~15–25, indicating that nighttime illumination at this site offsets most wet-road effects.

B. Angle 2: Condition-Wise Performance Comparison

Table II comprises ten paired contrasts; eight are significant ($p < 0.01$), while two are not (Clear Day–Rainy Day, $p = 0.495$; Clear Night–Rainy Night, $p = 0.310$). Absolute effect sizes span $|r| = 0.089$–0.843, with the strongest contrasts for Foggy Day–Rainy Night and Rainy Day–Rainy Night ($|r| = 0.843$) and the weakest for Clear Day–Rainy Day ($|r| = 0.089$).

TABLE II. ANGLE 2 PAIRWISE CONDITION COMPARISONS ON F1 (WILCOXON SIGNED-RANK), REPORTING r-EFFECT SIZE, HODGES–LEHMANN MEDIAN SHIFT, AND HOLM-ADJUSTED p-VALUES

Condition – Condition	Effect size (r)	HL shift	Significance
Clear Day – Clear Night	-0.679	-0.018	<0.001
Clear Day – Foggy Day	-0.546	0.0186	<0.001
Clear Day – Rainy Day	-0.089	0.0049	0.495
Clear Day – Rainy Night	-0.81	-0.0187	<0.001
Clear Night – Foggy Day	-0.798	0.0326	<0.001
Clear Night – Rainy Day	-0.738	0.022	<0.001
Clear Night – Rainy Night	-0.306	0.0	0.310
Foggy Day – Rainy Day	-0.402	-0.0085	0.002
Foggy Day – Rainy Night	-0.843	-0.0326	<0.001
Rainy Day – Rainy Night	-0.843	-0.0217	<0.001

HL shift magnitudes range from 0.0000 (Clear Night–Rainy Night) to 0.0326, attained by both Clear Night–Foggy Day (+0.0326) and Foggy Day–Rainy Night (-0.0326). Signs

follow the ordering reported in the table. After Holm correction, the significance pattern remains unchanged: only Clear Day–Rainy Day and Clear Night–Rainy Night are non-significant; all other contrasts remain significant.

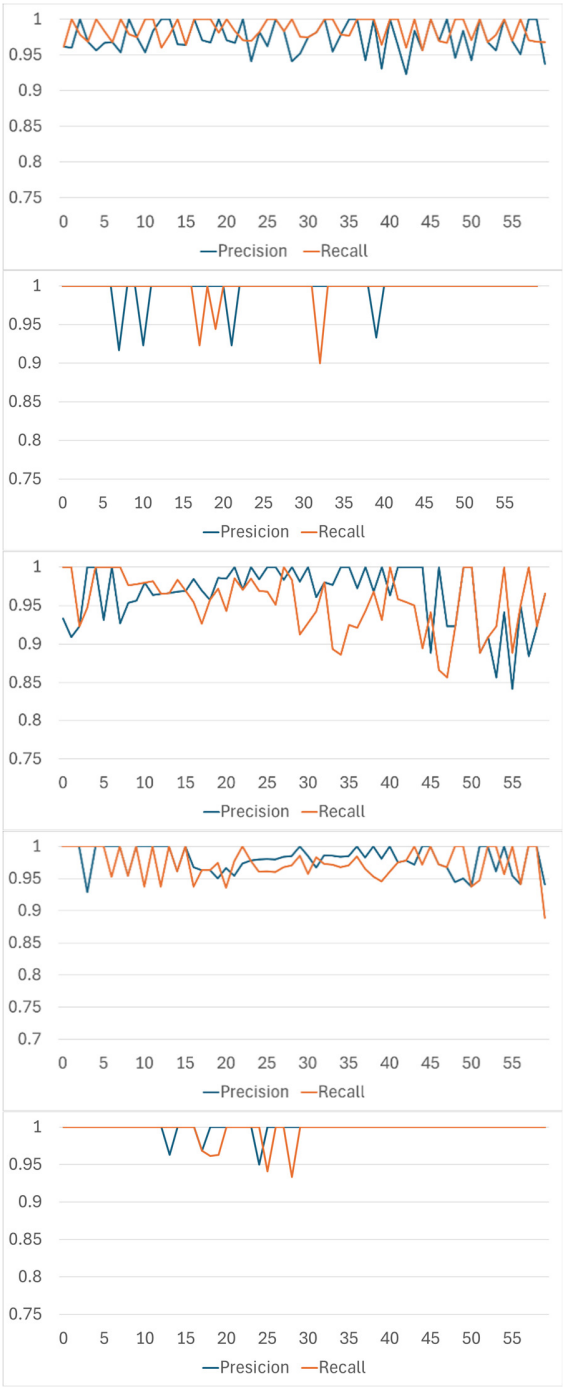

Fig. 4. Per-minute precision and recall for Angle 2 (a) Day Clear; (b) Night Clear; (c) Day Fog; (d) Day Rain; (e) Night Rain.

At Angle 2, per-minute precision–recall trends remain high across conditions but with distinct signatures, as illustrated in Figure 4. Day Clear (a) stays tight in the 0.95–1.00 band with only small oscillations. Night Clear (b) shows ceiling-level recall (≈ 1.0) with a few isolated precision dips. Day Fog (c) is the most volatile: after an early recovery, both curves fluctuate through the mid/late interval, with several brief precision troughs. Day Rain (d) holds near 0.96-0.99 for most of the hour, then softens in the final minutes, where both precision and recall decline. Night Rain (e) remains the most uniform: recall sits at the ceiling and precision varies narrowly around it, indicating that consistent lighting and lower night volumes mitigate rain effects at this viewpoint. Overall, Angle 2 maintains high accuracy, with variability concentrated in fog and at the tail of the rainy sequence.

C. Angle 3: Condition-Wise Performance Comparison

Table III reports ten paired contrasts, where eight are significant (all $p < 0.001$) and two are not (Clear Night–Rainy Night, $p = 0.052$; Foggy Day–Rainy Day, $p = 0.232$). Absolute effect sizes span $|r| = 0.157$–0.845, with the largest magnitude for Rainy Day–Rainy Night ($r = -0.845$) and the smallest for Foggy Day–Rainy Day ($r = -0.157$). HL shifts range from 0.0000 to 0.0303, peaking for Foggy Day–Rainy Night (HL = –0.0303) and minimal for Clear Night–Rainy Night (HL = 0.0000). Signs follow the "first-second" ordering shown in the table. After Holm correction, this pattern is unchanged: the two noted contrasts remain non-significant while the others remain significant.

TABLE III. ANGLE 3 PAIRWISE CONDITION COMPARISONS ON F1 (WILCOXON SIGNED-RANK), REPORTING R-EFFECT SIZE, HODGES–LEHMANN MEDIAN SHIFT, AND HOLM-ADJUSTED P-VALUES

Condition - Condition	Effect size (r)	HL shift	Significance
Clear Day – Clear Night	-0.49	-0.0174	<0.001
Clear Day – Foggy Day	-0.643	0.02	<0.001
Clear Day – Rainy Day	-0.595	0.0128	<0.001
Clear Day – Rainy Night	-0.74	-0.0177	<0.001
Clear Night – Foggy Day	-0.645	0.0301	<0.001
Clear Night – Rainy Day	-0.618	0.0275	<0.001
Clear Night – Rainy Night	-0.485	0.0	0.052
Foggy Day – Rainy Day	-0.157	-0.0042	0.232
Foggy Day – Rainy Night	-0.838	-0.0303	<0.001
Rainy Day – Rainy Night	-0.845	-0.0286	<0.001

As seen in Figure 5, precision-recall traces are uniformly high but exhibit condition-specific patterns. Day Clear (a) stays tightly clustered near 0.97–1.00 with only brief, shallow dips. Night Clear (b) holds recall at the ceiling and shows a few mid-segment precision softening that quickly recover. Day Fog (c) is the most variable: an early precision trough is followed by recovery to the high-0.9s, while recall oscillates just below 1.00—consistent with reduced contrast and intermittent occlusions. Day Rain (d) remains strong overall, though precision shows several early dips before stabilizing; recall trends upward through the middle of the hour and eases slightly near the end. Night Rain (e) is the most saturated of the rainy cases, with both metrics pinned close to 1.00 apart from a few short notches.

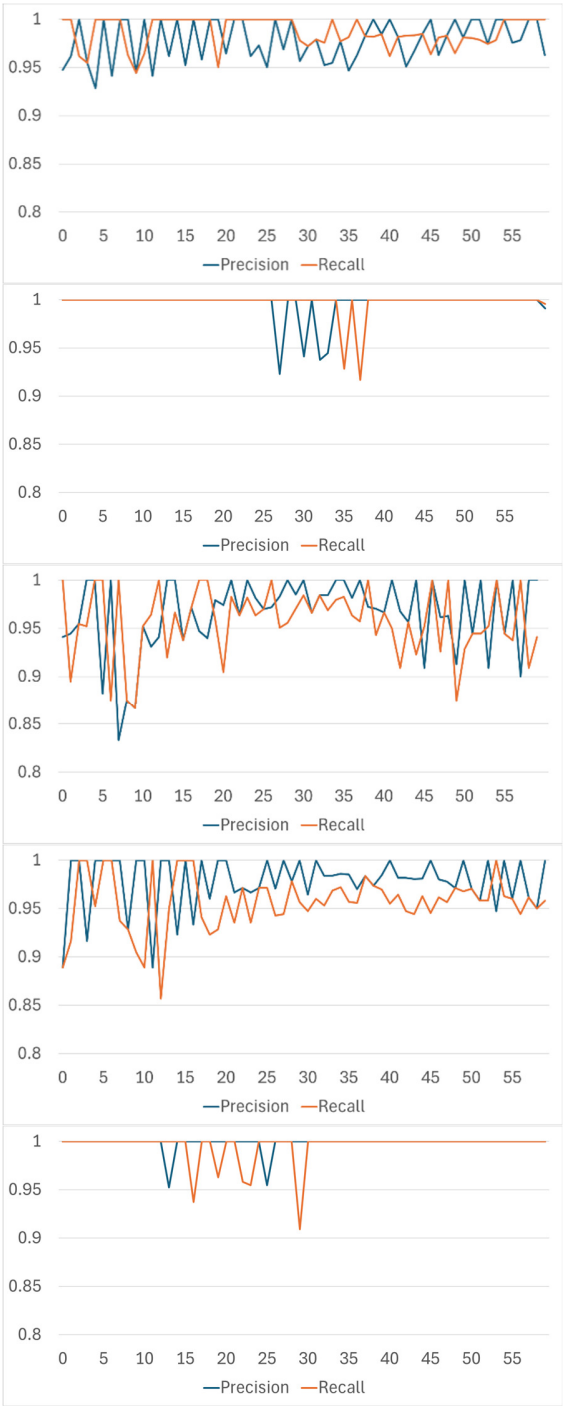

Fig. 5. Per-minute Precision and Recall for Angle 3 (a) Day – Clear; (b) Night – Clear; (c) Day – Fog; (d) Day – Rain; (e) Night - Rain.

D. Consolidated Condition-Wise Performance Summary

To complement the angle-specific analyses, Table IV presents a consolidated summary of the mean precision, recall, and F1-score for each environmental condition aggregated across all three viewpoints.

TABLE IV. CONSOLIDATED PERFORMANCE SUMMARY ACROSS CONDITIONS, REPORTING MEAN PRECISION, RECALL, AND F1-SCORE AGGREGATED OVER THE THREE VIEWPOINTS

Angle	Condition	Mean Precision	Mean Recall	Mean F1
Angle 1	Clear Day	97.7%	97.6%	97.7%
	Foggy Day	96.4%	94.6%	95.4%
	Rainy Day	97.3%	96.6%	96.9%
	Clear Night	99.8%	99.6%	99.7%
	Rainy Night	99.6%	99.4%	99.5%
Angle 2	Clear Day	97.4%	98.5%	97.9%
	Foggy Day	96.4%	95.4%	95.9%
	Rainy Day	98.0%	97.4%	97.7%
	Clear Night	99.5%	99.6%	99.5%
	Rainy Night	99.8%	99.6%	99.7%
Angle 3	Clear Day	97.8%	98.7%	98.2%
	Foggy Day	96.5%	95.7%	96.1%
	Rainy Day	97.8%	95.7%	96.7%
	Clear Night	99.6%	99.7%	98.8%
	Rainy Night	99.8%	99.5%	99.7%

E. Across-Angle Performance (Aggregated over Conditions)

Across-angle performance, shown in Table V, shows no statistically significant differences in per-sample mean F1 (all $p \geq 0.205$ after Holm). Effect sizes are negligible ($|r| \leq 0.163$), and HL median shifts are small in magnitude (≤ 0.0029): Angle 1 vs. Angle 2 (HL ≈ -0.0003), Angle 1 vs. Angle 3 (HL $\approx +0.0008$), and Angle 2 vs. Angle 3 (HL $\approx +0.0029$). Overall, the three viewpoints perform comparably within sampling variability.

TABLE V. ACROSS-ANGLE PAIRWISE COMPARISONS ON PER-SAMPLE MEAN F1 (WILCOXON SIGNED-RANK): R-EFFECT SIZE, HODGES–LEHMANN MEDIAN SHIFT, AND HOLM-ADJUSTED P-VALUES

Angle – Angle	Effect size (r)	HL shift	Significance
Angle 1 – Angle 2	-0.079	-0.0003	0.541
Angle 1 – Angle 3	-0.017	0.0008	0.895
Angle 2 – Angle 3	-0.163	0.0029	0.205

IV. DISCUSSIONS

The results indicate that the YOLO-SORT pipeline sustains high performance across all tested scenarios, but the influence of weather, illumination, traffic density, and camera geometry is not uniform. In the following subsections, these effects are examined in detail by relating condition-wise and angle-wise statistical contrasts to the observed precision-recall dynamics to clarify when the system remains stable and when it becomes more sensitive to visibility and occlusion constraints.

A. Conditions Effects at Angle 1

Condition-wise behavior at Angle 1 shows a clear hierarchy across environments, with nighttime recordings yielding the most stable performance and fog producing the greatest degradation. Both Clear and Rainy Night traces maintain recall at or near unity and exhibit only minor precision fluctuations, reflecting the favorable combination of consistent street lighting at the site and substantially lower vehicle density during evening hours. Clear Day follows closely, though its slight oscillations coincide with the periods of highest traffic volume within the session, where increased vehicle interactions

introduce brief occlusions that momentarily perturb precision and recall. Rainy Day exhibits a similar pattern: short precision dents occur during the early rainfall onset and in intervals of higher flow, after which both metrics converge toward the upper daytime range. Foggy Day remains the only condition with pronounced early instability; reduced visibility suppresses both precision and recall before a gradual recovery as conditions improve. This temporal pattern aligns with the large effect sizes and median shifts associated with fog in the statistical comparisons. Overall, Angle 1 presents a consistent ordering—Night > Clear Day > Rainy Day > Fog—driven primarily by illumination quality, traffic density, and the visibility constraints characteristic of this frontal viewpoint.

B. Conditions Effects at Angle 2

Angle 2 maintains strong overall accuracy across all environments, but its oblique geometry introduces sharper distinctions between stable and variable conditions. Both Night Clear and Night Rain remain highly consistent, with recall pinned at the ceiling and precision varying within a narrow band; this behavior reflects the steady artificial lighting at the site and markedly lower nighttime traffic volumes, which reduce occlusion and simplify the spatial arrangement of vehicles from this steep lateral perspective. Clear Day also stays near the top performance range, though with mild oscillations characteristic of daytime traffic flow. In contrast, Foggy Day shows the most volatile pattern: after an early recovery, both precision and recall fluctuate throughout the mid and late intervals, consistent with the large effect sizes and ±0.0326 HL shifts observed in the statistical comparisons. Day Rain exhibits a different form of instability—its performance remains high for most of the session but softens sharply in the final minutes, matching the weaker, non-significant contrast with Clear Day. The two non-significant comparisons (Clear Day–Rainy Day and Clear Night–Rainy Night) indicate that rain alone does not materially affect this viewpoint unless paired with visibility loss, while fog produces the most consistent degradation. Overall, Angle 2 preserves high precision-recall levels across conditions, with variability concentrated in fog and in the terminal segment of the rainy daytime sequence.

C. Conditions Effects at Angle 3

Angle 3 maintains high accuracy across conditions, yet its steep oblique geometry amplifies the influence of visibility and traffic interactions more than at the other viewpoints. Clear Day remains tightly clustered near the upper performance band, with only shallow, short-lived fluctuations. Clear Night shows a similarly stable pattern: recall stays at the ceiling and precision exhibits only brief softening in the mid-segment, consistent with steady artificial lighting and lower nighttime traffic volumes that reduce occlusions along this lateral view. Foggy Day is distinctly more volatile; reduced contrast and intermittent overlapping trajectories create an early precision trough and sustained recall oscillations, which align with the significant HL shifts and effect sizes reported for fog-related contrasts. Rainy Day follows a mixed trajectory—precision experiences several early dips before stabilizing, while recall strengthens through the mid-period and eases slightly toward the end—matching the moderate statistical separations from clear conditions and the non-significant comparison with Foggy Day. Night Rain is the most saturated of the rainy scenarios, with both metrics remaining close to 1.00 apart from a few isolated notches, reflecting the stabilizing effect of nighttime illumination even under precipitation. Overall, Angle 3 shows consistently high performance but with greater sensitivity to fog and early-sequence rain effects, driven by the compounded visibility and occlusion challenges inherent to this sharper oblique perspective.

D. Viewpoints Comparison across Conditions

Across-angle comparisons indicate that the three viewpoints perform equivalently once results are aggregated over environmental conditions. The Wilcoxon-Holm tests show no statistically significant differences among the angles, and all effect sizes remain negligible, with HL shifts well below one thousandth of an F1 point. This pattern suggests that, despite their geometric differences—Angle 1's frontal view, Angle 2's steep obliqueness, and Angle 3's intermediate lateral perspective—the YOLO–SORT pipeline maintains stable behavior when evaluated over mixed conditions. Small fluctuations observed at the angle level are therefore attributable to condition-specific interactions rather than structural differences in viewpoint geometry. In practice, this means that for the range of angles studied, the detector-tracker system generalizes robustly across perspectives, and performance variations arise chiefly from environmental factors such as lighting, visibility, and momentary traffic density rather than from the view angle itself.

E. Patterns in Precision-Recall Dynamics

Across all viewpoints and conditions, the precision-recall traces exhibit consistent high-level stability yet reveal characteristic patterns that reflect how environmental visibility, illumination, and traffic density interact with the detection-tracking pipeline. Nighttime recordings—whether clear or rainy—show the most uniform behavior: recall remains at or near the ceiling, and precision fluctuates only within a narrow band. These dynamics likely reflect two favorable properties of the recording site at night: consistent street lighting that preserves object contrast and lower traffic volumes that reduce occlusions and simplify vehicle trajectories. Because daytime and nighttime sessions were recorded under naturally different flow conditions, the present data do not fully separate the influence of illumination from that of traffic density.

Daytime sequences display broader oscillations, and in several sessions, the more pronounced dips in precision and recall coincide with the periods of highest vehicle density. Increased overlap among vehicles at these times produces transient association challenges for SORT and temporarily weakens detector separation of tightly clustered objects. Rain amplifies these effects during its onset, creating short-lived precision troughs before the system stabilizes as rain intensity or flow complexity subsides. Fog introduces the most persistent degradation. Reduced contrast and intermittent partial occlusions lead to extended stretches in which precision and recall oscillate, even after initial recovery, matching the large effect sizes associated with fog-related contrasts.

Importantly, the consistency of these temporal patterns across angles suggests that short-term fluctuations are governed more by environmental and traffic factors than by viewpoint geometry. Although sharper oblique angles (e.g., Angle 3) magnify the impact of visibility loss or early rain sequences, aggregated performance differences between angles remain negligible. Overall, the precision-recall dynamics indicate a system that is resilient under stable illumination and moderate flow but more sensitive to visibility reductions and peak-traffic occlusion events, irrespective of camera placement.

F. Error Mechanisms Observed Across Conditions

Qualitative inspection of the recordings further clarifies the mechanisms behind the quantitative patterns. In fog, softened edges and reduced contrast frequently caused YOLOv8s to miss distant or partially occluded vehicles, especially during the early minutes when fog density fluctuated. In addition to outright misses, fog occasionally produced bounding-box shrinkage, where vehicles were detected but with boxes too small to satisfy the minimum-area gate, resulting in suppressed detections. Rain introduced a different set of failure modes: reflections from wet pavement and headlight glare occasionally produced false positives near the lower RoI boundary, and during heavier rainfall, streak-like artifacts sometimes triggered transient spurious detections. Under both rain and fog, vehicles with low-reflectance colors (e.g., dark gray or black) were more prone to intermittent detection gaps.

Traffic density also shaped error behavior. In denser daytime segments, extended overlaps between vehicles led to brief ID switches in SORT, particularly when two vehicles of similar size crossed paths or when a larger vehicle temporarily occluded a smaller one. In a few cases, SORT prematurely terminated a track when a vehicle slowed abruptly near the count line, causing a short-lived fragmentation before reinitialization. At sharper oblique angles, partial side-view visibility occasionally caused YOLO to misclassify motorcycles as small cars or to momentarily lose track of narrow-profile vehicles. Although these events were infrequent and rarely affected the final count due to the trajectory-based crossing logic, they illustrate the specific visual and geometric conditions under which the pipeline becomes more sensitive.

Collectively, these representative failure modes align closely with the precision-recall dynamics described above and help explain why visibility loss, reflective surfaces, and peak-flow occlusions exert the strongest influence on system stability.

G. Tracker Alternatives and Expected Differences

Although SORT was chosen to isolate the influence of environmental and geometric factors without introducing appearance-based cues, it is informative to consider how alternative trackers might behave under the same conditions. DeepSORT, for example, augments IoU-based association with deep appearance embeddings, which could improve continuity in fog or rain by re-identifying vehicles after short detection gaps. This may reduce the brief ID switches observed during dense daytime traffic or partial occlusions. ByteTrack, in contrast, leverages low-confidence detections during association rather than discarding them, which could be advantageous in low-contrast environments where YOLOv8s occasionally produces borderline-score boxes. Under fog, where bounding-box shrinkage and intermittent misses were observed, ByteTrack's confidence-aware matching might preserve more stable trajectories.

However, these benefits come with trade-offs. Appearance-based trackers such as DeepSORT may be more sensitive to illumination changes, headlight glare, or color distortions—conditions that were common in the rainy and nighttime sequences. ByteTrack's reliance on low-confidence detections may also increase susceptibility to rain-induced reflections or transient artefacts. While evaluating these alternatives lies outside the scope of the present study, their differing mechanisms suggest that some of the observed failure modes—particularly ID switches in dense traffic and short detection gaps in fog—might be mitigated or altered under trackers that incorporate appearance or confidence-aware cues.

H. Comparison with Similar Studies

The results of this study are broadly consistent with prior YOLO-based traffic-monitoring studies showing that real-time vehicle detection and tracking pipelines can achieve strong practical performance under roadway surveillance conditions. For example, in [5], precision, recall, and mAP values of 91.25%, 93.52%, and 92.18%, respectively, were reported for a YOLOv5+DeepSORT traffic-monitoring scheme evaluated on BDD100K and PASCAL-based experiments, while a more recent YOLOv9s+DeepSORT system [42] reported mAP@0.5 of 91.4%, mAP@0.5:0.95 of 82.7%, average speed of 12.9 fps, and vehicle-counting error below 4% on urban traffic videos. In contrast, studies using uncontrolled open-webcam imagery have reported substantially lower out-of-the-box performance, with the best tested YOLOv8 variant reaching an F1-score of 0.75, highlighting the effect of image quality, occlusion, and environmental variability on monitoring accuracy [43]. These comparisons should be interpreted cautiously, since the underlying datasets, camera viewpoints, detector sizes, tracker choices, and evaluation metrics differ across studies. Within this broader literature, the contribution of this work lies less in proposing a new detector-tracker architecture than in showing, under a fixed YOLOv8s+SORT configuration, that fog is the strongest degradation factor, daytime rain causes milder short-term instability, and the tested viewpoint changes are less influential overall than environmental conditions.

I. Limitations of the Study

Several factors constrain the generalizability of the findings. The study site is a two-lane highway segment with relatively light and orderly traffic flow, resulting in fewer persistent occlusions than those typically encountered on wider or congested urban roads. This setting facilitates controlled analysis but does not fully represent high-density scenes where extended overlaps and multi-lane interactions challenge both detection and tracking. Nighttime traffic at the site was also substantially lower, which likely contributed to the uniformly strong nocturnal performance by reducing object overlap and simplifying association for SORT. In addition, the camera was mounted above the carriageway, providing an unobstructed line of sight across all three viewpoints. Such elevated perspectives minimize parallax, headlight artifacts, and roadside

obstructions relative to ground-level or offset installations, potentially underestimating the sensitivity of the pipeline to viewpoint degradation in less favorable placements.

Only one hour of data was available for each environmental condition, which constrains the statistical strength of the comparisons. This limitation is most evident in the fog session, where visibility fluctuated within the recording period. Such temporal variation reduces the stability with which fog-related degradation can be characterized. Longer recordings under comparable fog intensity would provide a more reliable basis for assessing performance. Also, the use of 60-minute-level observations per condition provides a practical basis for nonparametric testing, but it also constrains statistical power. This is evident in borderline contrasts such as the Clear Night vs. Rainy Night comparison at Angle 3 ($p = 0.052$), where the small observed difference may reflect either a genuinely negligible effect or insufficient sensitivity due to the one-hour sampling window. Longer or repeated recordings would strengthen the ability to detect subtle condition differences.

Ground-truth vehicle counts were obtained through manual frame-by-frame annotation performed by a single observer using the same RoI and count-line definitions applied to the model. While this ensured internal consistency, inter-annotator agreement was not measured, and the use of a single annotator introduces the possibility of human error. This limitation is particularly relevant under fog and nighttime conditions, where reduced visibility can make object boundaries less distinct. The results should therefore be interpreted with the understanding that some degree of annotation uncertainty may be present.

Environmental diversity is another constraint. Although the dataset includes clear, rainy, and foggy daytime conditions, fog did not occur during nighttime sessions, leaving this challenging combination untested. Therefore, the analysis cannot characterize how the system behaves under low-contrast, artificially lit fog environments, which often exacerbate haloing, scattering, and color distortion.

Overall, these limitations reflect a controlled recording environment that supports robust within-site comparisons but restricts the generalizability of the findings to denser or more heterogeneous deployments.

J. Future Work

A natural extension involves collecting data on wider multi-lane urban roads with higher vehicle density, where complex occlusion patterns, heterogeneous driving behavior, and diverse vehicle types may reveal failure modes not observed in this study. Incorporating nighttime fog, when available, would provide a more complete assessment of low-visibility conditions under artificial illumination. Additional camera placements—such as roadside, lower-elevation, or offset viewpoints—would help evaluate performance under more constrained geometries and stronger parallax effects. From a system perspective, exploring multiple detector-tracker combinations, adaptive thresholding strategies, and temporally aware methods (e.g., DeepSORT, ByteTrack) could clarify how architectural choices interact with environmental and geometric variability. Finally, assessing robustness under varied sensors, compression levels, and network bandwidth constraints would support more practical deployment recommendations across different infrastructure settings. In addition, a targeted sensitivity analysis of the detector confidence threshold would be useful for quantifying how the balance between false positives, missed detections, and track continuity changes under different environmental conditions. Finally, a volume-normalized or density-stratified analysis would be useful to separate the influence of traffic flow from that of nighttime illumination more clearly when interpreting the stronger nighttime performance observed in this study.

V. CONCLUSIONS

This study evaluated a compact real-time YOLO-SORT pipeline across three viewpoints and a range of environmental conditions using controlled, PPM-calibrated recordings from a highway segment in Győr, Hungary. By keeping the model parameters, thresholds, and geometric calibration fixed, the analysis isolated how visibility, illumination, and traffic density shape precision-recall behavior and event-level F1 performance. The within-angle comparisons revealed consistent condition hierarchies: nighttime recordings produced the most uniform performance, likely reflecting the combined influence of stable artificial lighting and lower vehicle density at the study site; daytime rain introduced short-lived fluctuations; and fog generated the largest and most persistent degradation. Despite the differing geometries of the three viewpoints, cross-angle comparisons did not show statistically significant differences once aggregated over conditions, indicating that viewpoint alone contributed little to overall variation relative to environmental and traffic factors.

Precision-recall dynamics further demonstrated that transient instability is closely tied to periods of increased vehicle density and reduced contrast, rather than to the underlying pipeline architecture. This suggests that lightweight detection-tracking systems can maintain reliable performance under stable illumination and moderate flow, but are more sensitive to visibility loss and peak-traffic occlusion events.

Although the controlled two-lane setting provided a clear testbed for isolating environmental effects, it also limits generalizability to denser, multilane, or more heterogeneous traffic environments. Future work should therefore expand the data domain toward more complex occlusion patterns, additional camera placements, and broader weather conditions, including nighttime fog, as well as alternative detector-tracker configurations and sensing hardware. Such extensions would help characterize the boundaries of real-time models under diverse deployment scenarios and support more robust guidance for infrastructure-scale traffic monitoring.

DECLARATION OF COMPETING INTERESTS

Not applicable to this work.

ACKNOWLEDGMENTS

Not applicable to this work.

DATA AVAILABILITY

The data supporting the findings of this study are available from the corresponding author upon reasonable request.

REFERENCES

[1] G. Hoxha, A. Fandaj, and X. Bajrami, "Quality of Automatic Traffic Volume Counting by Cameras and Impact on the Qualitative Indicators of Traffic," *Infrastructures*, vol. 8, no. 3, Feb. 2023, Art. no. 44, https://doi.org/10.3390/infrastructures8030044.

[2] P. Zheng and M. Mike, "An Investigation on the Manual Traffic Count Accuracy," *Procedia - Social and Behavioral Sciences*, vol. 43, pp. 226–231, 2012, https://doi.org/10.1016/j.sbspro.2012.04.095.

[3] M. Majumder and C. Wilmot, "Automated Vehicle Counting from Pre-Recorded Video Using You Only Look Once (YOLO) Object Detection Model," *Journal of Imaging*, vol. 9, no. 7, June 2023, Art. no. 131, https://doi.org/10.3390/jimaging9070131.

[4] C. J. Lin, S. Y. Jeng, and H. W. Lioa, "A Real-Time Vehicle Counting, Speed Estimation, and Classification System Based on Virtual Detection Zone and YOLO," *Mathematical Problems in Engineering*, vol. 2021, pp. 1–10, Nov. 2021, https://doi.org/10.1155/2021/1577614.

[5] S. Kumar *et al.*, "Fusion of Deep Sort and Yolov5 for Effective Vehicle Detection and Tracking Scheme in Real-Time Traffic Management Sustainable System," *Sustainability*, vol. 15, no. 24, Dec. 2023, Art. no. 16869, https://doi.org/10.3390/su152416869.

[6] S. Feng, X. Cai, L. Li, W. Wang, and S. Ying, "A review of research on vehicle detection in adverse weather environments," *Journal of Traffic and Transportation Engineering (English Edition)*, vol. 12, no. 5, pp. 1452–1483, Oct. 2025, https://doi.org/10.1016/j.jtte.2025.06.001.

[7] J. Redmon, S. Divvala, R. Girshick, and A. Farhadi, "You Only Look Once: Unified, Real-Time Object Detection," in *2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, June 2016, pp. 779–788, https://doi.org/10.1109/CVPR.2016.91.

[8] J. Redmon and A. Farhadi, "YOLO9000: Better, Faster, Stronger," in *2017 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, July 2017, pp. 6517–6525, https://doi.org/10.1109/CVPR.2017.690.

[9] W. Liu *et al.*, "SSD: Single Shot MultiBox Detector," in *Computer Vision – ECCV 2016*, vol. 9905, B. Leibe, J. Matas, N. Sebe, and M. Welling, Eds. Springer International Publishing, 2016, pp. 21–37.

[10] Z. Tian, C. Shen, H. Chen, and T. He, "FCOS: Fully Convolutional One-Stage Object Detection," in *2019 IEEE/CVF International Conference on Computer Vision (ICCV)*, Oct. 2019, pp. 9626–9635, https://doi.org/10.1109/ICCV.2019.00972.

[11] S. Ren, K. He, R. Girshick, and J. Sun, "Faster R-CNN: Towards Real-Time Object Detection with Region Proposal Networks," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 39, no. 6, pp. 1137–1149, June 2017, https://doi.org/10.1109/TPAMI.2016.2577031.

[12] T. Y. Lin, P. Goyal, R. Girshick, K. He, and P. Dollar, "Focal Loss for Dense Object Detection," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 42, no. 2, pp. 318–327, Feb. 2020, https://doi.org/10.1109/TPAMI.2018.2858826.

[13] M. Tan, R. Pang, and Q. V. Le, "EfficientDet: Scalable and Efficient Object Detection," in *2020 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, June 2020, pp. 10778–10787, https://doi.org/10.1109/CVPR42600.2020.01079.

[14] J. Huang *et al.*, "Speed/Accuracy Trade-Offs for Modern Convolutional Object Detectors," in *2017 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, July 2017, pp. 3296–3297, https://doi.org/10.1109/CVPR.2017.351.

[15] A. Bewley, Z. Ge, L. Ott, F. Ramos, and B. Upcroft, "Simple online and realtime tracking," in *2016 IEEE International Conference on Image Processing (ICIP)*, Sept. 2016, pp. 3464–3468, https://doi.org/10.1109/ICIP.2016.7533003.

[16] N. Wojke, A. Bewley, and D. Paulus, "Simple online and realtime tracking with a deep association metric," in *2017 IEEE International Conference on Image Processing (ICIP)*, Sept. 2017, pp. 3645–3649, https://doi.org/10.1109/ICIP.2017.8296962.

[17] Y. Zhang *et al.*, "ByteTrack: Multi-object Tracking by Associating Every Detection Box," in *Computer Vision – ECCV 2022*, vol. 13682, S. Avidan, G. Brostow, M. Cissé, G. M. Farinella, and T. Hassner, Eds. Springer Nature Switzerland, 2022, pp. 1–21.

[18] M. Humayun, F. Ashfaq, N. Z. Jhanjhi, and M. K. Alsadun, "Traffic Management: Multi-Scale Vehicle Detection in Varying Weather Conditions Using YOLOv4 and Spatial Pyramid Pooling Network," *Electronics*, vol. 11, no. 17, Sept. 2022, Art. no. 2748, https://doi.org/10.3390/electronics11172748.

[19] R. Wang *et al.*, "Real-time vehicle target detection in inclement weather conditions based on YOLOv4," *Frontiers in Neurorobotics*, vol. 17, Mar. 2023, Art. no. 1058723, https://doi.org/10.3389/fnbot.2023.1058723.

[20] R. Kejriwal, R. H J, A. Arora, and Mohana, "Vehicle Detection and Counting using Deep Learning basedYOLO and Deep SORT Algorithm for Urban Traffic Management System," in *2022 First International Conference on Electrical, Electronics, Information and Communication Technologies (ICEEICT)*, Feb. 2022, pp. 1–6, https://doi.org/10.1109/ICEEICT53079.2022.9768653.

[21] S. Bose, C. D. Ramesh, and M. H. Kolekar, "Vehicle Classification and Counting for Traffic Video Monitoring Using YOLO-v3," in *2022 International Conference on Connected Systems & Intelligence (CSI)*, Aug. 2022, pp. 1–8, https://doi.org/10.1109/CSI54720.2022.9924018.

[22] M. Guo, S. Song, W. Zhao, L. Chen, J. Zhou, and T. Zhang, "A Traffic Flow Monitoring System Based on YOLOv8 and DeepSORT for Real-Time Vehicle Tracking," in *2025 IEEE International Conference on Pattern Recognition, Machine Vision and Artificial Intelligence (PRMVAI)*, June 2025, pp. 1–5, https://doi.org/10.1109/PRMVAI65741.2025.11108516.

[23] B. Sathiyaprasad, P. A. R. Reddy, and D. M. Venkatesh, "TrackNCount - Intelligent Vehicle Tracking, Counting, and Speed Estimation Using YOLOv8 and DeepSORT Algorithms," in *2025 6th International Conference on Mobile Computing and Sustainable Informatics (ICMCSI)*, Jan. 2025, pp. 1738–1744, https://doi.org/10.1109/ICMCSI64620.2025.10883110.

[24] Z. Chu, "D-YOLO a robust framework for object detection in adverse weather conditions." arXiv, 2024, https://doi.org/10.48550/ARXIV.2403.09233.

[25] H. Abbasi, M. Amini, and F. R. Yu, "Fog-Aware Adaptive YOLO for Object Detection in Adverse Weather," in *2023 IEEE Sensors Applications Symposium (SAS)*, July 2023, pp. 1–6, https://doi.org/10.1109/SAS58821.2023.10254059.

[26] X. Meng, Y. Liu, L. Fan, and J. Fan, "YOLOv5s-Fog: An Improved Model Based on YOLOv5s for Object Detection in Foggy Weather Scenarios," *Sensors*, vol. 23, no. 11, June 2023, Art. no. 5321, https://doi.org/10.3390/s23115321.

[27] Q. Zhang, W. Guo, and M. Lin, "LLD-YOLO: a multi-module network for robust vehicle detection in low-light conditions," *Signal, Image and Video Processing*, vol. 19, no. 4, Apr. 2025, Art. no. 271, https://doi.org/10.1007/s11760-025-03858-6.

[28] Y. Miao, F. Liu, T. Hou, L. Liu, and Y. Liu, "A Nighttime Vehicle Detection Method Based on YOLO v3," in *2020 Chinese Automation Congress (CAC)*, Nov. 2020, pp. 6617–6621, https://doi.org/10.1109/CAC51589.2020.9326819.

[29] N. Ho, M. Pham, N. D. Vo, and K. Nguyen, "Vehicle Detection at Night Time," in *2020 7th NAFOSTED Conference on Information and Computer Science (NICS)*, Nov. 2020, pp. 250–255, https://doi.org/10.1109/NICS51282.2020.9335870.

[30] M. Chaman *et al.*, "A Real-Time Vehicle Detection System for ADAS in Autonomous Vehicles Using YOLOv11 Deep Neural Network on Embedded Edge Platforms," *Engineering, Technology & Applied Science Research*, vol. 15, no. 5, pp. 28077–28082, Oct. 2025, https://doi.org/10.48084/etasr.12138.

[31] X. Li, L. Chen, T. Huang, A. Yang, and W. Liu, "YOLO-edge: real-time vehicle detection for edge devices," *Cluster Computing*, vol. 28, no. 5, Aug. 2025, Art. no. 289, https://doi.org/10.1007/s10586-024-04963-w.

[32] Z. Han, Z. Yue, and L. Liu, "3L-YOLO: A Lightweight Low-Light Object Detection Algorithm," *Applied Sciences*, vol. 15, no. 1, Dec. 2024, Art. no. 90, https://doi.org/10.3390/app15010090.

[33] Y. Gu, L. Chen, and T. Su, "Research on Small Object Detection in Degraded Visual Scenes: An Improved DRF-YOLO Algorithm Based on YOLOv11," *World Electric Vehicle Journal*, vol. 16, no. 11, Oct. 2025, Art. no. 591, https://doi.org/10.3390/wevj16110591.

[34] "YOLOv10: Real-Time End-to-End Object Detection," *Ultralytics Docs*, May 25, 2024. https://docs.ultralytics.com/models/yolov10.

[35] "Ultralytics YOLO11," *Ultralytics Docs*, Sept. 30, 2024. https://docs.ultralytics.com/models/yolo11.

[36] "YOLO12: Attention-Centric Object Detection," *Ultralytics Docs*, Feb. 20, 2025. https://docs.ultralytics.com/models/yolo12.

[37] J. Yang, H. Xie, X. Zhang, J. Chen, and S. Sun, "Enhancing Mine Safety with YOLOv8-DBDC: Real-Time PPE Detection for Miners," *Electronics*, vol. 14, no. 14, July 2025, Art. no. 2788, https://doi.org/10.3390/electronics14142788.

[38] Y. Qin, J. Dong, W. Li, L. Zhang, K. Feng, and Z. Wang, "PSD-YOLO: An Enhanced Real-Time Framework for Robust Worker Detection in Complex Offshore Oil Platform Environments," *Sensors*, vol. 25, no. 20, Oct. 2025, Art. no. 6264, https://doi.org/10.3390/s25206264.

[39] "Hikvision Global - Leading the future of AIoT." https://www.hikvision.com/en/.

[40] F. Wilcoxon, "Individual Comparisons by Ranking Methods," *Biometrics Bulletin*, vol. 1, no. 6, Dec. 1945, Art. no. 80, https://doi.org/10.2307/3001968.

[41] S. Holm, "A Simple Sequentially Rejective Multiple Test Procedure," *Scandinavian Journal of Statistics*, vol. 6, no. 2, pp. 65–70, 1979.

[42] T. C. Vu *et al.*, "Vehicle Detection, Tracking and Counting in Traffic Video Streams Based on the Combination of YOLOv9 and DeepSORT Algorithms," *Journal of Future Artificial Intelligence and Technologies*, vol. 2, no. 2, pp. 255–268, June 2025, https://doi.org/10.62411/faith.3048-3719-115.

[43] D. Stiller *et al.*, "Open webcam data for traffic monitoring: YOLOv8 detection of road users before and during COVID-19," *Transportation Research Interdisciplinary Perspectives*, vol. 36, Mar. 2026, Art. no. 101774, https://doi.org/10.1016/j.trip.2025.101774.