Indonesian Journal of Electrical Engineering and Computer Science
Vol. 43, No. 3, September 2026, pp. 898~907
ISSN: 2502-4752, DOI: 10.11591/ijeecs.v43.i3.pp898-907

Event-level illegal parking detection using YOLOv8 and SSD models

Prachi Sodegaonkar[1], Rahat A. Khan[2]
[1]Department of Artificial Intelligence and Machine Learning, Deogiri Institute of Engineering and Management Studies, Sambhajinagar, India
[2]Department of Computer Science and Engineering, G. H. Raisoni International Skill Tech University, Pune, India

Article Info

Article history:

Received Jan 15, 2025
Revised Jun 29, 2026
Accepted Aug 19, 2026

Keywords:

Illegal parking detection
Intelligent transportation systems
Single shot multibox detector
Surveillance video analysis
YOLOv8

ABSTRACT

Unauthorized parking in urban areas results in traffic congestion and leads to the inefficient use of road infrastructure. This study presents an event-level illegal parking detection framework that combines deep learning–based object detection with region-based temporal logic to distinguish transient stops from actual parking violations. A comparative analysis of YOLOv8s and single shot multibox detector (SSD) is conducted using real-world surveillance footage spanning high traffic density and poor lighting conditions. Unlike traditional approaches that rely on hardware-intensive multi-object tracking, our method utilizes a 30-second stationarity threshold within designated regions of interest (ROIs) to detect violations. Experimental results indicate that YOLOv8s outperforms SSD in enforcement reliability, achieving a Recall of 1.00 on a limited-scale evaluation set of 12 confirmed violation events and an F1-score of 0.89, compared with SSD's Recall of 0.67 and F1-score of 0.77. Moreover, YOLOv8s achieves an inference speed of up to 97 frames per second (FPS), well beyond real-time requirements. These results indicate strong potential for real-time illegal parking detection under urban surveillance conditions.

This is an open access article under the CC BY-SA license.

Corresponding Author:

Prachi Sodegaonkar
Department of Artificial Intelligence and Machine Learning
Deogiri Institute of Engineering and Management Studies
Chhatrapati Sambhajinagar, Maharashtra, India
Email: prachisodegaonkar2009@gmail.com

1. INTRODUCTION

Urban areas increasingly face traffic congestion and parking management challenges, and manual enforcement remains labor-intensive and difficult to scale. Early automated approaches (background subtraction, motion detection, handcrafted features) proved highly sensitive to illumination and dynamic backgrounds; deep learning has since improved detection robustness. Single-stage detectors such as YOLO and single shot multibox detector (SSD) balance accuracy with real-time speed, but a direct comparison for illegal parking detection remains underexplored. A key challenge is "temporal ambiguity": distinguishing vehicles briefly halted by signals or congestion from those in violation, which frame-level presence alone cannot resolve. Multi-object tracking can help but adds overhead and may fail under occlusion; the proposed approach instead uses a lightweight, tracking-free, centroid-based displacement analysis to determine stationarity across frames.

This study introduces an event-level illegal parking detection framework that combines real-time deep learning–based vehicle detection with region-based temporal reasoning. We conduct a comparative evaluation of YOLOv8s and SSD on real-world surveillance sequences spanning high-density traffic and

low-light conditions, evaluating system performance at the event level using precision, recall, F1-score, and frames per second (FPS).

Contribution and significance of the study:

- An event-level illegal parking detection framework is developed that integrates deep learning–based vehicle detection with region-based temporal reasoning. This approach enables the system to differentiate between short-term vehicle stops and genuine parking violations in surveillance environments.
- A comparative evaluation of two widely used real-time object detection architectures, YOLOv8 and SSD, is conducted under identical experimental conditions using real-world surveillance footage, providing insights into their effectiveness for parking enforcement applications.
- The study adopts an event-oriented evaluation strategy in which parking violations are treated as behavioral events rather than isolated frame-level detections. This provides a more realistic assessment of enforcement reliability in intelligent transportation systems.
- The experimental results demonstrate that combining high-speed object detection with lightweight temporal analysis enables reliable real-time monitoring, indicating the practical potential of the proposed approach for urban traffic surveillance applications.

2. BACKGROUND AND RELATED WORK

The study of illegal parking detection has evolved from traditional image processing to deep learning–based approaches. Early methods relied on background subtraction and motion analysis to detect stationary vehicles [1], [2], while later work integrated handcrafted visual features with machine learning classifiers [3], [4]. Although effective in controlled settings, these approaches proved highly sensitive to lighting, shadows, camera vibration, and dynamic backgrounds, struggled in crowded or changing conditions, and required manual feature engineering that did not scale well across surveillance locations.

As deep learning matured, CNNs became dominant for traffic object detection. Two-stage detectors like Faster R-CNN improved accuracy but were too costly for real-time use [5], favoring faster one-stage detectors. YOLO is widely used for real-time vehicle detection [6], [7], with success in parking management and violation detection [8], [9], typically without addressing temporal factors distinguishing brief stops from violations. SSD uses multi-scale feature maps [10] and reduces false detections, but struggles with small or occluded vehicles in crowded scenes [11].

To address frame-level limitations, recent work incorporates temporal information—via object tracking [12], [13] or temporal-threshold and rule-based logic [14]—though tracking adds overhead and is sensitive to occlusion. Direct, systematic comparisons between YOLO-based detectors and SSD using region-based temporal logic remain limited, particularly across varied conditions and event-level (rather than frame-level or occupancy-based) metrics, motivating the present evaluation.

Recent iterations of YOLO (v4, v7, v8) have further improved detection accuracy while preserving real-time performance [15]–[17], and surveys confirm the suitability of single-stage detectors for intelligent transportation applications [18]. Other studies report robust deep learning–based vehicle detection and tracking across varying traffic densities and illumination [19], [20], and several have investigated vision-based illegal parking and parking-management systems [21], [22], affirming the broader effectiveness of deep learning for automated enforcement [23]-[25].

3. METHOD

This section outlines the proposed framework for identifying illegal parking through the examination of surveillance video data. The methodology is structured into three fundamental stages: the detection of vehicles, the application of region-based spatial filtering, and the assessment of illegal parking through temporal logic. A comparative analysis is conducted utilizing two deep learning–based object detection models, YOLOv8 and SSD, under identical experimental conditions.

3.1. System overview

Figure 1 illustrates the overall system architecture for illegal parking detection using deep learning and temporal logic. The pipeline extracts frames from the video stream, applies deep learning models to localize objects, filters detections spatially to isolate vehicles in high-risk “no-parking” areas, and finally applies a temporal logic engine to determine whether stationary presence constitutes a brief stop or a violation.

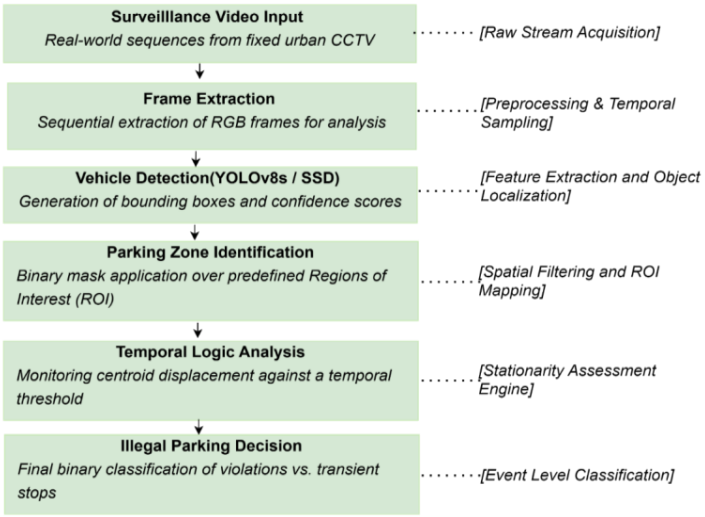


Figure 1. System architecture for detecting illegal parking utilizing deep learning-based vehicle detection and region-based temporal logic

3.2. Vehicle detection using deep learning

The vehicle detection phase employs two one-stage object detection models, chosen for their high-speed inference and localization accuracy: YOLOv8 frames detection as a single regression task, directly predicting bounding boxes and class probabilities for rapid, low-latency inference; SSD instead applies convolutional filters over multi-scale feature maps, enabling detection of both distant and nearby vehicles. Both networks are initialized with pre-trained MS COCO weights via transfer learning, and detections are filtered to retain only vehicle categories (cars, motorcycles, buses, trucks) to reduce false alarms and computational cost.

SSD was selected as a baseline because it remains a widely used, low-overhead real-time detector in embedded vision systems, providing a representative point of comparison for evaluating the improvements introduced by more recent architectures such as YOLOv8. This allows the study to analyze performance differences between earlier anchor-based and newer anchor-free detectors under identical conditions, though we acknowledge SSD's age relative to YOLOv8 as a limitation on comparison fairness.

3.3. Region-based parking zone definition

To reduce false positives from active traffic lanes, the framework segments each video frame into predefined regions of interest (ROI) corresponding to restricted roadsides or no-parking areas, informed by camera perspective, urban layout, and local parking regulations. The system computes the geometric centroid *(x,y)* of its bounding box; only vehicles whose centroid falls within a parking zone's polygonal boundary enter the temporal analysis pipeline, exempting active transit lanes and reducing computational load and erroneous detections. While the ROIs were defined based on scene-specific constraints, the approach is generalizable across different camera perspectives by adapting region boundaries according to the geometric layout of the surveillance environment.

3.4. Temporal logic for illegal parking detection

Conventional object detection methods are fundamentally inadequate for identifying parking violations, as they lack the capability to differentiate between illegal stationary vehicles and temporary stops resulting from traffic signals, pedestrian crossings, or congestion. To address this limitation, the framework incorporates a temporal logic engine specifically designed to assess vehicle behavior across a continuous time-series.

The system monitors the trajectory of the geometric centroid *(x,y)* for each vehicle identified within the pre-defined ROI across sequential frames. The classification of a parking infraction is determined by two primary parameters:

- Spatial displacement threshold (Δd): the system observes the centroid's movement between consecutive frames. When the displacement is below a defined minimum pixel threshold ($\Delta d \approx 0$), the vehicle is classified as "potentially stationary."
- Temporal limit (T_{limit}): a violation is initiated only when the condition of stationarity persists continuously for a period exceeding a specified time threshold (T_{limit}).

On the other hand, vehicles that either traverse a distance greater than Δd or depart from the ROI prior to the expiration of the time limit T_{limit} are categorized as "transiently stopped" and excluded from violation records, ensuring enforcement targets only vehicles with extended immobility and reducing false positives. In many urban parking regulations, brief stopping intervals such as passenger drop-offs are tolerated, whereas prolonged stationary presence typically indicates a violation. The 30-second threshold was chosen to distinguish short traffic- or signal-related stops from genuine violations. The threshold is designed to be adjustable to local regulations, with stricter "no-stopping" zones warranting shorter thresholds and "no-parking" zones longer ones.

3.5. Comparative evaluation framework

YOLOv8 and SSD are assessed under identical conditions: synchronized video sequences, standardized ROI definitions, and consistent temporal thresholds T_{limit}. Each model's performance is evaluated independently at the event level, with each vehicle's presence in a restricted zone treated as a distinct experimental event rather than a simple aggregation of frames.

The evaluation employs a comprehensive set of performance metrics to assess the models across three primary domains: accuracy, reliability, and operational viability. The metrics include,

- Detection accuracy: assessed using precision, recall, and the F1-score to evaluate the models' effectiveness in accurately identifying violators while minimizing false positives.
- Operational robustness: evaluated through the use of a confusion matrix and a comprehensive false detection analysis, aimed at understanding model behavior under challenging environmental conditions.
- Real-time applicability: assessed by evaluating the processing speed, measured in FPS, to ascertain the model's appropriateness for implementation in live smart city surveillance infrastructures.

This systematic framework enables a thorough comparison of the trade-offs between detection sensitivity and computational efficiency, thereby offering a technical foundation for selecting the most suitable architecture for automated traffic enforcement.

3.6. Implementation details

The system is implemented in Python, using OpenCV for video stream handling and frame extraction, PyTorch with the Ultralytics implementation for YOLOv8, and TensorFlow/Keras for SSD. All experiments are performed within a standardized computational environment:

- Hardware configuration: the system is configured on a workstation featuring an Intel Core i7 CPU and an NVIDIA GeForce RTX 30-series GPU. This setup facilitates hardware acceleration through CUDA and cuDNN kernels, thereby optimizing inference latency.
- Software environment: the implementation employs CUDA 11.8 and Python 3.10, with model weights initialized from MS COCO pre-trained checkpoints to establish a standardized baseline for feature extraction.
- Execution parameters: video processing is conducted at a native resolution of 1280×720 pixels, with the temporal logic engine functioning at a sampling rate that is synchronized with the frame rate of the source footage.

4. EXPERIMENTAL SETUP

This section outlines the dataset used for experimentation, the experimental setup, and the evaluation protocol employed to assess the performance of YOLOv8 and SSD in detecting illegal parking.

4.1. Dataset description

Performance is evaluated using the ISLab-PVD (Illegally Parked Vehicle Dataset), assembled from stationary CCTV footage across various urban road settings. Of the 16 sequences in the full repository, this study uses a subset of 11, selected to test the temporal logic engine and detection models across the following variables:

- Environmental variability: luminosity ranges from high-contrast daylight (ISLab-04) to low-light and nighttime conditions (ISLab-16).
- Traffic density and occlusion: the ISLab-13 sequences, characterized by high-density traffic flow, present a significant challenge for the system's ability to maintain centroid tracking in the presence of frequent inter-vehicle occlusions.
- Behavioural complexity: the selected footage contains a balanced distribution of lawful transit, transient stops (due to signalized intersections or congestion), and genuine parking infractions.

While the ISLab-PVD dataset provides diverse urban traffic scenarios, its limited size and restricted public availability may affect full reproducibility. Future studies will consider additional publicly available datasets to improve cross-dataset generalization.

Performance is evaluated at the event level rather than the frame level, since frame-level accuracy can be inflated in largely static scenes. Nonetheless, the number of distinct video sequences (11) and confirmed

violation events (12) is limited, which constrains the statistical generalizability of the results; this is treated as a limitation and addressed further in section 6.

4.2. Experimental configuration

To ensure an impartial benchmarking process, YOLOv8 and SSD were assessed within a rigorously controlled experimental setting. The integration of both architectures into a single video processing pipeline was executed to ensure that any observed performance differences were attributable solely to the model architectures themselves, rather than to external variables.

- Inference engine: the system processes video frame-by-frame, preserving high temporal resolution for stationarity analysis.
- Model initialization: both detectors use transfer learning with pre-trained MS COCO weights, establishing a robust feature-extraction baseline for common urban objects.
- Class filtering: detection is confined to vehicle categories (cars, motorcycles, buses, trucks), suppressing all other classes at the output layer to improve efficiency.

To maintain rigorous consistency across all comparative trials, the main experimental parameters used for both detection models are summarized in Table 1. Beyond Table 1, both models were trained with a learning rate of 0.01, batch size of 16, and 100 epochs, ensuring a fair and consistent comparison.

Table 1. System hyperparameters and operational constraints

Parameter	Value/Setting	Rationale
Confidence threshold	0.40	Balances detection sensitivity with false-positive suppression.
IoU threshold	0.45	Standardizes non-maximum suppression (NMS) for overlapping vehicles.
Input resolution	1280×720	Maintains pixel-level detail for distant vehicle detection.
Temporal threshold (T_{limit})	30 seconds	Aligns with standard urban "no parking" enforcement windows.
Stationarity tolerance (Δd)	5 pixels	Accounts for minor camera vibration or pixel jitter.

4.3. Ground truth annotation

Ground truth was established through manual, frame-by-frame annotation of each vehicle interaction within the ROI, guided by the following criteria:

- Event classification: each vehicle entering the ROI is an individual experimental unit, classified as a Violation (Illegal Parking) or Non-Violation (Transit or Temporary Stop).
- Temporal categorization: a "non-violation" designation is applied to vehicles that either remain in motion or are stationary for a duration shorter than the specified temporal threshold (T_{limit}). In contrast, an "illegal parking" designation is applied exclusively when a vehicle remains stationary for a period exceeding (T_{limit}).
- Definitive reference: these human-verified annotations constitute the gold standard for determining the rates of true positives (TP), false positives (FP), and false negatives (FN).

4.4. Evaluation protocol

To rigorously quantify performance, this study employs an event-based evaluation protocol, which avoids the inflated accuracy that frame-level analysis can produce in largely static scenes. Results are classified as follows:

- TP: an illegal parking violation is accurately detected after the temporal threshold T_{limit} has been surpassed.
- FP: a vehicle that is legally parked or temporarily stopped (e.g., at a traffic light) is incorrectly marked as a violator.
- FN: an actual instance of illegal parking occurs within the ROI but is not detected or flagged by the system.

From these classifications, the subsequent standard performance metrics are established:

- Precision: refers to the accuracy of violation alerts, quantified by the proportion of true violations relative to the total number of detections.
- Recall: assesses the system's sensitivity by determining the proportion of true violations detected relative to the total number of actual violations present.
- F1-score: defined as the harmonic mean of precision and recall, serving as a comprehensive metric to evaluate the equilibrium between detection accuracy and sensitivity.

Confusion matrices are also developed for each model, and processing speed is quantified in FPS to assess suitability for live smart city surveillance systems. The evaluation follows standard object detection assessment practices in intelligent transportation research, where event-level precision, recall, and F1-score are widely used to measure detection reliability in surveillance-based applications.

Although statistical variance across multiple runs was not explicitly computed due to dataset constraints, the event-level evaluation across diverse traffic scenarios provides a consistent and representative performance assessment. Future work will incorporate statistical validation through repeated trials and confidence interval estimation to further strengthen result reliability.

5. RESULTS AND DISCUSSIONS

Assessment is performed at the event level, with each vehicle regarded as a distinct event: a true positive is a correctly identified violator, while false positives and false negatives indicate incorrect and missed detections. Detection accuracy is evaluated via precision, recall, F1-score, false detection analysis, processing speed, and confusion matrices.

5.1. Comparative vehicle detection analysis

Evaluation began by comparing raw vehicle detection counts across all 11 sequences, since the reliability of the subsequent temporal logic depends entirely on the base model's recall.

5.1.1. Application of evaluation metrics

Per the evaluation protocol (Section 4.4), models are assessed on minimizing false negatives while sustaining high precision. "detection count" serves as the principal sensitivity indicator: a higher count corresponds to better recall, meaning fewer potential violations within the ROI go unidentified.

5.1.2. Quantitative performance and model sensitivity

Figure 2 presents a comparative analysis of raw vehicle detection counts for YOLOv8s and SSD models. Across all test videos, YOLOv8s detected a total of 51 distinct vehicle events, whereas SSD identified 25, indicating a substantial difference in detection sensitivity between the two architectures.

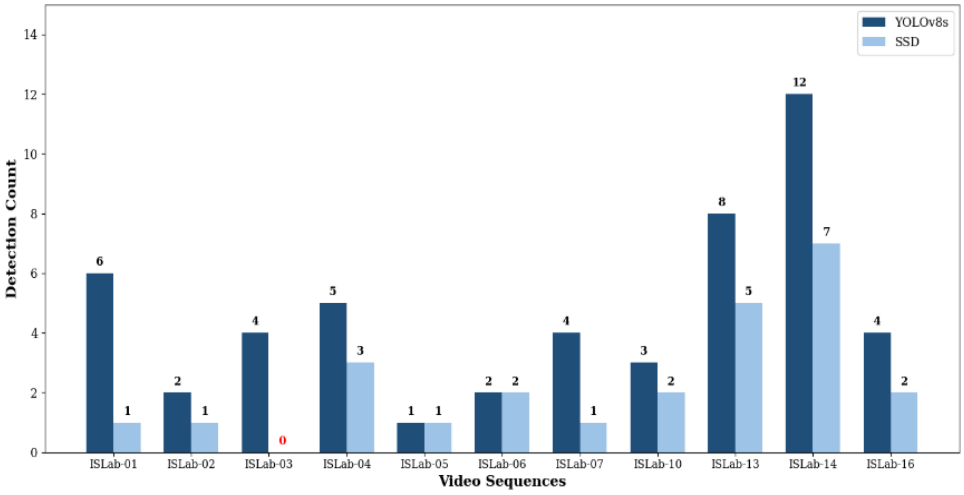


Figure 2. Comparative analysis of raw vehicle detection counts for YOLOv8s and SSD models across

The higher detection count of YOLOv8s demonstrates its stronger capability to identify vehicles under diverse urban conditions. For instance, in the high-density sequence ISLab-14, YOLOv8 successfully identified 12 vehicles, whereas SSD detected only 7. These results indicate that YOLOv8s provides the high recall capability required for a reliable enforcement system.

5.1.3. Analysis of detection philosophy: sensitivity vs. conservatism

This divergence reflects the two architectures' underlying designs: YOLOv8s' anchor-free head and enhanced feature representation improve detection of small or partially occluded vehicles, while SSD's anchor-based, multi-scale approach is more conservative, trading fewer false positives for a higher risk of missed detections. In illegal parking enforcement, where missed violations are more costly than occasional false alarms, this makes YOLOv8s' higher recall particularly valuable in complex urban settings.

5.2. Environmental stressor analysis

The detection models' performance was further categorized into four distinct environmental scenarios to assess their robustness against real-world urban stressors. This classification facilitates a more

nuanced understanding of how factors such as illumination, object density, and background noise affect detection accuracy across the 11 sequences. To further analyze model robustness under different environmental conditions, the detection results were grouped into four scenario categories.

5.2.1. Standard daylight (Baseline)

Across all four categories, YOLOv8s consistently outperformed SSD: 18 vs. 7 detections in high-visibility sequences (ISLab-01, 04, 07, 10; e.g., 6 vs. 1 in ISLab-01), 8 vs. 3 under complex backgrounds and occlusion (ISLab-02, 03, 06; SSD scored zero in ISLab-03 vs. YOLOv8s' 4), 21 vs. 13 in high-density traffic (ISLab-05, 13, 14; 12 vs. 7 in ISLab-14), and 4 vs. 2 under low light (ISLab-16). This consistent margin supports YOLOv8s' suitability for continuous, condition-independent surveillance.

5.3. Illegal parking identification results

While the preceding sections examined the models' fundamental detection capabilities, this section assesses the performance of the integrated system in identifying actual parking violations. This evaluation involves the application of a region-based temporal logic mechanism to distinguish transient vehicle stops from genuine illegal parking events.

5.3.1. Temporal rule logic validation

The application of a 30-second stationarity threshold has demonstrated significant efficacy in reducing false alarms triggered by non-violating events. In sequences such as ISLab-13 and ISLab-14, characterized by high traffic density and frequent signal-induced stops, the temporal engine effectively differentiated between:

- Transient stops: refers to vehicles that are temporarily halted due to factors such as congestion, traffic signals, or brief passenger drop-offs. These stops are excluded from analysis if the vehicles resume movement within a 30-second timeframe.
- Illegal parking: vehicles that remain stationary within the ROI beyond the specified threshold are identified as violations of parking regulations.

By mandating a stationary period of 30 seconds, the system ensures that enforcement targets vehicles causing immediate obstructions, thereby offering a responsive monitoring framework appropriate for high-traffic urban areas.

5.3.2. Temporal threshold sensitivity analysis

Selecting an appropriate temporal threshold is crucial for accurately differentiating between transient vehicle stops and actual parking violations. In the absence of empirical threshold sweeps in this study, the following outlines the qualitative reasoning behind candidate threshold values, drawing on commonly used temporal intervals reported in surveillance-based parking enforcement literature. Short thresholds may incorrectly classify temporary stops, such as those caused by traffic signals, passenger drop-offs, or congestion, as violations. Conversely, excessively long thresholds may delay violation and reduce system responsiveness. Therefore, multiple candidate thresholds were considered to reason through their expected behavioral impact on violation detection, as summarized below.

Candidate thresholds were reasoned through as follows: 10s - very sensitive to short stops, high false-positive risk; 20s - moderately sensitive, some transient stops still misclassified; 30s (selected) - balanced detection, reduces false positives while maintaining sensitivity; 60s - very conservative, higher risk of false negatives and delayed detection. Based on this reasoning, the 30-second threshold was selected as a practical operational compromise between detection sensitivity and reliability. We note this as a limitation of the current study: the threshold was not empirically validated through ablation across multiple values or datasets, and results should be interpreted accordingly. A systematic empirical ablation across threshold values and datasets is planned as future work.

5.3.3. Event-level performance metrics

The ultimate performance outcomes of the two models, following the application of the 30-second temporal logic, are presented in Table 2. Thesese results are derived from the classifications and formulas delineated in the evaluation protocol (Section 4.4).

Table 2. Final event-level performance comparison (30-second threshold)

Model	Total violations	True positives (TP)	False positives (FP)	False negatives (FN)	Precision	Recall	F1-score
YOLOv8s	12	12	3	0	0.8	1	0.89
SSD	12	8	1	4	0.89	0.67	0.77

5.3.4. Confusion matrix and error analysis

Confusion matrices were developed for both models to examine false positive and false negative behavior in more detail. As shown in Figure 3, YOLOv8s' errors are confined to the false positive quadrant (3 cases, from environmental noise misclassified as vehicles), with zero false negatives, yielding a recall of 1.00 and precision of 0.80.

As shown in Figure 3 (right), SSD's errors concentrate in the false negative quadrant (4 cases) rather than false positives (1 case), consistent with its conservative detection behavior: it frequently fails to flag partially occluded or low-light vehicles even after they exceed the 30-second threshold, yielding a recall of only 0.67. False positives in both models mainly stem from temporary halts at traffic signals or congestion, while false negatives concentrate in heavy occlusion or low-light scenarios, underscoring the practical challenges of real-world surveillance conditions.

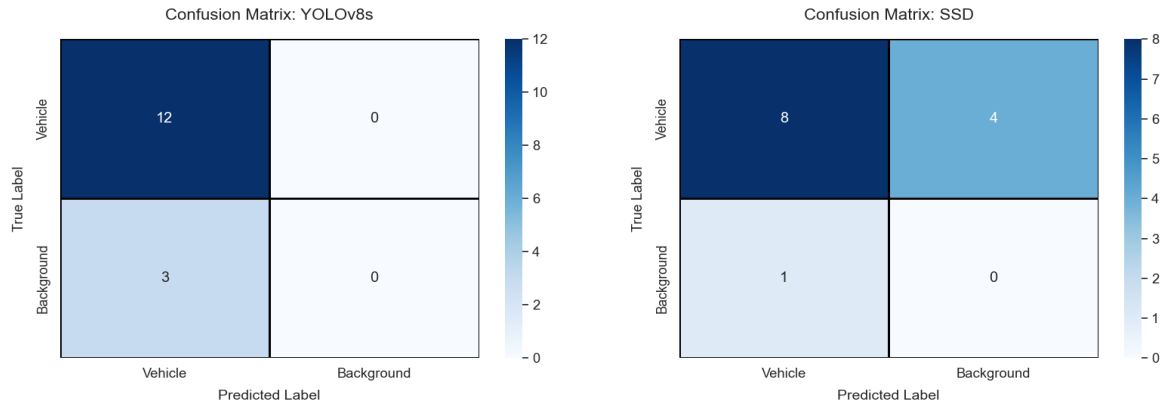


Figure 3. Confusion matrices for YOLOv8s (left) and SSD (right) models

5.4. Operational efficiency and real-time performance

The practical feasibility of an automated system for enforcing illegal parking is contingent upon its capacity to process high-resolution video streams in real-time. Both inference speed and detection accuracy are crucial; any substantial latency or frame loss may result in tracking discontinuities, thereby compromising the precision of the 30-second stationary timer.

5.4.1. Inference speed and throughput

To assess the computational efficiency of the ISLab-PVD system, the throughput of both YOLOv8s and SSD was evaluated. High processing speeds are essential to maintain consistent object IDs across frames. In terms of real-time processing efficiency, YOLOv8s achieved an average inference speed of 85–97 FPS (~10.3 ms per frame), compared to SSD's 22–46 FPS (~16.5 ms per frame), giving YOLOv8s excellent real-time capability versus SSD's marginal performance.

This speed advantage is primarily due to the anchor-free head and C2f backbone modules within the YOLOv8 architecture. In contrast to SSD's anchor-based architecture, which necessitates the evaluation of a large number of pre-defined boxes, YOLOv8s directly predicts object centres, significantly reducing the computational overhead per frame.

5.4.2. Comparative metric analysis

Table 2 summarizes the precision, recall, and F1-score comparison between YOLOv8s and SSD. As Table 2 shows, the most notable difference is in recall: YOLOv8s achieved a perfect 1.00, while SSD's 0.67 means it missed 33% of violations. SSD's marginally higher precision (0.89 vs. 0.80) reflects a more conservative approach that trades missed violations for fewer false alarms, and YOLOv8s' higher F1-score (0.89 vs. 0.77) confirms it as the more balanced model. Since a missed violation is a more critical failure than an occasional false alarm in this context, YOLOv8s shows stronger detection reliability. Future work may extend this comparison to two-stage models such as Faster R-CNN.

5.4.3. Resource efficiency

Although SSD is often noted for minimal memory usage, its throughput in high-density traffic (~22 FPS) is less reliable for real-time enforcement. YOLOv8s balances the speed needed for immediate alerts with efficiency suitable for contemporary edge-computing hardware in smart city deployments.

6. CONCLUSION

This research presented a comparative evaluation of YOLOv8s and SSD architectures for the automated identification of illegal parking events in complex urban environments. By implementing a region-based temporal logic—specifically a 30-second stationarity threshold—this study successfully transitioned from raw frame-level detection to a more meaningful event-level analysis.

YOLOv8s demonstrated the strongest performance, achieving a recall of 1.00 and F1-score of 0.89, with inference speed up to 97 FPS. SSD, despite marginally higher precision, showed a substantially higher failure rate in complex scenes (Recall of 0.67) and lower throughput, making it less suitable for safety-critical enforcement. These results should be read in light of the study's limitations: the 30-second threshold was not empirically validated across zones or datasets, the evaluation set was small (12 violations across 11 sequences), and hardware was limited to a high-performance GPU, leaving edge-device performance untested. Centroid-based filtering may also be less reliable under perspective distortion or severe occlusion. Future work will target empirical threshold ablation, larger cross-validated datasets, edge-device benchmarking, and privacy-preserving deployment (e.g., automated license plate masking) for city-scale surveillance.

Although the dataset is not publicly available due to privacy constraints, detailed methodological descriptions and experimental configurations have been provided to support reproducibility. Future work will focus on releasing annotated datasets and implementation resources.

FUNDING INFORMATION

Authors state no funding involved.

AUTHOR CONTRIBUTIONS STATEMENT

This journal uses the Contributor Roles Taxonomy (CRediT) to recognize individual author contributions, reduce authorship disputes, and facilitate collaboration.

Name of Author	C	M	So	Va	Fo	I	R	D	O	E	Vi	Su	P	Fu
Prachi Sodegaonkar	✓	✓	✓	✓	✓	✓		✓	✓	✓			✓	
Rahat A. Khan		✓		✓		✓		✓	✓	✓		✓		

C : **C**onceptualization
M : **M**ethodology
So : **So**ftware
Va : **Va**lidation
Fo : **Fo**rmal analysis
I : **I**nvestigation
R : **R**esources
D : **D**ata Curation
O : Writing - **O**riginal Draft
E : Writing - Review & **E**diting
Vi : **Vi**sualization
Su : **Su**pervision
P : **P**roject administration
Fu : **Fu**nding acquisition

CONFLICT OF INTEREST STATEMENT

Authors state no conflict of interest.

DATA AVAILABILITY

The data used in this study consist of real-world surveillance video sequences. Due to privacy and ethical restrictions associated with CCTV data, the datasets are not publicly available. However, the data supporting the findings of this study are available from the corresponding author upon reasonable request.

REFERENCES

[1] R. Cucchiara, C. Grana, M. Piccardi, and A. Prati, "Detecting moving objects, ghosts, and shadows in video streams," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 25, no. 10, pp. 1337–1342, Oct. 2003, doi: 10.1109/TPAMI.2003.1233909.

[2] C. Stauffer and W. E. L. Grimson, "Adaptive background mixture models for real-time tracking," in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, 1999, pp. 246–252, doi: 10.1109/CVPR.1999.784637.

[3] J. W. Hsieh, S. H. Yu, Y. S. Chen, and W. F. Hu, "Automatic traffic surveillance system for vehicle tracking and classification," *IEEE Transactions on Intelligent Transportation Systems*, vol. 7, no. 2, pp. 175–187, Jun. 2006, doi: 10.1109/TITS.2006.874722.

[4] N. C. Mithun, S. F. Rashid, and S. M. M. Rahman, "Detection and classification of vehicles from video using multiple time-spatial images," *IEEE Transactions on Intelligent Transportation Systems*, vol. 13, no. 3, pp. 1215–1225, Sep. 2012, doi: 10.1109/TITS.2012.2185803.

[5] S. Ren, K. He, R. Girshick, and J. Sun, "Faster R-CNN: towards real-time object detection with region proposal networks," in *Advances in Neural Information Processing Systems*, 2015, pp. 91–99.

[6] J. Redmon, S. Divvala, R. Girshick, and A. Farhadi, "You only look once: unified, real-time object detection," in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, 2016, pp. 779–788, doi: 10.1109/CVPR.2016.91.
[7] J. Redmon and A. Farhadi, "YOLOv3: an incremental improvement," *arXiv preprint arXiv:1804.02767*, 2018.
[8] L. Jayakumar, M. Dontabhaktuni, and S. Peddakrishna, "Performance evaluation of YOLOv5-based custom object detection model for campus-specific scenario," *International Journal of Experimental Research and Review*, vol. 38, pp. 46–60, 2024, doi: 10.52756/ijerr.2024.v38.005.
[9] J. Zhang *et al.*, "An improved YOLOv5-based object detection framework for traffic surveillance," *Sensors*, vol. 23, no. 7, p. 3693, 2023, doi: 10.3390/s23073693.
[10] W. Liu *et al.*, "SSD: single shot MultiBox detector," in *Proceedings of the European Conference on Computer Vision (ECCV)*, 2016, pp. 21–37, doi: 10.1007/978-3-319-46448-0_2.
[11] A. Hendrawan, R. Gernowo, O. D. Nurhayati, B. Warsito, and A. Wibowo, "Improvement object detection algorithm based on YOLOv5 with BottleneckCSP," in *Proceedings of the IEEE International Conference on Communication, Networks and Satellite (COMNETSAT)*, 2022, pp. 79–83, doi: 10.1109/COMNETSAT56033.2022.9994492.
[12] Bewley, Z. Ge, L. Ott, F. Ramos, and B. Upcroft, "Simple online and realtime tracking," in *Proceedings of the IEEE International Conference on Image Processing (ICIP)*, 2016, pp. 3464–3468, doi: 10.1109/ICIP.2016.7533003.
[13] N. Wojke, A. Bewley, and D. Paulus, "Simple online and realtime tracking with a deep association metric," in *Proceedings of the IEEE International Conference on Image Processing (ICIP)*, 2017, pp. 3645–3649, doi: 10.1109/ICIP.2017.8296962.
[14] G. Amato, F. Carrara, F. Falchi, C. Gennaro, and C. Meghini, "Car parking occupancy detection using smart camera networks and deep learning," in *Proceedings of the IEEE International Conference on Smart Computing*, 2016, pp. 121–128, doi: 10.1109/SMARTCOMP.2016.7501699.
[15] A. Bochkovskiy, C. Y. Wang, and H. Y. M. Liao, "YOLOv4: optimal speed and accuracy of object detection," *arXiv preprint arXiv:2004.10934*, 2020.
[16] G. Jocher, A. Chaurasia, and J. Qiu, "YOLOv8: ultralytics next-generation object detection model," *Ultralytics Documentation*, 2023.
[17] Y. Wang, A. Bochkovskiy, and H. Y. M. Liao, "YOLOv7: trainable bag-of-freebies sets new state-of-the-art for real-time object detectors," in *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, 2022, pp. 7464–7475, doi: 10.1109/CVPR52688.2022.00729.
[18] Z. Q. Zhao, P. Zheng, S. T. Xu, and X. Wu, "Object detection with deep learning: a review," *IEEE Transactions on Neural Networks and Learning Systems*, vol. 30, no. 11, pp. 3212–3232, Nov. 2019, doi: 10.1109/TNNLS.2018.2876865.
[19] X. Chen, H. Ma, J. Wan, B. Li, and T. Xia, "Multi-view 3D object detection network for autonomous driving," in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, 2017, pp. 1907–1915, doi: 10.1109/CVPR.2017.202.
[20] L. Li, D. Wen, and D. Yao, "A survey of traffic control with vehicular communications," *IEEE Transactions on Intelligent Transportation Systems*, vol. 19, no. 5, pp. 1491–1509, May 2018, doi: 10.1109/TITS.2017.2710701.
[21] Z. Shao, J. Cai, Z. Wang, and H. Zhang, "Vehicle detection and tracking in traffic surveillance using deep learning," *IEEE Transactions on Intelligent Transportation Systems*, vol. 22, no. 6, pp. 3755–3766, Jun. 2021, doi: 10.1109/TITS.2020.2967839.
[22] T. Tang, S. Zhou, Z. Deng, L. Lei, and H. Zou, "Vehicle detection in aerial images based on region convolutional neural networks and hard negative mining," *Sensors*, vol. 20, no. 1, p. 336, 2020, doi: 10.3390/s20010336.
[23] J. Wu, M. Xu, and C. Zhao, "Real-time illegal parking detection using deep learning and surveillance videos," *Journal of Advanced Transportation*, vol. 2021, Art. no. 6623714, 2021, doi: 10.1155/2021/6623714.
[24] F. Yang, Z. Li, and Y. Wang, "Intelligent parking management system based on deep learning and video surveillance," *Applied Sciences*, vol. 12, no. 9, p. 4312, 2022, doi: 10.3390/app12094312.
[25] Y. Zhang, L. Wang, and H. Liu, "A vision-based approach for detecting illegal parking in urban traffic scenes," *IEEE Access*, vol. 8, pp. 170456–170468, 2020, doi: 10.1109/ACCESS.2020.3023344.

BIOGRAPHIES OF AUTHORS



Prachi Sodegaonkar is a Ph.D. research scholar in Computer Science and Engineering at Dr. Babasaheb Ambedkar Marathwada University, Chhatrapati Sambhajinagar, researching intelligent parking systems and smart city surveillance. She is an assistant professor at Deogiri Institute of Engineering and Management Studies, with research interests in AI and computer vision. She can be contacted at email: prachisodegaonkar2009@gmail.com.



Dr. Rahat A. Khan is a senior academician and research supervisor in Computer Science and Engineering, specializing in AI and computer vision, with a Ph.D. in Computer Science and a role as research guide at Dr. Babasaheb Ambedkar Marathwada University, Chhatrapati Sambhajinagar. She can be contacted at email: rahat.khan@ghirstu.edu.in.