Vol. 23 (2026): Proceedings of 2026 International Symposium on Engineering Technology Innovation and Engineering Management
(ISETIM 2026)
E-ISSN: 3105-0522 P-ISSN: 3105-0514
DOI: https://doi.org/10.70267/ISETIM-2026.2026.1924

Research on Computer Vehicle Inspection and Its Impact on Urban Traffic Intelligence

Ruichen Chi*

Yan Shan University, QinHuangDao, China

**Corresponding author: Ruichen Chi.*

Abstract

With the acceleration of urbanization, urban traffic faces multiple challenges such as congestion and safety, accompanied by an explosive growth of traffic data. Promoting the intelligent development of urban transportation is of vital theoretical and practical value for improving traffic efficiency and travel experience. This paper adopts a literature review methodology to outline the technological shift from traditional machine learning classifiers to deep learning paradigms in vehicle detection. Traditional vehicle detection relying on SVM and Decision Trees suffers from manual feature extraction and severe operational bloat, yielding sluggish processing speeds that fail modern real-time demands. Conversely, the deep learning paradigm, led by YOLO, treats detection as a single regression problem to drastically enhance speed and accuracy. For comprehensive tracking, integrating YOLO with ByteTrack ensures continuous vehicle identification and precise lane occupancy analysis, while combining it with LSTM networks enables predicting traffic flow two hours in advance for dynamic signal control. To overcome optical limitations in extreme environments, multi-modal sensor fusion (combining YOLO with LiDAR/RADAR) and decentralized edge computing are reshaping smart city infrastructure. This evolutionary trajectory shifts urban mobility toward predictive governance, highlighting future needs for lightweight edge optimization, all-weather fusion, and model interpretability.

Keywords

YOLO, vehicle detection, urban traffic intelligence, sensor fusion, computer vision

1. Introduction

With the progress of rapid urbanization and the popularization of intelligent transportation equipment, the number of Registered Vehicles compiled by CEIC [1] is increasing year by year. Along with this comes a multitude of traffic issues. Urban traffic intelligence, as an important direction to solve urban traffic problems, must first take computer vision as its core to proceed.

According to Zheng et al. [2], for computational stereo vision algorithms, they should encompass the following processing capabilities: image acquisition, image matching, feature extraction, and depth information computation. In accordance with David F. Llorca's study, a fully integrated sensor system for mobile vehicle applications has been presented [3]. However, as traffic complexity grows, traditional sensor methods struggle with real-time demands. Vision-based sensors have emerged as the superior alternative, often implemented via widespread programming libraries like OpenCV. With increasing environmental complexity, such as unpredictable weather, nighttime glare, and heavy occlusion. Conventional computational vision

Copyright © 2026 The Authors. All individual papers remain the copyright of their respective author(s) · Published by Zeus Academic Research Publishing Group

methodologies are no longer sufficient to satisfy the stringent latency requirements of modern applications [4]. Modern ITS requires not just static detection, but continuous tracking and future flow prediction [5]. This paper adopts a systematic literature review methodology and categorizes the technological approaches into three main comparative frameworks: the historical baseline of traditional machine learning, the vanguard of modern deep learning architectures, and complex integration strategies utilizing multi-modal data., exploring how these modern technologies standardize and visualize vehicle characteristics to fundamentally support smart city governance.

2. Literature Review

2.1 Traditional Machine Learning Methods in Vehicle Detection

Historically, before the widespread availability of GPU-accelerated neural networks, vehicle detection within computer vision relied almost exclusively on traditional, supervised machine learning classifiers. Support Vector Machines (SVM) and Decision Trees (DT) represented the foundational phase of intelligent traffic monitoring, where mathematical rigidity and explicit rule-based programming were prioritized over autonomous feature learning. Researchers predominantly utilized the Histogram of Oriented Gradients (HOG) algorithm as the primary method to calculate gradient vectors across image pixels [6]. By capturing the edge structures, contours, and geometric shapes of vehicles, HOG provided a highly structured mathematical representation of visual data that could be fed into an SVM classifier. This method proved particularly pioneering for early autonomous driving research and static intersection monitoring, as it allowed systems to differentiate between generic vehicular shapes and background infrastructure with a reasonable degree of accuracy in highly controlled, well-lit environments. Concurrently, Decision Tree algorithms were widely deployed to construct the logical frameworks of early Intelligent Traffic Management Systems. As demonstrated by Ying, K. et al, DTs proved highly effective in classifying transit data based on specific, human-readable threshold rules [7]. These systems provided a transparent, easily interpretable logical framework for traffic flow analysis. A traffic engineer could explicitly see the decision pathway that led the algorithm to classify a moving cluster of pixels as a 'truck' rather than a 'sedan'. However, this reliance on rigid, hand-crafted features became their ultimate downfall. Traditional algorithms lacked the intrinsic plasticity required to process raw, unpredictable visual noise effectively. As urban traffic scenes became more chaotic, characterized by diverse vehicle models, erratic pedestrian movements, and varying lighting conditions, the manual tuning of HOG parameters and Decision Tree thresholds became an unsustainable engineering burden. Furthermore, the computational architecture of these traditional software-based solutions suffered from profound operational bloat. The necessity of passing a computationally expensive 'sliding window' across every possible scale, orientation, and location of an image frame resulted in exponential calculation times. Sharma, D.K., Pradhan, R., and Agrawal, R. performed a direct empirical comparison utilizing high-resolution highway data, finding that SVM frameworks utilizing HOG features achieved maximum processing speeds of merely 2 frames per second (fps) [8]. This severe latency created an insurmountable bottleneck for modern traffic control centers, which require instantaneous processing of multi-lane, high-speed urban traffic streams to prevent collisions and manage dynamic light signaling. The complete reliance on manual feature extraction fundamentally limited these models, rendering them incapable of scaling to the demands of contemporary Intelligent Transportation Systems.

2.2 The Deep Learning Paradigm: The Evolution of YOLO Architecture

To overcome the severe latency and accuracy limitations introduced by traditional sliding-window classifiers, the academic and industrial fields of computer vision underwent a massive shift toward Convolutional Neural Networks (CNNs). The introduction of the YOLO (You Only Look Once) algorithm revolutionized the discipline by fundamentally changing object detection. Instead of relying on a slow two-stage regional proposal process (like R-CNNs), YOLO treats vehicle detection as a single regression problem [8]. By dividing the input image into a defined grid and simultaneously predicting bounding box coordinates and class probabilities in a single forward neural pass, baseline YOLO models immediately achieved 45 fps, drastically outperforming the sluggish 2 fps constraint of SVMs while simultaneously improving Mean Average Precision [8, 9, 10]. This specific performance leap was the catalyst that allowed ITS to transition from retrospective, post-processed video analysis to real-time, proactive traffic management. As the urban

Copyright © 2026 The Authors. All individual papers remain the copyright of their respective author(s) · Published by Zeus Academic Research Publishing Group

landscape grew more complex, the YOLO architecture underwent rapid, iterative generations to address specific environmental challenges. Mid-generation models like YOLOv3 and YOLOv5 introduced critical advancements such as spatial pyramid pooling (SPP) and multi-scale feature fusion. These architectural enhancements were explicitly engineered to tackle the problem of severe vehicle occlusion and the detection of small, distant objects. According to prior research by Zhang, Y. et al., vehicles located hundreds of meters away are merely represented as pixel clusters [10, 11]. By fusing deep, semantic features with shallow, high-resolution features, these networks enabled traffic cameras to maintain accurate vehicle counts even during extreme rush-hour period where cars frequently overlap in the camera's field of view. Recent research in deep learning has further redefined the theoretical limits of real-time processing. Drawing from the studies conducted by Zhao et al., You et al. and their contemporaries, the YOLOv8 architecture discarded traditional anchor boxes entirely, introducing advanced anchor-free detection heads that drastically simplified the training pipeline and improved the algorithm's generalization capabilities across different sizes of vehicles [12, 13]. Pushing the boundaries further, YOLOv9c enhanced feature extraction capabilities in highly complex, noisy scenarios through the introduction of Programmable Gradient Information (PGI) and Context-Guided modules, guaranteeing that the network maintains precise spatial localization despite significant degradation in optical clarity [14]. Most recently, the release of YOLOv10 in 2024 marked a watershed moment in the pursuit of absolute minimum latency. YOLOv10 introduced a revolutionary NMS-free training paradigm. Historically, deep learning models generated multiple redundant bounding boxes around a single vehicle, requiring an expensive NMS post-processing step to filter them out. Wang and Liao reported that YOLOv10 thoroughly eliminates post-processing latency by employing consistent dual assignment during the training phase [15, 16]. This structural breakthrough makes it currently one of the fastest and most accurate real-time object detection systems in existence, easily exceeding 150 fps on modern hardware. This ultra-low latency is an absolute prerequisite for advanced ITS applications.

2.3 Advanced Tracking, Forecasting, and Multi-Sensor Fusion

The static vehicle detection provides the foundational data for traffic monitoring. However, this kind of detection capability is insufficient for the comprehensive realization of a truly intelligent urban traffic system. Continuous, persistent object tracking is paramount for understanding the temporal dynamics of urban mobility. For instance, You et al. and Nguyen Thi Thanh Thuy et al. demonstrated that integrating YOLOv8 or YOLOv9c with the ByteTrack algorithm enables systems to link objects by intelligently leveraging both high- and low-confidence bounding boxes [13, 14]. Unlike older DeepSORT models that simply discarded low-confidence detections, ByteTrack maintains the vehicle's identity. This reliable tracking capability is vital for calculating exact lane-level occupancy, measuring average route travel speeds, and automatically identifying unruly driving behaviors without changing from identity switches. Furthermore, modern research is aggressively addressing extreme environmental challenges that render standard optical cameras ineffective. Nighttime vehicle detection systems have evolved beyond simple exposure adjustments. Encío, L. *et al.* employ multi-head neural networks built on shared Hourglass backbones, allowing the system to dynamically adapt to the altered visual appearances of vehicles illuminated only by headlights and streetlamps in absolute darkness [4]. To mitigate the inherent physical limitations of optical lenses entirely, the vanguard of traffic intelligence relies heavily on multi-modal sensor fusion. By fusing the rich semantic classification data from YOLO with the geometric precision of LiDAR or RADAR, Dai et al. facilitate a comprehensive environmental perception that exceeds the capabilities of single-modality systems [17]. This decision-level fusion capitalizes on RADAR's ability to easily penetrate heavy rain, thick fog, and snowstorms, ensuring that autonomous driving systems and intersection controllers can make flawless real-time decisions. Beyond the immediate horizon of real-time detection and tracking, the ultimate goal of urban traffic intelligence is predictive governance. Wu, B. and Wang, X.'s studies have successfully integrated the spatial output of lightweight YOLO models with the temporal forecasting capabilities of LSTM recurrent neural networks [5]. By feeding real-time vehicle counts, trajectory matrices, and velocity statistics continuously into an LSTM framework, traffic management authorities can accurately predict urban traffic two hours in advance. By leveraging this anticipatory capability, city management systems can mitigate emerging congestion through dynamic signal control and traffic rerouting, thereby averting city-wide traffic collapse.

Copyright © 2026 The Authors. All individual papers remain the copyright of their respective author(s) · Published by Zeus Academic Research Publishing Group

3. Discussion & Synthesis

Through a comprehensive synthesis of extensive literature, several key insights regarding the evolutionary trajectory of Intelligent Transportation Systems (ITS) have been derived. What becomes immediately apparent in the comparative performance metrics is an irreversible paradigm shift. Traditional manual feature extraction methodologies, heavily reliant on algorithms like Support Vector Machines (SVM) and Decision Trees, suffer from a fatal inability to scale gracefully to the stochastic, dynamic variations inherent in real-world traffic scenarios. Whether facing sudden weather changes, extreme multi-vehicle occlusion, or diverse vehicular shapes, rule-based algorithms are too rigid [6, 7].

Deep learning models, particularly the YOLO family, overcome this by autonomously extracting hierarchical spatiotemporal features directly from raw data. Empirical studies demonstrate that traditional SVM frameworks based on HOG features achieved maximum processing speeds of merely 2 fps [8], creating an insurmountable bottleneck for real-time applications. In contrast, by treating detection as a single regression problem, baseline YOLO models immediately achieved 45 fps [8, 9, 10]. Furthermore, structural breakthroughs in the latest YOLOv10 architecture have completely eliminated post-processing latency, easily exceeding 150 fps on modern hardware [15, 16]. This exponential increase in processing speed—from a sluggish 2 fps to over 150 fps—unequivocally demonstrates deep learning's superior capacity to handle high-throughput ITS demands. However, the primary academic discourse now centers on the continuous trade-off between computational efficiency and absolute detection precision. While YOLOv10 achieves unmatched inferencing speed by eliminating NMS, transplanting such parameter-heavy architectures onto heavily constrained, low-power edge devices is rarely a plug-and-play endeavor. It demands significant engineering compromises, often forcing developers to employ aggressive model pruning, knowledge distillation, or tensor quantization just to make the local deployment viable [15, 16].

Modern ITS applications have definitively evolved from isolated detection algorithms into holistic, interconnected ecosystems. When the highly robust bounding-box outputs of these neural networks are piped directly into sophisticated trackers like ByteTrack, traffic control centers gain the unprecedented ability to seamlessly calculate exact intersection dwell times and autonomously analyze complex, dangerous behaviors such as illegal U-turns, wrong-way driving, or vehicle-pedestrian conflict trajectories [13].

A critical, overriding trend extracted from the most recent literature is the aggressive push toward decentralized Edge Computing. Streaming thousands of uncompressed, high-definition urban camera feeds to a centralized cloud server is no longer practical; the massive bandwidth costs and unacceptable transmission latency render such architectures obsolete. Consequently, the industry is pivoting toward deploying ultra-lightweight YOLO variants directly into Roadside Units (RSUs) and embedded smart cameras. This localized deployment facilitates instantaneous anomaly detection right at the source, lifting a tremendous burden from central cloud infrastructures and critically accelerating emergency response times [6, 18]. Rather than merely watching a traffic jam unfold on a monitor, these systems mathematically predict bottleneck formations hours in advance, executing preventative signal adjustments to keep arteries clear [5].

However, relying solely on optical data remains a severe blind spot for any safety-critical system, as even the most advanced cameras are rendered completely blind by zero-visibility conditions like dense fog or blinding direct sunlight. Therefore, it is imperative that the next generation of smart city infrastructure embraces multi-modal sensor fusion, synthesizing LiDAR, RADAR, and optical processing techniques [17].

4. Conclusion

This study realizes the standardization and visualization of urban vehicle data through the analysis of computer vehicle detection technology, which provides accurate data support for traffic planning and intelligent scheduling, and achieves the research purpose proposed in the paper.

In conclusion, the seamless integration of advanced, computer-based vehicle inspection represents the fundamental linchpin of modern urban traffic intelligence. Reflecting on the evolutionary trajectory of these systems, the technological leap from mathematically rigid algorithms—such as SVM, HOG, and Decision Trees—to highly adaptable neural networks is incredible. In practical field deployments, where legacy systems once choked on chaotic street data at sluggish, single-digit frame rates, the YOLO architecture (from v3 to the NMS-free v10) has completely rewritten the processing baseline. We have moved from sluggish, single-digit

Copyright © 2026 The Authors. All individual papers remain the copyright of their respective author(s) · Published by Zeus Academic Research Publishing Group

frame rates to real-time, ultra-high-precision analytics. By utilizing sophisticated single-stage detectors combined with robust tracking paradigms like ByteTrack, these technologies successfully standardize and continuously visualize massive streams of urban mobility data. By feeding this structured data into LSTM networks for dynamic congestion prediction, authorities can execute automated, intelligent dispatch. We have effectively revolutionized urban mobility, shifting the paradigm from passive camera monitoring to active, predictive city governance.

Yet, as we push the boundaries of what these intelligent systems can achieve, environmental visual limitations and hardware constraints remain exposed. No matter how advanced the YOLO model is, it remains fundamentally blind in extreme meteorological conditions. Future academic research and industrial development must prioritize the refinement of multi-modal sensor fusion techniques—integrating LiDAR and millimeter-wave RADAR with YOLO optical processing. To establish truly robust, all-weather autonomous traffic monitoring networks, advancing ultra-lightweight architectures optimized for edge deployment will be crucial for scaling these solutions globally across resource-constrained intersections. Finally, ensuring the interpretability and transparency of these deep learning models will be essential for their legal, ethical, and social acceptance as they take on increasingly autonomous roles in managing public safety within smart city infrastructure.

References

[1] Number of registered vehicles (no date) CEIC. Available at: https://www.ceicdata.com/en/indicator/number-of-registered-vehicles (Accessed: 10 May 2026).

[2] Zheng, L.-W., Chang, Y.-H. and Li, Z.-Z. (2010) 'A study of 3D feature tracking and localization using a stereo vision system', 2010 International Computer Symposium (ICS2010), pp. 402–407. doi:10.1109/compsym.2010.5685479.

[3] Llorca, D.F. et al. (2010) Vision-based traffic data collection sensor for Automotive Applications, MDPI. Available at: https://www.mdpi.com/1424-8220/10/1/860 (Accessed: 10 May 2026).

[4] Encío, L. et al. (2025) 'Enhanced Nighttime Vehicle Detection for on-board processing', IEEE Access, 13, pp. 44817–44835. doi:10.1109/access.2025.3548837.

[5] Wu, B. and Wang, X. (2026) 'Research on urban traffic flow detection and forecasting based on Yolo Model and LSTM', Frontiers in Computing and Intelligent Systems, 15(3), pp. 6–8. doi:10.54097/8wmzb464.

[6] Xu, Y. et al. (2016) A hybrid vehicle detection method based on viola-jones and hog + SVM from UAV images, MDPI. Available at: https://doi.org/10.3390/s16081325 (Accessed: 10 May 2026).

[7] Ying, K. et al. (2015) 'Decision tree-based machine learning algorithm for in-node vehicle classification', 2015 IEEE Green Energy and Systems Conference (IGESC), pp. 71–76. doi:10.1109/igesc.2015.7359454.

[8] Sharma, D.K., Pradhan, R. and Agrawal, R. (2022) 'Comparative study on real-time vehicle classification', 2022 11th International Conference on System Modeling & Advancement in Research Trends (SMART), pp. 433–437. doi:10.1109/smart55829.2022.10046885.

[9] Doria Usta, D. et al. (2025) 'Intelligent Traffic Management: Comparative Evaluation of yolov3, Yolov5, and yolov8 for vehicle detection in urban environments in Montería, Colombia', Future Transportation, 5(4), p. 191. doi:10.3390/futuretransp5040191.

[10] Zhang, Y. et al. (2022) 'Real-time vehicle detection based on improved Yolo V5', Sustainability, 14(19), p. 12274. doi:10.3390/su141912274.

[11] Li, A. et al. (2023) 'A multi-scale traffic object detection algorithm for road scenes based on improved Yolov5', Electronics, 12(4), p. 878. doi:10.3390/electronics12040878.

[12] Zhao, Z. et al. (2025) 'Research on the construction of real-time vehicle detection model based on Yolov8-BASW', IEEE Access, 13, pp. 183519–183533. doi:10.1109/access.2025.3625088.

[13] You, L. et al. (2024) 'Multi-object vehicle detection and tracking algorithm based on improved yolov8 and ByteTrack', Electronics, 13(15), p. 3033. doi:10.3390/electronics13153033.

Copyright © 2026 The Authors. All individual papers remain the copyright of their respective author(s) · Published by Zeus Academic Research Publishing Group

[14] Nguyen Thi Thanh Thuy et al. (2026) 'An integrated approach for multi-object detection and tracking in traffic monitoring using yolov9c and ByteTrack', Journal of Technical Education Science, 21(01), pp. 81–90. doi:10.54644/jte.2026.2077.

[15] Yolov10: Real-time end-to-end object detection (no date) Ultralytics Docs. Available at: https://docs.ultralytics.com/models/yolov10 (Accessed: 11 May 2026).

[16] Wang, C.-Y. and Liao, H.-Y.M. (2024) 'Yolov1 to yolov10: The fastest and most accurate real-time object detection systems', APSIPA Transactions on Signal and Information Processing, 13(1), pp. 1–38. doi:10.1561/116.20240058.

[17] Dai, Z. et al. (2024) 'Enhanced object detection in autonomous vehicles through LIDAR—camera sensor fusion', World Electric Vehicle Journal, 15(7), p. 297. doi:10.3390/wevj15070297.

[18] Kang, J. et al. (2025) 'Collaborative real-time single-object anomaly detection framework of roadside facilities for traffic safety and management using efficient Yolo', Applied Sciences, 15(20), p. 11139. doi:10.3390/app152011139.

Funding

This research received no external funding.

Conflicts of Interest

The authors declare no conflict of interest.

Acknowledgment

This paper is an output of the science project.

Open Access

This chapter is licensed under the terms of the Creative Commons Attribution-NonCommercial 4.0 International License (http://creativecommons.org/licenses/by-nc/4.0/), which permits any noncommercial use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons license and indicate if changes were made.

The images or other third party material in this chapter are included in the chapter's Creative Commons license, unless indicated otherwise in a credit line to the material. If material is not included in the chapter's Creative Commons license and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder.



Copyright © 2026 The Authors. All individual papers remain the copyright of their respective author(s) · Published by Zeus Academic Research Publishing Group