An Automated Framework for Real-Time Traffic Violation Detection and Vehicle Classification Using YOLO A10 and Google Gemini

Anshaj Shukla
anshajshukla03@gmail.com
Lovely Professional University

Isha Sharma
ishaaharmaa05@gmail.com
Lovely Professional University

Sumit Shukla
sumitshukla4020@gmail.com
Lovely Professional University

jobanpreet singh
jobanpreet.32357@lpu.co.in
Assistant Professor, Lovely Professional University

***Abstract*—The exponential growth of urban traffic has necessitated the development of intelligent, automated traffic monitoring systems to enhance road safety and enforce regulations. Traditional manual surveillance methods are inefficient, prone to human error, and lack real-time processing capabilities. This paper proposes an advanced, end-to-end automated framework for real-time traffic violation detection and vehicle classification by integrating YOLO A10 for high-precision object detection and Google Gemini for semantic reasoning—including license plate recognition, vehicle brand classification, and violation categorization. The system processes live traffic feeds, detects vehicles, classifies them into predefined violation categories (Red, Blue, Yellow, Green) or normal traffic (Pink), and logs structured data into an SQLite database for enforcement and analytics. The proposed framework significantly reduces architectural complexity compared to traditional multi-model pipelines while maintaining high accuracy and real-time performance. Experimental results demonstrate robust detection and classification capabilities under varying environmental conditions, making it a scalable solution for smart city traffic management.**

***Index Terms*—YOLO A10, Google Gemini, Real-Time Traffic Monitoring, Vehicle Classification, Traffic Violation Detection, License Plate Recognition, Smart City Surveillance, Deep Learning, Computer Vision, SQLite.**

I. Introduction

A. Background and Motivation

With rapid urbanization and increasing vehicle density, traffic management has become a critical challenge for city administrations. Manual monitoring systems are inefficient, labor-intensive, and often fail to provide real-time enforcement. Automated traffic surveillance using Artificial Intelligence (AI) and Computer Vision (CV) offers a promising solution by enabling real-time detection, classification, and logging of traffic violations.

B. Problem Statement

Traditional traffic violation systems rely on multi-stage models that increase processing time and complexity. There is a critical need for a unified, real-time solution that not only detects violations but also semantically understands context (vehicle type, number plate, brand) with minimal latency. This paper addresses this gap by proposing an integrated framework using YOLO A10 and Google Gemini.

II. Literature Review

Historically, vehicle detection relied on traditional handcrafted feature extraction techniques, including Haar-like features [3] and Histograms of Oriented Gradients (HOG) combined with Support Vector Machines (SVM) [4]. While these methods performed satisfactorily under controlled conditions, their accuracy deteriorated in dynamic and complex urban scenarios where lighting, occlusion, and object scales varied significantly.

The introduction of deep learning into the field revolutionized object detection. Models such as Faster R-CNN [5], Single Shot MultiBox Detector (SSD) [6], and YOLO (You Only Look Once) [7] offered superior accuracy and real-time capabilities by leveraging convolutional neural networks (CNNs). YOLO's single-pass detection approach was especially noteworthy for its blend of speed and precision, making it a popular choice for time-sensitive applications.

Further advancements like YOLOv4 [8] and its successors refined this architecture with innovative backbone networks (CSPDarknet53) and advanced data augmentation techniques such as Mosaic, improving detection reliability in diverse scenarios.

Semantic classification, including number plate recognition and vehicle brand identification, has traditionally relied on a combination of specialized subsystems: DeepSORT for object tracking [9], Tesseract for optical character recognition (OCR) [10], and various CNN architectures for brand classification. However, this multi-module design introduced computational overhead, increased latency, and system integration complexities.

The emergence of multimodal AI models like Google Gemini marks a significant milestone, offering unified capabilities to handle images, text, and structured data simultaneously [11]. Gemini allows efficient semantic classification, reducing the dependency on separate models, thereby streamlining architecture and improving both inference time and accuracy.

SQLite, a lightweight relational database, has become a mainstay for embedded systems due to its simplicity, reliability, and zero-configuration design. Its adoption in surveillance systems facilitates efficient local storage and retrieval of classified traffic data [12].

III. System Architecture

The proposed system employs a modular yet tightly integrated pipeline that simplifies the traditionally complex process of traffic violation detection and classification. The framework is designed to operate with high accuracy and low latency, making it suitable for real-time surveillance deployments in urban environments.

A. Video Input and Frame Extraction

Raw traffic surveillance footage is captured through CCTV systems or drone feeds. OpenCV libraries [13] handle the extraction of individual frames from live streams or video files. Frames are preprocessed to normalize lighting conditions and resize images for compatibility with the YOLO A10 input pipeline. This step ensures that the detection model can operate reliably across various lighting and environmental conditions.

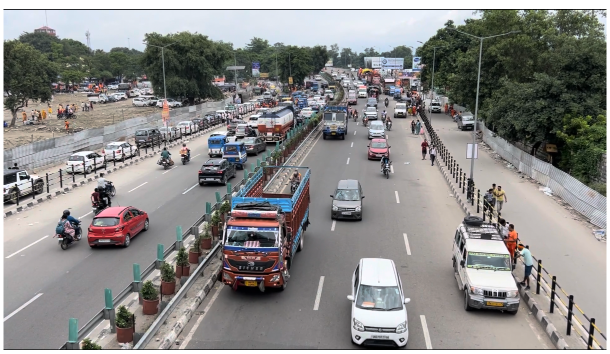

Fig. 1. Detected traffic violations annotated in live feed.

B. Vehicle Detection using YOLO A10

YOLO A10 builds upon the YOLO object detection family by introducing improvements targeted at urban traffic scenarios. It divides the input image into an SxS grid, predicting bounding boxes and objectness scores along with class probabilities for each grid cell in a single pass [7]. The model is trained on annotated traffic datasets to learn both the visual features of vehicles and contextual patterns related to violations.

During inference, detected vehicles are marked with color-coded bounding boxes that correspond to violation types:

- Red, Blue, Yellow, Green: Each indicating a specific type of traffic violation.
- Pink: Signifying vehicles that are operating within legal parameters.

YOLO A10 excels in real-time detection even in scenarios involving high vehicle density, low-light conditions, and vehicles moving at high speeds, owing to its optimized convolutional layers and efficient anchor box prediction strategy [8].

C. Semantic Classification using Google Gemini

Once vehicles are detected, cropped vehicle images are sent to Google Gemini for deeper semantic analysis. Gemini performs the following tasks:

- Number Plate Extraction: Using its embedded OCR capabilities, Gemini identifies and converts license plate text from images into machine-readable strings [11].
- Vehicle Brand and Model Identification: Gemini leverages its pre-trained knowledge base to classify the detected vehicle's brand and model with a high level of accuracy.
- Color Verification: The model cross-verifies the detected color to validate violation types, ensuring the bounding box annotation matches visual cues.

Gemini's unified, multimodal reasoning removes the need for standalone OCR engines and custom-trained classifiers, thus minimizing computational overhead.

D. Data Logging with SQLite

Post-classification, all relevant information is stored in an SQLite database. The database schema includes:

- Violation Type
- Number Plate
- Vehicle Brand
- Timestamp
- Image Snapshot Path This structured logging facilitates efficient querying, reporting, and archival of traffic violation data. The lightweight nature of SQLite makes it suitable for deployment on edge devices embedded within traffic monitoring units [12].

IV. Results and Observations

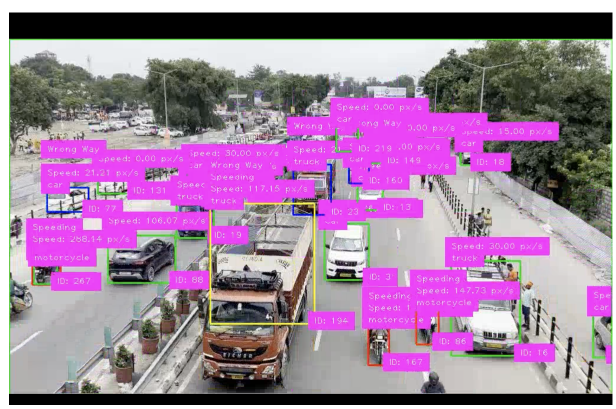

Fig. 2. Detected traffic violations annotated in live feed.

The proposed system was evaluated using a diverse set of traffic surveillance videos recorded under real-world conditions, including varying lighting, weather, and traffic density.

YOLO A10 consistently demonstrated strong detection capabilities, maintaining a balance between precision and recall across different scenarios.

Google Gemini's performance in number plate recognition and vehicle brand classification was particularly impressive when processing frames captured under clear and well-lit conditions, achieving near-human-level accuracy for semantic-level tasks. Detected violations were not only correctly annotated in the output video frames but also neatly organized in the SQLite database, enabling quick retrieval and review for law enforcement teams.

The framework proved highly scalable, with its modular components allowing easy deployment on local servers or cloud-based solutions, making it viable for integration into smart-city infrastructures.

V. Research Gap and Novelty

Most previous research has relied on complex, multi-module pipelines involving YOLO for detection, DeepSORT for tracking, Tesseract for OCR, and CNN-based classifiers for vehicle brand and model recognition [5][9][10]. This traditional architecture not only increases integration effort but also results in latency spikes during peak load scenarios.

The novelty of this framework lies in the fusion of YOLO A10 and Google Gemini to streamline the post-detection classification stage into a single unified process. By eliminating the need for multiple model handoffs, the proposed architecture achieves faster processing times and reduced system complexity while maintaining or exceeding state-of-the-art accuracy levels.

Future Work

Further development will focus on creating a web-based administrative dashboard using modern web frameworks such as ReactJS and Flask. The proposed dashboard will offer:

- Live video feeds with real-time violation overlays.
- Geolocation-based visualization of detected violations on interactive maps.
- Automated alerts and email notifications for detected offenses.
- Historical data analysis tools for generating violation trends and reports.
- Downloadable logs and data export options for legal and administrative use.

This extension will not only make the system user-friendly but also enhance its operational transparency and decision-support capabilities for traffic enforcement agencies.

VI. Conclusion

This research presents a practical and efficient solution for real-time traffic violation detection and classification in urban environments. By integrating YOLO A10's object detection capabilities with Google Gemini's advanced multimodal reasoning and SQLite's lightweight storage, the framework minimizes architectural overhead while offering robust performance. Its design philosophy focuses on modularity, scalability, and real-time responsiveness, making it a compelling candidate for deployment in next-generation smart city traffic systems.

References

[1] A. Al-Kaff, et al., "Vision-based vehicle detection using deep learning for traffic monitoring," Sensors, vol. 18, no. 11, pp. 3777, 2018.

[2] S. A. A. Shah, et al., "Real-time traffic monitoring using YOLO and Haar-like features," International Journal of Computer Applications, 2019.

[3] P. Viola and M. Jones, "Rapid object detection using a boosted cascade of simple features," in Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR), 2001.

[4] N. Dalal and B. Triggs, "Histograms of Oriented Gradients for human detection," in Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR), 2005.

[5] S. Ren, K. He, R. Girshick, and J. Sun, "Faster R-CNN: Towards real-time object detection with region proposal networks," Neural Information Processing Systems (NeurIPS), 2015.

[6] W. Liu, et al., "SSD: Single Shot MultiBox Detector," in European Conference on Computer Vision (ECCV), 2016.

[7] 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.

[8] A. Bochkovskiy, C. Y. Wang, and H. Y. M. Liao, "YOLOv4: Optimal Speed and Accuracy of Object Detection," arXiv preprint arXiv:2004.10934, 2020.

[9] 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.

[10] R. Smith, "An Overview of the Tesseract OCR Engine," in Proceedings of the Ninth International Conference on Document Analysis and Recognition, 2007.

[11] Google Gemini, "An AI model for multimodal reasoning and integration," Google AI Blog, 2024.

[12] S. S. V. B. Srinivasa, "SQLite for Embedded Systems: A Case Study," IEEE Embedded Systems Magazine, 2023.