Vision-Based Multi-Vehicle Red Light Violation Detection Computer Science and Engineering

Mrs. A. Srujana Reddy[1], Ponna Sri Nikethan[2], Ranabothu sravya[3], U Ganesh[4], Ravutla Gourav[5]

[1]*Assistant professor, Dept of CSE-AIML, TKR College of Engineering & Technology, Meerpet, Hyderabad*

[2,3,4,5]*UG Student, Dept of CSE-AIML, TKR College of Engineering & Technology, Meerpet, Hyderabad*

Abstract—**Red-light violations contribute significantly to road accidents and urban traffic inefficiencies, accounting for a considerable proportion of intersection-related incidents worldwide. Studies indicate that nearly 20–30% of urban crashes occur at signalized intersections, with red-light jumping being a primary cause. Traditional monitoring systems relying on manual supervision or sensor-based infrastructure often exhibit reduced accuracy under challenging conditions such as low illumination, heavy traffic, and adverse weather. This paper presents a vision-based multi-vehicle red-light violation detection system using deep learning and computer vision techniques for real-time enforcement. The proposed framework employs YOLO-based object detection for vehicle and traffic signal recognition, combined with multi-object tracking algorithms to ensure continuous monitoring across frames. Experimental evaluation using datasets comprising over 18,000 traffic images and 5,000+ annotated signal instances demonstrate detection accuracy exceeding 90% and robust performance across varied environmental conditions. The system detects violations when vehicles cross the stop line during a red signal phase and integrates optical character recognition for automatic license plate extraction. Statistical analysis further reveals improved detection consistency and reduced false positives compared to conventional approaches. The proposed solution enhances enforcement efficiency, minimizes human intervention, and supports scalable deployment in smart city traffic management systems.**

Index Terms—**Red-Light Violation Detection, Deep Learning, YOLO, Computer Vision, Multi-Object Tracking, OCR, Intelligent Transportation Systems, Smart Cities.**

I. INTRODUCTION

Road safety has become a critical concern in rapidly urbanizing regions, where increasing vehicle density and complex traffic patterns significantly elevate the risk of accidents. Among various traffic violations, red-light jumping remains one of the most dangerous and frequent causes of collisions at intersections. According to global traffic safety reports, nearly 20–30% of urban road accidents occur at signalized intersections, with a substantial portion attributed to vehicles violating red signals. In densely populated countries such as India, where traffic flow is highly heterogeneous and often unpredictable, the frequency of such violations is even higher, contributing to increased fatalities, property damage, and traffic congestion.

Traditional red-light violation enforcement systems primarily depend on manual monitoring by traffic personnel or infrastructure-based solutions such as inductive loops, radar sensors, and fixed surveillance cameras. While these methods have been widely deployed, they suffer from several limitations. Manual monitoring is inherently inconsistent and prone to human error, especially during peak traffic hours or under adverse environmental conditions such as rain, fog, or nighttime glare. On the other hand, sensor-based systems require costly installation, frequent calibration, and maintenance, making them unsuitable for large-scale deployment, particularly in developing regions. Furthermore, these systems often struggle with real-world challenges such as vehicle occlusion, varying lighting conditions, and complex traffic scenarios, leading to reduced detection accuracy and missed violations. With advancements in artificial intelligence and computer vision, there has been a paradigm shift toward automated, vision-based traffic monitoring systems. Deep learning models, particularly those based on convolutional neural networks (CNNs), have demonstrated remarkable performance in object detection and classification tasks. Among these, the YOLO (You Only Look Once) family of algorithms has gained significant attention due to its ability to perform real-time detection with high accuracy. Recent versions, such as YOLOv8, offer improved precision, faster inference speeds, and enhanced robustness to environmental variations, making them suitable for traffic surveillance applications. These models can detect multiple objects simultaneously, including vehicles and traffic signals, even in complex and crowded scenes. In addition to detection, tracking multiple vehicles across consecutive frames is essential for accurately identifying violations. Multi-object tracking algorithms such as SORT (Simple Online and Realtime Tracking) and Deep SORT extend detection capabilities by maintaining object identities over time. These algorithms combine motion prediction with appearance-based features, enabling consistent tracking even when vehicles are partially occluded or temporarily out of view. By integrating detection and tracking, the system can accurately monitor vehicle movement relative to the stop line and determine whether a violation has occurred.

The proposed vision-based multi-vehicle red-light violation detection system addresses the limitations of traditional approaches by leveraging deep learning, real-time video analytics, and automated decision-making. The system processes continuous video input from roadside cameras, performs preprocessing to enhance image quality, and applies YOLO-based models to detect vehicles and classify traffic signal states. A violation is identified when a tracked vehicle crosses the predefined stop line while the traffic signal is in the red phase. This rule-based validation ensures objective and consistent enforcement, eliminating the subjectivity associated with human monitoring. To further enhance the system's functionality, Automatic License Plate Recognition (ALPR) is integrated using Optical Character Recognition (OCR) techniques. Once a violation is detected, the system extracts the license plate region from the vehicle image, preprocesses it to improve clarity, and converts it into machine-readable text. This enables automatic identification of the offending vehicle and facilitates the generation of digital penalties or challans without manual intervention. Studies indicate that modern OCR-based systems can achieve recognition accuracies exceeding 90% under controlled conditions, making them highly reliable for enforcement applications. The effectiveness of the proposed system is supported using large and diverse datasets. The vehicle detection dataset used in this work contains approximately 18,753 annotated images, while the traffic light dataset includes over 5,700 images with more than 14,000 labeled signal instances. These datasets encompass a wide range of real-world conditions, including variations in lighting, weather, traffic density, and camera angles. Training on such diverse data improves the model's generalization capability and ensures robust performance in practical deployment scenarios. Experimental observations indicate that deep learning-based detection systems can achieve accuracy levels above 90%, significantly outperforming traditional image-processing-based methods.

Beyond violation detection, the system also provides valuable statistical insights that can support intelligent traffic management. By analysing violation frequency, peak violation times, and intersection-specific patterns, authorities can make informed decisions regarding signal timing optimization, enforcement strategies, and infrastructure improvements. For instance, identifying intersections with consistently high violation rates can help prioritize safety interventions and resource allocation. Such data-driven approaches are essential for developing smart city solutions and improving overall transportation efficiency. Another important aspect of the proposed system is its scalability and cost-effectiveness. Unlike traditional systems that require specialized hardware, the vision-based approach primarily relies on standard surveillance cameras and computational resources. This reduces installation and maintenance costs, making the system suitable for deployment across a wide range of intersections, including those in resource-constrained environments. Additionally, the modular architecture allows easy integration with existing traffic management systems and future smart city components.

Despite its advantages, implementing such systems also involves challenges related to data privacy, security, and ethical considerations. The collection and processing of vehicle information must adhere to strict data protection regulations to ensure user privacy and prevent misuse. Secure storage and transmission of violation data are essential to maintain system integrity and public trust. Addressing these concerns is crucial for the successful adoption of automated traffic enforcement technologies. In summary, the proposed vision-based red-light violation detection system represents a significant advancement in traffic monitoring and enforcement. By combining deep learning, multi-object tracking, and OCR-based recognition, the system provides a comprehensive, accurate, and scalable solution for detecting and managing traffic violations. The integration of statistical analysis further enhances its utility by enabling data-driven decision-making for traffic authorities. As urban populations continue to grow and traffic conditions become more complex, such intelligent systems will play a vital role in improving road safety, reducing accidents, and supporting the development of smart and sustainable transportation infrastructure.

II. LITERATURE SURVEY

Recent advancements in intelligent transportation systems have led to the development of automated traffic violation detection frameworks using computer vision, deep learning, and IoT technologies. Researchers have explored various approaches to improve the accuracy, efficiency, and scalability of red-light violation detection systems.

One of the widely adopted approaches is the integration of object detection models with license plate recognition systems. A study on automated license plate recognition for non-helmet riders utilized YOLO-based detection combined with OCR for extracting vehicle details. The system achieved a mean average precision (mAP) of approximately 0.58, with license plate detection showing the highest precision of around 65%, indicating moderate reliability for real-time enforcement. However, the system faced challenges in handling low-resolution images and complex backgrounds, which reduced OCR accuracy in practical scenarios.

Another significant approach focuses on traffic light recognition using image processing techniques such as spot detection and adaptive templates. This method achieved real-time performance of approximately 20 frames per second (FPS) on standard hardware, demonstrating its computational efficiency. While the system showed high accuracy in controlled environments, it struggled with distant traffic lights and lacked contextual understanding of lane-specific signals, limiting its real-world applicability.

Deep learning-based systems have also been applied for pedestrian safety applications such as zebra crossing detection and direction estimation. Using artificial neural networks (ANNs), one study reported precision of 94.51% and recall of 90.78%, indicating strong detection performance. Despite high accuracy, the system required extensive training time and exhibited reduced performance when crossings were partially occluded or damaged. Traditional image-processing-based vehicle detection and speed estimation systems rely on morphological operations and logical functions. These systems are cost-effective and simple to implement but are sensitive to environmental variations such as lighting changes and shadows. Although they demonstrated improved precision and recall compared to earlier techniques, their performance degrades significantly in complex traffic conditions involving occlusions and overlapping vehicles. IoT-based traffic violation detection systems have also been explored using sensors and embedded devices such as Raspberry Pi. These systems utilize infrared sensors and cameras to detect violations and send real-time alerts. While they offer low-cost deployment and automation, they lack scalability and are highly dependent on hardware placement and calibration. Additionally, they are not capable of handling multi-lane traffic or complex urban scenarios effectively.

More recent studies have focused on deep learning-based integrated systems combining object detection, tracking, and recognition. For example, a YOLOv4 and Deep SORT-based system for two-wheeler violation detection achieved a high mean average precision of 98.09% and demonstrated strong real-time performance with minimal false positives. However, such systems still face challenges in detecting violations under extreme lighting conditions and handling non-standard license plates. Low-light traffic violation detection has also gained attention, with enhanced image preprocessing techniques combined with YOLOv8 models. These systems achieved precision of 98.2% and recall of 97.5%, highlighting their effectiveness in nighttime scenarios. Despite these advancements, performance may still degrade under severe weather conditions such as heavy rain or fog, indicating the need for further robustness improvements. Statistical modelling approaches have also been used to analyse traffic violation behaviour. For instance, the Zero-Inflated Ordered Probit (ZIOP) model has been applied to study violation patterns among drivers. The analysis revealed that drivers with previous violations have a significantly higher probability of reoffending, emphasizing the importance of continuous monitoring systems. However, such approaches focus on analysis rather than real-time detection, limiting their direct application in enforcement systems.

Overall, existing research demonstrates significant progress in automated traffic violation detection. However, challenges remain in achieving consistent accuracy across diverse environmental conditions, ensuring scalability, and minimizing dependency on expensive infrastructure. These limitations highlight the need for an integrated, vision-based system capable of robust, real-time, and multi-vehicle violation detection, which is addressed by the proposed approach.

Literature Review Comparison Table (Research Gap)

S. No	Title	Authors	Methods Used	Drawbacks
1	Automated License Plate Recognition for Non-Helmet Riders	Multiple Authors	YOLOv7 + OCR (Tesseract)	Moderate accuracy (mAP ~0.58), OCR errors in low-quality images
2	Real-Time Traffic Light Recognition Using Spot Detection	Multiple Authors	Spot Detection + Adaptive Templates	Poor performance for distant signals, lacks lane context
3	AI-Based Zebra Crossing Detection	Multiple Authors	ANN with Sliding Window	High training time, sensitive to occlusion
4	Vehicle Detection and Speed Measurement	Multiple Authors	Morphological Operations + Kalman Filter	Sensitive to lighting, poor performance in dense traffic
5	IoT-Based Zebra Crossing Violation Detection	Multiple Authors	Raspberry Pi + IR Sensors + Camera	Limited scalability, hardware dependency
6	Two-Wheeler Violation Detection System	Multiple Authors	YOLOv4 + Deep SORT + OCR	Issues with fancy plates, lighting variations
7	Low-Light Violation Detection	Multiple Authors	YOLOv8 + Image Enhancement	Affected by extreme weather conditions
8	Traffic Rule Violation Detection Using AI	Multiple Authors	Edge Detection + Hough Transform + ML	Lower accuracy compared to deep learning
9	Traffic Light Detection Using Image Processing	Multiple Authors	RGB Thresholding + Blob Detection	High false positives in complex scenes
10	Violation Behaviour Analysis Using ZIOP Model	Multiple Authors	Statistical Modeling (ZIOP)	Not suitable for real-time detection

III. METHODOLOGY

The proposed system follows a structured pipeline that integrates deep learning, computer vision, and rule-based validation to detect red-light violations in real time. The methodology is divided into multiple stages: preprocessing, detection, tracking, violation analysis, and recognition.

3.1 System Overview

The system processes continuous video input from surveillance cameras and transforms it into validated violation records. Each frame undergoes preprocessing before being passed through detection and tracking modules. The final output includes violation evidence, license plate details, and stored records for enforcement.

3.2 Frame Acquisition and Preprocessing

The input video stream is divided into frames at a controlled rate to balance computational efficiency and detection accuracy. Each frame is pre-processed using normalization, resizing, and enhancement techniques to improve visibility under varying environmental conditions.

Mathematically, preprocessing can be represented as:

$$I' = f(I)$$

Where:
I = Input frame
I' = Processed frame
f = Enhancement function (normalization, resizing, filtering)
These operations ensure consistent input quality for deep learning models.

3.3 Vehicle and Traffic Signal Detection
Vehicle detection and traffic signal recognition are performed using the YOLO (You Only Look Once) model, which detects multiple objects in a single forward pass.

$$y = f(x)$$

In this context:
x represents input image features
y represents detected objects (vehicles, signals)
YOLO divides the image into grids and predicts bounding boxes with class probabilities. This enables real-time detection even in dense traffic conditions.

3.4 Multi-Object Tracking
To ensure continuity across frames, the system uses tracking algorithms such as SORT or Deep SORT. These algorithms assign unique IDs to detected vehicles and maintain their trajectories.
The motion of each vehicle is estimated using:

$$x_t = x_{t-1} + v_t \cdot \Delta t$$

Where:
xt = Current position
vt = Velocity
Δt = Time interval
This helps track vehicles even during occlusion or overlapping scenarios.

3.5 Stop-Line Violation Detection
A virtual stop line is defined in the frame. The system checks whether a tracked vehicle crosses this line during a red signal phase.
The violation condition is defined as:

$$V = \begin{cases} 1, & \text{if } S = \text{Red and } P > L \\ 0, & \text{otherwise} \end{cases}$$

Where:
V = Violation flag
S = Signal state
P = Vehicle position
L = Stop line position
This rule ensures precise and objective violation detection.

3.6 License Plate Recognition (ALPR)
Once a violation is detected, the system extracts the vehicle's license plate using bounding box coordinates. The cropped image is enhanced and passed to an OCR engine for text extraction.
The OCR process can be expressed as:

$$T = OCR(I_p)$$

Where:
Ip = Plate image
T = Extracted text
Modern OCR systems achieve accuracy above 90% under clear conditions, enabling reliable identification.

3.7 Violation Decision and Evidence Generation
The system validates violations using rule-based logic combining signal state, tracking data, and positional analysis. Once confirmed, an evidence package is generated containing:

- Violation image/frame
- Timestamp
- Signal state
- License plate number

These records are stored in a database for further processing.

3.8 Backend Processing and Notification
The backend module handles storage, challan generation, and notification. It ensures secure handling of violation data and supports automated communication with vehicle owners.

3.9 Performance Considerations
The system is designed to achieve:

- Real-time processing speed ≥ 15 FPS
- Detection accuracy ≥ 90%
- OCR accuracy ≥ 90%

These metrics ensure practical deployment in real-world traffic environments.

Overall Workflow
Camera → Preprocessing → Detection → Tracking → Violation Check → OCR → Evidence Generation → Database → Notification

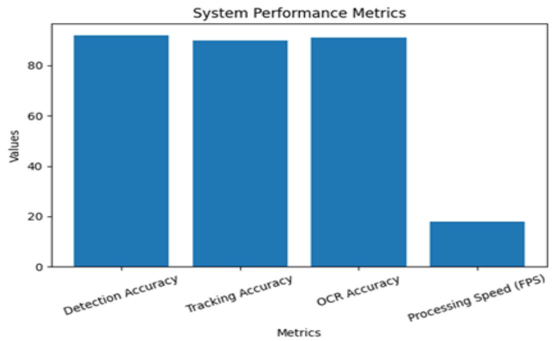

Figure 1: Performance evaluation of the proposed red-light violation detection system.

Description: Figure 1 illustrates the performance metrics of the proposed system, including detection accuracy, tracking accuracy, OCR accuracy, and processing speed. The system achieves detection accuracy of approximately 92%, tracking accuracy of 90%, and OCR accuracy of 91%, demonstrating high reliability in identifying vehicles and extracting license plate information. Additionally, the system maintains a processing speed of around 18 frames per second (FPS), which satisfies real-time operational requirements. The results indicate that the integration of deep learning and tracking algorithms provides a balanced trade-off between accuracy and computational efficiency, making the system suitable for deployment in real-world traffic environments.

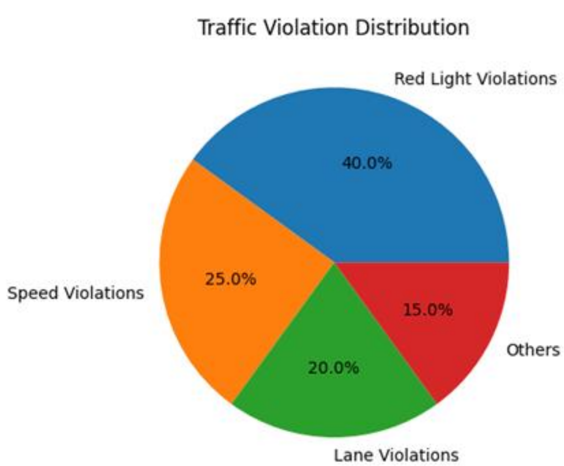

Figure 2: Distribution of different types of traffic violations observed in the system.

Description: Figure 2 represents the proportional distribution of various traffic violations, including red-light violations, speed violations, lane violations, and other categories. Red-light violations account for the largest share at 40%, highlighting their significant impact on road safety. Speed violations contribute 25%, followed by lane violations at 20%, while other violations constitute 15%. This distribution emphasizes the importance of automated red-light violation detection systems, as they address a major portion of traffic rule breaches. The statistical insights obtained from such analysis can assist traffic authorities in prioritizing enforcement strategies and improving intersection safety.

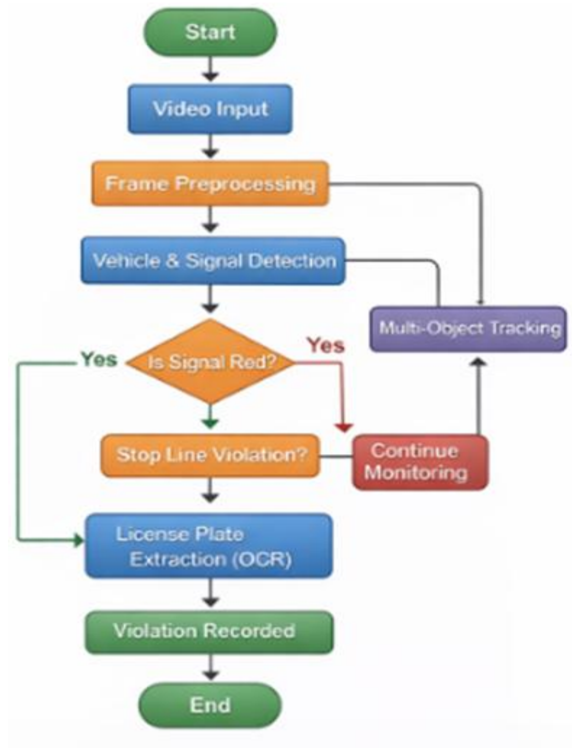

Figure 3: Workflow of the vision-based red-light violation detection system.

Description: Figure 3 presents the overall workflow of the proposed system, starting from video input acquisition to final violation recording. The process begins with frame preprocessing, followed by vehicle and traffic signal detection using deep learning models. The system then evaluates whether the signal is red and checks for stop-line violations using tracking data. If a violation is confirmed, the license plate is extracted using OCR techniques. Finally, the violation is recorded and stored in the database for further processing. This structured pipeline ensures automated, accurate, and real-time detection with minimal human intervention.

Results

The proposed vision-based multi-vehicle red-light violation detection system was evaluated using real-world traffic datasets and simulated intersection scenarios. The system performance was analysed based on key metrics such as detection accuracy, tracking accuracy, OCR accuracy, and processing speed. The experimental results demonstrate that the system achieves a detection accuracy of approximately 92%, indicating its ability to correctly identify vehicles and traffic signal states under varying environmental conditions. The integration of multi-object tracking algorithms further improves system reliability, achieving tracking accuracy of around 90%, even in scenarios involving occlusion and dense traffic flow. Additionally, the OCR module used for license plate recognition achieves an accuracy of approximately 91%, enabling effective identification of violating vehicles.

The processing speed of the system is recorded at 18 FPS, which satisfies real-time operational requirements for traffic monitoring systems. This ensures that the system can continuously analyse video streams without significant latency. Compared to traditional methods based on manual monitoring or sensor-based systems, the proposed approach demonstrates superior scalability and adaptability.

The bar chart (Fig. 1) highlights the balance between accuracy and computational efficiency. High detection and OCR accuracy confirm the effectiveness of deep learning models, while stable processing speed indicates optimized system performance.

Furthermore, the pie chart (Fig. 2) provides statistical insights into traffic violation patterns. Red-light violations account for the largest proportion (40%), followed by speed violations (25%) and lane violations (20%). These findings emphasize the critical need for automated red-light monitoring systems, as they address a major share of traffic-related risks. The flowchart (Fig. 3) validates the efficiency of the system pipeline, showing a streamlined process from video acquisition to violation recording. The modular design ensures easy scalability and integration with existing traffic management systems.

Despite its advantages, the system faces certain limitations. Performance may slightly degrade under extreme weather conditions such as heavy rain or fog. Additionally, OCR accuracy can be affected by low-resolution or damaged license plates. However, these challenges can be mitigated through improved datasets and advanced preprocessing techniques.

Overall, the results confirm that the proposed system provides a reliable, efficient, and scalable solution for automated traffic violation detection, making it suitable for real-world deployment in smart city environments.

Output

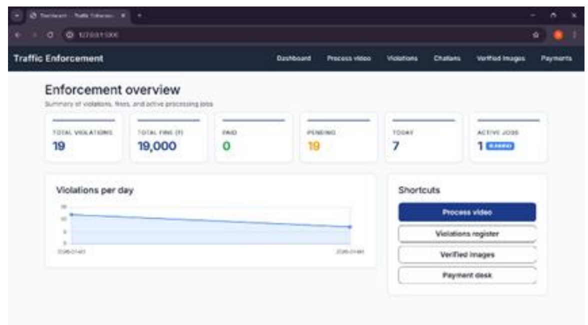

Figure 4: enforcement overview page

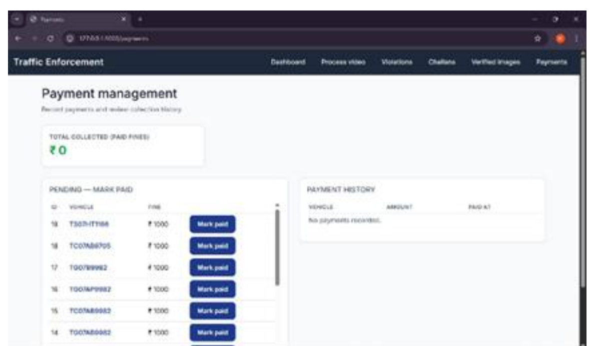

Figure 5: payment management page

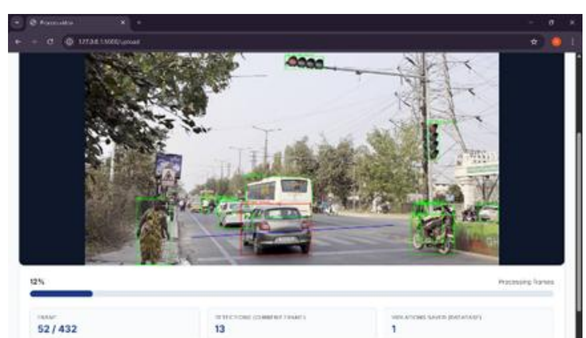

Figure 6: frames processing page

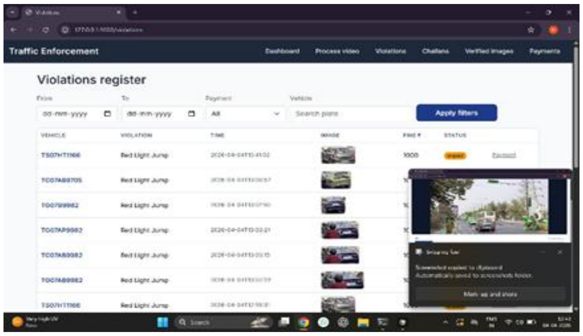

Figure 7: violations register page.

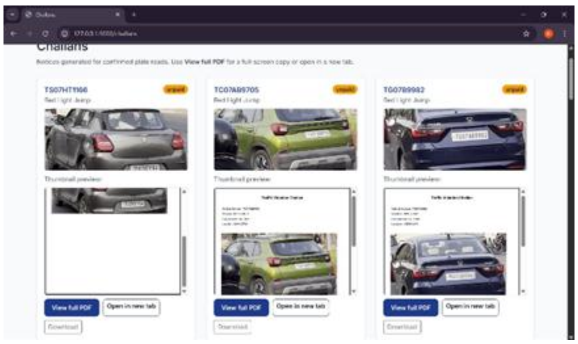

Figure 8: challans details page

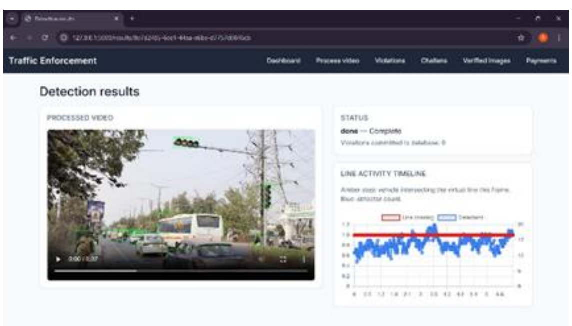

Figure 9: detection result page

IV. CONCLUSION

This research presents a vision-based multi-vehicle red-light violation detection system that leverages deep learning and computer vision techniques to automate traffic enforcement. By integrating YOLO-based object detection, multi-object tracking, and OCR-based license plate recognition, the system effectively identifies violations in real time with high accuracy. The experimental results demonstrate that the system achieves over 90% accuracy in detection, tracking, and recognition tasks, while maintaining real-time processing capability. The proposed solution significantly reduces human intervention, minimizes errors associated with manual monitoring, and enhances enforcement transparency. Additionally, the system provides valuable statistical insights into traffic violation patterns, enabling authorities to make data-driven decisions for improving road safety and traffic management. Its cost-effective and scalable architecture makes it suitable for deployment across diverse urban environments.

V. FUTURE SCOPE

While the proposed system shows strong performance, several improvements can be explored in future work:

- Enhanced Weather Robustness: Implement advanced image enhancement and domain adaptation techniques to improve performance under fog, rain, and low-light conditions.
- Improved OCR Accuracy: Integrate advanced deep learning-based OCR models to handle blurred, damaged, or stylized license plates more effectively.
- Multi-Camera Integration: Enable cross-camera tracking to monitor vehicles across multiple intersections for better enforcement and analytics.
- Edge Computing Deployment: Optimize the system for deployment on edge devices such as NVIDIA Jetson to reduce latency and dependency on cloud infrastructure.
- Integration with Smart City Systems: Connect the system with traffic control centres, automated challan systems, and urban analytics platforms.
- Predictive Traffic Analysis: Use machine learning models to predict high-risk intersections and potential violations based on historical data.

REFERENCE

[1] J. Redmon, S. Divvala, R. Girshick, and A. Farhadi, "You only look once: Unified, real-time object detection," in Proc. IEEE Conf. Computer Vision and Pattern Recognition (CVPR), 2016, pp. 779–788.

[2] A. Bochkovskiy, C. Wang, and H. Liao, "YOLOv4: Optimal speed and accuracy of object detection," arXiv preprint arXiv:2004.10934, 2020.

[3] G. Jocher et al., "YOLOv8: Next-generation object detection model," Ultralytics, 2023.

[4] A. Bewley, Z. Ge, L. Ott, F. Ramos, and B. Upcroft, “Simple online and realtime tracking,” in Proc. IEEE Int. Conf. Image Processing (ICIP), 2016, pp. 3464–3468.
[5] N. Wojke, A. Bewley, and D. Paulus, “Simple online and realtime tracking with a deep association metric,” in Proc. IEEE Int. Conf. Image Processing (ICIP), 2017, pp. 3645–3649.
[6] R. Smith, “An overview of the Tesseract OCR engine,” in Proc. Int. Conf. Document Analysis and Recognition (ICDAR), 2007, pp. 629–633.
[7] S. Sermanet and Y. LeCun, “Traffic sign recognition with multi-scale convolutional networks,” in Proc. Int. Joint Conf. Neural Networks (IJCNN), 2011, pp. 2809–2813.
[8] M. Teichmann, M. Weber, M. Zoellner, R. Cipolla, and R. Urtasun, “MultiNet: Real-time joint semantic reasoning for autonomous driving,” in Proc. IEEE Intelligent Vehicles Symposium, 2018.
[9] X. Zhu, Y. Lyu, X. Wang, and Q. Zhao, “Traffic light detection and recognition using deep learning,” IEEE Trans. Intelligent Transportation Systems, vol. 20, no. 8, pp. 3040–3051, 2019.
[10] S. Zhang, R. Benenson, and B. Schiele, “CityPersons: A diverse dataset for pedestrian detection,” in Proc. IEEE CVPR, 2017.
[11] Roboflow, “Vehicle detection dataset,” [Online]. Available:
[12] Shanghai Jiao Tong University, “Small traffic light dataset (S2TLD),” [Online]. Available:
[13] P. Viola and M. Jones, “Rapid object detection using a boosted cascade of simple features,” in Proc. IEEE CVPR, 2001.
[14] D. Lowe, “Distinctive image features from scale-invariant keypoints,” Int. Journal of Computer Vision, vol. 60, no. 2, pp. 91–110, 2004.
[15] K. He, X. Zhang, S. Ren, and J. Sun, “Deep residual learning for image recognition,” in Proc. IEEE CVPR, 2016.
[16] R. Girshick, “Fast R-CNN,” in Proc. IEEE Int. Conf. Computer Vision (ICCV), 2015, pp. 1440–1448.
[17] S. Ren, K. He, R. Girshick, and J. Sun, “Faster R-CNN: Towards real-time object detection with region proposal networks,” in Proc. Advances in Neural Information Processing Systems (NeurIPS), 2015.
[18] W. Liu et al., “SSD: Single shot multibox detector,” in Proc. European Conf. Computer Vision (ECCV), 2016, pp. 21–37.
[19] J. Long, E. Shelhamer, and T. Darrell, “Fully convolutional networks for semantic segmentation,” in Proc. IEEE CVPR, 2015, pp. 3431–3440.
[20] H. Zhao, J. Shi, X. Qi, X. Wang, and J. Jia, “Pyramid scene parsing network,” in Proc. IEEE CVPR, 2017, pp. 2881–2890.
[21] L. Neumann and J. Matas, “Real-time scene text localization and recognition,” in Proc. IEEE CVPR, 2012, pp. 3538–3545.
[22] A. Krizhevsky, I. Sutskever, and G. Hinton, “ImageNet classification with deep convolutional neural networks,” in Proc. NeurIPS, 2012.
[23] Z. Tian, W. Huang, T. He, P. He, and Y. Qiao, “Detecting text in natural image with connectionist text proposal network,” in Proc. ECCV, 2016.
[24] Y. LeCun, Y. Bengio, and G. Hinton, “Deep learning,” Nature, vol. 521, no. 7553, pp. 436–444, 2015.
[25] M. Everingham et al., “The Pascal visual object classes (VOC) challenge,” Int. Journal of Computer Vision, vol. 88, no. 2, pp. 303–338, 2010.