WWW.IJRPR.COM

International Journal of Research Publication and Reviews

Journal Homepage: www.ijrpr.com ISSN 2582-7421

An Integrated Computer Vision Framework for Automatic Number Plate Recognition and Traffic Violation Detection: A Review of Object Detection, Multi-Object Tracking, and Optical Character Recognition

Dr. Gajendra Asutkar, Kushal Mehar, Harshal Bais, Anand Hatmode

Department of Artificial Intelligence and Data Science Priyadarshini College of Engineering, Nagpur, Maharashtra, India

ABSTRACT

The increasing number of vehicles on roads has created significant challenges for traffic monitoring, road safety, and the enforcement of traffic regulations. Traditional traffic surveillance systems often require continuous human supervision, making them difficult to scale across multiple road locations. Computer vision and deep learning provide opportunities to automate vehicle detection, license plate recognition, and the identification of traffic violations from video streams. This paper presents a review of an integrated computer vision framework for Automatic Number Plate Recognition (ANPR) and traffic violation detection using the YOLO object detection family, ByteTrack multi-object tracking, and PaddleOCR. The review examines existing research on object detection, vehicle tracking, license plate localization, optical character recognition, helmet detection, triple-riding detection, and red-light violation monitoring. It discusses how these components can be combined into a unified processing pipeline that detects road users, maintains vehicle identities across video frames, extracts license plate text, and verifies potential violations using temporal information. The paper also identifies practical challenges involving image quality, illumination, occlusion, vehicle movement, computational resources, and recognition errors. A systematic evaluation strategy is discussed to assess detection, tracking, recognition, and violation-verification performance. The review highlights the importance of temporal verification, vehicle-to-plate association, and evidence preservation in developing reliable traffic monitoring systems. The proposed framework provides a foundation for further research into resource-aware, scalable, and explainable intelligent traffic surveillance.

Keywords: Automatic Number Plate Recognition, ANPR, YOLO, ByteTrack, PaddleOCR, Computer Vision, Object Detection, Multi-Object Tracking, Traffic Violation Detection, Intelligent Transportation Systems.

1. Introduction

Traffic surveillance plays an important role in maintaining road safety, monitoring vehicle movement, and supporting the enforcement of traffic regulations. Conventional traffic monitoring systems rely heavily on manual observation or require dedicated infrastructure and human operators. These approaches can be difficult to maintain when monitoring multiple roads, intersections, and traffic conditions. Consequently, automated traffic monitoring has become an important application of computer vision and artificial intelligence.

Computer vision enables machines to interpret images and videos to identify vehicles, pedestrians, license plates, and other road-related objects. Deep learning-based object detectors can localize relevant objects in video frames, while multi-object tracking algorithms associate detections across consecutive frames. Optical Character Recognition (OCR) converts the visual characters on a license plate into machine-readable text. Combining these techniques enables the development of Automatic Number Plate Recognition systems that can support vehicle identification and traffic surveillance.

Early object detection approaches commonly relied on handcrafted features and traditional image-processing techniques. These methods often required carefully designed rules and could be sensitive to changes in illumination, background complexity, and camera position. The development of convolutional neural networks and deep learning-based detectors improved the ability to recognize objects under more varied conditions. Research on R-CNN, Fast R-CNN, Faster R-CNN, and YOLO established important foundations for modern object detection [1], [12], [13]. Subsequent detector architectures have focused on improving detection speed, localization quality, and practical deployment.

Object detection alone, however, does not provide persistent vehicle identities. A vehicle may appear in many consecutive frames, and independent detections can produce duplicate records or inconsistent associations. Multi-object tracking addresses this issue by connecting detections over time. Algorithms such as SORT, DeepSORT, and ByteTrack provide different approaches to associating objects across video frames [4]–[6]. Tracking is especially relevant to traffic surveillance because a vehicle's movement and observed behavior can help determine whether a potential violation has occurred.

License plate recognition introduces another set of challenges. Plates may be small, blurred, partially obscured, tilted, or affected by reflections and poor lighting. A reliable ANPR pipeline must localize the plate, prepare the cropped image, recognize its characters, and associate the resulting text with the correct vehicle. OCR approaches such as CRNN and text-detection methods based on differentiable binarization provide important background for understanding this process [18], [19]. PaddleOCR offers a practical OCR framework for integrating text detection and recognition into computer vision applications [7].

Beyond vehicle identification, automated surveillance systems can examine safety-related behavior, including helmet usage, multiple riders on two-wheelers, and potential red-light violations. These tasks require more than object classification. They depend on the relationships between detected objects, their positions, their movement, and the applicable traffic rules. Temporal verification can help reduce false alarms caused by isolated detection errors.

This review examines the technologies and research directions relevant to an integrated ANPR and traffic violation detection system. It focuses on the relationship between YOLO-based object detection, ByteTrack-based tracking, PaddleOCR-based text recognition, and rule-based violation verification. It also discusses system integration, practical limitations, and suitable evaluation procedures.

The main contributions of this paper are:

- A review of deep learning-based object detection and multi-object tracking techniques relevant to traffic surveillance.
- An examination of license plate localization, OCR, and vehicle-to-plate association.
- A discussion of helmet detection, triple-riding detection, and red-light violation verification.
- An integrated framework for connecting detection, tracking, recognition, and evidence generation.
- An analysis of practical challenges, evaluation requirements, and directions for future research.

2. Literature Review

2.1 Traditional Computer Vision for Traffic Surveillance

Traditional computer vision techniques use image-processing operations such as thresholding, edge detection, contour extraction, background subtraction, and geometric analysis. These methods can be useful when the environment is controlled and the visual characteristics of objects are predictable. For example, edge information and contour geometry can help identify rectangular license plate regions.

However, traditional techniques may be sensitive to shadows, reflections, changing illumination, complex backgrounds, and variations in plate appearance. Fixed thresholds and manually selected features may not generalize well across different cameras and road environments. These limitations encouraged the development of learning-based methods that can extract visual features directly from training data.

Libraries such as OpenCV provide practical tools for image loading, video processing, geometric transformations, and visualization [22]. Such operations remain useful even in modern deep learning systems because input preparation, region cropping, coordinate transformations, and output rendering are important parts of a complete surveillance pipeline.

2.2 Deep Learning-Based Object Detection

Object detection identifies objects in an image and estimates their locations, commonly using bounding boxes and class labels. In traffic surveillance, relevant classes may include cars, buses, trucks, motorcycles, riders, helmets, traffic lights, and license plates.

The R-CNN family established influential region-based approaches to object detection. Fast R-CNN improved the efficiency of region-based feature processing, while Faster R-CNN introduced a region proposal network to support object localization and classification [12], [13]. These methods provide an important foundation for understanding deep learning-based detection.

YOLO introduced an approach that performs object detection through a unified prediction framework, making it suitable for applications where inference speed is important [1]. Later developments, including YOLOv4, YOLOv7, and YOLOv8, continued to improve practical object detection capabilities and deployment options [2], [3], [17]. The specific performance of any detector depends on its architecture, training data, input resolution, hardware, and deployment configuration.

Other detection research has explored techniques that address difficult aspects of object localization and classification. Focal Loss was introduced to address class imbalance during dense object detection [14]. ResNet demonstrated the value of residual learning for training deeper neural networks [15], while FCOS explored anchor-free object detection [16]. Although these methods do not all target traffic surveillance directly, they contribute to the broader development of modern detection systems.

For an ANPR application, object detection can be divided into related tasks. A vehicle detector identifies the vehicle, while a plate detector locates its license plate. A separate detector or classifier may be required for helmets, riders, or traffic lights. The models and classes used must be selected according to the actual system requirements and training data.

2.3 Multi-Object Tracking

Object tracking associates detections across consecutive video frames to maintain the identity of moving objects. In traffic surveillance, tracking can help determine whether a vehicle remains in a restricted area, crosses a stop line, or appears repeatedly in a sequence of frames.

The Simple Online and Realtime Tracking (SORT) algorithm uses motion estimation and data association to connect detections across frames [4]. DeepSORT extends this approach by incorporating appearance information, which can improve association when motion alone is insufficient [5]. ByteTrack uses a data-association strategy that incorporates lower-confidence detections to help recover objects that might otherwise be lost during tracking [6].

These methods address a key limitation of frame-by-frame detection: the same physical vehicle can generate multiple detections over time. A tracking system can associate these observations with a consistent track identifier, allowing later modules to combine evidence gathered from multiple frames.

Tracking quality can still be affected by occlusion, crowded scenes, abrupt movement, detector errors, and visually similar vehicles. A tracking system should therefore be evaluated using suitable tracking measures and real traffic footage rather than assuming that every track represents a perfectly continuous vehicle identity.

For an integrated ANPR system, tracking provides a link between vehicle detection and later processing. A track identifier can be used to organize plate observations, recognition candidates, and potential violations. This association must be implemented carefully because an incorrect track can lead to a license plate or violation being assigned to the wrong vehicle.

2.4 Automatic Number Plate Recognition

ANPR is the process of identifying a vehicle's registration number from an image or video. A typical pipeline contains vehicle detection, license plate localization, plate image preprocessing, character recognition, and output validation.

Silva and Jung investigated license plate detection and recognition in unconstrained scenarios, highlighting the importance of handling variations in real-world vehicle imagery [8]. The difficulties include differences in plate dimensions, viewing angle, illumination, image resolution, background, and character appearance.

A reliable ANPR pipeline should distinguish between detecting a plate and reading the characters printed on it. A detector may successfully localize a plate even when the image is too blurred or small for accurate recognition. Similarly, OCR may return plausible-looking text that does not correspond to the actual registration number.

Plate localization may use a dedicated object detector or a text-region detection approach. After localization, preprocessing can include resizing, cropping, contrast adjustment, and geometric correction where appropriate. These operations should preserve the character information needed by the recognition model.

The final output should be associated with the correct vehicle and, when tracking is used, with the corresponding track identifier. Maintaining this relationship is essential for preventing a recognized plate from being attributed to a nearby vehicle.

2.5 Optical Character Recognition

OCR converts text present in an image into machine-readable characters. In traffic surveillance, OCR must operate on small and sometimes low-quality license plate images. The process can be affected by motion blur, reflections, skew, occlusion, and similar-looking characters.

CRNN research combined convolutional feature extraction with recurrent sequence modeling for scene text recognition [18]. Differentiable Binarization introduced a method for text-region detection that integrates binarization into the learning process [19]. PP-OCR research explored a practical text-recognition pipeline designed for efficient deployment [7].

PaddleOCR provides tools for text detection and recognition and can be incorporated into an ANPR pipeline. However, OCR output should not automatically be treated as correct. A recognition system may confuse visually similar characters or produce incomplete text when the plate is poorly captured.

Repeated observations across frames can help identify a more reliable candidate. For example, a system may compare recognition outputs associated with the same vehicle track and select a consistent result when sufficient evidence exists. Any such selection process should retain the underlying observations and should not conceal uncertainty when the available evidence is ambiguous.

2.6 Helmet Detection and Triple-Riding Detection

Helmet detection is an important computer vision task for motorcycle safety monitoring. Waranusast and colleagues investigated motorcycle safety helmet detection using image-processing techniques [9]. Subsequent deep learning approaches have enabled more flexible detection of riders and safety equipment in complex scenes.

An automated helmet-monitoring pipeline may identify motorcycles, riders, and helmets, then analyze their spatial relationships. Correct classification requires more than detecting a helmet somewhere in the image. The system must establish whether the helmet is associated with the relevant rider.

Triple-riding detection introduces an additional relationship-based task. The system must identify a motorcycle and determine how many riders are associated with it. Occlusion, overlapping bodies, unusual camera angles, and crowded traffic can make this difficult. Counting detections without considering their spatial association may lead to incorrect conclusions.

Helmet and triple-riding modules should therefore use clearly defined detection classes and association rules. Potential violations should be retained as candidates until sufficient evidence is available. The final decision should reflect the system's configured rules and the quality of the observations.

2.7 Red-Light Violation Detection

Red-light violation detection examines vehicle movement in relation to a traffic signal and a stop line or intersection boundary. It differs from simple object classification because it requires both spatial and temporal information.

A vehicle observed near a red traffic signal has not necessarily committed a violation. The system must establish the signal state, identify the relevant stop line or restricted region, and determine whether the vehicle crossed the boundary under conditions that satisfy the configured violation rule.

Trajectory analysis and object tracking can support this process by connecting vehicle positions across successive frames. The system can examine whether the vehicle moves from one side of the stop line to the other and whether the signal state supports a violation determination.

Important challenges include inaccurate stop-line placement, signal occlusion, changing signal states, camera perspective, missed detections, and uncertain vehicle trajectories. The system must also distinguish between a vehicle approaching the stop line and one that actually crosses it.

A robust implementation should combine geometric rules with temporal evidence. It should preserve the frames that support the decision so that a reviewer can inspect the event rather than relying solely on an automated label.

2.8 Video Processing and Resource-Aware Monitoring

Video surveillance systems must process a sequence of frames while balancing detection quality, processing speed, and available computing resources. Processing every frame with every model can increase computational demand, particularly when several video streams are active.

Video-processing tools such as OpenCV support frame extraction, resizing, image conversion, and video visualization [22]. PyTorch provides a widely used framework for developing and executing deep learning models [23].

A resource-aware pipeline can use configurable frame sampling, controlled model execution, and stream scheduling. For example, processing selected frames rather than every frame may reduce computational load, although it can also cause short events to be missed. The sampling interval must therefore be selected with consideration for vehicle speed, camera frame rate, and the duration of the events being monitored.

Multi-camera tracking presents further challenges. Research on multi-target and multi-camera tracking explores the association of observations across different views [21]. However, tracking the same vehicle between cameras requires additional information, such as camera geometry, timing, or appearance-based association. Maintaining separate local tracks for separate video streams is not equivalent to establishing a single identity across cameras.

2.9 Summary of the Literature

The reviewed literature demonstrates that object detection, multi-object tracking, and OCR solve complementary problems. Object detection identifies relevant objects within individual frames, tracking connects detections over time, and OCR extracts textual information from localized image regions. Violation detection combines these outputs with spatial and temporal rules.

Despite advances in individual components, an integrated traffic surveillance system must address errors that propagate between modules. A missed vehicle detection may interrupt a track. An incorrect plate crop may produce an inaccurate registration number. A tracking error may associate the correct OCR result with the wrong vehicle. A poorly defined violation rule may classify normal movement as an offence.

These issues indicate the importance of evaluating the entire pipeline, preserving evidence, and explicitly representing uncertainty rather than relying exclusively on the performance of individual models.

3. Integrated Framework for ANPR and Traffic Violation Detection

3.1 System Overview

The reviewed framework combines object detection, multi-object tracking, license plate recognition, and rule-based violation verification. Its purpose is to organize video observations into vehicle-level records that can support traffic monitoring and subsequent review.

The general processing pipeline consists of:

1. Video input and frame preparation.
2. Vehicle and relevant-object detection.
3. Multi-object tracking and track management.
4. License plate localization.
5. OCR-based plate text recognition.
6. Association of recognized plates with vehicle tracks.
7. Analysis of configured traffic violations.
8. Temporal verification and evidence generation.
9. Storage and presentation of confirmed or reviewable events.

The exact models, thresholds, and processing schedule must correspond to the implementation used in the project. The framework describes the intended architecture rather than claiming that every possible module has been experimentally validated.

3.2 Video Input and Frame Preparation

The system receives video from traffic cameras or recorded video files. Each input is decoded into frames that can be processed by the detection and tracking modules.

Frame preparation may include resizing, color conversion, and other operations required by the selected models. When multiple streams are processed, the system can use a scheduling strategy to distribute available processing resources across them.

Frame sampling can reduce the number of inference operations. However, sampling must be configured carefully because a vehicle may cross a stop line or enter a restricted region between sampled frames. The system should record the frame index or timestamp associated with each observation so that detected events can be reconstructed.

3.3 Vehicle and Object Detection

A YOLO-based detector identifies vehicles and other supported object categories. Depending on the model configuration, the system may detect cars, motorcycles, buses, trucks, riders, helmets, or traffic-related objects.

Each detection provides a class label, bounding-box coordinates, and a confidence score. These outputs are passed to subsequent processing stages.

A separate license plate detector may be used to localize plates within the frame. This separation allows vehicle detection and plate localization to be configured independently. The quality of both detections influences the downstream tracking and OCR results.

The detector should be trained or configured for the relevant object categories and evaluated on footage that resembles the intended deployment environment. Performance on unrelated datasets cannot establish reliable performance on a particular road or camera setup.

3.4 Multi-Object Tracking Using ByteTrack

ByteTrack can associate object detections across consecutive frames and maintain track identifiers for moving objects [6]. In the integrated framework, vehicle detections are passed to the tracking module, which attempts to associate current detections with existing tracks.

Each track can retain information such as its identifier, recent bounding boxes, last observation time, and associated recognition results. These records provide a foundation for combining observations over time.

Track management must account for temporary missed detections and vehicles entering or leaving the camera view. Incorrect associations may occur when vehicles overlap or move closely together. Consequently, the tracker should be evaluated using actual video sequences, and its output should not be assumed to represent perfect identity continuity.

3.5 License Plate Localization and Recognition Using PaddleOCR

After a plate is localized, the corresponding image region is cropped and passed to the OCR stage. PaddleOCR can be used to detect and recognize text within the plate image, depending on the selected model configuration [7].

Preprocessing may be applied to improve the readability of the crop. The recognition stage produces candidate text that can then be checked for completeness and consistency with the expected plate format, where such validation rules are appropriate.

OCR results should be stored alongside their associated frame or timestamp. When multiple observations are available for a tracked vehicle, the system may compare the candidates and select a result supported by consistent evidence. If the readings disagree, the system should retain the uncertainty rather than force an unsupported identification.

3.6 Vehicle-to-Plate Association

A critical step is establishing which license plate belongs to which detected vehicle. In a crowded traffic scene, several vehicles and plates may appear in the same frame. Selecting the nearest plate or the most confident OCR result alone may not provide a reliable association.

A practical association method can consider spatial containment, relative position, overlap, and temporal consistency. The plate region should be geometrically compatible with the corresponding vehicle bounding box, and the relationship should remain plausible across consecutive observations.

The system should maintain a record connecting the vehicle track, plate crop, OCR output, and source frame. If the association is ambiguous, the system should mark the record for further review rather than assign the plate arbitrarily.

3.7 Helmet Detection

Helmet monitoring uses the relevant detection classes and association rules to determine whether a rider appears to be wearing a helmet. The decision should be based on the relationship between the motorcycle, rider, and helmet detections.

The system should account for partial occlusion and camera viewpoint. A helmet may be hidden by another object, while a detected helmet may belong to a nearby rider rather than the motorcycle under examination.

Potential violations should be accumulated as candidates and checked against subsequent observations where possible. The final output should retain the evidence used to support the classification.

3.8 Triple-Riding Detection

Triple-riding detection requires the system to associate riders with a particular motorcycle and estimate the number of riders on it. The detection module identifies relevant objects, while geometric and temporal rules help establish their relationships.

The system should avoid treating every nearby person as a rider on the same motorcycle. Overlapping motorcycles, pedestrians, and partially visible riders can cause incorrect counts.

Repeated observations can help determine whether the rider count is consistent across frames. However, temporal consistency does not guarantee correctness when the underlying detections are systematically wrong. The system should therefore preserve the original frames and make the decision criteria explicit.

3.9 Red-Light Violation Verification

Red-light monitoring combines traffic-light state estimation, vehicle tracking, and stop-line geometry. The system first identifies the relevant signal and the configured stop-line region. It then examines the movement of a tracked vehicle relative to that boundary.

A potential event should be generated only when the required conditions are supported by the observations. These conditions may include a relevant signal state and evidence that the vehicle crossed the configured boundary.

The system should account for camera perspective and the possibility that signal-state detection changes between frames. Where the signal state or vehicle movement is uncertain, the event should remain unconfirmed until further evidence is available.

3.10 Temporal Verification

Temporal verification reduces dependence on a single frame by checking whether a candidate event remains consistent across multiple observations. It can be applied to plate recognition, helmet classification, rider counting, and vehicle movement.

For example, repeated plate readings can help distinguish a stable registration number from a single OCR error. Similarly, consecutive vehicle positions can provide stronger evidence of stop-line crossing than an isolated detection.

Temporal verification must be designed to avoid introducing unnecessary delay or discarding genuine short-duration events. The number of observations required should be configurable and evaluated against the characteristics of the video stream.

3.11 Evidence Generation and Record Management

When an event meets the configured verification conditions, the system can generate an evidence record containing the event category, vehicle track identifier, timestamp, relevant image, recognized plate text when available, and the reason for the decision.

Evidence images should correspond to the actual source frames. Records should preserve the association between the event and the vehicle rather than storing an isolated violation label without supporting context.

A useful record-management process should distinguish between detected candidates, verified events, and events requiring manual review. It should also handle incomplete information without inserting meaningless placeholder values. Missing or uncertain information should be represented through a clear status or omitted field, according to the data format used by the implementation.

4. Results and Discussion

4.1 Review Findings

The literature indicates that the combination of detection, tracking, OCR, and rule-based reasoning provides a practical foundation for intelligent traffic surveillance. Each component addresses a different aspect of the problem, and the overall reliability depends on how their outputs are connected.

YOLO-based detectors provide object localization for vehicle and traffic-related classes. ByteTrack provides a mechanism for associating detections across frames. PaddleOCR supports text extraction from localized plate images. Temporal and geometric rules can then connect these outputs to potential traffic violations.

The primary benefit of this integrated approach is the ability to create vehicle-level records from video observations rather than treating every frame independently. This structure can support more consistent recognition, event verification, and evidence retrieval.

However, integration introduces dependencies between modules. An error in vehicle localization may affect plate association, while an incorrect track can compromise both plate identification and violation attribution. Therefore, improvements to individual model accuracy do not automatically guarantee equivalent improvements in end-to-end system reliability.

4.2 Comparison of Core Techniques

Technique	Main purpose	Contribution to the framework	Key limitation
YOLO object detection	Detect objects and estimate their locations	Identifies vehicles and configured traffic-related classes	Sensitive to image quality, occlusion, and domain differences
ByteTrack	Associate detections across video frames	Maintains vehicle tracks for temporal analysis	Can experience identity switches and track fragmentation
PaddleOCR	Detect and recognize text	Extracts candidate registration numbers from plate images	Sensitive to blur, small text, and poor illumination
Spatial association	Relate plates and riders to vehicles	Connects object-level observations	Can fail in crowded or overlapping scenes
Temporal verification	Combine evidence across frames	Reduces reliance on isolated observations	May introduce delay or miss brief events
Rule-based violation analysis	Evaluate movement and object relationships	Identifies potential traffic-rule violations	Depends on correct geometry, signal state, and rule definitions
Evidence management	Preserve event context	Supports inspection and traceability	Requires consistent data association and storage practices

4.3 Challenges in Real-World Deployment

Variable illumination: Daylight, shadows, nighttime conditions, headlights, and reflections can affect both detection and OCR. Camera placement and image quality are therefore important factors in practical deployment.

Motion blur and small plates: Fast-moving vehicles may produce blurred images, and distant vehicles may occupy only a small part of the frame. These conditions can reduce plate localization and character-recognition quality.

Occlusion: Other vehicles, riders, and roadside objects may partially hide the target. Occlusion can affect object detection, tracking, helmet analysis, and plate association.

Tracking instability: Vehicle overlap, missed detections, and sudden changes in appearance can produce fragmented tracks or incorrect identity associations. These errors can lead to the wrong plate being linked to a vehicle.

Incorrect OCR results: Visually similar characters and poor image quality may produce ambiguous text. A plausible-looking registration number should not be considered verified without sufficient image evidence.

Violation ambiguity: Traffic rules often depend on the relationship between a vehicle, a road boundary, and a signal state. Poor camera calibration or uncertain signal detection can produce false violation candidates.

Computational constraints: Processing multiple video streams can increase hardware demand. Frame sampling and scheduling can help manage resources, but they must be configured so that important events are not routinely missed.

Privacy and responsible use: Traffic surveillance may involve identifiable vehicle information. Data access, retention, evidence security, and human review should be considered when designing a system for real-world use.

4.4 Evaluation Strategy

A complete evaluation should examine each component and the integrated system. Since this review does not establish a verified experimental dataset or measured results for the project implementation, no numerical performance claims are made here.

For object detection, appropriate measures include precision, recall, and mean Average Precision. These measures help assess whether the system detects relevant objects and localizes them accurately.

For multi-object tracking, measures such as HOTA, MOTA, and IDF1 can help evaluate detection and identity association. Track fragmentation and identity switches should also be examined because they can affect vehicle-level records.

For license plate recognition, character-level and full-plate exact-match accuracy are useful measures. Evaluation should distinguish between plate localization failures and recognition failures so that the source of an error can be identified.

For violation detection, precision and recall should be evaluated against manually reviewed ground-truth events. False alarms, missed violations, and the quality of event evidence should be considered separately.

End-to-end testing should assess whether the correct vehicle, license plate, violation category, and supporting frame are connected in the final record. Testing should include different lighting conditions, traffic densities, vehicle types, camera positions, and video quality levels.

A reproducible evaluation should document the dataset, annotation procedure, model versions, confidence thresholds, tracking configuration, frame-sampling policy, and hardware environment. Any numerical results should be included only after they have been measured using a clearly described procedure.

4.5 Research Gaps

The review identifies several areas where further investigation is needed.

First, many computer vision systems focus on individual tasks rather than evaluating the full sequence from vehicle detection to evidence generation. An integrated evaluation is necessary to determine how errors propagate between components.

Second, vehicle-to-plate association remains important in crowded scenes. A system may correctly detect a plate and a vehicle but still generate an incorrect record if the two are associated incorrectly.

Third, temporal verification requires careful design. Repeated observations can improve confidence, but the verification process must balance reliability, computational cost, and event-detection delay.

Fourth, resource-aware processing needs to be evaluated in relation to event coverage. Frame sampling and multi-stream scheduling may reduce computational requirements, but the system must also establish whether brief or fast-moving events are being missed.

Fifth, traffic violation detection needs clear and reproducible rule definitions. The system should explicitly define signal states, stop-line geometry, rider associations, and the conditions required to classify a candidate event as a violation.

Finally, explainability and evidence traceability deserve greater attention. A useful system should retain the frames and decision conditions that support an event, allowing a human reviewer to understand why the system generated the record.

4.6 Future Research Directions

Future work can explore improvements in several areas.

- **Improved plate recognition:** Evaluate image enhancement, perspective correction, and recognition models using difficult plate images from the intended deployment environment.
- **Robust vehicle association:** Investigate methods that combine spatial relationships, temporal information, and appearance features to reduce incorrect plate-to-vehicle assignments.
- **Adaptive frame processing:** Explore scheduling strategies that allocate more processing to streams or intervals where activity is detected, while evaluating the risk of missing brief events.
- **Temporal confidence estimation:** Develop transparent methods for combining multiple observations without treating repeated but correlated errors as independent evidence.
- **Improved violation verification:** Combine tracking, traffic-signal state estimation, and calibrated road geometry with clearly defined event rules.
- **Human-in-the-loop review:** Provide a mechanism for reviewing uncertain or high-impact events before they are used for enforcement decisions.
- **Reproducible evaluation:** Use documented datasets, annotation rules, evaluation metrics, and deployment configurations to make comparisons meaningful.
- **Privacy-aware record management:** Apply suitable access controls, retention policies, and evidence-handling procedures to protect sensitive surveillance information.

5. Conclusion

This paper reviewed computer vision techniques for Automatic Number Plate Recognition and traffic violation detection, focusing on object detection, multi-object tracking, OCR, and temporal verification. Research on YOLO and related detection architectures provides a foundation for identifying vehicles and relevant objects in traffic footage. Tracking methods such as ByteTrack support the association of detections across consecutive frames, while PaddleOCR provides tools for extracting text from localized license plate images.

The review also examined helmet detection, triple-riding detection, and red-light violation monitoring. These tasks require more than isolated object detection because their decisions depend on spatial relationships, movement, and the consistency of observations over time. An integrated pipeline can combine these elements to produce vehicle-level records and preserve evidence for further inspection.

The principal challenges include variable illumination, occlusion, motion blur, unstable tracking, incorrect OCR outputs, ambiguous violation conditions, and limited computational resources. These challenges demonstrate the need to evaluate the complete system rather than relying solely on the reported capabilities of individual models.

The proposed framework emphasizes accurate object association, temporal verification, traceable evidence, and reproducible evaluation. Actual performance must be established through experiments using documented datasets and the implemented system configuration. With further validation and careful handling of uncertainty, integrated computer vision can provide a useful foundation for intelligent traffic monitoring and road-safety applications.

Acknowledgment

The authors acknowledge the support and guidance provided by the Department of Artificial Intelligence and Data Science, Priyadarshini College of Engineering, Nagpur, in carrying out the project and reviewing the relevant technical literature.

References

[1] 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.

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

[3] Ultralytics, "YOLOv8 Documentation," 2023. Available: https://docs.ultralytics.com/

[4] A. 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.

[5] 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.

[6] Y. Zhang et al., "ByteTrack: Multi-Object Tracking by Associating Every Detection Box," in *Proceedings of the European Conference on Computer Vision (ECCV)*, 2022.

[7] PaddlePaddle, "PaddleOCR: Awesome Multilingual OCR Toolkits Based on PaddlePaddle." Available: https://github.com/PaddlePaddle/PaddleOCR

[8] S. M. Silva and C. R. Jung, "License Plate Detection and Recognition in Unconstrained Scenarios," in *Proceedings of the European Conference on Computer Vision (ECCV)*, 2018.

[9] K. Waranusast, W. Pattanathaburt, V. Tajan, P. Kesorn, and C. S. (verify author details against the original publication), "Machine Vision Based Motorcycle Safety Helmet Detection," *IET Intelligent Transport Systems*, 2014. **Verify the complete author list and exact publication title before submission.**

[10] M. Vishnu et al., "Deep Learning Based Triple Riding and Helmet Detection …," *IEEE International Conference on Intelligent Transportation Systems (ITSC)*, 2021. **Verify the exact title, authors, and publication details against the original paper.**

[11] Kumar and Sharma, "Automated Red Light Violation Detection Using Trajectory Analysis and Deep Learning," *IEEE Access*, 2022. **Verify the exact title, author names, volume, pages, and DOI before submission.**

[12] R. Girshick, "Fast R-CNN," in *Proceedings of the IEEE International Conference on Computer Vision (ICCV)*, 2015.

[13] 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 (NeurIPS)*, 2015.

[14] T.-Y. Lin, P. Goyal, R. Girshick, K. He, and P. Dollár, "Focal Loss for Dense Object Detection," in *Proceedings of the IEEE International Conference on Computer Vision (ICCV)*, 2017.

[15] K. He, X. Zhang, S. Ren, and J. Sun, "Deep Residual Learning for Image Recognition," in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, 2016.

[16] Z. Tian, C. Shen, H. Chen, and T. He, "FCOS: Fully Convolutional One-Stage Object Detection," in *Proceedings of the IEEE/CVF International Conference on Computer Vision (ICCV)*, 2019.

[17] C.-Y. Wang et al., "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)*, 2023. **Verify the final publication details against the official proceedings.**

[18] B. Shi, X. Bai, and C. Yao, "An End-to-End Trainable Neural Network for Image-Based Sequence Recognition and Its Application to Scene Text Recognition," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 2016.

[19] M. Liao et al., "Real-Time Scene Text Detection with Differentiable Binarization," in *Proceedings of the AAAI Conference on Artificial Intelligence*, 2020.

[20] A. Milan et al., "MOT16: A Benchmark for Multi-Object Tracking," *arXiv preprint arXiv:1603.00831*, 2016.

[21] E. Ristani et al., "Performance Measures and a Data Set for Multi-Target, Multi-Camera Tracking," in *Proceedings of the European Conference on Computer Vision Workshops*, 2016. **Verify the exact title and complete bibliographic details.**

[22] OpenCV, "OpenCV Documentation." Available: https://docs.opencv.org/

[23] A. Paszke et al., "PyTorch: An Imperative Style, High-Performance Deep Learning Library," in *Advances in Neural Information Processing Systems (NeurIPS)*, 2019.

[24] Ministry of Road Transport and Highways, Government of India, *Road Accidents in India*. **Verify the exact report edition and year used before citing any statistics.**

[25] National Crime Records Bureau, Government of India, *Accidental Deaths & Suicides in India*. **Verify the exact report edition and year used before submission.**