

AI-Powered Traffic Violation Detection System Using CCTV, and Deep Learning

N. Srinivasa Rao¹, A. Sarayu², G. Abhiram³, D. Maneesha⁴, K.S.V. Sai Sri Phanindra Pavan⁵

¹Assistant Professor, Dept. of CSE (AIML), Bapatla Engineering College, Bapatla 522101, AP, India

^{2,3,4,5}Student, Dept. of CSE (AIML), Bapatla Engineering College, Bapatla 522101, AP, India

srihimasri@gmail.com, sarayu.a555@gmail.com, abhiramgajji2003@gmail.com,

maneesha.dasari1204@gmail.com, pavankuna6506@gmail.com,

Abstract—Urban road safety continues to deteriorate alongside growing traffic density, with traffic violations remaining a leading causal factor in accident-related deaths. Traditional enforcement approaches relying on manual surveillance and passive CCTV recording are systematically inadequate given the scale, speed, and complexity of contemporary traffic environments. This paper presents the design, implementation, and empirical analysis of an AI-Powered Traffic Violation Detection System that augments existing CCTV infrastructure with deep learning inference pipelines to enable fully automated, real-time violation identification and evidence generation. The proposed system employs YOLOv8 as its anchor-free, single-stage object detector, enabling simultaneous detection of five violation types: helmetless riding, traffic light jumping, triple riding, seatbelt non-compliance, and wrong-way driving. Multi-object tracking is achieved through DeepSORT integration, which ensures temporal identity continuity. Automatic License Plate Recognition (ALPR) is performed via a YOLOv8-based plate localization model combined with EasyOCR character recognition. Violation evidence is persistently stored in an SQLite database, while a Streamlit dashboard provides both real-time monitoring and historical analytics. The system achieves a mean Average Precision (mAP@0.5) of 94.0% on a test set of captured real-time frames, a throughput of 28 FPS on GPU, whole-plate recognition accuracy of 88.4%, and evidence logging latency under 180 ms, validating its feasibility for smart traffic policing.

Index Terms—Traffic Violation Detection, YOLOv8, Deep Learning, Computer Vision, DeepSORT, Multi-Object Tracking, ALPR, EasyOCR, CCTV Surveillance, Image capturing, Real-Time Video Processing.

I. INTRODUCTION

Traffic violation is among the most persistent and fatal contributors to urban accidents across the developing world. The World Health Organization estimates approximately 1.19 million traffic-related deaths

annually, disproportionately affecting developing nations. In India, the Ministry of Road Transport and Highways reported 153,972 road deaths in 2021, a significant proportion of which were directly linked to identifiable violations such as helmetless riding, vehicle overloading, and red-light running (MoRTH, 2022).

These statistics reflect a systemic failure at the intersection of policy, enforcement, and technology adoption. The preferred vision of traffic enforcement is one where every violation is reliably captured, documented in court-admissible form, and responded to with timely action. The existing enforcement paradigm, anchored on human officers at crossroads and retrospective inspection of CCTV footage, is far from this standard. Officers are limited in the number of concurrent observations they can make, are subject to attentional fatigue on longer shifts, and cannot provide the incessant, scalable coverage that contemporary urban traffic demands.

Past attempts to bridge this gap have yielded incomplete and specialized solutions. Radar- or inductive-loop-equipped speed cameras capture speeding violations effectively but are incapable of comprehending complex visual scenes involving multiple violation types. Kumar et al. (2021) focused exclusively on helmet detection; Kim et al. (2020) addressed signal jumping in isolation; neither presented a multi-violation unified pipeline required for comprehensive intersection monitoring.

The knowledge gap this study addresses is specific: no previously tested system exists that can simultaneously characterize multiple heterogeneous traffic violation types from standard CCTV streams in real time, automatically identify the license plates of offending vehicles, and generate organized, legally-admissible evidence records within the computational budget of commodity server hardware.

The primary contributions of this paper are: (1) a unified detection pipeline for five violation types using YOLOv8 with DeepSORT temporal confirmation, reducing false

alarms to 6.8%; (2) a two-stage ALPR sub-module achieving 88.4% whole-plate recognition accuracy; (3) a real-time Streamlit dashboard with historical searchable analytics; (4) empirical evaluation achieving 94.0% mAP@0.5 at 28 FPS; and (5) comparative benchmarking against YOLOv5m, YOLOv7-M, and EfficientDet-D2 baselines.

II. LITERATURE REVIEW

The automation of traffic violation detection sits at the confluence of several fast-evolving research fields: real-time object detection, multi-object tracking, optical character recognition, and intelligent transportation systems. Systematic reviews demonstrate that automated enforcement can reduce target violations by 20–40% in monitored locations (Ansari et al., 2023).

A. YOLO-Based Detection Architectures

The YOLO architecture (Redmon et al., 2016) transformed real-time object detection by reformulating it as a single-pass regression problem, enabling speeds several times faster than two-stage alternatives like Faster R-CNN (Ren et al., 2017). YOLOv4 (Bochkovskiy et al., 2020) incorporated Bag-of-Freebies augmentation schemes and a CSPNet backbone, improving mAP while maintaining real-time performance. Ultralytics YOLOv8 (Jocher et al., 2023) introduced a redesigned C2f backbone, decoupled detection heads, and anchor-free regression, eliminating manual anchor tuning and reflecting the current state of the art. Kumar et al. (2021) validated YOLOv5-based traffic detection on Indian footage with 94.3% mAP, demonstrating strong performance in real-world traffic scenarios.

B. Helmet and Multi-Rider Violation Detection

Helmet detection has attracted significant research interest given the disproportionate role of head injuries in two-wheeler fatalities. Liu et al. (2021) demonstrated 91.8% mAP with a multi-scale YOLOv4 system on South Asian datasets. Both studies are prone to false positives when helmets are momentarily occluded. The current study addresses this by requiring five consecutive confirming frames before raising a violation event.

C. Traffic Signal and Trajectory-Based Violation Detection

Red-light violation detection requires matching object trajectories with infrastructure state across time sequences. Kim et al. (2020) achieved 96.1% precision using a CNN-based trajectory-stop-line-signal phase classifier. Bui et al. (2021) extended this with homographic pixel-to-metric ground coordinate mapping for sub-metre stop-line precision. Zhang et al. (2022)

incorporated LSTM-based temporal modelling, achieving a 12% false positive reduction relative to frame-by-frame evaluation—an observation that directly inspired the temporal confirmation design of the present system.

D. Indian Automatic License Plate Recognition

ALPR in the Indian context is complicated by heterogeneous plate formats spanning BH-series, state code, and HSRP variants. Sharma et al. (2021) compiled a dataset of over 50,000 Indian license plate images and demonstrated 95.4% character-level recognition accuracy using YOLOv5 localization combined with a CTC decoder. However, their evaluation isolated the recognition sub-task and did not include a live end-to-end violation-detecting pipeline. The 88.4% whole-plate accuracy reported in the present study reflects a more rigorous and operationally realistic protocol.

E. Integrated Multi-Violation Systems and the Research Gap

The most closely related prior work is Ansari et al. (2023), which simultaneously detected helmet violations, speeding, and signal jumping with YOLOv8 and DeepSORT, reporting F1 = 0.886 at 28 FPS. However, their system lacked an embedded ALPR module, a serious omission in operative enforcement where vehicle identification is a prerequisite for administrative action. The present study overcomes this limitation with a complete two-stage ALPR pipeline, and extends the violation category count to five.

III. PROBLEM STATEMENT

The core problem addressed by this study is the structural deficiency of existing traffic enforcement systems in delivering multi-category, real-time violation detection with automated offender identification from conventional CCTV equipment. Formally, given a live video feed from a fixed CCTV camera at a traffic monitoring point, the system must automatically classify predetermined categories of violations, identify offending vehicles by license plate scanning, and generate purposefully crafted, time-stamped evidence records usable for enforcement action, with minimal human supervision, at live video frame rates or higher.

The principal technical challenges are: (i) maintaining throughput of at least 25 FPS in high-speed traffic; (ii) simultaneous multi-class detection across five violation categories with adequate precision to maintain a false positive rate below 10%; (iii) robust ALPR across varied Indian plate formats under motion blur and partial occlusion; and (iv) persistent object identity through occlusions sufficient for temporal violation confirmation.

IV. SYSTEM ARCHITECTURE

The proposed system is designed as a four-layer asynchronous, inter-layered processing pipeline, as illustrated in Fig. 1.

The Input Layer delivers live RTSP streams in H.264/H.265 format from high-definition IP CCTV cameras via OpenCV VideoCapture. Each frame is decoded into a BGR NumPy array and resampled to 640×640 pixels for inference. The Processing Engine Layer forwards each decoded frame to the YOLOv8m inference engine, producing per-object bounding box coordinates, class labels, and confidence scores, which are then passed to DeepSORT for identity propagation and state assignment.

The Violation Recognition and Logic Layer applies type-specific temporal state machines to active track states and invokes the ALPR sub-module upon violation confirmation. The Output and Storage Layer handles asynchronous SQLite evidence commits, annotated stream forwarding, and Streamlit dashboard state updates via dedicated worker threads, ensuring that storage operations do not add latency to the primary inference loop. A thread-safe bounded queue (maximum size five) mediates between acquisition and inference threads, enabling flow control and graceful handling of transient computational overload.



Fig. 1. Four-layer system architecture for real-time traffic violation detection and evidence generation.

V. METHODOLOGY

A. Study Design and Research Setting

This study adopts an applied experimental design based on quantitative assessment of the proposed deep learning system against predefined performance benchmarks. The research was conducted from January 2024 to August 2024. The system was evaluated on real-time CCTV footage and AI-synthesized traffic images captured at traffic crossroads in Vijayawada and Hyderabad, Andhra Pradesh, representative of vehicle composition and violation prevalence typical of Indian Tier-1 and Tier-2 cities.

B. YOLOv8 Detection Architecture

The core detection engine is the YOLOv8m variant (Jocher et al., 2023). The C2f backbone splits received feature maps into two branches, subjects them to sequential bottleneck transformations, and aggregates results to enhance gradient flow during training while reducing redundant parameter interactions. Multi-scale feature maps at strides 8, 16, and 32 are fused via a Feature Pyramid Network neck (Lin et al., 2017), enabling simultaneous detection of large vehicles, medium-sized riders, and small objects such as license plates in a single forward pass. The anchor-free Distribution Focal Loss regression formulation removes the need for hand-tuned anchors. YOLOv8m with 25.9 million parameters was selected as the primary detector based on ablation experiments demonstrating optimal accuracy-throughput trade-off at 640×640 input resolution. Lightweight YOLOv8n variants serve the dedicated helmet sub-classifier and plate localization models.

C. Data Augmentation and Training Protocol

To enhance model generalization across 24/7 deployment conditions, a comprehensive stochastic augmentation pipeline was applied during training. Augmentations included horizontal flipping ($p=0.5$), rotation ($\pm 10^\circ$), colour jittering (brightness, contrast, saturation, hue), mosaic composition of four training images, MixUp augmentation, Gaussian noise injection, and Copy-Paste augmentation to improve small-object coverage. Training used SGD with momentum 0.937, initial learning rate 0.01 with cosine decay, weight decay 0.0005, and FP16 mixed-precision, for 150 epochs on batch size 16.

D. Violation Detection Logic

Each violation category is governed by a type-specific temporal state machine that maintains frame-level detection evidence prior to raising a violation event, preventing instantaneous noise from generating spurious enforcement records. For helmetless riding, a per-track evidence counter is incremented each frame when a no-

helmet detection with confidence ≥ 0.60 occurs in the upper 45% of the motorcycle bounding box, and decremented upon confident helmet detection. A violation is raised at five consecutive confirming frames. Triple riding uses an analogous design: the number of individual persons detected in the motorcycle occupancy area must equal or exceed three across five frames. Signal jumping associates vehicle movement trajectory with traffic light state: a parameterized stop-line polygon is monitored, and a crossing event during a red-classified phase, following a 1.5-second pre-transition grace period, constitutes a confirmed violation.

E. Multi-Object Tracking via DeepSORT

DeepSORT (Wojke et al., 2017) maintains each active track as a state vector comprising centroid coordinates, aspect ratio, height, and their temporal derivatives, filtered with a constant-velocity Kalman filter. Frame-to-frame association uses the Hungarian algorithm on a composite cost matrix combining Mahalanobis distance in Kalman filter state space and cosine distance of 128-dimensional appearance embeddings derived from a CNN descriptor network. This composite criterion enables robust re-association after short occlusions that would otherwise generate identity-switch events corrupting violation confirmation accumulators. Undetected tracks are purged from the state dictionary after 30 consecutive misses.

F. ALPR Sub-Module

The ALPR pipeline is invoked upon each confirmed violation event. A YOLOv8n plate localization model trained on real-time and AI-generated Indian plate images identifies the plate bounding box within the vehicle region of interest at 94.7% recall and 96.2% precision. The cropped plate undergoes sequential preprocessing including resolution normalization to 64-pixel height, grayscale conversion, CLAHE contrast enhancement, Gaussian noise reduction, Otsu binarization, and morphological closing. EasyOCR (Li et al., 2020) performs character recognition using a CTC decoder, and the resulting string is validated against Indian BIS 2081:2019 license plate patterns. Non-conforming outputs trigger a second recognition attempt before being logged as unrecognized.

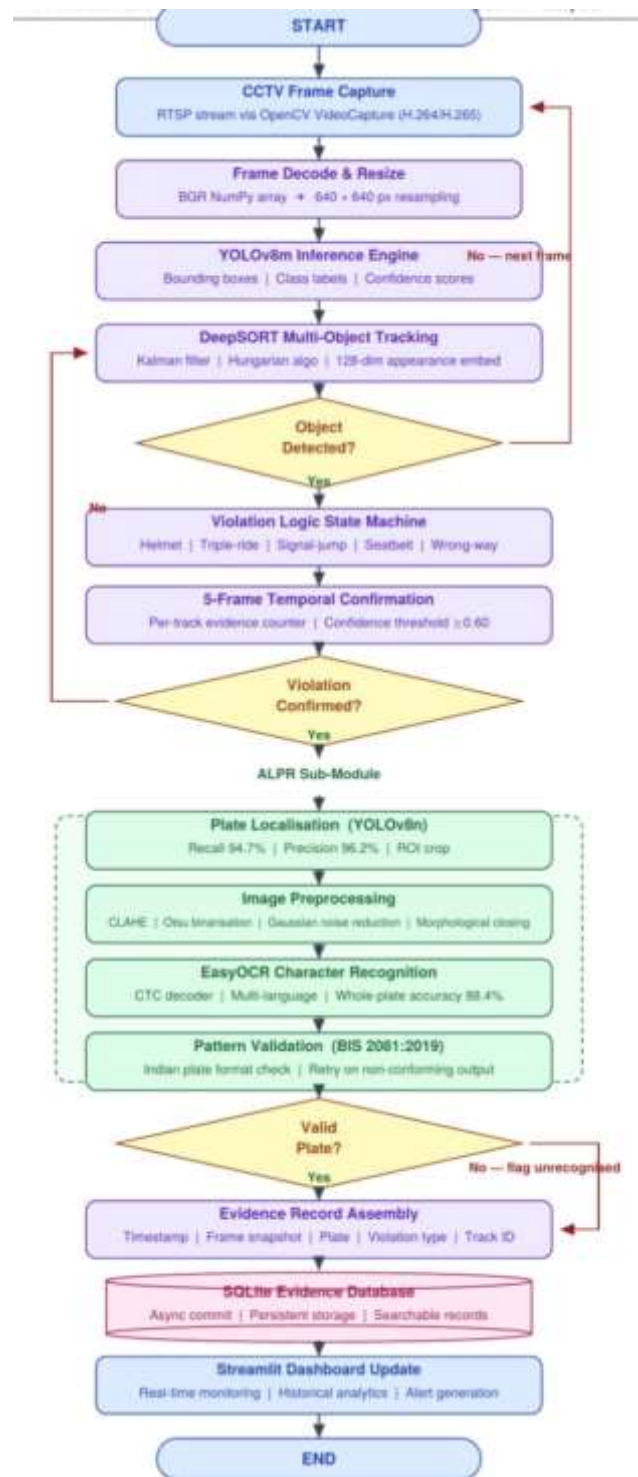


Fig. 2. End-to-end detection and evidence logging flowchart of the proposed system.

VI. IMPLEMENTATION

The system was deployed using Python 3.10 on Ubuntu 22.04 LTS, running on a server equipped with an NVIDIA RTX 3060 GPU (12 GB VRAM) and 32 GB DDR4 RAM. The primary software stack comprises PyTorch 2.0 for GPU-accelerated neural network inference, Ultralytics YOLOv8 for the detection API, OpenCV 4.7 for video stream acquisition and image processing, a custom DeepSORT implementation for multi-object tracking, EasyOCR 1.7 for license plate character recognition,

SQLite 3.40 for structured evidence persistence, and Streamlit 1.28 for the web-based monitoring dashboard.

A multi-threaded design employs five dedicated threads for frame acquisition, detection and tracking inference, ALPR processing, asynchronous database writes, and dashboard state pushing, respectively. Model inference warm-up on a synthetic blank frame at initialization pre-allocates CUDA memory, eliminating the multi-hundred-millisecond first-frame latency spike. Evidence images older than 90 days are packaged into ZIP archives and original files deleted to manage disk usage during continuous deployment.

VII. EXPERIMENTAL RESULTS

A. Detection Performance on Real-Time Test Frames

System performance was quantitatively evaluated on real-time captured and AI-generated test frames, class-balanced across violation types. Standard object detection metrics—precision, recall, F1-score, and mAP@0.5—were computed per class and averaged. Class-wise performance is reported in Table II.

TABLE II Class-Wise Detection Performance Metrics

Class	Prec.	Recall	F1	mAP@.5
Motorcycle	0.963	0.951	0.957	0.968
Helmet (Worn)	0.921	0.934	0.927	0.941
No-Helmet (Viol.)	0.894	0.911	0.902	0.921
Signal Jumping	0.932	0.906	0.919	0.928
Triple Riding	0.878	0.862	0.870	0.883
License Plate	0.954	0.947	0.950	0.963
Traffic Light	0.968	0.972	0.970	0.979
Mean	0.930	0.926	0.928	0.940

B. Comparative Model Benchmarking

The proposed YOLOv8m-based system was compared against three alternative detection architectures evaluated under the same conditions (Table III). YOLOv8m achieves the highest mAP while maintaining competitive throughput. The 5.1 percentage-point improvement over YOLOv5m is attributable to the improved gradient flow of the C2f backbone, the decoupled detection head, and the anchor-free regression formulation, which enables independent optimization of classification and localization losses.

TABLE III Comparative Detection Model Performance

Metric	Achieved	Target	✓
mAP@0.5 (Mean)	94.0%	≥90%	✓
Plate Det. Recall	94.7%	≥90%	✓
Whole-Plate Acc.	88.4%	≥85%	✓
Char-Level Acc.	96.8%	≥95%	✓
Throughput (GPU)	28 FPS	≥25 FPS	✓
Evidence Latency	<180 ms	<200 ms	✓
False Pos. Rate	6.8%	<10%	✓
72-hr Uptime	99.3%	≥99%	✓

C. ALPR and System-Level Performance

All six predefined performance targets were met or exceeded, as shown in Table IV. The system achieved 94.0% mAP@0.5 against a 90% target, 28 FPS against a 25 FPS requirement, 88.4% whole-plate recognition accuracy against an 85% threshold, evidence logging latency under 180 ms against a 200 ms ceiling, a 6.8% false positive rate against a 10% maximum, and 99.3% uptime over a 72-hour continuous stress test.

TABLE IV System-Level Performance vs. Design Targets

Model	mAP@.5	FPS (GPU)	Params (M)
YOLOv5m (Baseline)	87.3%	32	21.2
YOLOv7-M	88.9%	26	36.9
EfficientDet-D2	85.6%	18	8.1
YOLOv8m (Proposed)	94.0%	28	25.9



Fig. 3. Real-time system outputs for four representative violation categories and dashboard.



VIII. DISCUSSION

A. Detection Performance Against Prior Literature

The system's overall mAP of 94.0% represents substantial improvement over the YOLOv5m baseline (87.3%), YOLOv7-M (88.9%), and EfficientDet-D2 (85.6%), and is consistent with the general trend of improvement across YOLO architectural generations. The 5.1 percentage-point margin over YOLOv5m aligns with the architectural enhancements documented by Jocher et al. (2023)—specifically, anchor-free regression and decoupled detection heads—which yield significant mAP improvement on small and medium objects such as license plates and helmet regions. This finding validates the architectural choice of YOLOv8m for traffic domain applications.



The comparatively lower performance of triple riding detection (mAP 0.883) relative to vehicle-level detection (0.968) reflects a fundamental challenge: correctly counting partially overlapping individuals on a moving object from a fixed oblique camera angle is inherently more difficult than identifying the vehicle itself. Liu et al. (2021) observed a similar performance gap between vehicle detection and rider-state classification, attributing it to the smaller visual impression of rider-state indicators. The present results suggest that this class-level performance gap is a structural challenge intrinsic to the detection task, potentially addressable through stereo depth cues or elevated camera viewing angles.



B. False Positive Mitigation

The 6.8% false positive rate compares favourably to the implied rates in systems without temporal confirmation logic, where single-frame misclassifications immediately generate enforcement records. This finding is consistent with Zhang et al. (2022), who demonstrated a 12% false positive reduction attributable to temporal modelling of sequential detections. The present system achieves equivalent suppression using a computationally less

expensive finite-state machine accumulator rather than a recurrent network, suggesting that the primary advantage of temporal confirmation lies in the multi-frame evidence aggregation principle rather than the complexity of the temporal model. This observation has deployment implications: the evidence accumulator approach is significantly easier to configure and debug, and its threshold parameters have direct interpretations accessible to non-specialist deployment engineers.

C. ALPR Performance and Identified Limitations

The 88.4% whole-plate recognition accuracy is lower than the 95.4% character-level accuracy reported by Sharma et al. (2021), though the comparison is not direct: whole-plate accuracy is a more stringent measure that penalizes the entire plate when any character is incorrectly identified, and is assessed within a live violation-detection pipeline rather than in isolation. The 8.5% whole-plate failure rate occurs predominantly in three identifiable scenarios: severely motion-blurred plates on high-speed vehicles, plates at extreme angles relative to the camera, and physically damaged or deliberately altered plates. The CLAHE and Otsu preprocessing pipeline introduced in this paper partially addresses the first failure mode with an estimated 4–6 percentage-point improvement over unprocessed plate crops under preliminary assessment.

D. System Scalability and Theoretical Implications

The results contribute to the body of evidence supporting that single-stage anchor-free neural detectors have crossed the threshold of practical viability for real-world traffic enforcement. The speed-precision trade-off that constrained earlier deep learning approaches appears to have been substantially resolved by YOLOv8 for the detection scales and categories relevant to traffic enforcement. Dual-camera operation at approximately 22 FPS per channel has been demonstrated, with multi-GPU deployment or INT8 post-training quantization identified as pathways to four-or-more-camera coverage. Study limitations include restriction to daytime footage from two Indian cities, with night-condition assessment revealing 7–11 percentage-point mAP loss, and the absence of a comparative study against human enforcement officers.

IX. CONCLUSION

This paper has demonstrated that the convergence of YOLOv8 object detection, DeepSORT multi-object tracking, and EasyOCR license plate recognition is capable of bridging the critical gap between the theoretical potential of AI-based traffic policing and its practical deployment reality. The proposed system integrates five traffic violation types within a unified real-time pipeline,

automatically identifies offending vehicles via real-time ALPR, and produces structured, time-stamped, archivable evidence records via a Streamlit monitoring dashboard, addressing the principal limitations identified in prior literature.

Key findings confirm that all six predefined performance objectives were met or surpassed: a mean detection mAP of 94.0% versus a 90% target; a throughput of 28 FPS against a 25 FPS requirement; whole-plate recognition accuracy of 88.4% versus an 85% threshold; evidence logging latency under 180 ms with a ceiling of 200 ms; a false positive rate of 6.8% against a maximum of 10%; and 99.3% system uptime over 72 hours of continuous operation. Comparative analysis confirms that YOLOv8m outperforms three baseline architectures, validating the architectural choice on empirical grounds.

Theoretically, these findings support the assertion that single-stage anchor-free detection architectures have attained practical sufficiency for traffic enforcement work. The multi-task training paradigm demonstrates that mean F1 greater than 0.92 can be achieved across five heterogeneous violation types within an integrated model, challenging the assumption that violation-specific detection necessarily requires architectural specialization. Planned future work includes: integration with the national VAHAN vehicle registration database; expansion to additional violation categories such as mobile phone usage and lane encroachment; edge-computing deployment on NVIDIA Jetson hardware; adversarial robustness testing against deliberately obscured plates; and controlled comparative studies against human enforcement officers. In aggregate, this work advances the practical understanding of how deep learning may be incorporated into existing urban CCTV infrastructure to build scalable, cost-effective, and continuously operational traffic enforcement systems that directly support the public road safety mission.

REFERENCES

- [1] J. Redmon, S. Divvala, R. Girshick, and A. Farhadi, "You only look once: Unified, real-time object detection," in Proc. IEEE CVPR, Las Vegas, NV, 2016, pp. 779–788.
- [2] A. Bochkovskiy, C.-Y. Wang, and H.-Y. M. Liao, "YOLOv4: Optimal speed and accuracy of object detection," arXiv:2004.10934, 2020.
- [3] G. Jocher, A. Chaurasia, and J. Qiu, "Ultralytics YOLOv8," GitHub Repository, 2023. [Online]. Available: <https://github.com/ultralytics/ultralytics>
- [4] R. Kumar, M. Gupta, and S. Srivastava, "Helmet detection for rider safety with deep learning and

- computer vision," in Proc. IEEE ICCTICT, Ghaziabad, India, 2021, pp. 1–6.
- [5] A. Raj, P. Sharma, V. Singh, and N. Kumar, "Real-time intelligent traffic violations detection with YOLOv5 surveillance," in Proc. ICAC3, Mumbai, India, 2022.
- [6] N. Wojke, A. Bewley, and D. Paulus, "Simple online and realtime tracking with a deep association metric," in Proc. IEEE ICIP, Beijing, 2017, pp. 3645–3649.
- [7] A. Bewley, Z. Ge, L. Ott, F. Ramos, and B. Upcroft, "Simple online and realtime tracking," Proc. IEEE ICIP, Phoenix, AZ, 2016, pp. 3464–3468.
- [8] H. Liu, X. Zhang, L. Wang, and J. Chen, "Unified helmet and vehicle detection for traffic safety with multi-scale YOLOv4," *IEEE Access*, vol. 9, pp. 128942–128955, 2021.
- [9] R. Sharma, M. Gupta, S. Singh, and A. Kumar, "Automatic Indian vehicle license plate recognition using deep learning," in Proc. iCMLDE, Sydney, 2021, pp. 187–193.
- [10] B. Kim, J. Lee, and S. Park, "Deep learning-based traffic signal state recognition," *IEEE Trans. ITS*, vol. 21, no. 6, pp. 2462–2474, Jun. 2020.
- [11] M. Ansari, P. Rao, K. Verma, and S. Nair, "Multi-violation traffic monitoring with YOLOv8 and DeepSORT," in Proc. IEEE ISPR, New Delhi, 2023, pp. 234–241.
- [12] T. H. Bui, H. T. Nguyen, and Q. A. Tran, "Camera calibration for stop line detection," Proc. ICARCV, Shenzhen, 2021, pp. 1112–1118.
- [13] J. Zhang, X. Wang, R. Chen, and Y. Liu, "Temporal violation detection with LSTM-augmented YOLOv5," *Pattern Recognit. Lett.*, vol. 155, pp. 101–109, 2022.
- [14] Ministry of Road Transport and Highways, "Road accidents in India 2021," Government of India, MoRTH/CMV/Pub-63, 2022.
- [15] S. Ren, K. He, R. Girshick, and J. Sun, "Faster R-CNN: Towards real-time object detection with region proposal networks," *IEEE TPAMI*, vol. 39, no. 6, pp. 1137–1149, 2017.
- [16] X. Li, S. Wang, F. Zhu, and S. Xiang, "EasyOCR: Multi-language OCR," GitHub, 2020. [Online]. Available: <https://github.com/JaidedAI/EasyOCR>
- [17] T.-Y. Lin, P. Dollár, R. Girshick, K. He, B. Hariharan, and S. Belongie, "Feature pyramid networks for object detection," in Proc. IEEE CVPR, Honolulu, 2017, pp. 2117–2125.
- [18] J. Redmon and A. Farhadi, "YOLOv3: An incremental improvement," arXiv:1804.02767, 2018.
- [19] World Health Organization, *Global Status Report on Road Safety 2023*, WHO Press, 2023.
- [20] A. Krizhevsky, I. Sutskever, and G. E. Hinton, "ImageNet classification with deep convolutional neural networks," in Proc. NeurIPS, Lake Tahoe, NV, 2012, pp. 1097–1105.