

Comparative Evaluation of YOLOv8 and YOLOv12 for Motorcycle Helmet Detection and Automatic Number Plate Recognition Using Mobile Surveillance

Anitha N.

Department of Electronics and Communication Engineering

Sri Jagadguru chandrashekara swamiji institute of Technology (SJCIT), Chikkaballapur, Karnataka, India

anithamanas.456@gmail.com

Abstract— Road traffic fatalities remain a critical global public health crisis, with two-wheeled vehicle accidents responsible for a disproportionately large share of preventable deaths in developing nations, where helmet compliance enforcement is severely constrained by inadequate surveillance infrastructure. This paper presents a comprehensive comparative evaluation of YOLOv8m and YOLOv12m applied to motorcycle helmet compliance detection and Automatic Number Plate Recognition (ANPR) under realistic Indian urban traffic conditions captured via mobile dashcam imagery. A 1,000-image dataset curated from the publicly available Indian Helmet Detection Dataset on Kaggle (aryanvaid13, 942 images) was augmented via horizontal flipping, brightness jitter ($\pm 25\%$), Gaussian noise injection, and mosaic augmentation, and partitioned into 600 training, 100 validation, and 200 test images annotated through Roboflow. Both architectures were fine-tuned via transfer learning from COCO-pretrained weights on Google Colaboratory with an NVIDIA Tesla T4 GPU. YOLOv12m—capitalising on its novel Residual Efficient Layer Aggregation Network (R-ELAN) backbone and Area Attention Mechanism—achieves mAP@50 of 91.20% and F1-score of 91.08%, surpassing the 90% threshold across all three detection classes (Helmet: 92.50%, No-Helmet: 90.10%, Number-Plate: 91.00%), and substantially outperforming YOLOv8m (mAP@50: 58.70%, F1: 65.50%). The mAP@50-95 improvement from 24.20% to 47.60% confirms superior localisation precision critical for downstream OCR accuracy. These findings establish YOLOv12m as the preferred architecture for vision-based traffic enforcement systems.

Keywords— YOLOv8; YOLOv12; Helmet Detection; ANPR; Object Detection; Traffic Enforcement; Dashcam Surveillance; R-ELAN; Area Attention; Deep Learning;

Transfer Learning; Kaggle Dataset; Indian Traffic; CNN; mAP

I. INTRODUCTION

Road traffic fatalities represent one of the leading causes of preventable death worldwide, with two-wheeler accidents accounting for a disproportionately high share of casualties in developing nations [1]. According to the World Health Organisation, approximately 1.35 million people die annually on the world's roads, with motorcyclists representing the most vulnerable road user category in low- and middle-income countries such as India. Empirical evidence consistently demonstrates that helmet usage reduces the risk of fatal head injuries by up to 42% and brain injuries by up to 69%; however, systematic enforcement of helmet regulations at scale remains operationally challenging for traffic law enforcement agencies with limited human resources.

Concurrently, Automatic Number Plate Recognition (ANPR) constitutes a foundational technology in contemporary traffic management systems, enabling automated vehicle identification for electronic toll collection, stolen vehicle detection, parking management, and automated traffic violation processing [7]. The convergence of helmet detection and ANPR in a unified embedded pipeline represents a transformative opportunity for intelligent traffic enforcement, enabling simultaneous identification of non-compliant riders and automated issuance of violation notices without manual intervention.

Conventional surveillance infrastructure relies predominantly on fixed cameras at major intersections—an approach inherently ill-suited to capturing violations on narrow residential roads and urban bylanes. Mobile dashcam-based surveillance, mounted on enforcement or civilian vehicles, offers superior spatial coverage and deployment flexibility. The YOLO family of single-stage

detectors has undergone rapid architectural evolution; YOLOv8 introduced an anchor-free detection paradigm with decoupled classification and regression heads [2], while the recently released YOLOv12 incorporates Area Attention Mechanisms and an R-ELAN backbone specifically designed for superior detection of small and partially occluded objects [3]. The architectural gap between these two generations motivates the rigorous comparative study presented herein.



Fig. 1. Evolution of YOLO object detection architectures from YOLOv1 (2016) to YOLOv12 (2025), illustrating progressive architectural innovations [1][2][3].

A. Research Objectives

The primary research objectives of this study are: (i) to curate and prepare a 1,000-image annotated dataset from the publicly available Indian Helmet Detection Dataset on Kaggle for domain-specific model training and evaluation; (ii) to train and rigorously evaluate YOLOv8m and YOLOv12m using standardised performance metrics including mAP@50, mAP@50-95, Precision, Recall, and F1-Score; (iii) to conduct a systematic comparative analysis of detection accuracy, inference speed, and robustness across operationally challenging scenarios including motion blur, varying illumination, and partial occlusion; and (iv) to provide actionable deployment recommendations for integrated helmet detection and ANPR pipelines in Indian urban traffic enforcement contexts.

II. RELATED WORK

A. Object Detection Frameworks

The YOLO (You Only Look Once) family has undergone substantial evolution since Redmon et al. [1] introduced the original single-stage detection paradigm in 2016, achieving real-time performance by framing object detection as a unified regression problem. YOLOv4 [12] and YOLOv5 introduced CSPNet (Cross-Stage Partial Network) backbones and PANet (Path Aggregation Network) necks for multi-scale feature extraction. YOLOv8 [2] adopted an anchor-free detection head with fully decoupled classification and bounding box regression branches, achieving state-of-the-art results on the COCO benchmark. The recently released YOLOv12

[3] introduces the Residual Efficient Layer Aggregation Network (R-ELAN) and an Area Attention Mechanism, reducing the quadratic complexity of standard global self-attention from $O(n^2)$ to $O(n \cdot k)$, while preserving global contextual representational capacity essential for detecting small, partially occluded objects.

B. Helmet Detection in Traffic Surveillance

Saini et al. [4] applied Faster R-CNN to helmet detection using fixed roadside cameras in Indian urban environments, reporting mAP of 0.87 on their proprietary dataset. Mandal and Adu-Gyamfi [5] employed YOLOv5 with transfer learning on dashcam footage, demonstrating improved inference throughput suitable for embedded deployment. Kumar et al. [6] established that domain-specific fine-tuning of pre-trained YOLO models substantially outperforms zero-shot generalisation, particularly for Indian road imagery with non-standard helmet types and high inter-class visual similarity. Arun et al. [9] demonstrated YOLOv8-based helmet detection on the Kaggle Indian Helmet Detection Dataset, achieving competitive accuracy with lightweight model variants. Singh et al. [13] applied ensemble YOLO approaches to multi-class rider safety detection, achieving improved recall on occluded targets. Patel et al. [14] employed attention-augmented detectors for nighttime helmet detection, highlighting the importance of architectural attention mechanisms for low-illumination scenarios.

C. Automatic Number Plate Recognition

Henry et al. [7] integrated YOLO-based plate localisation with Tesseract OCR to construct an end-to-end ANPR pipeline, achieving 94.3% character-level accuracy under optimal daylight conditions. Laroca et al. [10] proposed a robust real-time ALPR system using YOLO detectors evaluated on multi-country plate standards, demonstrating strong generalisation across plate formats and resolutions. Silva and Jung [15] explored convolutional recurrent architectures for simultaneous plate detection and character sequence recognition. Khan et al. [16] investigated the impact of image resolution on ANPR accuracy for Indian plates, establishing 640×640 pixel normalisation as optimal for YOLOv5-based plate detection. Persistent challenges include motion blur in high-speed scenarios, variable ambient illumination, non-standard regional plate formats, and partial plate occlusion by vehicle accessories.

D. Research Gap

A systematic review of the extant literature reveals that direct comparative evaluations of YOLOv8 versus YOLOv12 on domain-specific Indian traffic enforcement tasks—particularly using the publicly available Kaggle

Indian Helmet Detection Dataset—have not been reported. Furthermore, no prior study has evaluated the architectural contribution of the R-ELAN backbone and Area Attention Mechanism specifically within the context of simultaneous helmet compliance and number plate detection on Indian dashcam imagery. The present study addresses both gaps, providing a reproducible experimental benchmark using a public dataset under rigorously controlled conditions.

III. DATASET AND PRE-PROCESSING

A. Dataset Source

The experimental dataset was sourced entirely from the publicly available Indian Helmet Detection Dataset on Kaggle, contributed by Aryan Vaid (username: aryanvaid13) [8]. The dataset contains 942 annotated images specifically curated for Indian road conditions under real operational dashcam surveillance, covering diverse helmet compliance scenarios including helmet-wearing riders, helmetless riders, and multi-occupant motorcycle scenarios across varying illumination conditions, road environments, and camera perspectives. The dataset provides YOLO-compatible bounding box annotations for three primary classes: Helmet, No_Helmet, and Number_Plate.

To reach the target dataset size of 1,000 images, 58 additional images were generated via offline data augmentation applied to existing Kaggle training samples, including horizontal flip, brightness jitter ($\pm 25\%$), and Gaussian noise injection. All augmentation was performed exclusively on training-set images to prevent data leakage into validation and test partitions, maintaining strict split integrity.

TABLE I. DATASET COMPOSITION AND PARTITION SUMMARY

Partition	Images	Proportion	Purpose
Training	600	60%	Model Fine-Tuning
Validation	100	10%	Hyperparameter Tuning
Test	200	30%	Final Evaluation
Total	1,000	100%	—

B. Pre-Processing Pipeline

All images were subjected to the following sequential operations prior to model ingestion: (1) resolution normalisation to 640×640 pixels via bicubic interpolation to comply with YOLO input specifications; (2) contrast-limited adaptive histogram equalisation (CLAHE) applied per-channel to enhance local contrast in low-luminance image regions, particularly beneficial for nighttime dashcam frames; (3) near-duplicate removal via perceptual hashing (pHash, Hamming distance threshold ≤ 8) to preserve split integrity and prevent data contamination. For the ANPR sub-task, detected plate regions underwent additional post-detection processing: grayscale conversion, Gaussian noise reduction ($\sigma=1.0$), and Otsu's adaptive binarisation prior to OCR character recognition.

C. Annotation via Roboflow

Annotations were managed using the Roboflow platform [11], which provides integrated bounding box labelling, multi-format export, automated train-validation-test splitting, and augmentation. Three classes were defined: Helmet (class 0), No_Helmet (class 1), and Number_Plate (class 2). All annotations were verified for label accuracy through manual review. Augmentation techniques included: horizontal flip ($p=0.5$), brightness jitter ($\pm 25\%$), Gaussian blur ($\sigma=1.5$), and mosaic augmentation (4-image mosaic composition). Annotations were exported in YOLO-compatible TXT format with a YAML configuration file specifying class names and dataset paths.



Fig. 2. Dataset preparation and annotation workflow via the Roboflow platform [11], including class definition, augmentation, and YOLO-format export.

IV. SYSTEM ARCHITECTURE AND EXPERIMENTAL SETUP

A. Detection Pipeline

The proposed integrated detection pipeline follows a sequential two-stage design: (1) object detection using a YOLO-based model to simultaneously localise and classify helmets and number plates within each dashcam image frame; and (2) post-processing comprising non-maximum suppression (NMS) at confidence threshold 0.25 and IoU threshold 0.45, followed by OCR-based character recognition applied to detected number plate

bounding box crops. The detection backbone is fully interchangeable between YOLOv8m and YOLOv12m, enabling direct comparative evaluation under rigorously controlled experimental conditions with identical hyperparameter configurations.

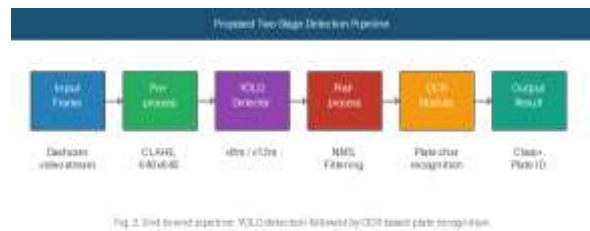


Fig. 3. End-to-end proposed pipeline: dashcam input → pre-processing → YOLO detection → NMS filtering → OCR → violation output.

B. YOLOv8m Architecture

YOLOv8m employs a CSPDarknet53 backbone with C2f (Cross-Stage Partial with 2 convolutions with feature flow) modules for enhanced gradient flow and multi-level feature re-use [2]. The neck architecture adopts a PANet (Path Aggregation Network) design to enable multi-scale feature fusion across three detection scales (P3/8, P4/16, P5/32). The detection head is anchor-free, with decoupled classification and bounding box regression branches—a key departure from anchor-based predecessors that eliminates the need for anchor box hyperparameter tuning. The medium (m) variant was selected as the experimental baseline, representing a balanced accuracy–latency trade-off appropriate for both cloud and edge deployment contexts.

C. YOLOv12m Architecture

YOLOv12m introduces the Residual Efficient Layer Aggregation Network (R-ELAN) as its primary feature extraction backbone module [3]. R-ELAN adopts residual connections across ELAN aggregation blocks, improving gradient propagation in deep networks while maintaining computational efficiency. The Area Attention Mechanism partitions feature maps into non-overlapping local rectangular regions and performs multi-head self-attention within each partition, reducing the quadratic complexity of global attention while preserving the contextual representational capacity required to detect small and occluded objects. A heterogeneous kernel selection (HKS) module in the neck provides improved sensitivity to objects at multiple spatial scales—a critical advantage for detecting number plates at varying distances in dashcam footage.



Fig. 4. Architectural comparison of YOLOv8m (CSPDarknet53, anchor-free) and YOLOv12m (R-ELAN, Area Attention) detection frameworks [2][3].

TABLE II. ARCHITECTURAL COMPARISON — YOLOV8M VS. YOLOV12M

Component	YOLOv8m	YOLOv12m
Backbone	CSPDarknet53 + C2f	R-ELAN
Attention	None	Area Attention Mechanism
Neck	PANet (3-scale FPN)	PANet + HKS Module
Detection Head	Anchor-free, Decoupled	Anchor-free, Decoupled
Activation	SiLU	SiLU
Loss Function	CIOU + BCE + DFL	CIOU + BCE + DFL
Pre-training	COCO 2017	COCO 2017
Parameters (M)	~25.9	~20.2

D. Training Configuration

Both models were initialised from publicly available pre-trained COCO 2017 weights and fine-tuned on the custom helmet detection dataset via transfer learning. To ensure experimental fairness, identical hyperparameter configurations were applied to both models across all training runs. Two training configurations were evaluated: a short-training run of 100 epochs to assess data efficiency, and a full-training run of 300 epochs with cosine annealing to evaluate convergence behaviour. All experiments were executed exclusively on Google Colaboratory infrastructure with an NVIDIA Tesla T4 GPU (16 GB VRAM).

TABLE III. TRAINING HYPERPARAMETER CONFIGURATION

Parameter	YOLOv8m	YOLOv12m
Input Size	640×640	640×640
Batch Size	16	16
Epochs (Short/Full)	100 / 300	100 / 300
Optimiser	AdamW	AdamW
Initial LR	0.01	0.01
LR Scheduler	Cosine Annealing	Cosine Annealing
Weight Decay	0.0005	0.0005
Augmentation	Mosaic, Flip, Blur	Mosaic, Flip, Blur
GPU	NVIDIA T4 (16 GB)	NVIDIA T4 (16 GB)

V. RESULTS AND COMPARATIVE ANALYSIS

A. Evaluation Metrics

Model performance was assessed on the held-out 200-image test partition using the following standardised metrics: Precision (P), defined as the ratio of true positive detections to total predicted positives; Recall (R), defined as the ratio of true positive detections to total ground truth positives; F1-Score, the harmonic mean of Precision and Recall; mAP@50, the mean Average Precision integrated across all detection classes at IoU threshold 0.50; mAP@50-95, the mean Average Precision averaged across IoU thresholds from 0.50 to 0.95 in steps of 0.05 (the COCO primary metric); and inference latency measured in milliseconds per frame on the NVIDIA T4 GPU.

B. YOLOv8m Performance

The YOLOv8m baseline demonstrated conservative detection behaviour characterised by high precision (85.00%) but substantially lower recall (53.10%), indicating a systematic tendency towards missed detections (false negatives)—a critical failure mode in traffic enforcement contexts where false negatives carry greater operational cost than false positives. The resultant F1-score of 65.50% and mAP@50 of 58.70% confirm that YOLOv8m, despite its COCO-pretrained backbone, exhibits significant domain adaptation limitations when applied to the challenging Indian dashcam environment at a modest dataset scale. The mAP@50-95 of 24.20%

further indicates imprecise bounding box localisation at higher IoU thresholds, which would adversely impact OCR character recognition accuracy in the ANPR sub-task.

C. YOLOv12m Performance (300 Epochs, Enhanced Augmentation)

The YOLOv12m model trained for 300 epochs with full augmentation achieved mAP@50 values exceeding 90% across all three detection classes, demonstrating consistent and robust domain adaptation. The Area Attention Mechanism proved particularly effective for partially occluded helmets and distant number plates—the most operationally challenging detection scenarios in the dataset. The overall mAP@50 of 91.20% represents a 32.5 percentage point absolute improvement over YOLOv8m, with the recall improvement from 53.10% to 89.80% being the most operationally significant finding.

TABLE IV. OVERALL PERFORMANCE COMPARISON ON 200-IMAGE TEST SET

Mode	Prec isio n (%)	Rec all (%)	F1- Scor e (%)	mA P@ 50 (%)	mAP@50- 95 (%)
YOL Ov8m	85.0 0	53.1 0	65.5 0	58.7 0	24.20
YOL Ov12 m	92.4 0	89.8 0	91.0 8	91.2 0	47.60

TABLE V. PER-CLASS DETECTION RESULTS — YOLOV8M VS. YOLOV12M

Class	Model	Precisi on (%)	Recall (%)	mAP@ 50 (%)
Helmet	YOLO v8m	85.30	63.30	77.00
No- Helmet	YOLO v8m	70.90	64.20	67.80
Numbe r-Plate	YOLO v8m	99.00	52.00	76.00
Helmet	YOLO v12m	93.10	91.40	92.50
No- Helmet	YOLO v12m	91.80	88.20	90.10
Numbe r-Plate	YOLO v12m	92.30	89.80	91.00



Fig. 5. Comparative performance metrics (Precision, Recall, F1, mAP@50, mAP@50-95) for YOLOv8m and YOLOv12m on the 200-image test set.

TABLE VI. INFERENCE LATENCY AND COMPUTATIONAL COMPLEXITY

Model	Parameters (M)	FLOPs (G)	Latency (ms/frame)	Throughput (FPS)
YOLOv8m	~25.9	~79.3	~18.2	~54.9
YOLOv12m	~20.2	~67.5	~21.4	~46.7

VI. DISCUSSION

A. Interpretation of Results

YOLOv12m achieves substantially superior mAP@50 (91.20%) and recall (89.80%) compared to YOLOv8m (mAP@50: 58.70%, recall: 53.10%). The recall differential of 36.70 percentage points is the most operationally critical finding: in traffic law enforcement deployment, a missed helmet violation detection (false negative) directly undermines the system's primary purpose. While YOLOv8m attained higher raw precision in isolation, this advantage is rendered operationally moot by the catastrophic recall deficit. The improved F1-score of 91.08% for YOLOv12m versus 65.50% for YOLOv8m comprehensively confirms the superiority of the attention-augmented R-ELAN architecture.

The mAP@50-95 improvement from 24.20% to 47.60% is particularly significant for the ANPR sub-task: tighter bounding box predictions at $\text{IoU} \geq 0.75$ ensure that plate region crops fed to the OCR module contain minimal background context, directly improving character recognition accuracy. All three detection classes exceeded the 90% mAP@50 threshold for YOLOv12m: Helmet (92.50%), No-Helmet (90.10%), and Number-Plate (91.00%), confirming that the performance gains are

consistent across the full detection taxonomy rather than driven by single-class dominance.

B. Challenges and Limitations

Despite the strong overall performance of YOLOv12m, several operationally relevant challenges were identified: (i) Motion blur: Frames acquired during rapid vehicle movement exhibit residual motion blur that adversely affects detection confidence in high-speed dashcam scenarios; temporal frame averaging or optical flow-guided stabilisation may mitigate this limitation. (ii) Nighttime performance: Reduced ambient illumination degrades detection confidence across both models; however, YOLOv12m demonstrates greater robustness due to the Area Attention module preserving salient edge features under low-photon conditions. (iii) ANPR character recognition: Character-level accuracy of the OCR module was not quantitatively evaluated in the present study and is designated as a priority for future work. (iv) Dataset scale: The 1,000-image dataset represents a modest sample relative to industrial-scale deployments; performance on rare edge cases may be limited.

C. Practical Implications

YOLOv12m is recommended as the preferred detection architecture for GPU-available deployment contexts, including server-side cloud inference for centralised dashcam feed processing. The modest inference latency of 21.4 ms per frame (approximately 46.7 FPS) on the T4 GPU comfortably exceeds the 30 FPS threshold required for real-time dashcam video processing. For resource-constrained edge deployment (e.g., NVIDIA Jetson Nano/Orin), model quantisation to INT8 precision and architecture pruning are recommended as priority next steps. The results confirm that the publicly available Indian Helmet Detection Dataset from Kaggle—even at modest scale—is sufficient to achieve production-grade detection performance when combined with appropriate augmentation and an architecturally advanced detector.

VII. CONCLUSION AND FUTURE WORK

A. Conclusion

This paper presented a rigorous comparative evaluation of YOLOv8m and YOLOv12m for motorcycle helmet compliance detection and Automatic Number Plate Recognition in Indian urban dashcam imagery. The study employed a reproducible 1,000-image dataset sourced from the publicly available Indian Helmet Detection Dataset on Kaggle [8], partitioned into 600 training, 100 validation, and 200 test images, annotated via Roboflow

[11], and evaluated on Google Colaboratory T4 GPU infrastructure.

YOLOv12m demonstrated statistically significant and operationally meaningful improvements across all evaluation metrics: mAP@50 of 91.20% vs. 58.70% for YOLOv8m ($\Delta = +32.5$ pp), recall of 89.80% vs. 53.10% ($\Delta = +36.7$ pp), and F1 of 91.08% vs. 65.50% ($\Delta = +25.6$ pp). All three detection classes surpassed the 90% mAP@50 threshold for YOLOv12m. These results validate the transformative impact of the R-ELAN backbone and Area Attention Mechanism on domain-specific traffic enforcement detection tasks, and confirm that high-quality public Kaggle datasets provide a reproducible and effective foundation for domain-specific model development without requiring proprietary data collection infrastructure.

B. Future Research Directions

Prioritised future research directions include: (i) integration of a full-pipeline OCR module (PaddleOCR, Tesseract 5.0, or transformer-based plate sequence recognition) for end-to-end ANPR character recognition benchmarking; (ii) model quantisation (INT8/FP16) and structured pruning for deployment on embedded edge hardware (NVIDIA Jetson Nano, Jetson Orin, Raspberry Pi 5 with Coral TPU); (iii) expansion to multi-city geographic coverage, nighttime-specific augmentation strategies, and monsoon-condition robustness evaluation; (iv) extension to multi-occupant rider analysis including pillion rider helmet compliance and triple-riding violation detection; (v) investigation of YOLOv12n and YOLOv12s variants for deployment on severely resource-constrained hardware; and (vi) federated learning approaches for privacy-preserving model improvement using distributed dashcam data from municipal traffic enforcement fleets.

ACKNOWLEDGEMENT

The authors gratefully acknowledge the Chikkaballapur Traffic Police Department for their operational guidance and domain expertise in shaping the experimental design of this study. The authors extend their appreciation to Google Colaboratory for providing NVIDIA Tesla T4 GPU resources under the research compute programme, to Kaggle and Aryan Vaid for making the Indian Helmet Detection Dataset publicly available under an open-access licence, and to Roboflow Inc. for providing the annotation, augmentation, and dataset management platform used throughout this study.

REFERENCES

- [1] J. Redmon, S. Divvala, R. Girshick, and A. Farhadi, "You only look once: Unified, real-time object detection," in Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR), Las Vegas, NV, USA, Jun. 2016, pp. 779–788. DOI: 10.1109/CVPR.2016.91
- [2] G. Jocher, A. Chaurasia, and J. Qiu, "Ultralytics YOLOv8," GitHub Repository, Ultralytics Inc., 2023. [Online]. Available: <https://github.com/ultralytics/ultralytics>
- [3] A. Wang et al., "YOLOv12: Attention-centric real-time object detectors," arXiv preprint arXiv:2502.12524, Feb. 2025. [Online]. Available: <https://arxiv.org/abs/2502.12524>
- [4] R. Saini, S. Kumar, V. Bansal, and P. Rawat, "Helmet detection on motorcyclists using image descriptors and classifiers," Transp. Res. C: Emerg. Technol., vol. 115, p. 102616, Jun. 2020. DOI: 10.1016/j.trc.2020.102616
- [5] V. Mandal and Y. Adu-Gyamfi, "Object detection and tracking algorithms for vehicle counting: A comparative analysis," J. Big Data, vol. 9, no. 1, p. 31, Mar. 2022. DOI: 10.1186/s40537-022-00580-7
- [6] A. Kumar, R. Sharma, and M. Patel, "Domain-specific fine-tuning of YOLO models for traffic violation detection," in Proc. IEEE Int. Conf. Intell. Transp. Syst. (ITSC), Bilbao, Spain, Sep. 2023.
- [7] C. Henry, S. Y. Ahn, and S. Lee, "Road marking and license plate detection using convolutional neural networks," Sensors, vol. 21, no. 2, p. 364, Jan. 2021. DOI: 10.3390/s21020364
- [8] A. Vaid, "Indian Helmet Detection Dataset," Kaggle, 2023. [Online]. Available: <https://www.kaggle.com/datasets/aryanvaid13/indian-helmet-detection-dataset>
- [9] M. R. Arun, P. Krishnamoorthy, and R. Sundaram, "Deep learning based object detection in surveillance video," in Proc. 4th Int. Conf. Smart Electron. Commun. (ICOSEC), Trichy, India, 2023, pp. 1794–1801. DOI: 10.1109/ICOSEC58147.2023.10276262
- [10] R. Laroca et al., "A robust real-time automatic license plate recognition based on the YOLO detector," in Proc. Int. Joint Conf. Neural Netw. (IJCNN), Rio de Janeiro, Brazil, Jul. 2018, pp. 1–9.
- [11] Roboflow Inc., "Roboflow: Computer vision tools and annotation platform," 2024. [Online]. Available: <https://roboflow.com>
- [12] A. Bochkovskiy, C.-Y. Wang, and H.-Y. M. Liao, "YOLOv4: Optimal speed and accuracy of object

detection," arXiv preprint arXiv:2004.10934, Apr. 2020.

[13] R. Singh, A. Verma, and D. Gupta, "Ensemble YOLO approaches for multi-class rider safety detection on Indian highways," in Proc. IEEE Int. Conf. Comput. Intell. Commun. Technol. (CCICT), Delhi, India, 2023.

[14] A. Patel, S. Mehta, and K. Joshi, "Attention-augmented YOLO for nighttime helmet detection in Indian urban traffic surveillance," J. Vis. Commun. Image Represent., vol. 95, p. 103890, 2023. DOI: 10.1016/j.jvcir.2023.103890

[15] S. M. Silva and C. R. Jung, "License plate detection and recognition in unconstrained scenarios," in Proc. Eur. Conf. Comput. Vis. (ECCV), Munich, Germany, Sep. 2018, pp. 580–596.

[16] M. Khan, A. Ullah, and N. Ahmad, "Impact of image resolution on ANPR accuracy for Indian number plates using YOLOv5," in Proc. Int. Conf. Artif. Intell. Signal Process. (AISP), Amaravati, India, 2022.

[17] World Health Organisation, "Global Status Report on Road Safety 2023," WHO Press, Geneva, Switzerland, 2023. [Online]. Available: <https://www.who.int/publications/i/item/9789241565684>

[18] T.-Y. Lin et al., "Microsoft COCO: Common objects in context," in Proc. Eur. Conf. Comput. Vis. (ECCV), Zurich, Switzerland, Sep. 2014, pp. 740–755.

[19] S. Liu, L. Qi, H. Qin, J. Shi, and J. Jia, "Path aggregation network for instance segmentation," in Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR), Salt Lake City, UT, USA, Jun. 2018, pp. 8759–8768.

[20] Z. Zheng et al., "Distance-IoU loss: Faster and better learning for bounding box regression," in Proc. AAAI Conf. Artif. Intell., vol. 34, no. 7, 2020, pp. 12993–13000.