DashCop: Automated E-ticket Generation for Two-Wheeler Traffic Violations Using Dashcam Videos

Deepti Rawat*, Keshav Gupta*, Aryamaan Basu Roy, Ravi Kiran Sarvadevabhatla
International Institute of Information Technology, Hyderabad, India
deepti.rawat@research.iiit.ac.in, keshav.gupta@students.iiit.ac.in, basuroyaryamaan@gmail.com, ravi.kiran@iiit.ac.in

Abstract

Motorized two-wheelers are a prevalent and economical means of transportation, particularly in the Asia-Pacific region. However, hazardous driving practices such as triple riding and non-compliance with helmet regulations contribute significantly to accident rates. Addressing these violations through automated enforcement mechanisms can enhance traffic safety. In this paper, we propose DashCop, an end-to-end system for automated E-ticket generation. The system processes vehicle-mounted dashcam videos to detect two-wheeler traffic violations. Our contributions include: (1) a novel Segmentation and Cross-Association (SAC) module to accurately associate riders with their motorcycles, (2) a robust cross-association-based tracking algorithm optimized for the simultaneous presence of riders and motorcycles, and (3) the RideSafe-400 dataset, a comprehensive annotated dashcam video dataset for triple riding and helmet rule violations. Our system demonstrates significant improvements in violation detection, validated through extensive evaluations on the RideSafe-400 dataset. Project page: https://dash-cop.github.io/

1. Introduction

Motorized two-wheelers have been widely used as an affordable and effective commuting choice, especially in the Asia-Pacific region [3, 24, 26, 27]. However, hazardous driving habits, such as not wearing helmets, exceeding passenger limits, and wrong-side driving, often pose significant safety risks. Annually, more than half a million two-wheeler accidents are reported in this region [39]. This alarming statistic underscores the need for promoting helmet compliance and deterring triple riding by imposing penalties. Such measures aim to curb dangerous driving practices, ensure rider safety, and maximize overall traffic safety goals [6, 48].

To address these safety concerns, there is a critical need for efficient enforcement mechanisms. Automated electronic traffic ticket (E-ticket) systems offer a promising solution by automating the detection of specific violations and using automatic number plate recognition (ANPR) to generate E-tickets [6, 36, 52]. This automation reduces reliance on manual enforcement, which is particularly beneficial in countries with a low ratio of traffic police to vehicles [10, 20, 41].

Detecting the violation of interest is an essential component of these E-ticket generation systems. Many works have individually addressed automating the detection of triple riding and helmet rule violations [2, 15, 17, 21, 50, 53, 54, 57]. These works process video frames from statically mounted cameras (*e.g.* CCTV). Despite the advantage of continuous monitoring, their scope is limited in terms of coverage over the road network and high installation costs. Moreover, the fixed and predictable positions of the cameras lack the element of unpredictability that is crucial for effective deterrence.

Vehicle-mounted dashcams offer a complementary solution to this challenge. They provide a broader range of road network coverage for traffic enforcement. Unlike static cameras, they have the advantage of unpredictability which is essential for effective deterrence. With the growing ubiquity of vehicle-mounted dashcams [19, 25, 43], it is essential to develop systems that can effectively utilize dashcam videos for automated detection of traffic violations.

Detecting two-wheeler traffic violations via dashcam footage presents numerous challenges. These include relative motion between the dashcam-equipped vehicle and other vehicles, the unpredictable dynamics of dense and unstructured traffic environments, occlusions caused by vehicles, riders, other road elements, and the small spatial footprint of objects of interest (*e.g.* motorcycles, riders, license plates). The current body of work addressing dashcam-based detection of two-wheeler traffic violations is extremely limited [21, 42], and no solution extends to the automated ticket generation stage.

*Authors contributed equally to this work.

Figure 1. Instances of triple riding and helmet rule violations from our RideSafe-400 dataset.

To bridge this gap, we propose an end-to-end system for automating the generation of E-tickets for two-wheeler traffic violations, utilizing dashcam videos as input. Our contributions include:

- **Segmentation and Cross-Association (SAC):** A novel segmentation module for associating riders with their corresponding two-wheeler.
- **Association-based Tracker**: A novel instance tracking module with a formulation optimized for joint tracking of riders and associated two-wheelers.
- **RideSafe-400 Dataset:** An annotated dashcam video dataset containing triple riding and helmet rule violations. The dataset includes extensive track-level mask and bounding box annotations for six object classes: `rider-motorcycle`, `rider`, `motorcycle`, `helmet`, `no-helmet`, `license plate`.

We use our dataset to evaluate the E-ticket generation system and to benchmark its various components. Please refer to the https://dash-cop.github.io/ for videos and additional media related to our work.

2. Related Works

Helmet Detection: Previous research on helmet detection predominantly utilizes static cameras [13, 15, 47, 55, 56, 61]. While some methods process the entire frame [13, 17, 47, 55, 56], others utilize cropped images of the riding instance as input [51, 61]. State-of-the-art detection frameworks such as YOLO [28] and DETR [9, 64] are commonly employed. In our approach, we perform detection using the full frame and leverage YOLO-v8x [28] as the base detector.

Triple Riding Detection: Similar to helmet detection, triple riding detection is typically performed using static cameras and established detection architectures [12, 14, 29, 31, 32, 38, 40]. Determining triple riders is done either by counting the number of riders belonging to a driving instance [12, 14, 21, 31, 38] or holistically, as a detection problem [11, 40]. Our approach treats the task as a classification problem.

Rider-Motorcycle Association: Determining the association between riders and their motorcycle is crucial for multiple downstream modules in the context of the violation detection task. Current methods often rely on the overlap between rider and motorcycle bounding boxes for rider-motorcycle (R-M) association [12, 14, 38]. Goyal *et al*. [21] introduce a learnable trapezium-based representation with IoU-threshold-based suppression for compact modeling. However, these methods treat rider and motorcycle detections independently, which can degrade association performance in dense traffic scenarios. Our approach explicitly models the spatial relationship between a motorcycle and its riders, resulting in more robust associations.

Tracking: In most works, trackers based on the SORT [8] family such as DeepSORT [58] are employed [14, 17, 21, 55]. However, tracking is typically done at the individual rider or motorcycle level. We introduce a unique extension of the basic SORT tracker to enable collective tracking of riders and their corresponding motorcycles as a single associated entity.

End-to-end systems: Dashcam footage has been used in various contexts [5, 30, 45, 46, 60], but its application for traffic violation detection is limited [21, 42, 51]. Automatic or semi-automatic systems for end-to-end E-ticket generation are usually based on static cameras [11, 38]. However, these approaches often rely on image processing heuristics and are prone to fixed viewpoint bias. Also, the datasets tend to be fairly small and end-to-end evaluation (*i.e*. E-ticket performance metrics) is absent. To the best of our knowledge, our work is the first to introduce an end-to-end E-ticket generation system for two-wheelers *using dashcam*

videos.

3. RideSafe-400 Dataset

Our dataset, RideSafe-400, comprises 400 driving videos specifically curated for identifying triple riding and helmet rule violators on the road. The traffic scenarios are captured using a DDPAI X2S Pro dashcam, with a resolution of 1920×1080 pixels and a frame rate of 25 fps. Each video spans between 60 to 72 seconds, culminating in approximately 600K frames. These videos contain multiple instances of triple riding or helmet rule violations, captured in a diverse range of traffic conditions, including varying traffic types, street types, illumination, weather, and other challenging scenarios (see Fig. 1).

Annotations in RideSafe-400 cover six object classes: 'R-M instance', 'rider', 'motorcycle', 'helmet', 'no-helmet', and 'license plate' (see Fig. 2-①). For each video frame, riders are associated with their respective motorcycles (termed an R-M instance) using rectangular bounding boxes. Each R-M instance is further annotated with the classes rider and motorcycle (see Fig. 2-② (a)). For each rider, the presence or absence of a helmet is annotated (Fig. 2-② (b)). Each motorcycle object is assigned an attribute 'license plate' or 'no license plate', depending on the visibility of the license plate, and the corresponding license plate bounding box is annotated along with its plate number (Fig. 2-② (c)). To uniquely identify an R-M instance across video frames, a *track id* is provided (see Fig. 2-①). Additionally, each R-M instance and its corresponding objects (rider, motorcycle, helmet, no-helmet, license plate) are assigned a common *group id* attribute (see Fig. 2-③).

Our dataset includes 354K R-M annotations and 23K R-M tracks, 356K motorcycle annotations and 24K motorcycle tracks, 311K rider annotations and 19K rider tracks, 194K helmet annotations and 10K helmet tracks, 149K no-helmet annotations and 8K no-helmet tracks. Additionally, there are 27K annotated license plates, each labelled with its corresponding license plate number.

4. Methodology

Figure 3 provides an overview of our proposed end-to-end E-ticket generation method for detecting triple riding and helmet rule violations. As the first step, each frame from the dashcam video is processed by our novel Segmentation and Cross-Association (SAC) module (Sec. 4.1). This module generates segmentation masks of riders and motorcycles using a YOLO-v8 framework [28]. We introduce novel extensions to the standard YOLO-v8 framework, which predicts the association of rider and motorcycle masks. The aggregation of all the riders on the same motorcycle comprises an R-M instance (see $R\text{-}M_1$, $R\text{-}M_2$, ... in Fig. 3 (a)).

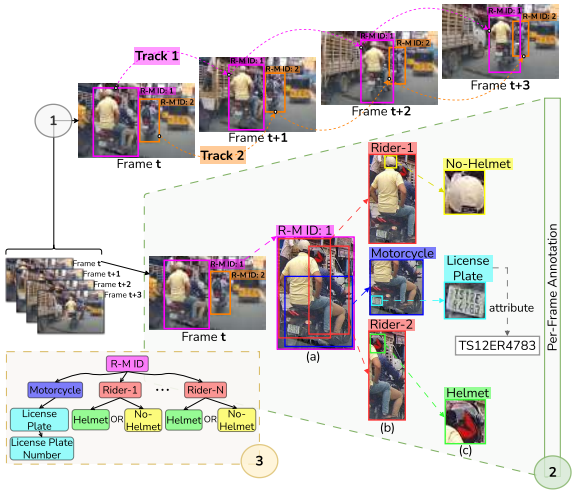

Figure 2. Annotation schema for our RideSafe-400 dataset. Refer to Sec. 3 for details.

Next, our novel cross-association-based tracker module (Sec. 4.2) operates across video frames and assigns unique track IDs to R-M instances (Fig. 3 (b)). Each tracked R-M instance is then processed to detect potential traffic violations. For this, the region-of-interest (ROI) image crop of each R-M instance is fed into the violation module (Fig. 3 (c)), which includes a helmet detection module (Sec. 4.3.1) and a triple rider classification module (Sec. 4.3.2). The per-frame predictions of these modules are consolidated to indicate the presence of a potential triple riding violation or a helmet rule violation for each track.

The tracks with detected traffic violations are further processed by the ANPR module. This module first detects the license plates within the R-M instance and then recognizes the license plate characters (Sec. 4.4, Fig. 3 (e)). Finally, the generated E-ticket (license plate data, along with the corresponding violation information and the R-M track frames) is made available to traffic authorities for verification and E-ticket generation (Fig. 3 (f)).

4.1. Segmentation and Cross-Association (SAC)

Given a frame $I \in \mathbb{R}^{3\times H\times W}$ from a video sequence, we first use a CNN backbone [28] to obtain multi-scale features $F_i, i = 1, 2, 3, \ldots$ (Fig. 4 (i)), where i denotes the scale index. The extracted feature maps are fed to the detection head and mask coefficient head (Fig. 4 (ii)). For each F_i, the detection head outputs $D_i = [B_i, K_i]$, where B_i is the bounding box coordinate prediction map of shape $(4, H_i, W_i)$, and K_i is the class scores prediction map of shape $(2, H_i, W_i)$, representing the object classes 'motorcycle' and 'rider'.

Instance segmentation: The mask coefficient head outputs an intermediate mask coefficient embedding map E_i of shape (C, H_i, W_i), where each E_i^{xy} is a vector along

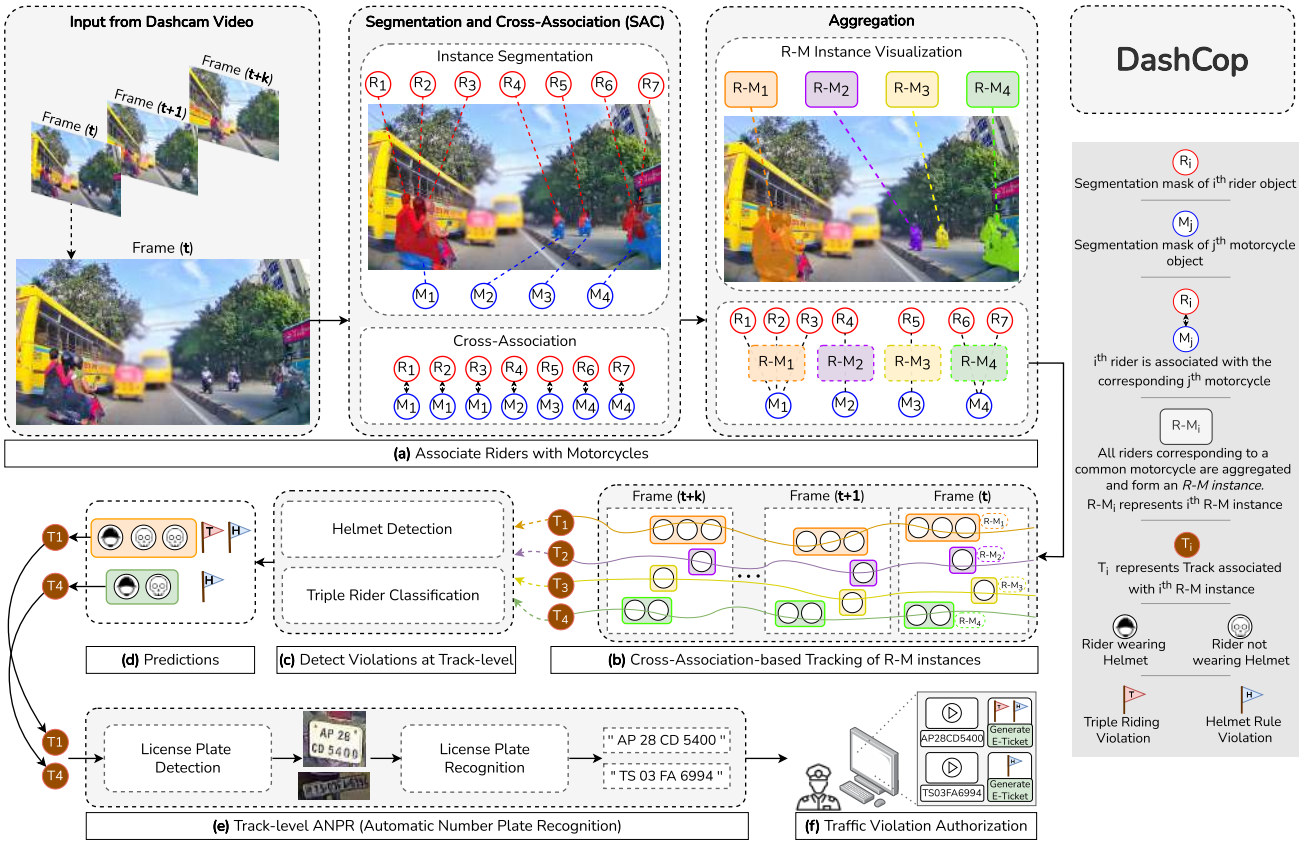

Figure 3. **Our Method:** (a) Dashcam video frames are input to our Segmentation and Cross-Association (SAC) module, which associates riders with their motorcycles to form rider-motorcycle (R-M) instances. (b) These R-M instances are then fed into a cross-association-based tracker, which robustly tracks them across frames and outputs track information T_i for each R-M instance $R\text{-}M_i$. (c) This track information is used by the violation module to predict triple riding and helmet rule violations. (d) In the illustration, R-M instance $R\text{-}M_1$ is flagged for both triple riding (red flag) and helmet rule violations, while $R\text{-}M_4$ is flagged for a helmet rule violation (blue flag). (e) The track information T_1 and T_4, corresponding to $R\text{-}M_1$ and $R\text{-}M_4$ respectively, are fed into the ANPR module, which detects and reads the license plates. (f) Finally, the generated E-ticket, with supporting evidence, is shared with traffic authorities for authorization.

the channel that encodes an instance's representation at cell (x, y) on i^{th} scale. A parallel branch processes the high-resolution feature map F_1 to output M of shape $(C, \frac{H}{4}, \frac{W}{4})$. To predict the instance segmentation mask for each spatial location (x, y) in the F_i, we take the dot product between M and E_i^{xy}

$$S_i^{xy} = M \cdot E_i^{xy} \tag{1}$$

where S_i^{xy} is of shape $(\frac{H}{4}, \frac{W}{4})$.

Intuitively, E_i^{xy} represents a query embedding, whereas M represents the key embeddings. E_i^{xy} attends strongly to spatial locations where the corresponding object is present in M. Consequently, the key locations which have a high similarity with E_i^{xy} become a part of the final segmentation mask S_i^{xy}.

Cross-Association: Traditional instance segmentation treats riders and motorcycles as independent entities. To model the association between riders and their corresponding motorcycles, we introduce a novel cross-object segmentation mask prediction head. This head follows the same segmentation mechanism as that of instance segmentation described above, but with a key modification. The query E_i^{xy} for a rider instance is now encouraged to attend to location(s) where the rider's motorcycle is present (see $M^{crossmask}$ in Fig. 4). Conversely, the query for the motorcycle is encouraged to attend to *all* locations corresponding to all the riders on the motorcycle ($R^{crossmask}$ in Fig. 4). The cross-object segmentation mask is predicted as

$$A_i^{xy} = M' \cdot E_i^{xy} \tag{2}$$

where A_i^{xy} is of shape $(\frac{H}{4}, \frac{W}{4})$.

During inference, Non-Maximum Suppression (NMS) is applied to identify spatial locations in F_i where the class scores exceed a predefined threshold. Segmentation and cross-object segmentation masks are then retrieved for these

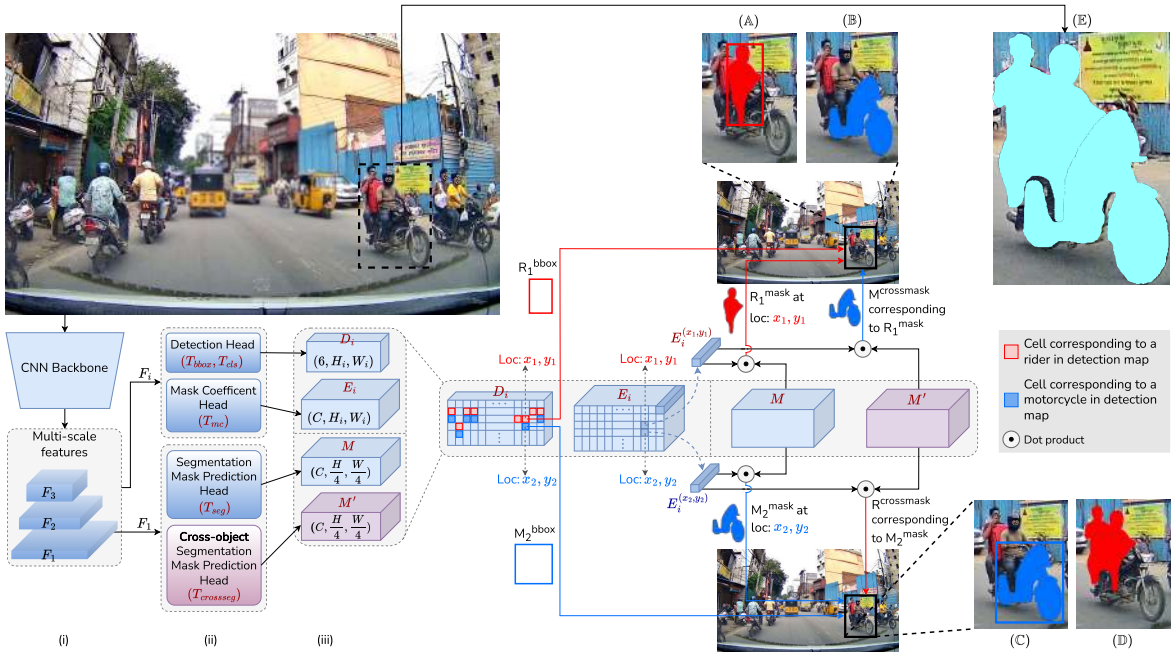

Figure 4. **Segmentation and Association (SAC):** The input frame (top-left) is processed by the SAC module to predict rider locations ($\mathbb{A}$), motorcycle locations ($\mathbb{C}$), and cross-object segmentation masks ($\mathbb{B}, \mathbb{D}$). ($\mathbb{E}$) visualizes the R-M instance. In the architecture diagram, the blue blocks are from YOLO-v8 [28]. The purple blocks are our novel addition which enable association. Refer to Sec. 4.1 for details.

spatial locations using Eqs. (1) and (2), resulting in a set of masks and cross-object masks for both riders and motorcycles. The resulting set of bounding boxes, B^k, segmentation masks, S^k, and cross-object segmentation masks, A^k, are expressed as $[(B_r^1, S_r^1, A_r^1), \ldots, (B_r^N, S_r^N, A_r^N)]$ for riders and $[(B_m^1, S_m^1, A_m^1), \ldots, (B_m^M, S_m^M, A_m^M)]$ for motorcycles. The association score $a_{k,l}$ between the k-th rider and l-th motorcycle detections is computed as

$$a_{k,l} = \frac{1}{2}(\text{IoU}(A_r^k, S_m^l) + \text{IoU}(S_r^k, A_m^l \cdot \text{mask}(B_r^k))) \quad (3)$$

where $\text{mask}(B)$ is a function that takes in a bounding box and returns a binary mask with values inside the box set to 1 and outside set to 0.

4.2. Rider-Motorcycle Association-based Tracking

SORT is a popular, efficient, and robust family of trackers [4, 8, 37, 58, 59, 62]. Tracking is performed via a bipartite matching between the existing tracks and current detections. This is done using a score matrix that represents the likelihood of assigning the i^{th} track to the j^{th} detection. This is modeled by assigning a binary variable $a_{ij} \in \{0, 1\}$ to each combination of the current tracks and the detections. If s_{ij} denotes the score of assigning the i^{th} track to the j^{th} detection, the goal is to maximize $a^* = \underset{a_{ij} \in \{0,1\}}{\arg\max} \sum_{i=1}^{i=N} \sum_{j=1}^{j=M} a_{ij} \cdot s_{ij}$ subject to the constraints

$$\sum_{i=1}^{i=N} a_{ij} \leqslant 1 \; \forall \; j \in \{1 \ldots M\} \quad (4)$$

$$\sum_{j=1}^{j=M} a_{ij} \leqslant 1 \; \forall \; i \in \{1 \ldots N\} \quad (5)$$

These constraints ensure that each detection can be matched to only one track, and vice versa. The score s_{ij} is a function of a motion model (*e.g.* a Kalman filter) and an appearance model (*e.g.* object re-identification).

Cross-AssociationSORT: In contrast to tracking object classes independently, our method aims to track both 'rider' and 'motorcycle' instances *jointly*, while keeping track of the evolving associations between members of these object classes. This approach offers two key advantages: firstly, it improves tracking accuracy for each object, and secondly, it establishes robust associations among R-M instances over time.

To implement cross-association-based tracking, we utilize R-M association scores defined earlier (see Eq. (3)). Let the rider tracks up to time t be represented by $T_t^r = \{r_1, r_2, \ldots r_{N_r^t}\}$ and the motorcycle tracks by $T_t^m = \{m_1, m_2, \ldots m_{N_m^t}\}$, where N_r^t and N_m^t denote the number of rider and motorcycle tracks at time t, respectively. The

observations/detections at time $t+1$ are denoted by the set D^r_{t+1} and D^m_{t+1}. Let the set of all possible assignments for the rider and motorcycle class be represented by

$$C^r = \{(r_i, d_j) \mid r_i \in T^r_t \ \& \ d_j \in D^r_{t+1}\} \quad (6)$$

$$C^m = \{(m_i, d_j) \mid m_i \in T^m_t \ \& \ d_j \in D^m_{t+1}\} \quad (7)$$

We define a binary variable t^r_i, for the i^{th} element in C^r, represented by C^r_i denoting its possibility of being a rider track at time $t+1$. Similarly t^m_j is another binary variable defined for C^m_j. Additionally, we define a likelihood score based on R-M association alone for every pair of combinations, one from C^r and the other from C^m. For the i^{th} rider hypothesis in C^r and the j^{th} motorcycle hypothesis in C^m, we define a binary variable e_{ij} with an associated score of a_{ij}, which reflects the likelihood of association of these two hypotheses. If and only if both C^r_i and C^m_j are chosen, only then there is a possibility of $e_{ij} = 1$, implying the association of the two tracks. This can be represented as

$$e_{ij}(1 - t^r_i \cdot t^m_j) = 0 \ \forall \ i, j \quad (8)$$

Additionally, to ensure that a rider can only be associated with one motorcycle, we impose the constraint

$$\sum_{j=1}^{j=|C^m|} e_{ij} \leqslant 1 \ \forall \ i \quad (9)$$

In our modified formulation, we maximize

$$(t^{r*}, t^{m*}, e^*) = \underset{\{t^r_i, t^m_j, e_{ij}\} \in \{0,1\}}{\arg\max} \left(\lambda_1 \sum_{i=1}^{|C_r|} t^r_i s^r_i + \lambda_2 \sum_{j=1}^{|C_m|} t^m_j s^m_j + \lambda_3 \sum_{i=1}^{|C_r|} \sum_{j=1}^{|C_m|} e_{ij} a_{ij} \right) \quad (10)$$

subject to the aforementioned (Eq. (8), Eq. (9)) and the original (Eq. (4), Eq. (5)) constraints. Here, s^r_i is the score for i^{th} rider hypothesis and s^m_j is the score for j^{th} motorcycle hypothesis. The values obtained by the variables, t^r_i and t^m_j after the optimization will reflect the next set of tracks at time $t+1$, and will be associated based on the value of e_{ij}.

4.3. Traffic Violation Detection

4.3.1 Helmet Detection Module

Each video frame is processed by a helmet detection module, trained to classify 'helmet' and 'no-helmet'. The module outputs detections for each frame. These detection boxes are then matched with corresponding rider tracks. For each rider track, we determine the predominant class ('helmet' or 'no-helmet') across multiple frames using majority voting.

4.3.2 Triple Rider Classification Module

The input to the triple riding classification model consists of R-M instance ROI crops output by the cross-association-based tracking module. The classifier is designed as a learnable convolutional layer on the frozen mask coefficient head of the SAC module (Sec. 4.1), followed by a learnable linear layer. The model classifies the rider count as 'single', 'double', 'triple' or 'none'. During training, the backbone of SAC is frozen. An R-M track is flagged as a triple riding violation if at least one instance ROI in the track is classified as 'triple riding'.

4.4. License Plate Recognition and E-ticket Generation

The license plate detection model is trained to identify license plate regions within the R-M instance crops. For tracks with detected violations, the ROIs from all frames are passed through the license plate detector. Detected license plate crops are then input to an OCR model. A lightweight OCR architecture [49] is trained on a dataset of real-world, augmented, and synthetic license plates. The license plate numbers detected across multiple frames for a given R-M instance track are consolidated using a majority vote, where the most frequently detected license plate number is selected and linked to the corresponding R-M instance track for E-ticket generation.

5. Training and Implementation Details

5.1. Segmentation and Cross-Association (SAC)

The bounding box and class predictions are matched to the ground truth using the Task-Alignment Learning (TAL) approach [18]. The matched boxes are evaluated against the ground truth boxes using the Complete IoU (CIOU) loss [63] and the Distribution Focal Loss (DFL) [34]. For classification, Binary Cross-Entropy (BCE) loss is used. For segmentation and cross-object segmentation, a per-pixel BCE loss is used. The overall loss function is formulated as $\mathcal{L}_{SAC} = \lambda_1 L_{CIOU} + \lambda_2 L_{DFL} + \lambda_3 L_{cls} + \lambda_4 L_{seg} + \lambda_5 L_{crossseg}$ where $\lambda_1 = 7.5, \lambda_2 = 1.5, \lambda_3 = 0.5, \lambda_4 = 7, \lambda_5 = 7$ are hyperparameters for balancing multi-task learning objectives. The model is trained for 200 epochs with a learning rate of 0.01, using the Adam optimizer on 4 NVIDIA GeForce RTX 3080 Ti GPUs over 4 days.

5.2. Rider-Motorcycle Association-based Tracking

We use the pre-trained Re-Identification (ReID) model SBS-R50 [23] for appearance modeling and a Kalman filter for motion modeling. The optimization process is handled using the Gurobi solver [22]. Segmentation masks from SAC module are stored for a fixed sized buffer of $k = 3$ time steps. Using these, for computing a_{ij}, we calculate the

sum of association scores for the rider and motorcycle detection for the k time steps, i.e. $\sum_{t=T-k}^{T} \mathrm{A}(C_{i^t}^r, C_{j^t}^m) - \theta$, where $C_{i^t}^r$ and $C_{j^t}^m$ represent the rider and motorcycle detections at the t^{th} time step, and θ is a threshold set to 0.5. The function A is implemented using Eq. (3). If there is no detection for either $C_{i^t}^r$ or $C_{j^t}^m$, then function A returns a value of 0. If the association IDs of the tracks is already established and the association ID of the i^{th} rider hypothesis is not equal to the association ID of the j^{th} motorcycle hypothesis, then a_{ij} is set to -∞. We use the validation set of RideSafe-400 dataset for tuning the tracker's hyperparameters.

5.3. Traffic Violation Detection

The Helmet Detection module is based on the YOLOv8-x [28], trained to classify two categories: 'helmet' and 'no-helmet'. The model is trained using full-resolution frames, with a dataset consisting of 37K helmet and 23K no-helmet bounding boxes. The training process spans 70 epochs, using an Adam optimizer with a learning rate of 0.01, on four NVIDIA GeForce RTX 3080 Ti GPUs for one day.

5.4. License Plate Detection and Recognition

A YOLOv8-x model [28] is trained for license plate detection. To improve performance during both training and inference, multi-line license plates of motorcycles are converted into a single-line format. License plate recognition is performed using an OCR model [49], which is trained on real-world, augmented, and synthetic license plate data.

6. Experiments and Results

We conducted experiments using a dataset of 400 videos, divided into 200 videos for training, 100 videos for validation, and 100 videos for final evaluation (test).

6.1. System-Level Performance (E-ticket Generation)

We evaluate the system using correctly tracked R-M instances. An R-M track is considered a true positive for E-Ticket generation if the system accurately predicts the traffic violation (*e.g.* no-helmet or triple riding) and correctly detects and recognizes the license plate. False positives and false negatives are calculated similarly (see Tab. 1). As shown in Tab. 2, the automated system achieves an F1-score of 72.18%, reflecting good overall performance. In practice, traffic enforcement personnel review E-Tickets and corresponding evidence before issuance. This human-in-the-loop approach eliminates false positives, raising the F1-score to 82.05%, improving system reliability. Some examples of system performance can be seen in Fig. 5. Empirical observations indicate that the system struggles with detecting triple riders in scenarios involving significant glare or motion blur, and helmet detection suffers from occlusions due to objects or vehicles, leading to false negatives. Furthermore, the triple riding classifier misclassified distant occluded R-M instances, contributing to false positives. License plate recognition is most accurate when R-M instances move in the same direction as the ego vehicle. However, the recognition rate drops for R-M instances approaching from the opposite lane.

Traffic Violation			LP Detection			LP Recognition		**E-ticket**		
TP	FP	FN	TP	FP	FN	Correct	Incorrect	**TP**	**FP**	**FN**
✓			✓			✓		✓		
✓			✓				✓		✓	
	✓		✓				✓		✓	
	✓		✓			✓			✓	
✓				✓			✓		✓	
	✓			✓			✓		✓	
		✓								✓
✓					✓					✓

Table 1. **Evaluation criteria for E-ticket generation:** For a given R-M track, the labelling (TP, FP, FN) at various stages of system pipeline is used to determine the final E-ticket level label of the track. For e.g., a True Positive (TP) prediction at all stages of the pipeline is considered a TP for E-ticket generation system. *Correct* means that all the characters of the license plate number match the ground-truth. Refer to Sec. 6.1 for details.

E-ticket System Output Level	**Precision %**	**Recall %**	**F1-score %**
Automatic	84.21	63.16	72.18
Human-In-The-Loop	100.00	69.57	82.05

Table 2. Overall E-ticket system performance (Sec. 6.1).

Method	**ROI Extraction Approach**	**Precision %**	**Recall %**	**F1-score %**
Goyal *et al.* [21]	R-M Instance	51.42	32.34	39.71
Cui *et al.* [16]	Frame	**76.34**	39.11	51.72
YOLOv8-x [28]	Rider Instance	67.15	53.57	59.60
	R-M Upper Instance	67.98	58.11	62.66
	R-M Instance	73.48	57.85	64.73
Ours	**Frame**	68.75	**62.78**	**65.63**

Table 3. Comparative evaluation of helmet violation detection.

6.2. Violation Detection

Helmet Violation: Tab. 3 presents comparative evaluation results for helmet violation detection at the rider track level are. Our YOLOv8-x-based frame-level approach achieves the best F1-score, striking a good balance between precision and recall. The results show that frame-level context is marginally more effective compared to detecting helmet violations at R-M instance crop level. Compared to prior works, YOLOv8-x based methods show better balance across different ROI extraction methods.

Triple Riding Violation: The results in Tab. 4 indicate that our R-M instance-based classification approach yields the best performance. Methods based on counting rider or helmet detections are less robust to occlusions. Combining our classifier with rider-counting greatly reduces false negatives. However, it increases the false positive rate.

Method	Precision %	Recall %	F1-score %
Goyal *et al*. [21]	32.30	01.46	02.79
Rider-Count	54.86	14.60	23.07
Helmet Detector	65.92	59.55	62.57
TR Classifier + Rider-Count	59.51	**75.28**	66.40
TR Classifier + Helmet Detector	57.88	43.82	49.88
TR Classifier	**69.22**	66.29	**67.72**

Table 4. Comparative evaluation of triple riding violation.

Detection Method	R-M Association	Association Score %
Goyal *et al*. [21]	IOU	58.50
	Trapezium Regressor	59.32
Ours	IOU	80.93
	Trapezium Regressor [21]	81.71
	SAC Module	**84.04**

Table 5. Evaluation of rider-motorcycle association.

Tracker	HOTA	MOTA	IDF1
ByteTrack [62]	56.86	48.60	60.04
DeepOCSORT [37]	58.21	50.78	62.57
BotSORT [4]	58.12	50.72	63.09
HybridSORT [59]	58.10	50.66	62.29
Cross-AssociationSORT (Ours)	**60.41**	**51.96**	**64.58**

Table 6. Evaluation of tracking rider-motorcycle instances.

6.3. Rider-Motorcycle Detection and Association

We define the association score as the percentage of correctly associated R-M instances. Our SAC module outperforms association methods involving geometric heuristics based on bounding box overlapping (Tab. 5). The trapezium regressor [21] does not take into account the visual information, and directly operates on the coordinates of the bounding boxes present. In contrast, our approach operates directly in the image space and learns a more robust association between the classes.

6.4. Rider-Motorcycle Instance Tracking

To evaluate the tracking of R-M instances, we use the widely accepted HOTA [35], MOTA [7], and IDF1 [44] metrics. We compare with various SORT trackers. From Tab. 6, we see that joint tracking of R-M instances results in better tracking performance across all metrics than independently tracking component objects (riders, motorcycles) and post-hoc aggregation.

6.5. License Plate OCR

For license plates, the metrics are Character Error Rate (CER) and accuracy (percentage of plates recognized correctly in their entirety). From Tab. 7, we see that CRNN [49] trained on a combination of synthetic and real data provides the best results. We found CNN-based approach to outperform more recent transformer-based approaches [33] and commercial OCR [1].

Method	Real	Synth	Aug	CER (↓)	Accuracy %
TrOCR [33]	✓	✓	✓	0.1691	16.41
GoogleOCR [1]				0.2698	33.34
CRNN [49]	✓			0.0507	71.42
	✓		✓	0.0602	71.43
	✓	✓	✓	0.0503	80.95
	✓	✓		**0.0400**	**85.71**

Table 7. Evaluation of license plate recognition methods.

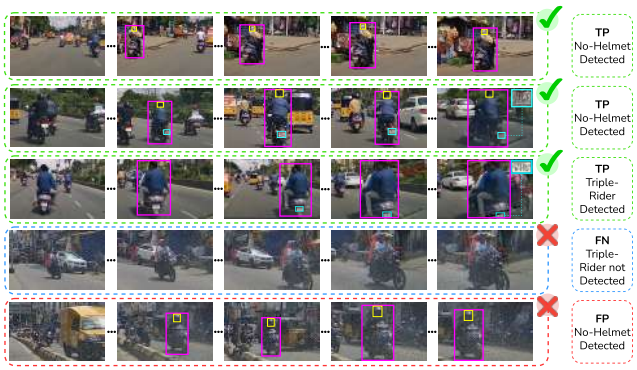

Figure 5. Qualitative results of our system and failure cases. R-M instances are indicated by pink bounding boxes. Yellow indicates a No-Helmet detection.

7. Conclusion

DashCop is a comprehensive end-to-end dashcam-video based system for automated E-ticket generation targeting traffic violations in motorized two-wheelers. The key innovations include the SAC module for precise rider-motorcycle (R-M) association, Cross-AssociationSORT for robust R-M instance tracking, and system-level performance evaluation criteria. The large-scale RideSafe-400 video dataset, featuring multi-level and multi-label annotations, is another key contribution. Validated on this dataset, DashCop shows reliable E-ticket generation capabilities, presenting a promising solution for enhancing traffic safety through automated enforcement and deterrence.

Societal concerns: While enhancing road safety, the proposed technology raises privacy concerns. To mitigate these risks, we adhere to relevant legal and ethical standards governing surveillance-based public traffic management systems. Our system is designed to relay only relevant information about detected traffic violations to authorities. Thus, we minimize the risks associated with sharing complete footage indiscriminately. Also, we employ anonymization techniques to obscure identifiable information (*e.g.* faces). Even for license plates, only the instances corresponding to traffic violations are shown. The data is stored securely, with access restricted to authorized personnel.

Acknowledgements: This work is supported by IHub-Data, IIIT Hyderabad.

References

[1] Vision AI: Image & Visual AI Tools — cloud.google.com. https://cloud.google.com/vision?hl=en. 8

[2] Armstrong Aboah, Bin Wang, Ulas Bagci, and Yaw Adu-Gyamfi. Real-time multi-class helmet violation detection using few-shot data sampling technique and yolov8. In *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*, pages 5349–5357, 2023. 1

[3] MarkNtel Advisors. Global two wheeler market research report: Forecast (2023-2028), 2023. 1

[4] Nir Aharon, Roy Orfaig, and Ben-Zion Bobrovsky. Bot-sort: Robust associations multi-pedestrian tracking. *arXiv preprint arXiv:2206.14651*, 2022. 5, 8

[5] AH Ariffin, MS Solah, A Hamzah, MS Ahmad, MA Mohamad Radzi, ZH Zulkipli, AS Salleh, and Z Mohd Jawi. Exploring characteristics and contributory factors of road crashes and near misses recorded via dashboard camera. *Journal of the Society of Automotive Engineers Malaysia*, 5(2):292–305, 2021. 2

[6] Fabian Barreto, Nesline D Almeida, Premraj Nadar, Ravneet Kaur, and Sanjana Khairnar. Object detection and e-challan generation systems for traffic violation: A review. *International Journal of Research in Engineering, Science and Management*, 4(5):54–57, 2021. 1

[7] Keni Bernardin and Rainer Stiefelhagen. Evaluating multiple object tracking performance: the clear mot metrics. *EURASIP Journal on Image and Video Processing*, 2008:1–10, 2008. 8

[8] Alex Bewley, Zongyuan Ge, Lionel Ott, Fabio Ramos, and Ben Upcroft. Simple online and realtime tracking. In *2016 IEEE international conference on image processing (ICIP)*, pages 3464–3468. IEEE, 2016. 2, 5

[9] Nicolas Carion, Francisco Massa, Gabriel Synnaeve, Nicolas Usunier, Alexander Kirillov, and Sergey Zagoruyko. End-to-end object detection with transformers. In *European conference on computer vision*, pages 213–229. Springer, 2020. 2

[10] Shubham Kumar Chandravanshi, Hirva Bhagat, Manan Darji, and Himani Trivedi. Automated generation of challan on violation of traffic rules using machine learning. *International Journal of Science and Research (IJSR)*, 10(3):1157–1162, 2021. 1

[11] R Shree Charran and Rahul Kumar Dubey. Two-wheeler vehicle traffic violations detection and automated ticketing for indian road scenario. *IEEE Transactions on Intelligent Transportation Systems*, 23(11):22002–22007, 2022. 2

[12] Pooja Chaturvedi, Kruti Lavingia, and Gaurang Raval. Detection of traffic rule violation in university campus using deep learning model. *International Journal of System Assurance Engineering and Management*, 14(6):2527–2545, 2023. 2

[13] Yunliang Chen, Wei Zhou, Zicen Zhou, Bing Ma, Chen Wang, Yingda Shang, An Guo, and Tianshu Chu. An effective method for detecting violation of helmet rule for motorcyclists. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 7085–7090, 2024. 2

[14] Rutvik Choudhari, Shubham Goel, Yash Patel, and Sunil Ghane. Traffic rule violation detection using detectron2 and yolov7. In *2023 World Conference on Communication & Computing (WCONF)*, pages 1–7. IEEE, 2023. 2

[15] Shun Cui, Tiantian Zhang, Hao Sun, Xuyang Zhou, Wenqing Yu, Aigong Zhen, Qihang Wu, and Zhongjiang He. An effective motorcycle helmet object detection framework for intelligent traffic safety. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 5469–5475, 2023. 1, 2

[16] Shun Cui, Tiantian Zhang, Hao Sun, Xuyang Zhou, Wenqing Yu, Aigong Zhen, Qihang Wu, and Zhongjiang He. An effective motorcycle helmet object detection framework for intelligent traffic safety. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 5470–5476, 2023. 7

[17] Viet Hung Duong, Quang Huy Tran, Huu Si Phuc Nguyen, Duc Quyen Nguyen, and Tien Cuong Nguyen. Helmet rule violation detection for motorcyclists using a custom tracking framework and advanced object detection techniques. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 5380–5389, 2023. 1, 2

[18] Chengjian Feng, Yujie Zhong, Yu Gao, Matthew R Scott, and Weilin Huang. Tood: Task-aligned one-stage object detection. In *2021 IEEE/CVF International Conference on Computer Vision (ICCV)*, pages 3490–3499. IEEE Computer Society, 2021. 6

[19] Yahoo Finance. Global dashboard camera market analysis 2023-2030, 2024. 1

[20] Rajdeep Ghosh, Akash Kotal, Raja Karmakar, and Krittika Das. A deep learning-based smart helmet detection and notification with e-challan issuance for traffic fines. In *2023 IEEE 4th Annual Flagship India Council International Subsections Conference (INDISCON)*, pages 1–6. IEEE, 2023. 1

[21] Aman Goyal, Dev Agarwal, Anbumani Subramanian, CV Jawahar, Ravi Kiran Sarvadevabhatla, and Rohit Saluja. Detecting, tracking and counting motorcycle rider traffic violations on unconstrained roads. In *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*, pages 4303–4312, 2022. 1, 2, 7, 8

[22] Gurobi Optimization, LLC. Gurobi Optimizer Reference Manual, 2024. 6

[23] Lingxiao He, Xingyu Liao, Wu Liu, Xinchen Liu, Peng Cheng, and Tao Mei. Fastreid: A pytorch toolbox for general instance re-identification. *arXiv preprint arXiv:2006.02631*, 2020. 6

[24] Fortune Business Insights. Two-wheeler market size, share covid-19 impact analysis, by type (scooter, motorcycle, and moped), by technology (ice and electric), and regional forecast, 2023-2030, 2023. 1

[25] Future Market Insights. Dashboard camera market by technology, video quality, application region | forecast 2022 to 2032, 2022. 1

[26] Global Market Insights. Electric two-wheeler market size - by vehicle type (electric motorcycle, electric scooter, e-

bikes, electric kick scooter), battery (sla, li-ion), motor placement (hub-motor, frame-mounted motor), motor power, motor speed global forecast, 2023 - 2032, 2023. 1

[27] Mordor Intelligence. Global two-wheeler market size share analysis - growth trends forecasts up to 2029, 2023. 1

[28] Glenn Jocher, Ayush Chaurasia, and Jing Qiu. Ultralytics YOLO, Jan. 2023. 2, 3, 5, 7

[29] Salna Joy, Ujwal A Reddy, Rishab C Lal, R Vinay, et al. Traffic rule violation recognition for two wheeler using yolo algorithm. In *2023 Second International Conference on Electronics and Renewable Systems (ICEARS)*, pages 1477–1480. IEEE, 2023. 2

[30] Gorkem Kar, Shubham Jain, Marco Gruteser, Fan Bai, and Ramesh Govindan. Real-time traffic estimation at vehicular edge nodes. In *Proceedings of the Second ACM/IEEE Symposium on Edge Computing*, pages 1–13, 2017. 2

[31] Nikhil Kumar, Gaurav Kumar Sahu, M Ravi, Sahil Kumar, V Sukruth, and AN Mukunda Rao. Triple riding and no-helmet detection. In *2023 4th IEEE Global Conference for Advancement in Technology (GCAT)*, pages 1–6. IEEE, 2023. 2

[32] Mrs J RATNA KUMARI, NADENDLA BHAVANI, SHAIK THALIB, VATLURU CHARAN NAGA SAI SURYA, and BATHULA Srikanth. An effcient system for detecting traffic violations such as over speed, disregarding signals, and instances of triple riding. *Turkish Journal of Computer and Mathematics Education (TURCOMAT)*, 15(1):104–108, 2024. 2

[33] Minghao Li, Tengchao Lv, Jingye Chen, Lei Cui, Yijuan Lu, Dinei Florencio, Cha Zhang, Zhoujun Li, and Furu Wei. Trocr: Transformer-based optical character recognition with pre-trained models. In *Proceedings of the AAAI Conference on Artificial Intelligence*, volume 37, pages 13094–13102, 2023. 8

[34] Xiang Li, Chengqi Lv, Wenhai Wang, Gang Li, Lingfeng Yang, and Jian Yang. Generalized focal loss: Towards efficient representation learning for dense object detection. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 45(3):3139–3153, 2023. 6

[35] Jonathon Luiten, Aljosa Osep, Patrick Dendorfer, Philip Torr, Andreas Geiger, Laura Leal-Taixé, and Bastian Leibe. Hota: A higher order metric for evaluating multi-object tracking. *International journal of computer vision*, 129:548–578, 2021. 8

[36] Dr. Shikha Tiwari M. Shaun Daniel. Advanced traffic monitoring and enforcement using yolov8. *International Journal of Research Publication and Reviews*, 5(1):2070–2073, 2024. 1

[37] Gerard Maggiolino, Adnan Ahmad, Jinkun Cao, and Kris Kitani. Deep oc-sort: Multi-pedestrian tracking by adaptive re-identification. In *2023 IEEE International Conference on Image Processing (ICIP)*, pages 3025–3029. IEEE, 2023. 5, 8

[38] Nikhil Chakravarthy Mallela, Rohit Volety, and Nadesh RK. Detection of the triple riding and speed violation on two-wheelers using deep learning algorithms. *Multimedia Tools and Applications*, 80(6):8175–8187, 2021. 2

[39] Md Mamun Miah, Biton Chakma, and Kabir Hossain. Analyzing the prevalence of and factors associated with road traffic crashes (rtcs) among motorcyclists in bangladesh. *The Scientific World Journal*, 2024(1):7090576, 2024. 1

[40] RG Nandhakumar, K Nirmala Devi, N Krishnamoorthy, S Shanthi, VR Pranesh, and S Nikhalyaa. Yolo based license plate detection of triple riders and violators. In *2023 International Conference on Computer Communication and Informatics (ICCCI)*, pages 1–6. IEEE, 2023. 2

[41] The Times of India. Just 72,000 traffic cops to manage 20 crore vehicles, 2019. 1

[42] Yan-Tsung Peng, Chen-Yu Liu, He-Hao Liao, Wei-Cheng Lien, and Gee-Sern Jison Hsu. Traffic violation detection via depth and gradient angle change. In *2022 IEEE 7th International Conference on Intelligent Transportation Engineering (ICITE)*, pages 326–330. IEEE, 2022. 1, 2

[43] Straits Research. Dashboard camera market, 2022. 1

[44] Ergys Ristani, Francesco Solera, Roger Zou, Rita Cucchiara, and Carlo Tomasi. Performance measures and a data set for multi-target, multi-camera tracking. In *European conference on computer vision*, pages 17–35. Springer, 2016. 8

[45] Arash Rocky, Qingming Jonathan Wu, and Wandong Zhang. Review of accident detection methods using dashcam videos for autonomous driving vehicles. *IEEE Transactions on Intelligent Transportation Systems*, 2024. 2

[46] Junaid Shaikh and Nils Lubbe. Evaluating forward collision warning and autonomous emergency braking systems in india using dashboard cameras. Technical report, SAE Technical Paper, 2024. 2

[47] K Shanmugam and Kakimanu Chandana. A swin transformer-based approach for motorcycle helmet detection. 2024. 2

[48] Bhuvanesh Kumar Sharma, Aman Sharma, Sanjay Kumar Sharma, Yogesh Mahajan, and Sneha Rajput. Do the two-wheeler safety harnesses effective in rider's safety- analysis of attitude and switching intention. *Case Studies on Transport Policy*, 15:101146, 2024. 1

[49] Baoguang Shi, Xiang Bai, and Cong 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, volume=39, number=11, pages=2298–2304, year=2016, publisher=IEEE.* 6, 7, 8

[50] Elham Soltanikazemi, Ashwin Dhakal, Bijaya Kumar Hatuwal, Imad Eddine Toubal, Armstrong Aboah, and Kannappan Palaniappan. Real-time helmet violation detection in ai city challenge 2023 with genetic algorithm-enhanced yolov5. In *2023 IEEE Applied Imagery Pattern Recognition Workshop (AIPR)*, pages i–x. IEEE, 2023. 1

[51] Richard Sugiarto, Evan Kusuma Susanto, and Yosi Kristian. Helmet usage detection on motorcyclist using deep residual learning. In *2021 3rd East Indonesia Conference on Computer and Information Technology (EIConCIT)*, pages 194–198. IEEE, 2021. 2

[52] Supriya S Thombre, Sampada S Wazalwar, Rakshanda Tarekar, Sakshi Ramgirwar, Shivani Zulkanthiwar, and Shubhra Dubey. Automated traffic rule violation detection with e-challan generation for smart societies. In *Decision Analytics for Sustainable Development in Smart Soci-*

ety 5.0: Issues, Challenges and Opportunities, pages 65–81. Springer, 2022. 1

[53] Duong Nguyen-Ngoc Tran, Long Hoang Pham, Hyung-Joon Jeon, Huy-Hung Nguyen, Hyung-Min Jeon, Tai Huu-Phuong Tran, and Jae Wook Jeon. Robust automatic motorcycle helmet violation detection for an intelligent transportation system. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 5341–5349, 2023. 1

[54] Chun-Ming Tsai, Jun-Wei Hsieh, Ming-Ching Chang, Guan-Lin He, Ping-Yang Chen, Wei-Tsung Chang, and Yi-Kuan Hsieh. Video analytics for detecting motorcyclist helmet rule violations. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 5365–5373, 2023. 1

[55] Thien Van Luong, Huu Si Phuc Nguyen, Duy Khanh Dinh, Viet Hung Duong, Duy Hong Sam Vo, Huan Vu, Minh Tuan Hoang, and Tien Cuong Nguyen. Motorcyclist helmet violation detection framework by leveraging robust ensemble and augmentation methods. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 7027–7036, 2024. 2

[56] Hao Vo, Sieu Tran, Duc Minh Nguyen, Thua Nguyen, Tien Do, Duy-Dinh Le, and Thanh Duc Ngo. Robust motorcycle helmet detection in real-world scenarios: Using co-detr and minority class enhancement. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 7163–7171, 2024. 2

[57] Bor-Shiun Wang, Ping-Yang Chen, Yi-Kuan Hsieh, Jun-Wei Hsieh, Ming-Ching Chang, JiaXin He, Shin-You Teng, HaoYuan Yue, and Yu-Chee Tseng. Prb-fpn+: Video analytics for enforcing motorcycle helmet laws. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 5476–5484, 2023. 1

[58] Nicolai Wojke, Alex Bewley, and Dietrich Paulus. Simple online and realtime tracking with a deep association metric. In *2017 IEEE international conference on image processing (ICIP)*, pages 3645–3649. IEEE, 2017. 2, 5

[59] Mingzhan Yang, Guangxin Han, Bin Yan, Wenhua Zhang, Jinqing Qi, Huchuan Lu, and Dong Wang. Hybrid-sort: Weak cues matter for online multi-object tracking. In *Proceedings of the AAAI Conference on Artificial Intelligence*, volume 38, pages 6504–6512, 2024. 5, 8

[60] Stephen A Zekany, Ronald G Dreslinski, and Thomas F Wenisch. Classifying ego-vehicle road maneuvers from dashcam video. In *2019 IEEE Intelligent Transportation Systems Conference (ITSC)*, pages 1204–1210. IEEE, 2019. 2

[61] Hongpu Zhang, Zhe Cui, and Fei Su. A coarse-to-fine two-stage helmet detection method for motorcyclists. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 7066–7074, 2024. 2

[62] Yifu Zhang, Peize Sun, Yi Jiang, Dongdong Yu, Fucheng Weng, Zehuan Yuan, Ping Luo, Wenyu Liu, and Xinggang Wang. Bytetrack: Multi-object tracking by associating every detection box. In *European conference on computer vision*, pages 1–21. Springer, 2022. 5, 8

[63] Zhaohui Zheng, Ping Wang, Dongwei Ren, Wei Liu, Rongguang Ye, Qinghua Hu, and Wangmeng Zuo. Enhancing geometric factors in model learning and inference for object detection and instance segmentation. *IEEE transactions on cybernetics*, 52(8):8574–8586, 2021. 6

[64] Zhuofan Zong, Guanglu Song, and Yu Liu. Detrs with collaborative hybrid assignments training. In *Proceedings of the IEEE/CVF international conference on computer vision*, pages 6748–6758, 2023. 2