International Journal of
DATA SCIENCE AND IOT MANAGEMENT SYSTEM
ISSN: 3068-272X www.ijdim.com Original Research Paper

VISION-DRIVEN TRAFFIC SAFETY: DETECTING RULE VIOLATIONS WITH DEEP LEARNING MODELS

[1]Kim Jenny,[2]Jung Woo Bin
Department Of ECE
Korea Advanced Institute of Science & Technology (KAIST), Daejeon, South Korea

Received: 14-05-2022 Accepted: 20-6-2022 Published: 27-6-2022

ABSTRACT

With the rapid growth of urbanization and increasing vehicular density, traffic rule violations have emerged as a major contributor to road accidents, congestion, and public safety risks. Traditional surveillance methods relying on manual monitoring and CCTV review are time-consuming, error-prone, and resource-intensive. To address these challenges, this work proposes a Smart Traffic Enforcement System powered by Deep Learning for the automated detection of traffic violations.

The system leverages computer vision and convolutional neural networks (CNNs) to analyze real-time traffic footage and accurately identify violations such as signal jumping, lane discipline breaches, helmet non-compliance, and overspeeding. By integrating advanced image processing with AI-driven classification models, the framework enables high-precision detection and reduces dependency on human intervention. Additionally, the detected violations can be linked to vehicle registration databases for generating automated challans, ensuring transparency and accountability in enforcement.

This approach not only strengthens road safety but also supports smart city initiatives by providing scalable, cost-effective, and reliable traffic monitoring. The proposed model demonstrates the potential of deep learning to revolutionize intelligent transportation systems (ITS) and pave the way for data-driven urban traffic governance.

I. INTRODUCTION:

Road safety has become one of the most pressing global concerns as the number of vehicles and population density in urban areas continues to rise. According to the World Health Organization (WHO), millions of accidents occur annually, with traffic rule violations such as signal jumping, overspeeding, not wearing helmets, and lane indiscipline being leading causes. Traditional methods of traffic monitoring rely heavily on human observation and manual CCTV review, which are both labor-intensive and prone to errors. The inefficiency of these methods highlights the urgent need for automated, intelligent solutions that can operate in real time.

Advances in computer vision and deep learning have paved the way for smart traffic enforcement systems that can detect violations with greater accuracy and minimal human involvement. Convolutional Neural Networks (CNNs), in particular, have revolutionized image recognition tasks, enabling models to learn and detect complex patterns in traffic environments. These models can identify vehicles, riders, and behaviors such as helmet absence, license plate visibility, and traffic signal compliance with high precision.

The proposed system, Vision-Driven Traffic Safety, applies deep learning models to analyze video feeds from surveillance cameras and automatically flag instances of rule violations. Unlike traditional approaches, this method leverages real-time detection, scalability, and adaptability to different urban traffic conditions. The integration of such a system into existing surveillance infrastructure supports smart city initiatives, enhances transparency in law enforcement, and helps in reducing the burden on traffic authorities.

Ultimately, this research aims to demonstrate that AI-powered traffic monitoring is not only feasible but also essential for improving road safety, traffic efficiency, and law enforcement effectiveness in modern cities.

II. LITERATURE SURVEY

1. Problem decomposition & pipeline

Typical automated violation systems break the problem into modular stages:

Sensing & capture — fixed CCTV / PTZ / traffic cameras, sometimes supplemented by radar/LiDAR.

Object detection — find vehicles, riders, pedestrians, helmets, traffic lights, lanes, etc.

Multi-object tracking (MOT) — associate detections across frames to compute trajectories and enable temporal rules (e.g., crossing during red).

Violation reasoning — rule-based or learned classifiers that use tracked positions + scene geometry (stop line, signal state) and temporal logic to detect violations (signal jumping, lane breaches, helmetless riding, illegal turns, etc.).

Identity/forensics — ALPR (automatic license plate recognition), evidence packaging (images/video snippets, timestamps), and e-challan integration.

Modularity helps testing and legal validation but introduces cascading error sources (detection errors → tracking errors → false violation flags).

2. Object detection: architectures & trade-offs

Two-stage detectors (Faster R-CNN, Mask R-CNN): higher accuracy, useful when precision is critical (e.g., helmet detection), but relatively slower.

Single-stage detectors (YOLO family, SSD, RetinaNet): very popular for traffic because they balance accuracy with real-time speed—suitable for edge deployment. Newer YOLO variants and transformer-based detectors push further on speed/accuracy.

Lightweight backbones and model compression (MobileNet, EfficientNet, pruning, quantization) are widely used to deploy on edge devices (Jetson, TPU).

Key concerns: small-object detection (distant vehicles/plates), occlusion in dense scenes, night-time and adverse weather robustness.

3. Tracking & temporal association

Tracking-by-detection using methods like DeepSORT, ByteTrack, and newer transformer-based trackers is common.

Good tracking is crucial to compute speeds, lane changes, and to avoid duplicate challans. Practical challenges include ID switches in crowded scenes and sustained occlusions.

Combining visual tracking with lightweight motion models and scene priors (lane maps, homography) improves temporal reasoning.

4. Violation detection & geometric reasoning

Rule-based geometric checks (line-crossing for red-light detection, region-of-interest for no-entry zones) remain reliable and interpretable; they require camera calibration or homography to map pixels → real-world coordinates.

Learned classifiers (LSTM/RNN or 3D-CNNs on tracklets) can capture complex temporal patterns (aggressive lane weaving), but need annotated temporal datasets.

Hybrid approaches (detectors + rule checks + small learned modules) often give the best trade-off between explainability and flexibility.

5. License plate recognition & identity linking

ALPR pipelines usually follow detection → plate cropping → OCR (CRNN/Transformer OCR models). Challenges include motion blur, dirt/damaged plates, multilingual formats, and low-resolution frames.

Super-resolution preprocessing, temporal aggregation (combining multiple frames), and

International Journal of
DATA SCIENCE AND IOT MANAGEMENT SYSTEM
ISSN: 3068-272X www.ijdim.com Original Research Paper

domain-specific OCR fine-tuning help increase accuracy.

III. PROPOSED METHODOLOGY

We'll talk about its recommended start-to-end plan in this section. There are four major parts to this approach. You can see the similar software architecture in. The creators of the proposed technique first provide a public formulation of the security camera footage as input, after which they use detection to do lane detection. Swag (you look just once) has been used for object tracking, such as attempting to identify vehicles within the public domain. Individual autos appear to be clipped out using position attained and by specified parameters in addition to detection approaches within a week of the system being built to detect automobiles. Additionally, a car would be tried to inspect the group for specific violations. The suggested method appears to contain violations of the zebra crossing (violating camel crossing) and head gear (two rowers not yet planning to wear one helmet). Several vehicles would check that head gear, but instead they would check for pedestrian crossing violations, and four more vehicles would check for intersection violations. A headgear breach was found using Fox News, a great application that tends to function well with visual information (convolutional neurones network). Using conceal resnet, which allows feature extraction to compare the precise location of an automobile's lower half, such as its tread, with that of the identified zebra crossing, junction violations can be easily detected. The registration plate of the roughly equivalent car would be found via detection even before the violation or violations appear to be intercepted for such a vehicle. Following that, swag was being utilised to identify a vehicle's license plate. The number plate number is first obtained from the registration plate using a classifier (optical protagonist recognition). Customers of automobiles appear to be informed of related violations, and violations appear to be ingrained in the DBMs. Its dbms could be utilised to obtain data analysis through violations of the aforementioned traffic laws. then we will carefully examine each analogue output separately.

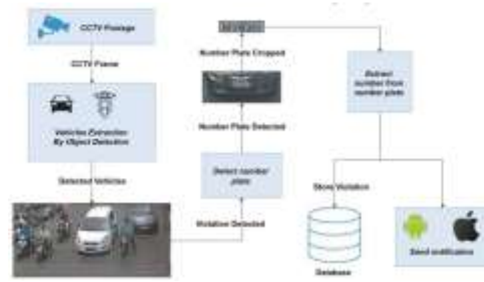

Figure 1: Architecture Diagram for proposed system

Below are our presentations that are pertinent to each device:

Video surveillance would be used to detect vehicles of a single structure when their message illuminates. In order to provide input to the object tracking device about lane detection, a shape has now been carried. For this purpose, we're using a classifier (you only need to look once). Justsayin is merely a heuristic for tracking objects [10]. The classifier uses a convolution ejector's fully-connected forward lid to directly precede the identical compass bearings after every frame. Attempting to forecast minimises the complexity of the situation rather than location and makes it far easier for the internet backbone to fully comprehend. Since classifiers employ pipelines rather than recognise artefacts that contain sequential stages, they are healthier than other computer vision systems that are somewhere around r-cnn, rapidly r-cnn, and quicker r-cnn. Therefore, these optimisation methods appear to be slow between drives, but they are actually quite challenging compared to maximise, and each component needs to be trained separately. Justsayin could surpass this algorithm by

International Journal of
DATA SCIENCE AND IOT MANAGEMENT SYSTEM
ISSN: 3068-272X www.ijdim.com Original Research Paper

performing this task with only a tuned artificial neural network.

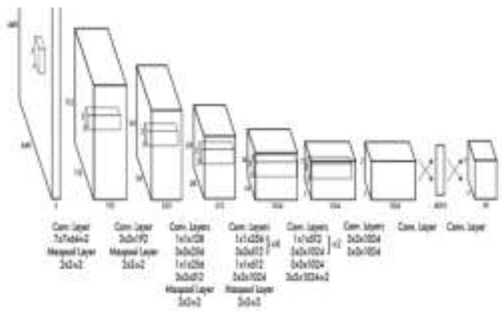

Figure 2: Architecture of YOLO [10]

Figure 2 illustrates its construction, which has two final completely connected and 18 convolutional layers. provide a competitive edge by lowering the same spatial properties of the single layer leading up to persons using a convolutional start [10].

After all, the pre-trained design is capable of detecting not only cars but also people or groups, which are like entirely other objects [12]. As can be seen in the image, researchers developed a customised framework in which a particular sensor detects the identical autos and the roadies as a single target. Strength training from the pre-trained design and the concept yeah learning algorithm were used to produce its customised model. Object detection has been used to conduct traffic surveillance.

Figure 3: Object detection with YOLO

Figure 2.2 shows that although the items appear to have been found, they are now categorised into "two" and "four" classrooms. School "two" is reflected in a sustainable boundary box, and school "four" is indeed reflected in the coloured projection.

B. Helmet Classification

Two vehicles appear to be considered for such a hard hat categorisation after cars have been recognised based on photos. If the car really belong to the two-rowe school, it might have been further pixellated, which scholars believe is the top half of both vehicles (upper 50% of the auto frame). Individual cars have been snipped first from closed-circuit television shapes with support such related projection. Even though it is statistically significant, faces (helmets) can be seen in the upper half of both images now that the top half is considered. A scheme's photo computation cost may be reduced while also taking into account a higher percentage of both. Whether the snapshot corresponds to full head gear or non-helmet lectures, the top part appears to be sent as input to the classifiers. A diagrammatical representation of headgear designation is shown in fig. 4.

Figure 4: Helmet classification using CNN. Input image is classified as “wearing helmet” or “not wearing helmet”.

International Journal of
DATA SCIENCE AND IOT MANAGEMENT SYSTEM
ISSN: 3068-272X www.ijdim.com Original Research Paper

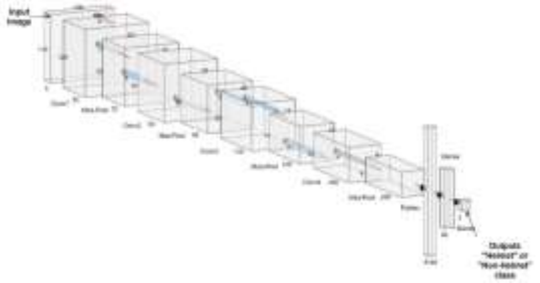
Figure 5: Architecture of CNN for Helmet Classification


Figure 6: Visualization of the trained representation by CNN for the classification of helmet and non-helmet class

C. Crosswalk violations at traffic signals

Feature extraction would be used to find the crosswalk breach. In fact, helmet r-cnn is an extension of faster r-cnn, which includes sets such as object-based compartmentalised and one classification conceal [11]. With every property, edge detection usually provides us with a sparse representation helmet. Training in feature extraction design was provided for two college courses, focussing on intersections rather than vehicles. Humans chose to utilise pre-trained models to hide r-cnn strength training because coaching helmet r-cnn takes a significant amount of time [13]. Now that researchers have used the segmentation method, each car has a helmet, but the camel and its bounding boxes are in the picture instead.

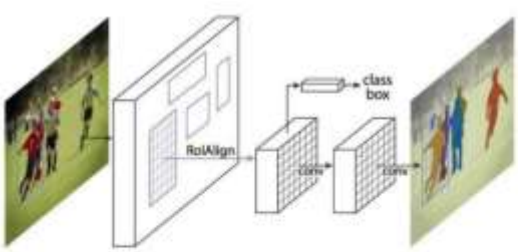

Figure 7: The Mask R-CNN framework for instance segmentation [11]

Crosswalk violations occur even when a car appears to be on a pedestrian crossing (zebra crossing). First, using a subset of different vehicles, researchers are able to pinpoint the precise location of the underside of a wheel. For human verification, it is currently possible to determine whether collected compass bearings cross both junctions without comparing them to the projection of both intersections. Even though the roots are (0,0), researchers assume the highest spot after all public formulations. They consider (zx1,zy1) but also (zx2,zy2), even though the location comes after all subsets like zebra crossings, which are high streets, and the lowest part edge between (vx1,vy1) but instead (vx2,vy2), even though the location comes after all projections of a specific car after all extremely high streets but also base streets including both. Researchers compare zy1 to vy2, but unless vy2 is more than zy1, in which case vy2 will only be clarified—yes, underside like a wheel—then an infringement has already been reported to have occurred. Every car would be subject to a pedestrian crossing violation in addition to needing to verify that the lower half's coordinates, such as wheel mistruths, are accurate within the crossings. Figure 3. Six shows that even a 4 (masked yellow) and a couple of two different (marked purple and grass) have breached junctions, but they are more concerned with the precise placement of the underside after all rubber false claims and across crosswalks.

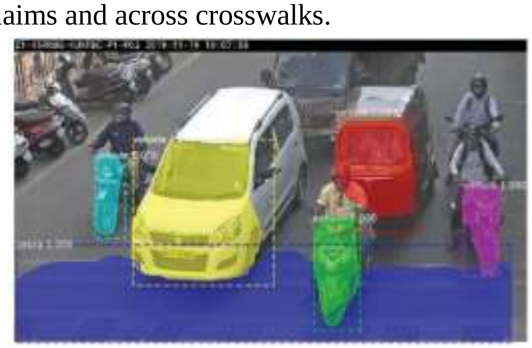
Figure 8: Instance segmentation using Mask R-CNN

D. License Plate Recognition

After attempting to identify a traffic law violation, the responsibility is to obtain the vehicle registration from the vehicle that violated the law. First, researchers use object recognition to locate an automobile's registration plate. Again, yolo[10] appears to be used for placement, such as plate numbers. restricted driver's licenses of vehicles that violate any laws that have been removed. In addition to being able to conduct opencv (optical protagonist recognition) on a snippet of the license plate, this identical vehicle's registration has been obtained. It appears that pattern matching comes before character recognition in the classifier. Heroes and villains have been divided into different groups, and actors have then been compared in DBMs using feature matching. as the vehicle's breach would be recorded in the data as soon as the registration was obtained, and the user would be notified via email and text message about the same.

Figure 9: shows the detection of the number plate of the vehicle, then the characters are read using OCR

IV. DATASET

The traffic police agency of Pune provided the CCTV footage of different signals, which was used to create the dataset. In order to prevent visual recurrence, the dataset was created from the CCTV footage by capturing image frames at regular intervals of roughly 15 seconds. To incorporate variation in the data, the photos included both sparse and busy traffic scenarios.

TABLE I: DATASET USED FOR DIFFERENT MODULES

Dataset	No. of images	No. of classes	Name of classes
Vehicle Detection	525	2	two, four
Helmet Classification	1079	2	helmet, non-helmet
Crosswalk Violation	200	2	vehicle, crosswalk
Number Plate Detection	450	1	numplate

a total of the 600 images collected for the purpose of detecting vehicles, which also includes all other kinds of vehicles, including bicycles, cars, and buses, but instead pedicabs. Out of the total number of images, 420 appear to have been considered for coaching, while another 105 were considered for testing. A full set of enlist pictures appear to have been considered for this type of headgear classification; 492 of those individuals wore headgear, while 628 wore hard hats instead. The images, including the classification of headgear, were acquired by applying traffic surveillance to its data set and then selecting its vehicles for the two-wheelers school. As a result, it appears that our enlisted images 683 were evaluated for teaching, while 180 were evaluated for checking. 1999 images have been identified as having geometric shapes that appear to have been created around artefacts to focus on the latter's dpi while also appropriately planning its headgear, such as coaching mask-rcnn internet backbone, in order to check for errors like camel infringement that employs segmentation. Over the course of 1999, 150 photos were identified as coaching and label shots related to test-taking. number. Both of those wheelers and four rickshaws will be taken into consideration when describing one's own vehicle's number plate number using the number plate recognition data set that has been prepared. anything that includes, in total, about 23 license plates, of which 4 have been used for training and 99 have been used for testing.

V. RESULT

In this section, we present experimental results of our modules viz. vehicle detection, helmet classifier and Image Segmentation. The experiments are carried on Ubuntu 18.04 machine consisting of i5 6th gen processor, 12 gb of ram and 256 gb SSD. Python 3.65 was used for evaluation of results along with libraries such as opencv-python for image processing, tensorflow as a Deep Learning framework (for training and prediction), numpy for mathematical operations and matplotlib for visualization.

1. Vehicle Detection:

Vehicle detection is performed using YOLO algorithm, which was trained on 420 images consisting of both the classes, two and four. This was evaluated against 105 test images considering metric as average precision per class and also calculating the mean average precision. The results for which are as follows:

TABLE II: RESULTS OF VEHICLE DETECTION

Class	Average Precision
Four	0.9406
Two	0.9594

The mean average precision(mAP) for the model is 0.95.

2. Helmet Classification:

Helmet Classifier performs image classification using CNN consisting of five convolutional layers with ReLu activation units, 4 max-pooling layers, and one fully connected dense layer with final softmax unit for classification into two classes. The proposed model architecture is evaluated using metrics such as precision, recall, accuracy, and F1-score on a test-set consisting of 215 images where 98 images belonged to helmet class and 117 belonged to non-helmet class.

TABLE III: RESULTS OF HELMET CLASSIFICATION

Metric	Value
Precision	0.8780
Recall	0.8925
Accuracy	87%
F1 score	0.8852

3. Crosswalk Violation:

Crosswalk Violation employs use of mask-rcnn network [11]. This network is trained on 150 images with 2 classes namely vehicles, and crosswalk(zebra) using the concept of transfer learning. This network was then evaluated against 50 images each consisting of all of the 3 classes. Metric used for evaluation is mean Average Precision(mAP) is computed on a testset consisting of 50 images.

TABLE IV:RESULTS OF CROSSWALK VIOLATION

Metric	Value
mean Average Precision (mAP)	0.96

4. Number Plate Detection:

Number plate detection is performed using YOLO again which was trained on 450 images. This was evaluated against 100 test images considering metric as average precision. The results for which are as follows:

TABLE V: RESULTS OF NUMBER PLATE DETECTION

Class	Average Precision
numplate	0.9305

VI. CONCLUSION

The proposed approach employs the use of concepts such as YOLO, CNN, Mask R-CNN and OCR for automatic detection of traffic rules violations. It achieves the desired goal with precision and ease, but requires high computational power as it makes use of concepts like image segmentation and object detection. The advantage of proposed system is it has the ability to catch more number of violations as compared to the human intervened system. In addition, the proposed methodology provides an end to end autonomous system, when implemented would prove an upper hand in detecting violations. Thus strict regulations of traffic rule violations can be implemented resulting in better road safety and bring awareness among vehicle users.

FUTURE SCOPE

This study demonstrates that deep learning–based vision systems can move traffic enforcement beyond reactive measures toward proactive safety management. By accurately detecting rule violations—such as signal jumping, helmet non-compliance, and lane misuse—these models not only provide evidence for enforcement but also enable real-time interventions that reduce accidents and improve overall road discipline.

While challenges remain in scaling across diverse lighting, weather, and infrastructure conditions, the results indicate strong potential for deployment in smart city environments. Integrating these systems with IoT-enabled surveillance and traffic management platforms could create a closed-loop framework where detection, alerting, and corrective measures happen seamlessly.

In essence, vision-driven deep learning models transform raw traffic footage into actionable intelligence, bridging the gap between technology and road safety. With continuous refinement, they can serve as a cornerstone for safer, more efficient, and more accountable transportation ecosystems.

REFERENCES

1. https://www.indiatoday.in/diu/story/two-wheeler-deathroad-accidents-helmets-states India-1602794-2019-09-24
2. J. Chiverton, "Helmet presence classification with motorcycle detection and tracking," Intelligent Transport Systems (IET), vol. 6, no. 3, pp. 259–269, September 2012
3. Gomathi, et al, "Automatic Detection of Motorcycle without helmet using IOT", South Asian Journal of Engineering and Technology
4. Samir Ibadov, et al, "Algorithm for detecting violations of traffic rules based on computer vision approaches", MATEC Web of Conferences 132, 05005 (2017)
5. Aaron Christian P. Uy, et al, "Automated Traffic Violation Apprehension System Using Genetic Algorithm and Artificial Neural Network", 2016 IEEE Region 10 Conference
6. Amey Narkhede, et al, "Automatic Traffic Rule Violation Detection and Number Plate Recognition", IJSTE - International Journal of Science Technology & Engineering | Volume 3 | Issue 09 | March 2017
7. Ana Riza, et al, "Localization of License Plates Using Optimized Edge and Contour Detection Technique ", 2017

IEEE Region 10 Conference (TENCON), Malaysia, November 5-8, 2017

8. Prem Kumar Bhaskar, et al, "Image Processing Based Vehicle Detection and Tracking Method ",Universiti Teknologi PETRONAS Seri Iskandar, 31750, Perak, Malaysia
9. Kunal, Dahiya ; Dinesh, Singh ; and C. Krishna, Mohan "Automatic Detection of Bike-riders without Helmet using Surveillance Videos in Real-time". 2016 International Joint Conference on Neural Networks (IJCNN)
10. Joseph Redmon, Santosh Divvala, Ross Girshick, Ali Farhadi, "You Only Look Once: Unified, Real-Time Object Detection"
11. Kaiming He, Georgia Gkioxari, Piotr Doll´ar, Ross Girshick, "Mask R-CNN", Facebook AI Research (FAIR) ,24 JAN 2018
12. http://pjreddie.com/media/files/yolo.weights
13. https://github.com/matterport/Mask_RCNN/releases/download/v2.0/mask_rcnn_coco.h5
14. Pooya Sagharichi Ha, Mojtaba Shakeri, "License Plate Automatic Recognition based on Edge Detection"
15. Abdullah Asım, et al, "A Vehicle Detection Approach using Deep Learning Methodologies"