IJCRT.ORG ISSN : 2320-2882

INTERNATIONAL JOURNAL OF CREATIVE RESEARCH THOUGHTS (IJCRT)

An International Open Access, Peer-reviewed, Refereed Journal

A Conceptual Framework for Real-Time Traffic Violation Detection Using YOLOv8-Small

[1]Abhishek K, [2]Chirag HS, [3]Dheeraj BU, [4]Krishna Kumar

[1, 2, 3, 4] B.Tech Students

Department of Computer Science & Engineering,

Reva University, Bengaluru, India

Abstract: Traffic offenses like signal jumping, unhelmeted riders, lane indiscipline, etc., occur frequently in urban areas and cannot be corrected manually as it is time consuming and prone to inconsistency and inaccuracy. This paper proposes an AI-based Traffic Violation Detection System with a conceptual design that is capable of detecting vehicles and riders, helmets and traffic lights using the YOLOv8-small deep learning model. The whole system is based on object detection, Tracking and Frame-by-Frame analysis of the vehicles, identifying the violations and also calculating the speed of the vehicle from the video input, without the need of any external hardware (radar, sensors, etc.). Combining different types of detections into a single pipeline also provides a more holistic traffic monitoring approach. The primary requirements for this pipeline from an implementation viewpoint are reliable real-time detection, violation automation, and a speed that can support practical use. The proposed approach is extensible for centralized monitoring dashboard for traffic authorities. The framework provides a proof of concept on the role of AI helped solutions in traffic law enforcement, reducing human intervention and making roads safer and more efficient in urban settings.

Index Terms - Traffic violation, YOLOv8, real-time object detection, smart transportation systems, computer vision.

I. INTRODUCTION

Traffic urbanization has increased the need for well-executed road safety and traffic laws enforcement. Existing monitoring techniques, including manually-based methods and hardware methods such as speed scanners and traffic sensors, are costly, have limited coverage, can be error-prone if manually operated, and cannot be easily scaled. This has created a need for smarter and more automated solutions.

With the introduction of computer vision and deep learning, it is now possible to perform smart traffic monitoring by streaming existing feeds from traffic surveillance cameras. This project seeks to build an integrated system to detect and control violations such as red light jumping, over-speeding, non-usage of helmets while driving, triple riding on two-wheeler vehicles and license plate recognition to provide a complete solution.

The subsequent traffic scene analysis is achieved through the fusion of object detection, object tracking, and image analytical techniques. The system accepts multiple inputs from sensors. Several different modules work in tandem to identify the target, including vehicle detection, rider detection, object tracking, speed estimation using spatial transforms and the extraction of vehicle license plate features.

One challenge of the system is to deal with real-world phenomena such as changing camera perspective, occlusion and the fact that detection and re-detection happen at different times. To this end, the systems rely on perspective correction, temporal tracking and custom inference pipelines that are optimized to their specific use case.

This model presents a workable and economical method for automatically detecting traffic violations from video input alone. It points out the great potential of expanding AI-powered methods in easing the reliance on human intervention and in aiding the growth of intelligent traffic control systems.

II. LITERATURE REVIEW

Recently, the use of computer vision for traffic monitoring has been attracting a lot of attention mainly because of the increasing availability of surveillance cameras and deep learning breakthroughs. Traditional image processing techniques like background subtraction, edge detection, motion tracking were used in earlier days. Although these techniques were effective in only controlled conditions, more realistic scenarios involving changes in illumination, occlusion, and shadows were very challenging for them. Following CNNs' emergence, traffic analysis became a lot more efficient and reliable.

YOLO and other object detection networks have become very well-known due to their capabilities of detecting objects in real-time with very good precision [1], [2]. These networks are used for the identification of automobiles pedestrians traffic lights, etc. in videos. By the same token, great improvements in OCR has resulted automatic vehicle license plate recognition which has led to easier identification of the violators [4].

Several research works have targeted specific traffic rule violations, for instance, red light violation detection systems work by defining a stop line and checking if a vehicle crosses it when the light is red. But, quite a number of these methods depend on unchanging coordinates and do not take camera perspective into account, which may result in incorrect outputs in case of angled views. Attempts trying to solve this issue by utilizing geometric transformations for a better estimation of the positions in the real world have emerged recently.

Video-based speed estimation has also been considered as a substitute for radar-based systems. Detection of objects together with tracking will be the core components that facilitate speed measurement from moving image data by tracking the position of the object frame by frame [5], [6]. Performing a transformation based on homography allows one to derive world coordinates from image coordinates thus enabling precise speed measurement. Still, problems such as obtaining camera parameters accurately and the occlusion of objects have an impact on results adversely.

Identifying helmets and analysing riders on motorbikes have become popular topics in violation-prone areas. The majority of them rely object detection for classifying the images of the riders with or without helmet as well as detecting failures to wear protective gear [3]. A few setups go beyond that by recognizing not only if there are multiple riders but also if the vehicle is overloading; however, occlusion and overlap of the images still pose serious hurdles.

Despite such advancements, lots of present-day systems concentrate on handling just one issue at a time. The whole gamut of violation detection typically remains scattered in various systems. Our project extends previous works by uniting detection tracking perspective correction, and OCR in a single system that can offer a fuller and more viable answer to live traffic monitoring.

III. PROBLEM STATEMENT

Traffic rule violations like jumping red lights speeding not wearing helmets, and triple riding are some of the main causes of accidents, particularly in high-density urban areas. To check these violations manually is not only very laborious but also inefficient since it requires continuous human involvement and is prone to mistakes. Even though some automated systems are available, a lot of them are reliant on costly equipment like radar guns, embedded road sensors, or special enforcement infrastructures, which is a major drawback for their wide-scale deployment.

Another problem is the shortcomings of current vision-based systems. A lot of the solutions that have been developed can detect only a single type of violation, which makes them much less feasible for actual traffic monitoring as multiple violations can happen at the same time. What is more, the accuracy of detection can be impacted to a great extent by different factors such as the camera angle, perspective distortion, occlusion between objects, lighting conditions changes, and processing delays. For instance, from a certain point of view, a vehicle may seem like it is crossing the stop line, or bikers may be only partly visible due to their overlapping.

So, there is a necessity for a single, scalable system that can efficiently track various types of traffic violations from standard video inputs without compromising on accuracy in real-world situations. This kind of system ought to be able to identify and follow several objects simultaneously, understand spatial relationships properly, and cope with timing differences during processing.

The main goal of the present work is to develop a unified computer vision-based system that can do the detection tracking perspective correction, and text recognition for enabling automatic identification and analysis of different kinds of traffic violations in one pipeline.

IV. METHODOLOGY

This paper introduces a simple and integrated system that detects multiple traffic violations and analyzes them individually with a high degree of accuracy using live camera input. It combines pedestrian, bicycle, and vehicle detection with tracking, spatial analysis, and text recognition as one unified workflow. Our methodology consists of the following components:

4.1. Data Collection

To depict real-life urban situations such as intersections, highways, and busy streets, a varied dataset composed of traffic videos and still images was employed. The dataset comprises various vehicle forms (cars, motorcycles, and trucks), people on motorcycles both with and without helmets, and different traffic signal conditions.

Concerning helmet detection, a specially compiled dataset of about 5000 annotated images was employed. In each image, there are labeled bounding boxes with class annotations indicating whether a helmet is present or not. This contributed to the development of a special model that can tell if the rider is following safety rules or not.

Most importantly, a variety of data augmentation methods such as rotation scaling flipping, and changing brightness were used to produce a stronger model. These changes allow the system to work even better in different lighting conditions, with changing camera angles, and even when there are partial occlusions.

4.2. Detection Framework

At the heart of the detection system, YOLOv8 is the main detection model used. This is because it can carry out fast and accurate multi-class detection [3]. Besides detecting vehicles and persons, it also detects helmets and traffic signals in different modules. License plate recognition is performed by cropping detected vehicle areas and then passing them to EasyOCR that extracts the textual information from the plate images [4]. This allows identifying the vehicles that are involved in the violations.

To make the system more efficient, region-based detection is used in some cases-for example, traffic signal detection is limited to a predefined region of interest (ROI), thereby reducing unnecessary computation and enhancing reliability.

4.3. Red Signal Violation Detection

Red signal jumping is detected using a spatial and temporal approach rather than relying on a fixed line check.

A virtual stop line is basically a line that marks a boundary in a scene, like an invisible border that you're not supposed to cross. When a vehicle is spotted, a key point on it is used to figure out where it is. This point is at the bottom and in the middle of a box that is drawn around the vehicle.

The system keeps an eye on this point and follows it from one frame to the next, constantly tracking its movement.

A violation is recorded only when:

- the traffic signal state is detected as red, and
- the tracked vehicle position crosses the defined stop line.

When a camera is at an angle, it can make things look distorted. To fix this, a special kind of transformation is used to get a better sense of how far apart things are and where they are. This helps us understand what's going on in the image more accurately. Also, to make sure we're looking at the right moment in the video, a buffered approach is used. This means that the results of what we're trying to detect are matched up with the correct frame in the video, which reduces mistakes that can happen because of delays in processing.

4.4. Helmet & Triple riding detection

Detecting helmets is done with a special computer program that can tell if a rider is wearing one or not. This program looks at each rider one by one and marks them if they're not wearing a helmet.

For triple-riding detection:

- Motorcycles are detected first.
- Persons (riders) are detected separately.
- The system links riders to a particular motorcycle by how close they are to it.

If more than two riders are associated with a single bike, it is classified as a triple-riding violation. When people or objects are blocked from view, the system looks at what it has detected over several frames, not just one. This makes it more reliable when a lot of things are happening at the same time or when things are overlapping.

4.5. Vehicle speed estimation

Vehicle speed is estimated using a combination of detection, tracking, and geometric transformation.

- Cars are found in each picture and followed from one picture to the next using a special tracking method [5], [6].
- A region of interest is defined to focus on relevant road areas.
- A perspective transformation (homography) is applied to convert the camera view into a top-down representation, allowing pixel distances to approximate real-world distances.

To figure out how fast a vehicle is going, you can measure how far it moves between frames in a video and then divide that by the time between those frames, which is determined by the video's frame rate. This way, you can estimate the speed of the vehicle without needing any extra sensors.

4.6. Violation reporting & Output

The system puts together what it finds in real-time, showing it on top of the video. This includes boxes around the things it finds, special IDs for each thing, and signs that something is not right.

Detected violations include:

- Red signal jumping
- Helmet absence
- Triple riding
- Over-speeding

When a traffic rule is broken, the system can log important details like the time it happened and information about the vehicle, which can be identified using its license plate. This system can also be expanded to include a dashboard that shows all the violations, saves pictures of the incidents, and gives a summary of the traffic data for further study and analysis.

V. CHALLENGES & LIMITATIONS

While the proposed system demonstrates the potential of AI-based traffic violation detection, it still faces several practical challenges when applied to real-world environments. It is important to note that the system is not 100% accurate, and its performance largely depends on the quality of the input video or images.

One major problem is that it's really sensitive to what's going on around it. For example, if it's dark or rainy, or if there are shadows, it can be harder to get a clear picture, which makes it tougher for the system to figure out what's going on - like whether someone's wearing a helmet or what the traffic lights are doing. Also, where the camera is and how it's angled can make things look weird, which can mess up calculations like how fast someone's going or if they've crossed a line, if it's not set up just right.

One big problem with this system is that it can't always see everything clearly. When there are a lot of vehicles and riders on the road, they can block each other's view. This makes it harder for the system to detect things correctly, especially when there are multiple riders on one vehicle and one of them is hidden from view. Because the system mostly uses 2D detection, it can miss things that are blocked by other objects, which can lead to wrong or missed detections of traffic violations. For example, if one rider is behind another, the system might not see them and think there's only one rider, when in fact there are two or three. This can cause problems with enforcing traffic rules and keeping the roads safe.

The performance of the system is also dependent on the quality and diversity of the dataset used for training. A limited or biased dataset may not capture the variability present in real-world traffic conditions, such as different vehicle types, rider positions, or environmental settings. This can affect the model's ability to generalize effectively when deployed in new locations.

When you're dealing with a lot of tasks at the same time, like detecting things, tracking them, reading text, and figuring out how fast something is moving, you need a lot of computing power. Even if the system is made to be efficient, it's still hard to get consistent results in real-time, especially when you're working with high-quality video or multiple cameras. This is because it needs really powerful graphics cards to handle all the work. If the system doesn't have enough resources, it can start to slow down, drop frames, or not be as accurate as it should be, because it has to make compromises on things like resolution or how complex the models are.

One of the tough problems is getting all the parts of the system to work together at the same time. This is because things like detection, tracking, and signal recognition don't always happen at the same time. Even if their timing is just a little off, it can lead to wrong conclusions. For instance, a car might be wrongly accused of running a red light when it actually crossed the road while the signal was changing. So, it's really important to make sure that all the parts of the system are properly aligned, like the frames and the results of the analysis.

To get the speed right, you need to make sure your camera is set up correctly and that you're looking at things from the right perspective. If you get these things wrong, it can throw off your measurements of how far apart things are, and that means your calculations of how fast something is moving will be wrong too. It's all about getting the basics right so you can trust the results you're getting.

It's really important to have good input data, a well-designed dataset, and a system that's been carefully calibrated. When you're deploying a system, you need to have a solid strategy in place. If you want to make things better in the future, you might think about combining vision-based methods with other ways of sensing things, and using more advanced models and hardware to make the system more reliable, even when things get tough. This could help improve performance and make the system more robust. By doing so, you can increase the accuracy and efficiency of the system, which is crucial for achieving good results.

VI. RESULTS

The proposed system is expected to provide a real-time, lightweight, and accurate traffic violation detection framework suitable for standard hardware, capable of identifying multiple violations and providing actionable insights to traffic authorities.

Table 6.1: Expected Outcomes of Proposed System

METRIC	RESULT
Detection accuracy	Good under normal conditions; affected by lighting and occlusion
Frame processing speed	Near real-time on optimized systems; slower on standard hardware
Violation logging	Automatic detection with basic details
Multi-violation detection	Helmet, signal, triple riding, and speed
Speed estimation	Frame-based calculation (depends on calibration)

VII. CONCLUSION & FUTURE WORK

This project uses computer vision and deep learning to detect traffic violations. It combines several techniques, including object detection, tracking, and text recognition, to identify different types of violations. For example, it can detect when a driver runs a red light, isn't wearing a helmet, or has too many people on a bike. It can also catch speeding drivers and extract license plate information from videos. By putting all these techniques together, the system can effectively identify multiple types of traffic violations from video footage. This can help make roads safer and reduce the number of accidents. The system is designed to be comprehensive, meaning it can handle ACKNOWLEDGMENT a wide range of traffic violation detection tasks. It's an important tool for traffic management and can be used to improve road safety.

The YOLOv8 system is really good at detecting objects in real time, and it can do this for many different types of objects. It also uses tracking and perspective-based methods to understand how things are moving and how they are related to each other in space. This is different from older systems that need special hardware to work. Because it only uses video data, it's more flexible and cheaper to use in existing surveillance systems. This makes it a great option for people who want to upgrade their surveillance without having to buy a lot of new equipment.

One key feature of this system is that it can deal with lots of traffic issues at the same time in the same area. By putting different parts together, it can do more than just spot individual problems - it can help with real-world traffic watching in a more practical way. It also has automatic logging and can be connected to a dashboard, which makes it more useful for traffic authorities as they can keep an eye on things all the time and make decisions based on data. This means they can respond quickly to what's happening on the roads.

But like we talked about before, the system still has some problems to deal with, such as things getting in the way, poor quality, and not being able to handle different environments and computing needs. These issues show us where we need to make some changes to make it better.

Future work will focus on enhancing the robustness and scalability of the system. This includes expanding the dataset to cover more diverse traffic scenarios, improving detection under occlusion using advanced tracking or pose-based methods, and optimizing performance for stable real-time deployment. Additionally, integrating more accurate calibration techniques can further improve speed estimation and spatial analysis. The system can also be extended with a fully functional dashboard for centralized monitoring, evidence storage, and statistical reporting.

Overall, this work demonstrates how AI-based solutions can play a significant role in modern traffic management. With further refinement, such systems have the potential to reduce dependency on manual enforcement, improve compliance with traffic rules, and contribute to safer and more efficient road environments.

VIII. ACKNOWLEDGMENT

The authors would like to express their sincere gratitude to **Dr. Sarvamangala D R**, Department of Computer Science and Engineering, REVA University, Bengaluru, for their valuable guidance, continuous support, and insightful suggestions throughout the development of this project. The authors also thank the faculty members of the department for providing the necessary resources and academic environment to carry out this work successfully.

IX. REFERENCES

[1] Joseph Redmon, S. Divvala, R. Girshick, and A. Farhadi, “You Only Look Once: Unified, Real-Time Object Detection,” in Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR), 2016.

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

[3] Glenn Jocher et al., “Ultralytics YOLOv8,” 2023. [Online]. Available: https://github.com/ultralytics/ultralytics

[4] Jaided AI, “EasyOCR: Ready-to-use Optical Character Recognition (OCR),” 2020. [Online]. Available: https://github.com/JaidedAI/EasyOCR

[5] Yifu Zhang, P. Sun, Y. Jiang, D. Yu, F. Weng, and W. Yuan, “ByteTrack: Multi-Object Tracking by Associating Every Detection Box,” in European Conference on Computer Vision (ECCV), 2022.

[6] Nicolai Wojke, A. Bewley, D. Paulus, and D. Zisserman, “Simple Online and Realtime Tracking with a Deep Association Metric,” in IEEE International Conference on Image Processing (ICIP), 2017.

[7] OpenCV, “Open Source Computer Vision Library,” 2023. [Online]. Available: https://opencv.org/

[8] IEEE, “Traffic Monitoring and Violation Detection using Computer Vision: A Review,” IEEE Access, recent studies.