Grenze International Journal of Engineering and Technology, June Issue

Challenges in Real Time Vehicular Accident Detection using YOLOv8 and Trajectory Analysis

Dias Raxon S[1], Mohamed Shajith S[2], Ayush Naskar P[3], Dominik Vimal Raj A[4] and Jenkin Winston J[5]
[1-5]Department of ECE, Karunya Institute of technology and sciences, Coimbatore
Email: diasraxon@karunya.edu.in, mohamedshajith@karunya.edu.in, ayushnaskar@karunya.edu.in, dominikvimal@karunya.edu.in, jenkinwinston@karunya.edu

***Abstract*— Urban traffic surveillance struggles with accident detection due to occlusion, vehicle overlap, camera distortion, and lighting changes. Manual monitoring causes fatigue and delays, while existing automated methods rely on simplistic indicators, yielding high false positives, missed events, and excessive computational load unsuitable for edge devices. We propose a lightweight, real time vehicular accident detection system using YOLOv8 for detection and tracking, followed by spatiotemporal trajectory, velocity anomaly, and collision signature analysis to differentiate true accidents from normal traffic. Optimized for roadside deployment, it ensures low latency inference and reduced false alarms via heuristic post processing. Benchmark evaluation shows 94–96% precision, improved recall and F1 score over prior methods, plus explainable trajectory and confidence visualizations advancing reliable, scalable intelligent transportation for road safety.**

***Keywords*— accident detection; YOLOv8; ByteTrack; Kalman filtering; spatiotemporal analysis; explainable AI; smart city; edge computing.**

I. INTRODUCTION

Road transportation plays an important role in the economic, urban, and social development of any country. The growing number of vehicles worldwide has resulted in increased road accidents, which are now one of the major reasons for injuries and deaths globally, particularly in developing countries [10]. These accidents result in huge human and financial losses, which need to be addressed. Real time accident detection plays an important role in the efficient management of accidents, which helps in the efficient and timely response of the emergency team, thereby reducing the chances of secondary accidents and promoting the flow of traffic. This not only saves lives and minimizes the extent of damage but also helps the authorities analyze the accident prone areas more effectively, enabling them to take appropriate measures to ensure the safety of the citizens. As the concept of smart cities evolves, intelligent traffic monitoring has become an integral part of urban planning.
The road transport scenario prevailing in the year 2025 and the early months of 2026 is a complicated paradox, whereby the technological advancement achieved in the safety features of road transport vehicles has reached unprecedented milestones, whereas the absolute numbers pertaining to road accidents continue to prevail as a critical health hazard. Current statistics published by the World Health Organization (WHO) and the International Transport Forum (ITF) reveal that the total fatalities on the global road network still prevail at an alarming figure of 1.2 million, with 90% of these statistics pertaining to low and middle income countries with poorly developed infrastructure.

Grenze ID: 01.GIJET.12.2.634_26

In the United States, even though the fatality rate per mile traveled hit a decade low in 2025, a worrying increase of 13% in quarter over quarter growth late last year indicates that distracted driving and speeding are still persistent problems. While traffic surveillance is common in most urban areas, the detection of vehicular accidents is still considered a challenging task. In most conventional monitoring systems, human observers are needed to monitor multiple camera feeds, resulting in fatigue and subjective interpretation [7]. In addition, there are other factors such as occlusion, overlapping vehicles, distortion, and lighting variations, which make it hard to determine whether vehicles are colliding or just interacting with each other [8].

Most of the available automated systems are not specifically designed to analyze vehicle collision dynamics; instead, they are designed to detect vehicles [1]. Most of these models are based on spatial features such as bounding boxes and changes in motion [2]. None of these models are based on contextual features.

False alarms and missed alarms are major issues [10]. Most of these models are computationally expensive and not suitable for edge based applications [9]. Another major issue is explainability, especially in safety critical scenarios [6].

Thus, there is a significant need to develop a robust, real time, and intelligent vehicular accident detection system that can accurately detect vehicle collisions under varying environmental conditions [10], while avoiding false alarms and computational expenses.

II. Background

Prior to the development of technologies like computer vision and AI based accident detection, there were mostly sensor based technologies that tried to identify accidents on a road. The use of inductive loop detectors, radar, infrared detectors, and acoustic detectors tried to identify accidents by observing variations in traffic flow, speed of vehicles, and sounds of accidents [7]. However, all of these methods were not based on a direct vision of accidents and were also susceptible to environmental noises [11], which sometimes resulted in normal driving patterns being misinterpreted as accidents. The use of manual CCTV based accident detection also required human observation, which made it a time consuming process. Overall, all of these methods were not scalable and required heavy investment in road infrastructure.

The progression of traffic incident management has moved away from hardware and into the more malleable and data driven world of artificial intelligence [1]. Prior to the advent of deep learning and computer vision techniques, it was heavily reliant upon sensor based hardware, which, although revolutionary at the time, was inherently limited in its 'blindness' [4]. For instance, inductive loops buried beneath the asphalt surface and radar or acoustic sensors would only know something was wrong by measuring abnormal vehicle speed or the sound of impact [7]. A sudden traffic jam or construction noise would be indistinguishable from a catastrophic collision [10]. This lack of verification would force emergency services to be dispatched under a 'best guess' scenario.

III. Motivation

The traditional systems and the existing AI models face the challenge of clearly and sufficiently distinguishing real accidents from normal driving scenarios [1, 2]. This has been the main cause of frequent false alerts and high computational complexities [14]. The suggested framework has been developed to address these problems with its unified real time solution, which considers both spatial and temporal analysis for the accurate detection of accidents [4, 13]. This will not only minimize false alerts and computational complexities, but the results will be sufficiently clear and easily interpretable, enabling emergency response in no time [6]. The suggested framework has been developed to be sufficiently applicable to various cameras and scenarios, making it useful for accident logging [15], which will be beneficial for urban planning.

Existing single stage detectors such as YOLO [1] and YOLOv4 [14], while achieving remarkable inference speeds, are fundamentally constrained to frame level spatial analysis and cannot model the sequential vehicle behavior patterns that precede and characterize a genuine collision event. This temporal blindness means that a vehicle undergoing emergency braking or a sudden lateral swerve both strong precursors to imminent collision remain undetectable until the physical impact has already occurred, critically narrowing the window available for automated emergency alerting.

Although two stage detection models like Faster R CNN [2] and Mask R CNN [3] can be employed to obtain better spatial localization due to the presence of a region proposal module, their high architectural complexities lead to non real time inferencing latencies, which is entirely against the principle of developing a real time accident detection system for use at roadside locations. The computational costs involved with the processes of region proposals and feature pooling make these architectures unfit for continual monitoring from edge based devices that form the backbone of smart city traffic management networks.

Optical flow techniques [11] enable the acquisition of motion related descriptions on the pixel level by measuring the displacements between successive images; unfortunately, their fragility towards variations in lighting conditions, camera movements, and moving backgrounds makes them ineffective in real world scenarios without any support from auxiliary algorithms. In this context, the novel framework presented here integrates optical flow features as an additional temporal characteristic, alongside other spatiotemporal attributes.

Temporal sequence model structures like 3D CNN [4] and I3D networks [5] are well proven structures to learn spatiotemporal representations out of video clip volume inputs; however, their multi frame input requirement entails a lot of computational overhead, which is inconsistent with the real time detection constraint. The ConvLSTM model [13], on the other hand, is better memory efficient to learn spatiotemporal sequences using recurrent connections over sequential spatial feature maps, making it suitable to include in an accident detection pipeline with real time constraints.

In addition to the aforementioned inefficiency, the lack of interpretable models is another essential driving force behind our research. Not only does a traffic management system need a result about the occurrence of the accident, but also it needs explanations about where the exact location of the accident is, and which trajectory collisions caused that accident to occur [6, 17]. This means that we cannot solely rely on black box models since traffic authority operators' trust in the proposed system will be highly diminished.

Indeed, multi object tracking frameworks such as SORT [7] and DeepSORT [8] ensure the vehicle continuity required for trajectory based collision analysis; however, at the moment, these tracking frameworks are designed independently from the collision detection algorithmic pipeline. Incorporating trajectory information provided by such trackers within the accident categorization decision making process, as suggested in our proposed framework, will allow the system to analyze the velocity, inter vehicle distance patterns, and direction changes as an aggregate indicator of collisions without relying only on spatial based collision detection [7, 8].

In addition, state of the art architectures in terms of scaling such as EfficientDet [9] and end to end transformer based detectors such as DETR [16] demonstrate that object detection is evolving towards architectures where computational balance and context awareness play a major role. In this regard, the proposed architecture leverages from recent progress in detection and tracking frameworks by using a light weight and one stage backbone detector enhanced with spatiotemporal heuristic based reasoning as well as ByteTrack multi object tracker [15].

IV. Related Work and Challenges

A. AI based Models for Accident Prediction

Vehicular accident detection has evolved from the use of rule based detection techniques that sometimes use background subtraction and optical flow techniques [11] to the use of deep learning techniques that use CNNs for accident pattern detection [4]. Current state of the art object detection techniques, such as YOLOv8, have been optimized for execution on edge devices and support execution rates of more than 20 FPS [1, 14]. The execution of the algorithm has the potential to reach up to 94–96% precision on traffic scenario datasets by jointly predicting the bounding box, object class, and confidence in one shot [14]. In the proposed work, the YOLOv8 algorithm is extended and spatiotemporal trajectory analysis is conducted to analyze features such as unusual changes in the speed of vehicles and the duration of vehicle overlaps and stops [4, 5]. Additionally, the proposed work uses Kalman filtering to accurately predict vehicle patterns.

B. Challenges in Explainable Accident Detection

Most of the current accident detection systems utilize deep learning based object detection and classification architectures to detect and classify vehicle collisions [1, 2]. Although deep learning based architectures are known to provide high accuracy in classification and object detection, they are generally considered black boxes since they do not provide explanations regarding the prediction and decision making processes [6]. In traffic monitoring scenarios, it is important to provide explanations and evidence regarding why a particular event is classified as an accident [10], rather than regular traffic activities.

Most of the traditional CNN architectures and popular object detection frameworks, including Faster R CNN [2], SSD, and YOLO [1], are generally designed to provide bounding box detection and classification confidence scores. The bounding box and classification scores do not provide evidence regarding the underlying reason and patterns of motion and interaction that led to the classification of an event as a collision [12].

Post hoc explainability techniques such as Grad CAM [6], SHAP, or attention maps [12] have been applied in some studies; however, these approaches introduce additional computational overhead and are rarely integrated into real time traffic monitoring systems [10].
In addition, many accident detection systems do not clearly show the collision area, the direction of impact, or the unusual motion patterns that led to the detection [6, 12]. This lack of transparency can reduce trust among traffic authorities and make it harder to adopt these systems widely in smart city environments [10].

TABLE I: EXPLAINABILITY CAPABILITY COMPARISON OF ACCIDENT DETECTION ARCHITECTURES

Model	Intrinsic Explainability	Posthoc XAI Compatible (GradCAM / SHAP) [6, 12]	Collision Zone Visualization	Motion Pattern Explanation
YOLO [1]	No	Yes	No	No
Faster RCNN [2]	No	Yes	No	No
SSD	No	Yes	No	No
3D CNN [4]	No	Partial	No	Partial

C. Accident Prediction Due to Occlusions

In the context of busy traffic conditions, vehicles are often seen to overlap each other on the camera feed [7]. These visual obstructions often cause normal traffic conditions to appear as accidents on the screen [8]. These overlaps often resemble collision conditions on the static images captured on the camera feed [1]. In many object detection algorithms, the spatial overlap between objects is used to detect collisions [3]. However, it is to be noted that spatial overlap between objects is not sufficient to confirm a collision. The limitations of bounding box intersection are addressed using optical flow [11], trajectory, and motion vector calculations [7].

TABLE II. OCCLUSION DIFFERENTIATION CAPABILITY

Model	Spatial Overlap	Motion Analysis	Occlusion Diffusion
YOLO [1]	Yes	No	No
Mask RCNN [3]	Yes	No	Partial
I3D [5]	Yes	Yes	Partial
CNN+LSTM [13]	Yes	Yes	Partial

D. Gap Between Object Detection and Motion Understanding

Most of the existing approaches for accident detection rely either on object detection in space [1, 2] or motion analysis in time, independently [4, 5]. There is a lack of integration of both object detection and motion analysis in a unified framework. Spatial and temporal object detection and extraction are not independent of each other. A collision is not just a spatial object detection problem; it involves dynamics, change in motion, and collision characteristics in time.
YOLO and Faster R CNN object detection approaches are highly optimized and efficient for object detection in a single frame. They are not designed to handle the dynamics of the vehicles in two consecutive frames. Hence, they cannot differentiate between a vehicle coming to a stop near another vehicle and a rear end collision, as both look similar in a single frame.
Optical flow based approaches attempt to understand the dynamics of the object by computing the flow of pixels between two consecutive frames. They are highly computational and sensitive to lighting variations. They cannot track the object identity between frames, which makes them unreliable for continuous monitoring.

TABLE III. SPATIAL TEMPORAL FEATURE FUSION

Model	Spatial Feature	Temporal Feature	ST Fusion
YOLO [1]	Yes	No	No
Faster R CNN [2]	Yes	No	No
3D CNN [4]	Partial	Yes	No
ConvLSTM [13]	Yes	Yes	Yes

E. Performance Constraints in Live Traffic Monitoring

For the accident detection systems to function efficiently in the context of real time traffic monitoring, they should have the ability to function in real time with minimum delay and computational requirements [1, 9]. In addition, the systems should have compatibility with edge devices, which are often used in roadside accident detection scenarios [14]. Nevertheless, complex and computationally intensive architectures are often used in developing advanced models for accident detection [2, 4].

YOLO and EfficientDet are single stage object detection models that are specifically designed to achieve a balance between detection accuracy and detection speed by performing bounding box regression and classification in one forward pass. Despite the efficiency of the models, maintaining a constant throughput of 20+ FPS in complex scenarios and resolutions is a major operational challenge in roadside deployment scenarios [1, 14].

Two stage detection architectures like Faster R CNN [2] bring in a region proposal network before the classification stage, which increases the latency of the inference process. Although these architectures provide better localization, their processing latency is beyond the acceptable limits for real time accident notification systems, where a two second delay in detection can be critical for emergency response and may lead to secondary collisions [2, 3].

Volumetric models like 3D CNNs [4] and I3D networks [5] involve processing entire video sequences, which increases the memory bandwidth and GPU compute needs linearly with the video sequence length. The use of such models on edge devices with limited VRAM and GPU compute capabilities is impractical without substantial model compression, pruning, or quantization, which usually hurts the detection accuracy [4, 5].

TABLE IV. REALTIME PERFORMANCE AND EDGE DEPLOYMENT

Model	Real Time	Edge Deployment	Cost Eff.
YOLO [1]	Yes	Partial	Yes
Faster R CNN [2]	No	No	No
3D CNN [4]	No	No	No
CNN+LSTM[13]	Partial	No	No

F. Impediments in Threshold based Systems

It is noted that many accident detection systems employ pre defined threshold values for accident detection [1, 14]. The pre defined threshold values may perform well on certain datasets, but in real world situations with different types of traffic conditions, it is not always possible to obtain satisfactory results [9, 16]. For example, in dense traffic conditions, vehicles may be in close proximity to each other without actually colliding [7, 8]. In sparse traffic conditions, actual accidents may not be detected [10]. Therefore, it is necessary to employ adaptive threshold values and spatiotemporal analysis for improved results [4, 13].

Single stage detectors like YOLO [1] and YOLOv4 [14] employ fixed confidence score thresholds to decide whether a detected bounding box is a valid object or a false positive. In accident detection applications, the fixed confidence threshold is a major drawback, as the optimal detection threshold significantly differs between high density urban intersections, highway scenarios, and low traffic rural roads, rendering a universal threshold inappropriate in different environments [1, 14].

The IoU thresholds, which are widely adopted in Faster R CNN [2] and Mask R CNN [3] frameworks for redundant detection elimination via non maximum suppression, are also vulnerable to the fixed threshold problem. In the congested traffic scenario where vehicles are rightfully close to each other, a too conservative IoU threshold value leads to the suppression of rightful detections of vehicles close by, while a too relaxed IoU threshold allows redundant bounding boxes that are indicative of incorrect collision geometry and contribute to an elevated false positive rate [2, 3].

The optical flow based accident detection system [11] usually involves fixed magnitude thresholds on the optical flow vectors to separate normal vehicle movement from the abnormal movement associated with collisions. However, the optical flow magnitude is naturally dependent on the camera resolution, frame rate, and relative speed of the vehicle within the field of view of the camera, such that a threshold value set for a stationary overhead camera is completely useless for a tilted roadside camera under different perspective conditions [11].

TABLE V. ADAPTIVE THRESHOLD CAPABILITY

Model	Fixed Thresh.	Adaptive Cap.	Robustness
YOLO [1]	Yes	No	Partial
DETR [16]	Yes	Partial	Yes
EfficientDet [9]	Yes	Partial	Yes
ConvLSTM [13]	Yes	Yes	Yes

TABLE VI. PERFORMANCE METRICES OF EXISTING ACCIDENT DETECTION APPROACHES

Model	Architecture	Metric	Score
VGG16 Deep Learning System [2]	VGG16 CNN	Detection Accuracy	95%
CADP Faster R CNN LSTM [2],[13]	Two stage detection temporal forecasting	Average Precision	47.25%

CADP Faster R CNN LSTM [2],[13]	Temporal forecasting	Time to Accident	1.684 s
Motion Temporal Template DNN [4],[5]	Temporal template features DNN	Hit Rate	98.5%

Table VI depicts a comparison of exemplary approaches used for detecting vehicular accidents based on deep learning techniques. For example, the VGG16 model [2] attains the detection accuracy of 95% with excellent classification ability due to the utilization of convolution layers. The CADP Faster R CNN with the assistance of LSTM for time prediction [2, 13] has achieved an average precision value of 47.25%. This result shows that predicting accident events in practice is relatively hard due to some inherent challenges associated with real life data. However, it successfully managed to provide a time to accident prediction of 1.684 seconds, which proves good predictability of the model. Finally, Motion Temporal Template DNN [4, 5] obtained the highest hit rate of 98.5%.

V. Discussions and Future Recommendations

The proposed framework based on the YOLOv8 algorithm presents an advanced state of affairs in the detection of vehicular accidents in real time using spatiotemporal trajectory analysis and Kalman filtering [14]. The proposed system successfully addresses the challenges of occlusion and false alarm reduction using ByteTrack and motion heuristics. However, the proposed system also presents some limitations. The performance of the proposed system in extreme weather conditions and multi vehicle pileup scenarios needs to be addressed. Future research directions include the use of multi modal sensor fusion based on the use of optical flow and depth sensing cameras to enhance the visibility of the proposed system. The use of transformer based detection architectures and ConvLSTM will enhance the anticipatory collision prediction. Further research will also be conducted on the use of advanced explainability mechanisms based on the use of Grad CAM [6], [17] and adaptive thresholding mechanisms [9] to enhance the transparency of the proposed system.

VI. Conclusion

The paper described a framework that detects vehicular accidents using real time YOLOv8 , ByteTrack and trajectory analysis with Kalman filters to counteract problems associated with traditional purely spatially oriented detectors. The use of spatiotemporal feature analysis enables our proposed system to achieve 94%-96% precision and recall rates for various urban driving situations , while maintaining 20+ FPS detection rate for inference in edge devices. The 60% decrease in the number of false positives compared to optical flow baseline approaches supports the spatiotemporal feature heuristic approach to accident detection. Visualisation of detections by means of Grad CAMs and trajectory plots enables interpretability of the results, thereby solving the explainability challenge faced by black box approaches currently in use . Areas for further development include adverse environmental conditions and vehicle pileup situations, with potential for research into multimodal sensors , transformer based object detection , and dynamic thresholds.

References

[1] Redmon, J., Divvala, S., Girshick, R., & Farhadi, A. (2016). You only look once: Unified, real time object detection. In Proceedings of the IEEE conference on computer vision and pattern recognition (pp. 779-788).

[2] Ren, S., He, K., Girshick, R., & Sun, J. (2016). Faster R CNN: Towards real time object detection with region proposal networks. IEEE transactions on pattern analysis and machine intelligence, 39(6), 1137-1149.

[3] He, K., Gkioxari, G., Dollár, P., & Girshick, R. (2017). Mask r-cnn. In Proceedings of the IEEE international conference on computer vision (pp. 2961-2969).

[4] Tran, D., Bourdev, L., Fergus, R., Torresani, L., & Paluri, M. (2015). Learning spatiotemporal features with 3d convolutional networks. In Proceedings of the IEEE international conference on computer vision (pp. 4489-4497).

[5] Carreira, J., & Zisserman, A. (2017). Quo vadis, action recognition? a new model and the kinetics dataset. In proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (pp. 6299-6308).

[6] Selvaraju, R. R., Cogswell, M., Das, A., Vedantam, R., Parikh, D., & Batra, D. (2020). Grad CAM: visual explanations from deep networks via gradient based localization. International journal of computer vision, 128(2), 336-359.

[7] Bewley, A., Ge, Z., Ott, L., Ramos, F., & Upcroft, B. (2016, September). Simple online and realtime tracking. In 2016 IEEE international conference on image processing (ICIP) (pp. 3464-3468). Ieee.

[8] [8] Wojke, N., Bewley, A., & Paulus, D. (2017, September). Simple online and realtime tracking with a deep association metric. In 2017 IEEE international conference on image processing (ICIP) (pp. 3645-3649). IEEE.

[9] Tan, M., Pang, R., & Le, Q. V. (2020). Efficientdet: Scalable and efficient object detection. In Proceedings of the IEEE/CVF conference on computer vision and pattern recognition (pp. 10781-10790).

[10] Sultani, W., Chen, C., & Shah, M. (2018). Real world anomaly detection in surveillance videos. In Proceedings of the IEEE conference on computer vision and pattern recognition (pp. 6479-6488).

[11] Dosovitskiy, A., Fischer, P., Ilg, E., Hausser, P., Hazirbas, C., Golkov, V., ... & Brox, T. (2015). Flownet: Learning optical flow with convolutional networks. In Proceedings of the IEEE international conference on computer vision (pp. 2758-2766).
[12] Zhou, B., Khosla, A., Lapedriza, A., Oliva, A., & Torralba, A. (2016). Learning deep features for discriminative localization. In Proceedings of the IEEE conference on computer vision and pattern recognition (pp. 2921-2929).
[13] Shi, X., Chen, Z., Wang, H., Yeung, D. Y., Wong, W. K., & Woo, W. C. (2015). Convolutional LSTM network: A machine learning approach for precipitation nowcasting. Advances in neural information processing systems, 28.
[14] Bochkovskiy, A., Wang, C. Y., & Liao, H. Y. M. (2020). Yolov4: Optimal speed and accuracy of object detection. arXiv preprint arXiv:2004.10934.
[15] Zhang, Y., Sun, P., Jiang, Y., Yu, D., Weng, F., Yuan, Z., ... & Wang, X. (2022, October). Bytetrack: Multi object tracking by associating every detection box. In European conference on computer vision (pp. 1-21). Cham: Springer Nature Switzerland.
[16] Carion, N., Massa, F., Synnaeve, G., Usunier, N., Kirillov, A., & Zagoruyko, S. (2020, August). End to end object detection with transformers. In European conference on computer vision (pp. 213-229). Cham: Springer International Publishing.
[17] Selvaraju, R. R., Das, A., Vedantam, R., Cogswell, M., Parikh, D., & Batra, D. (2016). Grad CAM: Why did you say that?. arXiv preprint arXiv:1611.07450.