Vehicle Collision Detection Model Based on Multiple Traffic Factors

Xin Cheng, Jingmei Zhou*, Lizhuoyi Mi, Can Cheng, Chongfeng Wei

***Abstract*- Real-world traffic scenes present significant visual complexity and dynamic variability, which challenge robust collision detection. This paper proposes a multi-factor collision detection framework combining an improved YOLOv11s detector with ByteTrack-based tracking. The detection backbone is enhanced by integrating Partial Convolution (PConv) and a Contextual Gated Linear Unit (CGLU) into a redesigned FC block to better capture small and distant objects. A multi-factor collision judgment model fuses IoU, abrupt velocity change, angle deviation, and track-ID discontinuity in a weighted manner to reduce false alarms and misses. Experiments on a constructed roadside surveillance dataset (8127 images; collision and non-collision video sequences) show that the proposed YOLOv11s_FC detector improves detection metrics over baseline YOLO variants and that the multi-factor collision model achieves 80.76% accuracy while reducing both false alarm and miss rates.**

***Index Terms*- collision detection, YOLOv11, multi-factor fusion**

I. INTRODUCTION

The rapid development of Intelligent Transportation Systems (ITS) has significantly improved the efficiency of urban traffic monitoring and management [1]. Among various ITS applications, automatic traffic accident detection has become an important research topic due to its potential to reduce accident response time and improve road safety. However, conventional traffic surveillance systems still rely heavily on manual video inspection, which is inefficient and prone to errors under long-term monitoring conditions [2].

Existing vehicle collision detection methods can generally be categorized into sensor-based and vision-based approaches. Sensor-based methods employ devices such as LiDAR or millimeter-wave radar to detect potential collision risks by perceiving the surrounding environment of vehicles [3]. Although these methods provide accurate environmental perception, they often involve expensive hardware and complicated installation processes.

Vision-based collision detection methods utilize roadside surveillance cameras and have become increasingly popular due to their low cost and flexible deployment [4]. However, these approaches still face two major challenges. First, object detection and tracking accuracy in traffic surveillance scenarios is limited, particularly for small or distant objects, where traditional models such as Faster R-CNN or YOLOv5 often suffer from missed detections [5]. Second, many existing collision detection models rely on single-factor indicators, such as bounding box overlap or velocity thresholds, which fail to capture the complex dynamics of vehicle collisions [6].

To address these issues, this paper proposes a multi-factor vehicle collision detection framework based on an improved YOLOv11s detection model and the ByteTrack tracking algorithm. By integrating multiple motion and spatial indicators, the proposed approach improves the reliability of collision detection in complex traffic environments.

II. RELATED WORK

A. Object detection and lightweight modules

In recent years, single-stage detectors based on deep learning have dominated traffic scenarios due to their inherent speed advantages. For instance, YOLOv5 achieves over 80% MAP@0.5 in traffic object detection through its CSPDarknet backbone [7]. YOLOv7 introduced "trainable freebies" to further boost performance [8], while the YOLOv8 series optimized multi-scale pooling to enhance feature extraction. However, these models still suffer from high false-negative rates for small, distant objects (e.g., pedestrians) in surveillance scenarios. To balance execution efficiency and detection accuracy, lightweight structures have been explored. Chen et al. [9] proposed the Partial Convolution (PConv) strategy, applying linear transformations only to selected channels to replace deep convolutions, significantly enhancing computational efficiency. Similarly, the Contextual Gated Linear Unit (CGLU) [10] mechanism has been shown to dynamically filter and recalibrate channel-wise contextual information, which is crucial for detecting occluded objects. While Transformer-based models like RT-DETR achieve state-of-the-art precision, their self-attention mechanisms incur excessive Memory Access Costs (MAC) and inference latency. These hardware bottlenecks prohibit their deployment on resource-constrained roadside edge devices requiring millisecond-level responsiveness. Recent studies [11-12] confirm that CNN-based YOLO architectures provide a superior latency-accuracy trade-off for high-speed traffic scenarios. Consequently, our research exclusively focuses on optimizing the CNN paradigm.

B. Multi-object tracking

The primary goal of multi-object tracking is to maintain consistent trajectories. Traditional Tracking-by-Detection (TBD) and Joint Detection and Embedding (JDE) paradigms typically discard candidate bounding boxes with confidence scores below a predefined threshold. This often leads to fragmented trajectories or target loss during short-term occlusions or lighting changes. To address this, the ByteTrack algorithm [13] significantly improves data association by categorizing all detection results into high-confidence and low-confidence groups for sequential

*This work was supported in part by National Natural Science Foundation of China under Grants 52302491 and 52472337, Research Funds for the Interdisciplinary Projects, CHU under Grant 300104240911.

X. Cheng and C. Cheng are with the School of Information Engineering, Chang'an University, Xi'an 710018, China.

J. Zhou and L. Mi are with the School of Electronics and Control Engineering, Chang'an University, Xi'an 710018, China. (corresponding author e-mail: jmzhou@chd.edu.cn)

C. Wei is with the James Watt School of Engineering, University of Glasgow, G12 8QQ, UK.

matching, effectively recovering heavily occluded targets and improving trajectory continuity.

C. Collision detection

Collision detection models rely on target trajectory features to quantify risk. Early methods favored single-factor detection due to simplicity. For example, Luo et al. [14] relied solely on velocity change rates. These approaches often yield high false alarm rates (e.g., triggering alerts during temporary occlusions) or miss low-speed congestion collisions. Fildes et al. [15] employed geometric analysis based on trajectory intersections, which was limited to straight roads. To improve adaptability, recent studies attempt multi-feature fusion. Bayoudh et al. [16] combined IoU and velocity to improve accuracy, but prior multi-feature models often apply equal weighting to all factors and ignore track-ID jumps caused by tracking interruptions, leaving room for improvement in complex urban scenarios.

III. Research Methodology

To achieve robust vehicle collision detection in highly dynamic and complex traffic environments, we propose a comprehensive three-stage framework. This framework consists of an edge-efficient object detection module, a spatio-temporal tracking pipeline, and a hierarchical multi-factor collision judgment model.

A. The Novel FC Block Design

Deploying real-time collision detection on edge devices demands high precision and low computational overhead. To meet this, we upgrade the YOLOv11s C3k2 modules with a novel "FC Block," integrating Partial Convolution (PConv) from FasterNet and the Contextual Gated Linear Unit (CGLU) from TransNext. Crucially, the FC Block provides a necessary synergy over using these modules independently. While PConv alone reduces FLOPs and memory latency, it risks suboptimal channel-wise feature interaction. Conversely, deploying CGLU with standard convolutions creates unacceptable computational bottlenecks for edge devices. Combined, PConv frees up essential computational headroom, allowing CGLU to dynamically recalibrate channel-wise attention and fully compensate for PConv's partial channel processing. Furthermore, this combination is specifically tailored for traffic collision detection. Roadside surveillance involves noisy backgrounds and the need to instantly identify fast-moving, distant, or overlapping vehicles. Here, PConv guarantees the high frame rates crucial for capturing split-second collision dynamics. Simultaneously, CGLU filters out complex environmental noise (e.g., weather, illumination changes) and highlights critical vehicle cues. This synergistic architecture ensures highly accurate bounding box localization for complex traffic targets without compromising real-time edge execution speeds.

B. Joint Detection and Tracking

In complex traffic scenes, vehicles frequently experience severe short-term occlusions, leading to trajectory fragmentation and identity (ID) switches in traditional tracking paradigms. To address this, we design a joint detection-and-tracking pipeline tightly integrated with the ByteTrack algorithm. As illustrated in Figure 1. , rather than prematurely discarding bounding boxes that fall below a predefined confidence threshold, our improved YOLOv11s network explicitly partitions all extracted bounding boxes, object categories, and confidence scores into "High-Score" and "Low-Score" data streams. Concurrently, a Kalman filter is employed to mathematically model the kinematics of the vehicles and predict their historical trajectories in the current frame. The core data association utilizes a robust two-stage Hungarian matching strategy. Initially, the predicted tracklets are matched against the high-score bounding boxes based on spatial Intersection over Union (IoU) similarity. Subsequently, any tracklets that remain unmatched-typically due to sudden occlusions or severe motion blur-are re-associated using the low-score detection stream. By actively rescuing these low-confidence but spatially relevant features, our pipeline effectively recovers heavily occluded vehicles and significantly mitigates ID switches, ensuring continuous spatio-temporal tracking without introducing additional computational overhead.

C. Multi-factor Collision Judgment

Traditional collision detection models that rely on a single factor, such as mere bounding box overlap (IoU) or simple proximity, frequently trigger false alarms during routine close-proximity driving (e.g., traffic jams or overtaking). To systematically eliminate these false positives, we formulate a hierarchical, temporal multi-factor collision condition model as shown in Figure 2.

For any vehicle i at frame t, the system continuously retrieves its forecasted trajectory states over a sliding observation window of N=15 frames. The model executes a primary risk filter by evaluating the minimum Euclidean spatial distance between vehicle i and any surrounding vehicle j. If this distance strictly falls below a predefined safety threshold $D_{threshold}$, a secondary, comprehensive kinematic assessment is triggered.

A definitive collision event (C_{event}) is confirmed only if all subsequent multi-dimensional criteria are simultaneously satisfied:

$$C_{event} = \mathrm{I}\left(\mathrm{IoU}_{i,j} > \tau_{iou}\right) \wedge \mathrm{I}\left(\max(|\Delta v|, |\Delta\theta|) > \tau_{kin}\right) \wedge \mathrm{I}\left(\mathrm{ID}_{switch}\right) \quad (1)$$

Where $\mathrm{I}(\cdot)$ is the indicator function. This models three physical crash phenomena: (1) Bounding box overlap exceeds τ_{iou} ; (2) Velocity or angular deviation undergoes abrupt structural mutation (τ_{kin}); (3) The tracking identity (track_id) experiences a sudden discontinuity due to severe morphological deformation during impact.

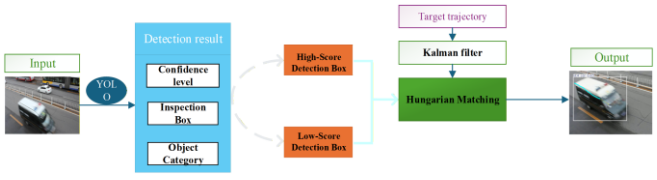


Figure 1. Joint Model Flowchart

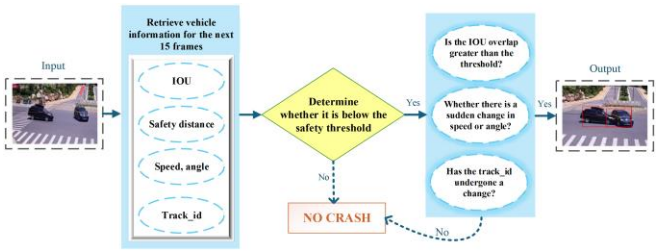


Figure 2. Overall Flowchart of the Collision Condition Model

IV. EXPERIMENTS

A. Dataset and Implementation Details

To establish a robust training and evaluation environment, we constructed a comprehensive roadside traffic dataset comprising 8127 images across 6 object categories.This dataset was built by augmenting several standard public datasets—including (UA-DETRAC [17], CityFlow [18], CADP [19], TAD [20]) with manually supplemented pedestrian annotations. To validate the collision condition logic, we further collected and processed a video dataset containing 2401 collision events and 1352 non-collision sequences. During the training phase, the proposed models were optimized using the AdamW optimizer with an initial learning rate of 0.001 (adjusted via a cosine annealing scheduler). The training was conducted with a batch size of 64 and an input resolution of 640×640 over 500 epochs. To comprehensively evaluate model performance, standard metrics including Mean Average Precision (mAP@0.5), Precision, Recall, F1-score, False Alarm Rate (FAR), and Miss Rate (MR) were utilized.

B. Detection and Tracking Performance

As shown in TABLE I. ,RT-DETR-L (31.01 M params, 108.34 GFLOPs) is impractical for edge deployment. Our proposed YOLO11s-FC overcomes this by requiring only 7.75 M parameters and 19.86 GFLOPs, yielding an 18.1% parameter reduction against standard YOLO11s. Despite this drastic compression, TABLE II. reveals that YOLOv11s-FC outperforms the baseline, boosting mAP@0.5 from 85.42% to 87.33%, alongside improved Precision (89.78%) and Recall (79.85%). These results validate that our FC Block (PConv + CGLU) successfully enhances small-object detection while maintaining an optimal Pareto balance for real-time roadside applications.

C. Collision Condition Model Analysis

Quantitative evaluations confirm the efficacy of our architectural improvements and the necessity of unequal feature weighting. TABLE III. presents an ablation study comparing different weight assignments. Assigning a higher weight to velocity variation (Group A: ω_1=0.5, ω_2=0.3, ω_3=0.2) yields a peak detection accuracy of 80.8%, significantly outperforming a balanced equal-weighting strategy (75.3%). This confirms our hypothesis that abrupt kinematic changes possess the most significant discriminative capability for identifying physical impacts. Furthermore, as shown in TABLE IV. , integrating YOLOv11s-FC with the collision condition model improves accuracy by 2.58% over the standard YOLOv11s. Notably, incorporating track_id jump detection provides an additional 2.54% accuracy boost while simultaneously reducing false negative and false positive rates by 2.21% and 1.86%, respectively. This demonstrates the tracking module's robust capability to resolve identity switches caused by severe collision-induced occlusions or appearance alterations.

D. Case Study

Qualitative visual evaluations further demonstrate the practical robustness of the proposed framework. Figure 3. illustrates the detection performance of YOLOv11s and the YOLOv11s_FC model from this section on the dataset. The comparison reveals that the baseline YOLOv11s model exhibits false negatives or bounding box misalignment when handling distant objects (e.g., pedestrians at a distance). The improved YOLOv11s_FC model, after integrating the FasterNet architecture with the CGLU module, enhances perception capabilities for distant small objects. In Figure 3. YOLOv11s_FC achieves more accurate detection of pedestrians and vehicles, outperforming the original model in both detection precision and recognition stability, demonstrating the effectiveness of the structural improvements.

Additionally, Figure 4. demonstrates the detection and tracking performance of the combined YOLOv11s_FC and ByteTrack model on the test dataset. Across diverse traffic scenarios, the model not only stably detects foreground vehicles and pedestrians but also performs continuous tracking of targets, indicating the proposed detection-tracking joint model exhibits robust spatio-temporal continuity. Finally, the definitive collision detection alerts, successfully triggered by the multi-factor logic, are presented in Figure 5.

TABLE I. COMPARISON OF MODEL EFFICIENCY

Model	Parameters(M)	GFLOPs
YOLOv8s	11.17	28.82
YOLOv11s	9.46	21.72
YOLOv12s	9.29	21.69
RT-DETR-L	31.01	108.34
YOLOv11s_FC(Ours)	**7.75**	**19.86**

TABLE II. PERFORMANCE COMPARISON OF OBJECT DETECTION MODELS

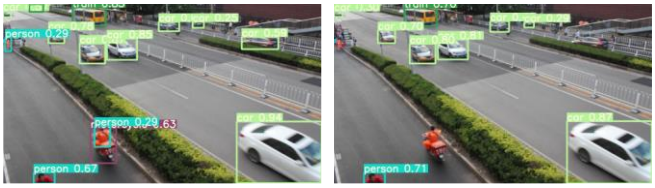
Model	mAP@0.5 (%)	Precision (%)	Recall (%)	F1-Score
Faster R-CNN	81.21	88.56	72.31	0.797
YOLOv5s	81.70	89.01	72.82	0.801
YOLOv8s	83.61	88.74	77.12	0.825
YOLOv11s	85.42	89.12	78.13	0.832
YOLOv12s	85.63	89.08	78.36	0.834
YOLOv11s_FC(Ours)	**87.33**	**89.78**	**79.85**	**0.854**

TABLE III. ABLATION STUDY ON MULTI-FACTOR WEIGHT ASSIGNMENTS

Group	Speed Var.ω_1	Angle Var.ω_2	Angle Change ω_3	Acc (%)
A	**0.50**	**0.30**	**0.20**	**80.8**
B	0.34	0.33	0.34	75.3
C	0.20	0.40	0.40	72.6

TABLE IV. EVALUATION RESULTS OF COLLISION DETECTION MODELS

Model Pipeline	Acc(%)	FPS	MR(%)	FAR(%)
YOLOv11s+ByteTrack	75.64	26	23.58	25.84
YOLOv11s_FC+ByteTrack	78.22	28	21.45	22.67
(Full Pipeline) Ours	**80.76**	**28**	**19.24**	**20.81**



(a) YOLOv11s_FC (b) YOLOv11s

Figure 3. Traffic Object Detection Results

Figure 4. Traffic Object Detection and Tracking Result

Figure 5. Collision Detection Results

V. CONCLUSION

This paper presents a robust, edge-friendly vehicle collision detection framework tailored for the visual complexity of real-world traffic environments. The importance of this work lies in providing a highly accurate, vision-based alternative to expensive sensor-based systems. By successfully addressing the chronic issues of small-object misdetection and trajectory fragmentation, and by abandoning simplistic single-factor collision logic in favor of a weighted multi-dimensional kinematic assessment, the proposed system significantly suppresses false alarms that plague conventional surveillance setups.

For future extensions, we suggest expanding the dataset to encompass extreme adverse weather conditions (such as heavy rain, snow, or dense fog) to further test model resilience. Additionally, incorporating 3D bounding box estimation from monocular vision could provide more precise spatial distance metrics, potentially elevating the collision judgment model's accuracy to an even higher level.

REFERENCES

[1] N. Nigam, D. P. Singh, and J. Choudhary, "A review of different components of the intelligent traffic management system (ITMS)," *Symmetry*, vol. 15, no. 3, Art. no. 583, pp. 1-21, 2023.

[2] H. F. Yang, J. Cai, C. Liu, and S. Li, "Cooperative multi-camera vehicle tracking and traffic surveillance with edge artificial intelligence and representation learning," *Transp. Res. Part C: Emerg. Technol.*, vol. 148, Art. no. 104052, pp. 1-22, 2023.

[3] M. S. Almutairi, K. Almutairi, and H. C. Roma, "Selecting features that influence vehicle collisions in the Internet of Vehicles based on a multi-objective hybrid bi-directional NSGA-III," *Appl. Sci.*, vol. 13, no. 4, Art. no. 2064, pp. 1-19, 2023.

[4] C. Wang, Y. Dai, W. Zhou, and Y. Geng, "A Vision-Based Video Crash Detection Framework for Mixed Traffic Flow Environment Considering Low-Visibility Condition," *J. Adv. Transp.*, vol. 2020, Art. no. 9194028, pp. 1-11, 2020.

[5] M. Kilicarslan and J. Y. Zheng, "Direct vehicle collision detection from motion in driving video," in *Proc. IEEE Intell. Vehicles Symp. (IV)*, 2017, pp. 1215-1221

[6] L. J. Zhang, "Detection method for collision events and collision damages in moving vehicles," *Comput. Syst. Appl.*, vol. 29, no. 11, pp. 157-162, 2020.

[7] B. Setiyono, D. A. Amini, and D. R. Sulistyaningrum, "Number plate recognition on vehicle using YOLO-Darknet," *J. Phys.: Conf. Ser.*, vol. 1821, no. 1, Art. no. 012049, pp. 1-10, 2021.

[8] C. Y. Wang, A. Bochkovskiy, and H. Y. M. Liao, "YOLOv7: Trainable bag-of-freebies sets new state-of-the-art for real-time object detectors," *arXiv preprint arXiv:2207.02696*, 2022.

[9] J. Chen, S. Kao, H. He, W. Zhuo, S. Wen, C. H. Lee, and S. L. Huang, "Run, don't walk: Chasing higher FLOPS for faster neural networks," in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit. (CVPR)*, 2023, pp. 12021-12031.

[10] Y. N. Dauphin, A. Fan, M. Auli, and D. Grangier, "Language modeling with gated convolutional networks," in *Proc. Int. Conf. Mach. Learn. (ICML)*, 2017, pp. 933-941.

[11] A. Parekh and M. Bauer, "Comparative analysis of YOLOv8 and RT-DETR for real-time object detection in advanced driver assistance systems," International Conference on Computational Science and Computational Intelligence, pp. 26-39, 2024.

[12] S.-E. Tsai and C.-H. Hsieh, "A real-time collision warning system for autonomous vehicles based on YOLOv8n and SGBM stereo vision," Electronics, vol. 14, p. 4275, 2025.

[13] Y. Zhang, P. Sun, N. Yi, Z. Jiang, S. Wang, et al., "ByteTrack: Multi-object tracking by associating every detection box," *arXiv preprint arXiv:2110.06864*, 2021.

[14] Q. Luo, X. Zang, J. Yuan, X. Chen, J. Yang, and S. Wu, "Research of vehicle rear-end collision model considering multiple factors," *Math. Probl. Eng.*, vol. 2020, Art. no. 8835848, pp. 1-11, 2020.

[15] B. Fildes, M. Keall, N. Bos, A. Lie, Y. Page, et al., "Effectiveness of low speed autonomous emergency braking in real-world rear-end crashes," *Accid. Anal. Prev.*, vol. 81, pp. 24-29, 2015.

[16] K. Bayoudh, F. Hamdaoui, and A. Mtibaa, "Transfer learning based hybrid 2D-3D CNN for traffic sign recognition and semantic road detection applied in advanced driver assistance systems," *Appl. Intell.*, vol. 51, no. 1, pp. 124-142, 2021.

[17] L. Wen, D. Du, Z. Cai, Z. Lei, M. C. Chang, et al., "UA-DETRAC: A new benchmark and protocol for multi-object detection and tracking," *Comput. Vis. Image Underst.*, vol. 193, Art. no. 102908, pp. 1-13, 2020.

[18] Z. Tang, M. Naphade, M. Y. Liu, X. Yang, S. Birchfield, et al., "CityFlow: A city-scale benchmark for multi-target multi-camera vehicle tracking and re-identification," in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit. (CVPR)*, 2019, pp. 8789-8798.

[19] P. Shah, J. B. Lamare, T. Nguyen-Anh, and A. Hauptmann, "CADP: A novel dataset for CCTV traffic camera based accident analysis," in *Proc. 15th IEEE Int. Conf. Adv. Video Signal Based Surveill. (AVSS)*, 2018, pp. 1-9.

[20] Y. Xu, H. Hu, C. Huang, F. Guo, and J. Li, "TAD: A large-scale benchmark for traffic accidents detection from video surveillance," *IEEE Access*, vol. 12, pp. 20235-20247, 2024.