Application And Analysis of Computer Vision Technology in Traffic Safety

Du Li

College of Engineering, The Ohio State University Columbus, Ohio, United States

li.13355@osu.edu

Abstract. In recent years, the integration of intelligent transportation systems and artificial intelligence has been continuously driving profound changes in the field of traffic safety. At present, computer vision technology is increasingly becoming a key tool for enhancing road behavior perception and accident early warning capabilities. Over the past five years, relevant research has achieved significant achievements in areas such as pedestrian and vehicle detection, accident early warning, and damage identification of traffic facilities. A series of algorithms such as YOLO, Faster R-CNN, and CenterNet perform outstandingly in terms of detection accuracy and real-time performance, while multi-object tracking algorithms such as DeepSORT further enhance the stable tracking ability for continuous targets. Despite this, the field still faces many challenges, such as weak system robustness in complex environments, high costs of large-scale labeled data, prominent privacy and security risks, and real-time constraints of edge devices in actual deployment. This paper systematically reviews the advantages and limitations of the existing mainstream methods and offers prospects for future development directions, including multi-sensor fusion, model lightweighting, edge computing advancement, and self-supervised learning applications, with the aim of providing useful references for enhancing the safety and adaptability of intelligent transportation systems.

Keywords: Artificial Intelligence, Intelligent Transportation, Computer Vision.

1. Introduction

Traffic safety has always been a key topic of concern in today's society. According to statistics from the World Health Organization (WHO), approximately 1.3 million people die in road traffic accidents each year, and the number of injured people reaches tens of millions. These accidents not only deal a heavy blow and have a huge impact on the families of the victims, but also bring economic losses to society. In recent years, with the continuous growth of the number of motor vehicles, not only has the traffic pressure been increasing day by day, especially during the morning and evening rush hours when road congestion occurs frequently, but also the safety hazards have been further increased, posing more severe challenges to traffic safety management.

The core objective of computer vision is to imitate human visual observation and comprehension abilities. This technology is based on deep learning algorithms and can automatically analyze the images and videos captured by cameras, achieving automatic target detection, behavior judgment, state analysis, and abnormal situation discovery. In traffic safety applications, computer vision systems can conduct 24-hour continuous monitoring of vehicles, non-motorized vehicles and pedestrians on the road, promptly capturing dangerous behaviors such as speeding, illegal lane changes and pedestrians running red lights, thereby assisting traffic management departments in implementing precise early warnings and rapid intervention, effectively preventing the occurrence of traffic accidents.

In recent years, researchers from various countries have conducted extensive explorations on the application of computer vision in traffic safety and achieved a number of significant advancements. Liu et al. [1] adopted street view video and computer vision technology to achieve precise tracking of pedestrians and vehicles in complex urban scenes, providing an effective solution for the multi-target tracking task of the transportation system. Zhang and Li [2] analyzed the high-risk pedestrian crossing behaviors based on surveillance videos and proposed a behavior recognition framework based on deep learning, providing a practical basis for the prevention of urban traffic accidents. Raj and Mehta [3] constructed a pedestrian perception anti-collision system based on YOLOv9, which significantly improved the detection speed and accuracy. The test results indicated that this system could be applied to real-world scenarios such as autonomous driving. Karim et al. [4] combined YOLOv8 with DeepSORT for real-time traffic accident identification, achieving an accuracy rate of 98.4% on multiple datasets, demonstrating a promising practical application prospect. Furthermore, Elbasha and Abdellatif [5] proposed an intelligent traffic management solution integrating artificial intelligence, which realized the functions of dynamic traffic control and rapid accident response relying on computer vision.

This article conducts a comprehensive review of the application research of computer vision in the field of traffic safety in recent years, and conducts a comprehensive analysis from multiple aspects such as technical basis, practical application and experimental verification. The specific research objectives include: (1) Introducing the currently commonly used computer vision technologies and their typical practices in traffic safety; (2) Compare the effects and application scopes of different methods in various practical environments; (3) Explore the current technical difficulties and practical bottlenecks; (4) Look forward to possible future research directions and put forward development suggestions.

2. Overview of Mainstream Technologies

2.1. YOLO Series Models

The You Only Look Once (YOLO) series is one of the most widely used object detection algorithms in the field of traffic safety in recent years. YOLOv5, YOLOv7, YOLOv8, and the latest YOLOv9 have continuously improved in detection speed and accuracy. Its core idea is to process the entire image in a single pass and directly output the location and category of each object. This approach not only saves time but also enables faster response in real-time monitoring, making it highly suitable for pedestrian and vehicle detection in traffic scenarios. Raj and Mehta [3] indicated that the anti-collision system based on YOLOv9 maintains high detection accuracy under complex lighting and occlusion conditions and is suitable for autonomous vehicles and intelligent transportation camera systems. Figure 1 shows the unified architecture of YOLO for object detection.

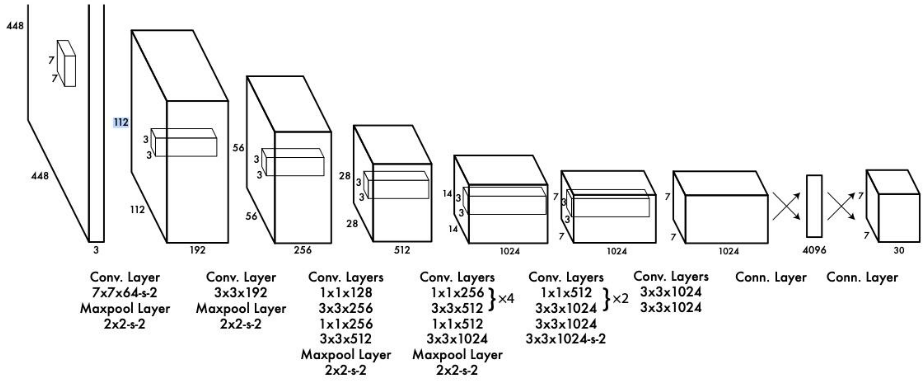

Figure 1. The unified architecture of YOLO for object detection [6].

2.2. Faster R-CNN

Faster R-CNN is an object detection method. It first uses the Region Suggestion Network (RPN) to quickly mark the possible positions of the target in the image, and then hands it over to the convolutional neural network for processing to determine the position of the target more accurately. The greatest advantage of this method is its high precision, making it very suitable for conducting detailed inspections, such as identifying cracks in guardrails, potholes on the road, or whether traffic signs are damaged. However, its drawbacks are also quite obvious: large computational load, slow speed, and insufficient real-time performance. In traffic safety scenarios that require immediate response, such as detecting accidents right away or monitoring vehicles in real time, its performance is not very satisfactory. Figure 2 shows the fast R-CNN model structure of the integrated region suggestion network.

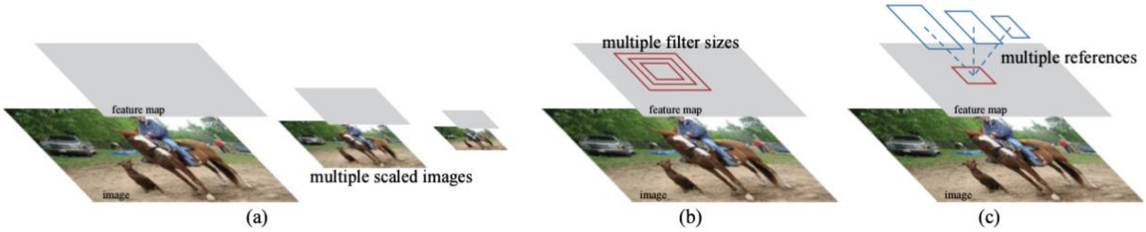

Figure 2. Structure of the Faster R-CNN model integrating Region Proposal Network (RPN) [7].

2.3. CenterNet and Anchor-Free Methods

Anchor-free methods (such as CenterNet) perform detection by predicting the target's center point and its related attributes (figure 3). Since they do not need to generate anchor boxes, the computational load is reduced. For example, in a recent study by Karim et al. [4], the YOLOv8 + DeepSORT system showed excellent performance in dense urban traffic scenes with many objects.

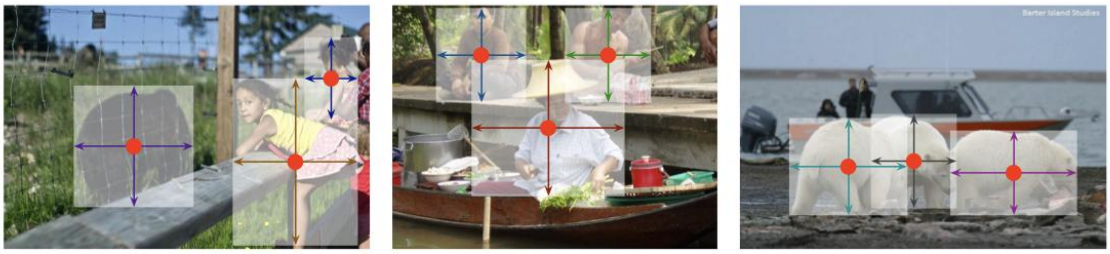

Figure 3. Structure of the CenterNet framework for object detection using center point estimation [8].

2.4. Multi-Target Tracking (MOT)

DeepSORT is a particularly commonly used multi-object tracking algorithm. In practical applications, it is often used in combination with detection models like YOLO. It can simultaneously detect targets and perform stable tracking, with clear results. Its working principle is also very intuitive (figure 4): on the one hand, it uses Kalman filtering to predict where an object will appear next, which is equivalent to "guessing" the future position of the target; On the other hand, it will also extract the appearance features of each target (just like remembering a person's clothing color or body shape), so that even if the target is briefly obscured, when it reappears, the system can still recognize it. It is precisely through this "prediction + human recognition" approach that DeepSORT effectively alleviates common problems such as trace loss and ID switching in tracking.

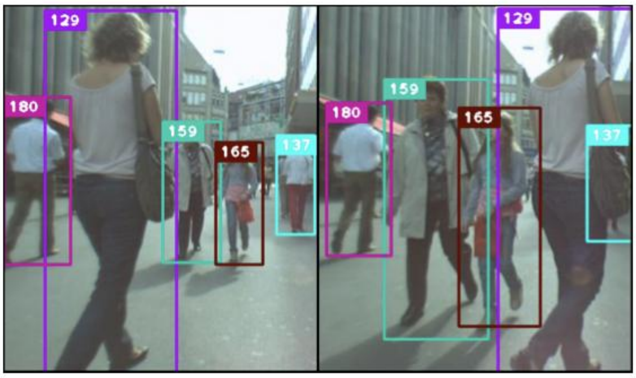

Figure 4. Operational flow of the DeepSORT algorithm with cascade matching and integrated appearance features [9].

2.5. Behavior Recognition and Anomaly Detection

In traffic safety, a behavior recognition system should not only be able to "see" where vehicles and people are and where they are going, but also be able to "understand" how they move - such as whether there are dangerous actions like sudden acceleration, random crossing of the road, or random lane changes. Once such abnormal behavior is detected, the system can immediately issue a warning to help prevent accidents. For instance, Zhang and Li applied the Long Short-Term Memory Network (LSTM) to analyze the behavior of pedestrians crossing the road [2], thereby identifying those dangerous behaviors that are prone to cause accidents - such as sudden acceleration, wandering or illegal crossing, etc.

3. Experimental Results

3.1. Dataset

Research on computer vision in traffic safety highly depends on high-quality datasets, and different tasks require different types of data resources.

UA-DETRAC: Features over 10 hours of video and 140,000 manually labeled vehicle instances. It covers various weather conditions, lighting, and traffic densities, making it suitable for vehicle detection and tracking research.

CityPersons: Focuses on pedestrian detection and is derived from the Cityscapes dataset. It provides pedestrian annotations under various density and occlusion conditions.

KITTI Vision Benchmark Suite: Covers vehicle, pedestrian, bicycle detection, and road semantic segmentation tasks. It is an important benchmark dataset for autonomous driving and traffic scene analysis.

VisDrone: Provides images and videos of traffic scenes from different heights and angles captured by drones, suitable for multi-object detection and tracking tasks.

Domestic customized datasets: For example, the surveillance video data of urban intersections collected by the Ministry of Transport in collaboration with universities, which contains behavioral annotations of various traffic participants, is used for research on risk behavior identification [2].

3.2. Evaluation Indicators

In computer vision tasks related to traffic safety, the following types of indicators are usually adopted to evaluate model performance:

mAP (mean Average Precision): It comprehensively measures the detection accuracy and calculates the average precision under different recall rates and intersection and union ratio thresholds.

Precision/Recall: Precision refers to the proportion of correctly detected targets among all detection results, while recall indicates the proportion of real targets that have been successfully detected.

F1-score: The harmonic mean of precision and recall, used to comprehensively evaluate the accuracy and stability of a model.

MOTA/MOTP: MOTA assesses the overall accuracy of multi-target tracking, while MOTP measures the precision of locating the tracked target.

FPS (Frames Per Second): It reflects the real-time processing capability of the model, that is, the number of image frames that can be processed per second.

3.3. Analysis of Experimental Results

The experimental results of Liu et al. show that when YOLOv8 is combined with DeepSORT [1], it performs excellently on the UA-DETRAC dataset: The comprehensive detection accuracy (mAP) reaches 96.2%, and the multi-target tracking accuracy (MOTA) is 94.8%. Meanwhile, it can process 35 frames of images per second, fully meeting the requirements of real-time monitoring.

Zhang and Li constructed a model integrating Faster R-CNN and LSTM, achieving an accuracy rate of 92.7% in the recognition of high-risk pedestrian crossing behaviors, and being able to issue early warning signals 1 to 2 seconds in advance, leaving valuable time for risk intervention [2].

Karim et al. reported that their YOLOv8+DeepSORT framework achieved a mAP of 98.4% on the VisDrone dataset [4]. This system still maintains stable performance under adverse weather conditions such as night and rain and snow, demonstrating strong environmental adaptability.

4. Challenges and Prospects

4.1. Insufficient Robustness in Complex Environments

At present, the recognition effect of computer vision systems is still not very stable when encountering sudden changes in weather, intense changes in light or targets being blocked, which makes them not reliable enough in real road safety applications.

In the future, it is necessary to further optimize the model, for instance, by adding more diverse training data, adopting cross-scenario adaptation techniques, and integrating multi-environment data for joint training, in order to enhance the system's adaptability and stability under different conditions.

4.2. High Computational Overhead and Limited Real-Time Performance

Although high-accuracy methods such as Faster R-CNN or improved transformer-based YOLO models [10] can achieve strong detection results, they consume substantial computing resources, making them difficult to deploy efficiently in real-time monitoring scenarios.

Future work should promote lightweight model design and the integration of edge computing. This approach enables direct deployment on roadside cameras or edge devices, which reduces computational burden, minimizes transmission latency, and enhances overall system responsiveness.

4.3. High Cost of Data Annotation and Dataset Limitations

The collection and preparation of large-scale labeled data are both time-consuming and labor-intensive, while existing public datasets fail to fully capture the diversity of real traffic scenarios.

Introducing self-supervised learning and transfer learning can reduce dependence on manual annotation. Meanwhile, integrating information from radar, lidar, and infrared imaging can enrich data sources and improve detection reliability. Abdelrahman et al. [11] also explored video-to-text pedestrian monitoring, providing complementary insights into multimodal traffic safety systems.

4.4. Privacy and Data Security Issues

With the growing scale of traffic monitoring data, concerns over personal privacy and data security have become increasingly prominent, posing ethical and regulatory challenges.

Privacy-preserving techniques such as federated learning, differential privacy, and secure multi-party computation need to be actively adopted to enable cross-institutional collaborative training and deployment while safeguarding personal data.

4.5. Insufficient Standardization and Interoperability

The absence of unified data formats and annotation standards leads to poor interoperability among different systems, limiting large-scale integration and application, as evidenced by applications such as helmet violation detection [12], pedestrian detection with YOLOv7 attention mechanisms [13], and deep learning-based accident detection [14].

Establishing unified data standards and evaluation benchmarks, and promoting the development of open-source traffic vision platforms, will strengthen collaboration between academia and industry and accelerate advances in traffic safety research.

5. Conclusion

This paper reviews the main applications of computer vision technology in traffic safety. We introduced common detection algorithms (such as YOLO, Faster R-CNN, CenterNet) and tracking methods (such as DeepSORT), and combined the experimental results to illustrate their advantages in pedestrian and vehicle detection, abnormal behavior recognition, traffic accident early warning, etc. Studies show that these methods can improve the accuracy and real-time performance of detection to a certain extent, providing effective support for traffic safety management. However, some deficiencies were also found in the research. For instance, the recognition effect is not stable enough in bad weather or complex lighting conditions, the computational cost is high, the data annotation cost is high, and there are still hidden dangers in privacy protection. Overall, computer vision technology has become a key support for enhancing traffic safety. It not only achieves remarkable results in accident early warning and traffic management but also provides a solid technical foundation for the development of intelligent transportation systems. In the future, by optimizing the lightweight model structure, promoting multi-sensor fusion perception, exploring self-supervised learning paradigms, and establishing a complete data privacy protection mechanism, the practicality and reliability of the system can be continuously enhanced, further empowering road safety management and leading the transportation system to continuously advance towards intelligence and safety.

References

[1] J. Liu, M. Han, & T. Zhang. Using computer vision and street-level videos for tracking pedestrians and vehicles in complex urban environments. Transportation Research Procedia, 97, 221–230 (2025). https://doi.org/10.1016/j.trpro.2025.02.009

[2] L. Zhang, & S. Li. Vision-based pedestrian’s crossing risky behavior analysis from CCTV cameras. Sensors, 22(9), 3451 (2022). https://doi.org/10.3390/s22093451

[3] D. Raj, & P. Mehta. A robust deep learning-based system for pedestrian-aware collision avoidance (YOLOv9). International Journal of Computer and Communication Engineering, 14(2), 67–74 (2025). https://doi.org/10.1016/j.ijcce.2025.01.005

[4] A. Karim, et al. YOLOv8 + DeepSORT for real-time traffic incident detection. Vehicles, 15(9), 382 (2024). https://www.mdpi.com/2032-6653/15/9/382

[5] A. M. Elbasha, & M. M. Abdellatif. AIoT-based smart traffic management system. arXiv preprint arXiv:2502.02821 (2025).

[6] J. Redmon, S. Divvala, R. Girshick, & A. Farhadi. You Only Look Once: Unified, Real-Time Object Detection. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR) (pp. 779-788) (2016).

[7] S. Ren, K. He, R. Girshick, & J. Sun. Faster R-CNN: Towards Real-Time Object Detection with Region Proposal Networks. IEEE Transactions on Pattern Analysis and Machine Intelligence, 39(6), 1137-1149 (2017).

[8] X. Zhou, D. Wang, & P. Krähenbühl. Objects as Points. arXiv preprint arXiv:1904.07850 (2019).

[9] N. Wojke, A. Bewley, & D. Paulus. Simple Online and Realtime Tracking with a Deep Association Metric. In 2017 IEEE International Conference on Image Processing (ICIP) (pp. 3645-3649) (2017).

[10] A. M. Roy, & J. Bhaduri. A computer vision enabled damage detection model with improved YOLOv5 based on transformer prediction head. arXiv preprint arXiv:2303.04275 (2023).

[11] A. S. Abdelrahman, M. Abdel-Aty, & D. Wang. Video-to-text pedestrian monitoring (VTPM) using computer vision. arXiv preprint arXiv:2408.11649 (2024).

[12] A. Aboah, B. Wang, U. Bagci, & Y. Adu-Gyamfi. Real-time multi-class helmet violation detection using computer vision. arXiv preprint arXiv:2304.08256 (2023).

[13] J. Hu. YOLOv7 with attention mechanisms for pedestrian detection. Sensors, 24(21), 6922. (2024).

[14] H. Ghahremannezhad, H. Shi, & C. Liu. Real-time accident detection in traffic surveillance using deep learning. arXiv preprint arXiv:2208.06461 (2022).