Research on Vehicle Detection Based on Faster R-CNN

Yuhao Zhao

School of Software, Nanchang Hangkong University, Nanchang 330063,China
22206134@stu.nchu.edu.cn

Abstract. While Faster R-CNN is a classical object detection framework widely used in vehicle detection, its performance tends to degrade when deployed in complex and dynamic traffic scenarios. This paper first introduces the fundamental principles of Faster R-CNN, providing a concise theoretical background. It then conducts an in-depth examination of its main limitations, such as limited adaptability to diverse environmental conditions, poor capability in detecting small or distant targets, insufficient real-time processing efficiency, and reduced accuracy when handling partially occluded vehicles. Building on recent advances in related research, several targeted improvement strategies are proposed. These include the adoption of domain adaptation techniques to enhance robustness across environments, network architecture optimization to improve detection precision, lightweight model design to strengthen real-time performance, and advanced strategies specifically tailored for occlusion-aware detection. By systematically analyzing Faster R-CNN's constraints and suggesting feasible solutions, this study aims to contribute both theoretical insights and practical guidance toward improving vehicle detection accuracy and reliability in increasingly complex traffic environments.

Keywords: Faster R-CNN, Vehicle Detection, Environmental Conditions, Real-Time Performance.

1 Introduction

With the rapid development of Intelligent Transportation Systems (ITS) and autonomous driving technology, vehicle detection, as a critical task in the perception layer, directly impacts the accuracy and safety of traffic decision-making. Accurately and efficiently detecting vehicle targets on roads provides crucial support for applications such as autonomous driving obstacle avoidance, traffic flow statistics, and violation monitoring, thereby significantly enhancing traffic efficiency and reducing accident rates, which holds substantial practical importance [1].

Among numerous object detection algorithms, Faster R-CNN has emerged as one of the mainstream frameworks in vehicle detection, owing to its innovative Region Proposal Network (RPN) and end-to-end training paradigm, which achieved an early favorable balance between detection accuracy and efficiency [2]. In recent years, improved models based on Faster R-CNN have demonstrated promising results in scenarios such as urban roads and highways. Examples include enhancing feature

S. Zhang (ed.), *Proceedings of the 2025 International Conference on Electronics, Electrical and Grid Technology (ICEEGT 2025)*, Adva nces in Engineering Research 292,
https://doi.org/10.2991/978-94-6463-986-5_69

extraction capabilities by optimizing the backbone network, or improving multi-scale object detection performance by incorporating Feature Pyramid Networks (FPN) [3] [4]. However, the complexity of real-world traffic environments imposes higher demands on vehicle detection technology: adverse weather conditions (e.g., fog, rain, snow) cause image feature degradation; frequent vehicle occlusions occur in urban congestion scenarios; the feature ambiguity of distant small-target vehicles remains prominent [5-7]. Meanwhile, the stringent real-time performance requirements of autonomous driving (typically requiring 30 fps or higher) also pose challenges to model efficiency [8]. These issues result in significant performance bottlenecks for Faster R-CNN in practical applications. For instance, its detection accuracy decreases by over 20% in complex Indian urban traffic scenarios compared to ideal conditions, while the insufficiency in real-time performance becomes more pronounced in intelligent city environments with composite traffic flows [9] [10].

In this context, this paper focuses on the application limitations of Faster R-CNN in vehicle detection. By reviewing previous relevant research, it systematically analyzes its core deficiencies in environmental adaptability, small object detection, real-time performance, and occlusion handling. Furthermore, it proposes targeted improvement strategies aimed at providing theoretical references and technical insights for optimizing Faster R-CNN-based vehicle detection models, thereby promoting their practical deployment in complex traffic scenarios.

2 Application of the Faster R-CNN Model in Vehicle Detection

Faster R-CNN, proposed by Ren et al. in 2015, is a typical representative of region-based proposal deep learning methods for object detection. The core architecture comprises two components: the Region Proposal Network (RPN) and the Fast R-CNN detector, which share convolutional features. Fig. 1 illustrates the workflow of Faster R-CNN for vehicle detection. During the processing of input images, feature maps are first extracted through a CNN backbone network (e.g., ResNet, VGG). The feature maps contain rich image information, where features at different layers capture low-level edges and textures as well as high-level semantic features. Subsequently, the Region Proposal Network (RPN) generates a set of region proposals that may contain vehicles based on these feature maps. This is achieved by predefined anchors of various scales and aspect ratios at each position on the feature map. Classification (determining whether a target is present) and regression (adjusting position and size) are then performed on these anchors, thereby filtering out high-quality region proposals. Finally, the Fast R-CNN detector employs the RoI pooling layer to convert the feature maps corresponding to region proposals into a fixed size. These are then fed into fully connected layers for category classification and bounding box regression, ultimately yielding the vehicle's category and precise location.

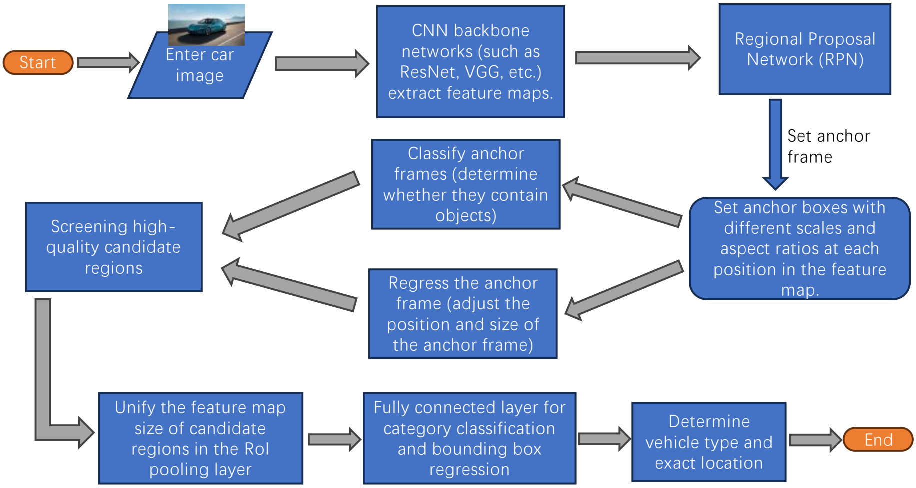

Fig. 1. Faster R-CNN vehicle detection process diagram

The advantage of Faster R-CNN lies in its unification of region proposal generation and object detection within a single deep learning framework, which significantly improves detection speed. Compared with its predecessors (R-CNN and Fast R-CNN), a significant performance improvement is achieved. In scenarios such as traffic surveillance, Faster R-CNN can detect vehicles with relatively high accuracy, providing reliable data for applications like traffic flow statistics. However, the model also has certain limitations. Due to its two-stage detection mechanism-first generating region proposals followed by classification and regression-the computational complexity remains relatively high, resulting in comparatively slower detection speeds. In scenarios with extremely high real-time requirements, such as real-time decision-making in autonomous driving, this approach may fail to meet the need for rapid response. Moreover, the numerous region proposals generated by the RPN contain redundancies, which not only increases computational overhead but also raises the probability of false positives. Furthermore, the model demonstrates suboptimal performance in detecting small-sized vehicles. These smaller targets tend to lose detailed information during the feature extraction process, resulting in higher rates of missed detections or false alarms.

3 Limitations of Faster R-CNN in Vehicle Detection

3.1 Poor Environmental Adaptability

In real-road environments, weather and lighting conditions often undergo rapid and drastic changes. Extreme weather events such as heavy fog, rainstorms, snowstorms, and sandstorms, as well as challenging lighting conditions like strong illumination, low light, and backlighting, can significantly degrade the image quality captured by cameras. Models such as Faster R-CNN are typically trained on data collected in fixed environments; consequently, their detection performance significantly deteriorates when exposed to unseen complex scenarios. For example, under foggy conditions, low visibility leads to blurred vehicle contours, making it difficult for the model to accurately extract features; under varying lighting conditions, significant changes in image brightness and contrast similarly interfere with detection outcomes. Particularly in complex urban environments, such as when street lighting is insufficient at night, the robustness of such models is more likely to be compromised. Hence, how to maintain stable detection performance in variable or even extreme environments remains a critical challenge that requires urgent resolution [11].

3.2 Weak Capability in Small Object Detection

When capturing vehicles from long distances or obtaining traffic footage using drones, vehicles typically appear as small objects in the image. Under these conditions, the detection performance of Faster R-CNN significantly deteriorates.This primarily stems from two reasons: Firstly, during the feature extraction process, certain operations in the model may lose critical details of small objects, thereby weakening its representational capacity for small vehicle features; secondly, small targets occupy very few pixels in the image, making them easily compromised by cluttered background information, which results in the model's difficulty in distinguishing vehicles from the background.This issue becomes particularly pronounced in traffic scenarios with dense vehicle flow and small-scale vehicles (e.g., Indian roads); in aerial traffic videos and multi-scale vehicle detection tasks involving both large and small vehicles, the inadequate capability for small object detection remains a persistent weakness of Faster R-CNN.

3.3 Insufficient Real-Time Performance

In application scenarios such as autonomous driving, vehicle detection must possess high real-time performance to provide effective information to the vehicle decision-making system in a timely manner. However, the network architecture of Faster R-CNN is relatively complex, comprising numerous convolutional layers, pooling layers, and a Region Proposal Network (RPN). This structure requires substantial computational resources when processing a single image, thereby limiting the potential for improving detection speeds. When dealing with high-resolution images or processing multiple images at once, this issue becomes especially evident and leads to a notable decline in detection efficiency. In complex traffic environments—such as those in smart cities—the practical usefulness of Faster R-CNN is limited due to its inadequate processing speed. Compared to more lightweight and faster object detection algorithms, Faster R-CNN suffers from a considerable disadvantage in terms of detection speed.

3.4 Poor Performance in Detecting Occluded Vehicles

In traffic congestion scenarios, vehicles often occlude each other, which can significantly degrade the detection performance of Faster R-CNN. When a vehicle is partially obscured, the model has difficulty capturing its complete feature representation. As a result, the Region Proposal Network (RPN) often struggles to produce high-quality proposals that accurately localize the occluded vehicle. Even if candidate regions are proposed, misjudgments are likely to occur during subsequent category classification and bounding box regression, resulting in occluded vehicles being either overlooked or misidentified. In complex traffic environments such as morning and evening peak hours, frequent occlusion occurrences significantly reduce the reliability of Faster R-CNN's detection results. The vehicle occlusion problem has thus become a critical factor constraining the model's practical performance in real-world applications.

4 Improvement Suggestions

To systematically enhance the vehicle detection performance of Faster R-CNN in complex traffic environments, this paper proposes a series of targeted improvement strategies across four key aspects: environmental adaptability, small object detection, real-time processing, and occluded vehicle recognition. These strategies cover not only algorithm-level adjustments but also technical innovations such as model architecture design and multi-modal information fusion. The following table (Table 1) provides a systematic summary of the specific measures, core objectives, and applicable scenarios for each improvement direction, offering a clear optimization pathway for subsequent engineering practices and theoretical research.

Table 1. Targeted Improvement Suggestions for Faster R-CNN Vehicle Detection Performance

Improvement Direction	Specific Measures	Core Objective	Applicable Scenarios
Introducing Domain Adaptation	1. Train a base model on source domain data (e.g., clear weather conditions); 2. Generate pseudo-labels for target domain data (e.g., fog, rain, snow);	Adapt the model to feature distribution shifts across environments, maintaining detection stability in varying conditions.	Urban dynamic traffic (frequent weather changes); Highway scenarios under diverse weather conditions (fog, rain, glare).
Optimizing Network Architecture (Improving Small Object Detection)	1. Employ MobileNetV3 or optimized ResNet-50 as the backbone network to reduce computational cost while preserving small object features; 2. Integrate FPN to merge multi-level features;	Enhance feature extraction and discrimination capability for small objects, reducing detail loss and background interference.	UAV aerial photography for small vehicle detection; Long-distance vehicle identification; Multi-scale vehicle detection

			tasks.
Real-Time Processing Strategies	1.Model Compression and Acceleration: Apply pruning and quantization to reduce redundant parameters; leverage GPU or FPGA platforms with acceleration libraries for hardware speed-up; employ lightweight convolutional layers to minimize redundant computations.	Reduce computational cost and improve processing speed for single or multiple images.	Smart city high-density traffic (large-scale image data); Real-time decision-making in autonomous driving.
	2. Multi-scale image optimization: Adaptive scale selection (coarse estimation of target size followed by dynamic selection of detection scale).		
Enhancing Occluded Vehicle Detection	1. Leverage contextual information: Use Graph Convolutional Networks (GCN) to model spatial relationships between targets, lane markings, and adjacent vehicles to infer the location/shape of occluded vehicles; 2. Multi-view fusion: Fuse images from multiple onboard cameras, or combine LiDAR and millimeter-wave radar data to supplement vehicle distance and shape information.	Compensate for missing features of occluded vehicles, reducing missed detections and false alarms.	Rush-hour congested roads; Complex intersections with vehicle interactions.

To more clearly present the practical limitations and targeted optimization strategies of derivative Faster R-CNN models across different improvement directions, Table 2 summarizes the core issues and specific enhancement approaches of five representative improved Faster R-CNN models.

The table focuses on the unique shortcomings exposed by each model in vehicle detection scenarios, such as the small-object feature fusion problem in FPN-enhanced models and the real-time performance trade-offs in attention mechanism improved models. For each category of problems, actionable technical optimization pathways are proposed, establishing a clear problem-solution correspondence that provides intuitive reference for subsequent model improvement practices.

Table 2. Core Limitations and Optimization Suggestions for Typical Improved Faster R-CNN Models

Model Name	Main Limitations	Improvement Suggestions
FPN-Enhanced Faster R-CNN	Insufficient fusion of small object features with significant noise interference in low-level features	-Introduce a cross-layer attention mechanism to enhance the weighting of small object features -Design a dynamic anchor generation strategy adapted to the scale of small targets
Attention-Enhanc-ed Faster R-CNN	High computational complexity of attention mechanisms leads to degraded real-time performance	-Employ lightweight attention modules (e.g., simplified Squeeze-Excitation) -Deploy attention mechanisms only in critical feature layers
Transformer Hybrid Faster R-CNN	The Transformer Module Exhibits Slow Training Convergence and Sensitivity to Small Samples	-Employ lightweight attention modules (e.g., simplified Squeeze-Excitation) -Deploy attention mechanisms only in critical feature layers
Domain-Adaptive Faster R-CNN	The domain adaptation effect is limited under extreme weather conditions (e.g., heavy rain, heavy snow)	- Integration of a Meteorological Feature Embedding Module to Enhance Environmental Robustness - Adoption of Meta-Learning for Dynamic Adjustment of Domain Adaptation Strategies
Lightweight Faster R-CNN	The deep feature extraction capability is weakened, and the detection accuracy in complex scenarios decreases	- Introduction of Residual Connections to Strengthen Deep Feature Propagation - Design of a Multi-Branch Lightweight Convolutional Structure to Balance Speed and Accuracy

5 Conclusion

This paper investigates the application of Faster R-CNN in vehicle detection. The research follows a structured approach of "analyzing principles — identifying problems — proposing improvements". First, it clarifies the two-stage working mode of Faster R-CNN, which consists of a Region Proposal Network (RPN) and a Fast Region-Based Convolutional Neural Network (Fast R-CNN). Subsequently, based on practical traffic scenarios, four main limitations are summarized: 1) suboptimal detection performance in complex environments; 2) insufficient capability in detecting small target vehicles due to limitations in the feature extraction mechanism; 3) high computational demands that hinder real-time detection requirements; and 4) frequent missed detections or false alarms when vehicles are occluded. To address these issues, corresponding improvement methods are proposed: domain adaptation technology is adopted to enhance cross-environment detection robustness; the backbone network is optimized and combined with the Feature Pyramid Network (FPN) to improve small target detection capability; model compression and hardware acceleration are employed to increase operational speed.

The value of this study lies in three dimensions: Academically, it advances research on problem identification and solution methods for Faster R-CNN in traffic scenarios; technically, the proposed improvement methods can be directly applied to practical projects, providing valuable references for engineering practice; in practical applications, it enhances the accuracy and reliability of vehicle detection in autonomous driving and traffic surveillance, thereby significantly contributing to improved traffic efficiency, enhanced travel safety, and promoting the implementation of intelligent transportation technologies.

References

1. Chen, X., Han, Y.: Multi-task CNN Model for Action Detection. In: 2018 IEEE Visual Communications and Image Processing (VCIP), pp. 1–4. IEEE, Taichung, Taiwan (2018).
2. Lin, T.Y., Dollár, P., Girshick, R., He, K., Hariharan, B., Belongie, S.: Feature Pyramid Networks for Object Detection. In: 2017 IEEE Conference on Computer Vision and Pattern Recognition (CVPR), pp. 936–944. IEEE, Honolulu, HI, USA (2017).
3. Singh, G., Sharma, M., Pratap, R.: Vehicle Detection and Classification in Indian Traffic. In: 2019 IEEE International Conference on Systems, Man, and Cybernetics (SMC), pp. 1–13. IEEE, Piscataway, USA (2019).
4. Chen, S., et al.: Fused Pyramid Attention Network for Single Image Super-Resolution. IET Image Processing 17(6), 1681–1693 (2023).
5. Slimani, I., Zaarane, A., Atouf, I.: Traffic monitoring system for vehicle detection in day and night conditions. Transport and Telecommunication 24(3), 256–265 (2023).
6. Alam, M.K., Ahmed, A., Salih, R., et al.: Faster RCNN based robust vehicle detection algorithm for identifying and classifying vehicles. Journal of Real-Time Image Processing 20(5), 93 (2023).
7. Sermsar, F., Moghadam, M.T., Koca, M.: Performance Comparison of Faster R-CNN and SSD in Vehicle Detection from Aerial Traffic Videos. In: 2024 Innovations in Intelligent Systems and Applications Conference (ASYU), pp. 1–6. IEEE, Ankara, Turkiye (2024).
8. Yin, G., Yu, M., Wang, M., et al.: Research on highway vehicle detection based on faster R-CNN and domain adaptation. Applied Intelligence 52(4), 3483–3498 (2022).
9. Kapoor, N., Aggarwal, P., Sharma, A.: Smart-VDCTC: Smart vehicular detection classification and traffic control for Indian urban traffic. Multimedia Tools and Applications, 1–25 (2025).
10. Musa, A.A., Malami, S.I., Alanazi, F., et al.: Sustainable traffic management for smart cities using internet-of-things-oriented intelligent transportation systems (ITS): challenges and recommendations. Sustainability 15(13), 9859 (2023).
11. Xu, Y., Chu, K., Zhang, J.: Nighttime vehicle detection algorithm based on improved faster-rcnn. IEEE Access 12, 19299–19306 (2023).