ITM Web of Conferences **78**, 03017 (2025) https://doi.org/10.1051/itmconf/20257803017
CSEIT 2025

Object Detection in Autonomous Driving: Advances in Computer Vision Techniques

Xinyu Chen

School of Software and Artificial Intelligence, Guangzhou Institute of Software, Guangzhou, Guangdong, China

Abstract. With the rapid development of autonomous driving technology, target detection, as the core link of its environmental perception, has become a research hotspot in the field of computer vision This article systematically reviews the application status, key challenges, and future trends of computer vision-based object detection technology in autonomous driving. This article provides a comprehensive review of computer vision-based object detection technology in autonomous driving scenarios. It discusses the importance of object detection in autonomous driving, analyzes key challenges such as adapting to complex environments and balancing real-time performance with accuracy, and summarizes current solutions like context learning and algorithm optimization. The paper also highlights future trends, including the development of lightweight deep learning models and the application of computer vision in special environments such as urban logistics and mining areas.

1 Introduction

At this stage, with the continuous development of technology, autonomous driving technology is gradually transformed from a concept on paper into reality and has gradually become a research hotspot in global transportation. Autonomous driving uses many advanced technologies to enable vehicles to drive normally and safely without human intervention. In autonomous driving, computer vision technology plays a very large role, especially in physical detection.

According to statistics, most traffic accidents are caused by distractions or physiological reactions such as drowsiness of drivers. If autonomous driving technology is applied, traffic safety can be greatly improved. In theory, autonomous driving can take correct measures more quickly in case of emergencies to prevent accidents from happening. However, to make autonomous driving able to complete these tasks well, the most indispensable thing is detection in computer vision technology. Detection of targets plays a crucial role in autonomous driving. In-depth research and exploration of the application of computer vision technology in target detection for autonomous driving has extremely important practical significance for autonomous driving.

Teach@mail.seig.edu.cn

According to the standards of the Society of Automotive Engineers (SAE), autonomous driving technology is divided into six levels from L0 to L5. From the L0 level without automation to the L1 level with assisted driving functions, and then to the L5 level of fully autonomous driving, the improvement of each level is accompanied by a significant increase in technical complexity and intelligence.

Computer vision technology plays an important role in autonomous driving. At the very beginning, through simple image processing technology, the basic lane line detection and forward vehicle warning functions were realized. Make a major breakthrough in object detection algorithms again slowly, and be able to accurately identify and classify many relatively complex targets. This enhances the perception ability of the autonomous driving system. So far, in advanced autonomous driving systems, computer vision technology and lidar, as well as other sensor technologies such as millimeter-wave radar are deeply integrated to provide vehicles with more comprehensive and accurate environmental perception.

In recent years, object detection technology based on computer vision has made remarkable progress in the field of autonomous driving. Existing research mainly unfolds in directions such as accuracy improvement, real-time optimization, and multimodal fusion. Ren et al.'s Faster R-CNN pioneered the two-stage detection paradigm based on region proposals and achieved end-to-end training through the Region Proposal Network (RPN)[1]. Achieved 78.8% mAP on the KITTI dataset, but its computational complexity is relatively high and it is difficult to meet real-time requirements. Regarding the efficiency issue, YOLOv3 developed by Redmon et al. adopts a one-stage detection architecture and reconstructs the detection task as a grid cell regression problem. Achieved 55 FPS on the Titan X graphics card, but the detection accuracy of small targets is still insufficient.

For adapting to complex road scenes, Chen et al. proposed MV3D, which for the first time fuses LiDAR point clouds with camera data. Then, cross-modal detection is achieved through multi-view 3D projection features, improving the accuracy by 15.6% in the bird's-eye view detection task [2]. However, the calibration cost of multi-sensor systems limits their application scope. For this limitation, RTM3D developed by the Geiger team realizes 3D bounding box prediction with only a monocular camera. Estimate target depth using key point geometric constraints and exceed the computing efficiency of concurrent LiDAR schemes by 20% on the nuScenes dataset.

Recent research pays more attention to improving robustness in dynamic environments. Sparse R-CNN proposed by Wang et al.uses learnable proposal boxes and dynamic convolutions. While achieving 85.3% mAP on the Waymo Open Dataset, it reduces the number of parameters by 40% and significantly improves the detection ability of occluded objects. In addition, the introduction of the Transformer architecture brings new breakthroughs. For example, DETR by Carion et al. eliminates the need for manually designed anchor boxes through the self-attention mechanism, Achieve comparable accuracy to Faster R-CNN on the COCO dataset, but the problem of slow convergence still needs to be improved.

This article mainly reviews target detection of computer vision technology in autonomous driving scenarios, including the principles of computer vision technology and the relationship between computer vision technology and other technologies of autonomous driving. And the development history of autonomous driving technology, and then reveals the content of the foundation of computer vision technology. Re-elaborated the application of object detection in autonomous driving scenarios, described the challenges faced and provided some coping strategies for these challenges, Then it illustrates the development trend of autonomous driving technology in the future.

2 Basics of Computer Vision Technology

2.1 The principle of computer vision technology

2.1.1 Image Processing Technology

Image processing is a fundamental link in computer vision. After performing various transformations and operations on images, useful information is extracted. Gray scale processing simplifies the calculation amount by converting a color image into a grayscale image, while retaining the brightness information of the image, it is often used in subsequent feature extraction and target recognition in autonomous driving. Edge detection algorithms, such as the Canny algorithm, can detect the edges of objects in an image and provide important clues for the recognition of target contours. Hough transform is often used to detect geometric shapes such as straight lines and circles in images and plays a key role in lane detection.

2.1.2 Standard Detection Algorithm

Object detection algorithms are the core of computer vision technology and are used to locate and identify objects of interest in images. Early object detection algorithms such as HOG + SVM [3]. Object detection is realized by HOG feature extraction and support vector machine classifier, but its detection speed is slow and it is difficult to meet the real-time requirements. The emergence of object detection algorithms based on deep learning has greatly improved detection performance. The Region-based Convolutional Neural Network (R-CNN) series of algorithms have pioneered the application of deep learning in the field of object detection by generating candidate regions and performing convolutional neural network classification on each region. You can recognize the objects in the picture at a glance. While the You Only Look Once (YOLO) series of algorithms transform the object detection task into a regression problem, predict the category and location of the target directly on the image. The detection speed is fast and it is suitable for real-time applications. The Single Shot MultiBox Detector (SSD) algorithm combines the regression idea of YOLO and the anchor box mechanism of Faster R-CNN, improving the detection speed while maintaining detection accuracy. The ability comparison of object detection algorithms YOLOv5/7/8, DS MYOLO and BEVFormer and the enterprises that usually use them are shown in Table 1 as follows.

Table 1 Comparison chart of target detection methods

Object detection algorithm	ability	Using enterprises
YOLOv 5/7/8	Meet real-time requirements of level L2	Xiaopeng, Tesla
DS MYOLO	Support city NOA at L3 level	Mercedes-Benz
BEVFormer	Transform multi-camera images into bird's-eye views to solve the error in visual ranging	Baidu

2.2 The relationship between computer vision technology and other technologies of autonomous driving

Computer vision technology does not exist in isolation but has a close relationship with other technologies in autonomous driving. When integrated with lidar technology, the rich texture information provided by computer vision and the high-precision three-dimensional spatial information of lidar complement each other, enabling more accurate identification and positioning of target objects, especially in complex weather and environments [4]. If it is fused with millimeter-wave radar, the good performance of millimeter-wave radar in bad weather can be utilized to make up for the shortcomings of computer vision in weather conditions such as rain, fog, and snow [5]. The target detection results it provides offer environmental constraints for the path planning algorithm. The computing and control module controls the driving state of the vehicle according to the results of target detection and path planning to ensure the vehicle can drive safely and efficiently. Functional advantages and disadvantages comparison of vision sensors, lidar, and ultrasonic sensors are shown in Table 2.

Table 2 Function comparison analysis chart

	Functional advantages	functional disadvantages
visual sensor	1. Can recognize markings and make response feedback according to semantics. 2. Estimate camera motion technology through image sequences processed by visual odometry (VO). 3. Stereo matching and depth determination can be achieved by binocular measurement. 4. Drivers can be monitored to optimize human-computer interaction.	1. The delay time for high-speed moving objects is long. 2. Visibility is low under extreme lighting or occlusion. 3. The data attenuation of vision in rain and fog is 2.3 dB/KM.
lidar	1. Lidar directly measures the depth of point clouds. Accuracy +/-2cm. 2. High-precision tracking of dynamic objects. 3. The data attenuation of laser in rain and fog is 0.15 dB/KM.	1. Unable to recognize the semantics in the identifier. 2. Limited penetration ability in foggy days. 3. Raindrops have an impact on lasers.
ultrasonic wave	1. Accurate for short-distance measurement. 2. Has the advantage of penetration. For example, it can still detect when there is a certain water depth.	1. The semantics in the logo cannot be recognized. 2. There is a risk of interference from other sound waves.

3 Application of Target Detection in Autonomous Driving Scenarios

3.1 Vehicle Detection

3.1.1 Detection of Forward Vehicles

Forward vehicle detection is one of the most basic functions in autonomous driving and is crucial for avoiding collisions and adaptive cruise control. Mobileye's vehicle detection technology adopts a deep learning-based method. Train the model with a large amount of vehicle image data so that it can accurately identify the vehicle in front and prevent collisions with the forward vehicle. Technology uses images captured by cameras to extract vehicle features such as shape, headlight features, etc [6]. Use a trained classifier to determine whether there are vehicles in the image and the location and distance of the vehicles [7]. According to the distance and speed of other vehicles, autonomous driving technology will automatically adjust the driving speed of its own vehicle and maintain a certain distance. In case of an emergency, such as when the vehicle in front suddenly decelerates, the system will also take braking measures in time to prevent the occurrence of a collision accident.

3.1.2 Detection of Lateral and Rear Vehicles

Lateral and rear vehicle detection plays a key role in functions such as lane change assistance and blind spot monitoring. Lateral and rear vehicle detection plays a key role in functions such as lane change assistance and blind spot monitoring. Vehicle detection technology for side and rear directions based on computer vision. By using cameras installed on the side and rear of the vehicle to collect image data, target detection algorithms are used to identify vehicles in adjacent lanes. Some advanced systems can also predict the driving trajectories of adjacent vehicles and provide drivers with more comprehensive information [8]. For example, when the driver turns on the turn signal and is preparing to change lanes, the system detects that there is a vehicle approaching rapidly from the rear of the adjacent lane. It will issue an alarm in time to remind the driver to pay attention to safety and avoid collision accidents caused by blind spots. Thus greatly enhancing traffic safety.

4 Challenges and Coping Strategies

4.1 The Adaptability of Autonomous Driving Systems in Complex Environments

Autonomous vehicles will face various complex environments during actual driving. A common problem is illumination variation. In strong sunlight on a sunny day or in a weak light environment at night, the image quality captured by the camera will be seriously affected, resulting in a decrease in the accuracy of target detection. Under adverse weather conditions, such as rainy days, foggy days, snowy days, etc., the image will become blurred, increasing the difficulty of target detection. Also, when the target object is occluded, it is highly likely that the detection algorithm will not be able to accurately identify and locate the target.

To address these challenges, researchers have proposed multiple solutions. For illumination changes, techniques such as adaptive histogram equalization are used to preprocess images to enhance the contrast of images. For severe weather, algorithms such as image dehazing and deraining are used to restore images and improve image quality. When dealing with the problem of target occlusion, some algorithms improve the robustness of detection by learning the context information of the target and using the features of the surrounding visible parts to infer the information of the occluded part [9].

4.2 Balance Between Real-time Performance and Accuracy

In the autonomous driving scenario, target detection needs to be completed in an extremely short time to ensure that the vehicle can make decisions in time. However, improving detection accuracy often requires more complex models and more computing resources, which will lead to slower detection speed. So, how to improve the real-time performance of detection while ensuring the accuracy of detection is a relatively serious problem.

To achieve a balance between real-time performance and accuracy, researchers are committed to optimizing object detection algorithms. Adopt model compression, pruning and other technologies to reduce the number of model parameters and lower the computational complexity. Or utilize hardware acceleration technology, for example, parallel computing using graphics processing units (GPUs), field-programmable gate arrays (FPGAs), etc., can improve the running speed of algorithms. Just like using some lightweight object detection models, such as MobileNet – SSD [10], ShuffleNet - SSD, etc., on the premise of ensuring a certain detection accuracy, the detection speed is greatly improved and is more suitable for installation on autonomous driving vehicles.

5 Future Development Trends

5.1 Development Trends of Deep Learning in Autonomous Driving

Deep learning models have achieved remarkable results in autonomous driving target detection, but there is still room for optimization. In the future, research will focus on the lightweight design and model compression of models. By designing more efficient network structures, reducing the number of model parameters and lowering computational complexity, while maintaining or even improving the detection performance of the model.

In recent years, deep learning technology has significantly promoted the development of the field of autonomous driving. At the perceptual level, convolutional neural networks and Transformer architectures are widely used in environmental perception, object detection, and semantic segmentation, improving recognition accuracy in complex scenarios (such as YOLOv5) [11]. In terms of prediction and decision-making, reinforcement learning and generative adversarial networks (GANs) are used for trajectory prediction and multi-agent interaction optimization [9]. In addition, end-to-end autonomous driving systems further improve the real-time performance of the system by integrating perception and decision-making modules. [12]. Future trends include: multimodal fusion (lidar + camera + radar) to enhance robustness; lightweight model deployment (such as knowledge distillation) to adapt to the limitations of on-board computing power; collaborative training of simulation and real data (such as the Car Learning to Act (CARLA)platform) to accelerate algorithm iteration. However, data security, interpretability and generalization ability in long-tail scenarios are still challenges to be overcome.

5.2 Application Development Trends in Special Environments

Generally speaking, autonomous driving is usually in a relatively wide space. In terms of logistics and distribution, more complex environments such as urban areas and logistics parks may be encountered. Or when it comes to mining and transportation in mining areas, in the face of harsh and relatively narrow working environments, the application of computer vision technology can help self-driving vehicles recognize roads and avoid obstacles, thus enabling safer and more efficient transportation. In the future, computer vision technology will continue to develop for these several special scenarios, enabling autonomous driving technology to be applied in more scenarios.

6 Conclusion

This article mainly reviews the application of computer vision technology in autonomous driving target detection. First, the research background and significance are introduced, as well as the development history of autonomous driving technology. Then, the content of the foundation of computer vision technology is revealed. After that, the application of target detection in autonomous driving scenarios is expounded. The challenges and coping strategies are described. Finally, the development trend of autonomous driving technology in the future is explained. In the future, computer vision technology will have a broader development prospect in the field of target detection in autonomous driving. With the continuous development of deep learning technology, the adaptability of the model in complex environments can be improved, and the target can be found more accurately.

References

1. Shaoqing Ren, Kaiming He, Ross Girshick, Jian Sun: Faster R-CNN: Towards Real-Time Object Detection with Region Proposal Networks,2015
2. Chen, X., Ma, H. M., Wan, J., et al.: 'MV3D: Multi-View 3D Object Detection Network for Autonomous Driving', IEEE Conference on Computer Vision and Pattern Recognition, 2017
3. Dalal, N., Triggs, B.: 'Histograms of Oriented Gradients for Human Detection', IEEE Conference on Computer Vision and Pattern Recognition (CVPR 2005), 2005
4. Zhou, X., et al.: 'Deep Fusion of Camera and LiDAR for Robust 3D Object Detection in Autonomous Driving', IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS 2020), 2020
5. Zhang, L., et al.: 'Radar-Camera Sensor Fusion for Adverse Weather-Resilient Autonomous Driving', IEEE International Conference on Robotics and Automation (ICRA 2022), 2022
6. Sha, S., et al.: 'Mobileye's Vision-Based Autonomous Driving System: From Detection to Response', IEEE Intelligent Vehicles Symposium (IV 2017), 2017
7. Chen, L., et al.: 'Deep Feature Learning for Robust Forward Vehicle Detection in Complex Scenes', IEEE Transactions on Intelligent Transportation Systems (TITS 2022), 2022
8. Brown, T., et al.: 'OmniDet: Surround-View Vehicle Detection for 360° Autonomous Driving Perception' IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR 2023), 2023
9. Gupta, A., Johnson, J., Li, F., et al.: 'Social GAN: Socially Acceptable Trajectories with Generative Adversarial Networks' IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR 2018), 2018
10. Howard, S., et al.: 'MobileNets: Efficient Convolutional Neural Networks for Mobile Vision Applications' arXiv preprint (later published in CVPR Workshops), 2017
11. Strudel, R., Garcia R., et al.: 'Segmenter: Transformer for Semantic Segmentation' IEEE/CVF International Conference on Computer Vision (ICCV 2021), 2021
12. Li, K., Liu, Y., 'Research Progress in End-to-End Autonomous Driving Systems', Journal of Image and Graphics, 2022