Advances in Object Detection for Intelligent Driving

Xuntong Hong[a]
Xiamen No.1 High School of Fujian, Xiamen, Fujian, China

Keywords: Intelligent Driving, Target Detection, You Only Look Once.

Abstract: Intelligent driving is at the forefront of modern transportation technology, with target detection playing a pivotal role in the safe and efficient operation of autonomous driving systems. This paper reviews the latest advancements in target detection for intelligent driving, focusing on the challenges posed by external factors such as weather conditions, illumination variations, and traffic density, as well as internal factors related to sensor technology. The paper highlights the importance of multi-sensor fusion, combining data from cameras, LiDAR, and millimeter-wave radar, to enhance detection accuracy and robustness. It also provides an in-depth analysis of popular target detection methods, particularly the You Only Look Once (YOLO) family of models, which have demonstrated significant improvements in real-time detection and accuracy. Other methods, such as Faster R-CNN, Single Shot Multibox Detector (SSD), and RetinaNet, are also discussed, emphasizing their strengths and limitations in intelligent driving applications. Despite progress, challenges remain, including robustness in complex environments, small object detection, and balancing accuracy with real-time performance. Future directions include multimodal data fusion, unsupervised learning, and hardware acceleration to further improve target detection capabilities. The advancements in sensor technology, deep learning, and computational power will drive the continued evolution of intelligent driving systems.

1 INTRODUCTION

Intelligent driving, as the frontier of current transportation technology, is gradually moving towards commercial applications and is expected to play a crucial role in future transportation systems. The realization of the core of an autonomous driving system relies on the synergistic operation of a number of technologies, among which the target detection technology plays a crucial role as one of the basic perception tasks. The main task of target detection is to recognize and locate various objects in the scene, and the ability to accurately complete target detection is a prerequisite for the safe and efficient execution of decision-making by the autonomous driving system.

With the rapid development of sensor technology, modern automatic driving systems can obtain richer and more accurate environmental information through a variety of sensors such as LiDAR, cameras, millimeter wave radar and so on. Through these sensors, the automatic driving system can perceive the surrounding environment in real time and respond accordingly to different target objects. However, target detection technology still faces many challenges, especially in complex and dynamic road environments. How to improve the accuracy, robustness and real-time performance of target detection has become an urgent challenge in autonomous driving research.

This paper reviews the latest research results in the field of target detection for intelligent driving, focusing on the various factors affecting the effectiveness of target detection, such as the quality of sensor data, the efficiency and accuracy of algorithms, and the complexity of the environment. Meanwhile, this paper will also analyze the current mainstream target detection methods in depth and discuss the advantages and limitations of these methods in practical applications. Finally, the article will look forward to the future development trend of intelligent driving target detection technology, including how to combine advanced technologies such as deep learning and reinforcement learning, as well as how to solve the problems of real-time and robustness in complex environments, so as to provide

[a] https://orcid.org/0009-0005-1231-356X

Hong, X.
Advances in Object Detection for Intelligent Driving.
DOI: 10.5220/0013679000004670

In *Proceedings of the 2nd International Conference on Data Science and Engineering (ICDSE 2025)*, pages 77-81
ISBN: 978-989-758-765-8

technical support for the comprehensive promotion and application of intelligent driving systems.

2 INFLUENCING FACTORS OF INTELLIGENT DRIVING TARGET DETECTION

The effect of target detection is affected by a variety of factors, both external and internal. This paper analyzes the main factors affecting target detection from both external and internal dimensions and discusses how to deal with these challenges.

2.1 External Factors

Weather conditions: different weather conditions (e.g., rain, snow, haze, etc.) have a significant impact on the effectiveness of target detection. Bad weather reduces the detection accuracy of sensors, e.g., cameras may suffer from blurred images and low contrast in rain and snow, and LiDAR may lose accuracy in hazy weather. To cope with this problem, current research directions include image denoising techniques based on deep learning, and sensor fusion techniques to enhance the sensing ability in adverse weather conditions.

Zhang et al. proposed an image denoising method based on convolutional neural networks aimed at addressing the effects of rain and snow on camera images. Studies have shown that deep learning techniques are significantly effective in enhancing image quality, thereby improving the accuracy of target detection (Zhang et al., 2019).

Kim et al. proposed a sensor fusion method combining LiDAR and a camera to address the effect of hazy weather on LiDAR accuracy. The method successfully improves target detection performance in harsh environments through multimodal data fusion (Kim & Lee, 2020).

Illumination variations: day and night variations as well as strong illumination (e.g., backlight, flash, etc.) pose challenges to target detection algorithms. Strong lighting conditions may lead to overexposure of the camera image, which affects the detection accuracy; while low lighting at night may make it difficult to recognize objects. In this regard, existing techniques mainly rely on image enhancement, low-light image processing, and the assistance of infrared sensors to enhance detection in low-light environments. Li et al. proposed a data fusion method based on infrared sensors and visible cameras, which can achieve efficient target detection in low-light environments, especially in nighttime driving conditions (Li et al., 2019).

Traffic Density and Environmental Complexity: Intelligent driving systems usually need to work in highly dynamic and complex traffic environments. With the increase in traffic density, especially on urban roads or highways, the gap between targets is small and occlusion may occur, which can greatly affect the accuracy of target detection.

2.2 Internal Factors

Target detection in autonomous driving systems relies on the collaborative work of multiple sensors, and different types of sensors have their own unique operating principles, accuracy, and adaptability, and these characteristics directly affect the effectiveness of target detection. Cameras are one of the most common sensors in autonomous driving, providing rich color, texture, and detail information, and are particularly good at identifying traffic signs, pedestrians, and other vehicle appearance features. However, cameras are more sensitive to conditions such as ambient lighting and weather changes, and image quality may suffer from blurring or low contrast. LiDAR, on the other hand, measures the distance of an object by emitting a laser beam and receiving an echo to construct highly accurate three-dimensional spatial information. It is not affected by lighting conditions and can work stably in inclement weather such as rain and fog, and is suitable for accurately locating the distance and shape of target objects. Millimeter-wave radar, on the other hand, detects the distance, speed and direction of objects by transmitting electromagnetic waves, has strong anti-interference ability, can work stably under various weather conditions, and is especially suitable for target detection in high-speed driving. Although it has a lower resolution and cannot provide as much detailed information as a camera, it has unique advantages in real-time and speed measurement. Since each sensor has its advantages and limitations, autonomous driving systems usually use multi-sensor fusion technology to synthesize data from LIDAR, cameras, and millimeter-wave radar to improve the accuracy, robustness, and adaptability of target detection. Through sensor fusion, the system can make up for the shortcomings of a single sensor and provide more comprehensive and accurate target detection results, thus enhancing the safety and reliability of the automatic driving system in complex environments.

3 TARGET DETECTION METHODS

In the field of target detection, especially in intelligent driving, target detection methods based on deep learning have made significant progress in recent years. The following discussion will focus on the You Only Look Once (YOLO) family of models and some other classical target detection methods.

3.1 Methods based on the YOLO Family of Models

Evolution and basic principle of YOLO: YOLO is a real-time target detection algorithm based on deep learning, and its main advantage is that it can simultaneously classify and localize targets through one forward propagation.YOLO achieves target detection by dividing the image into grids and predicting bounding boxes and category probabilities for each grid.YOLO series models have been evolving from YOLOv1 to YOLOv7, and their performance has been gradually improved. have been evolving and their performance has been gradually improved, especially in the balance between detection speed and accuracy, significant progress has been made.

Redmon proposed the YOlO model, which is a real-time target detection model that transforms target detection from a multi-stage processing method based on region candidates to a single regression problem, greatly improving the performance of target detection. The YOlO model avoids the cumbersome region proposition and region-by-region classification steps of traditional methods by dividing the image into fixed grids and letting each grid predict the target's category probability and bounding box parameters at the same time. and region-by-region categorization steps in traditional methods. This innovative design not only reduces the complexity of the model but also allows target detection to be adapted to real-time application scenarios while maintaining high accuracy (Redmon et al., 2016).

Bochkovskiy et al. proposed the YOLOv4 method, an approach that enhances the performance of Convolutional Neural Networks (CNNs) in target detection by introducing several new features such as Weighted Residual Connection (WRC), Cross-Stage Partial Connection (CSP), Cross-Minor Batch Normalization (CmBN), Self-Adversarial Training (SAT), and Mish activation. The method ultimately achieves 43.5% AP on the MS COCO dataset (65.7% AP50) and real-time detection on a Tesla V100 at about 65 FPS (Bochkovskiy et al., 2020).

YOLO in Intelligent Driving: YOLO series models are widely used in intelligent driving, especially in the detection of vehicles, pedestrians, traffic signs and other targets.YOLOv3 and YOLOv4 have achieved better results in urban traffic environments, and are able to achieve real-time detection in complex traffic scenarios.YOLOv5 further optimizes the model structure and improves the detection accuracy, especially in small object detection.

3.2 Methods Based on Other Models

In addition to the YOLO series, some other classical target detection methods have also been applied and studied in intelligent driving.

Faster R-CNN is a target detection method of region proposal network combined with the convolutional neural network, Faster R-CNN achieves high accuracy by generating candidate regions and subsequently performing classification and regression. However, due to its high computational complexity and poor real-time performance, it is more often used in non-real-time demanding scenarios. Girshick et al. proposed Faster R-CNN, which describes how to optimize the target detection process through a region proposal network and significantly improves the target detection accuracy (Girshick et al., 2015).

Xie et al. used Faster R-CNN to improve the detection accuracy of traffic signs and lane lines in smart driving applications with good experimental results (Xie et al., 2018).

Single Shot Multibox Detector (SSD): SSD is an improved target detection method that utilizes feature maps at different scales for target detection to achieve efficient detection of multi-scale objects.SSD has superior performance in dealing with small objects and is suitable for target detection in complex traffic environments. Liu et al. proposed SSD, which is an efficient solution to the problem of detecting multi-scale targets and achieved better detection accuracy (Liu et al., 2016). Zhang et al. successfully solved the detection problem of small objects using SSD in autonomous driving applications, especially in highway and urban intersection environments (Zhang et al., 2019).

RetinaNet: RetinaNet solves the category imbalance problem by introducing Focal Loss, which is outstanding in dealing with highly imbalanced datasets. The method has a better performance in detection accuracy, especially in small target detection with advantages. Lin et al. proposed RetinaNet and solved the category imbalance problem in traditional target detection methods by Focal Loss (Lin et al., 2017). Wei et al. applied RetinaNet to smart driving scenarios and proposed a deep learning-based small object detection method to optimize detection accuracy in low-light environments (Wei et al., 2020).

4. EXISTING LIMITATIONS AND FUTURE PROSPECTS

Although existing target detection techniques have made significant progress in many aspects, they still face some challenges and limitations. First, existing target detection methods are less robust in complex environments. For example, in bad weather (e.g., rain, haze) or poor lighting conditions, the image quality of sensors (especially cameras) degrades significantly, resulting in lower target detection accuracy. Although LiDAR and millimeter-wave radar can provide better detection performance in bad weather, their relatively low resolution and accuracy cannot yet fully replace visual perception. In addition, the detection of small objects is still a difficult problem. Especially in complex backgrounds, the detection of small objects (e.g., pedestrians, low obstacles, etc.) is not ideal. Finally, target detection algorithms usually need to find a balance between accuracy and real-time performance. High-precision models usually require more computational resources, resulting in slower detection; while models pursuing speed often sacrifice accuracy. Therefore, how to balance accuracy and real-time is still an important issue in autonomous driving platforms with limited computational resources.

In the future, target detection techniques are expected to overcome these limitations through multimodal data fusion. Multimodal fusion refers to the combination of different types of sensors, such as cameras, lidar, millimeter-wave radar, etc., to make up for the shortcomings of a single sensor. For example, the combination of visual sensors and radar can effectively deal with the detection problem under adverse weather conditions. In addition, unsupervised learning, as an emerging learning method, does not rely on manually labeled data but learns through the structure or contextual information of the data itself. Unsupervised learning can reduce the need for large-scale labeled data, thus accelerating model training and improving model adaptability in new scenarios. Hardware acceleration and model lightweight are also directions for future research. With the popularity of hardware platforms such as GPUs and TPUs, as well as the development of model optimization techniques such as quantization and pruning, the target detection model will be more efficient and adapt to the application scenarios of low-power devices and embedded systems.

5. CONCLUSION

This paper summarizes the development status and challenges of intelligent driving target detection technology, and deeply analyzes the key factors affecting the detection effect. With the rapid development of automatic driving technology, target detection has become one of the indispensable core technologies in intelligent driving systems. This paper analyzes the advantages and limitations of these methods in practical applications, especially their adaptability in automatic driving scenarios, through a detailed discussion of the YOLO series and other classical target detection methods. The YOLO series methods, by virtue of their efficient real-time detection capability and good accuracy, have become one of the mainstream target detection algorithms that are widely used in the field of automatic driving at present.

With the continuous progress of sensor technology, deep learning algorithms and computational power, the target detection technology is expected to make greater breakthroughs in accuracy, real-time and robustness. Meanwhile, the optimization of deep learning algorithms, especially the combination of cutting-edge technologies such as CNN and Reinforcement Learning (RL), may lead to significant improvements in the accuracy and efficiency of target detection. With the continuous progress of these technologies, the intelligent driving system will gain greater improvement in perception capability, thus promoting the rapid development and commercialization of automatic driving technology.

REFERENCES

Bochkovskiy, A., Wang, C. Y., & Liao, H. Y. M. 2020. YOLOv4: Optimal speed and accuracy of object detection. arXiv preprint arXiv:2004.10934.

Girshick, R., Donahue, J., Darrell, T., & Malik, J. 2015. Region-based convolutional networks for accurate object detection and segmentation. IEEE Transactions

on Pattern Analysis and Machine Intelligence, 38(1), 142-158.
Kim, Y., & Lee, K. 2020. A multi-sensor fusion approach for object detection in bad weather conditions using LiDAR and camera data. Sensors, 20(15), 4241.
Li, Y., Zhang, X., & Xie, L. 2019. An infrared and visible camera fusion system for robust vehicle detection in low-light conditions. Journal of Optical Society of America A, 36(5), 729-738.
Liu, W., Anguelov, D., Erhan, D., Szegedy, C., & Reed, S. 2016. SSD: Single shot multibox detector. European Conference on Computer Vision (ECCV), 21–37.
Lin, T. Y., Goyal, P., Girshick, R., He, K., & Dollár, P. 2017. Focal loss for dense object detection. IEEE International Conference on Computer Vision (ICCV), 2980-2988.
Redmon, J., Divvala, S., Girshick, R., & Farhadi, A. 2016. You only look once: Unified, real-time object detection. IEEE Conference on Computer Vision and Pattern Recognition (CVPR), 779-788.
Wei, X., Zhang, Y., & Zhang, L. 2020. RetinaNet for small object detection in autonomous driving. IEEE Transactions on Intelligent Transportation Systems, 21(1), 1-11.
Xie, L., Zhang, H., & Liu, S. 2018. Improving lane detection and traffic sign recognition for intelligent driving systems using Faster R-CNN. Sensors, 18(5), 1579.
Zhang, Z., Liu, L., & Zhao, Y. 2019. A convolutional neural network-based denoising method for camera images in rainy and snowy weather conditions. IEEE Transactions on Neural Networks and Learning Systems, 30(4), 1053-1064.