Proceedings of the 2023 International Conference on Machine Learning and Automation
DOI: 10.54254/2755-2721/42/20230670

An analysis of various deep learning-based target detection algorithms in the field of autonomous driving

Jiale Wei

Faculty of Science, Department of Mathematics, Xi'an University of technology, Xi'an, 710000, China

3210813034@stu.xaut.edu.cn

Abstract. Target detection is a crucial research objective within the domain of computer vision, finding extensive applications in areas such as robotics, autonomous driving, industrial inspections, and various other fields. Based on the foundation of deep learning theory, this paper systematically summarizes the application and prospect of each type of target detection algorithm (based on regression and based on candidate region) on automatic driving, compares the advantages and disadvantages of the two types of algorithms, as well as the results of detecting traffic signals, traffic vehicles, and pedestrians, and focuses on the application scenarios as well as the comparison of advantages and disadvantages of each method. A systematic summary of the current development results is made. Among them, the most prominent target detection in the field of transportation is undoubtedly the algorithms of various branches of the YOLO series.

Keywords: Self-Driving, Object Detection, Deep Learning, Neural Network.

1. Introduction

As the boundaries of artificial intelligence continue to expand, the field of autonomous driving has become a transformative innovation in modern transportation. Autonomous driving systems are moving towards seamless integration into our daily lives, giving vehicles the ability to drive autonomously on the road and promising to revolutionize the way we travel. This paradigm shift is expected to bring a range of societal benefits, including the potential to reduce traffic accidents, optimize traffic flow, and free up drivers' leisure time or pursuit of productivity. As the future direction of automobile development, self-driving cars have the ability to make autonomous judgments, which can reduce human errors to a greater extent can better save energy and reduce pollution and has a good application prospect. According to the degree of intelligence of the vehicle, China's Ministry of Industry and Information Technology (MIIT) approved the release of the GB/T 40429-2021 "Automotive Driving Automation Classification" standard on August 20, 2018 [1]. Automobile automated driving is divided into L0~L5 levels by this standard, and the specific grading standards are shown in Table 1. In this dynamic environment, object detection becomes key in the field of computer vision as it accurately identifies obstacles, pedestrians, vehicles, traffic signs, and other key elements through sophisticated analysis of image and point cloud data captured by sensors such as cameras and LIDAR. Accurate detection, recognition, and judgment of real-time targets is fundamental and central to its operation. The process of fine-grained target detection plays a key role in empowering autonomous driving systems to make

DOI: 10.54254/2755-2721/42/20230670

informed decisions, thereby ultimately increasing the level of driving automation and potentially transforming the urban landscape.

Table 1. GB/T 40429-2021 Automotive driving automation classification standard.

scale	name	define	Vehicle motion control	Target detection and response	Dynamic driving task takeover	Design operating conditions
L0	Emergency assistance	The system lacks the capability to consistently carry out vehicle control during a dynamic driving task, but it possesses the capacity to continuously identify and react to certain targets and events within such a dynamic driving task.	drivers	drivers + systems	driver	restrictive
L1	Partial Driving Assistance	The system consistently carries out dynamic driving tasks within predetermined operational parameters and possesses limited target and event detection and response capabilities suitable for both horizontal and vertical vehicle control.	drivers + systems	drivers + systems	driver	restrictive
L2	Combined Driving Assistance	The system consistently executes dynamic driving tasks within specified operational parameters, equipped with partial target and event detection as well as response capabilities tailored to both horizontal and vertical vehicle control.	systems	drivers + systems	driver	restrictive
L3	Conditional Autopilot	The system consistently executes all dynamic driving tasks within the specified operating conditions for which it was designed	systems	systems	dynamic driving tasks take over the user	restrictive
L4	Highly automated driving	The system consistently executes a wide range of dynamic driving tasks within its designated operational parameters and autonomously employs a low-risk approach.	systems	systems	systems	restrictive
L5	fully automatic driving	The system consistently executes all dynamic driving functions in any drivable condition and automatically employs a strategy with minimal risk.	systems	systems	systems	limitless

Amid the evolving autonomous driving landscape, the exploration and deployment of diverse object recognition algorithms bear testament to their seminal role in propelling technological advancement. Diverse algorithms, ranging from the pioneering Convolutional Neural Networks (CNN) to you only look once (YOLO) [2] and Fast R-CNN [3], furnish an expansive repertoire of solutions for autonomous

DOI: 10.54254/2755-2721/42/20230670

driving systems. These algorithms, adept at extracting salient features from images, are primed to pinpoint, categorize, and meticulously characterize the myriad of objects that adorn the urban thoroughfare. However, the intricate tapestry of real-world scenarios introduces many challenges—shifts in lighting conditions, inclement weather, and the concealing of objects due to obstructions or occlusions. As a result, researchers are galvanized to continuously refine and innovate object recognition algorithms, bolstering the system's stability, reliability, and adaptability to ensure safe and efficient navigation. This paper presents a comprehensive study of the complex interplay between autonomous driving and target recognition algorithms, particularly emphasizing the critical role of target detection. The study provides an in-depth comparison of the effectiveness and accuracy of algorithms in detecting different types of objects when confronted with them, examining aspects such as accuracy, real-time responsiveness, and the ability to navigate different environments seamlessly. In this survey work, this paper endeavours to shed light on the strengths and limitations inherent in different object recognition algorithms in the vast field of autonomous driving technology. In doing so, it enhances our understanding of the complex challenges and potential breakthroughs that contribute to the enhancement of object perception in autonomous driving systems, ultimately contributing to the overall goal of safer and more efficient transportation.

Target detection algorithms can be categorized into candidate region-based (two-stage) and regression-based (one-stage). The most significant contrast between these two methods lies in their approach to generating candidate bounding boxes. The former relies on sub-networks to assist in this process, whereas the latter directly causes candidate bounding boxes on the feature map. This research specifically centers on examining the effectiveness of these two types of object detection algorithms in the context of autonomous driving, with a thorough investigation into their performance in real-world road scenarios. By examining factors such as accuracy, real-time responsiveness, and adaptability to different environments, this paper aims to reveal the inherent strengths and limitations of different target recognition algorithms in the vast field of autonomous driving technology.

2. Method

Target detection algorithms have emerged as a prominent research focus within the realm of computer vision in recent years, comprising two main categories: candidate region-based and regression-based approaches. The inception of candidate region-based algorithms can be traced back to the introduction of R-CNN by Girshick et al. in 2014 [4], marking the first instance of deep learning integration into target detection. Subsequently, advancements in this domain led to the development and evaluation of algorithms like Faster R-CNN and Mask R-CNN. In 2016, Redmon et al. introduced the YOLO algorithm, while Liu et al. proposed the SSD algorithm, pioneering the field of regression-based algorithms. Detailed descriptions of these specific algorithms can be found in Figure 1.

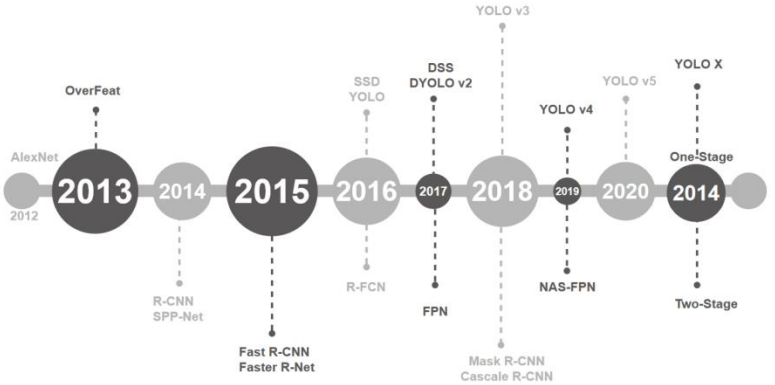

Figure 1. Representative object detection algorithms.

DOI: 10.54254/2755-2721/42/20230670

Candidate region-based algorithms are generally slow in detection and cannot meet the real-time detection but have relatively high detection accuracy when facing traffic scenes. On the contrary, regression-based algorithms are fast in detection but relatively inferior to the two-stage algorithms in terms of accuracy. Since efficient detection in traffic scenarios becomes more suitable, the application of target detection to real-time sensing in autonomous driving systems has become a reality with the proposal of a phase of target detection algorithms. The two most typical types of these algorithms are the SSD (single shot multibox detector) series and YOLO (you only look once) series.

3. Result

While the YOLO series of algorithms have successfully achieved real-time target detection, they appear to have some limitations when it comes to detection accuracy. For instance, in the case of YOLO v1, it tends to struggle with detecting small objects densely distributed, often leading to missed detections. However, in recent years, the Remon team has been actively enhancing the YOLO algorithm, progressing from v3 to subsequent versions like YOLO v4 and YOLO v5. Throughout these updates, both the accuracy and real-time performance of YOLO have seen notable improvements, as illustrated in Table 2.

Table 2. Advantage, disadvantages and applicable scenarios of One-Stage target detection algorithm.

	Backbone network	advantages and disadvantages	FPS	Usage Scenarios
YOLO v1	VOG-16	Simple network, fast detection, but poor localization, poor detection of small targets	45.0	target detection
YOLO v2	DarkNet-19	Reduced localization errors, high classification accuracy, but not very accurate	40.0	target detection
YOLO v3	DarkNet-53	Improved accuracy, more than 3 times faster, but not enough accuracy with tight bounding box predictions	19.6	Multi-scale target detection
YOLO v4	CSPDarknet-53	Improved detection of small targets with high model complexity	65.0	Highly accurate real-time target detection
YOLO X	Darknet-53	Improve anchor-based pipeline over-optimization with multiple network framework options	68.9	Highly accurate real-time target detection
SSD	VOG-16	Fast detection, low accuracy, poor detection of small targets	19.0	Multi-scale target detection

To ensure that autonomous driving does not cause harm to life safety and public property, it is necessary to apply object detection algorithms to traffic scenarios to achieve intelligent transportation and autonomous driving, avoiding casualties and property damage. In traffic scenarios, the targets that need to be detected mainly include traffic signals, vehicles, and pedestrians.

In traffic scenarios, the accuracy of traffic signal recognition is very important for autonomous driving systems. This is because the basic requirements for automatic driving can only be met if information such as the speed limit of the current roadway, whether the operation ahead is going straight or turning, etc. can be accurately detected.

Zhang et al [5] formed the letter's dataset CCTSDB by altering the Chinese Traffic Sign Dataset (CTSD) and trained YOLO v2 using an intermediate layer employing multiple 1x1 convolutional layers. Mohd-Isa et al [6] used Spatial Pyramid Pooling (SPP) to influence the YOLO v3 framework to further recognize traffic signs in real environments. Yang et al [7] trained the CCTSDB training set using YOLO v3 and YOLO v4 respectively. Awi et al [8] on the other hand used generative adversarial network to combine the synthetic image with the original image and trained using YOLO v3 and YOLO v4. Lui et

DOI: 10.54254/2755-2721/42/20230670

al. [9] on the other hand used EIOU (efficient intersection over union), a new loss function, to improve the accuracy of YOLO v5.

During the study of traffic signs using target detection, changes are made in two ways changing the data set and modifying the structure of the algorithm function. (The structure of the studies is shown in Table 3).

Table 3. Research work on the detection of traffic signs.

reference	dataset	algorithms	mAP%	Precision/%	Recall/%	Accuracy/%	IoU/%
[5]	CCTSDB	Improvement of YOLO v2		96.69	86.67		
[6]	MTSD(self-built)	Improvement of YOLO v3		82.50	92.15	91.00	
[7]	CCTSDB	YOLO v3				91.73	67.08
		YOLO v4				94.56	68.78
[8]	self-built	YOLOv3	99.83			96.00	
		YOLO v4	99.98				
[9]	CCTSDB	Improvement of YOLO v5	84.35	85.92	84.71		

Fast and accurate identification of other vehicles plays an important role in safe vehicle operation, however, other vehicles encountered on the road are susceptible to problems such as light intensity, weather changes, and occlusion, which become difficult to recognize. This creates a significant safety risk for autonomous driving applications. Therefore, how to achieve accurate and real-time detection and identification of vehicles on the road in complex natural traffic scenarios is a current research issue.

With the development of deep learning in recent years, target detection algorithms have become the mainstream method for traffic vehicle detection and identification. The target detection algorithm effectively overcomes the detection and recognition difficulties brought about by the changing appearance of traffic vehicles due to a certain degree of invariance to geometric transformations, shape changes, etc., and it can adaptively construct features in the samples, avoiding incomplete and omitted manually constructed features.

In 2016, the Faster R -CNN algorithm enabled end-to-end recognition, and the YOLO and SSD algorithms in the same year made traffic vehicle detection more efficient. Chen et al [10] proposed a Hybrid Deep Convolutional Neural Network for satellite image vehicle target detection and this algorithm can acquire variable scale features. Ye Jialin et al [11] used GIOU loss function to improve YOLO v3 and improved the accuracy of algorithm localization. Cao [12] realized real time target tracking by combining SSD algorithm, Camshift tracking and kalman algorithm. Lu et al [13] successfully combined edageBoxes and Fast R-CNN algorithms, which can obtain accurate target regions with improved recognition accuracy. SegNet used by Huazhong University of Science and Technology can save time and memory for traffic vehicle detection.

Many studies have shown that deep learning-based object detection algorithms have better detection performance than traditional detection and recognition methods and are well reflected in mAP values. Using object detection algorithms to detect nonmotorized vehicle targets in traffic scenes can avoid the limitations of traditional manual feature extraction, more effectively extract features, and accurately detect traffic vehicle targets.

As research on object detection algorithms continues to advance, the associated challenges and complexities are progressively mounting. These challenges encompass the quest for heightened detection accuracy, while concurrently addressing issues like reduced processing speed and subpar performance in detecting small objects. Traditional object detection algorithms are progressively falling short in meeting the requisites of applications in traffic scene object detection and recognition. Consequently, there is an imperative need to refine and enhance these conventional object detection

DOI: 10.54254/2755-2721/42/20230670

algorithms. Currently, optimization efforts in the realm of object detection algorithms primarily revolve around five key facets: augmenting features, integrating contextual information, optimizing anchor box design, enhancing non-maximum suppression algorithms, and refining loss functions.

The background of traffic scenes is complex and variable, and improving the robustness of object detection algorithms through optimization methods can achieve better application of object detection algorithms. However, the application of the algorithm based solely on the object detection model itself is singular and the improvement in detection performance is not significant. Therefore, further research is needed in conjunction with other methods.

Pedestrian detection stands as a significant research endeavour within the realm of object detection. Its primary focus lies in harnessing computer technology to ascertain the presence of pedestrians within an image and, if identified, annotate both the category and the location of the detected pedestrians. Traditional pedestrian detection methods have frequently grappled with issues like feature omission, limited detection precision, intricate operations, and the demand for substantial human and material resources. In recent years, the field of pedestrian detection has benefited from the adoption of object detection algorithms thanks to their remarkable detection capabilities.

At present, object detection algorithms have good performance in general pedestrian image datasets. However, accidents involving casualties and property damage in natural scenes mainly occur at night and in adverse weather conditions. How to detect and identify pedestrian targets at night and in adverse weather conditions is a current research challenge. Researchers have adopted multiple methods, among which the better one is multimodal object detection. The researcher successfully reduced the leakage rate by modifying the network structure of YOLO v2(as shown in Figure 2).

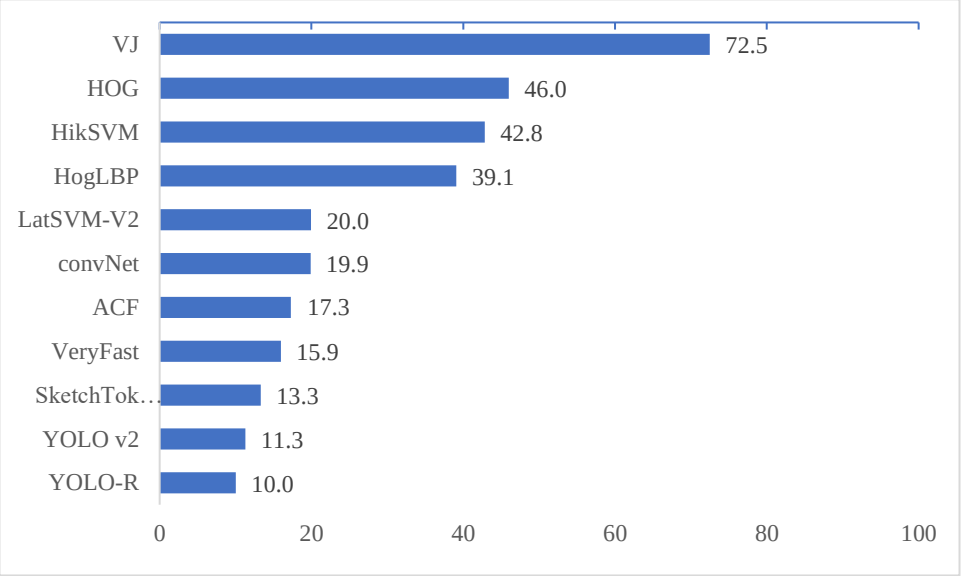

Figure 2. Comparative Analysis of Pedestrian Detection Algorithm Miss Detection Rates.

Zhang et al. developed a Cap-YOLO detection model using YOLOv3. They employed a dense connection to construct a dense block component, enhancing feature map utilization and effectively reducing detection model leakage through a dynamic routing mechanism. Additionally, Liu [14] utilized the K-means clustering method to compute dataset frame sizes directly, introducing the SE module and the DIOU loss function to enhance small target detection accuracy for pedestrians.

In addition, pedestrian re-recognition is also an important research branch of pedestrian detection. The main research content of pedestrian re-recognition is to determine whether a pedestrian in a certain camera has appeared in other cameras, which requires comparing a pedestrian feature with other pedestrian features to determine whether it belongs to the same pedestrian. At present, pedestrian re-recognition mainly relies on traditional methods, strongly supervised deep learning methods, and

DOI: 10.54254/2755-2721/42/20230670

unsupervised methods. Traditional methods mainly rely on feature extraction and metric learning methods, and most unsupervised methods are also based on research on traditional methods [15].

At present, pedestrian detection has achieved very high accuracy and recognition accuracy on public datasets, but for complex and dense traffic scenes, pedestrian detection still has a long way to go. Currently, pedestrian re-recognition is a focus of research in the field of pedestrian detection, and how to accurately identify occluded pedestrian targets in complex natural environments remains a research challenge.

4. Conclusion

Target detection represents a crucial research area with vast potential applications. This paper presents an exhaustive overview of the historical development and current research status of target detection algorithms, encompassing two major categories: those reliant on candidate regions and regression-based approaches. Building upon this foundation, the study examines and consolidates the present state of research and application in target detection algorithms, encompassing target detection, algorithm optimization, dataset enhancement, and more, with a focus on typical traffic scene objects like traffic signals, vehicles, and pedestrians. Lastly, the paper provides insights into commonly used target detection methods within traffic scenarios, facilitating a comparative analysis of the performance of various target detection algorithms.

Overall, target detection algorithms have been applied to traffic signal, vehicle and pedestrian detection and recognition, which are very rich in variety, but then there is no powerful algorithm that is highly efficient while maintaining high accuracy. Different target detection tasks have different requirements for the model, and the model should be improved accordingly to the specific scene and task characteristics. Currently, object detection algorithms have shown good performance in public traffic scene datasets, but there are still some problems in their application in actual traffic scenarios. Several research trends are proposed for this: 1. Investigate feature extraction networks better suited for target detection tasks. Presently, the feature extraction networks used in object detection algorithms predominantly rely on classification networks. However, the design principles for networks used in classification and detection tasks differ significantly. Furthermore, variations in datasets also give rise to challenges in object detection. Hence, it is imperative to begin from the object detection model itself and construct a feature extraction network tailored to the demands of object detection, thereby enhancing the detection performance of target objects. 2. Multimodal Object Detection. Data fusion plays a crucial role in accomplishing object detection tasks. While numerous algorithms for multimodal object detection have been continuously introduced, they predominantly rely on image data. Substantial variations in lighting conditions can result in distortion in camera recordings and hinder the ability to perceive scene information. Therefore, it is essential to explore the utilization of the complementary nature of multimodal data to bolster the model's resilience, encompassing the fusion of image, audio, text, and other information sources. 3. A model for target detection under weak supervision. Presently, most object detection algorithms predominantly rely on supervised learning, demanding substantial quantities of annotated data. The process of data annotation incurs significant labour costs. Consequently, the exploration of techniques like weakly supervised learning and small-sample learning to develop weakly supervised object detection models in the absence of annotated data has emerged as a prominent research focus.

References

[1] National Automotive Standardization Technical Committee 2021 Automotive driving automation classification GB/T 40429-2021

[2] Redmon, J., Divvala, S., Girshick, R., and Farhadi, A., 2016. You only look once: Unified, real-time object detection. 2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR).

DOI: 10.54254/2755-2721/42/20230670

[3] Ren, S., He, K., Girshick, R., and Sun, J., 2017. Faster R-CNN: Towards real-time object detection with region proposal networks. IEEE Transactions on Pattern Analysis and Machine Intelligence, 39(6), pp.1137–1149.
[4] Girshick, R., Donahue, J., Darrell, T., and Malik, J., 2014. Rich feature hierarchies for accurate object detection and semantic segmentation. 2014 IEEE Conference on Computer Vision and Pattern Recognition.
[5] Zhang, J., Huang, M., Jin, X., and Li, X., 2017. A real-time Chinese traffic sign detection algorithm based on modified yolov2. Algorithms, 10(4), p.127.
[6] Mohd-Isa, W.-N., Abdullah, M.-S., Sarzil, M., Abdullah, J.,, Ali, A., and Hashim, N., 2020. Detection of Malaysian traffic signs via modified Yolov3 algorithm. 2020 International Conference on Data Analytics for Business and Industry: Way Towards a Sustainable Economy (ICDABI).
[7] Yang, W., and Zhang, W., 2020. Real-time traffic signs detection based on Yolo Network Model. 2020 International Conference on Cyber-Enabled Distributed Computing and Knowledge Discovery (CyberC).
[8] Dewi, C., Chen, R.-C., Liu, Y.-T., Jiang, X., and Hartomo, K.D., 2021. Yolo V4 for Advanced Traffic Sign Recognition with synthetic training data generated by various gan. IEEE Access, 9, pp.97228–97242.
[9] Lv, H., and Lu, H., 2021. Research on traffic sign recognition technology based on YOLOv5 algorithm. Journal of Electronic Measurement and Instrumentation, pp.137–144.
[10] Xueyun Chen, Shiming Xiang, Cheng-Lin Liu, and Chun-Hong Pan, 2014. Vehicle detection in satellite images by hybrid deep convolutional neural networks. IEEE Geoscience and Remote Sensing Letters, 11(10), pp.1797–1801.
[11] Jialin Ye, Ziyi Su,, Haoyan Ma,, Xia Yuan, and Chunxia Zhao, 2021. Improved yolov3 method for non-motorized vehicle detection and recognition. Journal of Computer Engineering & Applications.
[12] Wei Cao. 2018 Research on SSD-based vehicle detection and tracking algorithm, Doctoral dissertation, Anhui University [Preprint].
[13] Xue Lu, Yongxiang Chen and Kun Liu 2019 A deep learning algorithm for non-motorized vehicle target detection, Computer Engineering and Applications [Preprint]. doi: 10.3778/j.issn.1002-8331.1801-0199.
[14] Liu, S. et al., 2021, Research on Small Target Pedestrian Detection Algorithm Based on Improved YOLOv3, International Conference on Genetic and Evolutionary Computing, pp. 203–214.
[15] Zhang, B., Zhou, Z., Cao, W., Qi, X.,, Xu, C., and Wen, W., 2022. A new few-shot learning method of bacterial colony counting based on The edge computing device. Biology, 11(2), p.156.