Proceedings of the 2023 International Conference on Machine Learning and Automation
DOI: 10.54254/2755-2721/39/20230575

Deep learning based multi-target detection for roads

Junlin Jiang

Department of Computer Science and Technology, Southwest Petroleum University, Chengdu, Sichuan, 610500, China

lbj38228@gmail.com

Abstract. The vehicle target detection algorithm based on deep learning has gradually become a research hotspot in this field. In recent years, with the significant breakthrough of deep learning in the field of target recognition, the vehicle target detection algorithm based on deep learning has gradually become a research hotspot in this field. For the task of vehicle target detection, this paper first briefly introduces the process of traditional target detection algorithms and some optimization methods. It summarizes the development process of YOLO, the current mainstream one-stage vehicle target detection algorithm, and the process of Faster R-CNN, the second-stage vehicle target detection algorithm, and its improvement. Then the characteristics of several types of representative convolutional neural network algorithms are analyzed in chronological development order. Finally, it looks forward to t he future research direction of vehicle target detection algorithms, and also provides new ideas for the optimization of the subsequent vehicle target detection algorithms, which have good engineering application value. Provides algorithmic support for the underlying logic of autonomous driving.

Keywords: Deep Learning, Target Detection, Autonomous Driving, Convolutional Neural Networks.

1. Introduction

With the arrival of 2023, artificial intelligence technology based on deep learning has become a hot research topic nowadays. Artificial intelligence technology is also utilized in various fields, such as natural language processing [1], autonomous driving [2], smart healthcare. In China, automobiles are the most common means of transportation for people's daily travel, and per capita holdings are climbing as shown in Figure1.1.

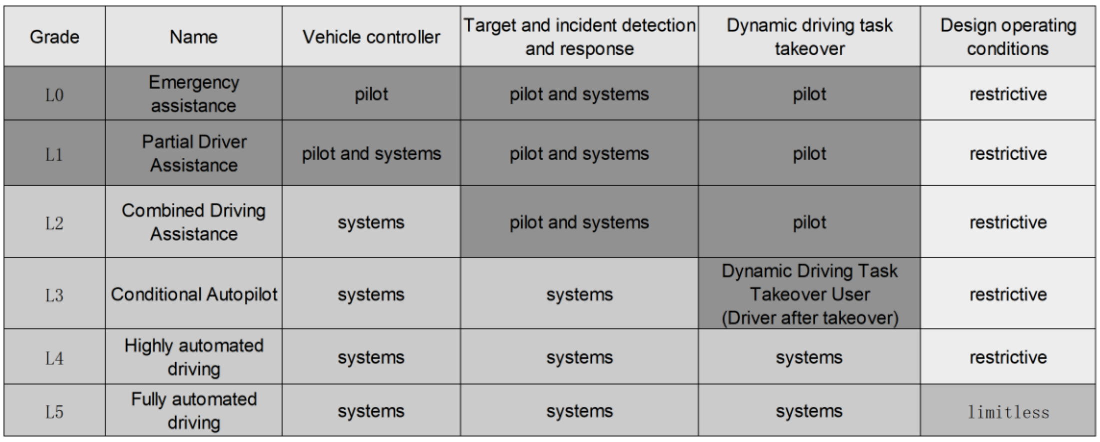
Table 1. China's highway mileage and vehicle ownership in recent years in China.

Particular year	Mileage (million kilometers)	Mileage increase year-on-year (%)	Ownership (billions of units)	Year-on-year growth in holdings (%)
2012	433.75	-	1.21	-
2013	435.82	2.85	1.37	13.22
2014	446.39	2.42	1.54	12.41
2015	457.73	2.54	1.72	11.69
2016	469.63	2.60	1.94	12.79

Table 1. (continued).

2017	477.35	1.64	2.17	11.86
2018	484.65	1.53	2.41	11.06
2019	501.25	3.43	2.50	3.73
2020	513.54	2.45	2.75	10.00
2021	519.58	1.22	3.02	9.80
2022	525.22	1.18	2.98	9.87
Aggregate growth	96.06	22.67	1.81	149.59

Its line automobile industry in the booming development at the same time also to the development of the city has created a large number of urgent needs to solve the traffic problems. These problems will cause a great threat to the safety of people's driving process. So how to reduce the incidence of traffic accidents, improve people's driving safety coefficient, reduce the economic loss of society, and so on has become an urgent problem to be solved [3]. The main reason for these problems is often due to the driver's unfamiliarity with the surrounding environment during driving, as well as the inability to make a timely and correct response to unexpected situations [4]. Therefore, Intelligentizing the Vehicle and how to make it autonomously perceive the surrounding road environment to achieve automatic driving, to improve the safety coefficient of people's driving process, are important research value and significance. To be more convenient and understand the automatic driving technology, the automatic driving classification standard has an important significance. In August 2021, the National Recommended Standard for Automated Vehicle Driving Classification was officially introduced, as shown in Figure 1.1 below. The domestic automated driving classification standard has also become a new reference standard for researchers.



Grade	Name	Vehicle controller	Target and incident detection and response	Dynamic driving task takeover	Design operating conditions
L0	Emergency assistance	pilot	pilot and systems	pilot	restrictive
L1	Partial Driver Assistance	pilot and systems	pilot and systems	pilot	restrictive
L2	Combined Driving Assistance	systems	pilot and systems	pilot	restrictive
L3	Conditional Autopilot	systems	systems	Dynamic Driving Task Takeover User (Driver after takeover)	restrictive
L4	Highly automated driving	systems	systems	systems	restrictive
L5	Fully automated driving	systems	systems	systems	limitless

Figure 1. National Recommended Standards for Automated Vehicle Driving Classification.

Target detection, as one of the three basic research areas of computer vision, is now widely used in the field of road traffic, e.g., it can be used in intelligent traffic signal control systems to optimize the control of signals according to the traffic flow and vehicle types on the road, thus improving traffic flow and reducing emissions; or, it can be used for real-time monitoring and analysis of traffic conditions in traffic monitoring systems. It can detect vehicle violations, congestion, traffic accidents, and pedestrian crossings, thus helping traffic management to take appropriate measures. It is crucial technology in the field of autonomous driving. It helps vehicles to recognize and track other vehicles, pedestrians, bicycles, traffic signs, and road markings, etc., thus enabling autonomous navigation and safe driving.

In this paper, a triple classification method containing a one-stage target detection algorithm, a two-stage target detection algorithm, and a neural network structure is proposed for the first time for road multi-target detection scenarios. Among the traditional methods, the method of using the feature vector calculated based on Haar-like features as the input parameter of the AdaBoost classifier is the best. Among the YOLO series, YOLOv5 is widely recognized by workers for its excellent detection rate and accuracy as well as its self-selectable model according to the environment. Among the two-stage target detection algorithms, the Faster R-CNN algorithm, which is comprehensively optimized from three aspects: training samples, feature extraction network, and feature fusion, is the optimal solution.

2. Related work

2.1. Detection of Vehicle Targets in Traditional Road Scenes

Prior to the advent of deep learning [5], most of the traditional feature engineering-based target detection methods were used for the task of vehicle target detection in road scenes. These methods typically involve manually designing features and using traditional machine-learning algorithms for target detection. One of the typical methods is the sliding window-based target detection method, which applies sliding windows of different sizes and positions to an image extracts the features within the window using a predefined feature extraction method (e.g., HOG or SIFT) and then performs the target detection by a classifier (e.g., SVM).

Viola and Jones made some improvements on this basis and introduced a cascade classifier based on Haar [15] features, which is a fast feature extraction method that can effectively detect features such as edges, line segments, and corners in an image. Cascade classifier is a method of cascading multiple classifiers, Each classifier is used to screen out the candidate targets, and the subsequent classifiers gradually carry out further validation of the candidate targets, thus realizing efficient target detection.

2.2. Detection of pedestrians in UAV aerial video with depth features

Nowadays, in the detection of pedestrians in UAV aerial video based on depth features [6], the mainstream method mainly consists of two-stage algorithms, which are ONE STAGE algorithm and TWO STAGE algorithm.

Typical representatives in the first category of ONE STAGE are the YOLOv1 algorithm proposed by Joseph Redmon in 2015, SSD proposed by Wei Liu et al. in 2016, and RetinaNet. Their common feature is to directly output the location and category information of target pedestrians through a neural network, thus reducing the computational complexity to a certain extent.2023 The algorithmic model of YOLOv8, developed by Ultralytics, employs an Anchor-Free-based detection method, compared to the traditional Anchor-Based detection method. It has higher detection accuracy and faster detection speed. It significantly reduces the time needed to localize and classify the target pedestrians during UAV aerial photography.

The second class of TWO-STAGE methods is faster to detect than the first class of ONE-STAGE methods, but the accuracy decreases. Its common algorithms are Fast-RCNN, Faster-RCNN, Cascade R-CNN, and so on. In 2013, after R-CNN [23] was first proposed by Ross Girshick et al. In 2015, Girshick et al. proposed Fast R-CNN, which improves on R-CNN. It feeds the entire UAV-captured image into the CNN network for feature extraction and then obtains the features of each character candidate frame from the feature map through the RoI pooling layer. This approach avoids extracting features for each character candidate frame individually, which improves computational efficiency and also introduces the idea of sharing features for classification and localization.

3. Method

Traditional target detection algorithms are generally divided into three steps: target candidate region generation, feature extraction, and feature matching or classification. Target candidate region generation refers to locating regions in an image that may contain targets such as vehicles, pedestrians, road signs, etc., which are usually called candidate regions. These regions can be generated by sliding window methods, image segmentation, region growing, etc. Feature extraction, on the other hand, extracts features from the candidate regions using an extraction method designed by the human himself. These appearance features include symmetry of the vehicle, underbody shading, etc. The vehicle target can be detected by the local feature that the rear part of the vehicle is symmetrical. These features capture the shape information in the vehicle image and play a key role in target detection classification. This type of detection method that directly uses vehicle appearance features is intuitive to understand, but it is not robust in real complex traffic scenarios and is affected by external environments such as weather, light, and road defacement. Meanwhile, the selection of algorithmic thresholds when extracting specific appearance features is also a difficult matter.

After obtaining manually designed features based on human perception of vehicles, it is common to use classifiers to do the detection task, two of the most widely used typical representatives of SVM and AdaBoost.AdaBoost is a classifier widely used in the field of target detection, and Yoav Freund et al. used AdaBoost based on the assumption of symmetry of the vehicle [7] classifiers for vehicle detection classifier to detect vehicles. Meanwhile, the HOG [8] feature operator extracts image features as input feature vectors for SVM classifiers and has many applications in vehicle detection. On the basis of both, Shujuan S et al. detect vehicles based on the idea of Haar-like features computed feature vectors as input parameters of the AdaBoost classifier, which also improves the rate of traditional target detection algorithms greatly improved. It is also the most recommended traditional target detection method in this paper.

However, these traditional vehicle target detection algorithms have a large number of redundant computations in the process, which are destined to make it difficult to improve the operation speed. Therefore they are gradually abandoned by autonomous driving which requires high accuracy and real-time performance.

Phase I Vehicle Target Detection Algorithm:

A real-time end-to-end object detection method, YOLO, was first presented at CVPR in 2016 by Joseph Redmon et al. Compared to the traditional target detection algorithms previously devised by Yoav Freund, Shujuan S , and others, YOLO requires only one pass through the network to complete the detection task and uses a more direct output that predicts the detection output based on regression only, which undoubtedly greatly ensures the real-time performance of the detection algorithm.

A few months following the release of YOLOv4, YOLOv5[9] was released in 2020 by Glenn Jocher. It was offered in four versions: YOLOv5s (small), YOLOv5m (medium), YOLOv5l (large), and YOLOv5x (extra large).

The biggest improvement over YOLOv4 is that it provides different version models, which can be flexibly deployed in many scenarios. Meanwhile, YOLOv5 adds additional modules such as PANet and SAM to enhance the feature fusion and sensory field while using CSPDarknet53 as the backbone network framework. These improvements help to improve the model’s perceptual ability and feature representation ability. YOLOv5 also uses more data enhancement techniques, such as Mosaic data enhancement and AutoAugment. These help to improve the generalization ability and robustness of the model.

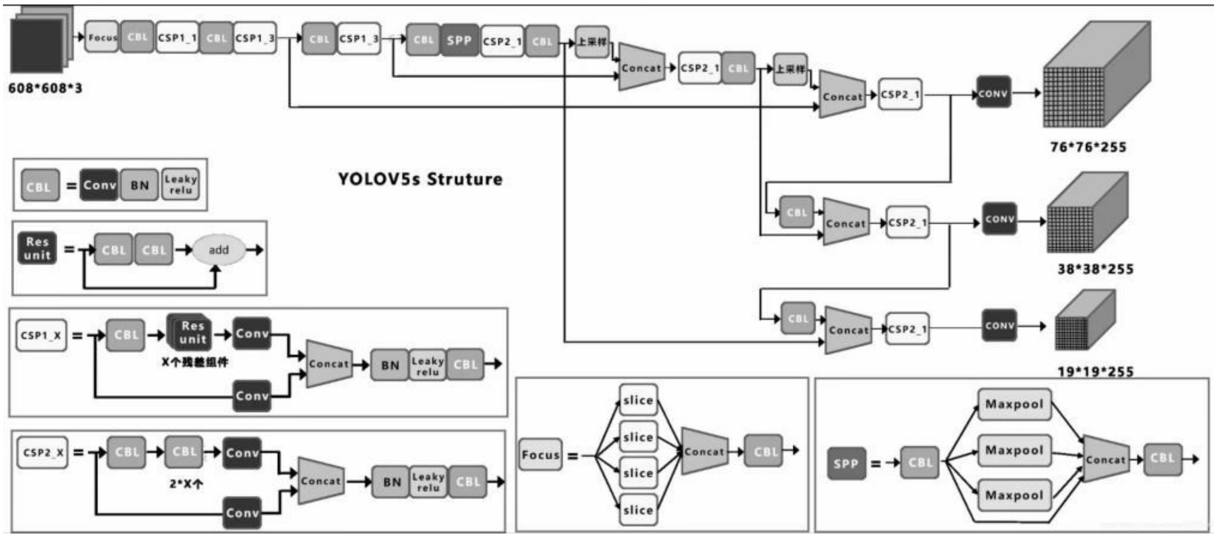


Figure 2. Network Architecture Diagram of Yolov5s.

YOLOv5 offers a range of lightweight models, including YOLOv5s, YOLOv5m, YOLOv5l, and others. Each of these models offers different tradeoffs between speed and accuracy, allowing workers to choose the right model for their specific needs.

In view of the excellent performance of YOLOv5 in target detection and its ability to select different versions of the model for optimal adaptation according to different scenarios. So in the method provided in this paper, YOLOv5 is used as the primary recommended algorithm for a one-stage target detection algorithm.

Two-stage target detection algorithm: (Faster R-CNN and its optimization)

The vehicle target detection algorithm based on the two-stage idea inherits the sliding window idea in the traditional target detection algorithm and divides the whole detection process into two key steps: candidate region extraction and target classification. The advantage of this method is that by selecting multiple candidate frames, the feature information of the vehicle target can be fully captured, which improves the accuracy of the detection results and enables more precise vehicle target localization. However, the division of the detection process into two stages makes the algorithm run slowly.

Faster R-CNN [10] was first proposed in 2015 by Ren et al. from the Microsoft team. It is the fastest of the R-CNN family of algorithms. It introduces a candidate frame extraction module called Region Proposal Network (RPN), which can extract candidate regions on an image with potential detection targets. This structure can share the convolutional layer features with the classifiers, thus avoiding repeated computation. The network structure of Fast R-CNN is shown in Fig. N, which consists of three main parts: CNN, RPN network, and Fast R-CNN detection network. The sharing of convolutional features between the RPN network and the Fast R-CNN detection network makes the Faster R-CNN a unified and independent framework so that each input can be detected in the same way. RPN network and Fast R-CNN detection network share convolutional features to make Faster R-CNN a unified and independent framework, so that each input image can get the labeled frames of different vehicle target locations on the image and the scores of the corresponding categories after Faster R-CNN operation.

Given that the VGG-16 network used in the Faster R-CNN algorithm contains multiple downsamplings, the loss of information about small targets on the road after multiple iterations is serious. At the same time, the anchor size setting in the RPN network does not match the target size, leading to the problem of low accuracy of the generated target area suggestion box. Deng et al. proposed the following solutions in 2021: (1) adding HF (High Frequency) enhancement of the image to the training samples to improve the target localization efficiency of the network; (2) replacing the feature extraction network with a deep residual network with stronger feature extraction capability, and further adjust the anchor point specification and aspect ratio in the RPN to adapt to the lanelet target size; (3)

propose a feature fusion strategy to obtain shallow details and deep semantic information as a way to obtain strong features for roadlet targets.

Given that Deng et al. comprehensively optimized the Faster R-CNN algorithm from three aspects: training samples, feature extraction network, and feature fusion. So in the method provided in this paper, it is used as the primary recommended algorithm for the two-stage target detection algorithm.

4. Conclusion

In this paper, for the task of road multi-target detection, the principles of traditional road target detection methods and optimization are first summarized. Then the deep learning-based road target detection methods are mainly categorized into one-stage target detection algorithms and two-stage target detection algorithms. Finally, the vehicle target detection algorithm under a convolutional neural network is summarized in a timeline. In general, traditional road target detection algorithms are unable to meet the real-time nature of autonomous driving because of the large amount of redundant computation in their process. The one-stage road target detection algorithm does not need to spend a lot of time because of the candidate region, so it has a great advantage in detection speed. However, the accuracy of the two-stage road target detection algorithm is significantly better than the former, and with the rapid development of deep learning technology, more lightweight network structures are also gradually integrated with the algorithmic model, which makes the efficiency of the vehicle target detection task further improved.

In the future, automatic driving technology and intelligent transportation systems will have higher requirements for accuracy and real-time, and the research direction of road target detection can also be optimized continuously around the above four directions to meet higher requirements.

References

[1] Gui Tao, Xi Zhiheng, Zheng Rui, et al. A review of research on the robustness of natural language processing based on deep learning[J/OL]. Journal of Computing:1-26[2023-08-27].http://kns.cnki.net/kcms/detail/11.1826.tp.20230727.0855.002.html

[2] Jiao Hongbin.Research and application of AI artificial intelligence technology in intelligent transportation[J]. China Xinxin,2023,25(12):74-76.

[3] Lv Nengchao, Wang Yugang, Zhou Ying, et al. A review of road traffic safety analysis and evaluation methods[J]. Chinese Journal of Highway,2023,36(04):183-201.DOI:10.19721/j.cnki.1001-7372.2023.04.016.

[4] YANG A-Rong, HUANG Wan-Yue, HOU Ming-Zhe. Review the causal analysis methods of road traffic accidents[J]. Shanghai Highway,2013(04):78-82+12.

[5] Wang Qian. Research on multi-target detection method of road scene based on deep learning[D]. Hebei University of Technology,2022.DOI:10.27105/d.cnki.ghbgu.2022.000702.

[6] Yang T. Research on target detection algorithm based on UAV aerial images[D]. Xi’an University of Technology,2023.DOI:10.27391/d.cnki.gxagu.2023.000802.

[7] Yoav Freund,Schapire R,Abe N.A short introduction to boosting[J].Joumal-Japaneses Society For Artificial Intelligence,1999,14 (5) : 771-780

[8] Dalal N,Triggs B.Histograms of oriented gradients for human detection[C]//2005 IEEE Conference on Computer Vision and Pattern Recognition (CVPR),June 20- 25,2005,San Diego, CA, USA.IEEE, 2005, 1: 886-893.

[9] Han K, Wang Y, Tian Q, et al. Ghostnet: more features from cheap operations[C]//Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. 2020: 1580-1589.

[10] Ren S, He K, Girshick R, et al. Faster R-CNN: towards real-time object detection with region proposal networks[J].IEEE Transactions on Pattern Analysis and Machine Intelligence, 2017, 39(6):1137-1149.