Mathematical Modeling and Algorithm Application
ISSN: 3006-0842 | Vol. 2, No. 1, 2024

Optimization of Unmanned Driving Obstacle Detection Method

Ziyan Li, Xinyi Liu

Wuhan University of Technology, Wuhan, China

Abstract: In order to effectively enhance the accuracy of obstacle detection in unmanned driving on roads, this paper proposes an improved Faster-RCNN object detection model. Diverging from conventional Faster-RCNN models that replace the feature extraction network with ResNet50 instead of VGG16 and deepen the convolutional layers, allowing for a more comprehensive utilization of feature information. The proposed model is trained and tested in comparison with the EfficientNet network on the same dataset, VOC2007. Experimental results indicate that the proposed model exhibits higher precision in detecting obstacles on roads, showcasing broad applicability and achieving effective target recognition.

Keywords: Unmanned driving; Obstacle detection; Faster-RCNN model.

1. Introduction

Intelligent vehicles represent a comprehensive system that integrates functions such as environmental perception, path planning, multi-level vehicle management, utilizing computer technology, advanced sensors, information fusion, wireless communication, artificial intelligence, and automation. This constitutes a highly technological infrastructure. Currently, research on intelligent vehicles focuses on enhancing automotive safety and ride comfort, striving to create a refined human-vehicle interface. In recent years, autonomous intelligent vehicles have become a globally recognized research hotspot in the automotive manufacturing industry, serving as a crucial area for driving industrial innovation and growth. Many developed countries have prioritized the development of related technologies.

With the rapid development of modern high technology, digitization, informatization, and intelligence have permeated every corner of human society's production and life. One day, we will witness the on-road operation of intelligent autonomous vehicles, and this cutting-edge technology will no longer exist solely in the realm of imagination. Various high-tech vehicles currently exhibit significant progress and success in performance, comfort, and safety. In intelligent autonomous vehicles, sensor devices are closely related to the surrounding environment. They are responsible for collecting and organizing various information, swiftly controlling and operating the vehicle system after sending data to highly intelligent computers. Therefore, the potential of functions such as autonomous driving and intelligent control can be fully realized.

As the socio-economic development progresses, the transportation industry is flourishing, and the number of vehicles is soaring. Traffic congestion is becoming increasingly severe, with frequent accidents resulting in inevitable casualties and economic losses. In response to this situation, designing a responsive, highly reliable, and economical collision avoidance and warning system for vehicles is imperative. Ultrasonic collision avoidance is the most common method of distance measurement, applied to short-range, low-speed collision prevention in the front, rear, left, and right of vehicle parking. In addition, it is widely used in the reverse collision warning system for vehicles. Ultrasonic waves, as a special type of sound wave, possess basic physical properties of sound wave transmission, including refraction, reflection, interference, diffraction, and scattering. Ultrasonic collision avoidance utilizes its reflective properties to detect obstacles behind the vehicle when it is in reverse. The ultrasonic distance sensor notifies the driver of the distance and position to the obstacle through indicator lights and a buzzer, ensuring safety.

2. Optimization of Obstacle Detection Methods in Unmanned Driving

In 2020, Chintakindi Balaram Murthy proposed an enhanced YOLOv3+ network aiming to achieve accurate real-time detection of small pedestrians in complex environments. In the proposed network, K-means clustering is applied before training to select the optimal K bounding boxes. The improved YOLOv3+ network introduces a reverse residual module to enhance feature extraction capabilities and refines the loss function to reduce bounding box loss errors. In terms of detection accuracy, this network exhibits stronger robustness, achieving an AP of 79.86%, compared to existing networks. However, there is a slight decrease in detection speed when dealing with smaller pedestrians.

This paper presents a road obstacle detection method based on an improved Faster-RCNN and partitions and trains on VOC2007. A comparison is made with the EfficientNet network, resulting in an enhanced accuracy in road obstacle detection.

2.1. Overall Scheme for Obstacle Detection

Initially, this paper scales an image of size P×Q to M×N. Subsequently, the scaled M×N image is fed into the Backbone network for feature extraction, obtaining a feature map. The chosen Backbone network employs the ResNet50 feature extraction network, where Conv Blocks and Identity Blocks constitute the two fundamental blocks of ResNet50. To alter the network's dimensions, Conv Blocks with disparate input and output dimensions are utilized, preventing continuous concatenation. Simultaneously, the Identity Blocks, with identical input and output dimensions allowing for concatenation, are employed to deepen the network.

The RPN (Region Proposal Networks) layer, upon obtaining the feature map of the image, employs the softmax activation function to categorize the generated anchors. To obtain precise proposals, bounding box regression corrects the generated anchors. The ROI (Region of Interest) Pooling layer has two inputs, namely the proposal layer and the feature map layer. After extracting information using the ROI Pooling layer, the data is fed into a fully connected layer to determine the target category, as illustrated in Figure 1.

2.2. Model Structure

2.2.1. Backbone

In this paper, the Backbone section adopts the ResNet50 network, where "50" indicates the presence of 50 layers distributed across 5 stages: Stage 0, Stage 1, Stage 2, Stage 3, and Stage 4. The Stage 0 assumes an image with channel (C), height (H), and width (W) dimensions represented as (3,224,224). After the first layer operations of CONV, BN (Batch Normalization), and RELU, the data proceeds to the second layer's max-pooling. The convolutional kernel size in the first layer is 7×7, with 64 kernels, and a stride of 2. The MAXPOOL in the second layer has a kernel size of 3×3, a stride of 2, resulting in a final output shape of (64,56,56). This can be understood as having 64 channels, a height of 56, and a width of 56. The two 56 values correspond to halving the original image's height and width twice, as both stages in Stage 0 have a stride of 2, leading to a reduction in input scale twice.

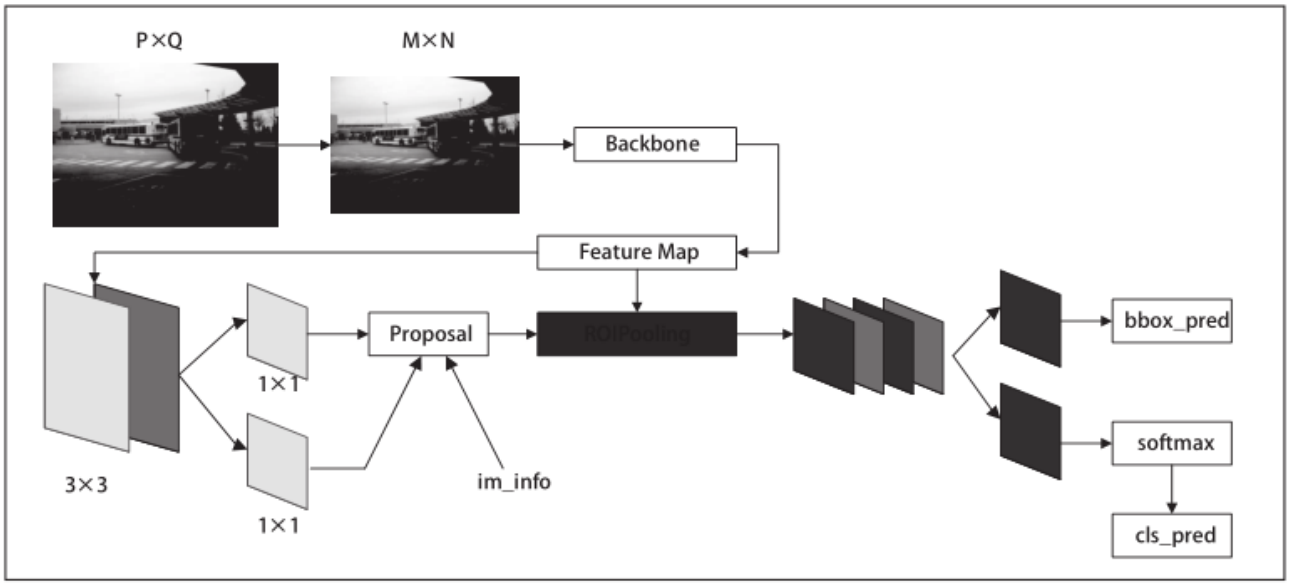

Figure 1. Overall Block Diagram

Stages 1, 2, 3, and 4 consist of two essential types of Bottlenecks: Different Input and Output Channel Numbers (BTNK1) and Equal Input and Output Channel Numbers (BTNK2). Since the input image in BTNK2 is (C,W,W), the two variable parameters in BTNK2 are denoted as C and W.

Assuming the input of the image (C,W,W) is x, the convolutional block on the left side of BTNK2, along with the corresponding activation function denoted as f(x), can be combined as (f(x)+x). This is achieved through a single ReLU activation function, producing the output of BTNK2 with an unchanged shape. Unlike BTNK2 with two variable parameters, BTNK1 has variable parameters C, W, C1, and S4. It also includes an additional convolutional layer denoted as g(x). The distinct input and output channel numbers in BTNK1 cause the convolutional layer to transform x into g(x), resulting in inconsistent input and output dimensions (where g(x) and f(x) have the same channel numbers). The sum of these components yields f(x)+g(x).

Therefore, Stage 1 comprises one BTNK1 layer and two sequential BTNK2 layers, resulting in a shape of (256,56,56). This output is then passed to the next stage, Stage 2, which consists of one BTNK1 layer and three sequential BTNK2 layers, yielding a shape of (512,28,28). This output is transmitted to Stage 3, where one BTNK1 layer and five sequential BTNK2 layers are combined, ultimately resulting in a shape of (1024,14,14). This output is then forwarded to the final stage, Stage 4, which combines one BTNK1 layer and two sequential BTNK2 layers, yielding a shape of (2048,7,7).

2.2.2. Region Proposal Networks (RPN)

In OpenCV, the use of sliding windows and image pyramid methods in AdaBoost can enhance detection discriminative box generation, yet the redundant computations involved consume considerable resources and time. Hence, this paper employs Region Proposal Networks (RPN) to generate detection boxes efficiently.

The RPN network consists of two branches. One branch undergoes softmax function classification for anchors to obtain correct and incorrect classifications. The other branch aggregates the bounding box regression offsets generated by anchors, adjusting them to achieve more accurate proposals. To achieve target localization, the Proposal layer is utilized to aggregate correct anchors, obtaining corresponding offsets and generating proposals. Finally, proposals that are too small or out of bounds are removed.

Anchors are a set of rectangles generated by the RPN, arranged in a 9x4 matrix. The coordinates of the four points of each rectangle are denoted as x1, y1, x2, y2, with the rectangles having aspect ratios of 1:1, 1:2, and 2:1. To rectify the position of the detection box, all Conv layers are iterated to calculate feature maps. All points are then matched with

these 9 initial boxes. The final correction results are obtained through two rounds of bounding box regression.

2.2.3. ROI Pooling

The ROI Pooling layer aggregates and computes proposal feature maps collected from proposals and forwards them to the next layer. ROI Pooling layer has two inputs: the original feature maps and proposal boxes outputted by RPN (of varying sizes). For traditional CNN networks like AlexNet and VGG, the input image size must be fixed after the network is trained, and the network output is also a fixed-size vector or matrix.

Due to the uncertainty of input image sizes and to preserve the original shape information, it is crucial to avoid cropping or warping parts of the image into a specified size, as it may result in the loss of the image's complete structure. The RoI Pooling layer processing steps are as follows: Firstly, using the spatial_scale parameter, the M×N image is adjusted to a feature map scale of (M/16)×(N/16). Secondly, each feature map region should be divided into a w×h grid. Lastly, to ensure a fixed output length, each grid is subjected to max-pooling processing, ensuring that proposals of different sizes yield output results of w×h.

2.2.4. Classification

The Classification section utilizes the acquired proposal feature maps, employing fully connected layers and the softmax function to calculate the specific category for each proposal (e.g., person, car, bicycle), resulting in the cls_prob probability vector. Simultaneously, bounding box regression is employed once again to obtain the position offsets (bbox_pred) for each proposal, facilitating the regression of more precise object detection boxes.

2.3. Model Training

In this study, the VOC2007 dataset is utilized for model training and testing. Initially, the training labels are placed in the Annotation folder, storing the label information for the training set. The images intended for training are stored in the JPEGImages folder. The training set is defined as a ratio of 9:1 for training to validation. As obstacles encountered during vehicle operation exhibit unpredictable, complex, and varied shapes, this study aims to differentiate between various detection categories as much as possible. Consequently, 12 common detection categories are designed, including: bicycle, boat, bottle, bus, car, cat, chair, dog, horse, motorbike, person, and train. These 12 detection categories are stored in txt format and named cls_classes.txt. The input shape size is set to [600,600], and the anchor sizes are set to [8, 16, 32].

Model training is divided into two stages: freezing and unfreezing. During the freezing stage, the backbone network is frozen to maintain unchanged feature extraction. The initial epoch is set to 0, the freezing epoch is set to 50, the freezing batch size is set to 4, and the freezing learning rate is set to 1e-4. During the unfreezing stage, the backbone network is unfrozen to allow feature extraction changes. The unfreezing epoch is set to 100, the unfreezing batch size is set to 2, and the unfreezing learning rate is set to 1e-5.

2.4. Prediction and Results

In this study, AP (Average Precision), mAP (mean Average Precision), and LAMR (log Average Miss Rate) are utilized as the primary evaluation metrics. The Faster-RCNN and EfficientDet networks, designed and trained in this study, are assessed using these metrics. Appropriate weight files are selected, and the models are evaluated on their ability to predict obstacle detection scenarios in road images, as shown in Figures 2 and 3. Based on the detection accuracy observed in Figures 2 and 3, the predictions from this study outperform the alternatives.

The study conducts an AP analysis for the 12 common obstacles on the road, as depicted in Figure 4. Additionally, an analysis of LAMR for the 12 common obstacles is presented in Figure 5. Table 1 compiles the mAP values for different models on the VOC2007 dataset.

Figure 2. EfficientDet Prediction Results

Figure 3. Prediction Results in This Study

Table 1. Test Results on VOC2007 trainval

	Backbone	Anchor boxes	mAP/%(VOC2007trainval)
Faster-RCNN	VGG16	12	68.2
Efficientdet	Efficientnet	9	69.83
this paper	Resnet50	9	80.35

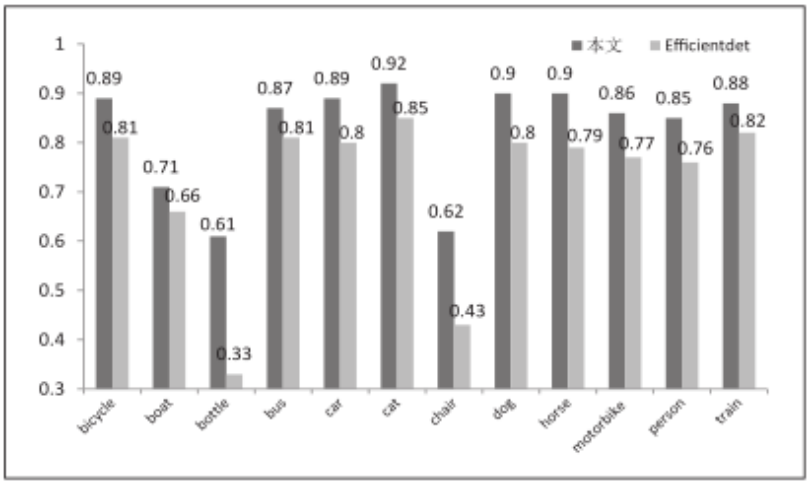

Figure 3. AP for 12 Common Road Obstacle Classes in Faster-RCNN and EfficientDet Networks

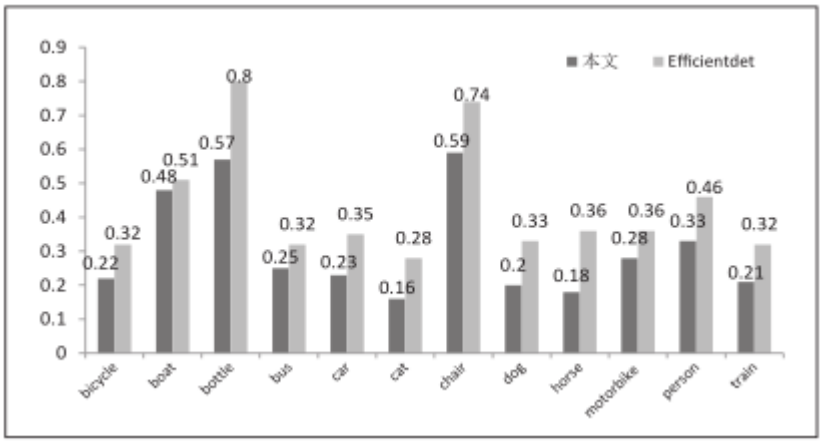

Figure 4. LAMR Values for 12 Common Road Obstacle Classes in Faster-RCNN and EfficientDet Networks

From Fig.4, it can be observed that the Faster-RCNN designed in this study outperforms EfficientDet significantly in detection accuracy. Moreover, it exhibits superior performance in detecting pedestrians, bottles, buses, cars, motorcycles, and bicycles. Figure 5 reveals that the log values of the average miss rate (LAMR) for Faster-RCNN are consistently lower than those for EfficientDet. Table 1 provides a comparison of the mAP values for the three different networks. The network designed in this study shows a 12.15% improvement in average precision compared to Faster-RCNN using VGG16 as the feature extraction network with 12 anchor boxes. Similarly, compared to EfficientDet using EfficientNet as the feature extraction network with 9 anchor boxes, there is a 10.52% improvement.

3. Conclusion

In order to effectively enhance obstacle detection accuracy on roads, this study proposes an improved Faster-RCNN target detection model. Diverging from the conventional approach of replacing the VGG16 feature extraction network with ResNet50 and deepening the convolutional layers, the proposed model maximizes the utilization of feature information. The model is trained and tested against the EfficientNet network on the VOC2007 dataset for comparison. The experiments indicate that the proposed model, compared to the traditional Faster-RCNN model, enhances average detection accuracy by 12.15%, demonstrating favorable metrics in the log values of average miss rate. Furthermore, the model designed in this study exhibits superior accuracy in detecting bicycles, buses, cars, cats, dogs, and pedestrians. This suggests that the proposed model has good applicability.

Acknowledgment

In the process of completing this paper, I experienced the joy of successful experiments and the frustration of data not yielding results, which taught me valuable lessons beyond the realm of academic knowledge, enriching my undergraduate journey.

First and foremost, I would like to express my gratitude to the fellow students who collaborated with me. Their unconditional support and encouragement during the project were unwavering, even in challenging times.

Secondly, I extend my thanks to the National Undergraduate Innovation and Entrepreneurship Training Program for funding the S202310497265 project. This support provided me with the platform and financial resources necessary for the successful completion of the experiments.

Lastly, I appreciate the reviewing professors who took the time from their busy schedules to review my paper. Your patient review and guidance have contributed to my progress and improvement.

References

[1] Zhao, X.Research on Path Optimization of Unmanned Agricultural Vehicles Based on Convolutional Neural Network. Agricultural Mechanization Research, 2024,46(07), 257-261.

[2] Li, X., Li, X., & Li, L.Development and Application of Environmental Information Processing in Autonomous Driving. Automation Review, 2023,40(12), 26-30.

[3] Jiao, Y., Su, C., & Huang, S.Research on Active Obstacle Detection System for Rail Transit Vehicles. Smart Rail Transit, 2023,60(06), 12-15.

[4] Yang, T., Guo, Y., Wang, S., & Ma, X.Obstacle Recognition of Unmanned Driving Trolley Locomotives in Underground Coal Mines. Journal of Zhejiang University (Engineering Science), 2024,58(01), 29-39.

[5] Hu, J. (2023). Simulation and Performance Analysis of Millimeter-Wave Radar under Unmanned Driving Conditions. Smart Rail Transit, 60(05), 6-11.

[6] Wu, H., & Cheng, Q.Research on Steering and Braking Obstacle Avoidance Control Strategy for Unmanned New Energy Vehicles. Automotive Testing Report, 2023,(12), 76-78.

[7] Yang, R.Research on Key Technologies of Unmanned Driving Based on ROS System. Dissertation, Xi'an University of Architecture and Technology,2023.

[8] Chen, Q.Research on Path Tracking Control of Unmanned Vehicles with Local Planning. Dissertation, Northeast Petroleum University,2023.

[9] Shen, B.Trajectory Planning and Obstacle Avoidance Design for Unmanned Vehicles Based on Model Predictive Control. Dissertation, North China University of Technology,2023.

[10] Tong, J.Research on Obstacle Detection Technology for Mining Electric Locomotives Based on Improved Instance Segmentation. Dissertation, Anhui University of Science and Technology,2023.