

A Vehicle Detection and Alert System Based on YOLOv4: Design, Implementation, and Performance Evaluation

Jia Zhou

Jincheng College, School of Computer and Software, Sichuan University, Chengdu 611731., Sichuan

Abstract: *As living standards continue to rise, transportation systems have become increasingly convenient; however, associated safety risks have also escalated significantly. Insufficient pedestrian safety awareness has contributed to a rising incidence of traffic accidents. In response, this study leverages the relative maturity of object detection technologies by retraining the YOLOv4 algorithm using the VOC dataset to design and implement a vehicle alert system tailored for unsignalized intersections. The proposed system aims to mitigate collision risks under low-light and poor nighttime visibility conditions by detecting approaching vehicles and providing timely alerts to support safer pedestrian crossing decisions. Evaluation of the model's performance demonstrates its effectiveness and practical applicability in real-world traffic environments.*

Keywords: Target detection; YOLOv4 algorithm; Pedestrian safety; Deep learning.

1. INTRODUCTION

With the improvement of people's living standards, the choice of transportation modes is becoming more and more diverse, and there are more and more cars on the roads and roads, which is accompanied by increasing safety risks, according to China's statistics yearbook in 2014. The number of traffic accidents in 2018 was 196812187781, 212846203 049 and 244937, and the overall traffic accidents showed a clear upward trend, and the traffic accidents reached the highest value in all years. City highway occupies an important position in the highway system of our country, is also the high incidence of traffic accidents area [1]. So in the face of this series of tragedies, we are constantly reminded of the importance of improving the safety of pedestrian crossing.

With the popularity of deep learning in recent years, target detection algorithms have adopted many improvements in recent years. At present, the mainstream algorithms include two categories. One is R_CNN [2]. Generate possible regions for representation and perform CNN extraction features and place them in a classifier and correct their location, Another type of regression of predicted objects is represented by the Yolo algorithm, which focuses on high accuracy and the latter on speed, which in this case can basically achieve rapid detection speed and high recall rate with several generations of updates.

2. FACE DETECTION YOLO ALGORITHM

2.1 Overview of the yolo algorithm

yolo algorithm is Redmon et al [3]. The first single-target detection method algorithm proposed in 2015, by sliding window to get different sizes of detection window, and then using non-maximum suppression algorithm [4] (Non-Maximum Suppression (NMS)) method gets a target detection box and a classification, Yolo divides the entire picture into $S * S$ grids, then determines the location of the central point of the detection target, writes the boxes value to the corresponding grid, and gets a three-dimensional dimension as an output result. Yolo algorithm is widely used in industrial practice because of its fast detection speed and high detection accuracy.

2.2 Improvements to yolov4

2.2.1 Network structure improvement

yolov4 [5] CSPD arknet53 is based on Darknet 53 [6]. On the basis of this, the backbone structure is modified according to the CSPNet algorithm, and the basic implementation idea is as shown in Figure 1. Send the feature map copied from the base layer to the next layer, This separates the feature map of the underlying layer and avoids

the problem of signal loss due to the network layer being too deep, and reuse features can reduce the computational amount, which can improve the accuracy of the model while reducing the number of computers.

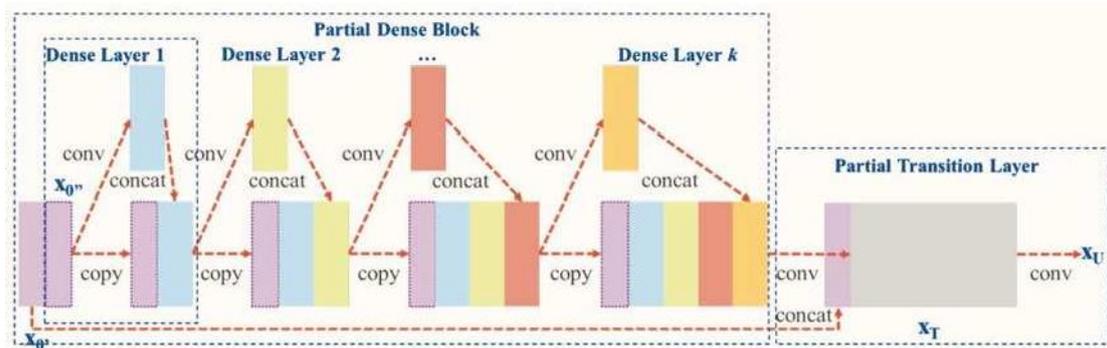


Figure 1: CSPDarknet53 Construction

2.2.2 Feature Extraction Improvement

The SPP module mainly increases the acceptance range of features by adding the maximum pooling layer, and the performance can be greatly improved by adding a small amount of computation cost. FPN is also known as feature pyramid. The FPN module of the original model transfers the strong semantic features of the target through top-down upsampling, but it transfers less location signals to the target. Aiming at the positioning problem of the original model, the new model adds a bottom-up pyramid PAN module to the back of the FPN module [7]. The structure was obtained, which not only solved the scale problem in the detection but also improved the robustness of the network. After SPPNet and PANet processing, the model outputs feature dimension sizes (13, 375), (26, 26, 75) and (52, 275), and features contain more detail and semantic information, which allows the feature processing layer to better utilize these features for classification and regression [8].

2.2.3 Detection box improvement

At the beginning, IOU_Loss [9] calculated the overlap area between the detection box and the target box. GIOU_Loss solved the problem of non overlapping IOU bounding boxes, while DIOU_Loss considered the distance information between the center points of the bounding boxes on the basis of IOU and GIOU; CIOU_Loss incorporates a scale information consideration factor for the aspect ratio of bounding boxes based on DIOU, resulting in formula 1. By improving the non maximum suppression algorithm, it has also achieved high detection of dense objects.

$$CIOU_Loss = 1 - CIOU = 1 - \left(IOU - \frac{Distance_2^2}{Distance_C^2} - \frac{v^2}{(1-IOU)+v} \right) \quad (1)$$

3. EXPERIMENTS AND RESULTS ANALYSIS

3.1 Experimental environment

This experiment is developed in Baidu's AiStudio platform, using the platform to provide GPU computing power, equivalent to Tesla V00, CPU: 4Cores RAM 32GB.

3.2 System implementation process

The system, as shown in Figure 2 below, consists of two parts, including an indicator light to inform pedestrians of the presence of a car near them, and a detection algorithm embedded in an infrared electronic camera. The infrared camera detects vehicle information on the road in real time and returns to the indicator light, and pedestrians can obtain the status of vehicles at intersections through the indicator light when vehicles with poor lighting conditions at night. If the indicator lights are turned red to prove that a car is approaching, and if the indicator lights turn green to prove that the car-free intersection is safe, thereby protecting the safety of pedestrians and vehicles.

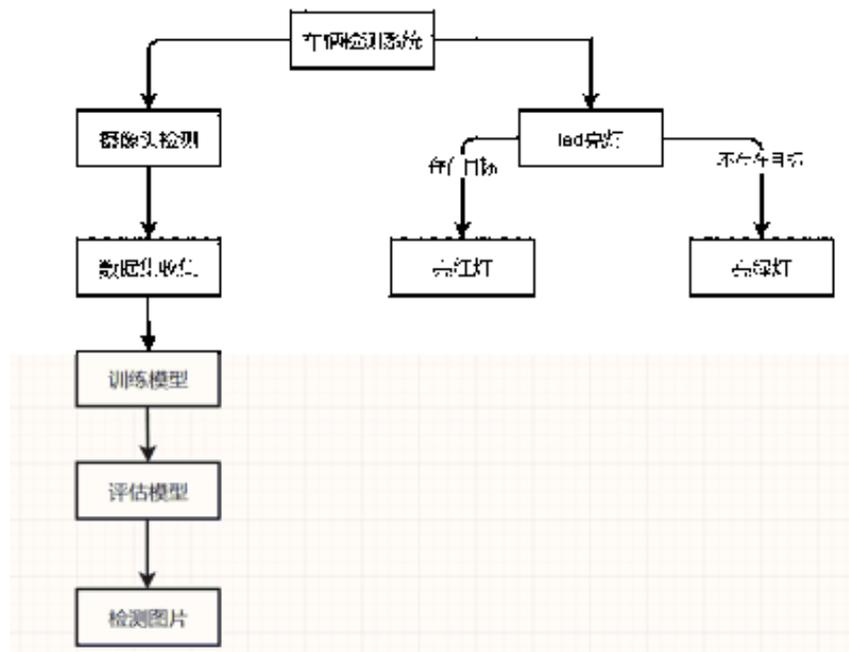


Figure 2: System flow diagram

3.3 Training set preparation

This experiment uses the VOC vehicle dataset on Baidu AiStudio, which contains a total of 60,000 images. Most of the images are taken from roadway monitoring, and most of the graphics consist mainly of multiple classes, as shown in Figure 3 below. Yolov4 provides different size of the anchor target detection, this experiment is mainly aimed at small target detection, the original model uses 20 categories, In this experiment, the classes car, bus, van, and others are used only for four classes, so after data cleanup, only the location and category of the target box for the data of car, bus and van in the training picture are extracted for training, and the other classes are others. The tag file contains the class names and location information for all the categories in the original model that appear in the photos because of the data set used in this article, For data mostly based on sunny days, data collection is more difficult at night, so the application is less accurate for night recognition. Some leakage problems may occur. The light conditions brought by different weather will have an impact on the real-time detection, which will lead to different detection accuracy and affect the detection results. Therefore, this paper is based on the above conditions chosen for having a good nighttime lighting system when the problem is selected to install intersections.



Figure 3: Data Set

3.4 Training Models

The total amount of data is 60,000 images, divided by a ratio of 9: 1 into training sets and test sets, and their paths are saved to the training text and test text respectively, but the Yolov4 algorithm performs some data enhancement and data processing on the data. For example, to increase the data volume, the training used a learning rate of 0.001, by obtaining the position of each photo in the picture for the corresponding category by comparing the cars, buses, and vans in the pictures of VOC vehicle data. Then re-train the model on top of the original Yolov4 model through the image path in the text to get the model weight, evaluate the model's strength through the model weight and finally evaluate the image to be detected through the model.

3.5 Apply models

The intersection monitoring video every 7 frames to save a picture, in the main function call video2frame (path0, path1,7) function, respectively, into the video where the path, the path to save the picture and sub-frame interval of three parameters, the video will be sub-frame, to facilitate the subsequent function of the processing[10]The resulting picture will be detected with the model, by judging the detection result, and returning the corresponding result to the LED warning light, if there is one, the corresponding LED light will light red, if there are no, the LED light will be green. This can indirectly accomplish the purpose of real-time monitoring.

3.6 Model effect display and analysis

In order to verify the accuracy and feasibility of the model, this paper analyzed the model's MAP function as shown in Figure 4 below, which can be seen that when the entire experiment is trained to the 30th round, the MAP accuracy reaches about 70%. The MAP generally showed an upward trend from the first round, and the loss value of the model remained around 0.94 after the fifth round, which can be deduced that the overall model tends to contract, and the subsequent contracting rate is slower, indicating that the model has good learning ability for vehicle data.

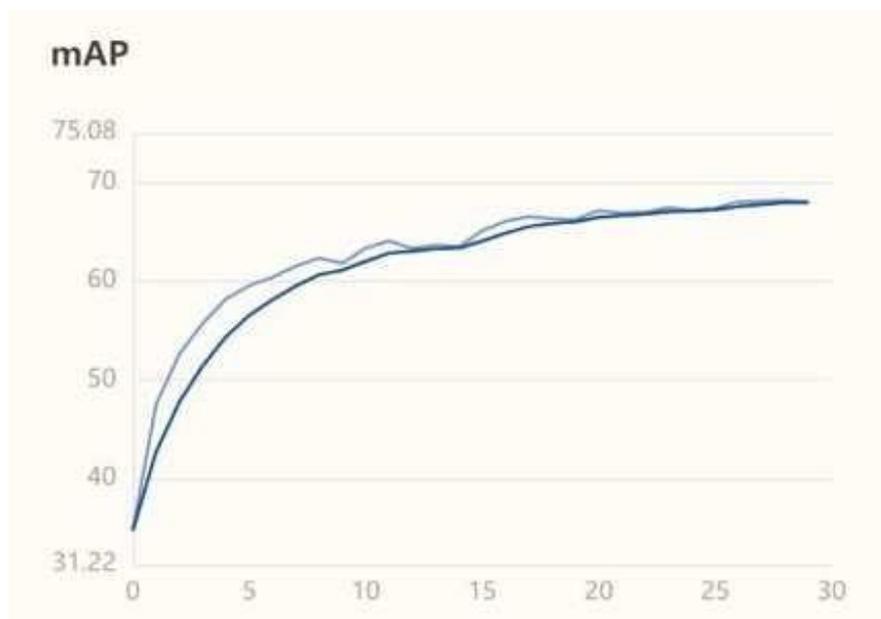


Figure 4: MAP map

Through the detection of the actual picture, the detection results in figure 5 can be seen that the model can detect all the vehicles, indicating that the detection effect of the model is still good. At the time of actual testing of the effect, the model can achieve detection accuracy of.: > 90%, Error rate: <5%, The error rate here is a vehicle type identification error, but in actual use we only need to detect whether there is a car, so it will not have an impact on actual application.



Figure 5: Actual testing diagram

Although the model has a good effect in detection, the data collection is insufficient and the datasets for multiple scenarios are missing, which makes the model accuracy not very high. At the same time, too many factors are considered in real-world conditions, such as pedestrian and vehicle travel routes, and the speed of the vehicle and the speed at which the camera is captured will have a huge impact on the system, which is where we will continue to improve the system.

4. CONCLUSION

This article obtained its own model by making a series of tweaks on the basis of Yolov 4. Some of the small tweaking techniques of Yolov 4 made the data set richer, and the introduction of CSPDarknet 53 did not increase the computation significantly, and the training speed was still fast compared to similar networks. By using VOC vehicle data sets to train classes to re-select classes on the basis of Yolov4's model, we can further obtain model effects compared to some other target detection models nowadays. Yolov4 is better than other models in vehicle recognition accuracy and recognition speed. With the improvement of people's living standards, people will pay more and more attention to road safety, so the accuracy of this model can be of great significance in practical application.

REFERENCES

- [1] Fang Su, Qin Yuan. Analysis of Association Rules of Urban Highway Traffic Accidents [J]. Journal of Wuhan University of Technology (Information and Management Engineering Edition), 2020, 42 (04) : 313-318 + 331.
- [2] GIRSHICK R, DONAHUE J, DARRELLT, et al. Rich feature hierarchies for accurate object detection and semantic segmentation [J]. Computer Science, 2013: 580- 587.
- [3] Redmon J, Divvala S, Girshick R, et al. You only look once: Unified, real-time object detection [C] // Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. New York: IEEE, 2016: 779-788.
- [4] Neubeck A, Gool L. Efficient Non- Maximum Suppression [C] // International Conference on Pattern Recognition. IEEE Computer Society, 2006.
- [5] Alexey Bochkovskiy, Chien-Yao Wang, Hong-Yuan Mark Liao. YOLOv4: Optimal Speed and Accuracy of Object Detection [C] // CVPR, Apr, 2020.
- [6] Wang H B, Zhang Z D. Text detection algorithm based on improved YOLOv3 [C]. Beijing: IEEE the Ninth International Conference on Electronics Information and Emergency Communication, 2019.

- [7] ZHENG Z,WANG P, LIU W,et al. Distance-IoU loss:faster and better learning for bounding box regression [C] // AAAI,2020:12993-13000.]
- [8] Shu Liu, Lu Qi, Haifang Qin, etal. Path aggregation network for instance segmentation [C] // Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition,2018:8759-8768.
- [9] Wei Gao, Chen Zhou, Moufa Guo. Research on insulator defect identification based on improved YOLOv4 and SR-GAN [J / OL].1-11 (2021-05-19 08: 34).
- [10] RAH MAN MA, WANG Y.Optimizing intersectionover-union in deep neural networks for image segmentation [J]. International Symposium on Visual Computing 2016.
- [11] IAN Goodfellow, YOSHUA Bengio,AARON Courville. Deep learning [M]. Beijing: People's Posts and Telecommunications Publishing House, 2017.