Object Detection for Traffic Management Based on YOLO

Xuebin Hong, Jubin Huang, Weiwei Zhao, Huiwen Zou, Zhe Lin*, Yuecheng Chen
College of Electronic Information, Shantou Polytechnic, Shantou 515078, China
*Corresponding author: gd1392@126.com

ABSTRACT

This paper presents an object detection method for traffic management based on the YOLOV7 model, using the bdd100k dataset for experimentation. The results show that the proposed method has good detection performance in traffic scenes. The main contribution of this paper is to improve the accuracy and efficiency of object detection in traffic scenes by applying the YOLOV7 model to the field of traffic management. The research results of this paper are of great significance for the optimization and improvement of traffic management systems. Future research can explore YOLOV7's performance on other target categories and consider algorithm optimizations to improve accuracy on new datasets.

Keywords: Object Detection, traffic management, YOLOV7

1. INTRODUCTION

Object detection for traffic management is the process of identifying and locating objects of interest in traffic scenes, such as vehicles, pedestrians, and traffic signs. Accurate and efficient object detection is a crucial component of many traffic management systems, including intelligent transportation systems, traffic monitoring systems, and autonomous vehicles. Object detection can provide important information for traffic analysis, congestion management, and safety improvement.

Accurate and efficient object detection is essential for effective traffic management systems. Object detection can provide real-time information about the location, speed, and direction of vehicles and pedestrians, which can be used to optimize traffic flow, reduce congestion, and improve safety. For example, traffic management systems can use object detection to detect accidents, identify traffic hotspots, and monitor traffic violations. Object detection can also be used in autonomous vehicles to enable them to navigate safely and avoid collisions. Therefore, accurate and efficient object detection is critical for the development of advanced traffic management systems and autonomous vehicles.

The YOLOV7 model is a state-of-the-art object detection model that uses a single convolutional neural network (CNN) to detect objects in an image. It is known for its high accuracy and fast processing speed, making it a popular choice for real-time applications. The model uses a grid-based approach to divide the image into smaller regions and predicts the bounding boxes and class probabilities for each region. This approach allows the model to detect multiple objects in a single pass, making it more efficient than other object detection models.

The bdd100k dataset is a large-scale dataset of diverse driving scenarios, containing over 100,000 images with object annotations. The dataset covers a wide range of weather conditions, lighting conditions, and traffic densities, making it ideal for training object detection models for traffic management applications. The dataset includes annotations for various object classes, including vehicles, pedestrians, and traffic signs. The dataset has been widely used in research on autonomous driving and traffic management systems.

2. RELATED WORK

After years of development, the target detection algorithm is mainly divided into two schools: the two schools represented by the R-CNN series (R-CNN[1], Fast-RCNN[2], Cascade R-CNN[3]) Two-stage target detection algorithm and one-stage target detection algorithm represented by YOLO series[4][5].

The two-stage target detection algorithm separates the two tasks of target positioning and classification. First, it locates the position of the object in the image. Usually, (x, y, w, h) is used to represent the object position box, where x, y is the center point of the target, and w and h are the length and width of the target object frame. After the target positioning is completed, the target category is determined by extracting features from the target area.

International Conference on Smart Transportation and City Engineering (STCE 2023),
edited by Miroslava Mikusova, Proc. of SPIE Vol. 13018, 130180M
 · 0277-786X · doi: 10.1117/12.3024069

Taking the classic R-CNN as an example, the algorithm first extracts 1k-2k candidate frames through the selective search method[6], then uses CNN to extract features for the candidate areas, and then uses SVM to classify the extracted features, and finally uses regression device, refine the positioning coordinates. The improved version of Fast-RCNN greatly optimizes the detection speed on the basis of R-CNN, while maintaining high accuracy. In addition to the two algorithms introduced above, the two-stage target detection algorithm has many other network models, such as SPP-Net[7], R-FCN[8] and various R-CNN variants[9][10]. But no matter what kind of algorithm it is, its process is to separate target frame detection and object classification into two tasks.

Different from the two-stage target detection, the one-stage target detection algorithm combines the target frame detection and object classification into one, which greatly improves the detection speed. Because of its speed advantage, the one-stage target detection is widely used in real-time image detection and video object tracking[11][12].

Among the many one-stage detection algorithms, the YOLO series is one of the most famous algorithms. Since its birth, the YOLO series has gone through 8 versions of model iterations, and its accuracy and detection speed have been greatly improved. With YOLOv7[13] as an example, the image is firstly preprocessed, and then the model is trained through the CNN network, and finally the object position, confidence and category are directly regressed. It can be seen that its process is much simpler than the two-stage target detection method represented by the R-CNN series, so its detection speed has been greatly improved, and it also maintains a high accuracy.

By comprehensively comparing these two types of target detection algorithms, we can see that the one-stage target detection algorithm has a higher detection rate while maintaining high accuracy, and has broad prospects in practical applications, especially real-time detection. Therefore, after comprehensive consideration, we chose the YOLOv7 algorithm as our experimental model. By using other public data sets, we can verify whether the YOLOv7 algorithm still has good generalization ability under the advantages of both accuracy and speed.

3. METHODS

YOLOv7 adopts a brand-new network structure, which divides the target detection task into two stages, namely object detection and object segmentation. In the object detection stage, YOLOv7 adopts a new convolutional neural network architecture called Multi-input Residual Attention Network (MiRANet), which can process multiple inputs at the same time, thereby improving the detection speed and accuracy. In the object segmentation stage, YOLOv7 uses a segmentation network called DeeplabV3+, which can finely segment images, thereby improving target recognition and positioning accuracy.The structure of YOLOv7 is shown in Figure 1.

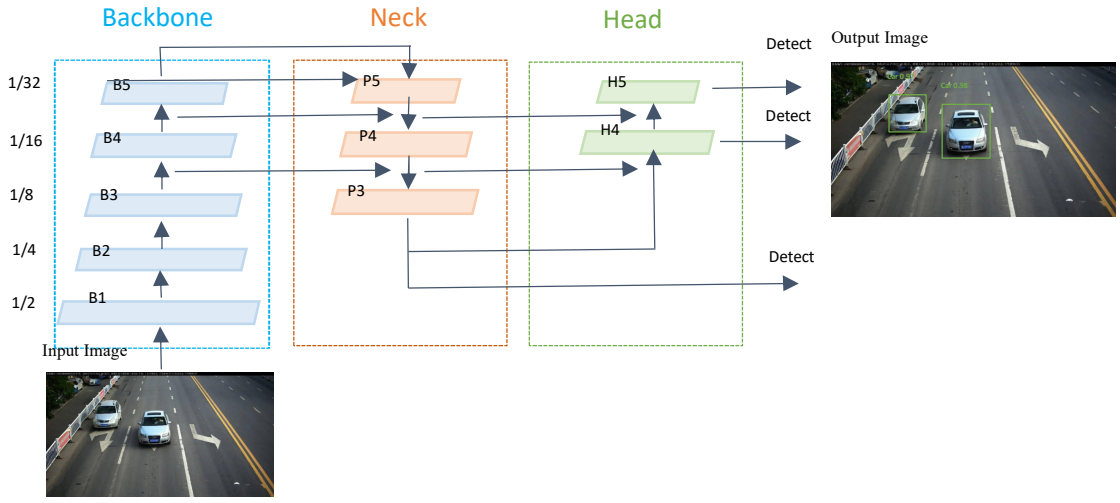

Figure 1. Structures Framework of the YOLOV7 model

According to Figure 1, the input image is first resized to 640x640 before being fed into the backbone network. The neck and head layers of the network then output three layers of feature maps with varying sizes. The Rep and conv layers are subsequently utilized to generate prediction results. For example, in the case of bdd100k, the output comprises 10 classes,

and each output tuple (x, y, w, h, o) represents the coordinate position and front/background information. The number 3 denotes the quantity of anchor boxes, resulting in each layer's output being (10+5)x3=255, which is then multiplied by the size of the feature map to yield the final output.

The entire model architecture exhibits a three-level pyramid structure. Colored blocks represent the output data, and arrows indicate their transformation through multiple conv layers. The leftmost blue pyramid structure represents the first pyramid processing the input image. As we progress from B1 to B5, they become smaller feature maps, with an increase in the number of nodes (channel count). The subsequent orange pyramid structure is an upsampled pyramid. Decreasing the index from P5 to P3 enlarges the images into feature maps while incorporating intermediate feature maps from B4, B3, and other blue pyramids along the way. Finally, the green pyramid structure further downsizes the feature maps, creating H4 to H5. Ultimately, detections are performed at three different resolutions using the feature maps from P3, H4, and H5. Due to YOLOv7's use of anchor boxes, three anchor boxes are assigned for each resolution's feature map.

YOLOv7 has achieved improved speed and accuracy through several architectural refinements. Similar to Scaled YOLOv4, YOLOv7's backbone does not utilize pre-trained ImageNet weights. The key innovations introduced in YOLOv7 are as follows:Firstly, A new method for re-parameterized models was designed. It was discovered that the identity connection in RepConv disrupts the residual connections in ResNet and the concatenation in DenseNet. To address this, the network architecture incorporates RepConv with no identity connections (RepConvN).Secondly, A novel dynamic label assignment strategy was proposed, leveraging hierarchical deep supervision and dynamic label assignment to enhance feature learning ability (coarse-to-fine lead guided label assignment).Thirdly, "Extend" and "compound scaling" methods were introduced, effectively utilizing parameters and memory usage. With the improvement of ELAN (Efficient Layer-wise Attention Network), Extend-ELAN (E-ELAN) was proposed, which continuously enhances the network learning ability without breaking the original gradient path. The parameters and computation of the state-of-the-art real-time object detection model were effectively reduced by 40% while achieving faster inference speed and higher detection accuracy.

4. EXPERIMENT RESULTS AND DISCUSSION

4.1 Experiment Environment

This study was conducted in an experimental environment that utilized the Ubuntu 18.04 operating system. The NVIDIA V100 graphics cards were used for computing and model training, while the deep learning framework was accelerated using CUDA 11.2. The primary deep learning framework utilized in this study was PyTorch (version 1.9.1).

4.2 Evaluation Index

To measure the performance and effectiveness of a defect detection model, common evaluation indicators include accuracy rate, recall rate, precision rate, and F1 score. However, in this paper, the evaluation will be based on mAP@0.5 and mAP@0.5:0.95. The former represents the average precision of the model when the intersection over union (IOU) is 0.5, while the latter represents the average precision when the IOU ranges from 0.5 to 0.95 in increments of 0.05. To calculate mAP@0.5:0.95, the accuracy values obtained at each IOU value are averaged. The formulas for all evaluation indicators are provided below.

$$P = \frac{T_p}{T_p + F_p} * 100\% \tag{1}$$

$$R = \frac{T_p}{T_p + F_N} * 100\% \tag{2}$$

$$AP = \int_0^1 p(r)dr \tag{3}$$

$$mAP = \frac{1}{n}\sum_{i=0}^{n} AP_i \tag{4}$$

4.3 Results and Discussions

The steps of the whole experiment process are as follows:Firstly, Write a script to convert the bdd100k images image set from voc format to YOLO format, and adjust the configuration of YOLOv7 to adapt to the new data set in terms of category and format.Secondly, Write a script to do random sampling, randomly select 1000 pieces of data in the training

set as the test set (the reason for this is that the test set part of this data set does not contain labels, and there is no human marking), the rest remains unchanged, and finally train The data volumes of the three data sets of , val and test are 69000, 10000 and 1000 respectively.Thirdly, Based on the pre-trained YOLOv7 model, fine-tuning was performed, and experimental results were obtained. The main parameters used in the experiment were as follows: epoch set to 100, batch size set to 128, and resize image set to 640*640. The remaining parameters were kept as default.

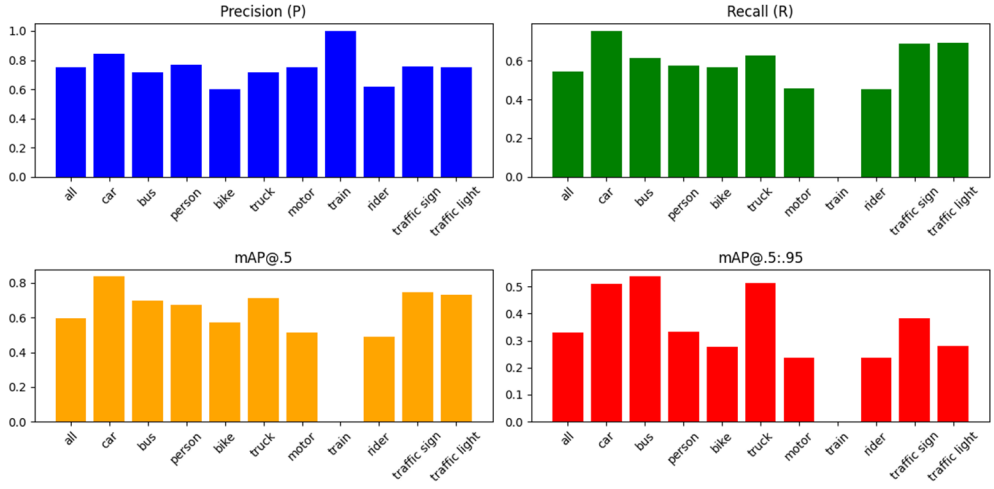

Figure 2. Precision, Recall, mAP test results for each category

According to Figure 2, it can be observed that the model performs well in certain categories, such as car, bus, person, and traffic sign. However, it shows relatively poor performance in other categories, such as train and motor.

In the end, we got the following experimental results. On the new data set, the map@.5 index is close to 60%, which is about 10 points lower than the coco data set, and the map@.5-.95 index is about 33.1%, which is about 33.1% compared with coco dataset is about 18 points lower.

Through the above experimental results, we found that although the accuracy of the YOLOv7 model has declined on the new data set, it still maintains a good effect overall, especially in vehicle target detection (cars, trucks, buses, etc.), Both mAP@.5 and mAP@.5:.95 have achieved significantly better results than other targets, indicating that the generalization ability of the YOLOv7 model on such targets is more significant.

5. CONCLUSIONS

The main findings of the study are that the proposed object detection method for traffic management based on the YOLOV7 model and the bdd100k dataset achieved good detection performance in traffic scenes. The proposed method improved the accuracy and efficiency of object detection in traffic scenes, making it a promising approach for optimizing and improving traffic management systems. The study showed that the YOLOV7 model is a suitable choice for object detection in traffic scenes and that the bdd100k dataset is a valuable resource for training and evaluating object detection models for traffic management applications. Overall, the study demonstrated the potential of using advanced object detection techniques for improving traffic management systems.

ACKNOWLEDGMENT

The authors gratefully acknowledge the Foundation for Scientific Research Projects of Universities in Guangdong Province (No.2021KTSCX287, No. 2021KQNCX215, No.2022KQNCX239).

REFERENCES

[1] Girshick, R., Donahue, J., Darrell, T., & Malik, J. (2014). Rich feature hierarchies for accurate object detection and semantic segmentation. In Proceedings of the IEEE conference on computer vision and pattern recognition (pp. 580-587).

[2] Girshick, R. (2015). Fast r-cnn. In Proceedings of the IEEE international conference on computer vision (pp. 1440-1448).

[3] Cai, Z., & Vasconcelos, N. (2018). Cascade r-cnn: Delving into high quality object detection. In Proceedings of the IEEE conference on computer vision and pattern recognition (pp. 6154-6162).

[4] Redmon, J., Divvala, S., Girshick, R., & Farhadi, A. (2016). You only look once: Unified, real-time object detection. In Proceedings of the IEEE conference on computer vision and pattern recognition (pp. 779-788).

[5] Redmon, J., & Farhadi, A. (2018). Yolov3: An incremental improvement. arXiv preprint arXiv:1804.02767.

[6] Uijlings, J. R., Van De Sande, K. E., Gevers, T., & Smeulders, A. W. (2013). Selective search for object recognition. International journal of computer vision, 104, 154-171.

[7] He, K. , Zhang, X. , Ren, S. , & Sun, J. . (2014). Spatial pyramid pooling in deep convolutional networks for visual recognition. IEEE Transactions on Pattern Analysis & Machine Intelligence, 37(9), 1904-16.

[8] Dai, J., Li, Y., He, K., & Sun, J. (2016). R-fcn: Object detection via region-based fully convolutional networks. Advances in neural information processing systems, 29.

[9] Chen, M. , Yu, L. , Zhi, C. , Sun, R. , Zhu, S. , & Gao, Z. , et al. (2022). Improved faster r-cnn for fabric defect detection based on gabor filter with genetic algorithm optimization. Computers in Industry, 134, 103551-.

[10] Uyar, K. , Tademir, A. , Lker, E. , M Ztürk, & Kasap, H. . (2021). Multi-class brain normality and abnormality diagnosis using modified faster r-cnn. International Journal of Medical Informatics, 155(11), 104576.

[11] Mokhtari, M. A., & Taheri, M. (2022). Real-time object detection and tracking using YOLOv3 network by quadcopter. Mechanics Based Design of Structures and Machines, 1-19.

[12] Pandiyan, P., Thangaraj, R., Subramanian, M., Rahul, R., Nishanth, M., & Palanisamy, I. (2022). Real-time monitoring of social distancing with person marking and tracking system using YOLO V3 model. International Journal of Sensor Networks, 38(3), 154-165.

[13] Wang, C. Y., Bochkovskiy, A., & Liao, H. Y. M. (2023). YOLOv7: Trainable bag-of-freebies sets new state-of-the-art for real-time object detectors. In Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (pp. 7464-7475).