Computer Life (CPL) ISSN: 1819-4818
DELIVERING QUALITY SCIENCE TO THE WORLD

Traffic Violations Detection Based on Improved Faster R-CNN

Lei Yu

Chongqing Jiaotong University College of traffic and transportation, China

---

Abstract: Aiming at the frequent passenger missed detection problems caused by factors such as complex traffic scenes, small head size, and overlapping passengers, an improved Faster R-CNN detection model is designed. This model uses GoogLeNet as the feature extraction network, uses the convolution kernels of different scales in GoogLeNet to extract the features under the image of the breakthrough receptive field, and uses the CA module to re-adjust the importance of the feature maps of different channels, so that the model is The target is more sensitive. Aiming at the missed detection problem caused by overlapping passengers, the model uses a soft non-maximum suppression algorithm to replace the non-maximum suppression algorithm in the original Faster R-CNN, and defines the loss function as the weighted sum of the positioning loss and the confidence loss. The experimental results show that the improved algorithm can solve the problem of passenger detection on two-wheeled vehicles. The detection accuracy of 92.31% and the recall rate of 96.96% are obtained on the two-wheeled vehicle manned data set.
Keywords: Target detection, area generating network, feature fusion.

---

1. INTRODUCTION

The two wheeled vehicle is more and more popular in the traffic, and because of its non-standard carrying, it is easy to cause traffic accidents, so the detection of two wheeled vehicle carrying is imminent. The key of the two wheel vehicle manned detection is to detect passengers on the two wheeled vehicles, which is similar to pedestrian detection, so we can learn from the research results of pedestrian detection.

The AlexNet [1, 2] convolutional network has won the championship with its performance higher than the traditional detection algorithm in the ImageNet [3,4,5] visual recognition competition. Since then, CNN (Convolutional Neuron Network, convolutional neural network) has attracted the attention of researchers. From 2012 to 2015, researchers continued to research and improve deep learning algorithms, and the accuracy of target detection continued to improve. For example, Girshick et al. designed the R-CNN algorithm, which can be divided into the following three parts. First use the SS (Selective Search, selective search) method or Edge Boxes to obtain the candidate area that probably contains the target from the picture; then use each candidate area with a fixed size as the input of CNN to obtain the features; finally, the second step The features are used as the input of the SVM classifier to implement classification and regression operations. Although the detection performance of R-CNN is higher than that of CNN, the training process is more time-consuming and laborious. He Kaiming [6] and others proposed SPP (Spatial Pyramid Pooling, spatial pyramid sampling layer)

in 2014. This method efficiently handles the problem of requiring separate operations for all regions in R-CNN. Although SPP is faster in detection speed than the former, the training process of SPP is also cumbersome. In 2015, Girshick [7] optimized the training steps of R-CNN and SPP and then designed Fast R-CNN to improve accuracy and speed. Compared with the training speed of the latter, the training speed of the former is about 64 times faster. Since R-CNN and Fast R-CNN both use SS to generate target candidate regions, the algorithm has a high time complexity. Therefore, in the same year, Ren Shaoqing [8] and others designed the Faster R-CNN target detection algorithm composed of the Fast R-CNN algorithm responsible for detection and the RPN (Region Proposal Network) network responsible for generating candidate regions. In 2018, He Yuming [9] and others optimized Faster R-CNN and then designed Mask R-CNN for semantic recognition.

In conclusion, faster CNN algorithm has good adaptability, high detection accuracy and fast detection speed, so this algorithm is selected to detect two wheeled vehicles. However, in real images, small head size and overlapping passengers often lead to missed detection of passengers on two wheeled vehicles, so the detection algorithm needs to be improved. The improved faster CNN algorithm adjusts the size of the anchor and the feature fusion structure in the region proposal network (RPN) to enhance the detection of small-scale targets and multi-scale targets. The improved algorithm also uses soft non maximum suppression (soft NMS) [14] to replace non maximum suppression (NMS) to improve the detection effect of overlapping targets.

2. RELATED WORK

In reference [10], the difference of skin color and hair color of human face is studied, and then the head model is established by clustering. Finally, the head is detected by template matching, which has high accuracy for pedestrian detection. In this paper, the sliding classifier is used to extract the features of the target. In reference [11], firstly, the histogram of gradient direction is calculated, and then the gradient model of the object is trained by support vector machine (SVM). Finally, the model is matched with the object to detect the object. In reference [12,13,14], by separating the moving object from the background and using fusion region matching and feature matching for the background, the head can be detected quickly. However, the features extracted by the above traditional algorithms are relatively single and are greatly affected by the environment. In recent years, deep learning technology has become the mainstream direction of object detection [5-10], and has also been applied to pedestrian detection. In reference [15,16], the pedestrian head model was established, the head features were extracted, and the Fast R-CNN training test was used, which showed excellent adaptability. Reference [17] uses pyramid network structure and feature fusion technology to improve faster CNN, which improves the effect of pedestrian detection in coal mine. In reference [18,19], we compared different feature extraction networks and detection algorithms through experiments, and found that faster CNN with perception V2 as feature extraction network had better effect in pedestrian detection at station, with detection accuracy of 81.08% and detection time of 0.5765s.

The single-stage algorithm has faster detection efficiency than the two-stage algorithm. In 2015, the YOLO (You Only Look Once, you only need to see it once) target detection algorithm was proposed by Redmon et al [20]. The main idea of the algorithm is that the target detection network takes pictures

as input, and then outputs the target category and detection frame. Compared with a series of algorithms based on R-CNN, the former has a significant improvement in detection speed, but the accuracy of detecting objects and small objects close to each other is not as good as the former. In 2016, edmon et al. designed an improved algorithm YOLOv2 based on YOLO [21]. The author designed a scheme for target classification and detection at the same time, so the detection effect was improved while maintaining a high detection speed. In the same year, Wei Liu [22] and others designed an end-to-end algorithm SSD (Single Shot MultiBox Detector), which can predict feature maps of different scales, so that the detection accuracy can reach the accuracy of Faster R-CNN. In 2018, Redmon et al. improved on the basis of YOLOv2 and designed YOLOv3 [24]. It draws on the idea of SSD, uses feature maps with different sizes as input for prediction, and uses a better backbone network. The R-CNN [23] series of algorithms have relatively weak accuracy in identifying objects, but it has achieved a relative balance between higher speed and higher accuracy.

3. METHOD

3.1 Network structure

This paper proposes an improved Fast R-CNN based E-bike manned detection method, using the detection model process: first detect the E-bike and head in the image; then find out all passengers according to whether the center point of the head detection frame is in the E-bike area; finally, mark the passengers on the two wheeled vehicle in the image. The technical scheme is as follows:

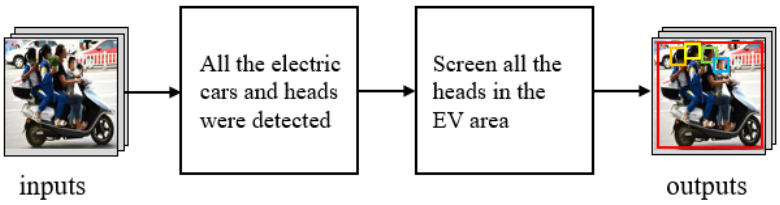

Figure 1 Flow chart of electric bicycle manned detection model

Although Fast R-CNN is widely used in image detection, target tracking and other visual tasks, its detection accuracy for small targets needs to be improved. Therefore, multi-scale convolutional neural network is introduced to solve this problem, and channel attention (CA) is introduced in the feature fusion stage to adjust the importance of each channel feature. The overall structure is shown in Figure 2.

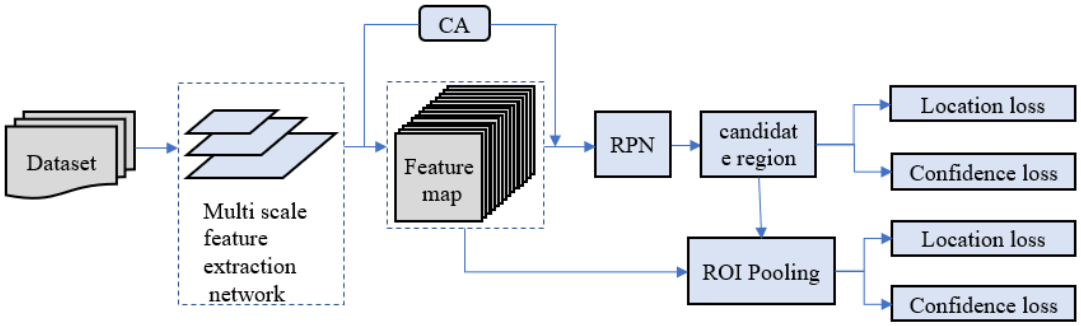

Figure 2 Improved overall structure framework of Fast R-CNN

The multi-scale feature extraction network uses googlenet. Compared with VGG network, googlenet has a deeper and wider network framework, which can extract richer features of two wheeled vehicles

and human head, so as to improve the training results. The convolution kernel of different scales in googlenet is used to extract the features in the burst receptive field of the image, and the CA module is used to adjust the importance of the feature maps of different channels, and the generated feature maps are sent to the RPN to generate candidate regions; then the features of the extracted candidate regions are sent to the ROI pooling layer to be processed into fixed size feature vectors; Finally, it is sent to the full connection layer to realize the regression of classification and border.

3.2 Inception V2 module

The Inception V2 module is used in googlenet. The multi-layer structure of convolutional neural network can automatically learn different levels of image features. The low-level feature map retains the image edge, contour, texture and other local detail information, which is conducive to target location. The high-level feature map contains more abstract semantic information, which is conducive to target classification, but the perception of details is poor. Scale feature fusion fuses low-level features and high-level features through top-down horizontal connection to construct a feature representation with fine-grained features and rich semantic information. The fused features are more descriptive and conducive to small target detection. Its structure is shown in Figure 3.

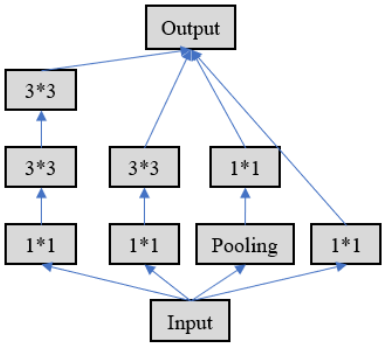

Figure 3 Inception V2

3.3 Loss function

Loss function of target detection method. The total loss function of the target detection method in the invention is defined as the weighted sum of the location loss (LOC) and the confidence loss (CONF)

$$\mathrm{L(x,c,l,g)} = \frac{1}{N}(L_{conf}(x,c) + \alpha \cdot L_{loc}(x,l,g))$$

Where N is the number of positive samples of a priori box; x is $x_{i,j}^{p} = \{0,1\}$, c is the confidence, l is the prediction box, g is the real box, α is the weight coefficient, L_{conf} is the confidence loss, L_{loc} is the location loss.

$$L_{conf}(x,c) = -\sum_{i\in pos}^{N} x_{i,j}^{p} \log(\hat{c}_i^p) - \sum_{i\in neg}^{N} \log(\hat{c}_i^0)$$

$$\hat{c}_i^p = \frac{\exp(\hat{c}_i^p)}{\sum_p \exp(\hat{c}_i^p)}$$

In order to minimize the error between the predicted value and the real value, the regression loss expression is as follows

$$L_{loc}(x,l,g) = \sum_{i\in pos}^{N} \sum_{m\in(cx,\ xy,\ w,\ h)} x_{i,j}^{k}\, Smooth_{L1}(l_i^m, \hat{g}_j^m)$$

$$\begin{cases} \hat{g}_j^{cx} = \dfrac{(g_j^{cx} - d_i^{cx})}{d_i^w} \\ \hat{g}_j^{cy} = \dfrac{(g_j^{cy} - d_i^{cy})}{d_i^h} \end{cases}$$

$$\begin{cases} \hat{g}_j^h = \log(\dfrac{g_j^h}{d_i^h}) \\ \hat{g}_j^w = \log(\dfrac{g_j^w}{d_i^w}) \end{cases}$$

Where, $Smooth_{L1}$ stands for smooth L1 loss function, $(g_{cx},\ g_{cy},\ g_{w,}\ g_h)$ indicates the prediction bounding box. $(d_{cx},\ d_{cy},\ d_{w,}\ d_h)$ Indicates the error bounding box, $\llbracket(l_{cx},\ l_{cy},\ l_{w,}\ l_h)$ represents the offset of the predicted bounding box from the error bounding box.

4. EXPERIMENT

4.1 Training

Our experiment is on 4 Titan XP GPU. The network is based on the Pytorch framework. We use Adam optimizer to optimize the parameters and set the original learning rate to le-5. The parameters are randomly initialized by Gaussian distribution. The average value is zero and the standard deviation is 0.01. In addition to the output layer, we also use batch standardization layer and RELU layer after each convolution layer In order to improve the training speed and effectively avoid the disappearance and explosion of gradient. The network parameters are set as follows.

Table 1 Network structure of feature extraction

Layers	Kernel/stride	Output size
Conv2d_1	Kernel_size=5, Stride=1	112×112×34
MaxPool_1	Kernel_size=3, Stride=2	56×56×64
Conv2d_2	Kernel_size=1, Stride=1	56×56×64
Conv2d_3	Kernel_size=3, Stride=1	56×56×192
MaxPool_2	Kernel_size=3, Stride=1	28×28×192
Mixed_1		28×28×256
Mixed_2		28×28×320
Mixed_3		14×14×576
Mixed_4		14×14×576
Mixed_5		14×14×576
Mixed_6		14×14×576

4.2 Result analysis

The comparison of the detection effect of the original algorithm and the improved faster CNN algorithm on the same image is shown in Figure 4. It can be seen from Figure 4 (a): in the first picture, the blocked rear passengers on the two wheeled vehicle are missed, and in the second and third pictures, the passengers with small visual distance are missed, which indicates that the effect of the original algorithm is not good. Figure 4 (b) can detect the blocked head and small size head, reflecting the better effect of the improved algorithm.

(a) Original faster CNN algorithm detection effect picture

(b) Detection effect picture of improved faster CNN algorithm

Figure 4 Comparison of detection effect of faster CNN algorithm before and after improvement

The comparison before and after the improvement is shown in Table 2. It can be seen from table 2 that there are 1 056 actual heads, and 35 targets in the improved algorithm are missed because the heads of these passengers are seriously blocked and the features are not obvious.

Table 2 Comparison of algorithm before and after improvement

Methods	N_{TP}	N_{FN}	N_{FP}	Recall	Accuracy
Original algorithm	951	105	122	90.09	90.09
Improved algorithm	1221	35	95	96.96	92.31

There are 95 false detections, some of which are the trunk of two wheeled vehicles being mistakenly detected as human heads, but mainly pedestrians near two wheeled vehicles being mistakenly detected as on-board personnel. The detection accuracy of the improved algorithm is 92.31%, and the recall rate is 96.96%.

5. CONCLUSION

A detection model based on improved faster CNN algorithm is proposed. The detection performance of the model is improved by optimizing RPN and using soft NMS. The improved algorithm improves the accuracy and recall rate of two wheel vehicle occupant detection, and the performance of the model is also effectively improved. The next step will focus on the pedestrian interference and serious occlusion near two wheeled vehicles, in order to further improve the performance of two wheeled vehicle manned detection.

REFERENCES

[1]Jinyong Chen,Jianguo Sun,Yuqian Li,Changbo Hou. Object detection in remote sensing images based on deep transfer learning[J]. Multimedia Tools and Applications,2021(prepublish).

[2]Bosquet Brais,Mucientes Manuel,Brea Víctor M.. STDnet-ST: Spatio-temporal ConvNet for small object detection[J]. Pattern Recognition,2021,116.

[3]Zhao Xiaoli,Chen Zheng,Hwang Jenq Neng,Shang Xiwu. AFLNet: Adversarial focal loss network for RGB-D salient object detection[J]. Signal Processing: Image Communication,2021,94(prepublish).

[4]Quang Tran Ngoc,Lee Seunghyun,Song Byung Cheol. Object Detection Using Improved Bi-Directional Feature Pyramid Network[J]. Electronics,2021,10(6).

[5]Lihua Guo,Kuiwei Xu,Jingmin Li,Chong Liu. A MEMS flow sensor based on fish lateral line sensing system[J]. Microsystem Technologies,2021(prepublish).

[6]Saji Ruhin Mary,Sobhana N V. Real Time Object Detection Using SSD For Bank Security[J]. IOP Conference Series: Materials Science and Engineering,2021,1070(1).
[7]Jiang Xiaoming,Xiang Fugui,Lv Minghong,Wang Wei,Zhang Zhonghua,Yu Yi. YOLOv3_Slim for Face Mask Recognition[J]. Journal of Physics: Conference Series,2021,1771(1).
[8]Jamal Jadaa Khalid,Munirah Kamarudin Latifah,Noori Hussein Waleed,Zakaria Ammar,Muhammad Mamduh Syed Zakaria Syed. Multi-Target Detection and Tracking (MTDT) Algorithm Based on Probabilistic Model for Smart Cities[J]. Journal of Physics: Conference Series,2021,1755(1).
[9]Kousik NalliyannaV.,Natarajan Yuvaraj,Arshath Raja R.,Kallam Suresh,Patan Rizwan,Gandomi Amir H.. Improved salient object detection using hybrid Convolution Recurrent Neural Network[J]. Expert Systems with Applications,2021,166.
[10]Surya Kant Singh,Rajeev Srivastava. A robust RGBD saliency method with improved probabilistic contrast and the global reference surface[J]. The Visual Computer,2021(prepublish).
[11]Nie Xinming,Chen Zhengyi,Tian Yaping,Chen Si,Qu Lulu,Fan Mengbao. Rapid detection of trace formaldehyde in food based on surface-enhanced Raman scattering coupled with assembled purge trap[J]. Food Chemistry,2021,340.
[12]Fonseca Erika,Santos Joao F.,Paisana Francisco,DaSilva Luiz A.. Radio Access Technology characterisation through object detection[J]. Computer Communications,2020(prepublish).
[13]Ding Lianghui,Xu Xin,Cao Yuan,Zhai Guangtao,Yang Feng,Qian Liang. Detection and tracking of infrared small target by jointly using SSD and pipeline filter[J]. Digital Signal Processing, 2020, 110 (prepublish).
[14]Li Fangyu,Jin Weizheng,Fan Cien,Zou Lian,Chen Qingsheng,Li Xiaopeng,Jiang Hao,Liu Yifeng. PSANet: Pyramid Splitting and Aggregation Network for 3D Object Detection in Point Cloud[J]. Sensors, 2020, 21(1).
[15]Xurshid Farhodov,Oh-Heum Kwon,Kwang-Seok Moon,Oh-Jun Kwon,Suk-Hwan Lee,Ki-Ryong Kwon. A New CSR-DCF Tracking Algorithm based on Faster RCNN Detection Model and CSRT Tracker for Drone Data[J]. Journal of Korea Multimedia Society,2019,22(12).
[16]Yang Inchul,Jeon Woo Hoon,Lee Joyoung,Park Jihyun. Development of an Integrated Traffic Object Detection Framework for Traffic Data Collection[J]. The Journal of The Korea Institute of Intelligent Transport Systems,2019,18(6).
[17]Frederik E.T. Schöller,Martin K. Plenge-Feidenhans'l,Jonathan D. Stets,Mogens Blanke. Assessing Deep-learning Methods for Object Detection at Sea from LWIR Images[J]. IFAC PapersOnLine,2019,52(21).
[18]Wang Shigang,Yang Shuyuan,Wang Min,Jiao Licheng. New Contour Cue-Based Hybrid Sparse Learning for Salient Object Detection.[J]. IEEE transactions on cybernetics,2019,PP.
[19] Jaejoong Kim, Jinkyu Ryu, Donggeol Kwak, Sunjun Byun. A Study on Flame Detection using Faster R-CNN and Image Augmentation Techniques[J]. Journal of IKEEE,2018,22(4).
[20]Pak Tae Young,Oh Hyun Soo,Chang Seong Rok. Analysis of Traps Incidents of Metro Train Door by Human Factors[J]. Journal of the Korean Society of Safety,2018,33(6).
[21]Kim Myung Eun,Kim Cheonyong,Yim Yongbin,Kim Sang Ha,Son Young Sung. An Origin-Centric Communication Scheme to Support Sink Mobility for Continuous Object Detection in IWSNs[J]. KIPS Transactions on Computer and Communication Systems,2018,7(12).
[22] Min Dong-eul, Kim Jeong-beom. Design of an Auto-Focus Module Using Target Object Detection Algorithm[J]. Journal of Korean Institute of Information Technology,2015,13(12).
[23]Souhail Guennouni,Ali Ahaitouf,Anass Mansouri,Aiguo Song. A Comparative Study of Multiple Object Detection Using Haar-Like Feature Selection and Local Binary Patterns in Several Platforms[J]. Modelling and Simulation in Engineering,2015,2015.
[24]Yoon Kyung Han,Jung Yong Chul,Cho Jae Chan,Jung Yunho. Design and Implementation of Optical Flow Estimator for Moving Object Detection in Advanced Driver Assistance System[J]. The Journal of Advanced Navigation Technology,2015,19(6).