Foreign body detection of urban rail train roof based on deep learning

Chen Song Wang¹, Wei Cui¹ and Xiao Zhu Wang¹

¹School of Electronic and Information Engineering, Changchun University of Science and Technology, Changchun 130022, China

* Corresponding author’s e-mail: 2019100589@mails.cust.edu.cn

Abstract. The identification of foreign body on the roof of urban rail train is an important link to ensure train safety. The foreign body on the roof has the characteristics of small size and irregular shape. In this paper, this paper proposes an improved YOLOV4 foreign body recognition method based on deep learning. It expands the sample of foreign body data through data enhancement to enhance the robustness of foreign body detection; DCN (Deformable Convolution) was added to the backbone network CSPDarkNet53 to enhance the receiver field of the feature extraction network and enhance the detection ability of irregular objects. The experimental results show that the improved YOLOV4 can obtain 94.5% MAP on NVIDIA RTX2080Ti. It can accurately and quickly detect the roof foreign body.

1. Introduction

The construction scale of urban rail transit in China has been expanding year by year, and rail transit has become the preferred mode of transportation for passengers in some medium and large cities [1]. It is of great significance to ensure the normal running of urban rail train to maintain the normal running of urban traffic. At present, the detection methods for foreign bodies on the roof are mainly divided into two kinds: inspector climbing operation method and roof image detection method. Ascending operation method is the inspection personnel on the roof of the car to naked eye for manual investigation, this way of high risk factor, accuracy can not be guaranteed, but also may leave wrench, gloves and other tools caused by secondary foreign bodies. The image detection method is mainly to shoot the local picture of the roof [2] with the planar array camera and the linear array camera, then get the panorama through stitching, and find the foreign body through template matching. However, the array camera is greatly affected by illumination and the image collection is also discontinuous, so it is difficult to Mosaic the panorama, resulting in low detection accuracy. The image brightness obtained by linear array camera is uniform, but there are also some image stretching and compression caused by speed adjustment, resulting in a certain image distortion, which affects the accuracy of template matching in detecting foreign bodies. In recent years, as deep learning networks [3] have made remarkable achievements in the field of target detection, more and more foreign body detection methods based on deep learning have been proposed. Xu Yan et al. [4] proposed a railway foreign body invasion limit detection algorithm based on the Faster R-CNN network model, replacing the full connection with the global average pooling layer, and improving the accuracy of the proposed target area by increasing the number of anchor points. Yu Xiaoying [5] proposed a detection method of railway invasion foreign body based on YOLO algorithm, which used 24 convolutional layers, 4 maximum pooling layers and 2 full connection layers to complete the extraction, dimensionality reduction and recognition output of foreign body images, and finally realized the function of foreign body positioning and recognition. Wang Yang [6] et
al. proposed a train operation safety detection method based on deep learning algorithm. By simulating the transmission structure of multi-layer neural network, they input and identify hidden trouble scanning materials, so as to achieve the purpose of detecting suspended foreign bodies in trains. However, these methods are not effective for detecting small targets and irregular foreign bodies.

An advanced target detection network, YOLOV4, is used in this paper. Data enhancement expands standard data samples to enhance the robustness of foreign body detection. The variable convolution kernel is introduced to enlarge the receptive field of feature extraction network, enrich the extraction of foreign body features, and enhance the detection ability of irregular foreign body.

2. Related work
YOLOV4 [7] adopts the end-to-end detection model of feature fusion based on CSPDarknet-53 and Panet network. The foreign body images are divided into several grids, and the candidate box prediction is made on each grid. Finally, the probability and coordinates of the foreign body category predicted by each candidate box are output to obtain the location of the foreign body. The size of the input foreign body image is 608x608. The foreign body feature extraction network CSPDarkNet-53 network is mainly composed of 1×1 and 3×3 convolutional layers. The Darknet Conv2D_BN_MISH convolution block was used to deal with the same input but different output problems caused by parameter optimization during training, and the speed of network training was accelerated. In the feature pyramid part, the Leaky relu function is used to enter the DarknetConv2D_BN_Leaky convolution for the last feature layer of CSPDarkNet-53 to fit the nonlinearity and avoid the problem of gradient disappearance. The SPP module uses the pool core size of 13x13, 9x9, 5x5 and 1x1 for processing, which can greatly increase the receptive field and isolate the most significant context features. Foreign body feature extraction is completed in Panet network with three effective feature layers. First, traditional feature extraction is carried out from top to bottom and then sampling is carried out from bottom to top. The three effective feature layers of the network are fully extracted by feature fusion for several times. Finally, YOLO HEAD was used to obtain the results of category, confidence and prediction box parameters.

![YOLOv4 network structure](image-url)
2.1. Deformable convolution
The traditional convolution kernel is generally square, and the shape of the deformable convolution kernel is variable, so that it can only see the image region of interest, so that the target features can be recognized better. CSPDarknet53 is composed of many CSP blocks. In order to fully extract the features of foreign bodies, DCN[8] is introduced into the backbone network CSPDarknet53 of YOLOV4. Replace the 3 by 3 convolution in each CSP block with DCN. In this way, the improved network backbone enhances the ability to extract the features of irregular targets.

2.2. Data enhancement
YOLOV4 network is a data-driven target detection model. A large number of samples can enrich the types of foreign body features. In the task of foreign body detection of urban rail train, the original data of small size foreign body is less, so it is necessary to enhance the data to increase the robustness of the model for small size foreign body detection. In this paper, image rotation, cropping and random occlusion are adopted to expand the original data set, and Aug is used to represent it.

2.3. Model training
The training server of the network uses RTX2080Ti graphics card, and the software environment is Pytorch1.2 and CUDA10., the whole training process is 200 epochs, and the weight attenuation coefficient is set as 0.0005. The first 50 Epoch training sessions had a learning rate of 0.001 and a Batch Size of 16. The learning rate of the last 150 Epoch training sessions was 0.0001, and the Batch size was set to 8.

3. Experiment and analysis
The experiment used 854 pictures of foreign bodies on the roof as training samples, 80% of which were used for training, 10% for Val, and 10% for testing. The experimental results are as follows:

![Figure 2 Comparison of multi-scale foreign body detection effect](image)

Figure 2 Comparison of multi-scale foreign body detection effect
Table 1  Comparison of experimental results

| The model name | Aug | DCN | mAP  | FPS |
|----------------|-----|-----|------|-----|
| YOLOv4         |     |     | 85.2%| 76  |
| YOLOv4 √       |     |     | 92.4%| 76  |
| YOLOv4 √       |     |     | 92.8%| 61  |
| Our √ √        |     |     | 94.5%| 61  |

The effect of original network in foreign body detection of urban rail train is not ideal, there are many missed detection targets. As the data is enhanced, the detection efficiency is improved. Smaller foreign bodies are detected, but irregular foreign bodies are ignored. After the deformable convolution was added into the backbone, the detection accuracy was improved by 7.6%, some irregular foreign bodies were detected, and small target foreign bodies were missed. With the addition of AUG and DCN in YOLOV4, the detection accuracy reached 94.5% and the detection speed reached 76FPS. As can be seen from the figure, smaller and irregular foreign bodies were detected. Therefore, the improved network can well adapt to the scene of foreign body detection on the roof of urban rail train.

4. Conclusion
An improved YOLOV4 network based on Aug +DCN was proposed for the detection of foreign bodies on the roof of urban rail train. Data enhancement is used to increase the data samples and improve the robustness of the detector. The variable appearance convolution is used to increase the feature extraction of irregular foreign bodies and improve the detection accuracy of such foreign bodies. The experimental results show that the improved network can accurately detect foreign bodies on the train roof, and the mAP reaches 94.5%.

References
[1] Xiao T W, Xu Y G, Yu H M. Research on Obstacle Detection and Train Control Technology for Urban Rail Transit[J]. Journal of Ordnance Equipment Engineering, 2020, 41(10):161-165.
[2] Chang L, Liu Z, Shen Y, et al. Novel Multistate Fault Diagnosis and Location Method for Key Components of High-Speed Trains[J]. IEEE Transactions on Industrial Electronics, 2020, 68(4): 3537-3547.
[3] Li J Y, Zhao Y K, Xue Z E, et al. A survey of model compression for deep neural networks[J].Chinese Journal of Engineering, 2019, 41(10): 1229-1239. doi: 10.13374/j.issn2095-9389.2019.03.27.002
[4] Xu Y, Tao H Q, Hu L L. Railway Foreign Body Intrusion Detection Based on Faster R-CNN Network Model[J]. Journal of the China Railway Society, 2020, 42(05):91-98.
[5] Yu X Y, Su H S, Jiang Z. Detection Method of Railway Intruding Obstacle Based on YOLO Algorithm. Journal of Lanzhou Jiaotong University, 2020, 39(02):37-42.
[6] Wang Y, Wang J G. Study on safety inspection of railway train operation based on deep learning algorithm[J]. China Safety Science Journal, 2018, 28(S2):41-45.
[7] Bochkovskiy A, Wang C Y, Liao H Y M. Yolov4: Optimal speed and accuracy of object detection[J]. arXiv preprint arXiv:2004.10934, 2020.
[8] Dai J, Qi H, Xiong Y, et al. Deformable convolutional networks[C]//Proceedings of the IEEE international conference on computer vision. 2017: 764-773.