Research On Grooved Rail Garbage Identification Algorithm Based On Improved YOLOv3

Xiaojie Huang¹, Kangquan Ye¹, Zhongbin Fang¹, Yongjun Xie*, Xizhe Ma¹, Jing Ji¹ and Qiantong Wu¹

¹Rail Transit Research Institute, Jinan University, Zhuhai, Guangdong Province, 519070, China
*Corresponding author’s e-mail: xieyongjun@jnu.edu.cn

Abstract. To solve the problem that existing modern tram track cleaning vehicles cannot automatically and accurately identify grooved rail garbage, this paper proposes a grooved rail garbage identification algorithm based on improved YOLOv3. This algorithm firstly extracts the groove rail region in the image, then adjusts and optimizes the residual block and convolutional layer network in the basic network of YOLOv3, and adopts k-means algorithm for clustering analysis to obtain anchor value of the adaptive dataset, which effectively improved the average identification accuracy of grooved rail garbage. The experimental results showed that the improved YOLOv3 model mAP reached 89.16% on the self-made Grooved Rail Garbage Dataset. For the image of 416×416, the recognition speed reached more than 14 frames per second. Compared with the traditional grooved rail garbage identification algorithm, this algorithm not only has higher recognition rate and accuracy, but also has good robustness to environmental changes.

1. Introduction

In recent years, with the rapid development of modern tram and the increase of the total mileage of modern tram, the cleaning of groove rails has begun to attract people's attention. Therefore, the automatic development of track cleaning vehicle is imperative[1].

However, the current way of track cleaning still give priority to with manual cleaning. In the future with the rapid growth of the mileage of modern tram, the traditional way of manual cleaning is bound to be unsustainable. As for the way to use track cleaning vehicle, only a few enterprises at home and abroad have developed the relevant track cleaning vehicle, which just adopt the semi-automatic cleaning[2].

In the Rail Transit Research Institute, Jinan University, the members of our project team have installed a control system on track cleaning vehicle, which can achieve the automatic cleaning of it. However, based on inter-frame difference method, the detection algorithm used in the control system is easy to identify track rust as garbage, cannot adapt well to environmental change, cannot classify the garbage, and has low recognition rate and high false alarm rate in the real track environment[3]. Therefore, this paper proposes a grooved rail garbage identification algorithm based on YOLOv3, and develops a grooved rail garbage data set, so as to realize the identification and classification of the garbage on groove rail. Compared with inter-frame difference method, this method improves the recognition rate and accuracy, which is of great significance to promote the development of track cleaning mode towards automation.
2. Algorithm overall design
The grooved rail garbage identification algorithm proposed in this paper mainly consists of the groove rail extraction module and the improved YOLOv3 module. Algorithm overall design is shown in Figure 1.

3. Groove rail extraction
The groove rail extraction module can extract the groove rail region in the image, so that the apparent properties of each image are as consistent as possible.

First of all, the image is graying and bilateral filtering, which can reduce the influence of image noise caused by illumination, temperature and ISO sensitivity on the accuracy of extraction. Secondly, the Canny edge detection algorithm is used to detect the edge region in the image. The algorithm can also filter other redundant noises. Then, we use Hough transform to detect the straight line of groove rail in the edge feature. Finally, the area outside the straight line of the groove rail is applied Padding by RGB to the gray background (128,128,128).
4. YOLOv3

The identification algorithm based on YOLOv3 network model transforms the target detection problem into a regression problem. The garbage location and category information are obtained through the regression analysis of the input garbage image[4-5].

4.1. Feature extraction

YOLOv3 uses Darknet-53 basic network to extract image features. On the one hand, it basically adopts full convolution structure, using 1×1 convolutional layer and 3×3 convolutional layer to form the basic network alternately. On the other hand, it adds residual modules to the network to solve the problem of deep network gradients. Each residual module consists of two convolutional layers and one shortcut connection.

The implementation of each convolutional layer contains convolution, BN and LeakyReLU. Each time the convolutional layer is passed, the size of the image is reduced by half. Finally, the learned features will be passed to the classifier for the prediction of coordinates and category labels of bounding boxes.

In Convolutional Neural Networks, more convolutional layers and deeper network structures tend to extract target features better. But when the network deepens, the extra convolutional layer will cause too many network model parameters, thus increasing the network computation. At the same time, in the field of object detection, although deep network can respond to semantic features, it is not conducive to object detection due to the small size of feature map and the limited geometric information contained in it. Although the shallow network contains a lot of geometric information, the semantic features of the image are not much, which is not conducive to image classification.

Considering the accuracy of the subsequent multi-scale feature detection, this paper partially improved the YOLOv3 basic network. The improved network structure is shown in Figure 3. The middle convolutional layer network is divided into two parts for optimization:

- The weighted pruning method is adopted to optimize the YOLOv3 convolutional neural network, that is, the convolutional layer that can be clipped in the network is selected first, and after the sparse training of BN layer parameters, unimportant connections in the network are removed according to the parameter range.
- Quantitative perception algorithm is adopted to quantify the weight and activation value of the optimized network model, so as to accelerate the network training speed and reduce the loss of network accuracy[6].

![Figure 3. Improved YOLOv3 basic network structure.](image-url)
4.2. Multiscale feature fusion
YOLOv3 uses the ideas of SSD and FPN for reference, extracts multiple feature maps, and uses low-level high-resolution information and high-level semantic feature fusion for detection. YOLOv3 adopts the method of three-scale feature map fusion for object detection, which enhances the accuracy of YOLOv3 network for small target detection. When the input is 416×416, the scales of these three feature maps are 13×13, 26×26 and 52×52.

Gao Xing and Liu Jianfei et al. considered that training with anchor adapted to the data set could make the model have better detection performance[7]. Therefore, K-means algorithm is used in this paper to carry out clustering analysis on the self-made Grooved Rail Garbage Dataset. K value is selected as nine. The obtained result corresponds to the size of nine anchors set in the training process. During the target detection, nine anchors will be evenly divided by three output layers of different scales, and the garbage of different sizes will be further predicted according to the three scales of large, medium and small.

4.3. Bounding box prediction
Image is divided into 13×13, 26×26, 52×52 grid cells according to three limits, each grid cell within each scale predicts three bounding boxes, each containing five parameters: \( TX, TY, TW, TH \) and Confidence. \( TX \) and \( TY \) are used to determine the upper left corner of bounding box. \( TW \) and \( TH \) are used to determine the width and height of bounding box. These five parameters can make bounding box predictions in combination with the image's upper-left offset and bounding box's wide and high.

Each grid cell is predicted to obtain the bounding box, confidence fraction and four category probabilities, and then decoded to obtain the bounding box coordinates under the image size of 416×416. Finally, it is adjusted to the bounding box coordinates under the original image size.

5. Experiment and analysis

5.1. Experimental environment
The experimental platform configuration environment of this paper is shown in Table 1. Model parameters are set as follows: Input_shape is (416,416); Num_classes is 4; Load_pretrained is True; Batch_size is 8; Learning_rate is 0.001; Ignore_thresh is 0.5; Momentum is 0.9; Policy is steps; Max_batches is 20000.

| Object                  | Related configuration          |
|-------------------------|--------------------------------|
| Operating system        | Windows 7                      |
| Graphics card           | NVIDIA TITAN X                 |
| CPU                     | Intel(R)Xeon(R)E5-2620,2.10GHz |
| GPU                     | CUDA 10.0.132,CUDNN 7.6.4      |
| Internal storage        | 112GB                          |
| Deep learning framework | Tensorflow                     |

5.2. Self-made Grooved Rail Garbage Dataset
In order to obtain the dataset that can be used for the training of YOLOv3 model, this paper adopts the method of artificial photographing to obtain the images on the real groove rail and establishes the self-made Grooved Rail Garbage Dataset. There are 1730 original images of various types, all with resolution of 1080×1440, including cigarette butts, plastic bags, bottle caps and leaves, each with about 430 pieces. For the annotation of dataset, labelling software is selected in this paper to manually label the garbage area of each image, and the tagged data information will be automatically saved as an XML file.
Since the images in the dataset are all based on groove rail, there are certain similarities in the data of the same category. Therefore, in order to increase the number and diversity, five methods, namely brightness enhancement, contrast enhancement, left rotation by 30°, right rotation by 30°, and mirror reversal, were adopted to enlarge the dataset. The resulting Grooved Rail Garbage Dataset including 8650 images is randomly generated into training set and test set at a ratio of 9:1.

5.3. Results analysis

The groove rail extraction and improved YOLOv3 network were used to identify and classify the grooved rail garbage, achieving good recognition effect. The mAP of the algorithm in this paper to identify the four types of garbage is shown in Table 2. The recognition effect of four types of garbage is shown in Figure 4.

Table 2. The mAP of the algorithm for four types of garbage.

| Category      | mAP   |
|---------------|-------|
| Cigarette butts | 91.54%|
| Plastic bags   | 83.95%|
| Bottle caps    | 90.81%|
| Leaves         | 86.37%|

Figure 4. The recognition effect of four types of garbage.

In order to evaluate the method of grooved rail garbage identification in this paper, the control variable method is used for experimental comparison. In the case that the input images are all 416×416, YOLOv3 algorithm and the improved YOLOv3 algorithm are trained and tested respectively. The experimental test results of different algorithms are shown in Table 3.

The experimental results show that although the inter-frame difference method has high recognition speed, its mAP is low and false alarm rate is high, so it is not suitable for practical track cleaning vehicle. Compared with YOLOv3, the improved YOLOv3 has improved mAP and average recognition speed. Meanwhile, the improved YOLOv3 meets the needs of modern tram track cleaners for normal operation.
Table 3. Experimental test results of different algorithms.

| Algorithm                        | mAP   | FAR  | FPS |
|----------------------------------|-------|------|-----|
| YOLOv3                           | 87.89%| 0.21%| 11  |
| Improved YOLOv3                  | 89.16%| 0.24%| 14  |
| Inter-frame difference method    | 64.77%| 48.12%| 21  |

6. Conclusions

This paper mainly researches the grooved rail garbage identification algorithm. Aiming at the low recognition rate of the traditional identification algorithm, this paper presents a grooved rail garbage identification algorithm based on the improved YOLOv3. Based on YOLOv3, this algorithm optimizes the convolutional layer network by using the weighted pruning method. It also adjusts the residual module in the YOLOv3 basic network and uses k-means algorithm for cluster analysis. The experimental results show that the algorithm has higher accuracy and better robustness than the traditional frame difference method. Its recognition speed can also meet the real-time requirements of track cleaning vehicle cleaning operation. Compared with YOLOv3, the improved YOLOv3 proposed in this paper has improved mAP and recognition speed.

Acknowledgments

This research was supported by Natural Science Foundation of Guangdong Province, China (NO: 2017A030310184), the National Undergraduate Innovational and Entrepreneurship Training Program (NO: 202010559061), the key project of the "Climbing Plan" of Guangdong College Students' Science and Technology Innovation Cultivation Special Fund (NO: pdjh2020a0058) and the Zhuhai Collaborative Innovation Center.

References

[1] Zhao, X., Gu, B. (2019) Statistics and analysis of urban rail transit operating lines in China in 2018. Urban Mass Transit, 22(01): 1-7.
[2] Liu, H. Master's thesis, Southwest Jiaotong University. (2017) Research and development of track cleaning technology and equipment for modern trams. https://kns-cnki-net-443.webvpn.jnu.edu.cn/KCMS/detail/detail.aspx?dbname=CMFD201702&filename=101729656.nh
[3] Dong, H., Ma, W., Huang, J., Li, H., Huang, Y., Liu, X., ... & Xie, Y. (2019, November). Development of Rail Precision Cleaning Vehicle for Modern Tramway. In IOP Conference Series: Materials Science and Engineering (Vol. 688, No. 2, p. 022016). IOP Publishing.
[4] Redmon, J., & Farhadi, A. (2017). YOLO9000: better, faster, stronger. In Proceedings of the IEEE conference on computer vision and pattern recognition (pp. 7263-7271).
[5] Redmon, J., & Farhadi, A. (2018). Yolov3: An incremental improvement. arXiv preprint arXiv:1804.02767.
[6] Sun, Y., & Ye, J. (2020) A convolutional neural network compression method based on pruning and quantization. Computer Science, 47(8): 261-266.
[7] Gao, X., Liu, J., Hao, L., & Dong, Q. (2020). Study on training set optimization and detection Method based on YOLOv3 algorithm. Computer Engineering and Science, 1.