Journal of Computers Vol. 34 No. 4, August 2023, pp. 145-152
doi: 10.53106/199115992023083404012

Computer Vision Aided Pantograph Fault Identification Method for Multiple Units

Peng-Jie Du*, Mu-Zhuo Zhang

School of Automation Engineering, Tangshan Polytechnic College,
Tangshan City 063299, Hebei Province, China
{dpj20220701, zmz19920923}@126.com

Received 1 July 2023; Revised 20 July 2023; Accepted 25 July 2023

Abstract. In order to solve the technical requirements for automatic recognition and judgment of pantograph wear degree of Multiple Units, this paper designs a network structure based on Mask R-CNN structure. At the same time, in order to improve the ability of image feature extraction in the network, the original backbone network is replaced with ResNet-50, a residual network with more prominent feature extraction ability. Secondly, in order to improve the ability to search for targets in the image, the detection head is reconstructed, to improve the recognition ability of targets. Finally, the effectiveness of the algorithm and its ability to identify pantograph faults were verified through simulation experiments.

Keywords: pantograph, Mask R-CNN, residual network

1 Introduction

High speed rail is a major infrastructure for national strategic development and plays a very important role in China's comprehensive transportation system and economic and social development. High speed Multiple Units rely on electric drive and electric marshalling, and the maximum operating speed can reach 300-350 km/h. Since the beginning of the 21st century, China has begun to study high-speed Multiple Units, and has conquered the problem of complete autonomy of ten supporting facilities, such as pantograph, for 20 years.

High speed operation of high-speed Multiple Units cannot be achieved without the high-voltage traction system, which can obtain electric energy from the contact power grid to drive the traction motor to drive the whole vehicle operation under the working condition of 300km/h. Therefore, the high-voltage traction system is one of the most important systems of high-speed Multiple Units, which determines the operation of the EMU.

The pantograph takes current by relying on the reliable contact between the carbon sliding plate on the pantograph head and the catenary. During the operation of high-speed railway, the two are in the state of Sliding friction. Constant wear and tear. As the operating mileage increases, the wear of the carbon sliding plate gradually increases. When the wear reaches a certain level, it needs to be replaced, otherwise there is a risk of not being able to raise the pantograph due to air leakage. Therefore, the content of this article is how to accurately identify and judge the degree of pantograph wear.

In order to improve the real-time image detection ability of the pantograph, the work done in this article is as follows:

1) Improved network structure by incorporating residual networks to improve the recognition speed of the entire algorithm

2) In order to increase the recognition accuracy of the network, an article was made in the image segmentation algorithm, and the detection head was reconstructed to improve the accuracy and recognition efficiency in image recognition.

3) The effectiveness of the algorithm was verified through recognition simulation using a simulation system and real images

This chapter is composed of the following chapters. Chapter 2 mainly focuses on the research achievements of relevant scholars. Chapter 3 mainly describes the improvement process of the recognition network, and presents the improvement process in detail through renderings. Chapter 4 describes the image segmentation processing method, with the aim of improving the recognition ability of target pixels. Chapter 5 introduces a detailed experi-

* Corresponding Author

mental simulation process and verifies the effectiveness of the algorithm. Chapter 6 is mainly the conclusion part, Summarize the article and propose further research directions.

2 Related Work

Slawomir Judek proposed a laser scanning detection system for measuring the installation error and position displacement of the pantograph during use, while also playing a role in identifying pantograph wear and defects [1]. Sacchi has developed an online detection system for the pantograph, known as Pantobot 3D, and installed it on both sides of the track. It uses advanced 3D imaging technology and machine learning technology to evaluate the condition of the pantograph. It can achieve real-time detection of carbon sliding plate wear and peeling, carbon sliding plate rotation, and bow head structure at speeds below 300km/h [2]. Ping Tan used the improved YOLOv4 as the basic model and integrated fuzzy and dirt detection algorithms to reduce the problem of interference in traditional detection methods. Then, he improved the detection accuracy of the fuzzy algorithm through spot detection and improved Breyne method, ultimately significantly improving the efficiency of the algorithm [3]. Xiukun Wei, focusing on the research and application of machine vision in the detection of the safety status of urban rail transit system, combed and summarized the specific application and research of machine vision in the detection of abnormal behaviors in the safety monitoring of escalators and platforms [4]. Chuanhui Wu, in response to the increasing demand for accuracy and speed in detecting the wear of train pantograph slide plates, has designed an automatic detection system for pantograph slide plate wear based on multiple visual sensors. While maintaining high measurement accuracy, it also has high repeatability, stability, and reliability, meeting the requirements of practical engineering applications [5]. Haowei Kou, aiming at the problem that the deformation and damage of the pantograph head that affects the listed travel causes the reliable operation of the train, proposed a method that uses the image Classful network to classify by intercepting the image of the pantograph head area. After simulation verification, the method accuracy, accuracy, and recall rate can almost all reach more than 90% [6]. Chen Yang proposed a visual tracking method for pantograph, based on YOLOv3 framework, which improved the recognition speed of the framework. The visual model mainly identifies faults such as pantograph detachment and missing sheep horns [7].

3 Establishment of a Pantograph Detection Network Model

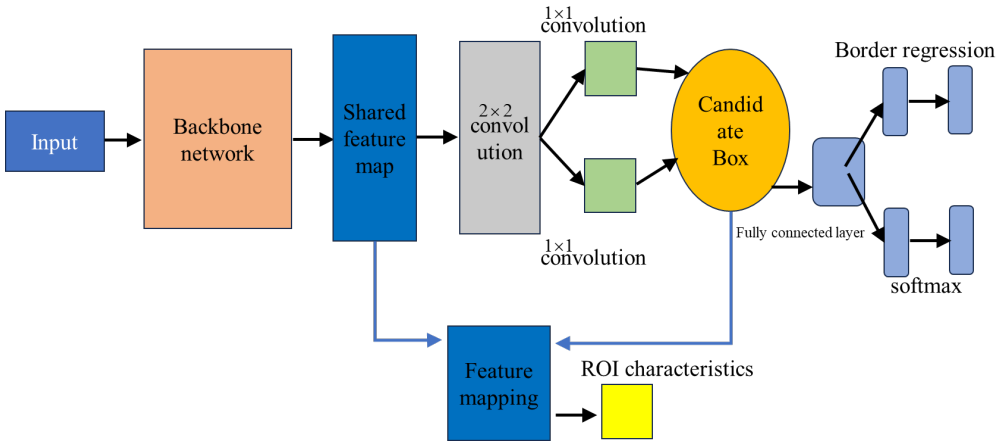


Fig. 1. Mask R-CNN model structure

Firstly, the detection model needs to be accurately positioned. By locating the bow head position, the image detection range can be reduced, and the detection speed can be improved. Accurate positioning can narrow the recognition range. In addition, accurate positioning directly eliminates the influence of complex backgrounds on

fault recognition. Taking advantage of the detection accuracy advantages of the Mask R-CNN framework and considering the detection characteristics of pantograph positioning in transit, the Residual Network (ResNet) structure with outstanding feature extraction ability and the Feature Pyramid Networks (FPN) multi-scale detection structure are applied to the bow head positioning model based on the Mask R-CNN detection framework.

The basic structure of the model includes a backbone network and a candidate area recommendation network. The backbone network is used to extract image features. The candidate area recommendation network uses a small network to evaluate the feature convolution mapping in the sliding window 2×2, and the detection head uses a nonlinear function *soft*max to classify and regress the input candidate Areal feature features, so as to obtain the target category and location correction. The model structure is shown in Fig. 1.

3.1 Replacement of Backbone Network

Because the backbone network of the original model structure has the problem of insufficient feature extraction, in order to obtain the feature expression ability of sufficient depth, the network structure with smaller parameters and stronger feature extraction ability is used to replace the original backbone network for feature extraction. In the collection of on-board video surveillance for pantograph, a 200 or 500 pixel visual camera is usually used. Considering the balance between actual pantograph image size, detection accuracy, and speed, and referring to the original Mask R-CNN commonly used object detection model, the input image size is set to 226×226.

The ResNet50 [8] network is composed of 5 residual modules, with residual module 3 parameters. It represents the inclusion of 4 building blocks, each containing 3 convolutional layers. 1×1×128 represents the use of 128 convolutional kernels of size 1×1 in the convolutional layer; In addition to residual module 1, the network uses the 1×1 convolution kernel for dimensional control, and the 3×3 convolution kernel for convolution and downsampling to control the parameter quantity. The ResNet50 parameter table is shown in Table 1.

Table 1. ResNet50 network architecture

Network layer name	Output Size	ResNet50
Conv1	64×64	7×7, 64, stride 2 3×3 max pool, stride 2
Conv2_x	32×32	$\begin{bmatrix} 3\times3 & 64 \\ 3\times3 & 64 \end{bmatrix}\times2$
Conv3_x	16×16	$\begin{bmatrix} 3\times3 & 128 \\ 3\times3 & 128 \end{bmatrix}\times2$
Conv4_x	8×8	$\begin{bmatrix} 3\times3 & 256 \\ 3\times3 & 256 \end{bmatrix}\times2$
Conv5_x	4×4	$\begin{bmatrix} 3\times3 & 512 \\ 3\times3 & 512 \end{bmatrix}\times2$
	1×1	Average pool, 6-d fc, Softmax

3.2 Design of Candidate Area Boxes

Taking the feature map output at any level of FPN as input, after $3\times3\times512$ convolution, a 256 dimensional feature vector will be obtained. The feature map composed of feature vectors will be convolutioned with 1×1 instead of the fully connected layer. Then, the combination of coordinates and classification will be calculated through $soft\max$ classifier and bounding box regression [9]. RPN_Head will output the position coordinates of candidate boxes represented by 4 values (x, y, w, h), and each candidate box is the foreground/background probability. Based on the foreground/background probability and candidate box coordinates, the target candidate region in the original image can be obtained. The structural design is shown in Fig. 2.

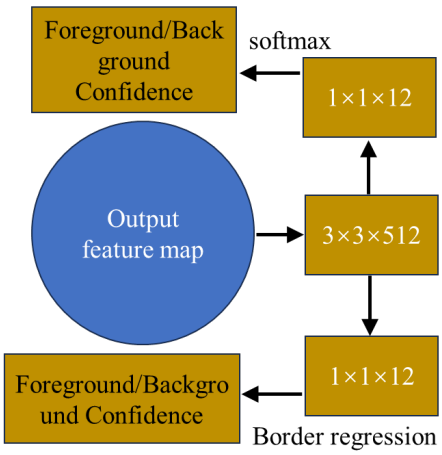


Fig. 2. Candidate box structure

4 Image Segmentation Processing

In order to facilitate the identification and localization of faults in the image, it is necessary to segment the image to obtain the appearance contour of the target. Traditional image segmentation methods, such as commonly used threshold based segmentation and edge detection based segmentation, are susceptible to image background interference and directly segment the target and background pixels without distinguishing them. In complex environments, the segmented image is relatively chaotic. In addition, traditional segmentation methods are difficult to ensure the continuity and closure of the segmented target contour [10]. This article addresses the difficulties of traditional methods for image segmentation of pantograph heads, improves the Mask R-CNN mask segmentation branch network structure, and redesigns the detection head. The improved detection structure is shown in Fig. 3.

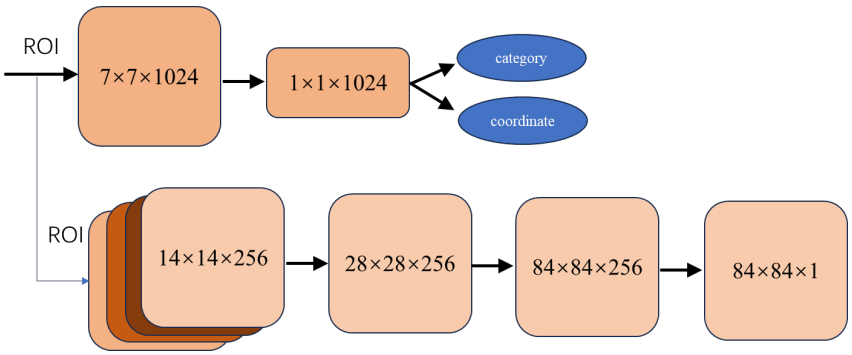


Fig. 3. Structural design of detection head

On the basis of inheriting Faster R-CNN, Mask R-CNN designed parallel segmentation, classification, and localization detection tasks in the second stage of the network, cleverly achieving segmentation. By adding a three-dimensional output that acts on each ROI $m \times m \times k$, where k is the number of categories and each category has a binary mask with a resolution of $m \times m$. All convolutions except for the output use 2×2, deconvolution uses 2×2, and the upsampling time step is 2.

As shown in Fig. 4, in order to improve the detection head design proposed for the algorithm proposed in this article, the fully connected layer of the classification and positioning branches is replaced by the average pooling layer to reduce model parameters; The mask segmentation branch is a fully convolutional network. If the mask output size is significantly increased, it will result in a significant increase in the computational complexity of the mask branch. However, if the mask size is too small, it cannot meet the segmentation accuracy requirements. Therefore, after balancing the quality of the output mask with the model's computational cost, the output mask size is set to. Structurally, the feature maps obtained from pooling are reduced in dimensionality, deconvoluted to, and then deconvoluted to, resulting in an output feature map size of, Reduced memory and computational losses.

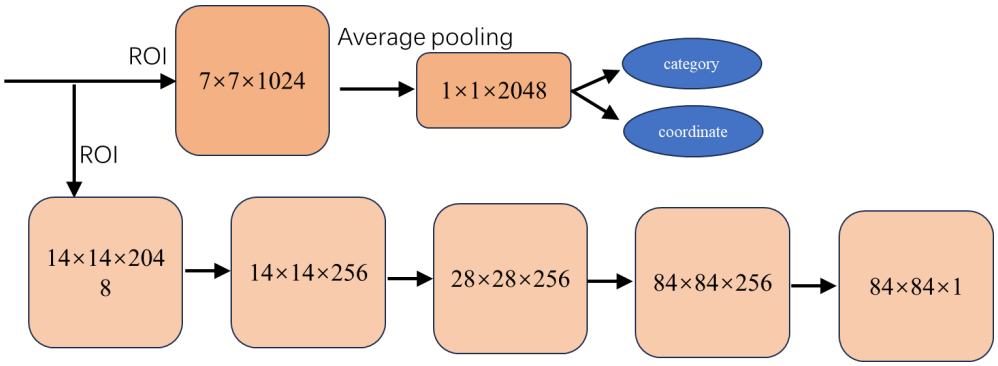


Fig. 4. Improved detection head

5 Simulation Experiment and Analysis

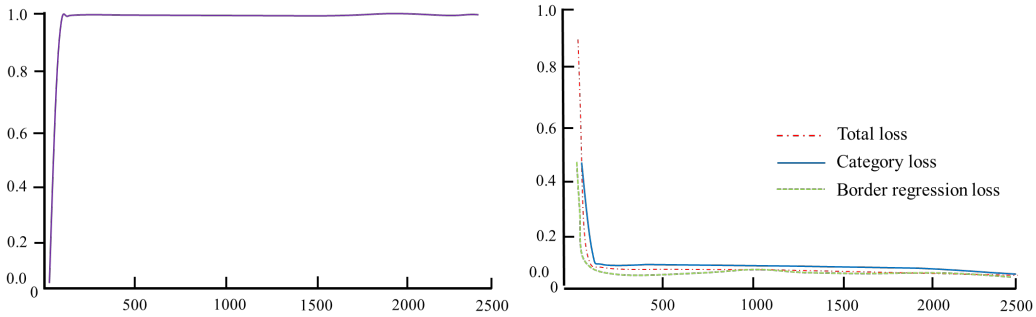
First, the performance of the improved algorithm model is analyzed, and the model is trained and tested by using the existing data set. Then, the training progress and Loss function comparison of the model structure in this paper are verified by comparing with other models to illustrate the effectiveness of the model. The hyperparameter settings are shown in Table 2.

Table 2. Hyperparameter setting

Parameter	Value
Initial learning rate	0.01
Attenuation index	0.1
Single training volume	8
Momentum parameter	0.9
Intersection to parallel ratio threshold	0.8

5.1 Model Training

A training model was built using a Python machine learning library on an Ubuntu 18.4 system with an I5-9600 processor and NVIDIA RTX-2060S graphics card. A small batch of gradient descent is used. The total number of iterations is 2500, and the Learning rate is attenuated in 1000 and 1500 rounds. The training accuracy and training loss of the model are shown in Fig. 5.



(a) Training accuracy (b) Training losses

Fig. 5. Improved detection head

From the record of training loss and training accuracy of the model, it can be seen that as the number of iterations in the model training process continues to increase, the relevant training loss of the model rapidly decreases to a certain level before stabilizing, and the recognition accuracy steadily improves before approaching 1. There are no significant fluctuations or oscillations, and the fitting state is good, without obvious overfitting or underfitting. When the model is iterated about 80 times, the model reaches its best recognition state, Similarly, when the model is iterated about 60 times, the loss of the model tends to stabilize.

5.2 Detection Performance Testing

The test data consists of a total of 300 images, with image types derived from three types of pantographs, as shown in Fig. 6. The images cover three types of pantographs with different sizes and backgrounds, and a total of three sets of data were set: 120 small target data, 60 medium target data, and 120 large target data. The performance indicators of each model are compared as shown in Table 3. For the convenience of comparison in subsequent articles, the improved Mask R-CNN is referred to as i-Mask R-CNN.

Fig. 6. Pantograph type diagram

Table 3. Hyperparameter setting

Algorithm name	Accuracy	Speed
YOLOv4	0.87	0.017 fra/s
Mask R-CNN	0.89	0.032 fra/s
Faster R-CNN	0.91	0.029 fra/s
i-Mask R-CNN	0.962	0.019 fra/s

After comparison, it was found that the improved Mask R-CNN algorithm has a recognition speed of 0.019 fra/s. Although not the fastest, it has the highest recognition accuracy, reaching 96.2%.

5.3 Analysis of Pantograph Identification Results

The pantograph is bent and slender, making it difficult to accurately segment. The model was trained under the original Mask R-CNN design parameters and the predicted results were obtained. The pantograph head was accurately positioned, but the mask segmentation results were extremely poor. By analyzing the algorithm principle, it can be concluded that during the model inference process, the predicted mask is upsampled to the actual bow head size, and the low resolution mask is distorted after upsampling multiple times; During the training phase, the label mask was reduced to a size that serves as the original training model. The recognition of the pantograph photo was performed, and the recognition results are shown in Fig. 7.

Fig. 7. Recognition result

After the above comparison, it can be seen that the recognition speed and accuracy of the algorithm proposed in this article have reached the expected level, and the detection of the pantograph can be evaluated through the corresponding system.

6 Conclusion

In this paper, a network structure based on Mask R-CNN structure is designed. At the same time, in order to improve the ability of image feature extraction in the network, the original backbone network is replaced with ResNet50, a residual network with more prominent feature extraction ability. Secondly, in order to improve the ability of searching for targets in the image, the detection head is reconstructed to improve the ability of target recognition. In the future research, we will continue to further study the recognition of pantograph, The research directions are as follows:

1) Due to the limitation of the installation angle of the pantograph camera, it is not possible to achieve comprehensive detection of the pantograph;

2) The detection speed is further improved. The main research object of this article is high-speed trains with a speed of 300km/h. As the speed continues to increase, trains with higher speeds will operate normally. Therefore, the algorithm detection speed needs to match higher vehicle speeds.

References

[1] J. Slawomir, J. Leszek, Analysis of Measurement Errors in Rail Vehicles' Pantograph Inspection System, Elektronika ir Elektrotechnika 22(3)(2016) 20-23.

[2] M. Sacchi, L. Ascari, S. Cagnoni, A. Piazzi, D. Spagnoletti, PANTOBOT: A computer vision system for the automatic inspection of locomotive pantographs, in: Proc. of the First International Conference on Railway Technology: Research, Development and Maintenance, Stirlingshire, 2012.

[3] P. Tan, Z.-S. Cui, W.-J. Lv, X.-F. Li, J. Ding, C.-Y. Huang, J.-E. Ma, Y.-T. Fang, Pantograph Detection Algorithm with Complex Background and External Disturbances, Sensors 22(21)(2022) 8425-8425.

[4] X.-K. Wei, D. Suo, D.-H. Wei, X.-M. Wu, S.-Y. Jiang, Z.-M. Yang, A survey of the application of machine vision in rail transit system inspection, Contral and Decision 36(2)(2021) 257-282.

[5] C.-H. Wu, J.-W. Ren, J. Liao, H.-K. Li, G. Yang, Desion of Wear Detection System of Pantograph Slide Plate Based on Multi-vison Sensor, Instrument Technique and Sensor (11)(2021) 78-87.

[6] H.-W. Kou, Y.-C. Su, H.-K. Li, G.-J. Chen, Y. Deng, Application of deep learning in the detection and classification of pantograph head fault, Laser Journal 43(6)(2022) 53-58.

[7] C. Yang, L-L. Shao, H.-M. Niu, T. Li, T. Wang, X.-G. Gong, Detection of pantograph fauits based on YOLOv3-tiny dection and kcf tracking algorithm, Intelligent Computer and Applications 10(5)(2020) 47-51.

[8] P. Wan, J.-W. Zhao, M. Zhu, H.-Q. Tan, Z.-Y. Deng, Y.-Y. Huang, W.-J. Wu, A.-Z. Ding, Freshwater fish species identification method based on improved ResNet50 model, Transactions of the Chinese Society of Agricultural Engineering 37(12)(2021) 159-168.

[9] H.-C. Liao, H.-X. Zhao, Infrared image target recognition method based on decision fusion of classifiers, Infrared and Laser Engineering 51(8)(2022) 282-287.

[10] Z.-Y. Qin, Y.-F. Lin, Y.-P. Chen, F.-Q. Lin, An Improved Spectral Clustering Image Segmentation Method Based on Adaptive Superpixel, Modern Computer (18)(2021) 103-108.