An Improved Faster R-CNN for Railway Fastening System Detection

Xinfeng Peng
School of Automation, Southeast University, Nanjing 210096, P.R. China; and Key Laboratory of Measurement and Control of Complex Systems of Engineering, Ministry of Education, Nanjing, 210096, P.R. China
1660482098@qq.com

Shuzhen Tong
School of Automation, Southeast University, Nanjing 210096, P.R. China; and Key Laboratory of Measurement and Control of Complex Systems of Engineering, Ministry of Education, Nanjing, 210096, P.R. China
114464169@qq.com

Xiaobo Lu*
School of Automation, Southeast University, Nanjing 210096, P.R. China; and Key Laboratory of Measurement and Control of Complex Systems of Engineering, Ministry of Education, Nanjing, 210096, P.R. China

Yun Wei
Beijing Mass Transit Railway Operation Corporation Limited, Beijing 100044, P.R. China

ABSTRACT

In the automatic railway anomaly inspection technology based on image processing and deep learning, an effective algorithm used for high-precision detection of the fastening system is very important, especially in turnout sections. It is challenging because the background of the turnout sections is complicated with various types of targets. This paper improved the Faster R-CNN model, used multi-scale feature map fusion for small targets. And modified predefined anchor to generate region proposals, added attention module to make the network focus on meaningful feature. Besides, this paper used cross-entropy function and SmoothL1 loss function for training and labeled 1200 image samples as dataset. Compared with the original Faster R-CNN model, the experimental results (AP) of the improved model in this paper increased from 96.3% to 98.9%, which effectively reduced the fault detection and missed detection and improved the accuracy of location.

CCS CONCEPTS

• **Computing methodologies**; • **Artificial intelligence**; • **Computer vision**; • **Computer vision problems**; • **Object detection**;

KEYWORDS

Railway, Fastener, Faster R-CNN, Multi-scale fusion, Attention module

ACM Reference Format:
Xinfeng Peng, Shuzhen Tong, Xiaobo Lu, and Yun Wei. 2021. An Improved Faster R-CNN for Railway Fastening System Detection. In *The 5th International Conference on Computer Science and Application Engineering (CSAE 2021), October 19–21, 2021, Sanya, China.* ACM, New York, NY, USA, 5 pages. https://doi.org/10.1145/3487075.3487184

1 INTRODUCTION

Fastener is one of the most important components of railway. Their main function is to keep the railways stable. When the fastening system is abnormal, such as fracture, torsion, loss, it will cause huge safety risk to the normal operation of the railway, even lead to serious consequences. Therefore, the condition of the fastening system has an important effect on railway safety [1]. Recently, with the continuous development in image processing and deep learning technology, a kind of cameras installed vehicle began to be used for automated inspection. By running on the railway, the camera will capture images, locate the fastening system and analysis the abnormality through computer processing [2]. This technology can effectively reduce the cost of manual work and improve efficiency.

In this technology, accurate locating of the fastening system is very significant. And high precision locating is an essential basis for subsequent abnormality detection [3]. In regard to locating work of the fastening system, Zhang et al. [4] proposed a method based on scale-invariant features to extract the target features, then used Support Vector Machine (SVM) to classification and detection, although the location accuracy of this method is low. Wei et al. [5] adopted Faster R-CNN model which used VGG-16 as backbone network performed a good location accuracy, but the type of target to be detected in their paper is single. Qi et al. [6] proposed a new network model called MYOLOv3-Tiny, which can reduce memory consumption and has a higher detection speed. Their paper pays more attention to real-time detection. The locating work in above papers is all around the targets in regular sections. In fact, the content of the image in turnout sections is more complex. The image in turnout sections is more challenging, as it not only has the situation of multi-railway, but also need to detect the bolt.

The scenes of the railway image data used in this paper can be divided into regular sections and turnout sections. Among them, there have more types of targets and more complicated interference in turnout sections. In order to ensure the detection accuracy in regular sections and resolve the difficulty in turnout sections, this paper proposed an improved model based on Faster R-CNN that

*Corresponding author: xblu2013@126.com

CSAE 2021, October 19–21, 2021, Sanya, China
© 2021 Association for Computing Machinery.
ACM ISBN 978-1-4503-8985-3/21/10…$15.00
https://doi.org/10.1145/3487075.3487184

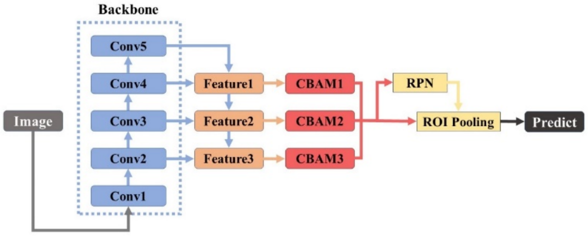

Figure 1: The Architecture of the Model In This Paper.

used multi-scale feature fusion, modified predefined anchor, and add attention model. Experimental results (AP) of the improved Faster R-CNN model reached 98.9%. It achieved high-precision fastening system locating performance both on regular sections and turnout sections.

2 METHOD

The original Faster R-CNN model consists of two stages, including region proposal generation and target prediction. This paper improved Faster R-CNN model, so that it can obtain better performance on railway dataset.

2.1 Improved Faster R-CNN Model

The improved Faster R-CNN model [7] proposed in this paper modified the backbone network and the region proposal network (RPN). The complete improved model architecture is as follows:

The flowchart of the algorithm in Figure 1 shows two stages in this model. Firstly, feed the image into the backbone network based on ResNet-50 [8] for feature extractions. Secondly, the output feature map from the backbone will do multi-scale fusion and enhance the advantageous feature value by using the attention module. Then, send the enhanced feature to RPN and get their region proposal. Thirdly, do ROI pooling on feature map by region proposal received from RPN, acquired proposal feature from every feature map, then send proposal feature to prediction network for classification and regression [9].

2.2 Multi-scale fusion

Generally, in the process of image feature extraction by convolutional neural networks, as the number of layers deepens and with downsampling, the extracted features changed from low-level basic features to high-level semantic features [10]. Prediction through the last layer of feature maps from the backbone may ignore many low-level information, and cause errors in the target location. Therefore, multi-scale feature map information fusion can improve the accuracy of the model. As shown in Figure 2, He et al. [11] proposed a feature pyramid networks (FPN) structure. This structure did not fuse one multi-scale feature map for prediction but predicted independently at every fused feature map in different scales.

In this paper, targets like fasteners and bolts all have the characteristics of small size, single scale change, and large position change. If the feature map with a larger scale on the top layer is used for prediction, it will lead to a prominent location error for small targets in prediction stage [12]. Therefore, this paper used a strategy similar to FPN to adapt small targets.

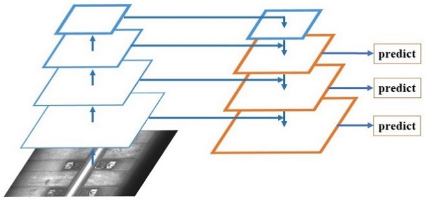

Figure 2: The Strategy of Multi-Scale Fusion in This Paper.

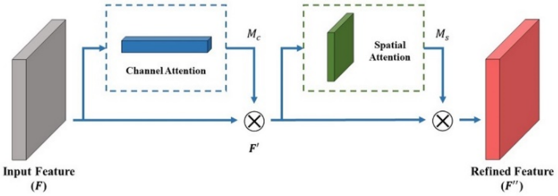

Figure 3: The Structure of CBAM.

In this paper, ResNet-50 is used as the backbone network, but not all four-layer fusion feature maps are used to predict. In order to adapt small targets, only the lower-scale three-layer fusion feature maps are sent to the prediction network. Although the top layer's feature map is not used, its information is still retained and combined with the lower-scale layer, which provides abundant background information [13]. This strategy can both inhibit the fault detection case of large-size objects and avoid the location error due to high-scale feature maps.

2.3 Attention module

In order to further increase the model's attention to the target features in fastening system, suppress the complex features of background and false targets, this paper add attention module to the feature map. The convolutional block attention module (CBAM) is composed of two complementary modules connected to the channel attention module and the spatial attention module, which can effectively suppress background features and enhance target features [14].

As shown in Figure 3, the input feature map$F \in R^{C\times H\times W}$ passes the channel attention module to obtain a one-dimensional channel attention map $M_C \in R^{C\times 1\times 1}$. After doing the weighting process of M_C and F along the channel direction, a new feature map F' is obtained. Then F' pass the spatial attention module is a two-dimensional spatial attention map $M_S \in R^{1\times H\times W}$. After the multiplying process of M_S and F', the final output optimized feature map can be obtained, the formulas are as follows:

$$F' = M_C(F) \otimes F \tag{1}$$

$$F'' = M_S(F') \otimes F' \tag{2}$$

$C$is the number of channels of the input feature map, $H$and W represents the height and width, and $\otimes$ is multiply the corresponding elements of the matrix.

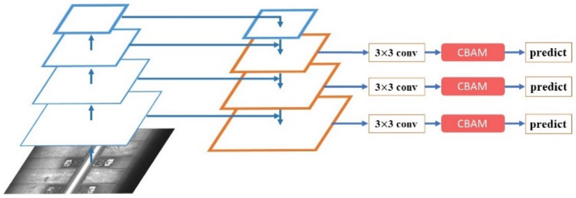

Figure 4: The Structure of Backbone Network with CBAM.

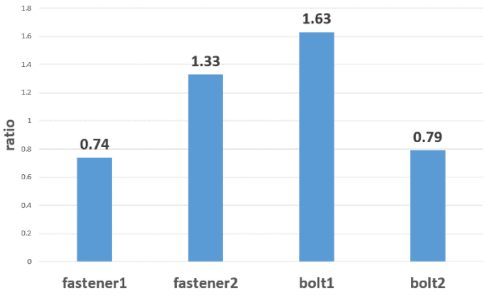

Figure 5: The Actual Ratio of Targets in the Dataset.

This paper added the attention module after three output feature maps (as shown in Figure 4) to make the network focus on meaningful features [15], and finally sent feature maps to the prediction network. The improved feature extraction network structure is as follows:

2.4 Modified Anchor

The RPN of original Faster R-CNN predefined anchor size is {128, 256, 512}, and ratio is {0.5, 1.0, 2.0}. There are 9 types of anchors. Such a setting is more suitable for large target detection. In the scene of this paper, the target to be detected in the fastening system has a small size. Therefore, based on the anchor's original size and ratio, a reasonable anchor setting can make them more similar to the actual target [16]. In this paper, the actual ratio of several types of targets to be detected in the dataset are shown in Figure 5.

Therefore, combined with the actual ratio of targets to be detected, the anchor ratio in this paper is modified to {0.8, 1.2, 1.6}. In view of each downsampling ratio of ResNet-50 model and the actual size of the target, the anchor size is modified to {8, 16, 32, 64, 128}. There are 15 types of modified anchors in the RPN to improve the accuracy of region proposal.

2.5 Loss function

After generating the proposal regions, the model needs to send these proposal regions to classification and bounding boxes regression. Thus, the loss function of RPN training process includes classification loss L_{cls} and regression loss L_{reg}. The formula is as follows:

$$L(p_i,t_i)=\frac{1}{N_{cls}}\sum_i L_{cls}(p_i,\tilde{p}_i)+\lambda\frac{1}{N_{reg}}\tilde{p}_i L_{reg}(t_i,\tilde{t}_i) \qquad (3)$$

$\tilde{p}_i$and $\tilde{t}_i$ represents the label of anchor in point i,$N$is the number of samples.

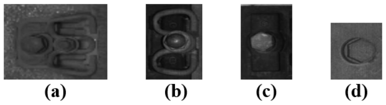

Figure 6: The Type of Targets in the Dataset.

This paper used the cross-entropy function as the classification loss function and the smooth L1 loss function as the regression loss function.

3 EXPERIMENTAL RESULTS

3.1 Experimental dataset

The experiment in this paper used railway inspection images as the dataset and carried out labeled work. The dataset contained 1200 images from the regular sections and turnout sections. The details are shown in Table 1.

There are 13142 labeled targets in this dataset. As shown in Figure 6, 4 types of targets to be detected in the fastening system: (a) is the fastener to be detected in the regular sections, (b) is the fastener in the turnout sections, and (c) (d) are the two types of bolts in the turnout sections:

In order to facilitate the subsequent application of the anomaly detection, all types of targets are labeled "fastener", and the dataset according to 7:1:2 ratio randomly divided into training set, validation set and test set.

3.2 Experimental environment

Experimental hardware configuration is Intel E5-2678 V3 Xeon processor with 64 GB of memory, GPU processor is NVIDIA GeForce RTX2080. Experimental platform is UBUNTU 20.04.2. The Software environment is Python 3.7.9, CUDA 10.0, Pytorch 1.6. And the pretrained weight file of ResNet-50 was used in model training.

3.3 Experimental results and analysis

The experiment in this paper adopts the Average Precision (AP) to evaluate the model. The experiment separately uses three improved methods to create and test models (only use multi-scale fusion, only modified anchor, only add CBAM) and combines the three improvements to create and test the final model (Faster R-CNN-In this Paper) and the model evaluate results are shown in Table 2. The content of turnout sections images is complex, and the predict results are shown in Figure 7-8:

The content of regular sections images is relatively simple, and the improved model also has a high detection accuracy, as shown in Figure 9:

It can be seen from the model evaluation results, the three kinds of modifications of the model in this paper have different degrees of improvement in terms of accuracy. When the IOU is 0.5, the improved Faster R-CNN model compared to the original Faster R-CNN Model, its AP has increased from 96.3% to 98.9%. When the IOU is from 0.5 to 0.95, the improved Faster R-CNN model has

Table 1: The Information of Dataset in This Paper

	Image size	Number	Type of target
Regular	1638×1565	513	1
turnout	2048×2000	687	3

Table 2: The Result of Experimental in This Paper

	AP	
	IOU=0.5	IOU=0.5:0.95
Faster R-CNN	96.3%	77.2%
Faster R-CNN-Multiscale-only	97.4%	81.4%
Faster R-CNN-Anchor-only	98.4%	85.2%
Faster R-CNN-CBAM-only	98.8%	88.1%
Faster R-CNN-In this Paper	**98.9%**	**89.2%**

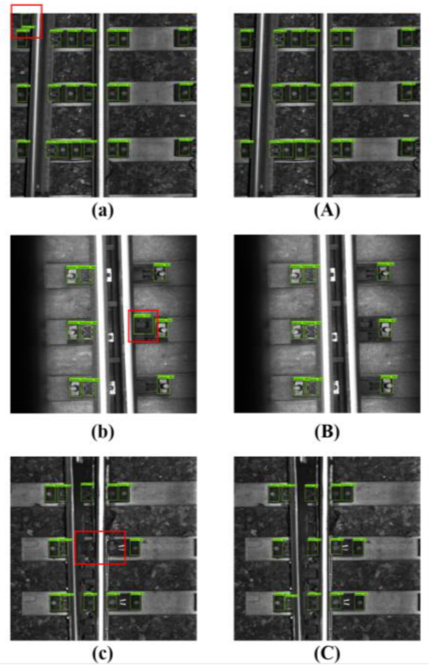

Figure 7: (Left) Original Model (Right) Faster R-Cnn-In This Paper. The Fault Detection and Missed Detection.

increased from 77.2% to 89.2%. Therefore, the improved model has been significantly improved.

As shown in Figure 7-8, it can be seen that the model in this paper has been improved in terms of fault detection, missed detection, and locating accuracy in turnout sections. It is observed that the addition of attention modules can make the model pay more attention to the target area and improve the problem of target missed detection. Besides, modifying the predefined anchors and canceling prediction at the highest scale features can make the location result more accurate. Finally, as shown in the Figure 9, the improved model also shows an accurate detection in regular sections.

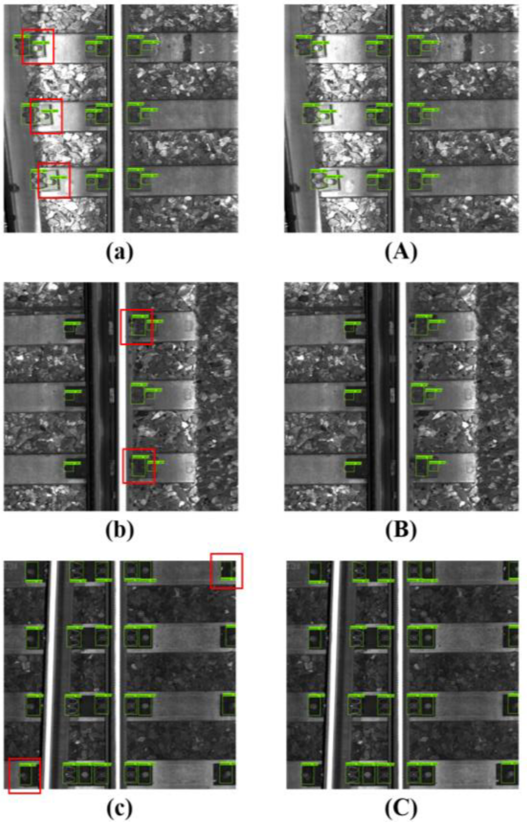

Figure 8: (Left) Original Model (Right) Faster R-CNN in This Paper. Location of Bounding Box.

4 CONCLUSIONS

This paper proposed a method to solve the problems of fastening system locating in different railway sections. For the four types of targets in regular and turnout sections, this paper adopted a

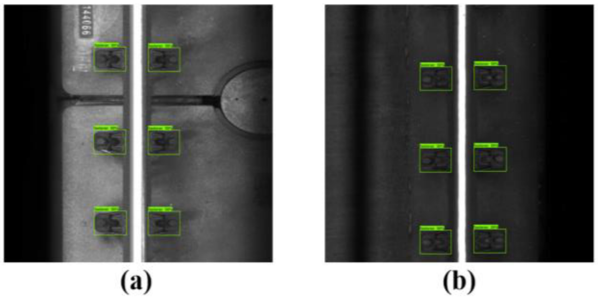
(a) (b)

Figure 9: Detection Results of Regular Sections.

multi-scale fusion strategy similar to FPN based on the Faster R-CNN model, which improved the small target detection accuracy. Besides, the anchor was modified and the attention module was added to the network. Compared with the original Faster R-CNN, the improved model's experimental results (AP) increased from 96.3% to 98.9%. It showed better accuracy and provided an accurate and stable method for target detection of the fastening system in railway. However, this paper only tested four types of targets in regular and turnout sections. In fact, the types of targets to be detected in turnout sections are complicated. The algorithm in this paper needs to be further optimized and tested for more types of targets.

ACKNOWLEDGMENTS

This work is supported by National Key R&D Program of China under Grant 2020YFB1600700 and Major scientific research projects of China Railway Group under Grant K2019G046.

REFERENCES

[1] Sadeghi, J., Najar, M. M., Zakeri, J. A., & Kuttelwascher, C. (2019). Development of railway ballast geometry index using automated measurement system. Measurement, 138, 132-142.

[2] Gibert, X., Patel, V. M., & Chellappa, R. (2016). Deep multitask learning for railway track inspection. IEEE transactions on intelligent transportation systems, 18(1), 153-164.

[3] Chen, J., Liu, Z., Wang, H., Nunez, A., & Han, Z. (2017). Automatic defect detection of fasteners on the catenary support device using deep convolutional neural network. IEEE Transactions on Instrumentation and Measurement, 67(2), 257-269.

[4] Anzhong, Z., Xinyang, H., Minyu, J., & Xiukun, W. (2020, August). Multi-target defect detection of railway track based on image processing. In 2020 Chinese Control And Decision Conference (CCDC) (pp. 3377-3382). IEEE.

[5] Wei, X., Yang, Z., Liu, Y., Wei, D., Jia, L., & Li, Y. (2019). Railway track fastener defect detection based on image processing and deep learning techniques: A comparative study. Engineering Applications of Artificial Intelligence, 80, 66-81.

[6] Qi, H., Xu, T., Wang, G., Cheng, Y., & Chen, C. (2020). MYOLOv3-Tiny: A new convolutional neural network architecture for real-time detection of track fasteners. Computers in Industry, 123, 103303.

[7] Ren, S., He, K., Girshick, R., & Sun, J. (2015). Faster r-cnn: Towards real-time object detection with region proposal networks. arXiv preprint arXiv:1506.01497.

[8] He, K., Zhang, X., Ren, S., & Sun, J. (2016). Deep residual learning for image recognition. In Proceedings of the IEEE conference on computer vision and pattern recognition (pp. 770-778).

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

[10] Cao, C., Wang, B., Zhang, W., Zeng, X., Yan, X., Feng, Z., ... & Wu, Z. (2019). An improved faster R-CNN for small object detection. IEEE Access, 7, 106838-106846.

[11] Lin, T. Y., Dollár, P., Girshick, R., He, K., Hariharan, B., & Belongie, S. (2017). Feature pyramid networks for object detection. In Proceedings of the IEEE conference on computer vision and pattern recognition (pp. 2117-2125).

[12] Xu, H., Yao, L., Zhang, W., Liang, X., & Li, Z. (2019). Auto-fpn: Automatic network architecture adaptation for object detection beyond classification. In Proceedings of the IEEE/CVF International Conference on Computer Vision (pp. 6649-6658).

[13] Ghiasi, G., Lin, T. Y., & Le, Q. V. (2019). Nas-fpn: Learning scalable feature pyramid architecture for object detection. In Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (pp. 7036-7045).

[14] Woo, S., Park, J., Lee, J. Y., & Kweon, I. S. (2018). Cbam: Convolutional block attention module. In Proceedings of the European conference on computer vision (ECCV) (pp. 3-19).

[15] Fu, H., Song, G., & Wang, Y. (2021). Improved YOLOv4 Marine Target Detection Combined with CBAM. Symmetry, 13(4), 623.

[16] Eggert, C., Brehm, S., Winschel, A., Zecha, D., & Lienhart, R. (2017, July). A closer look: Small object detection in faster R-CNN. In 2017 IEEE international conference on multimedia and expo (ICME) (pp. 421-426). IEEE.