ingegneria sismica
International Journal of Earthquake Engineering
Anno XLIII - Num. 2 - 2026

Research on the application of deep convolutional neural network in the defect detection of subway bogie

Baohua Jing[1,*] *and Yufeng Jiang*[2]

[1] School of Mechanical Engineering and Transportation, Changzhou Vocational Institute of Industry Technology, Changzhou, Jiangsu, 213164, China
[2] Crrc Changzhou Tech-mark Industrial Co., Ltd., Changzhou, Jiangsu, 213125, China

SUMMARY: *Due to the dark and humid underground environment, large subway loads and frequent braking, the bogies of subway trains are prone to failures that seriously jeopardize the safety of public transportation. In this paper, the shortcomings of RCNN's time-consuming training are solved by introducing Fast, which uses a shared convolutional layer and extracts the region where the target may exist for learning. It is proposed to improve the feature extraction process of Faster-RCNN algorithm by lightweight network by changing the traditional convolutional network structure, using MobileNet, attention mechanism model, and Unet network structure, and experimentally analyze the subway bogie defect detection of the improved Faster-RCNN. The processed Faster-RCNN image is mainly distributed between gray value 50 and 125, the proportion of which is 0.153 and 0.075, respectively, and the visual effect of local details of the image is more obvious, and the defects are easier to identify. The improved Faster-RCNN network model has improved the leakage detection rate of large-scale defects such as A4, A5, C4, C5, etc. The leakage detection rate of A4 and A5 is 0%, and the leakage detection rate of C4 and C5 is 2.63% and 3.57%, respectively, which has a better effect of detecting defects of subway bogies.*

KEYWORDS: *Faster-RCNN; MobileNet; Attention mechanism; Unet network structure; Metro bogie; Defect detection*

1 Introduction

In the context of the swift advancement of the urban rail transit sector, the subway has emerged as an essential mode of transportation for people's daily commutes. The bogie, which serves as the running component of the subway car, has an operational state that is directly linked to the safety of train operations [1]. Nevertheless, during the extended subway operation period, the environment inevitably exerts an influence on the bogie. Conducting continuous and efficient defect detection on the bogie is a crucial safeguard for ensuring the safe operation of the subway [2, 3].

The fault detection approach serves as the cornerstone of the entire detection system. Conventional detection methods typically depend on manual expertise. Maintenance staff conduct regular inspections of the crucial parts of the bogie to assess the wear and tear of components, as well as the presence of cracks and other issues [4, 5]. Nevertheless, this approach is highly subjective and inefficient, and it is challenging to identify certain latent faults [6].With the advancement of Artificial Intelligence (AI), detection methods leveraging

*jbhjava@163.com
https://doi.org/10.65102/is2026599
Published: 30 April, 2026

AI technology are gradually coming to the fore. Regarding the application of AI technology in detecting defects in metro bogies, literature [7] underscores the significance of early detection of loose bolts in metro bogies. To address the problems of leakage and a high false - detection rate in traditional machine - vision methods, a two - stage deep - learning detection method is put forward. First, the bogie and bolts are located using the SSD. Subsequently, semantic segmentation is carried out via U - Net to determine loosening. Validation using images from Shanghai Metro Line 9 demonstrates that this method can notably enhance the detection accuracy and ensure the stable operation of the vehicle.Literature [8] indicates that the bogie is a vital component of the train, and its malfunction can impact operational safety. A deep neural network diagnosis method is proposed. By comparing the traditional neural network with tests based on vibration data, it is discovered that the average diagnostic accuracy of this method reaches 98.3%, and it has a rapid convergence speed. This highlights the irreplaceable advantage of the deep neural network in the intelligent diagnosis of bogies.In view of the large scale and high training costs of deep - learning networks, literature [9] presents a multi - channel signal analysis method based on dynamic time - regularized KMedoids clustering. A lightweight clustering blueprint separable convolutional neural network (CBS - CNN) is constructed for bogie fault diagnosis. Through end - to - end training, the effectiveness of this network in detecting axle performance degradation, faults, and composite faults is verified. Experiments show that the CBS - CNN has good environmental adaptability.To tackle the problems of noise interference, insufficient features, and poor generalization with small samples in bogie fault diagnosis, literature [10] proposes a three - phase deep - learning method that combines variational mode decomposition, continuous wavelet transform, and a multi - level feature fusion network. Experiments show that the accuracy rate of this method reaches 98.98% with 70% of training samples and remains at 95.96% when the sample size is less than 20%. This indicates strong noise resistance, feature discrimination ability, and adaptability to small samples.

To meet the requirements of bogie condition monitoring under high - speed and heavy - load scenarios, the research in literature [11] developed a flexible dual - function sensor for gathering vibration and pneumatic signals. A recognition model founded on a two - channel one - dimensional residual network was also put forward. Experimental results indicated that this model achieved an average recognition rate of 92.3% for seven typical conditions, outperforming other algorithms and demonstrating potential in structural health monitoring. Literature [12] emphasized that fatigue cracks in metro bogies over their lifespan present a risk to operational safety. A reliability assessment approach that combines dynamic stress measurement and probabilistic methods was proposed. It was discovered that curve segments and rail wave abrasion can induce elastic vibration and modal stress, accelerating fatigue damage. Predictions showed that welded joints would fail after only 340,000 kilometers with 99% reliability, which aligns with the actual situation.Due to the insufficient accuracy of subway bogie strength testing, literature [13] explored testing methods in a big - data environment. First, a three - dimensional model was created using CATIA software to obtain parameters. A test bench was constructed, and strength testing was carried out through a finite element model. Simulation experiments verified that this method has higher detection accuracy compared to traditional methods.Literature [14] explained that the subway bogie axle box bearing is a crucial component for safe operation. A rapid abnormal state identification framework based on machine learning was proposed. After evaluating specific line data, it was found that the gradient boosting method performs best in terms of accuracy, precision rate, and other indicators, making it suitable for practical use.The objective of literature [15] is to evaluate the fatigue failure probability of bogie frames using a data - driven proxy model. This model takes into account parameter variations such as load and

endurance limits. By combining mechanical testing and machine learning to generate a dataset, it was found that the CatBoost model can optimally map safety coefficients and conduct reliability assessment with limited data points. The results show that the fatigue reliability of the frame under normal conditions is 99.34%, providing guidance for applying machine learning in the fatigue reliability analysis of complex vehicle structures.Literature [16] points out that manual inspection suffers from issues such as low efficiency and a high error rate. An intelligent inspection method based on machine vision and image processing was proposed. Through real - time data acquisition and online monitoring, this method overcomes the limitations of traditional manual inspection, enabling rapid feedback of inspection results and efficient and accurate inspection, thus significantly improving efficiency.However, existing AI detection methods in fault location, diagnosis, signal processing, and fatigue assessment still face problems such as detection accuracy being affected by noise interference and high model training costs. On this basis, the deep convolutional neural network (DCNN) method has been gradually adopted.DCNN is a vital technology for image processing and pattern recognition. It is a special neural - network - based structure mainly composed of multiple convolutional layers, pooling layers, and fully connected layers [17, 18]. In image processing, shallow convolutional layers can extract low - level features like edges and textures, while deeper convolutional layers can extract more abstract high - level features, such as object shapes and semantic information [19, 20]. For instance, when processing natural scene images, a shallow convolutional layer can detect edges and lines, while a deep convolutional layer can recognize object classes [21, 22]. In subway bogie defect detection, DCNN combines data enhancement and migration learning techniques by constructing a multi - scale feature extraction network, which can accurately identify defects such as cracks and corrosion on the bogie surface, and is of great importance for the safe operation of subways [23 - 25].

In this research paper, the Faster - RCNN algorithm is developed to process a whole image. First, the image undergoes convolution, followed by the extraction of areas where the target could potentially be located. The Region Proposal Network (RPN) is employed to identify regions within the input image where the algorithm can learn the target's features. Subsequently, the Faster - RCNN target detection module is utilized for detection purposes.To optimize the detection process, MobileNet is applied to modify the conventional convolutional network architecture. This modification enables the achievement of superior detection outcomes while consuming the same amount of resources.When taking into account the spatial relationship between the object and its surroundings, a better network performance can be achieved through the implementation of an attention model structure. Additionally, the semantic segmentation approach commonly used in medical images is incorporated. Moreover, the relu activation function in the original Faster - RCNN algorithm is replaced with the elu activation function. These changes further enhance the effectiveness of defect detection.

2 Metro Bogie Defect Detection Based on Faster-RCNN

2.1 Faster-RCNN modeling

In the case of conventional neural networks, when the resolution of the input image is high, the number of parameters needed for fitting grows exponentially, and the training process takes a longer time. In contrast to these conventional neural networks, convolutional neural networks offer distinct benefits, namely local perception, parameter weight sharing, and the presence of a pooling layer within the network. As a result, the number of parameters in a

convolutional neural network is significantly lower compared to neural networks of the same scale. Moreover, the training time is also reduced to some degree. This enhancement has enabled convolutional neural networks (CNNs) to become the most prevalently used deep - learning algorithms in the domains of image classification and image recognition within the field of computer vision.

RCNN is the first algorithm that introduces convolutional neural network into the field of target detection, and its effect is much better than that of the same period of target detection algorithms. Since the release of RCNN, its epoch-making target detection effect has made the overall research ideas of algorithms in the field of target detection dominated by convolutional neural network detection.RCNN's main feature is the fast detection speed, and traditional target detection algorithms use a sliding window to scan all the regions of the input image once, and make judgments on classification while scanning. The RCNN algorithm, on the other hand, pre-selects some regions where the target may exist, and then uses a convolutional neural network to make judgments on these regions.

Fast-RCNN is an upgraded version of the R-CNN algorithm, Fast-RCNN mainly solves the shortcomings of RCNN training time consuming too long, R-CNN in each possible target candidate box through the convolutional neural network for feature learning, Fast-RCNN algorithm uses a shared convolutional layer, a complete picture first convolution and then extract the possible target region for learning.

The FasterR - CNN algorithm can be split into two components: the RPN network and the FastR - CNN object detection. Initially, the RPN network is employed to identify the area of the input image within which the algorithm can potentially acquire the target features. Subsequently, the FastR - CNN object detection module is utilized for the detection process.

Loss function:

$$L\left\{\left\{p_i\right\},\left\{t_i\right\}\right\}=\frac{1}{N_{cls}}\sum L_{cls}\left(p_i,p_i^*\right)+\lambda\frac{1}{N_{reg}}\sum_i p_i^* L_{reg}\left(t_i,t_i^*\right) \tag{1}$$

$$L_{cls}\left(p_i,p_i^*\right)=-\log\left[p_i^* p_i+\left(1-p_i^*\right)\left(1-p_i\right)\right] \tag{2}$$

In Equation (1) λ denotes the balance weight, which is set to 10. i is the number of each anchor, p_i denotes the probability that the numbered anchor is a target, p_i^* is the label, which has the value of 0 or 1, t_i denotes the coordinates of the four parameters of the prediction frame, t_i^* denotes the coordinates of the four parameters of the calibration frame, and L_{cls} is the classification loss function. L_{reg} denotes the regression loss function, $L_{cls}(p_i,p_i^*)$ denotes the logarithmic loss of the detected and non-detected targets, N_{cls} denotes the number of anchors, and N_{reg} denotes the size of the small batch of samples, and Fig. 1 shows the flow chart of the Faster-RCNN algorithm.

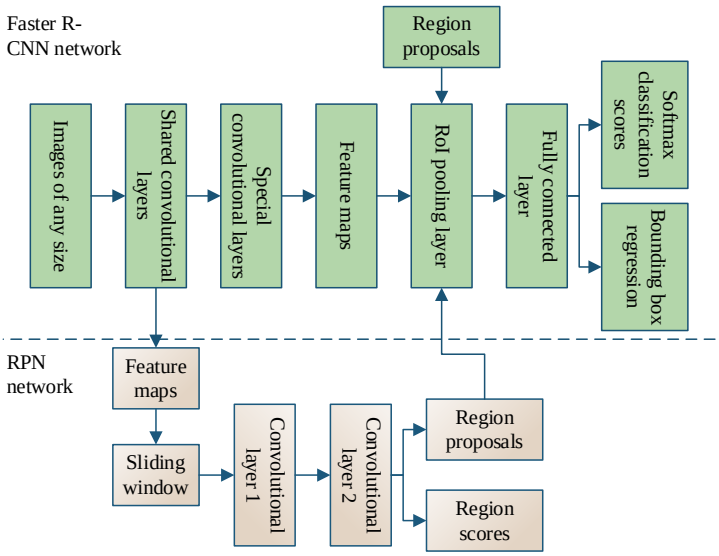

Figure 1: Faster R-CNN Algorithm Flow Chart

2.2 Feature Extraction Improvement Process of Faster-RCNN Algorithm

2.2.1 MobileNet Network Architecture

Lightweight networks drastically reduce the detection time and model size with very small accuracy loss by changing the traditional convolutional network structure, which is favorable for deployment in hardware environments with scarce computational resources.MobileNet is a representative lightweight network.

In the field of target detection, the fundamental convolutional network is employed to extract crucial features from the initial input image. This stage is the primary process influencing both the detection accuracy and the detection time. MobileNet, a lightweight network introduced by Google Inc. in 2017, is designed for deployment on portable devices. Its most notable characteristic is that it employs a depth - separable convolutional module in place of the conventional convolutional layer. This approach effectively decreases the computational load and attains superior detection outcomes with the same level of resource usage.

In the conventional convolutional layer, every channel of the convolutional kernel undergoes a bit - by - bit multiplication with the aggregate of all channels of the input features, followed by an addition operation. The quantity of channels in the convolutional kernel dictates the number of channels in the output features. The overall computational process is:

$$M_c = D_F \cdot D_F \cdot D_k \cdot D_k \cdot M \cdot N \tag{3}$$

where D_F is the feature map size, D_k is the convolution kernel size, and M and N are the number of input and output channels, respectively.

In the depth - separable convolutional layer, the standard convolution process is divided into two phases. First, the number of channels in the depth - separable convolution kernel is equal to the number of input feature channels, and its size is the same as that of the conventional convolution kernel. The convolution operation is performed independently on the corresponding channels. After that, a convolution kernel is used to perform channel merging. The number of channels in the output features depends only on the number of channels in this convolution kernel. At this time, the computational burden is as follows:

$$M_D = D_F \cdot D_F \cdot D_k \cdot D_k \cdot M + D_F \cdot D_F \cdot 1 \times 1 \cdot M \cdot N \tag{4}$$

The ratio of the computational effort of the two is:

$$c = \frac{1}{N} + \frac{1}{D_k \cdot D_k} \tag{5}$$

Thus when $N = 512$ and $D_k = 3$, the computational effort of depth-separable convolution will be reduced by a factor of 8.8 compared to conventional convolution. Of course this comes at the same time at the cost of some loss of accuracy, but the loss of accuracy is small compared to the huge speedup.

2.2.2 Structure of the Attention Model

In real life, when people are carrying out object recognition, the relative positional relationship of objects can be used to assist in the judgment, so as to improve the recognition accuracy in. In convolutional neural networks, the local receptive field size is usually determined by the convolutional kernel size. Simply put, each feature point response value is related to its surrounding pixel points, and when performing object recognition, the network usually only considers the object itself and the information around it, not fully utilizing the positional relationship between the object and the environment.

Although each feature point will be able to correlate to a larger range of feature information as the convolutional layers continue to be stacked, this process is accompanied by a large amount of information loss. On the other hand, increasing the volume and kernel size can enhance the size of the sensory field, but this implies an exponential increase in the amount of computation. Especially when the size of the input image becomes large, too small a receptive field will make the network model very inefficient.

The role of the attention mechanism is to establish the remote dependency between pixels in the image, as well as to extract the weighting coefficients between different features, and to get a better network performance by taking into account the relative position model between objects and the key feature model in the trained network model. The commonly used attention models are channel attention model and spatial attention model.

(1) Principle of Non-local

Non-local is an attention mechanism model that pounces on the dependency of any two positions in space without being limited to neighboring points. Its basic principle is as follows:

$$y_i = \frac{1}{C(x)} \sum_{\forall j} f(x_i, x_j) g(x_j) \tag{6}$$

where x represents the input picture signal and y represents the output response value. $f(*)$ represents the similarity relationship between the current position i and any other position j, which is a distance-dependent function. Commonly used Gaussian function, embedded Gaussian function, Dotproduct, Concatenation to portray. $C(x)$ is the normalization factor that makes the input features at the same scale as the output ones. $g(x_j)$ can be expressed as $g(x_j) = W_g \times x_j$, which represents the eigenvalue of position j plus a learned weight value.The process of the action of the Non-local Attention mechanism can be represented in Figure 2.

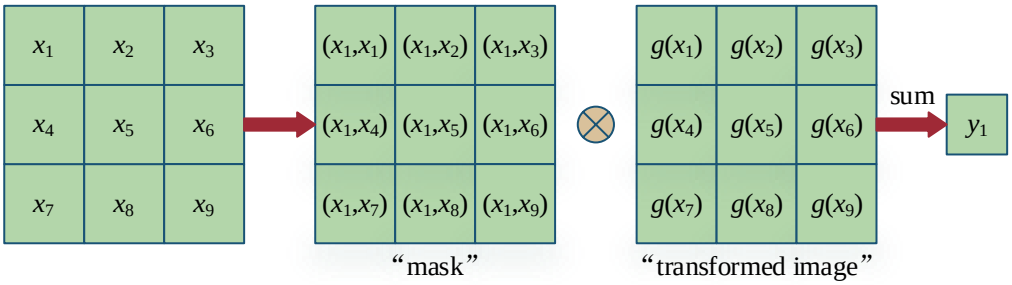

Figure 2: The process of Non-local action

For the response value of input feature x_1, first calculate $f(*)$ to get the "mask" matrix, then calculate $g(*)$ of all input features, and get the output value y_1 by convolution operation and normalization.

The above principle is encapsulated into a module to establish the structure of the non-local attention mechanism as follows:

$$z_i = w_z \cdot y_i + x_i \tag{7}$$

where x denotes the feature input and z denotes the feature output. y denotes the attentional attention module. Its structure is shown in Figure 3.

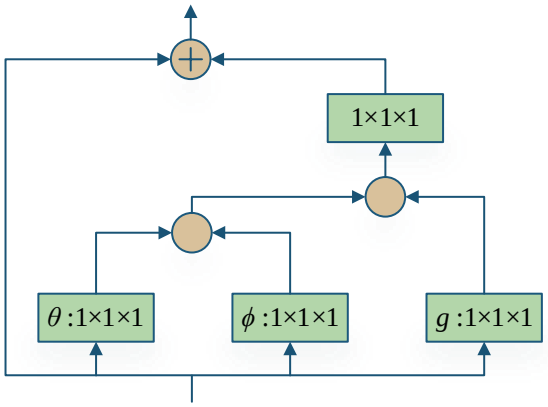

Figure 3: Non-local structure diagram

It represents an embedded Gaussian-based spatial attention module. The input is $T \times H \times W \times 1024$ and the output is of the same size as the input layer for easy embedding into the target network. The $1\times1\times1$ convolution kernel serves to perform compression and stretching of the number of channels. $\oplus$ denotes pixel summation and $\otimes$ denotes matrix multiplication.

(2) CBAM attention module

The Non-local network module calculates the pairwise relationships between the query point and other points, which are weighted to form a global relationship, thus forming a remote feature dependency with the original feature aggregation of the query point. It utilizes the relative positional relationship of objects to effectively improve the detection effect, but pays a lot of computational overhead.

The CBAM module proposes an attention mechanism that focuses on important features in the image and suppresses irrelevant features to improve the model's expressiveness.The

structure of the CBAM network integrates the channel attention module and the spatial attention module in a tandem manner. This can greatly save the parameters and reduce the computation, and also has the convenience of embedding and integrating in other network frameworks.

2.2.3 Improved Faster-RCNN network structure

The loss function of the network includes the loss of RPN and the loss of Faster-RCNN. In the process of network training, the two types of losses are trained separately using an alternating training method. Where the loss of RPN is a measure of the gap between the foreground background and position of the target corresponding to the generated anchor frame in the RPN network and the real situation, and the loss of Faster-RCNN is the gap between the category judgment and position prediction of the output of the classification network and the real sample labeling data. Both types of loss functions can be expressed by the following equation:

$$L(\{p_i\},\{t_i\}) = \frac{1}{N_{cls}} \sum_i L_{cls}(p_i, p_i^*) + \lambda \frac{1}{N_{reg}} \sum_i p_i^* L_{reg}(t_i, t_i^*) \tag{8}$$

where the first term to the right of the equal sign represents the classification loss and the second term represents the regression loss. N_{cls} represents the number of candidate frames and L_{cls} is the cross-entropy loss function. p_i and p_i^* denote the true labeling and predicted categories, respectively. N_{reg} is the number of anchor frames selected in RPN training. p_i^* is used to eliminate regression calculations on the background. L_{reg} denotes the smoothL1 loss function, and t_i and t_i^* denote the true target position versus the predicted target position. Thus the training process produces five loss values, which are RPN classification loss, RPN regression loss, Faster-RCNN classification loss, Faster-RCNN regression loss, and total loss. The total loss represents the sum result of the above four loss values.

2.3 Unet-based improved network

Faster-RCNN can achieve good results for common object recognition, but for the special target such as the signal map of welded seams on the aluminum alloy body of a subway train, the detection effect of directly using Faster-RCNN is not good. Therefore, this paper decides to introduce the same method as semantic segmentation in medical images to improve the detection effect.

The original Faster-RCNN algorithm uses the relu activation function, whose formula is shown in Equation (9). relu can have a very fast computational speed compared to other activation functions, but it may lead to neuron inactivation, and neuron inactivation has no effect in forward propagation, but the gradient will become 0 during backward propagation, resulting in the disappearance of the gradient:

$$relu(x) = \max(0, x) \tag{9}$$

The elu function is a modified version for the relu function, which is defined in equation (10):

$$f(x)=\begin{cases} x & x>0 \\ a(\exp(x)-1) & x\le 0 \end{cases} \tag{10}$$

elu can solve the problem of neuron inactivation in relu and has all the advantages of relu, so in this paper, the relu activation function of the original Faster-RCNN algorithm is changed to elu.

3 Metro Bogie Defect Detection Experiment Based on Improved Faster-RCNN

3.1 Metro Bogie Defect Detection

The ASTM E446 standard is often used to grade the radiographic inspection negatives of subway bogies in NDT. The standard will be subway bogie defects are divided into seven categories, respectively: porosity (A class), sand and slag (B class), shrinkage (C class), cracks (D class), hot tearing (E class), cold iron (F class), mottled (G class). In these seven categories, according to the severity of defects produced by A, B, C classified into 1 ~ 5 a total of 5 levels, while D, E, F, G defects do not have a class distinction.

3.1.1 Image data analysis

In the dataset extracted in this paper, there are two types of defects, namely, porosity class (Class A) and shrinkage class (Class C) defects. Porosity is a hole class defect, which usually shows black circular or elliptical spots in Faster-RCNN detection, and shrinkage is a cluster of tiny and scattered holes, which usually shows a cluster of large black spots in Faster-RCNN detection. Each type of defect is categorized into 5 classes, for a total of 10 defects.

The defect dataset used in this paper has a total of 5,000 images, of which 4,500 are used as the training set and 500 as the test set. Each image in the dataset is 3-channel with a pixel size of 2240 × 2048. Each image may have porosity and shrinkage defects, denoted by A and C, respectively, which are categorized into classes 1 to 5. The 5000 sample data were statistically analyzed for defect type morphology and number, and the results are shown in Fig. 4, which reveals that the images have a high percentage of defects of porosity level 2 and 3, while the defects of shrinkage level 2, 3, and 4 take up a high percentage, and the number of defects of shrinkage level 2, 3, and 4 are 3786, 4896, and 2945, respectively. To address this phenomenon, in order to balance the training samples of porosity and shrinkage and to improve the training effect, the sample equalization process is carried out.

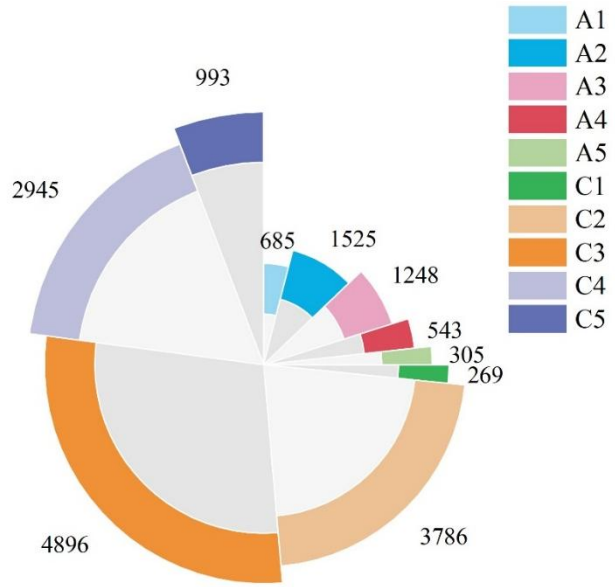

Figure 4: Statistics of the number of different defects

3.1.2 Image pre-processing

The noisy subway bogie image is processed by Faster-RCNN followed by image enhancement, Fig. 5 shows the unenhanced image and its grayscale histogram, and Fig. 6 shows the enhanced image and its grayscale histogram, and a series of defects existing in the subway bogie image can be easily observed from Fig. 6. From comparing the gray level histograms of the images before and after enhancement, it can be seen that the histogram of the enhanced image is more balanced compared to that of the image before enhancement, and the whole image has a wider range of pixel values, which are mainly distributed between gray level values of 50 and 125, with ratios of 0.153 and 0.075, respectively, and the local details of the image are more visually obvious, and the defects are easier to identify.

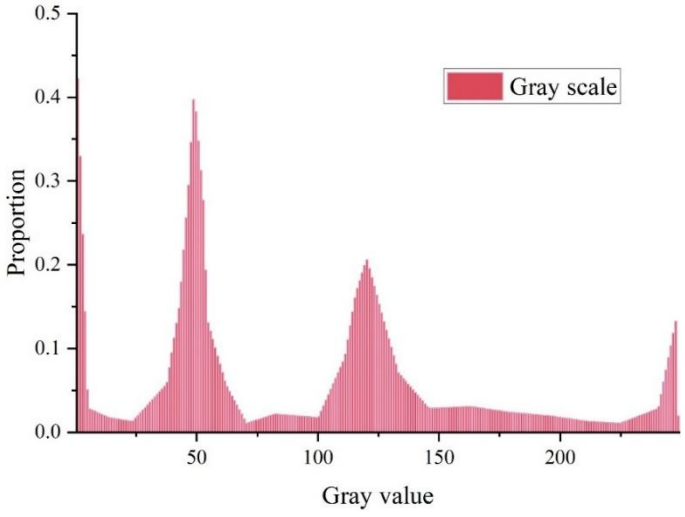

Figure 5: Original image and its grayscale histogram

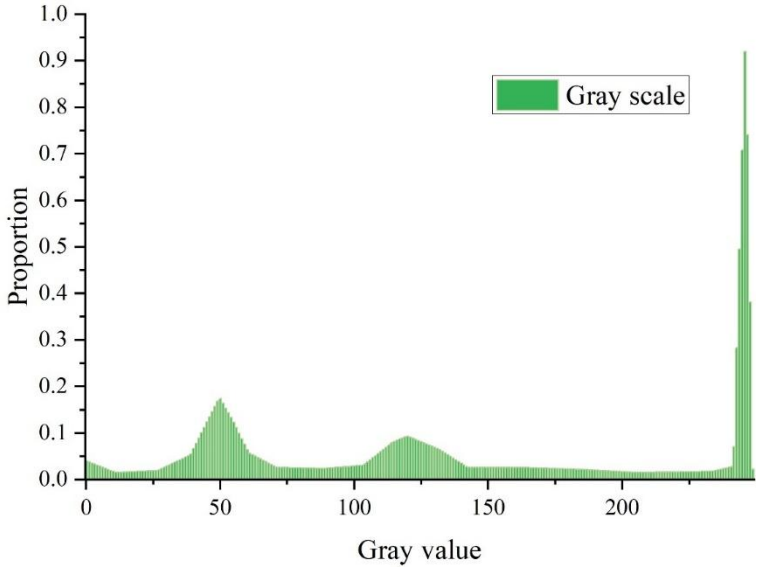

Figure 6: The enhanced image and its grayscale histogram

3.2 Experimental results and analysis

3.2.1 Recall rate

Only for the improvement of the propagation process, this paper has done a comparison experiment on the basis of the original Faster-RCNN, comparing the effect before and after the improvement of the semantic segmentation of the propagation process, and at the same time, also comparing the effect of its experiments in the model of this network, and the comparison results are shown in Fig. 7, which obviously can be found that the improvement of semantic segmentation of the propagation process improves the defective recall rate in the Faster-RCNN model by 1.09%. The defect recall rate in this network model after improving the semantic segmentation of the propagation process is improved by 4.15%. Therefore, in this paper, semantic segmentation is introduced for subway bogie images, which effectively improves the average accuracy of subway bogie part defects and its defect recall rate.

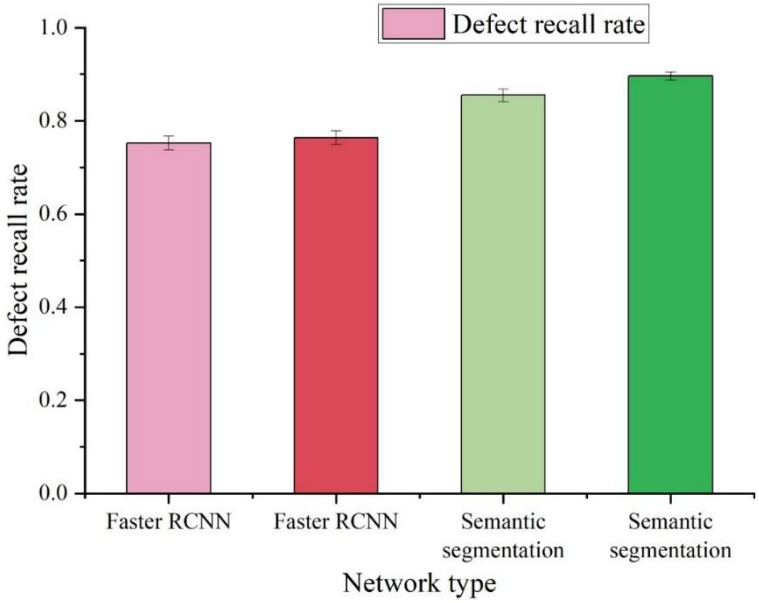

Figure 7: Effect comparison before and after ROI improvement

3.2.2 Loss function

Figure 8 shows the improved network loss function, the loss function did not show large fluctuations at the beginning of convergence, with the increase in the number of training iterations, the training loss accuracy of the model is gradually reduced, the model's loss function curve gradually converged, when the model iterated to 5800 times, the model's loss function value tends to be stabilized, about 1.13 or so, indicating that the model training is more effective to meet the lack of casting training requirements.

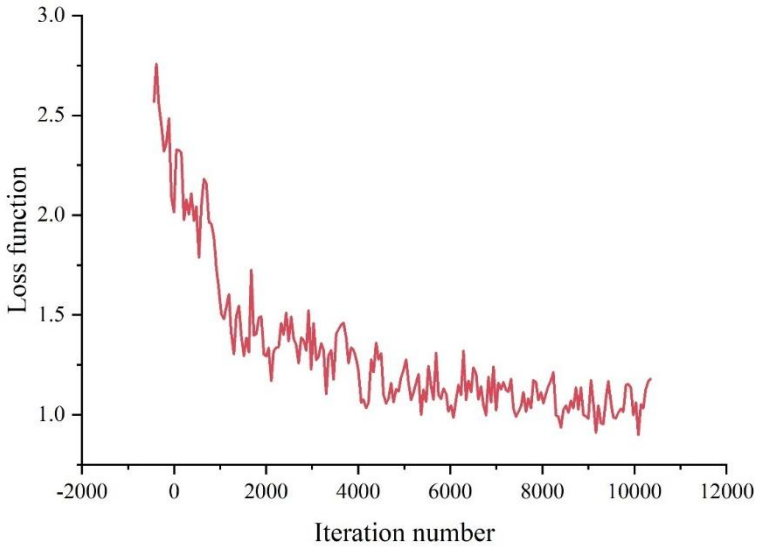

Figure 8: Improved network loss function

3.2.3 Network testing results

100 images are randomly taken out from the original sample without data enhancement as the test set. In this paper, a total of 444 defects are detected and statistically detected, and the detection results are obtained in Tables 1 and 2, where Table 1 shows the detection results of the Faster-RCNN network and Table 2 shows the detection results of the improved network. The horizontal labels in the table represent the correct types of defects and the vertical labels represent the types of detected defects. The bold numbers are the number of correct detections and the rest are the number of misdetected defects. In this paper, the Faster-RCNN is improved for the Faster-RCNN, which effectively improves the model's subway bogie defect detection performance. Before the improvement, for the defects with larger scales such as A4, A5, C4, C5, it is easy to be misdetected, and the misdetection rates are 27.78%, 18.75%, 10.53% and 28.57%, respectively. As for the defects with smaller scales such as A1, A2, A3, C1, C2, C3, etc., both leakage and misdetection will exist, but more often there will be a leakage, and the leakage rate and misdetection rate are between 6.67% to 31.58% and 8.16% to 30% respectively. After the improvement, the leakage rate for large-scale defects such as A4, A5, C4 and C5 is improved, the leakage rate of A4 and A5 is 0%, and the leakage rate of C4 and C5 is 2.63% and 3.57%, respectively.

Table 1: Faster-RCNN network detection results

Detection type	Correct type									
	A1	A2	A3	A4	A5	C1	C2	C3	C4	C5
A1	**10**	4	2	0	0	0	0	0	0	0
A2	2	**8**	4	0	0	0	0	0	0	1
A3	0	2	**20**	3	0	0	0	2	0	0
A4	0	0	0	**10**	2	0	0	0	0	0
A5	0	0	0	2	**11**	1	0	0	0	0
C1	0	0	0	0	0	**9**	6	0	7	0
C2	1	0	2	0	0	1	**74**	8	0	0
C3	0	0	0	0	0	0	2	**85**	0	0
C4	0	0	0	0	0	0	0	5	**60**	7
C5	0	0	0	0	1	0	0	1	1	**20**
Total number of detections	19	20	30	18	16	15	98	124	76	28
Number of missed detections	6	6	2	3	2	4	16	23	8	0
False detections	3	6	8	5	3	2	8	16	8	8

Table 2: Improved network detection results

Detection type	Correct type									
	A1	A2	A3	A4	A5	C1	C2	C3	C4	C5
A1	**13**	3	0	0	0	0	0	0	0	0
A2	1	**16**	1	1	0	0	0	0	0	0
A3	0	0	**29**	0	0	0	0	0	0	0
A4	0	0	0	**15**	2	0	0	0	0	0
A5	2	0	0	0	**14**	1	0	0	0	0
C1	0	0	0	0	0	**14**	3	0	0	0
C2	0	0	0	0	0	0	**95**	2	2	0
C3	0	0	0	0	0	0	0	**120**	0	0
C4	0	0	0	0	0	0	0	2	**69**	2
C5	0	0	0	2	0	0	0	0	3	**25**
Total number of detections	19	20	30	18	16	15	98	124	76	28
Number of missed detections	3	1	0	0	0	0	0	0	2	1
False detections	3	3	1	3	2	1	3	4	5	2

3.3 Distribution of Sample Abnormal Scores

Figure 9 shows the distribution of abnormal scores for positive and negative samples of the improved Faster-RCNN algorithm, which demonstrates the distribution of abnormal values for normal and abnormal samples of the subway bogie. 141 positive samples have abnormal scores that are greater than 3.0 in only 5 samples, and the abnormal scores for the rest of the 136 positive samples are located in the range of 1.1-3.0. In contrast, the anomaly scores of the negative samples are all larger than 3.0, and the anomaly scores of 25 negative samples are even higher than 4.0, which indicates that under the training of positive samples only, the improved Faster-RCNN still has excellent results in categorizing the positive and negative samples for the detection of defects in subway bogies.

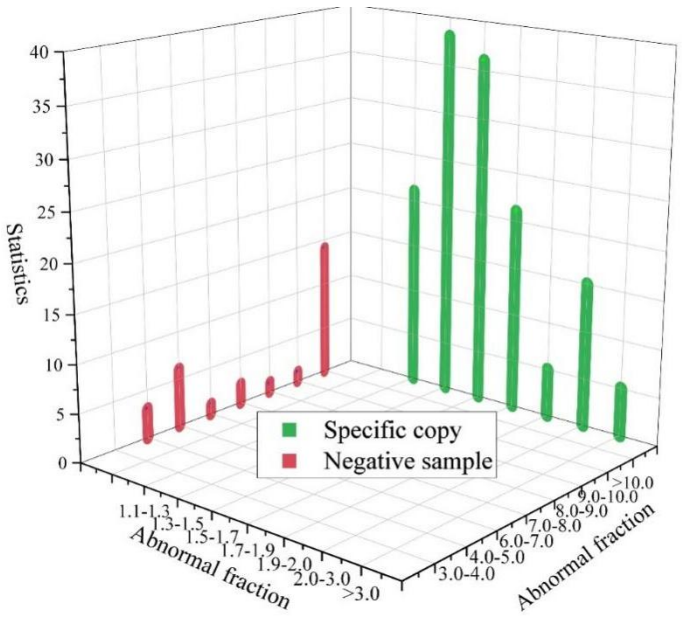

Figure 9: Improved frequency RCNN plus sample anomaly score distribution

3.4 Bogie key components of the subway intermodulation experiment

For the experiments of the clamp item point and the electrical box cover item point, this project adopts a two-stage approach to realize the defect detection of the subway bogie by first locating and then classifying the parts. The YOLOv5-GE algorithm has been used to realize the localization of the parts in the two points, and the localization accuracy is good, and the improved Faster-RCNN algorithm has been used to realize a classification under the condition of less negative samples for the spring parts in the clamp. Three Faster-RCNN networks are also used for the electrical box cover to realize the defect detection for the latch, rectangular marking and triangular marking, in which the coupling method for the electrical box cover item is to count the number of the three kinds of parts in the corresponding category pictures whether they correspond to the category database or not, so as to determine whether there are defects or not. The co-conditioning experiment, as a complete process experiment, needs to combine the target detection stage with the defect detection stage. Since the evaluation indexes of the algorithm performance in each stage are different, the co-tuning experiment adopts the leakage rate and the false alarm rate as the evaluation indexes. False alarm rate here refers to the ratio of the number of positive samples with wrong prediction results to the positive samples in the data set. The underreporting rate refers to the ratio of the number of negative samples with wrong prediction to the negative samples in the dataset. For the clamp item point, due to the one-classification algorithm used for defect detection, selecting a lower threshold will lead to an increase in the false alarm rate, and a higher threshold may lead to the occurrence of underreporting, so the clamp item point adopts a minimum threshold of zero underreporting and focuses on the false alarm rate. The number of negative samples for the joint commissioning experiments of the clamp item point is also added with the negative samples of manual simulation, and the experimental results are shown in Table 3, which show that the false alarm rate is lower than 5% and the leakage rate is 0 in the joint commissioning test set.In summary, finally, after the joint commissioning, the leakage rate of the clamp item point and the electrical box cover item point are both 0, and the false alarm rate is 4.58% and 2.96%, respectively, which has a good effect on the defect detection.

Table 3: Statistics of joint debugging results

Project Name	Positive sample size	Negative sample size	False alarm rate	Underreporting rate	Single sheet test time(GPU)
Top of the clamp	348	45	4.58%	0	0.569s
Electrical box cover items	296	25	2.96%	0	1.025s

4 Conclusion

In this research paper, the Faster - RCNN algorithm is developed by enhancing the RCNN model. The model can be divided into two parts. First, the RPN network is employed to identify the areas in the input image where the algorithm can potentially learn the target features. Subsequently, the Faster - RCNN target detection module is utilized for detection.To efficiently decrease the computational load and attain superior detection outcomes with the same resource utilization, this paper successively refines the feature extraction process of the Faster - RCNN algorithm through methods such as MobileNet, the attention mechanism, and semantic segmentation. Additionally, the relu activation function in the original Faster - RCNN algorithm is replaced with the elu activation function.A feature detection algorithm based on the improved Faster - RCNN is proposed. In the subway bogie defect simulation detection experiment, image data analysis reveals that shrinkage defects at levels 2, 3, and 4 account for a large proportion. The quantities of these defects are 3786, 4896, and 2945 respectively.Coupling experiments on the key components of the subway bogie are conducted. For the clamp item point and the electrical box cover item point, the leakage rate is 0, and the false - alarm rates are 4.58% and 2.96% respectively. Evidently, the Faster - RCNN algorithm demonstrates a favorable defect detection effect.

About the Author

Baohua Jing was born in Zhenjiang, Jiangsu, P.R. China, in 1979. He obtained a master's degree from Nanjing University of Science and Technology in China. His main research direction is Data mining and urban rail transit informatization.

Yufeng Jiang was born in Changzhou, Jiangsu, P.R. China, in 1982. He obtained a bachelor's degree from Changzhou Institute of Technology in China. His main research direction is Assembly and debugging of key components for Fuxing high-speed EMU.

References

[1] Wang, B., Xie, S., Jiang, C., Song, Q., Sun, S., & Wang, X. (2020). An investigation into the fatigue failure of metro vehicle bogie frame. Engineering Failure Analysis, 118, 104922.

[2] Li, J., Wang, J., Li, X., Yang, J., & Wang, H. (2015). The experiment study for fatigue strength of bogie frame of Beijing subway vehicle under overload situation. The Open Mechanical Engineering Journal, 9(1), 260-265.

[3] Sun, H., He, D., Zhong, J., Jin, Z., Wei, Z., Lao, Z., & Shan, S. (2023). Preventive maintenance optimization for key components of subway train bogie with consideration

of failure risk. Engineering Failure Analysis, 154, 107634.

[4] Wróbel, J., Bury, P., Zając, M., Kierzkowski, A., Tubek, S., & Blaut, J. (2024). Fault Detection and Diagnostic Methods for Railway Systems–A Literature Survey. Advances in Science and Technology. Research Journal, 18(6).

[5] Shadfar, M., & Molatefi, H. (2022). Detection of rail local defects using in-service trains. Proceedings of the Institution of Mechanical Engineers, Part F: Journal of Rail and Rapid Transit, 236(9), 1114-1123.

[6] Li, X., Zhang, W., Wang, Z., Hua, S., & Luo, W. (2025). Research on multiaxial fatigue life of metro bogie frame with welding defects. Proceedings of the Institution of Mechanical Engineers, Part F: Journal of Rail and Rapid Transit, 239(10), 831-841.

[7] Qi, W., Zheng, S., Li, L., & Yang, Z. (2022). Loosening bolts detection of bogie box in metro vehicles based on deep learning. IEICE TRANSACTIONS on Information and Systems, 105(11), 1990-1993.

[8] Zhao, Y., Guo, Z. H., & Yan, J. M. (2017). Vibration signal analysis and fault diagnosis of bogies of the high-speed train based on deep neural networks. Journal of vibroengineering, 19(4), 2456-2474.

[9] Jia, X., Qin, N., Huang, D., Zhang, Y., & Du, J. (2022). A clustered blueprint separable convolutional neural network with high precision for high-speed train bogie fault diagnosis. Neurocomputing, 500, 422-433.

[10] Lv, H., Liang, R., Gu, Z., Miao, H., Zhang, D., & Jiang, Z. (2025). Fault diagnosis method for train bogie axlebox based on deep learning feature fusion of CWT-MVGG. Journal of Vibration and Control, 10775463251399722.

[11] Dong, W., He, Z., Xiong, K., & Yang, L. (2025). Dual channel 1D residual network for condition monitoring of high-speed train bogie with flexible dual functional sensors. Sādhanā, 50(3), 224.

[12] Wang, B., Sun, S., Ma, S., & Wang, X. (2020). Fatigue damage and reliability assessment of subway train bogie frames under operating conditions. Advances in mechanical engineering, 12(1), 1687814020903590.

[13] Shi, W., Hai-tao, H., Ye-ming, Z., Yu-guang, W., Wei, Z., & Feng, J. (2020, September). Strength Detection Method for Subway Vehicle Bogie Frame in Big Data Environment. In International Conference on Advanced Hybrid Information Processing (pp. 442-451). Cham: Springer International Publishing.

[14] Liu, Y., Wang, Y., Han, Y., & Hu, H. (2025, July). Research on Abnormal State Identification of Axle Box Bearings in Subway Vehicle Bogies Based on Feature Extraction. In Proceedings of the 2025 2nd International Conference on Image Processing, Intelligent Control and Computer Engineering (pp. 339-343).

[15] Sanikhani, S. A., Soroush, M., Alizadeh Kiani, K., Ravandi, M., & Rezaeian Akbarzadeh, M. (2024). Probabilistic machine learning approach to reliability analysis of a bogie frame under dynamic loading. International Journal of Rail Transportation,

12(5), 958-978.

[16] Mao, K., Mo, L., Zhan, H., Jia, S., & Yang, Y. (2024, September). Intelligent Detection Scheme Design of Crucial Components of High Speed Train Bogie Based on Machine Vision. In 2024 IEEE 7th International Conference on Information Systems and Computer Aided Education (ICISCAE) (pp. 81-87). IEEE.

[17] Rawat, W., & Wang, Z. (2017). Deep convolutional neural networks for image classification: A comprehensive review. Neural computation, 29(9), 2352-2449.

[18] Liu, M., Shi, J., Li, Z., Li, C., Zhu, J., & Liu, S. (2016). Towards better analysis of deep convolutional neural networks. IEEE transactions on visualization and computer graphics, 23(1), 91-100.

[19] Zhou, D. X. (2020). Theory of deep convolutional neural networks: Downsampling. Neural Networks, 124, 319-327.

[20] Khan, A., Sohail, A., Zahoora, U., & Qureshi, A. S. (2020). A survey of the recent architectures of deep convolutional neural networks. Artificial intelligence review, 53(8), 5455-5516.

[21] Xin, M., & Wang, Y. (2019). Research on image classification model based on deep convolution neural network. EURASIP Journal on Image and Video Processing, 2019(1), 1-11.

[22] Zhang, F., Cai, N., Wu, J., Cen, G., Wang, H., & Chen, X. (2018). Image denoising method based on a deep convolution neural network. IET Image Processing, 12(4), 485-493.

[23] Man, J., Dong, H., Jia, L., & Qin, Y. (2022). AttGGCN model: A novel multi-sensor fault diagnosis method for high-speed train bogie. IEEE Transactions on Intelligent Transportation Systems, 23(10), 19511-19522.

[24] Yang, C., Yang, L., Guo, W., & Xu, P. (2023). Deep learning based structural damage identification for the strain field of a subway bolster. Alexandria Engineering Journal, 81, 264-283.

[25] Wang, J., Li, L., Zheng, S., Zhao, S., Chai, X., Peng, L., ... & Tong, Q. (2023). A Detection Method of Bolts on Axlebox Cover Based on Cascade Deep Convolutional Neural Network. CMES-Computer Modeling in Engineering & Sciences, 134(3).