Track Defect Detection Based on Neural Network

Zhengfeng Nan, Yan Xu*, Wenrui Cao

*Electronics and Information Engineering School, Lanzhou Jiaotong University, No.88, Anning Road, Lanzhou 730070, China

*15735169321@163.com

Abstract. The maintenance and inspection of the high-speed railway is the core problem of railway safety. To solve the problems of insufficient network depth and low defect extraction ability, an improved Unet network model was proposed. In view of the relatively small defect area, the attention mechanism is used to effectively suppress the background and highlight the significant features of the defect. The experimental results show that the improved Unet model improves the comprehensive detection accuracy of defect targets by 6% compared with similar segmentation models.

Keywords: Visual detection, neural network, Unet, Attention

1. Introduction
By the end of 2019, the length of China's railways in operation had reached 139,000 kilometers, including 35,000 kilometers of high-speed railways, an increase of 6,000 kilometers over the end of last year. Therefore, railway safety is particularly important, and the inspection and maintenance of the railway also become one of our concerns. On the other hand, due to the continuous expansion of the railway scale, the manual inspection mode commonly used at present not only consumes a lot of human resources but also has a long maintenance cycle, which cannot match the current railway scale in China. Therefore, according to the current needs of railway overhaul and in combination with the direction of intelligent development, Therefore, according to the current needs of railway maintenance, combined with the direction of intelligent development, it is the current research trend to propose an efficient and accurate detection method that adapts to the development of railways [1].

As computers are widely used in the field of image processing, railway detection technology is also rapidly updated. Foreign countries apply image processing technology in the field of rail parameter and defect detection. Hajizadeh S. et al. used Gabor filters to extract railway track surface features from multiple directions and then compared the characteristics of the non-destructive area of the track surface with the defect area, finally achieve the purpose of accurate segmentation of the rail surface defects[2]. In recent years, with the continuous deepening of artificial intelligence research, visual inspection has also been converted from theoretical research to practical applications, and detection models and algorithms suitable for various applications have been proposed, and a large number of results have emerged. In terms of railway inspection, developed countries such as Germany, Italy, and France have also introduced track inspection equipment [3]. Compared with foreign countries, the research time of visual detection in China is relatively short. Meng Jia et al. proposed a rail defect detection method, which is divided into three parts: rail image acquisition, defect parameter...
extraction, and track defect recognition [4]. After that, the Chinese Academy of Railway Sciences applied image processing technology on the track inspection car to develop an intelligent defect inspection system [5]. However, due to the complex and changeable actual railway operation conditions, and the need to process massive amounts of image information, at the same time, it is necessary to improve the speed and accuracy and reduce the missed detection rate and false alarm rate, the current detection methods are difficult to adapt to practical needs.

Due to the progress of deep learning in the field of target detection and image segmentation and its demonstrated advantages, more and more researchers apply it to the field of target detection [6] and image segmentation [7]. Aiming at Unet network [8], this paper integrates attention networks to improve network segmentation accuracy and extract higher semantic information. At the same time, it makes it easier for the low-level semantic information to fuse with the high-level semantic information, and also reduces the incoming useless parameters. Finally, the experimental results show that the improved network structure can improve the recall rate and detection accuracy of defect targets, in addition, the mIOU can also be improved.

2. Unet network
As shown in Figure 1, similar to the FCN model [9], the U-Net network model also uses a network structure with two stages of code. The down sampling represents feature extraction, solves the locating problem of determining pixels, and can also increase the robustness of the input image. The up-sampling solves the problem of image classification and recovers the image details by fusing the corresponding layer information of the two parts of encoding and decoding to improve the sensitivity to the segmented object edge details [10]. Different from the FCN model, the upper and lower sampling stages of U-Net use the same number of convolutional layers, connected in the middle by a jump structure, so as to combine the low-level high-resolution and low-level semantic information with the high-level, low-resolution, and high-semantic information, so that the overall segmentation effect of U-Net network is better.

![Fig.1 Unet network model](image)

3. Improved Unet network model
The essence of Unet network training is to predict the test set according to the characteristics of the training set provided, so once there is a big difference in data distribution between the two sets, the generalization ability of the network will be weakened. On the other hand, for improper data distribution among batches, each iteration of the network needs to update the distribution and reduce
the training speed. Therefore, this paper added batch Normalization to the input data before convolution and then sent it to the convolutional layer through relu activation function, so as to better fit the target and accelerate the convergence speed of the network.

Calculate the mean of the batch:

$$\mu_B = \frac{1}{m} \sum_{i=1}^{m} x_i$$  \hspace{1cm} (1)

the batch input is $x : B = \{x_1, \ldots, x_m\}$

Calculate the variance of Mini-batch:

$$\sigma_B^2 = \frac{1}{m} \sum_{i=1}^{m} (x_i - \mu_B)^2$$  \hspace{1cm} (2)

Normalization:

$$\tilde{x}_i = \frac{x_i - \mu_B}{\sqrt{\sigma_B^2 + \varepsilon}}$$  \hspace{1cm} (3)

$\varepsilon$ is a constant that keeps the data stable in the variance.

4. Attention mechanism

The image attention mechanism is similar to the human visual attention mechanism [11]. When humans read an article, they will pay attention to its title or illustration at first, then they will pay more attention to the key areas, suppress other useless information, quickly collect valuable information from the complex materials, and improve efficiency.

Image attention also takes advantage of the feature of focus attention. When it is aimed at the predicted image, local information will be highlighted, while background and noise information will be suppressed. In this paper, attention is used to improve the weight of the location of the defect features by suppressing the feature activation of the background area during detection, so that more information can be collected and the sensitivity and accuracy of the model to the defect area can be improved. Figure 2 shows the attention mechanism.

![Fig. 2 Attention mechanism](image)

The expression of attention is as follows:

$$q_{\text{att}}^l = \psi^T \left( \sigma_1 (W_{x}^T x_{\text{att}}^l + W_{g}^T g_{\text{att}} + b_{g}) \right) + b_{\psi}$$  \hspace{1cm} (4)

$q_{\text{att}}$ is the attention output, $l$ is the layer $l$, $\sigma_1$ is the RELU excitation function, $W_{x}, W_{g}$ is the matrix associated with $x, g, b_{g}, b_{\psi}, \psi$, is the bias term.

$$\alpha_{\text{att}}^l = \sigma_2 (q_{\text{att}}^l (x_{\text{att}}^l; \theta_{\text{att}}))$$  \hspace{1cm} (5)

$\alpha_{\text{att}}$ is the attention coefficient, $\sigma_2$ is the Sigmoid excitation function and $\theta_{\text{att}}$ is the network learning parameter.

5. Experimental results and analysis
5.1. **Experimental environment and parameters**

The improved Unet network experiments, FCN network, and Unet network comparison experiments in this paper are all conducted under the deep learning framework Keras. The network model was trained by the Geforce GTX 1080Ti 11GB GPU, combined with python3.6 and OpenCV image processing, and programming experiments were carried out on the network model.

The network learning rate was set to 0.0001, and the Adam optimization algorithm was selected for iterative optimization. The small batch random gradient descent algorithm was selected. The batchsize was set as 4 and the epoch was set as 50.

5.2. **Experimental data set selection**

In order to make the improved Unet model mentioned in this paper can be better applied to the field of railway defect detection, the use of special data sets in related fields is the key to the quality of training. The data set of railway defects includes images without defects and cracks, a total of 1000 images, all of which are 960×1000 in size. Among them, 100 are defect-free images; 900 are crack images. Before network training, the data set was divided into a training set, verification set, and test set according to 8:1:1.

5.3. **Performance indicators**

In order to ensure the objective evaluation of the experimental results, the mathematical model is used to analyze the results. MIoU was used to evaluate the experimental results.

The True value was divided into Positive value and Negative value, and the prediction result was recorded as True and the prediction error as False. A true example (TP) means that a defect has been detected correctly, a false positive example (FP) means that no defect has been detected incorrectly, a true negative example (TN) means that no defect has been detected correctly, and a false negative example (FN) means that a defect has been detected incorrectly.

Mean Intersection over Union (MioU): The ratio of the Intersection and Union of a class of predicted result values and real values, and the result of summation and reaverage.

\[ MioU = \frac{TP}{TP + FP + FN} \]  

(6)

5.4. **The experimental results**

![Image](image.png)

**Fig. 3** Experimental contrast diagram of crack test set

The MIoU values of the test set in different network structures are shown in the table.
| Models          | MIoU |
|-----------------|------|
| FCN-8s          | 0.58 |
| Unet            | 0.67 |
| Paper network   | 0.73 |

Tab.1 Comparison of MIoU values in test sets under different network models

6. Conclusion
In this paper, an improved Unet network segmentation model is proposed to effectively segment the railway defects. Integrate the attention mechanism, suppress the background part, highlight the defects, and make the final defect extraction accuracy higher. In the final processing, the noise is eliminated by adding post-processing to make the edge of the image smoother. In this paper, the network model is used to segment crack defects in the data set. Then, compared with the existing network models FCN-8S and Unet, the improved model is more accurate in locating the defects, with higher extraction precision and better segmentation effect. Compared with traditional defect detection, although the neural network is faster and can segment defects more accurately, a large number of data samples are needed in the early stage of the neural network to improve the final segmentation results. Secondly, the setting of parameters also has a certain influence, which requires multiple experiments.

References
[1] Guiyun T, Bin G and Yunlai G J 2016 Review of railway rail detect non-destructive testing and monitoring *Chinese Journal of Scientific Instrument* **37**(8) pp 1763-1780
[2] Hajizadeh S, Nuñez A and Tax D M J J 2016 Semi-supervised Rail Defect Detection from Imbalanced Image Data *IFAC-Papers OnLine* **49**(3) pp 78-83
[3] Guoqing L, Xiubo L J 2017 Track Dynamic Detection of High Speed Railway *Journal of Rail 

Science* **2** pp 50-55
[4] Jia M D 2005 Design and Research of recognition system on railway surface defect *Southwest Jiaotong University* (Chengdu) pp 31-81
[5] Guiyang X, Tianyun S and Shengwei R J 2013 Development of the On-board Track Inspection System Based on Computer Vision *China Railway Science* **1** pp 139-144
[6] Krizhevsky A, Sutskever I, and Hinton G 2012 ImageNet Classification with Deep Convolutional Neural Networks Conf *Neural Information Processing Systems* (Nevada) pp 1106-1114
[7] Zhongyue F D 2019 Technology of Image Semantic Segmentation *Beijing University of Posts and Telecommunications* (Beijing) pp 32-56
[8] Ronneberger O, Fischer P and Brox T 2015 U-Net: Convolutional Networks for Biomedical Image Segmentation Conf. *Proceedings of International Conference on Medical Image Computing and Computer-Assisted Intervention* (Heidelberg) pp 234-241
[9] Long J, Shelhamer E, and Darrell T J 2015 Fully Convolutional Networks for Semantic Segmentation J. *IEEE Transactions on Pattern Analysis and Machine Intelligence* **39**(4) pp 640-651
[10] Feng H, Jiang Z and Xie F J 2014 Automatic Fastener Classification and Defect Detection in Vision-Based Railway Inspection Systems *J. IEEE Transactions on Instrumentation and Measurement* **63**(4) pp 877-888
[11] Wang F, Jiang M and Qian C 2017 Residual Attention Network for Image Classification Conf *IEEE Conference on Computer Vision and Pattern Recognition* (Honolulu) pp 6450-6458