Rail Surface Defect Detection Method Based on Deep Learning Method with 3D Range Image

Geng Ming, Bo Zhou, Xiaohua Luo, Ren Ling, and Mingxiang Zhou

Abstract In the methods of using images for detecting surface defects of rails, the interaction such as light, stains, and water stains will cause false alarms. This paper proposals a method to detect surface defects of rails using 3D range line scan cameras combined with deep learning. By using the 3D range camera to acquire the information and 2D image information, and optimizing The original internet neural network structure, combined with the channel attention mechanism, a twin unet & 3D+ neural network model is proposed. First, the required database was established by using the 3D range camera, and then the comparison experts provided that the neural network proposed in this paper can effectually eliminate false alarms caused by light, Stains and water stains compared with other neural networks, and effectually promoted the rail surface. The correct rate of default detection.

Keywords Rail surface defects · 3D line scan camera · Twin Unet & 3D+ neural network

1 Introduction

Rail surface defects mainly refer to, Due to the wheel-rail pressure and centrifugal force of high-speed trains, the rail contact surface will be scratched, fallen off, blocks and other problems. If it is not rectified and polished at this time, after further wear, it may cause vicious time such as rail breakage, which will eventually lead to train safety accidents.

The traditional rail defect detection method is mainly through ultrasonic detection method, through the probe as the medium to detect the rail surface and internal defects [1]. Although the inspection items of rail include the defects on the rail surface and

G. Ming (✉) · B. Zhou · X. Luo · M. Zhou
China Railway Siyuan Survey and Design Group Co. LTD., Wuhan, China
e-mail: mynameisdc@yeah.net

R. Ling
Ningbo Rail Transit Group Co., Ltd. Smart Operation Branch, Ningbo, China
e-mail: sclead315@yeah.net

Y. Yang (ed.), *Advances in Frontier Research on Engineering Structures*, Lecture Notes in Civil Engineering 286, https://doi.org/10.1007/978-981-19-8657-4_5

inside the rail, the inspection efficiency is greatly discounted because of the need for the probe to move for inspection. At the same time, due to the limitation of the probe, the speed is often slow in the inspection process.

Compared with the traditional ultrasonic method to detect rail defects in service, Chen et al. [2] proposed to use image method to detect rail surface defects, and locate the defect position by finding the relatively large change area in the image pixel histogram. However, this method has higher requirements for illumination environment, and it requires that the refraction and reflection of rail surface are basically consistent and will not change too much, so it is difficult to meet the requirements of engineering application. Jin et al. [3] In order to solve the problems of inconsistent reflection on rail surface and large change of illumination in rail shooting environment, a special line scanning camera was used for shooting. At the same time, deep learning method is used for detection. Compared with the previous histogram image method, this method has higher accuracy, and can better adapt to different ambient light and rails with different reflection and refractive index. However, in order to ensure that the image will not appear jitter and other problems, and also to ensure the location and recognition effect of rail surface defects by neural network, its detection speed is relatively slower and its efficiency is poor. Among them, the most important thing is that there are many stains, dirt and other attachments on the surface of in-service rails due to the site environment, which often leads to false positives in this way.

In order to solve the problem of false alarm caused by rail surface stains and slow processing speed caused by neural network method, this paper proposes to use 3D camera to photograph rail surface, and at the same time, use the optimized neural network structure to detect rail surface defects.

2 Background

2.1 Defect Detection Based on Deep Learning Method

Under the background of using deep learning method to solve practical engineering problems, many successful deep learning frameworks have been put forward one after another, and have been applied to vehicle defect detection [4], pantograph wear detection [5], railway bridge bolt loss detection and railway irregularity, etc. Through the above engineering practice, it has been proved that compared with the manual detection method or the original detection method, the defect detection based on deep learning method has a more obvious improvement in detection accuracy.

The work of others has completely verified that the neural network method of continuous deep learning has a relatively good detection effect on rail surface defects, However, there is a large consumption of time and cost, and there may be some false positives. Therefore, further improvement is needed to improve the detection accuracy, and further reduce the time and cost required for rail surface defect detection and improve the detection efficiency.

2.2 3D Images

Compared with 2D images, 3D images are imaged in the form of structured ray scanning. In addition to the normal 2D image, the output of the 3D camera also contains data in the depth direction. Not only that, The 3D camera uses a laser transmitter with a specific band as the light source in image acquisition, It effectively avoids the interference of external ambient light, can effectively adapt to different illumination environments, and ensures that rail surface images with good imaging effect and relatively stable image quality can be obtained, which lays a good image foundation for rail surface defect detection.

2.3 Net Network Architecture

Unet [6] network structure was originally used for segmentation of lesions in medical images, and the whole network structure adopted encoding–decoding [7] network structure. Experiments show that the network structure of Unet neural network coding and decoding can extract the structural information and spatial information of objects in images relatively effectively. At the same time, because the network structure is relatively simple, the network parameters are relatively few. Therefore, compared with some complex networks such as DeepLabV3 [8], DeeplLabV3+ [9] and other neural network structures, the Unet network structure needs less processing time and higher efficiency, and has the value of engineering implementation.

But for rail surface defect detection, not only the types of defects are different, but also the morphological characteristics of the same defect are different, and there is a big gap between classes. Therefore, based on the analysis of rail surface defects and the images and depth data taken by 3D camera, this paper proposes a twin Unet-3D+ neural network framework to extract and alarm rail surface defects. Finally, compared with the previous rail surface defect detection methods, the neural network framework proposed in this paper has better accuracy and higher efficiency, and has stronger engineering practicability.

3 Twin Unet & 3D+ Neural Network

In order to solve the problems of rail surface defect recognition, such as leak recognition, long processing time and difficulty in real-time detection, twin Unet-3D+ neural network is proposed in this paper. In this network, the original image data collected by

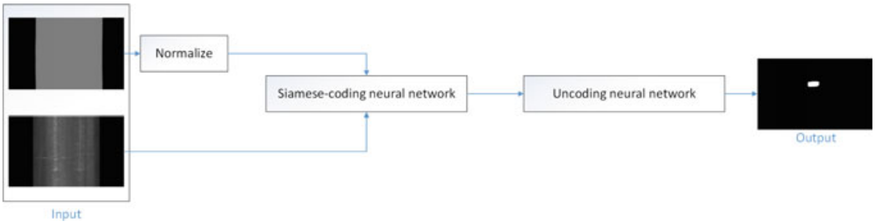


Fig. 1 Flow chart of overall algorithm

3D camera and the depth direction data are jointly input into the network. Compared with 2D cameras, 3D cameras have the following characteristics:

(1) Collected data, such as stripping, falling block, crushing and other rail surface defects, will not only be reflected in 2D images, but also show differences with normal rail surface in depth direction. But for the surface water stains, stains and other differences, although the 2D image with other rail surface area there is a big difference, but in the depth direction of the 3D image with the normal rail surface area there is no obvious difference.

(2) 3D cameras can only sense the specific band of light emitted by their lasers, so they can avoid the spot area formed on the image by the reflected light caused by external ambient light and smooth rail surface. Like stains and water stains, if we only rely on 2D images for recognition, it will often cause false recognition of defects and lead to false alarms. For 3D cameras, the external ambient light will not have a great impact on the image quality, whether it is for 2D images or images in depth direction, so as to ensure that real and reliable rail surface data can be obtained.

Based on the above input mode through the combination of original 2D image information and depth image information, not only the defect part can be effectively located, but also the light spot, water stain, dust and the like that may cause false alarm can be effectively eliminated, and more accurate rail surface defect alarm information can be obtained. The overall flow chart of the algorithm is shown in Fig. 1.

3.1 Overall Network Framework

Aiming at the problems existing in rail surface defect detection, this paper proposes to use twin Unet-3D+ neural network framework to solve them. The overall network framework is shown in Fig. 2.

As far as the original input image depth information is concerned, it mainly reflects the relative distance information between the photographed object and the camera position. At the same time, even if there are defects such as peeling and falling off the rail surface, its depth will not have a huge deviation, so it is necessary to

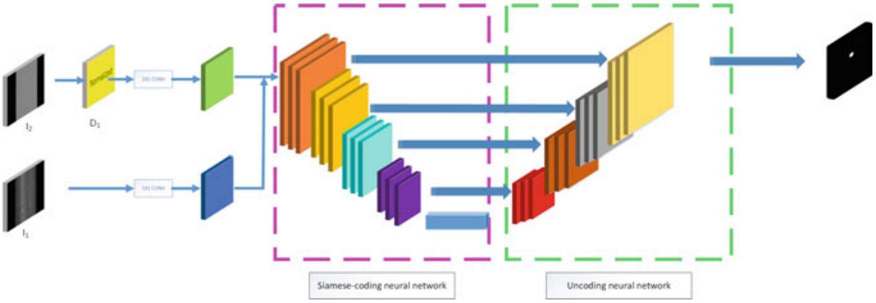


Fig. 2 Overall network framework

carry out preliminary normalization before inputting into the neural network, and the normalization method is shown in Sect. 2.2.1.

At the same time, in order to ensure the high unity of 2D image information and depth information, the encoder extraction information module at the front end of UNET is modified to be a twin [10] coding information module, which ensures that the feature information of 2D information and depth information extracted by the network has a high degree of consistency and reduces the discrete degree of 2D feature information and depth feature information as much as possible.

For the decoding information module, besides the information extracted by the preceding twin coding information module, it is necessary to further enhance the difference between depth information and 2D image information in the whole network. Therefore, in this layer, in addition to the feature information extracted by the previous twin coding information module, the extracted 2D feature information and depth information are further differentiated and input into the decoding network at the same level. The specific method is shown in Sect. 2.3.

As far as the loss function is concerned, You can't simply cross-bar the difference between the output image and Ground Truth, Instead, we should further improve the difference and unity between beam depth image and 2D image and the relationship between input and output to further optimize the back propagation link of the network, so as to improve the defect segmentation effect of the network and accelerate the convergence and training speed of the network.

3.2 Twin Coded Information Module

The network structure of twin coded information modules is shown in Fig. 3.

In addition to the normal 2D information, depth information is also used as the input of the whole module into the twin coded information module. But before input to the network, it needs to be normalized through Batch-Normalized layer to ensure that the depth information can be fully applied in the network, and at the same time,

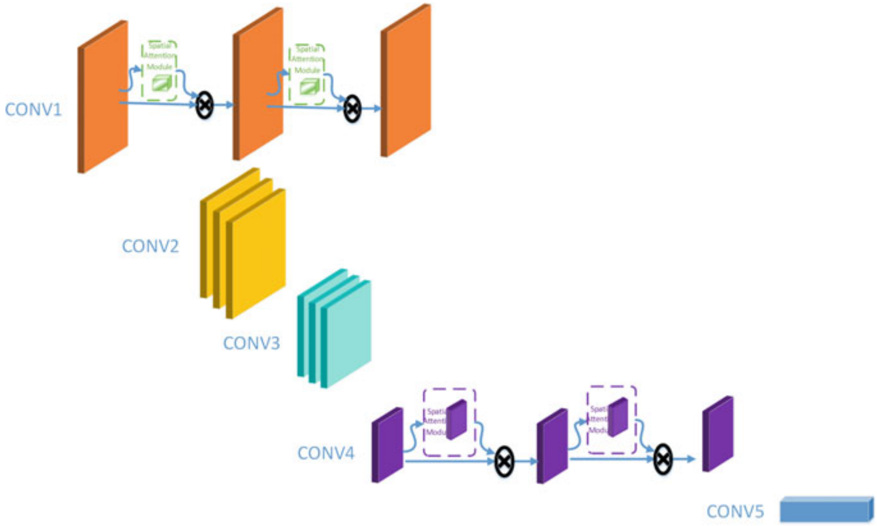


Fig. 3 Twin coding information module

the original depth value will not be far larger than the pixel value of the 2D image, which will lead to the coverage of the 2D image information.

As shown in Fig. 3, in order to ensure that the convolution network can extract consistent information content in the spatial position and object structure in I1 and D1 images, the twin network structure is mainly used for information extraction in the preceding network CONV1-CONV3 network structure. At the same time, in order to further emphasize the spatial position information and object structure information, the spatial attention mechanism is further adopted in the twin network structure of this level to extract its spatial information.

On the basis of the consistent extraction of image spatial position information and object structure information by the network in the preceding item, CONV4 to CONV5 in the latter item mainly extract feature information, so twin network structure is no longer used. On the premise that the spatial position information is consistent with the object structure information, the neural network with channel attention mechanism [11] is mainly used to extract the information. It can effectively ensure that the depth information and the original 2D image information can be fully fused and applied.

Depth information normalization layer

Depth information normalization layer is mainly used to normalize the original depth information image input into neural network and extract simple information. In this layer network, firstly, the normalized information matrix of depth information images needs to be obtained by the following formula 1.

$$f_{\text{Normalize}}(x, y) = \frac{f_{org}(x, y) - f_{\min}}{f_{\max} - f_{\min}} \tag{1}$$

where X and Y represent the horizontal and vertical coordinates of depth information and $f_{org}(x, y)$ represent the size of its original depth information. $f_{\max}$ and $f_{\min}$ represents the maximum and minimum value of depth information in the whole depth information image; $f_{\text{Normalize}}(x, y)$ represents the normalized depth information.

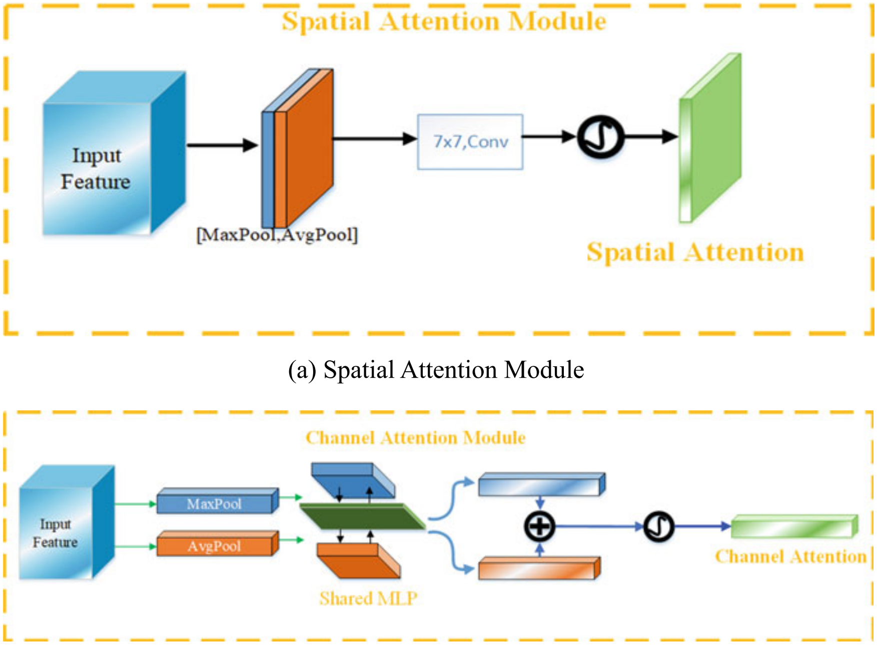
In addition to the linear normalization formula of Formula 1, the 3 × 3 convolution neural network is used to extract the normalized feature information and input it to the subsequent twin network.

Spatial Attention and Channel Attention Mechanisms

As shown in Fig. 4, it is the spatial attention and channel attention module. The main principle of this module is to evaluate the key information of the image in spatial domain and channel domain, and assign different weights.

The main uses of the above two attention mechanisms are divided into the following two aspects:

(1) Spatial attention module: As shown in Fig. 4a, the main purpose is to ensure a higher correspondence to the rail area during convolution. That is to say, it keeps a high response level and a high weight for the rail spatial area, and ensures



(a) Spatial Attention Module

(b) Channel Attention Module

Fig. 4 Spatial attention and channel attention mechanism

that the twin network can extract better spatial position information and object structure information from the 2D information image and depth information image of the rail area.

(2) In this module, the feature matrix input into the network is compressed at the channel level by using Average pooling and max pooling. Then, the average pooling matrix and the feature matrix output from the maximum pooling matrix are Concat operated, and then a convolution kernel with the size of 7 × 7 and Sigmoid activation layer are used to ensure that the final output feature matrix is consistent with the input feature matrix in dimension.

(3) Channel attention module: As shown in Fig. 4b, the main purpose of this module is to better fuse the information of different channels of the feature matrix input into this module, and ensure the integration and extraction of depth information and 2D image information.

In this module, the maximum and average pooling operations are performed on the input feature matrix, and the one-dimensional maximum pooling matrix and average pooling matrix are obtained. The global average pooling matrix mainly reflects the information of each element in the original input matrix, while the maximum pooling matrix mainly responds to the information of elements with greater weight. After that, the elements in the corresponding positions in the two pooled matrices are activated by sharing multi-layer perceptrons, and a one-dimensional channel attention weight output is obtained.

The weight is output and multiplied with the original feature matrix. On the basis of not affecting the dimensions of the original feature matrix, the information weights in different dimensions are strengthened to improve the network training speed and the fusion degree of key depth information and 2D image information.

3.3 Decoding Information Module

As shown in Fig. 2, after the depth information and 2D image information input to the neural network are deeply fused and extracted by the encoded information module, it is necessary to further restore the information to the original image size, so as to obtain the size and position of rail surface defects.

In this module, the horizontal connection part in the original UN is improved. The channel attention mechanism is mainly added to each layer of network. The original horizontal single input mode is changed to the joint input of depth information feature matrix and 2D information feature matrix. In addition, the channel attention mechanism is used to further enhance the weight of different channels, so as to ensure that the network can achieve better information fusion effect between depth information and 2D information.

3.4 Loss Function

Because depth information is used to identify rail surface defects in network input and network flow, it is impossible to simply feed back the difference between network output and label data as loss to the network mentioned in the preceding paragraph. In the process of measuring the loss, it is necessary to further consider the influence of depth information on the network. Therefore, the following formula 2 pairs are mainly used as the loss function of this network.

$$f_{loss} = \frac{\sum_{i=0}^{n-1} (Y_i^D \cdot Y_i^{GT} - Y_i^O)^2}{n} \tag{2}$$

Y_i^D represents the normalized depth data value; Y_i^{GT} represents a valid label information value; Y_i^O represents the output probability diagram of neural network. Firstly, Y_i^D point multiplication is carried out on and Y_i^{GT}. Then, Y_i^O the mean square difference is obtained by combining the output of neural network as the loss function of neural network for back propagation.

4 Experiment and Analysis

On the basis of the above twin Unet-3D+ neural network, further verification and comparative experiments are carried out to prove the effectiveness of the neural network proposed in this paper.

4.1 Experimental Environment

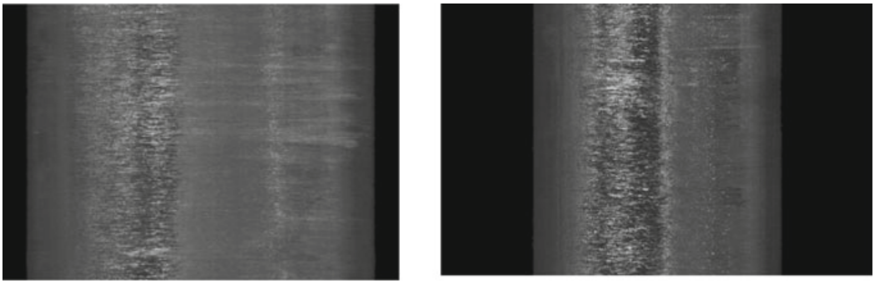
In order to ensure the repeatability of the experiment, the experimental environment is explained now. The experimental environment is: processor IntelE5-2670, memory 8G, graphics card GTX1080Ti, video memory 12G, operating system Windows10 Professional Edition, and deep learning framework PyTorch. At the same time, in order to improve the convergence speed of the model, some basic parameters of network training are shown in Table 1.

Table 1 Basic parameter setting of neural network

Parameter name	Parameter value	Parameter name	Parameter value
Learning rate	10e-5	Learning momentum	0.9
Iterative algebra	400	Batch size	20
Optimization mode	Adam	Learning strategy	Steplr

Table 2 Basic parameters of camera

Camera name	Accuracy of camera in X direction	Accuracy of camera in Z direction	Acquisition frequency
Kearns 3D line scanning camera	0.5 mm	0.1 mm	1600 Hz



(1) Normal Rail Surface

(2) Worn Rail Surface

Fig. 5 2D rail surface image

In this network, the depth image and 2D image data are used as the input of neural network to judge the rail surface defects. Therefore, the rail surface defect data set used for training and testing is mainly collected by Keens 3D line scanning camera. The main parameters of the camera are shown in Table 2.

4.2 Establishing a Dataset

The accuracy index of the collected data needs to ensure that the distance between the camera and the photographed object, that is, the rail surface, needs to be between 300 and 500 mm. Its 2D image data is shown in Fig. 2 (Fig. 5).

It can be seen from the figure relatively intuitively that the collected rail surface images can clearly see the rail surface texture, rust and stains, etc., which can fully meet the image quality requirements for rail surface defect recognition.

4.3 Comparative Experiment

Comparative experimental standard for defect detection

The objective, accurate and unified evaluation criteria of the algorithm can help people better compare the results of the algorithm and promote the improvement of the effect of the algorithm. For segmentation algorithms, MIoU (Mean Intersection over Union) is mainly used for comparative evaluation. As shown in Formula 3.

$$MIoU = \frac{1}{k+1}\sum_{i=0}^{k}\frac{p_{ii}}{\sum_{j=0}^{k} p_{ij} + \sum_{j=0}^{k} p_{ji} - p_{ii}} \tag{3}$$

In Formula 3, p_{ij} represents the number of real values i and predicted j, , and K + 1 is the number of classes (including empty classes). p_{ii} is the true positive. p_{ij} is a false positive and p_{ji} is a false negative, respectively.

On the premise of clear comparison experimental standards, the neural network proposed in this paper is compared with several popular neural networks. The data set of other methods mainly uses 2D data collected by camera as training set for training.

Experiments show that compared with other methods, the neural network method proposed in this paper is not only better in detection effect, but also better than other neural networks. At the same time, compared with other networks, the time required for detection is less, which is more in line with the actual needs of engineering.

Comparative experiment of defect detection

As shown in Table 3, the comparison effect between this method and other methods on objective indicators of MIoU.

It can be seen that compared with other detection methods, the method proposed in this paper has better performance and actual detection effect on MIoU.

It can be seen intuitively from Table 3 that compared with other methods, the index of MIoU has been greatly improved.

At the same time, in order to further explain the influence of eliminating rail surface stains on network recognition effect, this paper further recalculates p_{ij} of the MIoU formula, and the recalculated formula is shown in formula (4).

$$MIoU_{(False\,\mathrm{Positive})} = \frac{1}{k+1}\sum_{i=0}^{k}\frac{\sum_{j=0}^{k} p_{ij}}{\sum_{j=0}^{k} p_{ij} + \sum_{j=0}^{k} p_{ji} - p_{ii}} \tag{4}$$

The results are shown in Table 4.

Table 3 MIoU contrast experiment (Unit:%)

Method	MIoU
FCN	55.2
Unet	58.3
Unet++	58.5
DeeplabV3	68.7
Deeplabv3+	70.9
DPLBD	33.5
Ours	**80.3**

Table 4 MIoU (False Positive) contrast experiment (Unit:%)

Method	MIoU
FCN	38.4
Unet	40.3
Unet++	44.6
DeeplabV3	25.8
Deeplabv3+	28.2
DPLBD	60.5
Ours	**10.3**

Table 5 Comparison of elapsed time (unit: ms)

Method	CostTime
FCN	65.9
Unet	**35.5**
Unet++	74.5
DeeplabV3	786.6
Deeplabv3+	503.4
DPLBD	87.6
Ours	40.3

It can be seen that compared with other neural networks which only use 2D image information as network input, the twin Unet & 3D+ neural network proposed in this paper has better performance in reducing the false positives due to non-defects such as rail surface stains, water stains and oil stains, and can effectively reduce the workload of manual review.

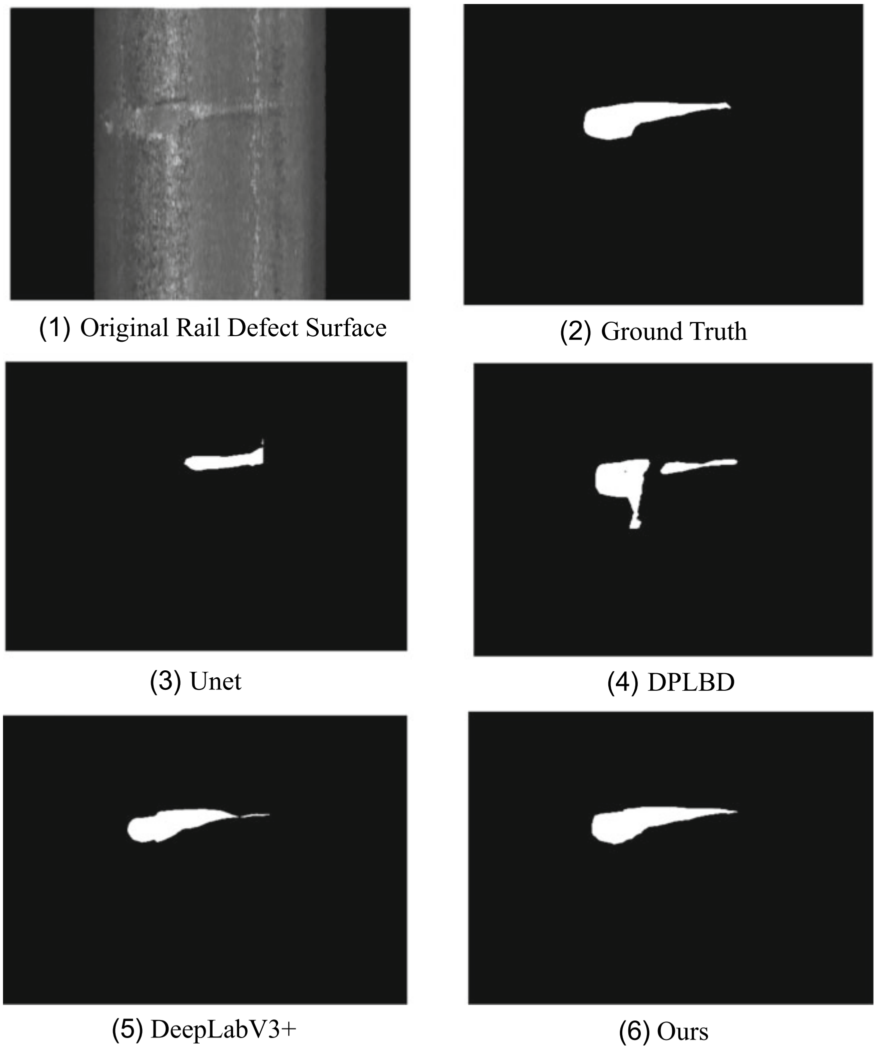
As shown in Fig. 6, the detection results of some rail surface defects are compared and displayed.

In addition to the comparison of rail surface defect detection results, this paper also compares the time required for each neural network to process single image data, as shown in Table 5.

4.4 Experimental Analysis

Through the above-mentioned comparative experiments, it is obvious that the network proposed in this paper has a better effect on rail surface defect detection than other algorithms that only use 2D images as input. Through the analysis of MIOU and MIOU (False Positive) data, the improvement of the detection effect of neural network proposed in this paper is mainly due to the effective reduction of the probability of false recognition by combining 2D image information with depth information. At the same time, for the real rail surface defects. By means of "strengthening" depth image information, Compared with the previous network, it has better segmentation effect in edge details, By further judging these information, we can determine whether the defects are rail head fracture, large-scale block falling, etc., and then transform the surface defect information into rail defect hazard grade information, thus helping public works maintenance personnel to determine the emergency degree of maintenance and ensure the safety of train running (Fig. 6).

In addition to relatively good detection results, the neural network proposed in this paper has a good performance in detection time. This is mainly due to the selection of



(1) Original Rail Defect Surface

(2) Ground Truth

(3) Unet

(4) DPLBD

(5) DeepLabV3+

(6) Ours

Fig. 6 Comparison of detection effects

the network, mainly using a simple Unet as the main structure of the neural network proposed in this paper. At the same time, the twin network structure further reduces the calculation parameters needed by the neural network. Therefore, even if the depth image information is added as the input of neural network, compared with the original Unet neural network structure, its reasoning speed is basically consistent, and there is no obvious time consumption.

5 Conclusions

Combining the actual engineering situation and the advantages of neural network detection method in defect detection, this paper proposes a way to detect rail surface defects by combining depth image information with 2D image information. Through the method proposed in this paper, compared with the previous methods of rail surface defect detection, the twin Unet-3D+ neural network proposed in this paper has better performance in detection effect. Compared with other methods, because of the combination of depth image information, a large number of stains, water stains, oil stains and light spots can be effectively eliminated, which greatly improves the accuracy of detection. At the same time, the time cost of the neural network proposed in this paper is also within the acceptable range, which ensures its engineering practicability and has great significance for improving the work efficiency of public works inspection. However, due to the limitation of data set, some rail surface defects that have not learned similar features are difficult to identify accurately. Therefore, in the follow-up work, in addition to further increase the rail surface defect data set, it is necessary to further improve the generalization of the neural network model and the recognition effect of rail surface defects with large gaps between classes.

Acknowledgements This paper is supported by the National Key R&D Program of China (2018YFB2100903) and Service capacity building of Hubei innovative method promotion and application base (2020IM020800).

References

1. Zhang S (201) Research on ultrasonic propagation simulation and defect detection method in rail. Central South University
2. Chen S, Chang H, Zhang R et al (2021) Rail surface defect detection based on projection histogram. Mod Ind Econ Inf 11(11):19–20, 76
3. Xiating J, Yaonan W, Hui Z et al (2019) Rail surface defect detection system based on Bayesian CNN and attention network. Chin J Autom 45(12):2312–2327
4. Deqiang H, Jian M, Jiajun Z et al (2020) Weld defect detection of subway vehicles based on improved Faster R-CNN. Chin J Railway Sci Eng 17(4):996–1003
5. Xiaobin T (2013) Research on pantograph wear detection system based on machine vision. Guangdong University of Technology, Guangdong

6. Brox T, Fischer P, Ronneberger O (2015) U-net: convolutional networks for biomedical image segmentation. International conference on medical image computing and computer-assisted intervention. Springer, Cham, pp 234–241
7. Chen LC, Kokkinos I, Papandreou G et al (2014) Semantic image segmentation with deep convolutional nets and fully connected crfs. arXiv preprint arXiv:1412.7062
8. Chen LC, Papandreou G, Kokkinos I et al (2017) Deeplab: Semantic image segmentation with deep convolutional nets, atrous convolution, and fully connected crfs. IEEE Trans Pattern Anal Mach Intell 40(4):834–848
9. Chen LC, Papandreou G, Zhu Y et al (2018) Encoder-decoder with atrous separable convolution for semantic image segmentation. In: Proceedings of the European conference on computer vision (ECCV), pp 801–818
10. Koch G, Salakhutdinov R, Zemel R (2015) Siamese neural networks for one-shot image recognition. In: ICML deep learning workshop, vol 2
11. Woo S, Park J, Lee J Y, et al (2018) Cbam: convolutional block attention module. In: Proceedings of the European conference on computer vision (ECCV), pp 3–19