

A Method for Train Rail Surface Condition Recognition Based on Local Inference Constrained Network

Jihong Zuo¹, Lili Liu², Yan Li¹, Chuanyin Yang²

¹College of Rail Transit Locomotive and Rolling Stock, Hunan Railway Professional Technology College, Zhuzhou, China

²College of Intelligent Control, Hunan Railway Professional Technology College, Zhuzhou, China

Email: 277644902@qq.com

How to cite this paper: Zuo, J.H., Liu, L.L., Li, Y. and Yang, C.Y. (2025) A Method for Train Rail Surface Condition Recognition Based on Local Inference Constrained Network. *Journal of Computer and Communications*, 13, 140-152.

<https://doi.org/10.4236/jcc.2025.134009>

Received: March 12, 2025

Accepted: April 24, 2025

Published: April 27, 2025

Copyright © 2025 by author(s) and Scientific Research Publishing Inc. This work is licensed under the Creative Commons Attribution International License (CC BY 4.0).

<http://creativecommons.org/licenses/by/4.0/>



Open Access

Abstract

The condition of the rail surface plays a crucial role in the wheel-rail contact behavior. Accurately recognizing the rail surface condition can provide important support for the high-performance control of trains. However, the small amount of rail surface condition data leads to the lack of the feature space of the rail surface adhesion medium. Therefore, this paper proposes a method for recognizing the rail surface condition based on the local inference constrained network. Firstly, an adversarial network is used to generate diverse data samples while ensuring the semantic consistency of these samples. Meanwhile, an attention mechanism module is added to the feature extraction network to prominently highlight the information features of the target area and enhance the adaptability of the model. Secondly, a residual spinal network for local input decision-making is constructed, and the activation function is improved to accelerate the learning speed of the network and enhance the recognition accuracy. Finally, experiments are carried out to verify the effectiveness and feasibility of the proposed method.

Keywords

Rail Surface Condition Recognition, Residual Spinal Network, Adversarial Network

1. Introduction

The traction/braking control performance of rail transit trains depends on the wheel-rail contact behavior, and the rail surface condition is a key factor affecting the wheel-rail contact behavior. Effectively recognizing the rail surface condition can provide a crucial basis for the high-performance control of trains [1]. Cur-

rently, scholars at home and abroad have carried out extensive research on the automatic detection methods of the rail surface condition [2]. However, most of their research objects are local close-up images of the rail surface adhesion medium area that have been manually selected. These images are abstracted into simple one-dimensional feature vectors, and a large amount of information is compressed. As a result, the detection effect on the target is poor, and there are also problems such as low efficiency and poor real-time performance, making it difficult to meet the requirements of rail surface condition detection work that requires high accuracy and efficiency [3].

With the rapid development of machine learning, deep learning has shown great advantages and potential in feature extraction. The application of deep learning methods to the research of rail surface condition recognition has emerged. Classic deep learning methods such as the R-CNN algorithm [4], YOLO algorithm [5], SSD algorithm [6], ION algorithm [7], DSSD algorithm, and deep convolutional neural networks have successively appeared. Target detection methods such as R-CNN (Region-CNN) [8] and later Fast R-CNN [9], Faster R-CNN [10] apply deep convolutional neural networks to the field of target detection through the method of candidate windows. However, this kind of method increases the target detection time by using candidate windows (in R-CNN, 2000 candidate regions need to perform 2000 single image classifications), making it difficult to carry out real-time detection, and it cannot be effectively applied to the recognition of rail surface condition images that require high real-time performance. In order to improve the shortcoming of the excessively long detection time of the R-CNN series, the YOLO (You Only Look Once) model abandons the time-consuming scheme of candidate windows and adopts the method of directly performing target detection on a single image. Ref [11] proposes a rail surface defect detection method based on an improved YOLOv5s. By setting image anchors and transforming the structure of the deep convolutional neural network, YOLO makes its output become the position of the target relative to the preset anchors and the size of the detection box, so as to locate the target. This method greatly improves the efficiency of target detection. However, since the YOLO model only uses the information in the deep layer of the network as the feature for target detection and loses the detailed information of the image, the positioning error of the target is relatively large.

The application practices at home and abroad have proved that a large number of annotated training images (including the category and location of the target) are the basis for the successful application of deep convolutional neural networks in the field of target detection. The rail surface condition is greatly affected by external factors. Especially for the railway system in an open environment, the rail surface will inevitably be interfered by media such as water, oil, ice, snow, frost, and fallen leaves, and condensates will form on the rail surface. The weather conditions will affect the collection of rail surface image data samples, and some rare rail surface conditions are even more difficult to collect. For example, the occurrence of an oily state is very rare during the data collection process. The collection

of railway data must be approved by relevant departments such as railway operators. At the same time, considering the limited railway operation time, all these will lead to a small number of available data samples. When directly using small samples for deep learning, problems such as overfitting of model training and low model accuracy will occur due to the insufficient feature space of the rail surface adhesion medium. There is rarely any relevant research on how to solve the problem of rail surface condition recognition under small samples.

In view of this, this paper focuses on the three key aspects of sample expansion, feature enhancement, and network decision-making, and makes every effort to overcome the deep learning problems involved in the field of train rail surface condition recognition in the context of small samples, and then constructs a rail surface condition recognition model based on the local inference constrained network. In terms of sample expansion, with the help of the generative adversarial network technology, the limited number of real sample data is expanded to obtain a data set with both semantic coherence and sample distribution diversity, laying a solid data foundation for subsequent research. In terms of feature enhancement, by ingeniously integrating the attention mechanism into the network architecture, the key position of the features of the rail surface adhesion medium is highlighted, and the adaptability level of the model is significantly improved, enabling it to better deal with the complex and changeable rail surface conditions. In terms of the network decision-making link, a spinal network with a residual structure is innovatively constructed. With the accurate input of local information, it helps the deep network to achieve more accurate result output while reducing the number of operations, greatly optimizing the operation efficiency and recognition accuracy of the model. Finally, the proposed method is rigorously verified and deeply analyzed through experiments to comprehensively ensure the scientificity and effectiveness of the method.

2. Modeling of the Local Inference Constrained Network for Rail Surface Condition Recognition

Dry, wet, and oily are relatively common types of rail surface conditions, as shown in **Figure 1**.



Figure 1. Common types of rail surface conditions.

Given that the data scale of rail surface condition images is relatively small, it is very likely to lead to the situation of feature space deficiency. Therefore, after completing the conventional data preprocessing process, on the one hand, starting from sample expansion, the number and types of samples are enriched to make

up for the scarcity of the original data; on the other hand, focusing on the feature enhancement link, through ingenious design and meticulous operation, various unique features of the rail surface are deeply explored and strengthened, highlighting their key position in the subsequent model recognition process. At the same time, bold innovation and optimization are carried out at the network decision-making level, and an efficient network system suitable for the small sample situation is carefully constructed. Combining all the above efforts, we are committed to building a train rail surface condition recognition model specifically for the condition of small samples, effectively improving the accuracy and reliability of rail surface condition recognition, as shown in **Figure 2**.

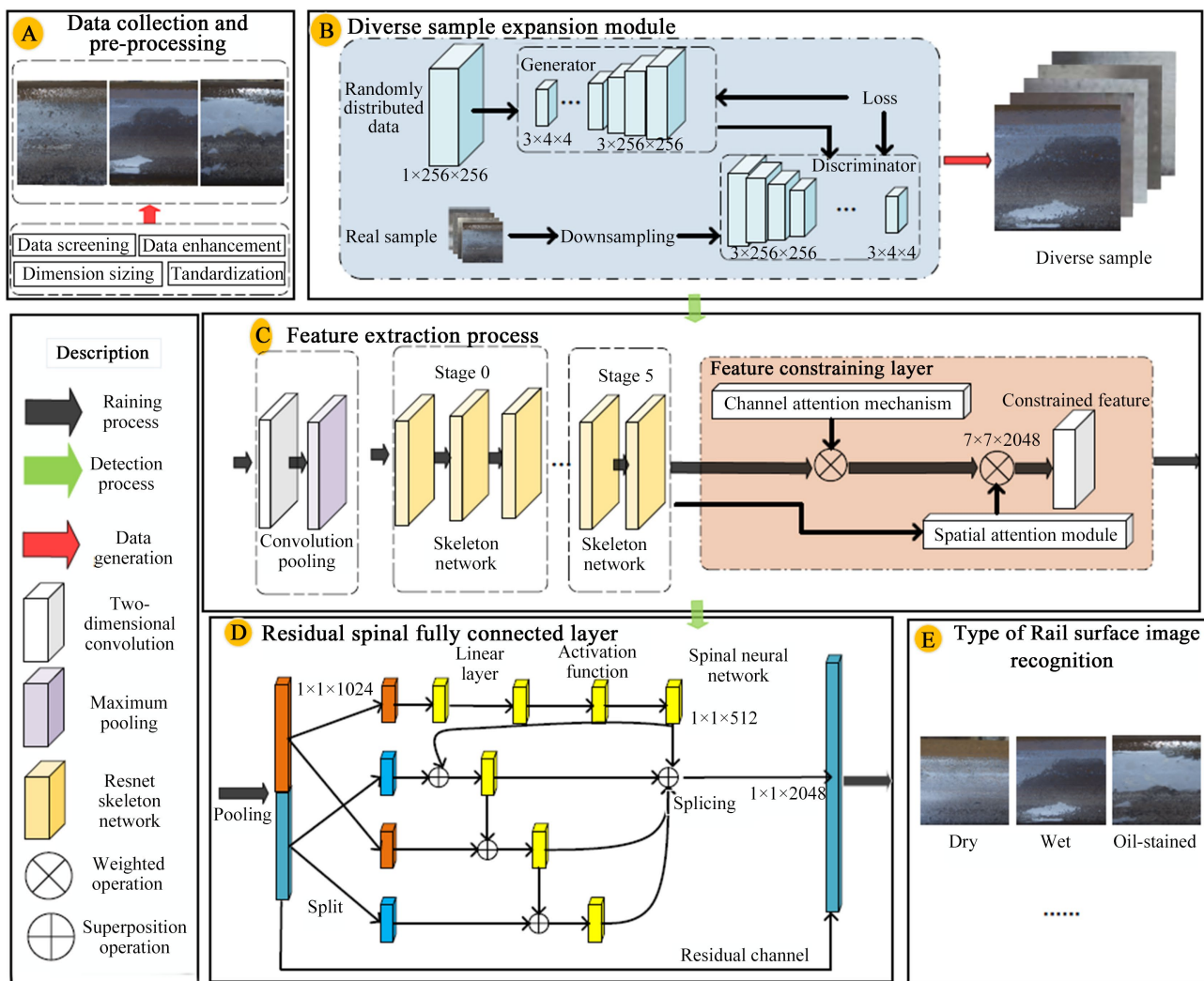


Figure 2. Recognition model of the local inference constrained network for rail surface conditions.

In the recognition model of the local inference constrained network in **Figure 2**, the diverse sample expansion module makes use of a generative adversarial network model to fit the real sample space, generating diverse and accurate rail surface image samples so as to expand the sample feature space; the feature extraction process module first acquires discriminative feature vectors and then embeds an

attention mechanism to assign different weights to the feature vectors, enhancing their discriminability; the residual spinal fully connected layer module reorders and assigns values to the importance of local features, allowing the network to achieve more accurate results with fewer computations.

2.1. Sample Expansion

The purpose of sample expansion is to provide an adequate number of samples for the learning and training of the deep network. In the basic generative adversarial network model, the generator receives a random sample space with a Gaussian distribution and noise data, while the discriminator receives the samples generated by the generator and real data samples. Both of them use the loss function to perform backpropagation of gradients to update the parameters of their own network models. However, this model uses random samples with a Gaussian distribution and noise data, ignoring the characteristics of the information distribution of the actual application objects, resulting in poor quality of the generated samples. In view of this, during the sample expansion stage of this paper, noise data is added based on the AdaIN mechanism, enabling the generator to generate rail surface condition images according to the image size. Moreover, independent Gaussian noise is input into each layer, which only causes slight changes in the visual features.

Among them, AdaIN is defined as

$$\text{AdaIN}(x_i, y) = \delta(y) \left(\frac{x_i - \mu(x_i)}{\sigma(x_i)} \right) + \mu(y) \quad (1)$$

In the formula, $\mu(x_i)$ and $\sigma(x_i)$ represent the mean and variance of the real samples respectively, $\mu(y)$ and $\delta(y)$ represent the mean and variance of the latent sample space respectively.

The data network generated by using formula (1) can better fit the data distribution of the rail surface images and generate image information that is more in line with the real samples. For the specific network structure, you can refer to reference [12].

2.2. Rail Surface Feature Enhancement Based on Attention Mechanism

The features of the rail surface condition are subtle, randomly located, and greatly interfered by background noise, which leads to the small differences in the feature subsets extracted by the traditional convolutional neural network and the unobvious weight comparison. Based on this, this paper designs a feature extraction module embedded with a multi-dimensional attention mechanism. The specific approach is as follows: channel attention is added in the channel dimension to highlight the information of the feature channels; attention is applied in the spatial dimension to strengthen the information in the spatial dimension. The module is shown in **Figure 3**.

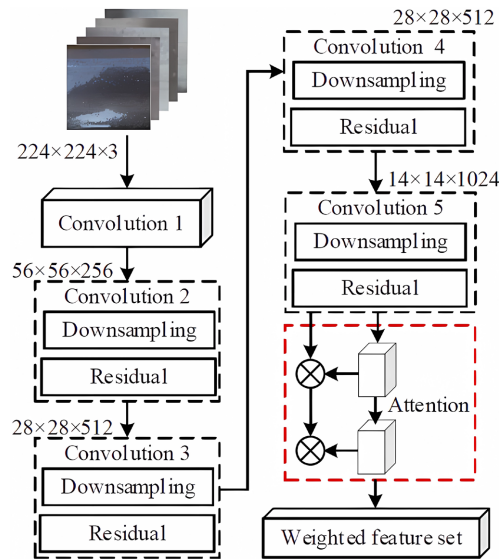


Figure 3. Structure diagram of rail surface feature enhancement by attention mechanism.

As shown in **Figure 3**, in the process of enhancing the rail surface condition features, the input consists of rail surface condition image samples. These samples are first processed by a convolutional neural network to obtain feature vectors containing high-dimensional semantic information. Then, the weighted attention mechanism parameters are calculated through the channel attention mechanism and the spatial attention mechanism. Subsequently, the feature maps composed of feature vectors are multiplied element-by-element with the attention weights obtained from the attention mechanism, and finally, the output result after attention weighting is obtained. Among them, the expressions of the channel attention mechanism and the spatial attention mechanism are as follows

$$Mc(x) = d(MLP(Conv5(x)) + MLP(Conv5(x))) \quad (2)$$

$$Ms(x) = d(Conv(Mc(x)) \oplus Mc(x)) \quad (3)$$

$$At(x) = Conv5(x) \otimes Mc(x) \quad (4)$$

Among them, Conv5 is the feature space formed after passing through five convolutional layers. Mc and Ms represent the channel attention weight and the spatial attention weight respectively in sequence, and At is the weighted feature space. By using the channel attention mechanism, the spatial dimension can be compressed, the features of the rail surface condition can be prominently highlighted, and high weights can be assigned to them. By applying the spatial attention mechanism, the channel dimension can be compressed, and the spatial position information can be supplemented to the model, so as to improve the utilization rate of the feature vectors.

2.3. Residual Spinal Fully Connected Layer

The fully connected layer is generally used in the decision-making layer of a neural

network. This module is characterized by a large number of parameters and difficult training. In this paper, the idea of residuals is introduced into the spinal fully connected layer, and a residual spinal fully connected layer is proposed. This module innovatively improves the input-output connection and adds a residual channel to alleviate the overfitting problem of the spinal fully connected layer. Moreover, the activation function is reset based on the quantity characteristics of the data set, making this module more suitable for the small sample data set of rail surface condition images. The model structure is shown in **Figure 4**.

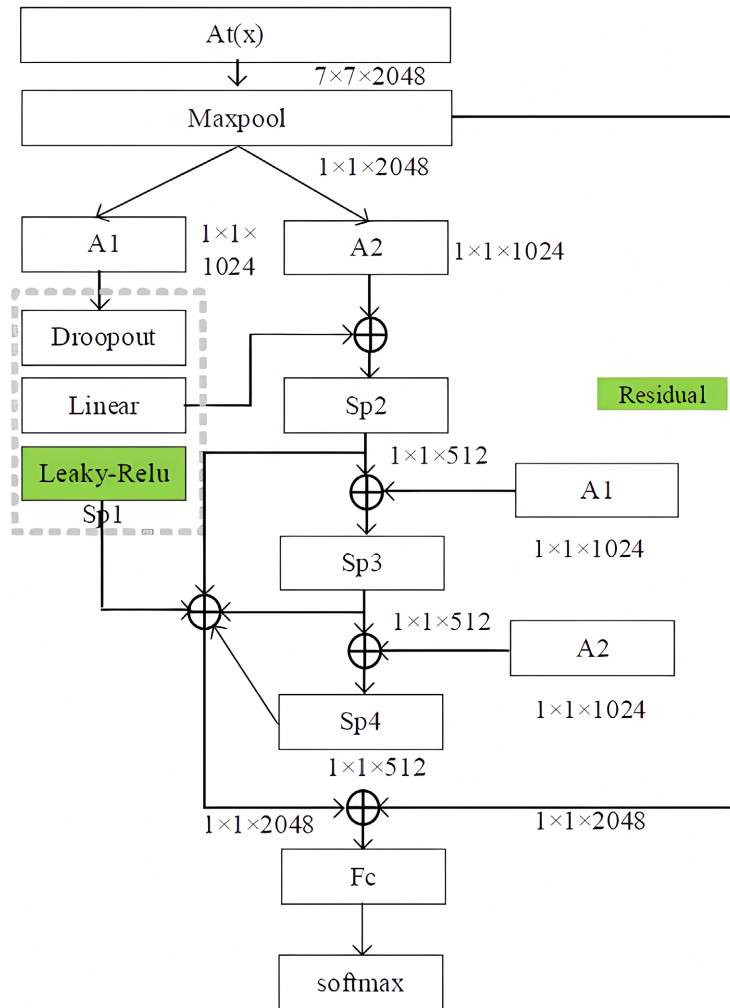


Figure 4. Structure diagram of the residual spinal fully connected layer.

Among them, the feature A_t extracted by the module with the attention mechanism introduced is split into two parts, $A1$ and $A2$.

$$A1 = A2 = \{A_{t1}, A_{t2}, \dots, A_{t640}\} \tag{5}$$

The spinal neural network in this paper is composed of four modules, namely $Sp1$, $Sp2$, $Sp3$, and $Sp4$. Each module contains a dropout layer, a linear layer, and an activation function. The settings of the dropout layer and the activation function are the same, and the output of the linear layer is 512 for all of them. The

difference lies in that the input size of the linear layer in the $Sp1$ module is 1024, while the input size of the linear layer in the $Sp2$, $Sp3$, and $Sp4$ modules is 1536. Such a setting is to keep the feature dimensions input to the fully connected layer consistent

$$Sp1 = \sigma_2(\text{Linear}(\text{Droop}(A_1))) \quad (6)$$

$$Sp2 = \sigma(\text{Linear}(\text{Droop}(Sp1 \oplus A2))) \quad (7)$$

$$Sp3 = \sigma(\text{Linear}(\text{Droop}(Sp2 \oplus A1))) \quad (8)$$

$$Sp4 = \sigma(\text{Linear}(\text{Droop}(Sp3 \oplus A2))) \quad (9)$$

$$FC = \text{linear}(Sp1 + Sp2 + Sp3 + Sp4) + \text{Maxpool}(At(X)) \quad (10)$$

Here, Droop represents a random zeroing operation, Linear stands for the linear layer, and σ is the activation function of the spinal neural network. The output of each spinal block, when combined with the feature vector after max pooling, yields a fine-tuned feature vector of size $1 \times 1 \times 2048$. Eventually, this enables the accomplishment of the pattern recognition task for rail surface condition images.

3. Experiments and Analyses

3.1. Experimental Conditions

3.1.1. Datasets

The dataset of rail surface condition images used in this paper was captured by a SONY-ILCE-6000 camera at different time periods in a local train depot. A total of 210 rail surface condition images were selected and the image categories were labeled. Among them, there were 70 images each for dry rail surfaces, wet rail surfaces, and oil-contaminated rail surfaces. Since the original images captured by the camera contain a large amount of background information and noise interference, which increases the difficulty of model recognition, pre-processing operations were carried out on the rail surface condition dataset. These operations include median denoising, geometric correction, and extraction of rail surface area images, so as to improve the robustness and recognition accuracy of the network.

To objectively measure the performance of the generative adversarial network and address the problem of insufficient training samples, this paper constructs a rail surface condition image dataset, which covers various forms of data enhancement: the original dataset without enhancement (Non), the dataset processed by geometric transformation (GT), the dataset processed by pixel transformation (PT), and the dataset enhanced by the generative adversarial network (GAN). Meanwhile, for the convenience of network training and testing, all images were uniformly reconstructed to a specification of 256×256 pixels. The specific distribution of the enhanced dataset is shown in **Table 1**. Finally, the self-built dataset was approximately divided into a training set and a test set at a ratio of 4:1.

Table 1. Enhance datasets.

Type of Rail Surface	oily	wet	dry
Non	70	70	70
GT	350	350	350
Gan	400	400	400
PT	350	350	350

3.1.2. Experimental Configuration

The experimental program in this paper uses PyTorch as the operating platform, and the running environment is configured as follows: The processor is Intel Core i9-9900K, the running memory is 20G, and the graphics card is NVIDIA RTX 4090. The code operating environment is Torch = 1.90, Python = 3.12, and the CUDA version is CUDA12.4. Stochastic Gradient Descent (SGD) is selected as the Optimizer, the batch size is set to 32, and the training is carried out for 50 iterations.

3.1.3. Evaluation Indicators

The evaluation indicators of the experiments in this paper include accuracy, recall rate, precision rate, and F1 score. Among them, the accuracy refers to the proportion of correctly predicted samples in the total number of samples; the recall rate R refers to the proportion of positive samples that are correctly predicted; the precision rate P refers to the proportion of truly positive samples among the samples predicted as positive; the F1 score comprehensively considers the recall rate and precision rate, and it is the weighted harmonic mean of the two.

3.2. Ablation Experiments

3.2.1. Ablation Experiment of Generative Adversarial Network

In the above three datasets, Resnet is used as the feature extraction module to verify whether the data augmentation method is effective. The results are shown in **Table 2**.

Table 2. Dataset validation.

model	(kb)	Accuracy (%)			
	parameter	Non	GT	PT	GAN
Resnet-34	73287	71.42	87.62	82.92	90.75
Resnet-50	82163	53.6	89.5	88.42	91.35
Resnet-101	66699	75.1	89.19	88.42	91.35

It can be seen from **Table 2** that data augmentation can expand the number of training samples, improve sample diversity, and expand the spatial distribution of samples. Therefore, in all models, whether using geometric transformation, pixel transformation, or generative adversarial networks to expand the dataset, the

model performance can be improved and the error rate can be reduced. Comparing geometric transformation and pixel transformation, the rail surface adhesion media are mostly distributed in blocks and their colors are darker than the background, which is similar to the noise added by pixel transformation and can easily mislead the model into making classification errors. So, the dataset expansion effect of geometric transformation is slightly better than that of pixel transformation. When applying the generative adversarial network, a realistic and class-balanced dataset can be generated based on real operation and maintenance data. The recognition accuracy on the ResNet pre-trained network reaches 91.35%, which is better than that of the datasets using geometric transformation and pixel transformation. Compared with traditional data augmentation methods, the generative adversarial network can generate almost an infinite amount of data through random seeds, meeting the large-scale data requirements of deep convolutional neural networks, alleviating the model's data dependence, and enhancing the anti-overfitting ability.

3.2.2. Ablation Experiment of Residual Spinal Fully-Connected Layer

To verify the impact of the improved fully-connected layer on the model performance, the fully-connected layer in Resnet-50 is replaced with the improved structure proposed in this paper. The performance differences between the spinal fully-connected layer and the residual spinal fully-connected layer are tested, and a comparative experiment is conducted on these two models to analyze the performance differences of the model before and after replacement. In the experiment, except for the fully-connected layer being different, the rest of the settings are the same, and the GAN dataset images are used for verification. The results are shown in **Table 3**.

Table 3. Residuals spinal fully connected laminar ablation experiment.

Feature network	decision-making module	Accuracy (%)	Precision (%)	recall rate (%)	Specificity (%)	Parameter (kb)
Resnet	Fc	91.35	91.87	92.16	91.54	82163
	Spinal	91.35	91.87	92.16	91.54	87374
	Spinal-res	92.67	92.99	93.61	92.80	87374

In the experimental settings, “Non” represents Resnet50 using a fully-connected layer, and “Spinal” refers to the situation where the fully-connected layer of Resnet50 is replaced with a spinal fully-connected layer. Although the spinal neural network can optimize the performance of some network models to a certain extent, when applied to Resnet50, it will lead to an increase in the number of model parameters. “Spinal-res” is the model form after replacing the fully-connected layer in Resnet50 with a residual spinal fully-connected layer. By looking at the data in **Table 3**, it can be found that compared with the original fully-connected layer of Resnet-50, the residual spinal fully-connected layer improves the model's recognition accuracy by 1.32%, precision by 1.12%, recall by 1.45%, and

specificity by 1.26%. This means that the residual spinal fully-connected layer has better recognition ability for the rail surface state image features. The reason is that its dual-branch decision-making network has both local and global perspectives, which can effectively reduce the interference of local information, enhance the model's stability and anti-interference ability, and improve the model's robustness.

3.2.3. Ablation Experiment of Attention Mechanism

The dimensions and weights of the feature layers can affect the network training results. During the training of the Resnet network, since the influence of feature information in different dimensions on the model recognition results varies, an attention mechanism module is added after feature extraction. In this experiment, two attention mechanism modules, the single-channel attention mechanism (SE) and the multi-channel attention mechanism (CBAM), are set up, and the GAN dataset images are used for verification. The results are shown in **Table 4**.

Table 4. Residuals spinal fully connected laminar ablation experiment.

Feature network	attention	Accuracy (%)	Precision (%)	recall rate (%)	Specificity (%)	Parameter (kb)
Resnet	Non	91.35	91.87	92.16	91.54	82163
	SE	92.89	93.39	93.95	93.14	98389
	CBAM	93.46	94.23	94.87	94.53	84211

It can be seen from **Table 4** that in terms of the attention mechanism, when it is set to SE, compared with the situation without the attention mechanism module, the model's accuracy rate is increased by 1.54%, the precision rate is increased by 1.52%, the recall rate is increased by 1.79%, and the specificity rate is increased by 1.6%. When the attention mechanism is set to CBAM, compared with the case without this module, the accuracy rate is increased by 2.11%, the precision rate is increased by 2.36%, the recall rate is increased by 2.71%, and the specificity rate is increased by 2.99%. Obviously, the multi-channel attention mechanism (CBAM) has better performance. It can assign corresponding weights according to the importance of different feature information, thereby strengthening the expression of the rail surface state features and improving the overall performance of the model.

4. Conclusion

In the task of rail surface state recognition, due to the small number of samples, the feature space is scarce. To this end, we propose a new method based on a local inference constraint network: First, use a generative adversarial network to construct a rail surface state image dataset, which not only ensures the semantic coherence of the data but also enriches the diversity of sample distribution. Second, use a pre-trained network to accurately extract the feature information of the rail surface state to obtain feature vectors containing high-dimensional semantics. Third, introduce the attention mechanism, focusing on strengthening the weights

of the limited feature vectors and highlighting the key position of the state features. We also innovatively propose a local input decision layer based on the residual network, simulating the information processing mode of the human brain and spine. With the input of local information, the neural network can quickly obtain accurate results with less computation. At the same time, optimize the activation function, fully tap the potential of features, add residual modules, and enhance the model performance while maintaining the stability of the number of parameters. Finally, through comparative experiments, it is strongly confirmed that this method has remarkable results in solving the problem of small-sample image recognition.

Conflicts of Interest

The authors declare no conflicts of interest regarding the publication of this paper.

Fund

The research is funded by the Scientific research project of Hunan Provincial Department of Education, China (Wheel-Rail Adhesion Depth Modeling Based on Rail Surface Condition Identification, under Grant 22B1013).

References

- [1] Pichlik, P. and Bauer, J. (2021) Adhesion Characteristic Slope Estimation for Wheel Slip Control Purpose Based on UKF. *IEEE Transactions on Vehicular Technology*, **70**, 4303-4311. <https://doi.org/10.1109/tvt.2021.3072484>
- [2] Liu, J., Liu, L., He, J., Zhang, C. and Zhao, K. (2020) Wheel/Rail Adhesion State Identification of Heavy-Haul Locomotive Based on Particle Swarm Optimization and Kernel Extreme Learning Machine. *Journal of Advanced Transportation*, **2020**, 1-6. <https://doi.org/10.1155/2020/8136939>
- [3] Yu, H., Peng, C., Liu, J., Zhang, J. and Liu, L. (2024) Improved Metric-Learning-Based Recognition Method for Rail Surface State with Small-Sample Data. *IEEE Access*, **12**, 4985-4996. <https://doi.org/10.1109/access.2023.3347634>
- [4] Chen, Y., Deng, C., Sun, Q., Wu, Z., Zou, L., Zhang, G., *et al.* (2024) Lightweight Detection Methods for Insulator Self-Explosion Defects. *Sensors*, **24**, Article 290. <https://doi.org/10.3390/s24010290>
- [5] Alruwaili, M., Atta, M.N., Siddiqi, M.H., Khan, A., Khan, A., Alhwaiti, Y., *et al.* (2024) Deep Learning-Based YOLO Models for the Detection of People with Disabilities. *IEEE Access*, **12**, 2543-2566. <https://doi.org/10.1109/access.2023.3347169>
- [6] Yin, Z., Liu, F., Geng, H., Xi, Y., Zeng, D., Si, C., *et al.* (2024) A High-Precision Jujube Disease Spot Detection Based on SSD during the Sorting Process. *PLOS ONE*, **19**, e0296314. <https://doi.org/10.1371/journal.pone.0296314>
- [7] Bell, S., Zitnick, C.L., Bala, K. and Girshick, R. (2016). Inside-Outside Net: Detecting Objects in Context with Skip Pooling and Recurrent Neural Networks. 2016 *IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, Las Vegas, 27-30 June 2016, 2874-2883. <https://doi.org/10.1109/cvpr.2016.314>
- [8] Tsyganov, S. and Ibadullayev, D. (2023) On the Sharing Signals Measured with DSSD Detector.
- [9] Hong, Q., Dong, H., Deng, W. and Ping, Y. (2024) Education Robot Object Detection

with a Brain-Inspired Approach Integrating Faster R-CNN, Yolov3, and Semi-Supervised Learning. *Frontiers in Neurorobotics*, **17**, Article 1338104.

<https://doi.org/10.3389/fnbot.2023.1338104>

- [10] Yu, C., Sun, Y., Cao, Y., He, J., Fu, Y. and Zhou, X. (2023) A Novel Wood Log Measurement Combined Mask R-CNN and Stereo Vision Camera. *Forests*, **14**, Article 285. <https://doi.org/10.3390/f14020285>
- [11] Luo, H., Cai, L. and Li, C. (2023) Rail Surface Defect Detection Based on an Improved YOLOv5s. *Applied Sciences*, **13**, Article 7330. <https://doi.org/10.3390/app13127330>
- [12] Karras, T., Laine, S. and Aila, T. (2019) A Style-Based Generator Architecture for Generative Adversarial Networks. 2019 *IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, Long Beach, 15-20 June 2019, 4396-4405. <https://doi.org/10.1109/cvpr.2019.00453>