Structural Durability & Health Monitoring

Tech Science Press

https://doi.org/10.32604/sdhm.2026.078140

ARTICLE

An Intelligent Assessment of Rail Surface Defects over the Life-Cycle Based on Improved Transformer Networks

Ziliang Yang[1], Mykola Sysyn[2], Jin Li[1], Jizhe Zhang[1], Jian Liu[1] and Lei Kou[1,3,*]

[1]School of Qilu Transportation, Shandong University, Jinan, China
[2]Institute of Railway Systems and Public Transport, TU-Dresden, Dresden, Germany
[3]Key Laboratory of Rail Infrastructure Durability and System Safety, Tongji University, Shanghai, China

*Corresponding Author: Lei Kou. Email: lei.kou@sdu.edu.cn

Received: 24 December 2025; Accepted: 03 March 2026; Published: 18 May 2026

ABSTRACT: Accurate assessment of the failure stage of rail rolling contact fatigue (RCF) is critical for guiding timely maintenance by track personnel, ensuring safe rail operations, and reducing maintenance costs. Although various methods have been developed to detect rail damage and classify surface defects, the rolling contact fatigue failure state of rails has not yet been comprehensively and objectively evaluated. This paper introduces the application of image processing and improved deep-learning network algorithms in rail failure evaluation and judgment. Based on Swin Transformer, a deep learning network is developed. By dividing the rail rolling contact fatigue failure process into five life-cycle stages, the proposed network can identify the current stage of the rail contact surface within its service life. Finally, compared with the commonly used neural network model, the recognition rate of the improved Transformer can reach 98.48%, which is far better than other network structures. The enhanced neural network forms a simple system for evaluating the life of the orbit. The system identifies potential failure hazards on rail surfaces. The results also provide early warning predictions for rolling contact fatigue failure.

KEYWORDS: Defect detection; swin transformer; rail surface; rolling contact fatigue; full life cycle; rail damage evaluation

1 Introduction

Railways remain a cornerstone of global economic development, where track safety is paramount for operational stability. Despite significant resources allocated to maintenance, nearly half of railway accidents still stem from track and fastener defects. Accurate assessment of rail surface damage is thus essential for transitioning from reactive repairs to proactive, reliability-centered maintenance, ultimately optimizing maintenance efficiency and costs. However, according to statistics from the Federal Railroad Administration of the United States, nearly half of railway accidents are caused by track and fastener defects. Accurate assessment of rail surface damage can help railroad maintenance personnel plan and perform maintenance tasks more effectively, reduce unnecessary maintenance costs, and improve maintenance efficiency.

With the development of artificial intelligence technologies, the inspection and maintenance of rails are gradually becoming automated and intelligent. Through automated classification of rail surface damage, data-driven decision-making can be introduced into railroad maintenance work to improve the accuracy and objectivity of decision-making and to discover and repair safety hazards promptly.

In recent years, railroad surface damage detection has become a research hotspot. Early systems focused primarily on detection, such as the eddy current sensor developed by Oukhellou et al. for rail fractures [1]. Subsequent studies utilized 3D laser cameras with ResNet50 [2], and feature extraction algorithms for cracks covering small cross-sectional areas [3]. Further research, including unsupervised learning on vibration signals [4–6] and machine vision-based surface defect detection [7,8], has made significant progress in identifying visible faults, though challenges remain in predicting continuous surface damage patterns.

Various techniques have been explored to further investigate the laws of rail damage. Wu et al. used ultrasonic detection to detect internal rail defects and classify them based on the position of cracks [9], achieving 89% accuracy in controlled lab environments but dropping to 72% in field conditions with complex noise interference. Xing et al. developed an improved YOLOv3 framework for detecting surface defects on track wheels, achieving classification detection of four types of wheel surface defects [10]. Prof. Ping Wang's team develops a meshing method combined with an accurate conjugate gradient (CG) method for contact mechanics in arbitrary 3D contact geometries. To analyze wheel-rail contact contact problems and thus detect rail failures [11]. Acikgoz and Korkmaz from Turkey proposed an effective multi-scale residual convolutional network (MSRConvNet) model for classifying different types of railway track defects [12], its fixed 3 × 3/5 × 5 kernel combination restricts adaptability to varying defect scales. A team from the Nanjing University of Aeronautics and Astronautics [13] combined laser ultrasonic technology with mixed intelligent methods to achieve the rapid deep classification of artificial rolling contact fatigue (RCF) defects. Aydin et al. [14] fused features from SqueezeNet and MobileNetV2 deep learning models to detect and classify surface defects on railways. Yang et al. [15] proposed a contour and semantic feature alignment fusion network (CSANet) with bidirectional feature alignment to enhance rail surface defect detection by exploring cross-modal features from contour and semantic perspectives, outperforming 12 state-of-the-art algorithms in evaluation metrics on an industrial RGB-D dataset. Ye et al. [16] present a laser-based 3-D semantic segmentation method for rail surface defect detection, combining precise laser measurement with deep learning, achieving high detection accuracy and enabling end-to-end 3-D defect characterization for safer and more efficient railway maintenance. Wang et al. [17] introduce a 3D tensor-based point cloud and image fusion (T-PCIF) method for robust rail surface defect detection and 3D measurement, achieving an accuracy rate of 86.27% MDPA and 0.7018 MDIoU values through tensor analysis and eigenvalue decomposition, but computational cost exceeds 3.2G FLOPs per sample. Guo et al. [18] propose RailFormer, a novel Transformer-based system for precise and efficient detection of rail surface defects that outperforms other deep learning models like SegFormer, Swin Transformer, ViT, and UNet on public and customized datasets, highlighting its potential for future deployment in railway track inspection applications. Xie et al. [19] present a novel data-driven convolutional regression scheme named RCNet that can accurately and efficiently detect rail corrugation roughness.

However, traditional rail damage classification mainly focuses on differences in damage location, depth, and size. Such research is of great help for rail damage detection and fine-grained research, but it does not provide significant insights into the evolution patterns of rail damage. Although scholars studying the evolution mechanisms of rail rolling fatigue and fracture from the perspective of materials science have made many outstanding contributions [20–22], there has been limited breakthrough research applicable to railway maintenance workers for rapid judgment of the entire lifecycle damage stage of tracks in engineering practice. A research team from Dresden University of Technology has determined the fatigue stage of tracks through long-term research on cracks and vibrations, aiming to achieve early prediction of faults. However, this research is limited to the final stage of fault occurrence and does not reflect the evolution patterns of the entire lifecycle. The analysis method combined the research foundation of the research team from Dresden University of Technology [23,24] and used an improved neural network to classify and identify surface fatigue

damage on railway tracks throughout the entire lifecycle. However, it has 78% accuracy for terminal failure prediction but ineffective for early-stage detection. This paper solves the problem that the German Railway Group has long-term required experts to manually and subjectively determine the damage status of rails, and also overcomes the problem that the complexity of the Chinese railway rail damage evaluation standards and the numerous parameters make it impossible to comprehensively evaluate the damage status of rails in the field promptly at the first time.

To address these limitations, this paper proposes an intelligent assessment framework for the entire life-cycle of rail rolling contact fatigue (RCF). The study utilizes a task-oriented MultiScale-Swin Transformer architecture to fuse hierarchical features through dynamic window attention, enhancing the discrimination of RCF stages. The core work includes: (1) establishing a comprehensive rail surface damage dataset categorized into five life-cycle stages based on expert consensus; (2) developing the MultiScale-Swin Transformer network with an implicit cross-stage feature recycling strategy; and (3) validating the model's performance against traditional architectures to ensure high-precision and real-time assessment capabilities.

2 Data Acquisition and Experiments

2.1 Experiments

Continuous spot monitoring of rail fatigue damage and precise evaluation by experts in the field employed by Deutsche Bahn constitute the rail surface damage data set. Such data sets are rare in the industry and are an objective evaluation solution based on competent critics. The experimental data in this paper mainly comes from the daily maintenance team of the operating railway line, which is the most reflective of the actual engineering of the data source, Wuhan Railway Bureau has a complex terrain, including small radius curves, mountains, culverts, heavy railroads, high-speed railroads and so on. This section is mainly to meet the demand for data diversity. The main data of the Hohhot Engineering Section comes from the heavy-load railroad for coal transportation, mainly to meet the demand of acquiring data in a short time. Part of the data acquisition environment and acquisition results are shown in Fig. 1. The data analysis mainly adopts the machine vision method, so the data acquisition is mainly through the high-definition camera to take pictures of the rail surface state. More than 1000 pictures of rail damage data were collected after two years of data accumulation, and the pictures that can reflect the rolling fatigue state of the rail surface were finally selected as the data set.

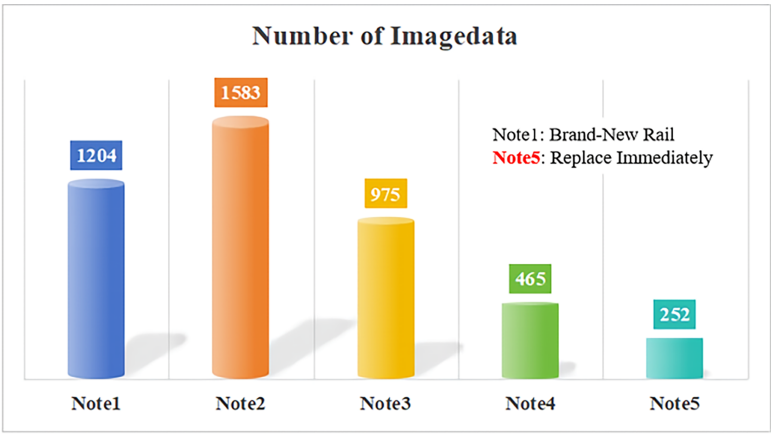

Figure 1: Experimental site in Wuhan and selected data.

2.2 Data Preparation

When assessing rail damage in a given area, Deutsche Bahn AG used to define the level of rail damage by inviting renowned experts in the railroad field to discuss the problem together. However, this method is inefficient and costly. Therefore, the development of an algorithm to directly determine the degree of damage on the surface of rails through images is the research objective of this paper. Since the original images of the expert-identified grades do not specify the physical parameters such as damage size and crack length, we train the images of the expert-identified grades as a reference dataset, with the aim of giving the damage grades of the railroads through the automatic recognition of the images without considering the specific physical parameters, so that the workers can determine the results initially through portable testing instruments when they perform the inspections.

The algorithm for evaluating the fatigue status of railway track surfaces is achieved by classifying images of surface defects on the tracks. Therefore, based on subjective evaluation (in the form of ratings from 1 to 5), the classification of Note1–5 is mainly based on the subjective evaluation of the collected data by experts who have long been engaged in rail damage detection. The damage annotations were conducted by more than three experienced railway inspection experts, and consensus labeling was adopted to ensure reliability.

In this study, the severity of rail surface damage is categorized into five grades, labeled as Note 1 to Note 5. This nomenclature is strictly aligned with the official condition assessment standards (Zustandsnoten) of the German Railway (Deutsche Bahn AG). These grades represent a unified professional evaluation system used by DB and our research consortium to guide maintenance decisions, where Note 1 signifies a 'very good' condition and Note 5 indicates 'failed/replacement required' status. The results of this evaluation serve as reference annotations for artificial intelligence. Note1 indicates a brand new track, Note2 states that the rails are not brand new but also do not show more obvious defects, only some minor damage, Note3 indicates a track that requires grinding and maintenance, Note4 represents a track with significant safety hazards, and Note5 indicates a track that needs immediate replacement. This work mainly captures the most severely damaged area on the surface of each track as the research object. The captured images are then saved in the corresponding independent datasets. Finally, the five datasets need to be labeled and established in the same dataset. Data was collected using a handheld high-definition camera by the experimenter. Data was collected monthly.

In order to ensure the wide applicability and accuracy of the algorithm railway workers collected a total of 4479 images that have been appraised and rated by a committee of experts as the basic dataset for this study. Fig. 2 illustrates the amount of data for each evaluation level (Note) that make up the original data distribution. Thus, the data collection has completed the initial preparation, but due to the limited number of image data, data augmentation is needed. In the preliminary stage of this study, several data augmentation strategies, including illumination variation and random cropping, were experimentally evaluated. However, these operations were found to potentially alter the physical appearance and boundary characteristics of rail surface damage. To ensure engineering rigor and preserve the authenticity of fatigue-related visual patterns, only rotation-based augmentation (90°, 180°, and 270°) was finally adopted in the formal experiments. This conservative augmentation strategy reduces the risk of introducing non-physical artifacts while maintaining sufficient data diversity for model training. Future work will further investigate physically consistent data augmentation strategies to enhance model robustness under varying inspection conditions. After augmentation, the dataset size increased from 4479 images to nearly 20,000 images. In order to make the results uniform and widely applicable, the rails for data collection have different materials, as shown in Table 1.

Figure 2: Distribution of note images.

Table 1: Rail material.

Header 1	Steel No.	σ_0	σ_S	C	Mn	Si
CHN60	U71Mn	1079	686	0.65~0.77	1.1~1.5	0.15~0.35
	PD2	1175	800	0.74~0.82	0.70~1.00	0.15~0.35
UIC60	900A	880		0.60~0.80	0.80~1.30	0.10~0.50
	1100	1080		0.60~0.82	0.80~1.30	0.30~0.90

Before the calculation, here also need to standardize the size of the captured image. For the input data of the convolutional neural network, the format of the image must be a square where the side length is divisible by two. Most neural networks are 512 × 512, 256 × 256 or 224 × 224 units in size. In this article, the size of the image is uniformly changed to 224 × 224 as images of this size are suitable for most networks. An example of the data set for each note is shown in Fig. 3 below.

Thus the main steps in the operation of the neural network are, importing the prepared data into the training database, loading the required neural network and tuning it to fit the network parameters for the purpose of this paper. Finally, the output layer needs to be modified and the output results are categorized into five categories. Of course, other parameters can be adjusted according to the actual situation. Finally, the designed network needs to be transmitted for training.

During training, the Adam optimizer was employed with an initial learning rate of 1×10^{-4}. The batch size was set to 16, and categorical cross-entropy was used as the loss function. All models were trained for 100 epochs, and early stopping was applied based on validation accuracy to prevent overfitting.

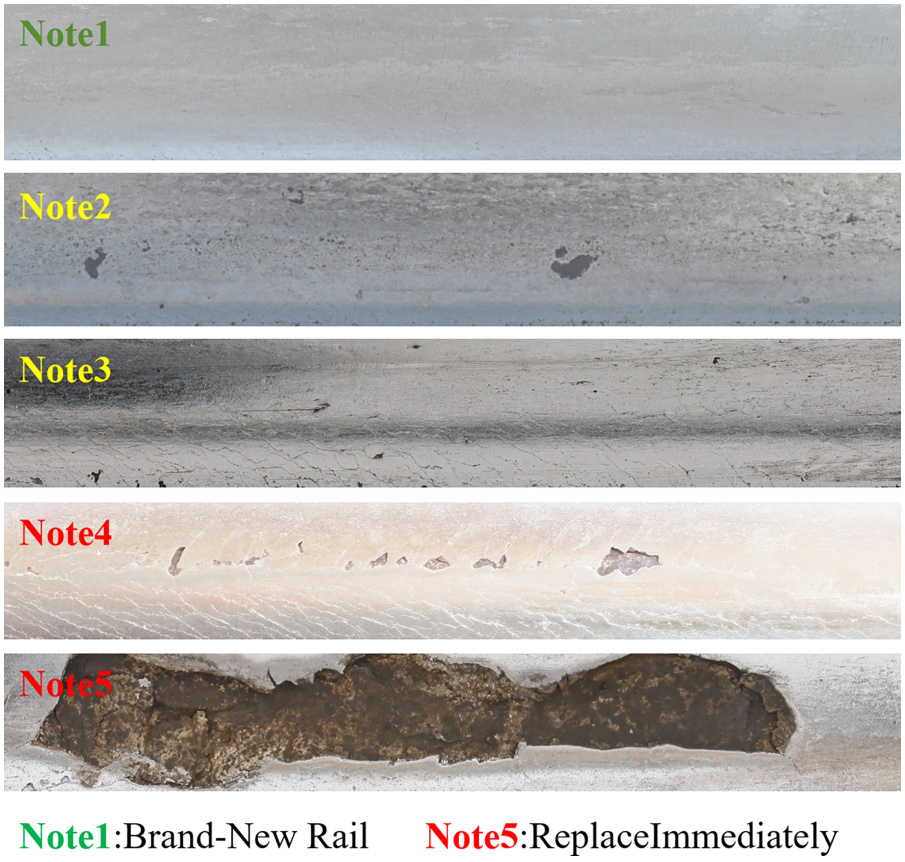

Figure 3: Example of note data.

The proposed intelligent assessment framework demonstrates significant potential for multi-scenario engineering applications. Due to its high computational efficiency (83 FPS) and robust feature extraction capabilities, the method can be implemented in two distinct operational modes. On one hand, it is suitable for integration into vehicle-mounted high-speed inspection systems, enabling continuous 'online diagnosis' and large-scale monitoring of rail networks. On the other hand, the model's architecture supports deployment on handheld portable devices. This allows maintenance personnel to perform instantaneous 'point-and-shoot' assessments during routine manual patrols, where a single surface image can be automatically processed to provide an objective life-cycle stage classification. Such dual-mode flexibility effectively bridges the gap between automated system-wide monitoring and flexible, localized site inspection.

3 Neural Network Algorithm

Building on ResNet, YOLO and Transformer, the state-of-the-art frameworks in the machine vision application research domain, the network architecture has undergone optimization and comparative evaluation. All experiments were conducted on a workstation equipped with an NVIDIA RTX 4090 GPU and an Intel Core i9-14900K CPU.

3.1 Improved YOLO and ResNet

With the in-depth research of artificial intelligence in the field of computer vision, many network models have been widely applied due to their excellent performance. Among them, ResNet50 [25], Parallel Spatial Attention Module [26], and Yolov5 [27] neural networks are important components in the field of deep learning and play a crucial role in computer vision tasks.

To establish a robust baseline for rail defect assessment, the YPResNet50 is developed by integrating the residual learning capability of ResNet50 with advanced spatial attention and detection heads. Instead of standard feature mapping, the network utilizes a Parallel Spatial Attention (PSA) module to capture complex spatial contexts of rail fatigue across different network levels. This parallel mechanism allows the model to prioritize critical damage regions while suppressing irrelevant background noise, which is essential for characterizing subtle fatigue states on complex rail surfaces. Furthermore, the YOLOv5 detection header is fused into the architecture to leverage its multi-scale feature fusion and lightweight reasoning capabilities. This integration ensures a high inference speed suitable for real-time rail inspection without compromising accuracy, providing a flexible and efficient backbone for assessing the rolling fatigue lifecycle.

Thus combining the respective features and advantages of ResNet50, Parallel Spatial Attention Module (PSA-Module) and Yolov5 neural network, more powerful and efficient deep learning models can be built. It is worth to make a scientific attempt will be to use ResNet50 as a framework to add the PSA-Module's attention model to increase the classification accuracy, and finally utilize Yolov5's detection head to complete the evaluation. In this study, they are combined into a new and improved ResNet50 model called YPResNet50 as shown in Fig. 4 and parameter descriptions for the corresponding modules in the neural network in Figs. 5–7.

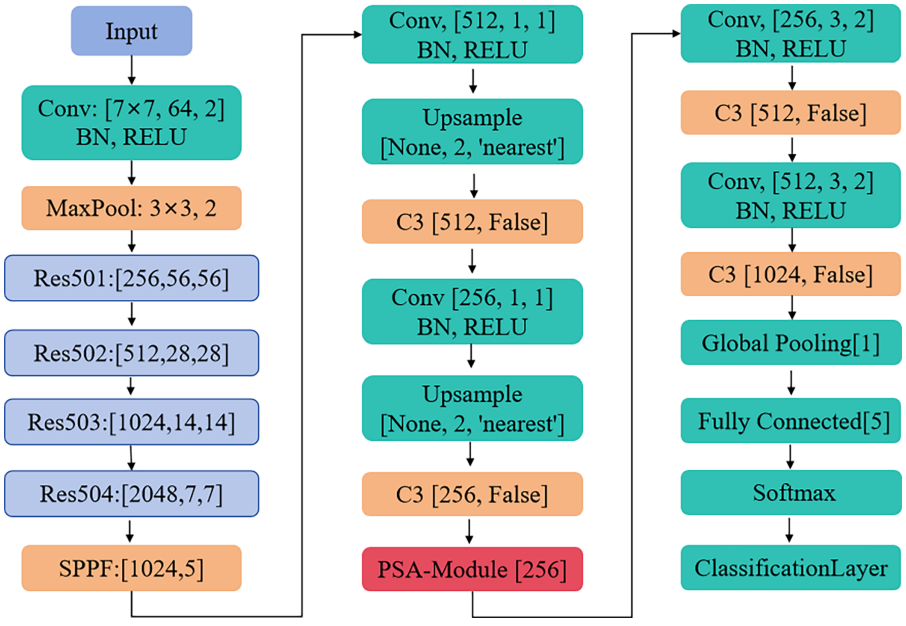

Figure 4: YPResNet50 neural network architecture.

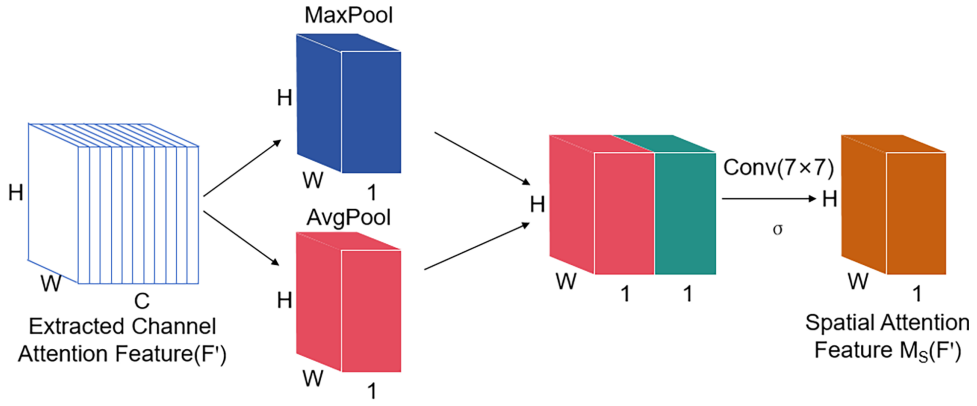

Figure 5: Structure of PSA-Module.

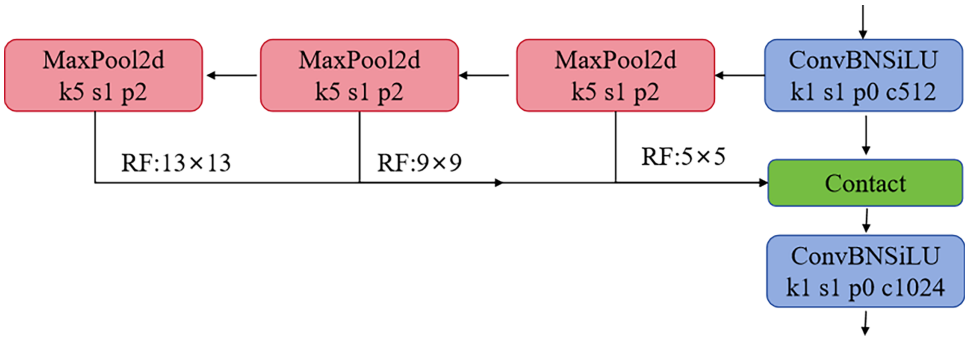

Figure 6: Structure of SPPF.

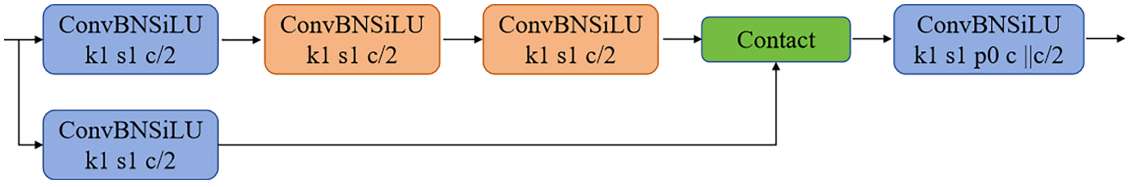

Figure 7: Structure of C3.

In the above structure, Backbone includes ResNet501, ResNet502, ResNet503, and SPPF modules. YOLOv5 Head includes a series of Conv, Upsample, Concat, and C3 modules that are used to generate the detection results. Detect is the final detection module that is used to output the target detection results. The PSA module is inserted after the first downsampling operation of the YOLOv5 Head structure. The PSA module should take as input a feature map of 256 input channels and generate the same number of output channels. The output feature maps will then continue to flow through the remainder of the YOLOv5 head structure for processing. In the paper, the detection part behind the YOLOv5 head structure is replaced with a global average pooling layer, and then a fully connected layer outputting 5 categories is connected and the final classification output is obtained by a Softmax activation function.

3.2 Improved Transformer Network

The major creative contribution of this study is the integration of an adaptive reweighting mechanism within the MultiScale-Swin Transformer. Unlike standard Transformers, our model distinguishes the importance of features across various scales specifically for RCF evolution stages, allowing for the capture of both micro-pitting (fine-scale) and macro-crack propagation (large-scale). Transformer [28] is a deep learning architecture centred on self-attention mechanisms. It abandons the sequential processing of sequence data by traditional Recurrent Neural Networks (RNN) and Long Short-Term Memory Networks (LSTM). This architecture enables more effective capturing of long-range dependencies within sequential data. Its core elements include self-attention mechanism, forward propagation layer and encoder-decoder structure. High parallelization capabilities significantly accelerate both training and inference processes. It is also capable of modelling long sequential data and can be easily scaled to larger model sizes for more complex tasks. Building upon the enhanced ResNet, this study further optimizes the Transformer architecture to identify the optimal classification algorithm through comparative performance analysis.

To address the challenge of identifying rail surface defects with varying sizes—ranging from sub-millimeter micro-cracks to large-scale spalling—the Swin Transformer [29] is employed for its distinctive hierarchical architecture. Unlike standard Transformers, this design mimics the multi-scale feature extraction of CNNs, where shallower layers preserve fine-grained spatial details while deeper layers extract global semantic contours. To overcome the lack of inter-window communication in independent patch mapping, the model utilizes Shifted Window Multi-head Self-Attention (SW-MSA), as illustrated in Figs. 8 and 9. This facilitates feature exchange across non-overlapping windows, which is critical for capturing the continuous propagation patterns of rolling contact fatigue (RCF). By integrating this hierarchical structure, the network can simultaneously process high-resolution local details and global crack morphologies, making it inherently more suitable for the dense prediction and pixel-level sensitivity required in the rail life-cycle assessment.

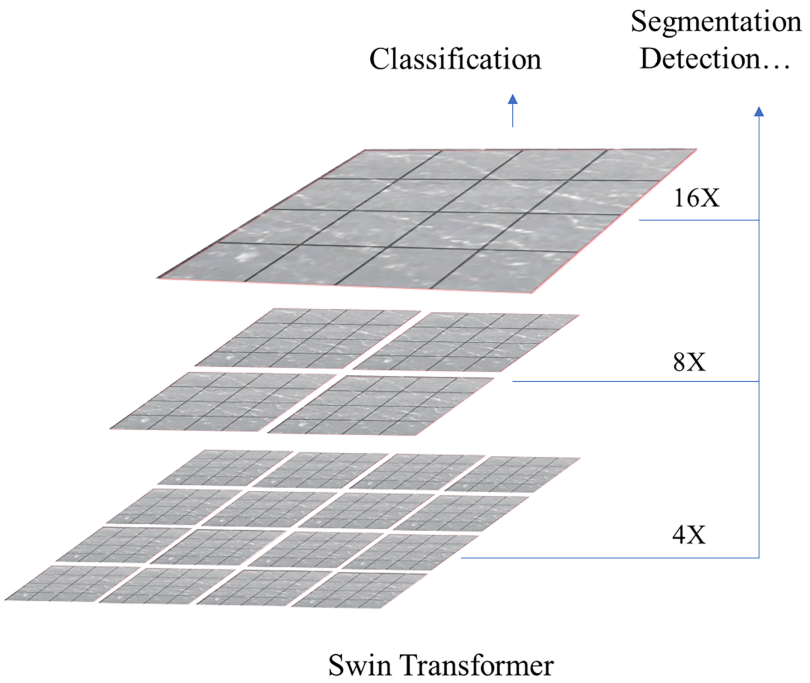

Figure 8: Swin transformer's image processing.

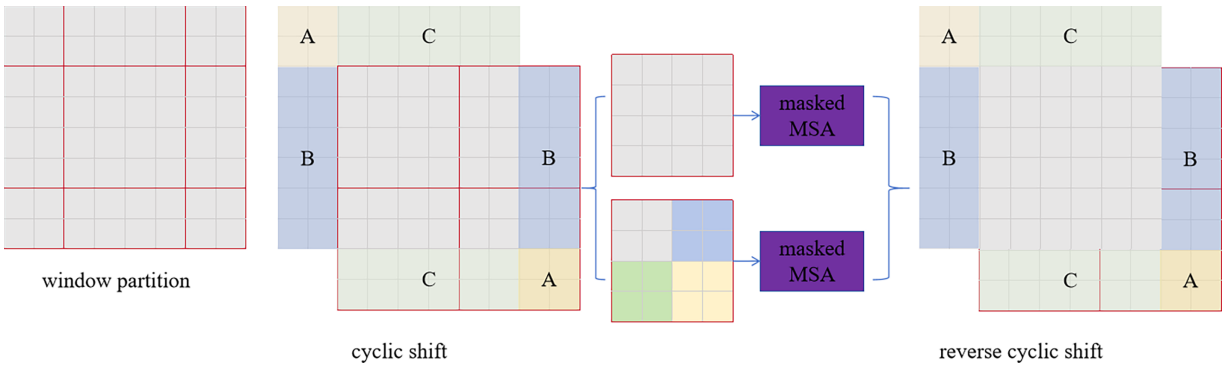

Figure 9: Illustration of an efficient batch compution approach for self-attention in shifted window.

In this paper, an attention mechanism is employed to reweight multi-scale features adaptively. These weights are determined by the relevance of features to different fatigue evolution stages rather than simple spatial variations. Subsequently, the features are fused based on these calculated weights. In the multi-scale Transformer model, the attention mechanism weights and fuses feature maps of various scales to highlight critical information. This approach addresses the limitations of standard Transformers [30]. Specifically, it introduces the ability to differentiate feature importance during multi-scale fusion. The specific optimisation network architecture is shown in Fig. 10, which demonstrates the powerful feature representation capability. Multi-scale feature fusion can combine different levels of information, providing details at a shallow level and extracting semantics at a deeper level.

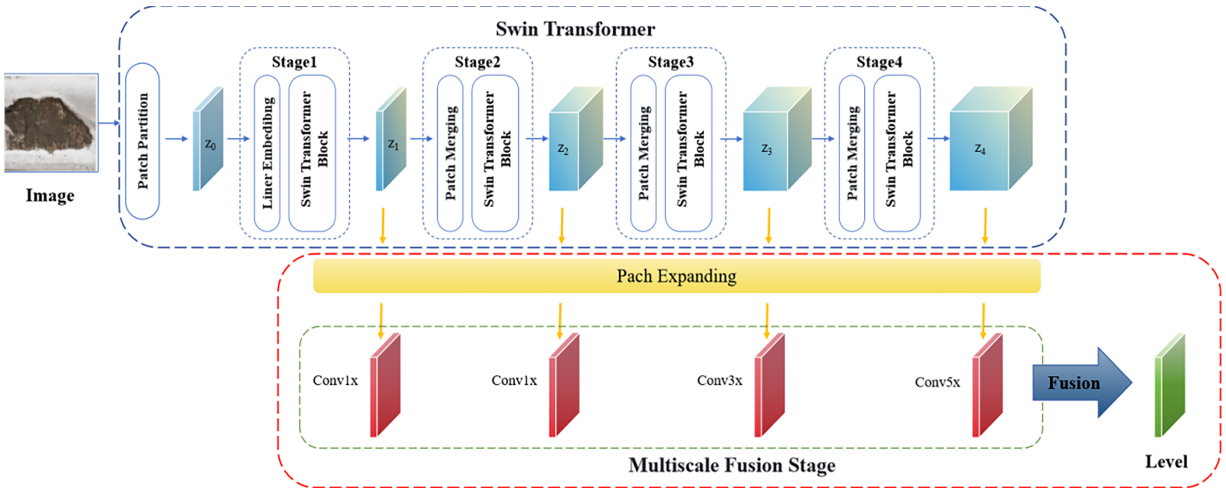

Figure 10: Structure of MultiScale-SwinTransformer.

The network integrates multi-scale feature relationships during global modeling. Simultaneously, it distinguishes the weighting ratios of these features. This distinction facilitates a clearer understanding of the relationship between local regions and the global context. Consequently, recognition accuracy is significantly improved. The combination of these mechanisms provides the model with both detailed and semantic information. This dual access enhances the representation of complex data. Furthermore, it supports diverse tasks and effectively boosts overall model performance.

Multi-scale feature fusion allows the model to learn diverse data representations. This enhances adaptability and reduces the risk of overfitting. Additionally, multi-scale features can be processed in parallel.

By leveraging the parallel computing advantages of the Transformer, hardware resources are fully utilized to accelerate training and inference. It should be noted that the so-called cross-stage feature recycling mechanism does not correspond to an independent network module. Instead, it is implicitly realized by the hierarchical architecture of the Swin Transformer, where shallow-stage features related to early fatigue patterns are progressively preserved, reweighted, and integrated into deeper representations through inter-layer attention operations. This design enables information continuity across different fatigue stages without introducing additional parameters.

3.3 Network Performance Comparison

The comparison of VGG16, AlexNet, ResNet50, YOLOv5, and Swin Transformer stems from systematic evaluation of railway defect detection requirements. The comparison of VGG16, AlexNet, ResNet50, YOLOv5, and Swin Transformer is intended to provide representative engineering baselines rather than to claim state-of-the-art performance for all compared models, as in Table 2. More recent architectures such as Vision Transformer and Swin Transformer are included to reflect the performance level of modern deep learning models for rail surface defect analysis.

Table 2: Network performance comparison.

Model Characteristic	VGG16	AlexNet	ResNet50	YOLOv5	Swin-T
Feature Hierarchy	Single-scale	Shallow	Multi-level	Multi-scale	Hierarchical
Receptive Field	Fixed	Limited	Adaptive	Pyramid	Dynamic
Real-Time Speed	38	45	62	83	71
Recall	68%	72%	79%	85%	88%

This quantitative comparison proved to highlight the main shortcomings of earlier network architectures such as VGG and AlexNet, firstly the limited ability to handle sub-millimetre defects at multiple scales, secondly the fixed receptive field is not suitable for variable crack morphology, and lastly the high computational cost makes it unsuitable for real-time detection.

However, with the ResNet50 backbone, the feature abstraction capability is balanced by a 50-layer depth optimisation while residual connectivity prevents gradient loss during backpropagation of fine surface textures, low visibility contrast is enhanced by using the Parallel Spatial Attention (PSA) module, and the elimination of channel excitations suppresses ballast interference in the rail bed image.

The layered Swin transformer solves three major problems in rail inspection, firstly cross-scale attention is achieved by moving the window partitions, combined with local windows to capture micropitting and global connections to detect crack propagation paths, eliminating scale differences. Secondly, multiple self-attention is used to model contextual relationships. Finally a material degradation model is considered with hierarchical feature cycling to simulate metal fatigue processes.

4 Results

4.1 Performance Evaluation and Comparative Analysis

The training process of the constructed neural networks provides a scientific basis for evaluating their learning capabilities. As illustrated in Fig. 11, the MultiScale-SwinTransformer and YPResNet50 demonstrate stable convergence and efficient loss reduction. The automatically generated training curves objectively

reflect the robustness of our modified structures throughout the iterations. Specifically, the hierarchical feature learning of the Transformer-based network allows for a more rapid adaptation to the rail surface damage dataset, ensuring that the model captures essential discriminative features early in the training phase, which lays the foundation for high-precision assessment.

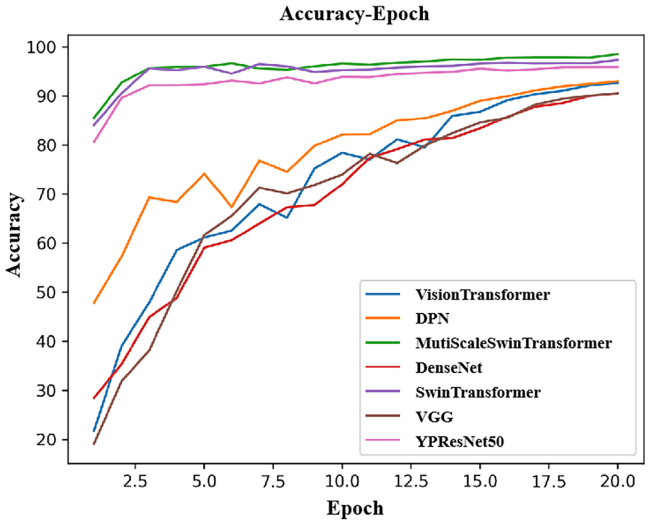

Figure 11: MultiScale-SwinTransformer training process and results.

Quantitative results summarized in Table 3 further validate the superiority of the proposed methods against several baseline architectures. The MultiScale-SwinTransformer (Ours) leads all models with a peak validation accuracy of 98.48%, significantly outperforming traditional CNNs like ResNet50 (71.93%) and VGG16 (92.64%). While the Transformer model involves a higher parameter scale (89.3M), it achieves the highest F1-score of 96.2% and mAP@0.5 of 95.7%, confirming its precision in complex damage identification. Meanwhile, the YPResNet50 (Ours) delivers an optimal balance with an inference speed of 83 FPS, representing a 29.7% improvement over standard ResNet50, which is crucial for real-time inspection requirements.

Table 3: Results comparison.

Model	Accuracy (%) ↑	Recall (%) ↑	F1-Score (%) ↑	mAP@0.5 ↑	FPS ↑	Params (M) ↓
VGG16	92.64	84.7	88.4	80.3	38	138
DenseNet121	93.62	86.1	89.6	82.7	52	20.1
DPN92	95.04	88.3	91.4	86.9	58	37.2
Vision Transformer	95.16	89.5	92.1	87.3	47	86.4
YPResNet50	96.20	91.7	93.8	90.5	83	26.8
Swin Transformer	97.05	92.4	94.6	93.1	71	28.0
Multi-Scale SwinTransformer	98.48	94.2	96.2	95.7	63	89.3

To verify the practical reliability of the system, the trained MultiScale-SwinTransformer was tested on challenging samples, particularly focusing on the transition between Note 3 and Note 4 stages. As

shown in Fig. 12, the model demonstrates high confidence (exceeding 99%) in most scenarios, with results consistently aligning with subjective expert judgments from China Railway Group. Although a rare misclassification occurred in Fig. 12c due to surface rust interference—identifying a Note 3 defect as Note 4—the model successfully flagged the area as a hazard. This localized error can be further mitigated through targeted retraining with diverse environmental data, proving that the network's accuracy is sufficient for objective, integer-value evaluation of rail fatigue stages in field practices.

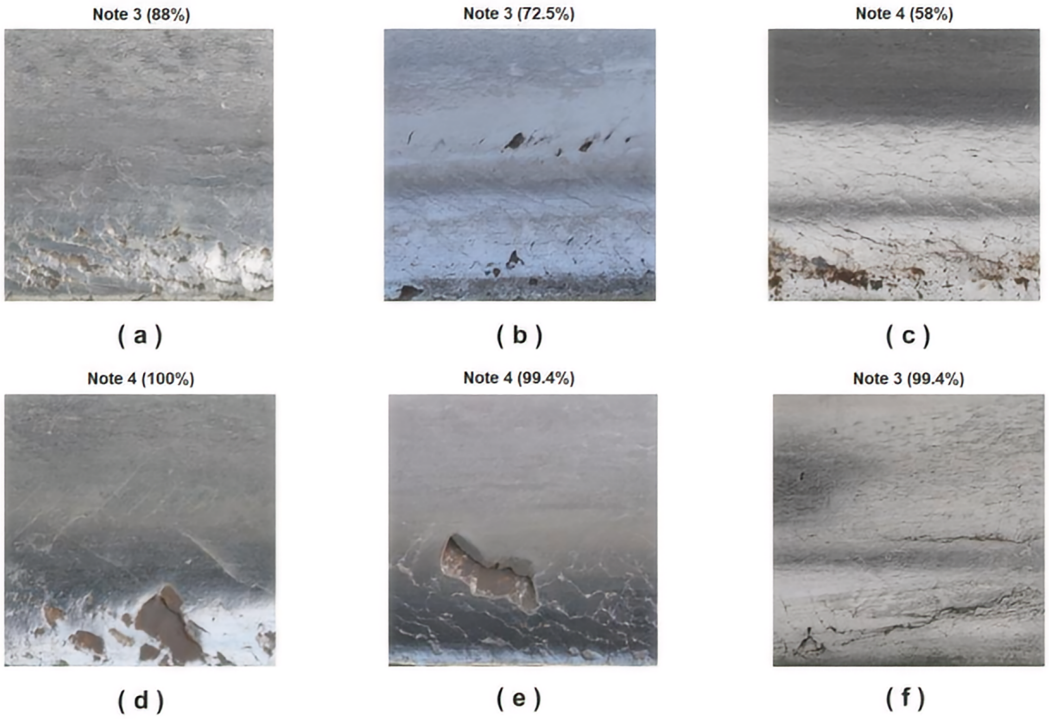

Figure 12: MultiScale-SwinTransformer training process and results.

The superiority of the core method, MultiScale-SwinT, is attributed to its hierarchical structure which mimics the physical process of metal fatigue. While ResNet-based models struggle with sub-millimeter defects due to fixed receptive fields, our improved Transformer utilizes shifted window partitions to capture multi-scale crack morphologies, leading to a 2.28% accuracy lead over the second-best model (YPResNet50).

4.2 Methodological Advantages and Engineering Impact

With the help of the improved network structure proposed in this article, the research team has proposed a new strategy for rail inspection and maintenance, taking into account the practical experience of the Deutsche Bahn Group. When the integrated rail inspection vehicle with machine vision inspection lens travels through the key monitoring and inspection area of the rails, it captures images of the rails in the area. The collected images are processed by the camera front-end pre-processing module and then entered into the MultiScale-SwinTransformer diagnostic evaluation. For maintenance decision-making, the five damage levels are further grouped into three broader categories. The results of the evaluation are classified into three categories according to the safety level: corresponding to Note1 and 2, the area is considered to be safe without manual inspection and maintenance, and the node of the road section is marked in green. When the rating is Note3 the area is judged to be of limited safety and needs to be scheduled into a manual inspection

and maintenance programme, the node in the area is marked yellow. When the rating is Note4 and 5 it is considered that the area rails are in a state of hazardous operation and need to be maintained or replaced immediately, the area is marked in red, as shown in Fig. 13.

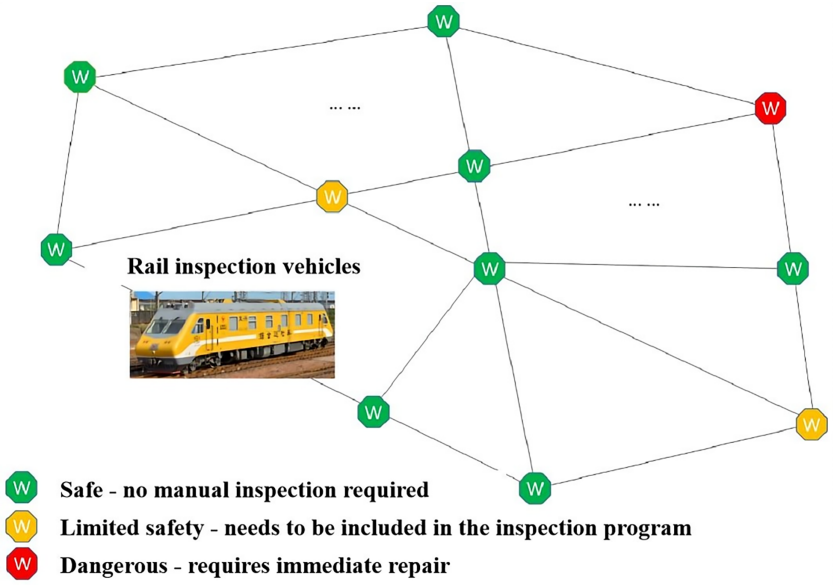

Figure 13: AI-based inspection schematic.

The economic impact of this strategy is quantified in Fig. 14, comparing it with conventional manual inspection costs. Field deployment data and estimations based on 2023 German Railway Engineering Standards indicate that the system can reduce annual inspection costs by 39%. As shown in the cumulative cost analysis (Fig. 14), the amortized expenditure of the AI system (approx. €620/km) is significantly lower than the manual inspection cost (€2840/km). By implementing this scoring mechanism, the frequency of manual inspections is optimized, effectively extending the rail service life and reducing the total life-cycle maintenance expenditure.

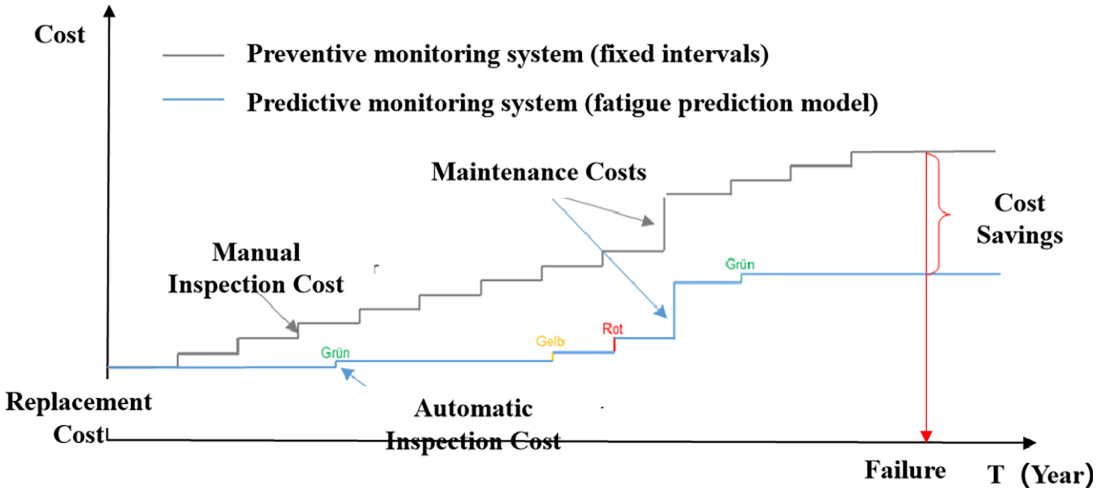

Figure 14: Conservation strategies based on objective evaluation of rails.

5 Conclusions

This study develops an intelligent assessment framework for the entire life-cycle of rail rolling contact fatigue (RCF) by integrating expert-driven stage classification with advanced deep learning architectures. The core contribution lies in the development of the MultiScale-Swin Transformer, which addresses the limitations of traditional CNNs in capturing hierarchical features of fatigue evolution. Achieving a validation accuracy of 98.48%, this model effectively identifies the transition between micro-pitting and macro-crack propagation, providing a level of precision that surpasses standard residual networks. Furthermore, to meet the requirements of practical engineering applications, the auxiliary YPResNet50 model was optimized to achieve a detection speed of 83 FPS, ensuring that high-accuracy assessment can be realized in quasi-real-time during mobile track inspections.

Beyond algorithmic improvements, this research establishes a systematic link between automated image recognition and infrastructure maintenance strategy. By digitizing the RCF evolution into five objective stages, the proposed system enables a transition from reactive repair to proactive life-cycle management. The classification of rail conditions into safe, limited danger, and hazardous categories provides a quantitative basis for railway authorities to optimize manual inspection intervals and allocate maintenance resources more efficiently. Consequently, this methodology not only reduces subjective bias in damage assessment but also offers a scalable solution for lowering the overall life-cycle costs of railway infrastructure. Future work will focus on the fusion of traffic load data and multi-sensor inputs to further enhance the predictive capabilities of the framework for remaining useful life estimation.

Acknowledgement: None.

Funding Statement: This work was supported by the National Key Research and Development Program of China [grant numbers 2022YFB2603300, 2022YFB2603303]; the Shanghai Key Laboratory of Rail Infrastructure Durability and System Safety [grant number 2024R1]; the National Natural Science Foundation of China Youth Program [grant number 52408493].

Author Contributions: The authors confirm contribution to the paper as follows: Conceptualization, Ziliang Yang and Lei Kou; methodology, Lei Kou and Mykola Sysyn; software, Jin Li; validation, Jizhe Zhang, Jian Liu and Lei Kou; formal analysis, Ziliang Yang; investigation, Jian Liu; resources, Mykola Sysyn; data curation, Jin Li; writing—original draft preparation, Ziliang Yang; writing—review and editing, Lei Kou; visualization, Jizhe Zhang; supervision, Mykola Sysyn and Lei Kou; project administration, Lei Kou; funding acquisition, Jizhe Zhang and Lei Kou. All authors reviewed and approved the final version of the manuscript.

Availability of Data and Materials: The data that support the findings of this study are available from the Corresponding Author, [KL], upon reasonable request.

Ethics Approval: Not applicable.

Conflicts of Interest: The authors declare no conflicts of interest.

References

1. Oukhellou L, Aknin P, Perrin JP. Dedicated sensor and classifier of rail head defects. Control Eng Pract. 1999;7(1):57–61. doi:10.1016/S0967-0661(98)00163-4.
2. Santur Y, Yilmazer M, Karakose M, Akin E. A new rail surface defects detection approach using 3D laser cameras based on ResNet50. Trait Du Signal. 2022;39(4):1339–45. doi:10.18280/ts.390427.
3. Scalea L, Rizzo P, Coccia S, Bartoli I, Fateh M, Viola E, et al. Non-contact ultrasonic inspection of rails and signal processing for automatic defect detection and classification. Insight. 2005;47(6):346–53. doi:10.1784/insi.47.6.346.66449.

4. Wu X, Zhang Z, Cai W, Yang N, Wang W, Liu K, et al. A study of the axle box ASD spectrum of high-speed rail vehicles based on field test measurements in China. Eng Fail Anal. 2023;154(8):107681. doi:10.1016/j.engfailanal.2023.107681.
5. Deng F, Li SQ, Zhang XR, Zhao L, Huang JB, Zhou C. An intelligence method for recognizing multiple defects in rail. Sensors. 2021;21(23):8108. doi:10.3390/s21238108.
6. Zhao Y, Hou Y, Wang X, Wei J, Shi H, Sha C, et al. A multiaxial fatigue method for rolling contact fatigue life prediction of axle box bearing under wheel-rail excitation conditions. Eng Fail Anal. 2024;159(9):108086. doi:10.1016/j.engfailanal.2024.108086.
7. Zhang R, Zheng C, Lv B, Wang X, Li X, Li Y, et al. Research progress on rolling contact fatigue damage of bainitic rail steel. Eng Fail Anal. 2023;143(2021):106875. doi:10.1016/j.engfailanal.2022.106875.
8. Gibert X, Patel VM, Chellappa R. Deep multitask learning for railway track inspection. IEEE Trans Intell Transp Syst. 2017;18(1):153–64. doi:10.1109/TITS.2016.2568758.
9. Wu F, Li Q, Li S, Wu T. Train rail defect classification detection and its parameters learning method. Measurement. 2020;151(2):107246. doi:10.1016/j.measurement.2019.107246.
10. Xing Z, Zhang Z, Yao X, Qin Y, Jia L. Rail wheel tread defect detection using improved YOLOv3. Measurement. 2022;203(8):111959. doi:10.1016/j.measurement.2022.111959.
11. An B, Sun Y, Liu J, Tao G, Qian Y, Wang P. The role of 3D contact geometry in modeling dynamic wheel-rail interaction at short-wave irregularities on rail surface. Eng Fail Anal. 2023;153(1):107559. doi:10.1016/j.engfailanal.2023.107559.
12. Acikgoz H, Korkmaz D. MSRConvNet: classification of railway track defects using multi-scale residual convolutional neural network. Eng Appl Artif Intell. 2023;121:105965. doi:10.1016/j.engappai.2023.105965.
13. Jiang Y, Wang H, Tian G, Yi Q, Zhao J, Zhen K. Fast classification for rail defect depths using a hybrid intelligent method. Optik. 2019;180(4):455–68. doi:10.1016/j.ijleo.2018.11.053.
14. Aydin I, Akin E, Karakose M. Defect classification based on deep features for railway tracks in sustainable transportation. Appl Soft Comput. 2021;111(7):107706. doi:10.1016/j.asoc.2021.107706.
15. Yang J, Zhou W, Wu R, Fang M. CSANet: contour and semantic feature alignment fusion network for rail surface defect detection. IEEE Signal Process Lett. 2023;30:972–6. doi:10.1109/LSP.2023.3299218.
16. Ye J, Stewart E, Chen Q, Roberts C, Hajiyavand AM, Lei Y. Deep learning and laser-based 3-D pixel-level rail surface defect detection method. IEEE Trans Instrum Meas. 2023;72:2513612. doi:10.1109/TIM.2023.3272033.
17. Wang Q, Wang X, He Q, Huang J, Huang H, Wang P, et al. 3D tensor-based point cloud and image fusion for robust detection and measurement of rail surface defects. Autom Constr. 2024;161:105342. doi:10.1016/j.autcon.2024.105342.
18. Guo F, Liu J, Qian Y, Xie Q. Rail surface defect detection using a transformer-based network. J Ind Inf Integr. 2024;38(6):100584. doi:10.1016/j.jii.2024.100584.
19. Xie Q, Tao G, Lo SM, Yang X, Wen Z. A data-driven convolutional regression scheme for on-board and quantitative detection of rail corrugation roughness. Wear. 2023;524:204770. doi:10.1016/j.wear.2023.204770.
20. Juboori A, Zhu H, Li H, McLeod J, Pannila S, Barnes J. Microstructural investigation on a rail fracture failure associated with squat defects. Eng Fail Anal. 2023;151(3):107411. doi:10.1016/j.engfailanal.2023.107411.
21. Zhang SY, Zhao XJ, Ding HH, Spiryagin M, Guo J, Liu QY, et al. Effects of dent size on the evolution process of rolling contact fatigue damage on defective rail. Wear. 2021;477:203894. doi:10.1016/j.wear.2021.203894.
22. Nguyen BH, Al-Juboori A, Zhu H, Zhu Q, Li H, Tieu K. Formation mechanism and evolution of white etching layers on different rail grades. Int J Fatigue. 2022;163:107100. doi:10.1016/j.ijfatigue.2022.107100.
23. Kou L, Sysyn M, Liu J, Fischer S, Nabochenko O, He W. Prediction system of rolling contact fatigue on crossing nose based on support vector regression. Measurement. 2023;210(4):112579. doi:10.1016/j.measurement.2023.112579.
24. Kou L, Sysyn M, Liu J, Nabochenko O, Han Y, Peng D, et al. Evolution of rail contact fatigue on crossing nose rail based on long short-term memory. Sustainability. 2022;14(24):16565. doi:10.3390/su142416565.

25. Panda MK, Sharma A, Bajpai V, Subudhi BN, Thangaraj V, Jakhetiya V. Encoder and decoder network with ResNet-50 and global average feature pooling for local change detection. Comput Vis Image Underst. 2022;222(28):103501. doi:10.1016/j.cviu.2022.103501.
26. Zhang M, Zheng H, Gong M, Wu Y, Li H, Jiang X. Self-structured pyramid network with parallel spatial-channel attention for change detection in VHR remote sensed imagery. Pattern Recognit. 2023;138(1–2):109354. doi:10.1016/j.patcog.2023.109354.
27. Li C, Yan H, Qian X, Zhu S, Zhu P, Liao C, et al. A domain adaptation YOLOv5 model for industrial defect inspection. Measurement. 2023;213(16):112725. doi:10.1016/j.measurement.2023.112725.
28. Vaswani A, Shazeer N, Parmar N, Uszkoreit J, Jones L, Gomez AN, et al. Attention is all you need. In: Proceedings of the 31st International Conference on Neural Information Processing Systems; 2017 Dec 4–9; Long Beach, CA, USA. p. 1–11.
29. Liu Z, Lin Y, Cao Y, Hu H, Wei Y, Zhang Z, et al. Swin transformer: hierarchical vision transformer using shifted windows. In: Proceedings of the 2021 IEEE/CVF International Conference on Computer Vision (ICCV); 2021 Oct 10–17; Montreal, QC, Canada. p. 10012–22.
30. Guo C, Fan B, Zhang Q, Xiang S, Pan C. AugFPN: improving multi-scale feature learning for object detection. In: Proceedings of the 2020 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR); 2020 Jun 13–19; Online. p. 12595–604.