Sensors and Materials, Vol. 38, No. 2 (2026) 857–870
MYU Tokyo

S & M 4352

Deep-learning-based Ultrasonic Guided Wave Detection of Turnout Switch Rail Cracks

ChengYue Lv,[1,2*] Ping Wang,[1] and Hao Sun[2]

[1]School of Automation, Nanjing University of Aeronautics and Astronautics,
29 Jiangjun Avenue, Jiangning District, Nanjing 211106, China
[2]The Fourteenth Institute, China Electronics Technology Group Corporation (CETC)
268 Guorui Road, Yuhuatai District, Nanjing 210000, China

(Received September 5, 2025; accepted November 4, 2025)

Keywords: ultrasonic guided waves, cracks in turnout switch rails, wavelet transform, deep learning

The turnout switch rail is a type of variable-cross-section rail, and its irregular structural characteristics result in complex ultrasonic guided wave detection signals. When employing the reflection method to detect cracks in the rail base, the amplitude of the echo signal cannot represent the size of the crack. To quantitatively analyze the crack signals, a method that combines deep learning and ultrasonic guided wave technology is employed to quantitatively assess the depth of cracks in the rail base of the turnout switch rail. By applying wavelet transform to obtain wavelet time–frequency diagrams, four deep learning models—GoogLeNet, Mobilenetv1, Mobilenetv2, and Mobilenetv3—are utilized to classify the depth of cracks in the rail base, and the performance of these models is assessed using experimental data. The experimental results show that the combination of the Mobilenetv3 deep learning model and ultrasonic guided wave technology achieves a 95% recognition accuracy for the quantitative detection of cracks in the rail base of turnout switch rails. This research work provides a foundation for the feasibility and reliability of combining deep learning models with ultrasonic guided wave technology for the quantitative detection of crack depths in turnout switch rails.

1. Introduction

The turnout switch rail constitutes a critical element within railway track transition systems. During operational service, it experiences environmental temperature variations and train loads, which may frequently result in surface and internal defects. Among these, defects in the rail base are predominant and remain challenging to detect using conventional ultrasonic testing methods.

To prevent train derailment incidents, nondestructive testing methods are essential for assessing railway track safety conditions. Common rail damage detection methodologies include eddy current testing,[1,2] magnetic flux leakage testing,[3] machine vision,[4,5] radiographic testing,[6] laser ultrasonic testing,[7,8] and ultrasonic testing.[9] These techniques typically require extended inspection periods and may disrupt regular train operations.

*Corresponding author: e-mail: sj18269770464@126.com
https://doi.org/10.18494/SAM5934

ISSN 0914-4935
https://myukk.org/

Ultrasonic guided wave technology[10,11] offers an extensive detection range with minimal attenuation, enabling long-range, large-area rail inspection. Although possessing a century-long history, its application to rail inspection has only emerged within the last two decades. Research in this domain has yielded significant advancements: de Abreu[12] and Mariani and co-workers[13,14] designed air-coupled ultrasonic transducers to excite rail head guided waves and capture corresponding echo signals, employing data analysis techniques for feature extraction; di Scalea and Xuan[15,16] developed a methodology utilizing wheel–rail noise signals, applying deconvolution to eliminate stochastic effects and reconstruct stable track acoustic transfer functions for health monitoring, enabling defect detection in rail joints and welds; Lee *et al.*[17] established a nonlinear guided wave approach for steel fatigue crack identification using second-harmonic-based nonlinear indices, demonstrating capability in detecting crack initiation and propagation at rail joints; Sadeghi and Rahimizadeh[18] formulated a novel rail condition index incorporating both surface/visual defects and all documented internal defects, utilizing data from visual inspections and automated ultrasonic measurements.

Given rail structural complexity, ultrasonic guided wave signals exhibit dispersion and mode conversion during propagation, necessitating signal processing to isolate specific modes. Established processing techniques include time-domain analysis, frequency-domain analysis, and wavelet transforms. The integration of deep learning[19] and machine learning algorithms[20] with ultrasonic guided wave technology represents an emerging research focus. Deng *et al.*[21] implemented segmented Principal Components Analysis (PCA) to extract features from ultrasonic guided waves received at multiple rail head positions, achieving classification and identification of rail head defects.

Compared with traditional machine learning, deep learning algorithms in image classification and rail defect detection bypass the need for manual feature extraction, instead autonomously learning features through networks. This substantially enhances algorithmic versatility and facilitates high-precision, efficient online rail damage assessment. We propose integrating deep learning models with ultrasonic guided wave technology to identify crack depths in turnout switch rail bases.

2. Detection Principle and Dataset Acquisition

2.1 Detection principle

Turnout switch rails are characterized by a continuously varying cross section, which causes complex reflections, scattering, and mode conversions in propagating ultrasonic guided waves. These phenomena result in significant background clutter interference, thereby complicating the detection of defects. To overcome these challenges and effectively detect damage in critical areas, particularly the variable-cross-section working edges, the reflection wave method is employed. In this method, ultrasonic guided wave sensors are installed at the root end of the switch rail.

In the reflection wave method, ultrasonic sensors are placed on the same side of the rail to detect defects by capturing the echoes they generate, as illustrated in Fig. 1. When ultrasonic

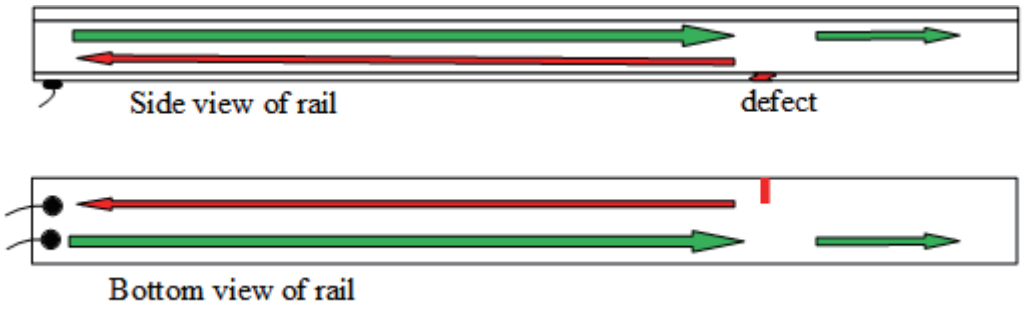


Fig. 1. (Color online) Schematic diagram of the reflection wave method detection principle.

guided waves encounter an abrupt change in the rail medium, such as a defect, an echo is produced. The extent of the damage can be determined by analyzing the energy of the defect echo. Furthermore, the location of the defect is determined from the time delay of the echo and the known propagation speed of the guided wave. Considering the limited length of the switch rail and the minimal requirement for precise crack localization, the primary focus of this work is the quantitative analysis of cracks in turnout switch rails.

2.2 Dataset acquisition

Figure 2 shows a schematic of the experimental system for detecting rail damage using ultrasonic guided waves. The key components include a signal generator, power amplifier, transmitting sensor, rail, receiving sensor, signal conditioner, digital oscilloscope, and computer.

The system operates as follows. The signal generator produces a Hanning-modulated, five-cycle sinusoidal excitation signal at a center frequency of 65 kHz. This signal is amplified by the power amplifier and sent to the transmitting sensor. The sensors employ piezoelectric ceramics as the energy conversion device, offering high sensitivity and efficient transduction for precise wave detection. They are mounted at a 45° incident angle on the rail bottom using metal fixtures with Vaseline as the coupling agent. The transmitting sensor generates ultrasonic guided waves that propagate along the rail. The receiving sensor detects these waves, which are then conditioned, acquired by the oscilloscope, and transferred to the computer for analysis.

In the experiment designed to detect cracks in the rail base of a turnout switch rail using ultrasonic guided waves, the switch rail was configured with a length of 8 m. Cracks were artificially introduced on the nonmachined side of the rail base, positioned at a distance of 2.3 m from the tip of the switch rail and 5.7 m from the excitation position. These cracks were configured as 90° through-cracks with depths of 0 mm (representing the undamaged condition) and from 1 to 12 mm in steps of 1 mm, with a uniform width of 1.0 mm. For each crack depth, 200 sets of signal data were acquired. The process of damage is depicted in Fig. 3.

To simulate noise effects present under actual operating conditions, data augmentation was employed by incorporating Gaussian white noise with variances of 0.1, 0.2, and 0.3 into the original dataset. This process expanded the dataset to a total of 2,600 signal sets, encompassing both undamaged and damaged conditions. Subsequently, the augmented dataset was randomly

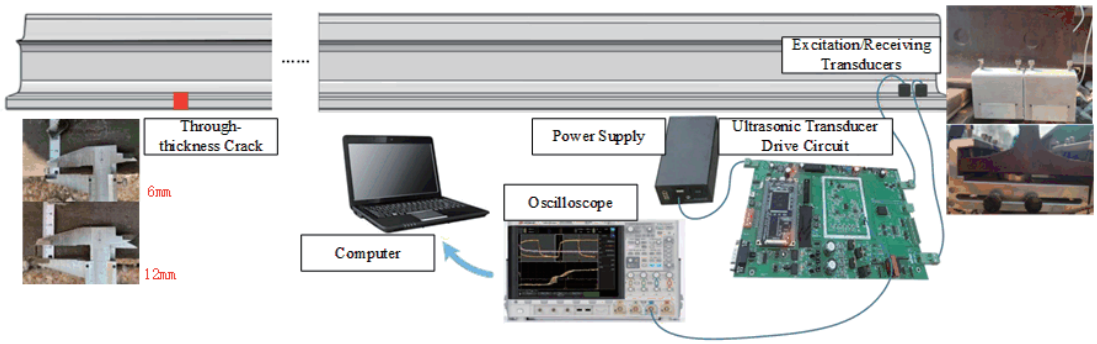


Fig. 2. (Color online) Schematic diagram of the turnout switch rail base crack detection system based on ultrasonic guided waves.

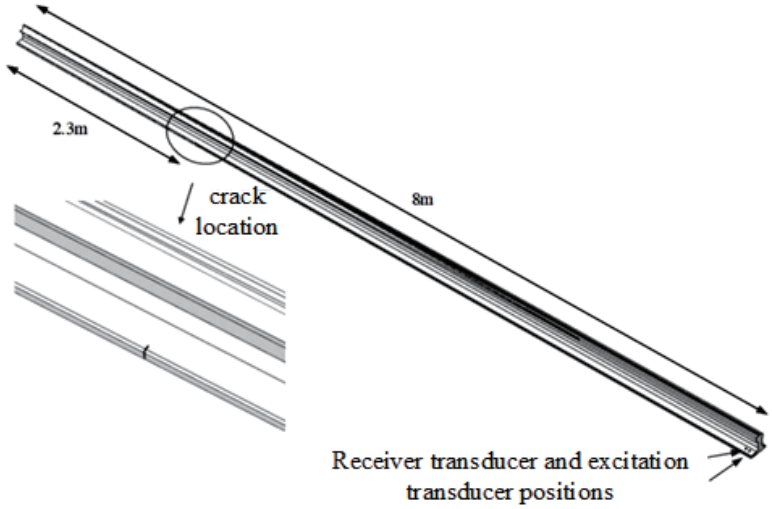


Fig. 3. Schematic diagram of crack processing in the rail base of the turnout switch rail.

divided into training and test sets in an 8:2 ratio, comprising 2080 and 520 signal sets, respectively. Furthermore, to validate the model's accuracy, artificial damage was induced in another switch rail under identical conditions. Data were collected for 13 conditions, encompassing 12 damage types and the undamaged state, with 50 sets per type, totaling 650 sets, which served as experimental validation sets.

3. Signal Processing Methods

3.1 Wavelet transform algorithm

In comparison with the short-time Fourier transform (STFT), the wavelet transform algorithm transforms time-domain waveforms into time–frequency domain representations, offering enhanced capabilities for time–frequency analysis. Consequently, ultrasonic rail signals are converted into time–frequency domain images via the wavelet transform algorithm to facilitate the extraction of crack-related features in the rail base.

The selection of the wavelet basis function, denoted as $X(t)$, is paramount in the wavelet transform algorithm. For continuous ultrasonic rail signals $x(t)$, the continuous wavelet transform is mathematically expressed as

$$CWT_x(\alpha,\beta) = \int_{-\alpha}^{+\infty} x(t) X^*_{(\alpha,\beta)}(t) dt \,, \tag{1}$$

where a represents the scale factor ($a > 0$), and β denotes the time-shift factor. $X^*(t)$ is the complex conjugate of the wavelet basis function $X(t)$. The choice of an appropriate wavelet basis function is crucial when applying the continuous wavelet transform to ultrasonic rail signals, as it directly impacts the effectiveness of extracting crack features from the rail base.

Among the commonly used wavelet basis functions—such as Morlet, Coiflet (Coif), Daubechies 4 (Db4), and Meyer—we select the Coiflet wavelet basis function to transform ultrasonic rail signals into the time–frequency domain. This selection is based on its suitability for capturing the transient features of ultrasonic signals in rail damage detection.

3.2 GoogLeNet learning model

The GoogLeNet network model,[22] which achieved victory in the 2014 ImageNet competition, incorporates the Inception module. This module improves the classification accuracy of the network model while decreasing the number of parameters, as illustrated in Fig. 4. GoogLeNet enhances the Inception Model A to develop the improved Inception Model B [depicted in Fig. 4(b)]. Inception Model A primarily employs dense components to approximate

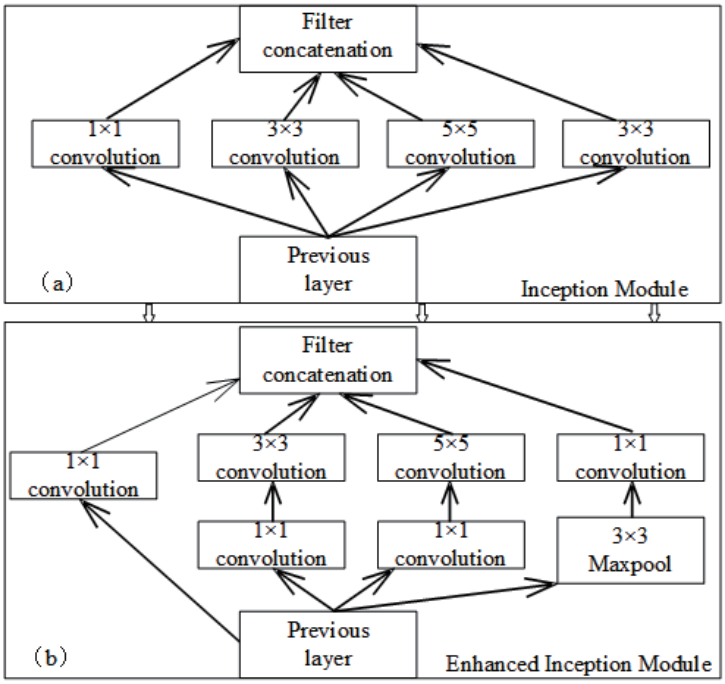


Fig. 4. Inception model and improved inception model

the optimal local sparse structure. In contrast, the enhanced Inception Model B significantly improves the feature extraction capability of convolutions without increasing the computational load of the model. Research results suggest that networks utilizing the Inception module can experience a 2–3-fold increase in operational speed.

3.3 MobileNetv1 learning model

In 2017, Google developed the lightweight MobileNetv1 model to facilitate the real-time application of deep learning network models on mobile devices.[23] The MobileNetv1 model decreases computational complexity while preserving accuracy and detection speed. It utilizes a streamlined architecture, employing depthwise separable convolutions and incorporating hyperparameters α and β to improve model performance. For a 128 × 128 pixel, three-channel color image (dimensions: 128 × 128 × 3), conventional convolutional neural networks generally necessitate four convolutional layers with filters, generating four feature maps, as depicted in Fig. 5.

In contrast, depthwise convolution applies a separate filter to each input channel, processing each channel independently. For an input image of dimensions 128 × 128 × 3, depthwise convolution involves a two-dimensional convolution on each channel, producing a set of feature maps corresponding to each input channel, as illustrated in Fig. 6(a). Subsequently, pointwise

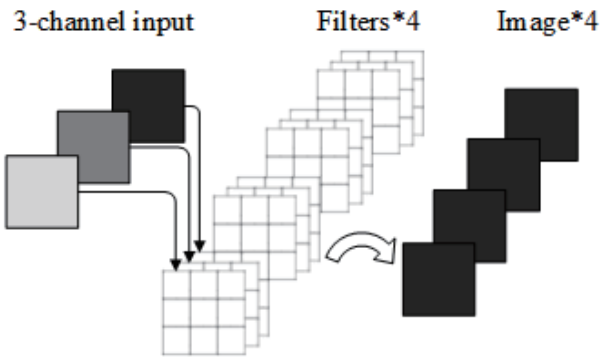


Fig. 5. Schematic diagram of conventional convolution

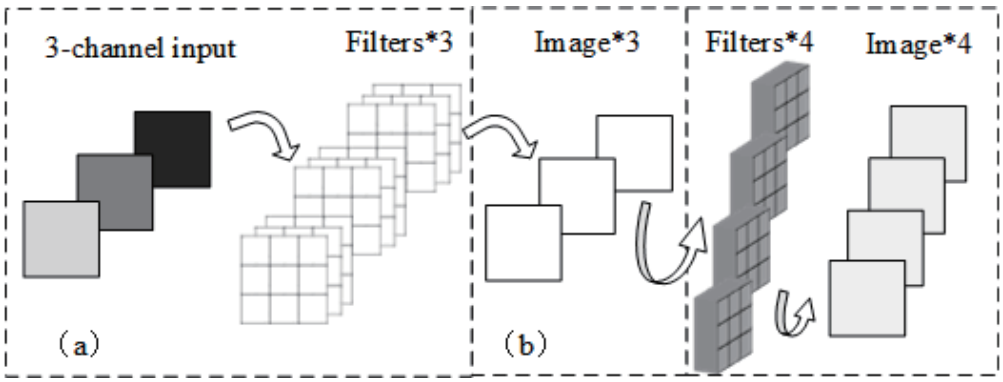


Fig. 6. Schematic diagrams of (a) depthwise convolution and (b) pointwise convolution.

convolution, utilizing 1 × 1 filters, combines these feature maps to generate the final output channels, a process akin to conventional convolution but with reduced complexity.

Depthwise separable convolution integrates these two steps: depthwise convolution to filter each input channel separately, followed by pointwise convolution to merge the resulting feature maps. The computational efficiency of this approach is demonstrated through the following calculations for a scenario with a 3 × 3 filter and four output channels.

Depthwise separable convolution consists of two steps: depthwise convolution and pointwise convolution, as illustrated in Fig. 6. The computational loads can be calculated as follows.

Computational load for conventional convolution: $C = 4 \times 3 \times 3 \times 3 = 108$

Computational load for depthwise convolution: $D = 3 \times 3 \times 3 = 27$

Computational load for pointwise convolution: $P = 1 \times 1 \times 3 \times 4 = 12$

Total computational load for depthwise separable convolution: $S = D + P = 39$

Thus, depthwise separable convolution reduces the computational load to approximately one-third of that required by conventional convolution This significant reduction in computational complexity enables neural networks employing depthwise separable convolution to incorporate deeper architectures.

3.4 MobileNetv2 learning model

The MobileNetV2 model,[24] an extension of the MobileNetV1 architecture, is designed to optimize performance for mobile and embedded vision applications. Traditional residual structures typically employ a compress-then-expand approach, reducing the number of channels in the feature map via convolutional layers and subsequently expanding them, as depicted in Fig. 7(a). In contrast, MobileNetV2's inverted residual structure utilizes an expand-then-compress strategy, characterized by narrower input and output dimensions with a wider intermediate dimension, as illustrated in Fig. 7(b).

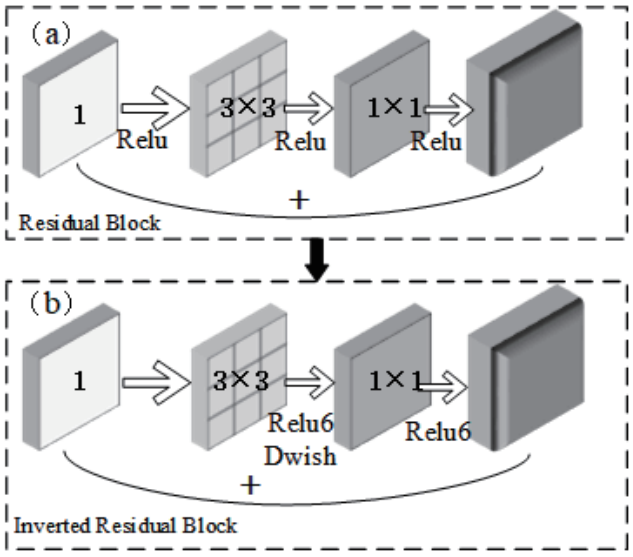


Fig. 7. (a) Residual module. (b) Inverted residual module.

MobileNetV2, similar to MobileNetV1, employs depthwise separable convolutions, which consist of depthwise and pointwise convolution operations. Furthermore, MobileNetV2 incorporates the ReLU6 activation function within its inverted residual blocks, as defined below.

$$Relu6 = \min(\max(0, x)) \tag{2}$$

3.5 MobileNetv3 learning model

In comparison with the block structure of MobileNetV2, as illustrated in Fig. 8(a), the block structure of MobileNetV3,[25] depicted in Fig. 6(b), incorporates a squeeze-and-excitation (SE) module and employs modified activation functions. The activation functions utilized in MobileNetV3 primarily consist of Hard-Swish and ReLU. Specifically, the Hard-Swish activation function serves as an optimized variant of the Swish activation function. Although the Swish activation function demonstrates superior performance relative to ReLU, it is characterized by increased computational complexity and greater resource demands. To address these challenges, MobileNetV3 employs the Hard-Swish activation function, which leverages ReLU6 to approximate Swish effectively, thereby mitigating precision loss during computation. This approach not only reduces computational complexity but also enhances inference speed and supports the quantization process of the MobileNetV3 model. The Hard-Swish activation function is expressed as

$$Swish = x * \sigma(x) \tag{3}$$

$$H - Swish = x * \frac{Relu6(x+3)}{6}. \tag{4}$$

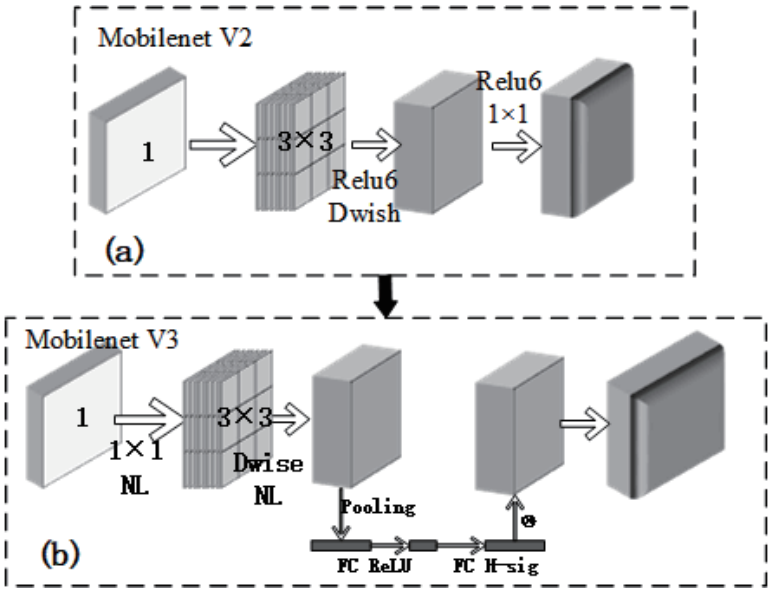
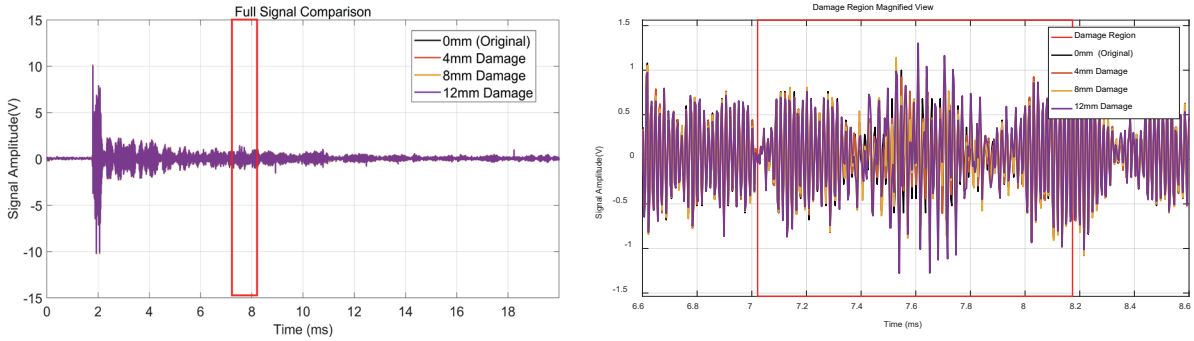


Fig. 8. Schematic diagrams of (a) MobileNet V2 module and (b) MobileNet V3 module.

4. Data Processing and Results

As the crack depth in the rail base (0, 4, 8, 12 mm) changes, the signal waveforms of the received ultrasonic guided waves are as presented in Fig. 9. With increasing crack depth in the rail base of the switch rail from 0 to 12 mm, the temporal variations in the ultrasonic signal remain minimal. This phenomenon is primarily attributed to the dispersion of various wave modes during the propagation of ultrasonic guided waves in the rail, accompanied by mode conversions. These factors pose significant challenges to detecting cracks in the rail base employing ultrasonic guided wave techniques.

Figure 10 shows time-frequency diagrams generated from the application of wavelet transform to ultrasonic signals corresponding to four distinct crack depths (0, 4, 8, 12 mm) in the



(a) (b)

Fig. 9. (Color online) (a) Overall signal time-domain diagram and (b) local time-domain diagram of damage echo.

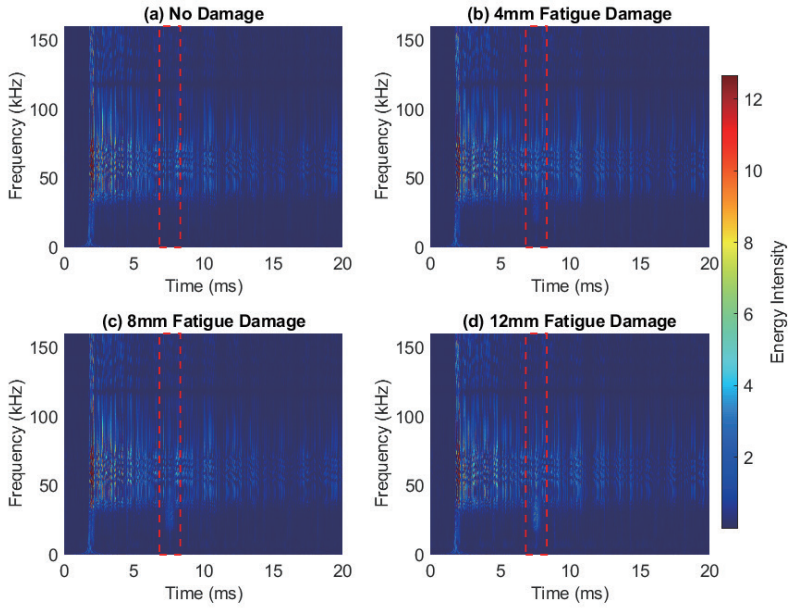


Fig. 10. (Color online) Wavelet time–frequency diagram (Coif).

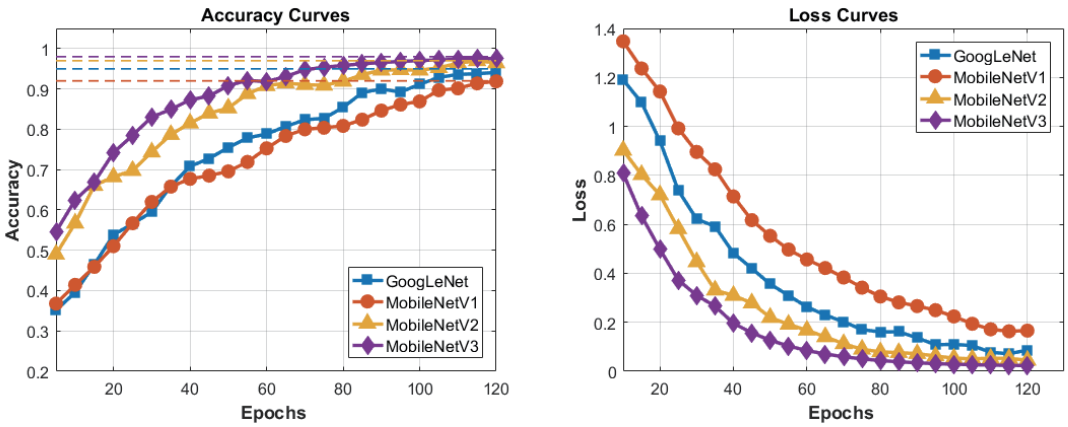
rail base of a switch rail. As depicted in Fig. 10, the energy distribution in the time–frequency spectrum of ultrasonic guided wave signals for various crack depths is predominantly concentrated around 65 kHz, with variations observed in the energy distribution across different time points. Despite discernible differences in the time–frequency spectrum energy distribution across crack depths, the precise quantification of crack depth remains challenging. Consequently, in this study, we utilize wavelet time–frequency diagrams as input data for deep learning models to facilitate the accurate identification of crack depths.

We employ four deep learning models—GoogLeNet, MobileNetv1, MobileNetv2, and MobileNetv3—to train on a simulated dataset for detecting cracks in the rail base of turnout switch rails.

Figure 11(a) shows the accuracy curves. The accuracy of all models rises with the number of training iterations. MobileNetv1 improves the slowest, reaching a maximum of 0.92 (Fig. 11, Table 1). MobileNetv2 increases faster and outperforms GoogLeNet. MobileNetv3 converges the fastest, achieving 0.975 (Fig. 11, Table 1).

Figure 11(b) presents the loss curves. The losses of all models decrease with the number of iterations. The loss of MobileNetv1 declines the slowest, reaching ~0.18. That of MobileNetv2 decreases faster and is less than that of GoogLeNet. The loss of MobileNetv3 decreases the fastest, reaching 0.02.

As evidenced by Fig. 12, the MobileNetv3 model utilized in this investigation demonstrates superior stability and accuracy in identifying the depth of crack damage in the rail base of



(a) (b)

Fig. 11. (Color online) (a) Accuracy curves and (b) loss rate curfves of deep learning models GoogLeNet, MobileNetv1, MobileNetv2, and MobileNetv3.

Table 1
Performance of four deep learning models.

Algorithm performance	Accuracy	Precision	Recall	F1 Score
GoogLeNet	0.953	0.967	0.957	0.965
MobilenetV1	0.920	0.924	0.917	0.920
MobilenetV2	0.971	0.972	0.976	0.977
MobilenetV3	0.975	0.985	0.977	0.984

turnout switch rails. The overall performance of the models decreases in the sequence of MobileNetv3, MobileNetv2, GoogLeNet, and MobileNetv1. Consequently, the MobileNetv3 deep learning model exhibits the highest recognition performance among the evaluated models.

5. Experimental Validation

Wavelet time–frequency diagrams derived from artificial damage data, collected under identical conditions from another switch rail on the same line, served as input for the deep learning models to identify cracks in the rail base. The final results, generated by the four deep learning models—GoogLeNet, MobileNetv1, MobileNetv2, and MobileNetv3—are presented in Fig. 13.

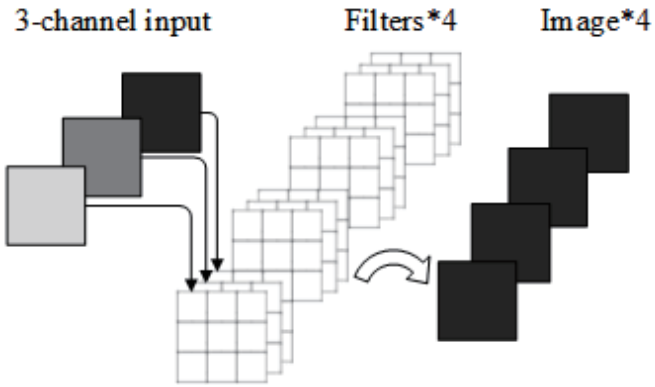


Fig. 12. (Color online) Performance metrics (accuracy, precision, recall, and F1 score) of GoogLeNet, MobileNetv1, MobileNetv2, and MobileNetv3 models.

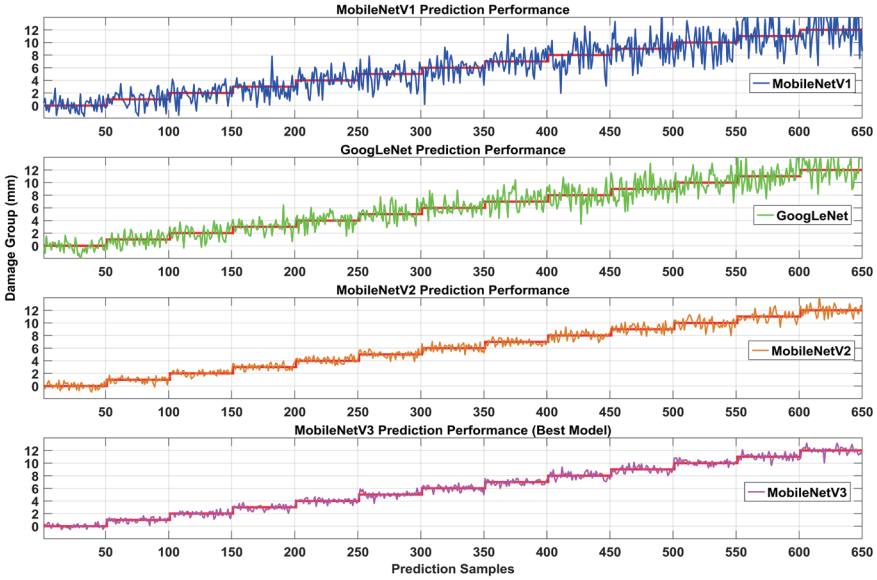


Fig. 13. (Color online) Output results of four deep learning models.

Table 2
Model prediction accuracy and performance.

Algorithm performance	Accuracy (±0.3 mm)	*RMSE*
GoogLeNet	0.798	0.890
MobilenetV1	0.709	0.983
MobilenetV2	0.906	0.453
MobilenetV3	0.957	0.310

The experimental dataset comprises 50 sets of data for each damage type, totaling 650 sets of damage data. In terms of stability and accuracy in crack identification, the MobileNetv3 model demonstrates superior performance relative to the other models. Employing an expected prediction interval of ±0.3 mm of actual damage, the MobileNetv3 model achieves the highest prediction accuracy as measured by the root mean square error (*RMSE*). Table 2 lists the prediction accuracies of the four models, with their performances evaluated using the *RMSE* metric.

6. Conclusions

A method integrating the MobileNetv3 deep learning model with ultrasonic guided waves was proposed for the quantitative assessment of crack depth in the rail base of turnout switch rails. Initially, an ultrasonic guided wave detection system was established to acquire ultrasonic guided wave signals from the rail base. These signals were subsequently processed via wavelet transform to generate time–frequency diagrams. To enhance the generalization capability of the deep learning model and ensure robustness under real-world noise conditions, we employed data augmentation by incorporating Gaussian white noise with variances of 0.1, 0.2, and 0.3 into the original dataset. The augmented image samples were then randomly divided into training and test sets in an 8:2 ratio. Finally, the performance of four deep learning models—GoogLeNet, MobileNetv1, MobileNetv2, and MobileNetv3—was evaluated using experimental data from cracks in another switch rail of the same turnout. The results indicated that the proposed method, which combines the MobileNetv3 deep learning model with ultrasonic guided waves, demonstrates superior performance in the quantitative assessment of crack depth in the rail base. This approach facilitates the precise quantitative evaluation of crack depth in the rail base of turnout switch rails.

In future work, we plan to extend the validation of our method by incorporating field-collected data from in-service turnout switch rails exhibiting natural cracks. These cracks, formed under long-term operational conditions involving temperature variations and dynamic train loads, may present more irregular morphologies. By augmenting the dataset with such real-world samples, we aim to further assess and enhance the model's robustness and practical relevance for railway safety monitoring.

References

1 R. Yang, Y. He, and B. Gao: Measurement **66** (2015) 54. https://doi.org/10.1016/j.measurement.2015.01.024
2 Y. Gao and G. Y. Tian: Sens. Actuators, A **270** (2017) 8. https://doi.org/10.1016/j.sna.2017.12.027
3 V. Nichoha, V. Shkliarskyi, and V. Storozh: Problemy Kolejnictwa-Railway Reports **180** (2018) 105.
4 K. Zhao, L. Luo, and Z. Ren: J. Phys. Conf. Ser. **1678** (2020) 012002. https://doi.org/10.1088/1742-6596/1678/1/012002
5 C. Xie and Z. Xie: J. Chongqing University **36** (2013) 16.
6 A. McNulty and J. McNab: Insight Non-Destr. Test. and Condition Monit. **44** (2002) 348.
7 Y. Jiang, H. Wang, and S. Chen: Optik **237** (2021) 166732. https://doi.org/10.1016/j.ijleo.2021.166732
8 M. Pathak, S. Alahakoon, and M. Spiryagin: Meas. **148** (2019) 106922. https://doi.org/10.1016/j.measurement.2019.106922
9 A. Spada, M. Capriotti, and F. L. di Scalea: Struct. Health Monit. **21** (2022) 370. https://doi.org/10.1177/14759217211000863
10 P. Hu, H. Wang, and G. Tian: IEEE Sens. J. **19** (2019) 11050. https://doi.org/10.1109/JSEN.2019.2934159
11 X. Xu, Z. Lu, and B. Xu: IEEE Access **6** (2018) 60414. https://doi.org/10.1109/ACCESS.2018.2875123
12 L. de Abreu Correa, R. Cottereau, and B. Faure: J. Sound Vib. **464** (2020) 115010. https://doi.org/10.1016/j.jsv.2019.115010
13 S. Mariani, T. Nguyen, X. Zhu, and F. L. di Scalea: Struct. Health Monit. **17** (2018) 684. https://doi.org/10.1177/1475921717715429
14 S. Mariani, T. Nguyen, and X. Zhu: J. Transp. Eng. Part A. Syst. **143** (2017) 04017007. https://doi.org/10.1061/JTEPBS.0000026
15 Z. Xuan, S. Simone, and M. Capriotti: Proc. 11th Int. Workshop on Structural Health Monitoring Systems (IWSHM, 2017) 2842–2846. https://doi.org/10.12783/shm2017/14188
16 F. L. di Scalea and Z. Xuan: Proc. 12th Health Monitoring of Structural and Biological Systems (SPIE, 2018) 106000J. https://doi.org/10.1117/12.2295634
17 Y. F. Lee, Y. Lu, and R. Guan: Smart Mater. Struct. **29** (2020) 694. https://doi.org/10.1088/1361-665X/ab6fe7
18 J. Sadeghi and Y. Rahimizadeh: J. Transp. Eng. Part A. Syst. **146** (2020) 04020078. https://doi.org/10.1061/JTEPBS.0000390
19 H. Mahajan and S. Banerjee: Mach. Learn. Appl. **7** (2021) 100216. https://doi.org/10.1016/j.mlwa.2021.100216
20 X. Wang, M. Lin, and J. Li: Mech. Syst. and Signal Process **169** (2022) 108761. https://doi.org/10.1016/j.ymssp.2021.108761
21 F. Deng, S. Q. Li, and X. R. Zhang: Sensors **21** (2021) 8108. https://doi.org/10.3390/s21238108
22 C. Szegedy, W. Liu, and Y. Jia: Proc. IEEE Conf. Computer Vision and Pattern Recognition (IEEE, 2015) 1–9. https://doi.org/10.1109/CVPR.2015.7298594
23 M. Abadi, A. Agarwal, and P. Barham: TensorFlow: Large-scale machine learning on heterogeneous systems (2015) http://tensorflow.org (accessed October 2025).
24 O. Russakovsky, J. Deng, and H. Su: Int. J. Comput. Vision **115** (2015) 211. https://doi.org/10.1007/s11263-015-0816-y
25 A. Howard, M. Sandler, and B. Chen: Proc. IEEE/CVF Int. Conf. Computer Vision (ICCV) (IEEE, 2019). https://doi.org/10.48550/arXiv.1905.02244

About the Authors

Chengyue Lv is currently a Ph.D. candidate at Nanjing University of Aeronautics and Astronautics. Since 2007, he has been working at the Nanjing Research Institute of Electronics Technology. He professional experience primarily encompasses data processing and research project planning and management. (sj18269770464@126.com)

Ping Wang received his Ph.D. degree from Southeast University, Nanjing China, in 2004. He is currently a professor with the Nanjing University of Aeronautics and Astronautics, Nanjing. His research interests include electromagnetic nondestructive testing and advanced signal processing. (Zeitping@nuaa.edu.cn)

Hao Sun received his master's degree in power engineering from Southeast University. Since 2021, he has been working at the Nanjing Research Institute of Electronics Technology. His professional experience there includes research management and market management responsibilities.