Commun Transp Res

Research Article

Ultrasonic denoising for intelligent operation and maintenance of heavy-haul railways: Noise mechanisms and suppression methods

Jiangtao Zhang[1,2,3], Yuan Cao[1,2,✉], Yuntong An[1,2], Feng Wang[1,2], Yongkui Sun[1,2], Shuai Su[1,2]

Cite this article: https://doi.org/10.26599/COMMTR.2026.9640021

ABSTRACT: Heavy-haul railways are critical for transporting freight. However, prolonged wheel–rail interactions cause frequent rail defects, particularly in small-radius curve sections. Ultrasonic A-scan signals are essential for the non-destructive evaluation of internal rail defects. In real heavy-haul environments, these signals suffer from strong non-Gaussian coupled noise. Such noise includes structural noise, low-frequency irrelevant components, and high-frequency electrical noise. Noise aliasing obscures defect echoes and increases the risk of missed detections. Conventional denoising methods are limited by poor noise–signal separability, mode mixing, and inadequate adaptability to complex non-Gaussian signals. To address these challenges, an A-scan signal model under noise-coupled conditions is constructed by analyzing the statistical and time–frequency characteristics of different noise components. Based on this model, a multi-feature fusion filtering framework is developed within the ideal binary mask (IBM) paradigm. This framework is designed to enhance defect echo extraction from ultrasonic A-scan signals under strong non-Gaussian interference. Tests on field inspection data show that the proposed method effectively suppresses coupled noise and achieves accurate extraction of defect echoes.

Keyword: Heavy-haul railway; ultrasonic A-scan signal; Multi-feature fusion; mask filtering

1 Introduction

Heavy-haul railways are a key part of the railway system, transporting bulk commodities such as coal and ore and playing a crucial role in regional economic development. Its high-density, long-formation transportation organization can effectively reduce costs (Shi et al., 2024; Cao et al., 2025). However, with the continuous increase in traffic volume, heavy-haul lines are subjected to long-term wheel-rail interaction forces, leading to frequent occurrence and accelerated deterioration of rail defects (Xing et al., 2020). Statistical data indicate that, over the past three years, a certain heavy-haul railway line in China has mainly experienced severe rail defects such as head checks and bolt hole cracks (Zhang et al., 2025). Head checks and bolt hole cracks are classified as severe rail defects that must be removed once detected. If not treated in time, they can quickly develop into rail breaks and severely threaten train operation safety.

To ensure transportation safety under high-intensity service conditions, intelligent operation and maintenance has become an essential approach for heavy-haul railways. In practical engineering applications, maintenance decision-making strongly depends on the accuracy of defect detection data obtained from routine inspections. Ultrasonic non-destructive testing (NDT) is widely used as a primary data source for defect identification in heavy-haul railway maintenance (Mićić et al., 2023; Chen et al., 2022). The extracted defect information directly affects defect assessment, maintenance scheduling, and risk control. Therefore, improving the accuracy of ultrasonic defect detection under complex service environments is a critical prerequisite for effective intelligent maintenance of heavy-haul railways.

In current practice, the electrical output of the ultrasonic piezoelectric transducer is digitized to form an A-scan signal. This signal contains rich but abstract features and therefore relies heavily on operator experience for interpretation. Converting A-scan data into B-scan images visualizes defect regions, morphologies and channel information and improves defect recognition efficiency (Yu et al., 2024). Ultrasonic rail inspection mainly relies on hand-pushed detectors, dual-rail inspection trolleys, and rail inspection vehicles, among which inspection vehicles are the primary tools for full-line scanning on heavy-haul

[1] School of Automation and Intelligence, Beijing Jiaotong University, Beijing 100044, China. [2] National Engineering Research Center of Rail Transportation Operation and Control System, Beijing Jiaotong University, Beijing 100044, China. [3] School of Automation and Electrical Engineering, Lanzhou Jiaotong University, Lanzhou 730000, China.

✉ Corresponding author. E-mail: ycao@bjtu.edu.cn

Received: December 17, 2025; Revised: February 2, 2026; Accepted: March 20, 2026

railways (Xiong et al., 2023). In A-scan signal, particular attention is paid to defect echoes. A defect echo can be modeled as a modulated cosine signal, which is converted through analog-to-digital sampling into a waveform with a triangular-shaped envelope.

Defect detection in small-radius curve sections is a key technical challenge in line inspection. Although small-radius curves account for only about 20% of the total line length, they contain more than 50% of head checks (Zhang et al., 2025). This is because the high-rail gauge corner is subjected to long-term wheel-rail interaction forces (Chen and Yang, 2024; Fu et al., 2022), leading to fatigue and frequent head checks (Cannon et al., 2003).Moreover, strong flange noise generated by wheel–rail friction in small-radius curves leads to severe noise aliasing in ultrasonic echoes (Thompson and Jones, 2000), which obscures defect signals and increases the risk of missed detections. Consequently, accurate and reliable extraction of defect echoes from ultrasonic A-scan signals under strong noise interference remains an unresolved and critical problem in heavy-haul railway inspection.Previous studies have demonstrated that flange noise in small-radius heavy-haul curves exhibits pronounced non-Gaussian characteristics (Zhang et al., 2025). In the time domain, this noise typically appears as multi-peaked, grass-like echoes. In the frequency domain, its energy distribution is highly overlapped with that of defect echoes and lacks a single dominant frequency band. As a result, noise components cannot be effectively separated from defect-related signals, which significantly increases the difficulty of accurate defect echo localization. In contrast, defect echoes are the target signals in ultrasonic testing and usually exhibit a triangular envelope morphology, within which critical defect-related information is embedded (Li et al., 2020).

From a physical perspective, the dominant noise source in small-radius curves is generally attributed to flange noise generated by wheel–rail friction during inspection vehicle operation. Under strong lateral wheel–rail forces, intense friction occurs at the gauge corner, producing vibration-induced structural responses and surface plastic deformation (Wang et al., 2013). This type of structural noise exhibits waveform envelopes and temporal patterns similar to defect echoes, making it difficult to distinguish between noise and defect-related signals based solely on time-domain morphology.

In addition to structural noise, previous studies have shown that flange friction also induces low-frequency vibration components, with dominant frequencies ranging from several to tens of kilohertz (He et al., 2023). These components further complicate the signal composition and contribute to the observed multi-peak phenomena in A-scan signals. High-frequency components are generally attributed to thermal and electronic noise from ultrasonic acquisition systems and are commonly modeled as stochastic electrical noise. The coexistence and coupling of these heterogeneous noise sources lead to complex, non-Gaussian, and multi-scale interference patterns in ultrasonic A-scan signals. In current ultrasonic rail inspection systems, onboard low-pass filtering is commonly applied to attenuate high-frequency noise. However, due to substantial frequency band overlap between flange noise and ultrasonic echoes, defect signals remain difficult to separate effectively. Under such conditions, advanced signal processing methods, including wavelet decomposition and variational mode decomposition (VMD), also exhibit inherent limitations (Qi et al., 2022; Xu et al., 2025). VMD is prone to mode mixing, which smears defect-related components across multiple modes. Wavelet-based methods are constrained by the short duration of A-scan records, which limits decomposition depth. Consequently, for complex A-scan signals in small-radius heavy-haul curves, a single denoising technique is often insufficient to reliably extract defect echoes.

In recent years, the industrial inspection field has been developing towards high precision, automation, and intelligence. To address multi-source coupled interference, researchers commonly adopt multi-feature fusion strategies to enhance signal expression capability and defect recognition performance. Related studies show that multi-scale, multi-dimensional joint features can significantly improve diagnostic robustness. Currently, the signal feature evaluation system for such problems can be broadly divided into three categories.

The first category is the waveform-based evaluation system, which focuses on statistical features of signals in the time domain, such as amplitude, mean, variance, kurtosis, and margin factor. These features are simple to compute, have clear physical meanings, and are suitable for preliminary screening and anomaly detection of signals. For example, in rotating machinery fault diagnosis, kurtosis is highly sensitive to impact-type faults. However, for signals with strong non-stationarity, non-Gaussian properties, and strong background noise, relying solely on time-domain statistics often leads to a significant decrease in discriminative ability. Dong constructed a multi-feature vector containing multiple statistics such as time-domain kurtosis value, peak factor, clearance factor, and shape factor, combined with a one-dimensional convolutional network for fault identification (Dong and Lotfipoor, 2023). Zhao extracted six-dimensional time-frequency feature factors, effectively reflecting the intrinsic characteristics and physical meanings of different modal components, providing a basis for quantifying the feature differences between signal and noise wavefields and separation tasks (Zhao, 2025). Wu developed an improved semi-supervised learning algorithm based on a deep autoencoder (DAE) for ultrasonic time-domain signals of internal rail defects, used in ultrasonic rail defect detection. The DAE algorithm can identify observations in the dataset that significantly deviate from the rest (Wu and Zhu, 2023).

The second category is the time-frequency scale-based evaluation system. This system utilizes time-frequency analysis methods such as short-time Fourier transform (STFT) and continuous wavelet transform (CWT)to convert one-dimensional signals into two-dimensional time-frequency distributions, typically visualized as heatmaps. Time-frequency diagrams contain both time and frequency information and depict the joint distribution of energy on the time-frequency plane, making them suitable for analyzing non-stationary dynamic signals. Features based on time-frequency diagrams, such as energy centers and texture indicators, significantly enhance the characterization capability for complex fault modes. However, this method has a large computational load, and parameter settings such as time-frequency resolution and mother wavelet selection significantly affect the analysis results. For example, Zhou used STFT to convert vibration signals into time-frequency diagrams, which were then input into a convolutional neural network (CNN)for lifetime prediction and fault identification, representing a typical example of feature extraction using time-frequency heatmaps (Zhou et al., 2020). Cheng used CWT to generate time-frequency diagrams and input them as images into a CNN, demonstrating the advantages of CWT-based time-frequency diagram features in non-stationary fault diagnosis (Cheng et al., 2021). Ye combined STFT time-frequency two-dimensional image

features with convolutional networks and transfer learning, indicating that time-frequency diagrams have become an important carrier for feature extraction and intelligent diagnosis under complex working conditions (Ye et al., 2021).

The third category is the evaluation system based on modal decomposition. Its core idea is to adaptively decompose a complex signal into a series of intrinsic mode functions (IMFs). Typical methods include empirical mode decomposition (EMD) and its improved algorithms, as well as VMD. Taking VMD as an example, this method adaptively decomposes the signal into a set of narrowband modal components with well-defined center frequencies by constructing and solving a variational problem, thereby suppressing mode mixing to some extent. Based on this, researchers can extract features such as energy and center frequency from each modal component to construct multi-resolution feature vectors, enabling finer discrimination of different signal constituents. For instance, to address the strong backscattering noise in ultrasonic signals from rail welds, Hu proposed an adaptive ultrasonic signal reconstruction method named VSKR (VMD-Singular Value Decomposition-Kurtosis Reconstruction). This method leverages the differences in frequency characteristics between noise and defect signals decomposed by VMD, and effectively highlights defect echoes in practical experimental signals, demonstrating the effectiveness of VSKR (Hu et al., 2024). Wei proposed a joint algorithm named SMA-VMD-WTD (Slime Mould Algorithm-VMD-Wavelet Threshold Denoising), in which the SMA is employed to optimize the VMD parameters, and wavelet thresholding to suppress strong non-Gaussian noise in transient electromagnetic signals (Wei et al., 2022). Xing proposed multivariate VMD (MVMD) for denoising transient electromagnetic data, pointing out that traditional single-channel VMD and conventional filtering fail to fully utilize the spatial correlation of multi-channel data and are prone to mode mixing. Through joint decomposition of multi-channel data, MVMD achieves cleaner signal reconstruction and more stable anomaly identification on real data (Xing et al., 2024). Yan proposed a joint denoising framework called RSA-VMD-DNN (Reptile Search Algorithm-VMD-Deep Neural Network), which uses an intelligent optimization algorithm to adaptively find the optimal parameters for VMD to alleviate mode mixing, and then combines a deep neural network to achieve a nonlinear mapping from noisy to "clean" signals. Experiments show that this method outperforms traditional filtering, wavelet thresholding, and standard VMD in terms of signal to noise ratio (SNR) improvement and detail preservation (Yan et al., 2023). These works indicate that, when dealing with complex non-stationary and non-Gaussian noise, relying solely on a single method like VMD or wavelet analysis is often insufficient to achieve optimal results, and it is necessary to combine more complex models such as intelligent optimization and deep learning.

Existing research has established relatively mature feature-extraction frameworks in multiple domains, including the time, time-frequency and modal domains. However, most studies still treat multi-feature fusion as a simple concatenation of heterogeneous features into high-dimensional vectors, without deeply exploring the intrinsic physical relationships among features. Consequently, in the strongly non-Gaussian coupled-noise environment of small-radius heavy-haul curves, the core requirement is no longer merely to suppress noise, but to accurately represent defect echo information in A-scan signals while satisfying engineering constraints on signal amplitude, with the ultimate goal of generating clear and accurate B-scan images.

The core problem addressed in this study is how to construct a physically meaningful multi-feature correlated filtering paradigm. This paradigm should enable robust extraction of defect echoes from A-scan signals under complex noise conditions. It should also ensure accurate defect localization in the reconstructed B-scan images.

First, motivated by engineering practice in small-radius heavy-haul curves and supported by existing research findings, a noise-coupled ultrasonic A-scan signal model is established. The model explicitly characterizes defect echoes together with multiple non-Gaussian noise components, including structural vibration noise, low-frequency irrelevant components, and high-frequency electrical noise. This model provides a clear physical interpretation of noise aliasing phenomena observed in field A-scan signals and serves as the theoretical foundation for subsequent feature design.

Second, guided by the proposed signal model and the statistical and time–frequency characteristics of different signal components, a multi-feature extraction and fusion scheme is developed. Complementary features are extracted from the modal domain, time–frequency domain, and time domain, including VMD-based inter-layer correlation features, clustered multi-scale energy features from time–frequency representations, and waveform morphology descriptors derived from statistical distributions. These features are jointly integrated to enhance the discriminability between defect echoes and coupled noise.

Third, inspired by the ideal binary mask (IBM) concept, an IBM-based filtering framework is designed. The fused multi-dimensional features are used to construct an indexed binary mask with spatial adjacency constraints for ultrasonic signal reconstruction, enabling effective suppression of coupled noise while preserving defect echo amplitude information. The reconstructed A-scan signals are further converted into B-scan images for defect localization and performance evaluation.

The rest of this paper is organized as follows. Section II introduces the proposed IBM-based multi-feature fusion filtering framework. Section III presents the data description and experimental results. Section IV concludes the paper.

2 Methodology

2.1. A-scan Signal Model

Ultrasonic A-scan signal is the basic display form in ultrasonic testing, and its accurate modeling is crucial for signal processing, defect detection, and identification. For the noise-free case, the A-scan defect signal in Fig.1 (a) exhibits a clear defect echo around 48–50 μs. Its spectrum in Fig.1(b) shows that most spectral energy lies below 0.5 MHz, while the amplitudes in other frequency bands are very small. However, traditional idealized signal models often fail to fully describe the complex characteristics observed in real inspection environments.

In field inspections of small-radius heavy-haul curves, ultrasonic A-scan signals are affected by multiple disturbance sources, such as flange-related noise, and therefore cannot be adequately characterized by an ideal Gaussian white-noise assumption. A typical noisy A-scan signal is shown in Fig. 1(c). From the spectral comparison, the noise-free A-scan signal exhibits a relatively clear defect pulse within the gated interval, with its spectral energy being comparatively concentrated. The corresponding spectrum is shown in Fig. 1(d). After superimposing in-situ noise, the waveform becomes highly

fluctuating over the entire time range and forms grass-like multi-peaked structures, in which the defect pulse is strongly masked. Meanwhile, the amplitude spectrum remains dominated by low-frequency components and changes only marginally compared with the noise-free case. This observation indicates that, under strongly coupled non-Gaussian noise, Fourier amplitude spectra alone are insufficient to separate defect-related components from noise.

To quantify the statistical properties of the noise, the time-domain amplitude distribution of noise segments is analyzed. The histogram and the corresponding cumulative distribution function (CDF), shown in Fig. 1(e) and Fig. 1(f), respectively, exhibit a pronounced right-skewed and heavy-tailed distribution. Most samples lie within a low-to-medium amplitude range, while a small fraction of high-amplitude outliers contributes disproportionately to the overall energy and extreme-value statistics. This asymmetric and heavy-tailed behavior suggests impulsive non-Gaussian noise, which can severely degrade threshold-based detection and mean-square-error–driven filtering methods.

To further interpret the multi-scale noise structure, we apply VMD to a representative noise sample, shown in Fig. 1(g). The decomposed modes suggest three typical components:

(1) a low-frequency component exhibiting slow-varying baseline behavior, which is consistent with low-frequency irrelevant disturbances induced by flange friction vibration.

(2) a mid-frequency structural noise component with dispersed energy and multi-peaked diffuse structures, whose temporal scale overlaps with defect echoes and therefore constitutes a major source of aliasing and missed detections.

(3) a high-frequency component resembling fine-grained random perturbations, which can be attributed to high-frequency electrical noise with approximately white-noise characteristics.

Ultrasonic A-scan signal is represented in the following composite form, as shown in Eq. (1):

$$f_{\text{actual}}(t) = f_{\text{defect}}(t) + f_{\text{clutter}}(t) + f_{\text{low-freq}}(t) + n_{\text{high-freq}}(t) \quad (1)$$

where $f_{\text{defect}}(t)$ represents the defect echo signal (target signal), $f_{\text{clutter}}(t)$ represents structural noise (clutter), $f_{\text{low-freq}}(t)$ represents the low-frequency irrelevant component, and $n_{\text{high-freq}}(t)$ represents high-frequency electrical noise.

The defect echo signal mathematical model can be expressed as follows, as shown in Eq. (2):

$$f_{\text{defect}}(t) = \beta_{\text{de}} \cdot e^{-\alpha\left(t-\tau_{\text{de}}-\delta_{\tau}\right)^2} \cdot \cos\left(2\pi f_c\left(t-\tau_{\text{de}}-\delta_{\tau}\right)+\phi_{\text{de}}\right) \quad (2)$$

where β_{de} is the defect echo amplitude, related to the defect reflection intensity; τ_{de} is the theoretical echo arrival time; f_c is the ultrasonic probe center frequency; α is the envelope shape parameter; ϕ_{de} is the phase angle; and δ_{τ} is the additional time delay.

Structural noise originates mathematical model can be expressed as follows, as shown in Eq. (3):

$$f_{\text{clutter}}(t) = \sum_{i=1}^{F} \beta_i \cdot e^{-\alpha(t-\tau_i)^2} \cdot \cos\left(2\pi f_c(t-\tau_i)+\phi_i\right) \quad (3)$$

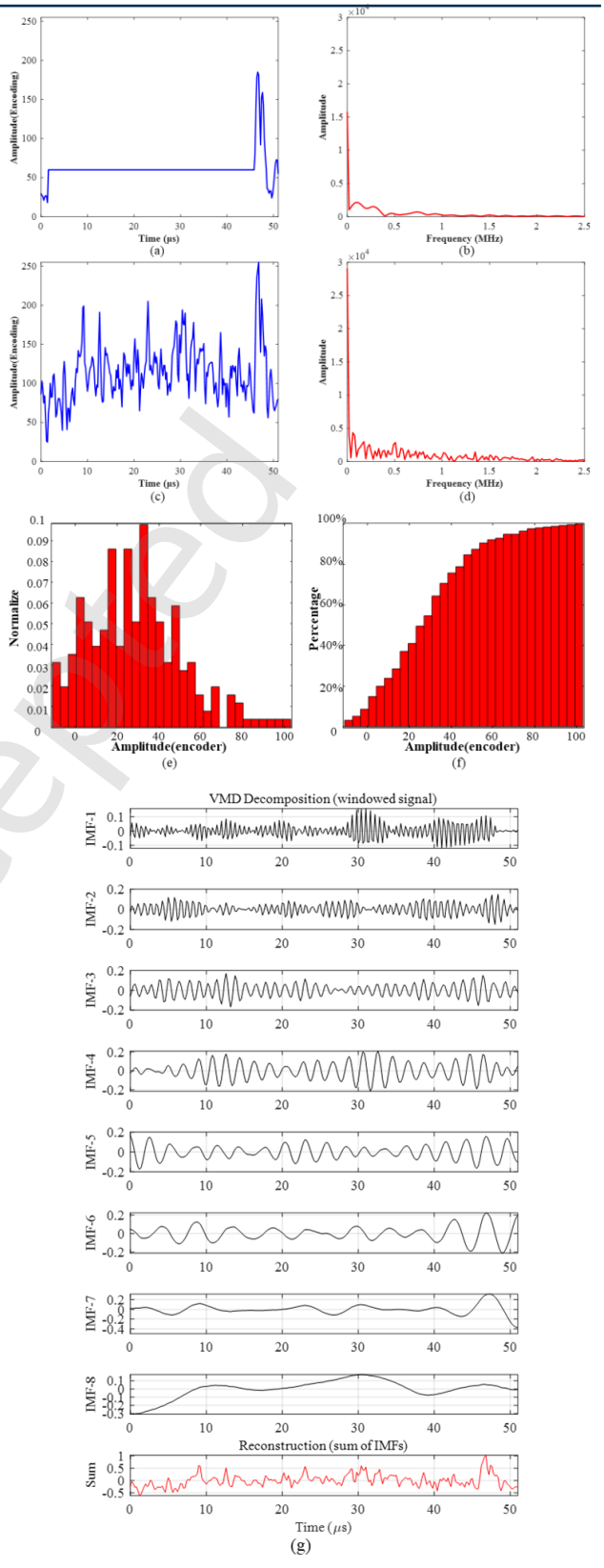

Fig. 1. A-scan and noise signals. (a) Noise-free ultrasonic A-scan defect signal. (b) Amplitude spectrum of the noise-free A-scan signal. (c) Ultrasonic A-scan signal with in-situ noise. (d) Amplitude spectrum of the noisy A-scan signal. (e) Amplitude histogram. (f) Cumulative distribution function (CDF). (g) VMD decomposition of a noisy A-scan signal

where F denotes the number of scatterers, and β_i, τ_i, and ϕ_i are random variables.

The low-frequency irrelevant component mathematical model can be expressed as follows, as shown in Eq. (4):

$$f_{\text{low-freq}}(t) = A_{\text{low}} \cdot \sin(2\pi f_{\text{low}} t + \phi_{\text{low}}) \tag{4}$$

where A_{low} is the amplitude, f_{low} is the frequency, and ϕ_{low} is the phase.

High-frequency noise mathematical form is shown in Eq. (5):

$$n_{\text{high-freq}}(t) = \text{N}(0, \sigma^2) \tag{5}$$

Comprehensive analysis shows that noise interference in ultrasonic A-scan signals can be divided into two categories according to its similarity to defect echoes. The first category is separable noise, including high-frequency electrical noise and low-frequency irrelevant components. Since their spectra are separated from the main frequency band of defect echoes, they can be effectively suppressed by frequency-domain methods. The second category is homologous interference, i.e., structural noise with time-domain characteristics similar to defect echoes, because the term $\beta e^{-\alpha(t-\tau_i)^2}$ shares the same envelope form. This overlap causes traditional filtering to fail, so distinguishing structural noise from defect echoes remains a key challenge.

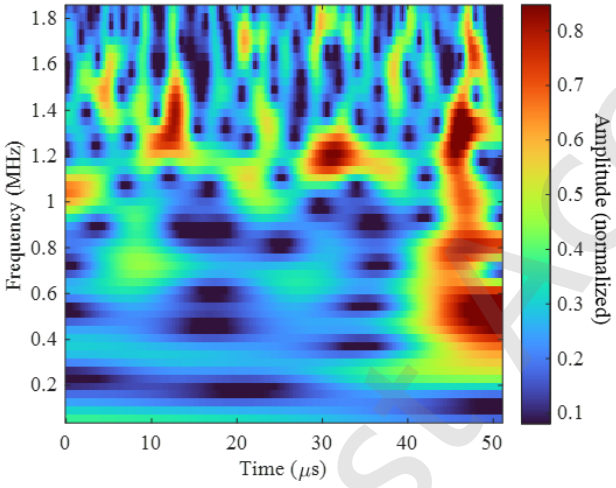

Fig. 2. CWT scalogram of the noisy A-scan signal.

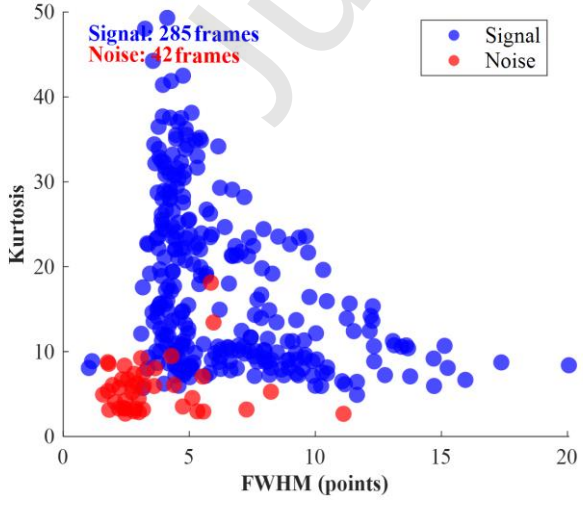

Fig. 3. Kurtosis-FWHM distribution of defect and noise samples in A-scan signals.

Nevertheless, further analysis based on Eq. (1) and statistical observations indicates that defect echoes and structural noise still exhibit distinguishable characteristics when examined jointly across multiple observation domains. At the modal level, structural noise in the mid-frequency range is typically concentrated in fewer decomposition layers, whereas defect echoes tend to span multiple adjacent modes, showing richer inter-layer responses as shown in Fig. 1(g). In the time–frequency domain, defect echoes usually form continuous and concentrated high-energy ridges, while structural noise appears as scattered and fragmented energy spots as shown in Fig. 2. From a time-domain statistical perspective, defect echoes behave as short-duration, high-amplitude impulsive signals, whose kurtosis and peak distributions show measurable separation from background noise, whereas structural and random noise exhibit more dispersed statistical characteristics as shown in Fig. 3.

These observations indicate that, although defect echoes and structural noise are difficult to separate using a single feature or a single domain, they exhibit intrinsic separability when examined jointly across modal, time–frequency, and statistical perspectives. Therefore, the core challenge lies not in simple noise suppression, but in effectively exploiting these multi-dimensional observational characteristics to enable reliable separation between defect echoes and structural noise, which motivates subsequent signal processing design.

2.2. *VMD Feature Extraction Algorithm Based on Inter-layer Feature Association*

VMD is an adaptive time-frequency analysis method that decomposes an input signal into several band-limited IMFs with distinct, non-overlapping center frequencies via variational optimization, enabling a fine multiscale representation of non-stationary signals.

This study exploits the inter-mode similarity structure revealed by VMD to develop a defect feature extraction algorithm based on inter-layer association. Applying VMD to the noise-contaminated A-scan signal yields multiple IMFs. In this decomposition, high-frequency IMFs are dominated by electrical noise, whereas low-frequency IMFs describe slow structural trends. Defect echo energy is concentrated in intermediate IMFs, where time-aligned local peaks appear at the same impulse position $t - \tau_i$. Exploiting these synchronized multi-mode peaks, the proposed inter-layer association algorithm searches for consistent extrema across intermediate IMFs and thereby separates coherent defect components from noise.

The ultrasonic A-scan signal $f_{\text{actual}}(t)$ is sampled and digitized to obtain the discrete-time sequence $f_{\text{actual}}(n)$, where $n = 0, 1, \ldots, N_t - 1$, and the sampling frequency is denoted by f_s. According to the multi-component model in (1), $f_{\text{actual}}(n)$ can be viewed as a superposition of defect echoes, structural vibration noise, low-frequency irrelevant components and high-frequency electrical noise.

For each gate, VMD is then applied to $f_{\text{actual}}(n)$ to obtain K IMFs $u_k(n)\}_{k=1}^{K}$ with non-overlapping center frequencies $\{f_k\}_{k=1}^{K}$. The discrete A-scan signal can be approximated, as shown in Eq. (6):

$$f_{\text{actual}}(n) \approx \sum_{k=1}^{K} u_k(n) \tag{6}$$

To further exploit the 'synchronized multi-mode response' of defect echoes, an inter-layer association criterion is constructed. For each time index n, a peak-indicator vector is defined, as shown in Eq. (7):

$$\mathbf{p}(n) = [p_1(n), p_2(n), \ldots, p_K(n)] \tag{7}$$

where $p_k(n)$ indicates whether mode k exhibits a significant peak at sample n.

The inter-layer association score $A_{VMD}(n)$ is then computed by aggregating the contributions of all mid-frequency modes, as shown in Eq. (8):

$$A_{\mathrm{VMD}}(n) = \sum_{k=l_{\mathrm{low}}}^{l_{\mathrm{high}}} p_k(n) \tag{8}$$

where l_{low} and l_{high} denote the lower and upper bounds of the mid-frequency band obtained from VMD decomposition.

For each candidate sample index n within this band, a composite criterion is constructed to evaluate its significance as a defect-related feature point.

Specifically, the feature score at sample n is defined as a weighted combination of two factors: the normalized amplitude of the reconstructed mid-frequency signal, and the normalized inter-layer association strength, represented by the number of correlated VMD modes at that sample, as shown in Eq. (9).

$$L_s(n) = \sum_{k=l_{\mathrm{low}}}^{l_{\mathrm{high}}} \begin{cases} 1, & P_k(n) \geq \delta \max(P_k) \\ 0, & \text{otherwise.} \end{cases} \tag{9}$$

where $L_s(n)$ denotes the inter-layer association count at sample index n, $\delta \in (0,1)$ is a layer-wise significance threshold.

By normalizing both terms and combining them in a weighted-sum form, the sample exhibiting the maximum composite score is selected as the representative feature point of the VMD output. The associated index set is defined as shown in Eq. (10).

$$\mathrm{V}_{\mathrm{VMD}} = \left\{ n \,|\, \max \left(\frac{A_{\mathrm{VMD}}(n)}{\max A_{\mathrm{VMD}}} + \frac{L_s(n)}{l_{\mathrm{high}} - l_{\mathrm{low}}} \right) \right\} \tag{10}$$

which represents candidate defect-echo locations with strong multi-mode support, while the remaining samples are regarded as noise or trend-dominated. In this way, the inter-layer association mechanism exploits the synchronized multi-mode response of defect echoes and suppresses single-mode noise, preserving time-frequency-consistent defect features for subsequent masking and reconstruction.

2.3. *CWT Time-Frequency Peak Energy Clustering Algorithm*

The CWT time-frequency scalogram provides a compact representation of the joint distribution of time, frequency, and energy, allowing clear visualization of echo-energy evolution in a two-dimensional plot. When the noise-contaminated A-scan signal from Fig.1(c) is transformed via CWT (Wu et al., 2023), the resulting time-frequency map in Fig.2 exhibits a distinct bright triangular ridge around the defect echo. This indicates that impact energy is simultaneously distributed over multiple mid-frequency scales, consistent with Fig. 1(g) where defect energy spreads across several modes. Inspired by mean-based clustering ideas (Hu et al., 2021), a triangular ridge clustering algorithm is proposed on the CWT scalogram to extract defect-related energy clusters.

The model in Eq. (1) further implies that defect echoes exhibit localized time-frequency energy concentrations, whereas structural noise is more diffuse. On this basis, an adaptive CWT time-frequency peak-energy clustering algorithm is designed on the scalogram to automatically extract triangular ridge-shaped energy clusters associated with defect echoes. In summary, the proposed CWT time-frequency peak-energy clustering algorithm first restricts the time-frequency search region via time-domain masking and band-limited energy, then identifies triangular bright ridges in the CWT scalogram using amplitude and width constraints, and finally collects their centers as CWT feature indices for subsequent IBM construction. The CWT using the Morlet mother wavelet is computed as shown in Eq. (11):

$$W_f(j,n) = \sum_{m=0}^{N_t-1} f_{\mathrm{actual}}[m] \psi_j^*(n-m) \tag{11}$$

The normalized amplitude scalogram is obtained as shown in Eq. (12):

$$A(j,n) = \frac{|W_f(j,n)|}{\max_{j,n} |W_f(j,n)|} \tag{12}$$

Let f_j denote the equivalent frequency corresponding to scale j; then $A(j,n)$ forms the CWT time-frequency energy distribution. For impact-type defects, $A(j,n)$ typically exhibits a bright-band ridge concentrated in the center and widening toward both sides, which provides the basis for clustering.

To suppress low-amplitude noise, the time-domain signal is first normalized, as shown in Eq. (13):

$$\tilde{f}[n] = \frac{f_{\mathrm{actual}}[n]}{\max_n |f_{\mathrm{actual}}[n]| + \varepsilon} \tag{13}$$

A binary mask is defined as shown in Eq. (14):

$$m_t[n] = \begin{cases} 1, & \tilde{f}[n] \geq A_{\mathrm{thr}}, \\ 0, & \text{otherwise.} \end{cases} \tag{14}$$

The mask is then dilated by convolution, as shown in Eq. (15):

$$\hat{m}_t[n] = \begin{cases} 1, & (m_t * h)[n] > 0, \\ 0, & \text{otherwise,} \end{cases} \qquad h = [1,1,1,1,1] \tag{15}$$

The main defect frequency band is restricted according to the transducer and rail response, as shown in Eq. (16):

$$\mathrm{F} = \{ j \;/\; f_{\min} \leq f_j \leq f_{\max} \} \tag{16}$$

Then, the time-domain energy distribution within the effective band is calculated as shown in Eq. (17):

$$E_t[n] = \hat{m}_t[n] \sum_{j \in \mathrm{F}} A(j,n)^2 \tag{17}$$

The K_{cand} largest $E_t[n]$ values are selected as candidate time indices, as shown in Eq. (18):

$$\mathrm{T}_{\mathrm{cand}} = \{ n_{(1)}, n_{(2)}, \ldots, n_{(K_{\mathrm{cand}})} \} \tag{18}$$

For each candidate column $n_c \in \mathrm{T}_{\mathrm{cand}}$, the frequency-energy vector is defined as shown in Eq. (19):

$$a_j = A(j, n_c), \quad j \in \mathrm{F} \tag{19}$$

The main peak and its amplitude are determined as shown in Eq. (20):

$$j^* = \arg\max_j a_j, \qquad p_{\max} = a_{j^*} \tag{20}$$

The contiguous ridge region J_c is then identified using a full width at half maximum (FWHM) type threshold $\ell = \alpha p_{\max}$ $(0 < \alpha < 1)$, as shown in Eq. (21):

$$\mathrm{J}_c = \{ j \in \mathrm{F} \;/\; a_j \geq \ell \} \tag{21}$$

Only ridges with width $W = |J_c|$ satisfying, as shown in Eq. (22):

$$W_{\min} \le W \le W_{\max} \tag{22}$$

which are considered valid triangular bright spots, whose centers are defined as shown in Eq. (23):

$$f_c = f_{j^*}, \quad (n_c, f_c) \in \mathrm{R} \tag{23}$$

To prevent repeated detection of the same ridge, exclusion rules in the time-frequency plane are applied, as shown in Eq. (24), Eq. (25):

$$|n_c - n_m| > \Delta t_{\min} \tag{24}$$

$$|f_c - f_m| > \Delta f_{\min} \tag{25}$$

The remaining ridge centers R thus form the CWT clustering feature set in the time-frequency plane. For subsequent IBM construction in the time domain, the corresponding time-index set is defined as shown in Eq. (26):

$$C_{\mathrm{CWT}} = \{n_c / (n_c, f_c) \in \mathbf{R}\} \tag{26}$$

where C_{CWT} collects the CWT-based defect feature indices obtained by the peak-energy clustering algorithm and serves as one branch of the feature set used for IBM construction.

2.4. Sliding Window Waveform Morphology Based Defect-Noise Discrimination Method

This section proposes a defect -noise discrimination method based on sliding-window waveform morphology features. In the time domain, defect echoes in the Eq. (1) behave as short-duration, high-amplitude pulses, in contrast to the grass-like noise background composed of structural echoes, coupling noise and instrument noise. In small-radius heavy-haul curves, this strong noise contamination makes traditional fixed-amplitude thresholds or single-feature criteria insufficient. To exploit the waveform differences, the proposed method computes local kurtosis and FWHM within a sliding window and constructs point-level masks, so that samples consistent with impulsive, narrow triangular pulses are retained as defect echoes, while background noise is effectively suppressed in complex inspection environments.

The kurtosis formula is defined as shown in Eq. (27):

$$Ku = \frac{E\left[(x-\mu)^4\right]}{\sigma^4} \tag{27}$$

Based on the above statistical laws, an explicit waveform morphology feature domain is constructed in this study. Let (Ku_{th} , $M_{\min}$, $M_{\max}$) be the kurtosis threshold and peak width upper and lower limits respectively. If a window satisfies:

$$\begin{cases} Ku \ge Ku_{th} \\ M_{min} \le \mathrm{FWHM} \le M_{max} \\ A_{max} \ge A_{th} \end{cases}$$

Then the window is judged as a defect candidate region, where A_{th} is an amplitude constraint based on engineering experience.

Since extracting a single feature from the entire A-scan signal cannot capture local impact structures, this section adopts a window-based analysis method.

The window length L and step size S (overlap rate) are set with reference to Shannon's theorem. If $L \ge 2S$ is satisfied, it can significantly reduce boundary effects.

Let the discrete ultrasonic A-scan signal have length N_t . The total number of windows as Eq. (28)

$$J_W = \left\lfloor \frac{N_t - L}{S} \right\rfloor + 1 \tag{28}$$

which is approximately N_t / S mask positions. For the j_W -th window, centered at $n_{jW} = j_W S$, local morphology features $(Ku^{(j_W)}, \mathrm{FWHM}^{(j_W)}, A_{\max^{(j_W)}})$ are computed, and a window-level indicator is defined as Eq. (29).

$$m_j = \begin{cases} 1, & Ku^{(j_W)} \ge Ku_{\mathrm{th}}, M_{\min} \le \mathrm{FWHM}^{(j_W)} \le M_{\max}, A_{\max^{(j_W)}} \ge A_{\mathrm{th}} \\ 0, & \mathrm{otherwise} \end{cases} \tag{29}$$

The corresponding time-domain feature set T_{KF} as Eq. (30)

$$T_{\mathrm{KF}} = \bigcup_{j_W : m_j = 1} \{n \mid n_{j_W} \le n < n_{j_W} + L\} \tag{30}$$

which collects all sample indices covered by defect-candidate windows and serves as the time-domain feature set for subsequent IBM construction.

2.5. Multi-Feature Fusion Mask and Edge Judgment

To enhance defect-echo extraction under complex noise interference, an index-based ideal binary mask (IBM) filtering algorithm is constructed by fusing multi-domain features and spatial edge judgment.

Three feature extraction modules are employed, including time-domain statistical analysis, CWT-based time–frequency ridge detection, and VMD-based inter-layer correlation analysis. Each module outputs a set of candidate sample indices, denoted as T , C , and V , respectively. These index sets are mapped to binary indicator sequences of identical length, where a sample is assigned a value of 1 if its index belongs to the corresponding set and 0 otherwise.

The initial union mask is defined by the logical OR operation of the three binary sequences, as shown in Eq. (31):

$$M_{\mathrm{IBM}}(n) = \begin{cases} 1, & T(n) + C(n) + V(n) \ge 1 \\ 0, & T(n) + C(n) + V(n) = 0 \end{cases} \tag{31}$$

which ensures that any sample supported by at least one observation domain is retained, effectively reducing the risk of missed detections.

In practice, the union mask may contain contiguous segments of activated samples. To avoid edge truncation and preserve the physical morphology of defect pulses, a spatial refinement is applied. For each connected segment, the sample with the maximum amplitude is selected as the representative peak, and an edge judgment is performed to ensure the presence of a complete rise–fall structure. The effective support of each peak is then expanded from the local minimum preceding the rising edge.

Based on the validated peak indices and refined supports, the final IBM mask $M_{\mathrm{IBM}}(n)$ is constructed and applied to the original gated A-scan signal via pointwise multiplication, as shown in Eq. (32):

$$f_{\mathrm{IBM}}(n) = f_{\mathrm{actual}}(n) \cdot M_{\mathrm{IBM}}(n) \tag{32}$$

where retained defect peaks preserve their original amplitudes, while unsupported noise samples are explicitly suppressed. This non-linear masking strategy avoids peak attenuation inherent to linear filtering and ensures that only samples supported by multi-domain evidence and valid edge morphology are preserved.

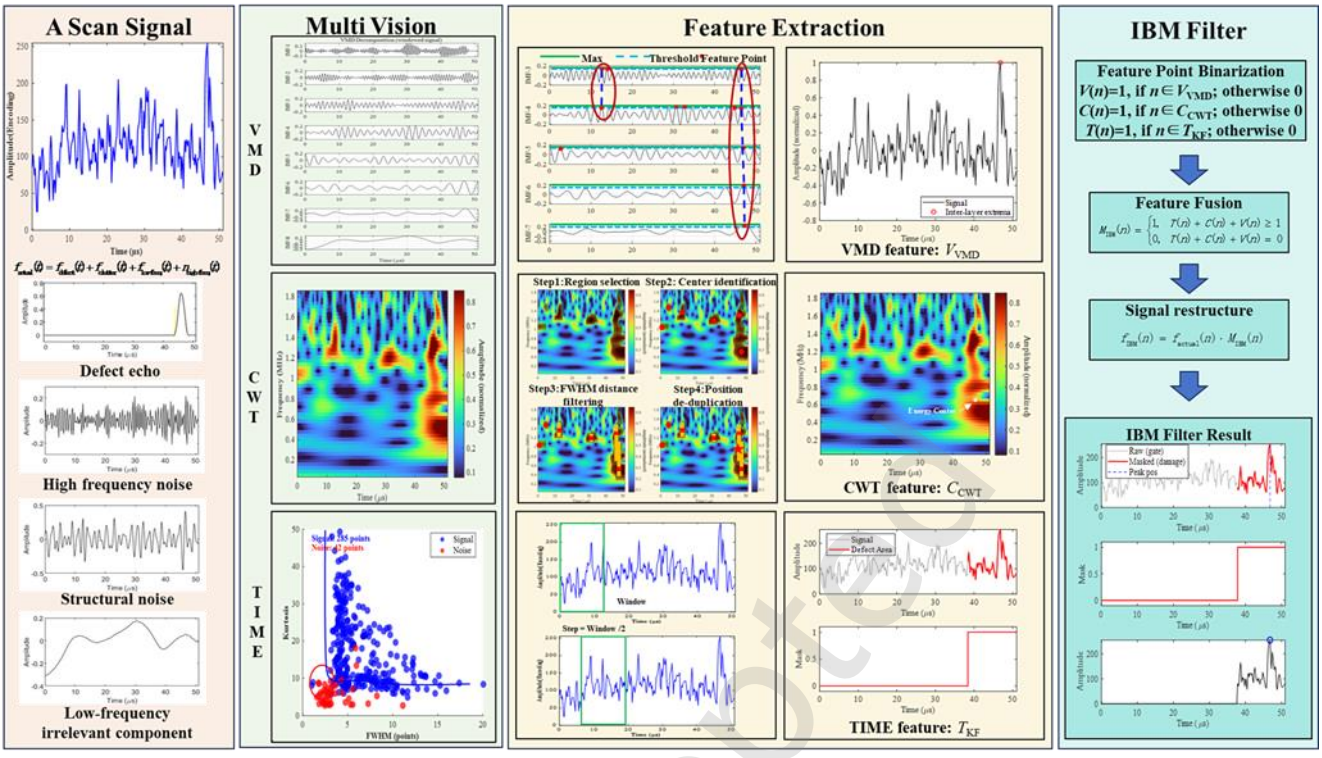

Fig. 4. Flow chat of multi-feature fusion masking method.

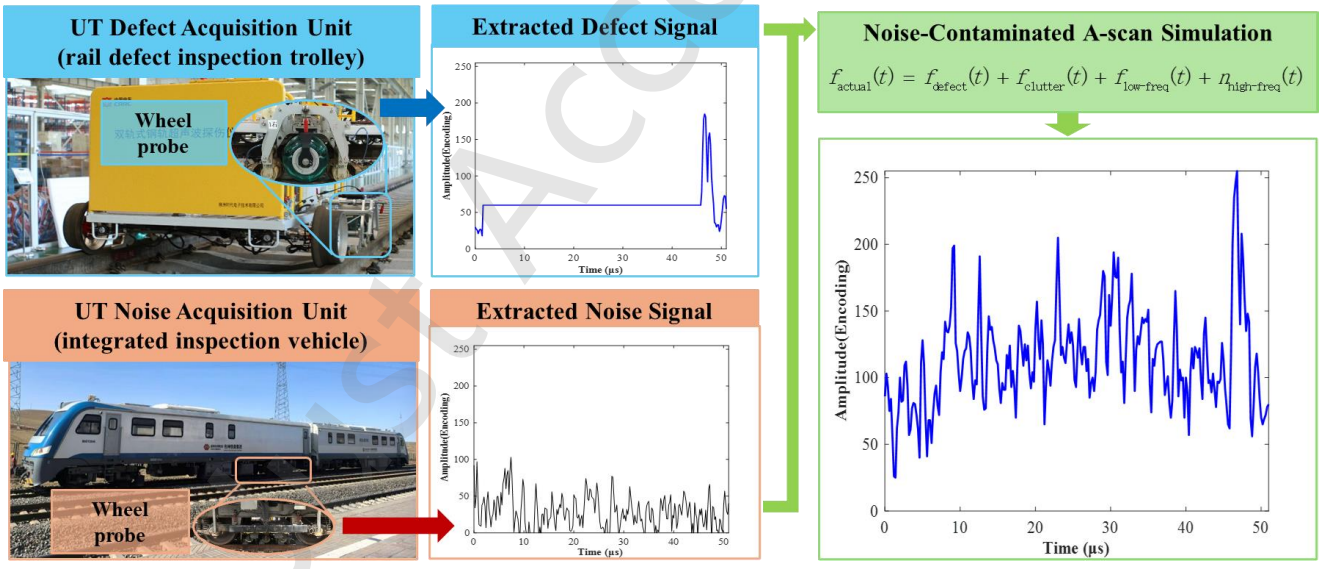

Fig. 5. Schematic diagram of A-scan data.

2.6. Presented Method

This paper proposes a multi-feature fusion filtering algorithm for heavy-haul railway ultrasonic A-scan signals, whose overall workflow is illustrated in Fig.4. Initially, taking the raw A-scan signal as input, three sets of features are extracted in parallel within the feature extraction module.

Specifically, the module comprises three parallel processing paths: VMD based on inter-layer feature association is employed to obtain the mode-related feature set V ; CWT time-frequency analysis is applied to the signal, and the feature set C is formed via energy-peak clustering; sliding-window waveform morphology analysis is utilized to extract time-domain features such as kurtosis and peak width, constituting the feature set T .

Subsequently, the three types of features are fused on a unified time axis to generate a point-level binary mask. Within the IBM mask filtering framework, this mask is used to perform spatially-varying filtering on the A-scan signal. The final output is a filtered A-scan signal that effectively preserves defect echoes while significantly suppressing noise.

3 Experiments and Results

3.1. Data Description

The dataset used in this study consists of ultrasonic A-scan signals collected from two types of real-world inspection equipment, as illustrated in Fig.5. defect signals were acquired from 60 kg/m rails on an experimental railway line in China,

yielding 285 valid defect echoes, including rail head nuclear defects and joint-related reflections. Noise-only signals were obtained from small-radius curves on a heavy-haul railway line in China using an integrated inspection vehicle operating under actual service conditions. The noise contains strongly coupled components such as structural vibration, wheel--rail interaction, and electronic interference. A total of 42 valid noise frames were recorded using a Tektronix TBS 1072B oscilloscope.

Owing to the sparsity of defects and the lack of real-time A-scan acquisition interfaces on inspection vehicles, it is difficult to obtain real 'defect + noise' A-scan signals simultaneously on heavy-haul lines. Therefore, this study simulates noise-contaminated defect signals by superimposing standard defect signals from the experimental line with real flange noise from the heavy-haul line. To ensure consistency, the noise signals were up-sampled for sampling-frequency alignment. Each of the 285 defect signals was randomly paired with one of the 42 noise signals, producing 285 A-scan signals with varying noise levels for evaluating feature extraction and filtering performance.

3.2. Evaluation metrics

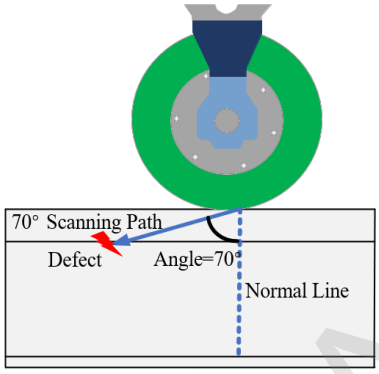

Fig. 6. Physical model of ultrasonic scanning for defect detection.

Fig. 6 illustrates the physical model of single-channel ultrasonic inspection using a 70° probe. When the ultrasonic wave encounters a defect, an echo is generated and received by the wheel-mounted transducer. By analyzing the echo time-of-flight in the A-scan signal, the spatial position of the defect can be estimated and further mapped into a B-scan image.

In heavy-haul railway ultrasonic inspection, the objective of signal processing is not merely to enhance signal quality, but to ensure accurate defect localization in B-scan reconstruction. Accordingly, this study adopts an imaging-oriented evaluation metric to directly assess whether the extracted A-scan features can effectively support reliable defect detection and positioning. In ultrasonic B-scan imaging, the defect position is determined by the echo time-of-flight and can be expressed as shown in Eq. (33):

$$D = \frac{1}{2} v_s t \tag{33}$$

where D denotes the ultrasonic propagation path length in rail, v is the shear wave velocity of the rail material, and t is the echo time-of-flight. Based on field calibration, the shear wave velocity is set to v_s =3.26 mm/μs.

Considering sampling quantization, sensor installation errors, and field condition variations, millimeter-level localization accuracy is adopted as the criterion for correct detection. The localization error is defined as shown in Eq. (33):

$$\Delta D = \left| D_{\text{det}} - D_{\text{gt}} \right| \tag{33}$$

where D_{det} is the detected defect position and D_{gt} is the ground-truth position obtained from manual verification or historical records. An allowable localization error range of ΔD=0–3 mm is adopted, corresponding to a time tolerance of approximately 0–5μs. With a sampling rate of 5 MHz (0.2 μs per sample), this tolerance is equivalent to approximately 0–9 sampling points.

A defect is regarded as correctly detected if $\Delta D \leq 3$ mm. otherwise, it is considered a missed detection. Accordingly, detection accuracy is defined as the proportion of true defects that are correctly localized within the allowable error range. This metric directly reflects the capability of the proposed method to support B-scan imaging and maintenance decision-making, rather than relying solely on signal-level quality measures.

3.3. Sliding Window Waveform Morphology Feature Experiment

This study adopts a sliding-window waveform morphology–based method for defect–noise discrimination to improve signal identification accuracy in rail defect detection.

Table 1. Comparison Experiment of Sliding Window Waveform Morphology Features (Unit: signal frames)

Identification Accuracy	No window FWHM=2, *Ku*=2		*L*=64, *S*=32 FWHM=2, *Ku*=3		*L*=64, *S*=32 FWHM=3, *Ku*=3		*L*=64, *S*=32 FWHM=5, *Ku*=3		*L*=64, *S*=32 FWHM=2, *Ku*=2	
	Frames	Ratio	Frames	Ratio	Frames	Ratio	Frames	Ratio	Frames	Ratio
Exact match	120	**42.11%**	85	29.82%	84	29.47%	23	8.07%	111	38.95%
0-4 points error	198	**69.47%**	152	53.33%	143	50.18%	42	14.74%	196	68.77%
0-9 points error	210	**73.68%**	161	56.49%	151	52.98%	43	15.09%	209	73.33%
0-20 points error	220	77.19%	183	64.21%	170	59.65%	48	16.84%	222	**77.89%**
0-5% error	224	78.60%	197	69.12%	183	64.21%	49	17.19%	239	**83.86%**
0-10% error	228	80.00%	219	76.84%	196	68.77%	49	17.19%	249	**87.37%**
0-25% error	234	82.11%	268	94.04%	228	80.00%	53	18.60%	279	**97.89%**
Error above 25%	51	17.89%	17	5.96%	57	20.00%	232	81.40%	6	**2.11%**

Table 2. Comparison Experiment of VMD Feature Extraction(Unit: signal frames)

Identification Accuracy	VMD		VMD (layer 7)		VMD (layer 8)		VMD (layer 9)		VMD (layer 10)	
	Frames	Ratio	Frames	Ratio	Frames	Ratio	Frames	Ratio	Frames	Ratio
Exact match	49	17.19%	54	18.95%	61	**21.40%**	62	21.75%	66	23.16%
0-4 points error	196	68.77%	204	71.58%	204	**71.58%**	199	69.82%	196	68.77%
0-9 points error	229	80.35%	232	81.40%	234	**82.11%**	229	80.35%	223	78.25%
0-20 points error	251	88.07%	259	**90.88%**	257	90.18%	248	87.02%	251	88.07%
0-5% error	265	92.98%	271	**95.09%**	269	94.39%	263	92.28%	266	93.33%
0-10% error	274	96.14%	274	**96.14%**	272	95.44%	264	92.63%	267	93.68%
0-25% error	278	97.54%	279	**97.89%**	279	**97.89%**	277	97.19%	277	97.19%
Error above 25%	7	2.46%	6	**2.11%**	6	**2.11%**	8	2.81%	8	2.81%

The corresponding quantitative results are summarized in Table 1. Compared with the window-free global analysis, the global method achieves acceptable accuracy in the low-error interval (0–9 points), with an accuracy of 73.68%. However, it exhibits a relatively high missed-detection ratio of 17.89% in the severe-error interval (error > 25%), indicating that global statistics tend to dilute the local characteristics of transient defect pulses under strong noise interference.

By contrast, the sliding-window strategy segments long A-scan signals into short frames for localized analysis, which enhances the representation of transient morphological features. Using a multi-dimensional morphology feature combination based on kurtosis and FWHM, the method enables finer discrimination between defect echoes and noise. Parameter sensitivity analysis shows that when the kurtosis threshold is set to 2 and the FWHM threshold is set to 2, the recognition accuracy in the 0–9 point interval reaches 73.33%, which is comparable to the global analysis method, while significantly reducing the proportion of signals with errors exceeding 25%. Under this parameter setting, the overall performance of the algorithm is optimal.

Fig. 7 illustrates the processing results of the proposed sliding-window waveform morphology–based method applied to the noise-contaminated A-scan signal corresponding to Fig. 1(c). The upper plot shows the original noisy signal, in which the defect echo within the range of 48–50 μs is correctly identified despite strong noise fluctuations. The lower plot presents the corresponding point-level mask, where the mask value rises to 1 at the defect location, indicating that the local kurtosis and FWHM satisfy the predefined defect morphology criteria. These results demonstrate that the proposed method can reliably extract defect-related features under strong noise interference.

3.4. VMD Feature Extraction Experiment

Fig.8 illustrates the inter-layer extrema detection result obtained from the VMD-based feature association algorithm. Although the noisy A-scan signal exhibits strong fluctuations across the entire time axis, the algorithm successfully identifies a single time point where multiple VMD modes simultaneously produce aligned local extrema. This point corresponds precisely to the defect echo location around 48-50 μs. The stability of this multi-mode consensus peak-despite significant noise contamination-demonstrates that the inter-layer association criterion effectively suppresses noise and enhances the accuracy of defect localization.

This confirms the suitability of VMD inter-layer extrema as a robust defect indicator under strongly coupled noise.

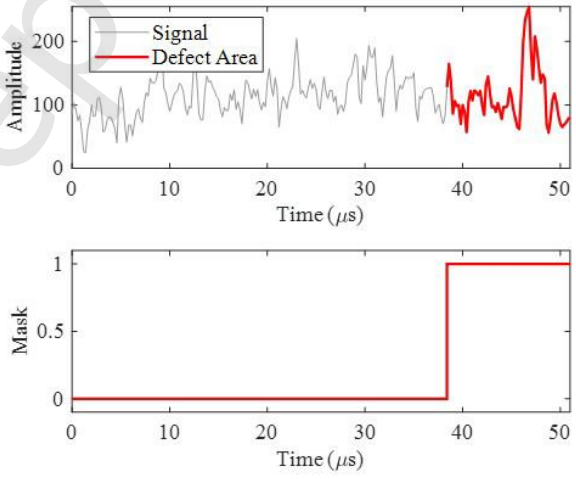

Fig. 7. Sliding-window waveform morphology results of A-scan signals.

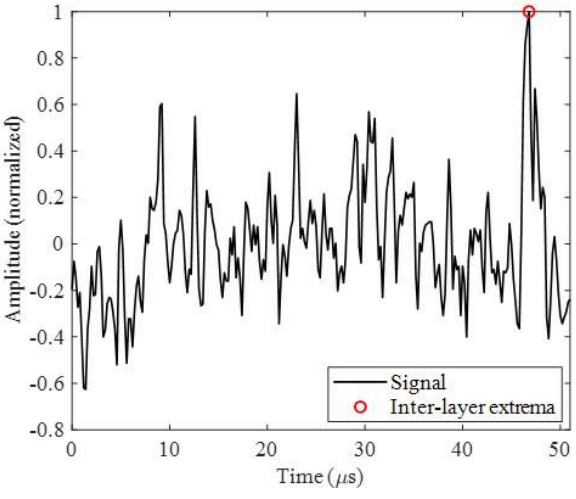

Fig. 8. Inter-layer VMD association results for A-scan signals.

First, this study analyzes the influence of the decomposition layer number K in VMD, the result show in **Table 2**. If K is too small, mode mixing easily occurs, compressing different feature components into the same mode; if K is too large, redundant modes are generated, dispersing energy and amplifying noise. Comparing the error distributions of four configurations (K=7, 8, 9, 10), when K=8, the cumulative proportion of errors less than 10 points reaches 82.11%, which is the optimal scheme. This indicates that the signal energy is more concentrated in the IMFs with stable center frequencies and clear boundaries, achieving a balance between under-decomposition and over-decomposition, providing a high-quality basis for subsequent feature extraction.

Further comparison across different numbers of VMD decomposition layers shows that the scheme with K=8 achieves the best overall performance. For K=8, the cumulative proportion of samples with localization error within 0–9 points reaches 82.11%, the highest among all schemes, while the misjudgment rate in the error range ≥25% is as low as that for K=7. Considering both the proportion of high-precision localization and the suppression of large errors, the inter-layer correlation VMD scheme with K=8 offers the best compromise between accuracy and stability, thereby improving the accuracy of rail defect echo localization and identification.

3.5. *Multi-Feature Fusion Algorithm*

Fig.9 presents the CWT time-frequency representation of the noise-contaminated A-scan signal. Although significant noise components are distributed across different frequency bands, a distinctive energy concentration emerges near the defect location at approximately 48-50 μs. This region exhibits a bright triangular structure spanning multiple mid-frequency scales, indicating that the defect echo maintains coherent energy distribution across both time and frequency. The marked energy center further highlights the dominant frequency component associated with the defect. Compared with the severe fluctuation in the waveform domain, the CWT map provides a clearer and more stable signature for identifying defect-related features. These results confirm that CWT effectively enhances defect visibility and supports multi-scale peak energy clustering for robust defect extraction.

However, analysis shows that each single-feature method has inherent limitations under different noise conditions. The CWT method is sensitive to continuous bright bands of defect energy in the time-frequency domain, but is prone to miss low-amplitude or noise-shielded defects. The TIME method can effectively distinguish triangular defect pulses from low-kurtosis wide-peak noise through sliding window statistics of peak kurtosis and peak width, but is susceptible to positioning deviation for coupled signals with elongated main peak morphology. The VMD inter-layer correlation algorithm performs well in handling low-frequency drift and peak position offset, but has limited suppression capability for pure high-frequency electrical noise and isolated scatter points. Therefore, no single observation domain can guarantee both robustness and accuracy under all working conditions.

To address this issue, a multi-feature fusion framework is introduced. The CWT ridge centers, TIME morphological features, and VMD inter-layer correlation indices are treated as three complementary observation channels. Candidate defect locations are first identified in each domain. These candidates are then aligned along the time axis, and a joint decision is performed within a local neighborhood. Points supported by multiple features are marked as high-confidence defects, while partially supported points are also retained according to physically consistent rules to ensure defect integrity.

Fig.10 illustrates the processing results of the Multi-Feature Fusion Algorithm. In the top panel, the raw gated A-scan signal shows that the defect echo is heavily masked by strong noise, making single-feature detection unreliable. The middle panel presents the multi-feature fusion mask, which rises sharply to 1 around the defect region, indicating accurate localization of defect-related samples. The bottom panel displays the reconstructed signal after applying the fusion mask, where noise components are significantly suppressed and the defect peak is clearly enhanced, aligning well with the expected impulse position. These results demonstrate that integrating VMD, CWT, and time-domain morphology features enables robust multi-scale defect detection under severe noise interference.

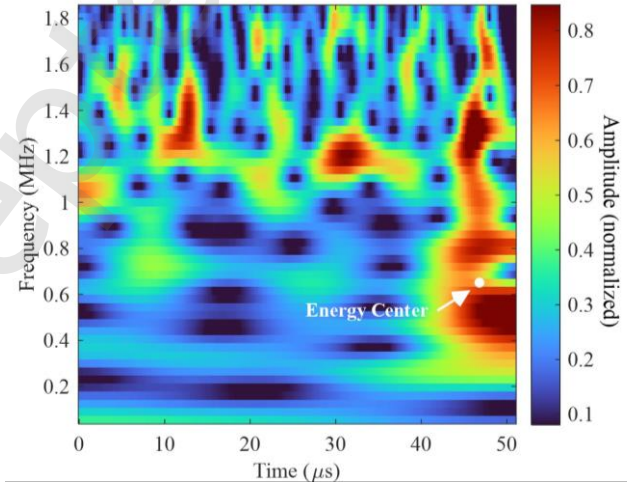

Fig. 9. CWT clustering results of A-scan signal.

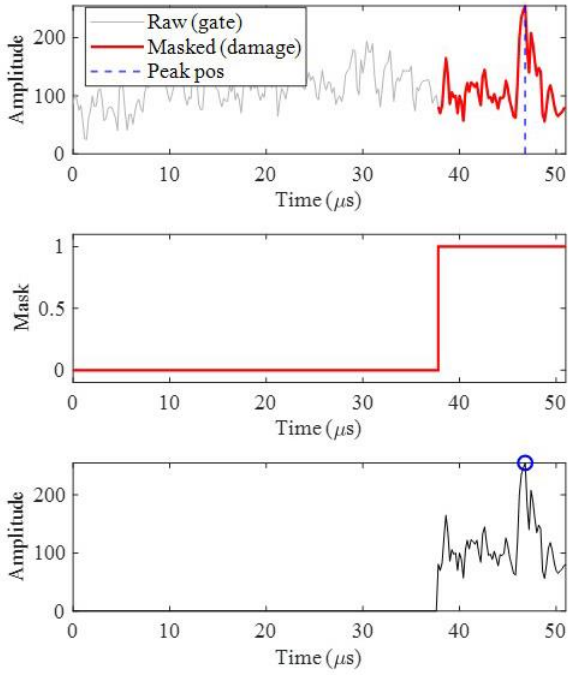

Fig. 10. Results of the proposed multi-feature fusion masking method for A-scan signals

Table 3. Single Algorithm Comparison Experiment (Unit: signal frames)

Identification Accuracy	CWT		TIME		VMD		Low-pass Filter		**ours**	
	Frames	Ratio	Frames	Ratio	Frames	Ratio	Frames	Ratio	Frames	Ratio
Exact match	19	6.67%	111	38.95%	74	25.96%	69	24.21%	**154**	**54.04%**
0-4 points error	127	44.56%	196	68.77%	207	72.63%	185	64.91%	**241**	**84.56%**
0-9 points error	160	56.14%	209	73.33%	225	78.95%	223	78.25%	**258**	**90.53%**
0-20 points error	203	71.23%	222	77.89%	250	87.72%	256	89.82%	**268**	**94.04%**
0-5% error	224	78.60%	239	83.86%	265	92.98%	272	95.44%	**276**	**96.84%**
0-10% error	232	81.40%	249	87.37%	270	94.74%	275	96.49%	**276**	**96.84%**
0-25% error	270	94.74%	279	97.89%	280	98.25%	280	98.25%	**282**	**98.95%**
Error above 25%	15	5.26%	6	2.11%	5	1.75%	5	1.75%	**3**	**1.05%**

Table 4. Fusion Experiment (Unit: signal frames)

Identification Accuracy	CWT-TIME		TIME-VMD		CWT-VMD		**ours**	
	Frames	Ratio	Frames	Ratio	Frames	Ratio	Frames	Ratio
Exact match	121	42.46%	145	50.88%	89	31.23%	**154**	**54.04%**
0-4 points error	215	75.44%	236	82.81%	217	76.14%	**241**	**84.56%**
0-9 points error	233	81.75%	252	88.42%	235	82.46%	**258**	**90.53%**
0-20 points error	254	89.12%	265	92.98%	258	90.53%	**268**	**94.04%**
0-5% error	264	92.63%	274	96.14%	268	94.04%	**276**	**96.84%**
0-10% error	270	94.74%	275	96.49%	271	95.09%	**276**	**96.84%**
0-25% error	279	97.89%	**282**	**98.95%**	280	98.25%	**282**	**98.95%**
Error above 25%	6	2.11%	**3**	**1.05%**	5	1.75%	**3**	**1.05%**

The experimental results demonstrate that the proposed multi-feature fusion algorithm consistently outperforms any single method as well as all pairwise fusion schemes. As shown in Table 3, taking frame-level recognition accuracy as an example, when CWT, TIME, VMD (layer 8), and low-pass filtering are applied individually, the numbers of exact match frames are 19, 111, 74, and 69, respectively. After applying multi-feature fusion, the number of exact match frames increases significantly to 154. When the high-accuracy interval defined by a localization error of 0–9 sampling points is considered, the cumulative proportions of all single methods remain below 78.95%, whereas the proposed fusion scheme achieves a cumulative proportion of 90.53%.

From an engineering perspective, this result indicates that the proposed method reaches a defect localization accuracy of approximately 3 mm in more than 90% of cases, which is consistent with the minimum spatial resolution required for practical ultrasonic rail inspection. Under a 3 mm error boundary, the missed-detection rate is reduced from about 21% to 9.47%, demonstrating a substantial improvement in inspection accuracy.

In addition, for the severe deviation interval (error ≥25%), the proportion of erroneous frames produced by the multi-feature fusion method is only 1.05%, which is significantly lower than the 1.75%–5.26% observed for single-feature methods. This implies that false localization outside the actual inspection region is effectively suppressed.

Further comparisons with pairwise fusion schemes in Table 4 show that the numbers of exact match frames for CWT–TIME, TIME–VMD, and CWT–VMD are 121, 145, and 89, respectively, with corresponding severe misjudgment ratios of 2.11%, 1.05%, and 1.75%. In contrast, the full CWT–TIME–VMD fusion further increases the number of exact match frames to 154 while maintaining the lowest severe misjudgment ratio of 1.05%.

4 Conclusion

This paper proposes a multi-feature fusion masking method for defect echo extraction in heavy-haul railway ultrasonic A-scan signals under non-Gaussian coupled noise. Targeting the noise interference and echo aliasing issues in complex operational environments, this study integrates complementary feature extraction mechanisms (VMD inter-layer correlation, CWT time-frequency clustering, and sliding window morphology discrimination) with an Indexed Binary Mask (IBM) filtering algorithm, achieving precise extraction of defect echoes.

Experimental validation on 285 simulated defect-noise signals demonstrates that the proposed method outperforms single-feature and pairwise fusion schemes: the proportion of signals with positioning errors $\leq$ 9 points reaches 90.53%, and the misjudgment rate for errors exceeding 25% is only 1.05%. It effectively suppresses non-Gaussian coupled noise while preserving the amplitude and morphological integrity of defect echoes, addressing the false/missed detection problems in traditional methods.

From an engineering perspective, the proposed method provides a reliable solution for millimeter-level defect localization

in heavy-haul railway ultrasonic inspection. By operating directly at the A-scan level and maintaining peak amplitude consistency, it supports accurate B-scan reconstruction and defect positioning, offering practical value for condition-based maintenance and transportation safety.

Discussion

It should be noted that the proposed model adopts an additive coupling assumption between defect echoes and noise.

In cases with strong phase interference or nonlinear coupling, this assumption may no longer hold, and model extensions would be required.

In addition, the current algorithm is designed for single-channel ultrasonic inspection.

Future work will focus on multi-channel scenarios, where coupled noise exists across channels.

Since different channels correspond to different inspection regions and defect types, their echo waveforms also exhibit distinct characteristics.

Extending the proposed framework to such heterogeneous multi-channel conditions requires further investigation.

CRediT authorship contribution statement

Jiangtao Zhang: Writing – original draft, Visualization, Methodology, Investigation, Conceptualization.

Yuan Cao: Supervision, Resources, Formal analysis, Investigation, Writing – review & editing.

Yuntong An: Writing – review & editing, Validation, Project administration, Investigation, Conceptualization.

Feng Wang: Conceptualization, Methodology, Writing – review & editing.

Yunkui Sun: Conceptualization, Methodology, Writing – review & editing.

Shuai Su: Supervision, Resources, Conceptualization.

Replication and data sharing

The replication package, including the dataset and the source code used in this study, is available on ETS-Data at https://ets-data.sciopen.com/ETSD.2026.9190009.

Declarations of competing interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Acknowledgements

This work was supported by the National Natural Science Foundation of China (Grant Nos. 52525214 and U2368202).

References

Cannon, D.F., Edel, K.-O., Grassie, S.L., Sawley, K., 2003. Rail defects: An overview. Fatigue & Fracture of Engineering Materials & Structures 26(10), 865–886.

Cao, Y., He, S., Wang, F., Su, S., Sun, Y., 2025. A large-model-enhanced method for rail surface defect detection in heavy-haul railway. IEEE Transactions on Intelligent Transportation Systems 26(12), 23328–23341.

Chen, L., Yang, H., 2024. Adaptive composite anti-disturbance control for heavy haul trains. Transportation Safety and Environment 6(1), tdad009.

Chen, Z., Wang, Q., He, Q., Yu, T., Zhang, M., Wang, P., 2022. Cufuse: Camera and ultrasound data fusion for rail defect detection. IEEE Transactions on Intelligent Transportation Systems 23(11), 21971–21983.

Cheng, Y., Lin, M., Wu, J., Zhu, H., Shao, X., 2021. Intelligent fault diagnosis of rotating machinery based on continuous wavelet transform–local binary convolutional neural network. Knowledge-Based Systems 216, 106796.

Dong, K., Lotfipoor, A., 2023. Intelligent bearing fault diagnosis based on feature fusion of one-dimensional dilated CNN and multi-domain signal processing. Sensors 23(12), 5607.

Fu, Y., Rao, W., Yang, H., 2022. Linear active disturbance rejection control of heavy-haul train operation based on an interval type-2 fuzzy logic system model. Transportation Safety and Environment 4(4), tdac037.

He, Y., Zhou, Q., Xu, F., Sheng, X., He, Y., Han, J., 2023. An investigation into the effect of rubber design parameters of a resilient wheel on wheel–rail noise. Applied Acoustics 205, 109259.

Hu, J., Cao, Y., Huang, J., Yu, T., Yao, J., Wang, P., He, Q., 2024. Adaptive signal reconstruction based on VMD for rail welding joint defect detection. Russian Journal of Nondestructive Testing 60(11), 1249–1262.

Hu, X., Niu, R., Tang, T., 2021. Pre-processing of metro signaling equipment fault text based on fusion of lexical domain and semantic domain. Journal of the China Railway Society 43(2), 78–85.

Li, Y., Yao, F., Jiao, S., Huang, W., Zhang, Q., 2020. Identification and classification of rail damage based on ultrasonic echo signals. In: Proceedings of the 39th Chinese Control Conference (CCC 2020), 3077–3082. IEEE.

Mićić, M., Brajović, L., Lazarević, L., Popović, Z., 2023. Inspection of RCF rail defects-Review of NDT methods. Mechanical Systems and Signal Processing 182, 109568.

Qi, T., Wei, X., Feng, G., Zhang, F., Zhao, D., Guo, J., 2022. A method for reducing transient electromagnetic noise: Combination of variational mode decomposition and wavelet denoising algorithm. Measurement 198, 111420.

Shi, J., Shi, H., Wu, Z., Li, J., 2024. An on-board damage detection method for heavy-haul railway bridge based on sensitivity analysis of bogie responses. IEEE Sensors Journal 24(4), 4642–4655.

Thompson, D.J., Jones, C.J.C., 2000. A review of the modelling of wheel/rail noise generation. Journal of Sound and Vibration 231(3), 519–536.

Wang, W., Guo, H., Du, X., Guo, J., Liu, Q., Zhu, M., 2013. Investigation on the damage mechanism and prevention of heavy-haul railway rail. Engineering Failure Analysis 35, 206–218.

Wei, X., Feng, G., Qi, T., Guo, J., Li, Z., Zhao, D., et al., 2022. Reduce the noise of transient electromagnetic signal based on the method of SMA-VMD-WTD. IEEE Sensors Journal 22(15), 14959–14969.

Wu, X., Yang, W., Cao, J., 2023. Fault diagnosis of railway point machines based on wavelet transform and artificial immune algorithm. Transportation Safety and Environment 5(4), tdac072.

Wu, Y., Zhu, X., 2023. Rail defect detection using ultrasonic A-scan data and deep autoencoder. Transportation Research Record 2677(7), 62–73.

Xing, K., Li, S., Qu, Z., Zhang, X., 2024. Time-domain electromagnetic noise suppression using multivariate variational mode decomposition. Remote Sensing 16(5), 806.

Xing, Y., Li, W., Liu, Z., 2020. Research on the performance of heavy-haul rail breakage detection using track circuit. Transportation Safety and Environment 2(4), 292–304.

Xiong, L., Jing, G., Wang, J., Liu, X., Zhang, Y., 2023. Detection of rail defects using NDT methods. Sensors 23(10), 4627.

Xu, Z., Chen, X., Xu, J., 2025. Multi-modal multi-sensor feature fusion spiking neural network algorithm for early bearing weak fault diagnosis. Engineering Applications of Artificial Intelligence 141, 109845.

Yan, F., Pi, S., Yu, Q., Lin, J., 2023. Transient electromagnetic data noise suppression method based on RSA-VMD-DNN. IEEE Geoscience and

Remote Sensing Letters 21, 1–5.

Ye, L., Ma, X., Wen, C., 2021. Rotating machinery fault diagnosis method by combining time–frequency domain features and CNN knowledge transfer. Sensors 21(24), 8168.

Yu, C., Liu, Y., Cao, Y., Sun, Y., Su, S., Yang, W., Wang, W., 2024. Improved YOLOv8 for B-scan image flaw detection of the heavy-haul railway. Measurement Science and Technology 35(7), 076106.

Zhang, J., Cao, Y., Wang, F., Sun, Y., Su, S., An, Y., 2025. Research on ultrasonic A-scan filtering algorithm for small radius curved rails in heavy haul railway. China Railway Science 46(2), 49–59.

Zhao, N., 2025. Feature deconstruction and physical coupling mechanism of complex noise fields in DAS-VSP. Ph.D. dissertation. Changchun, China: Jilin University.

Zhou, S., Xiao, M., Bartos, P., Filip, M., Geng, G., 2020. Remaining useful life prediction and fault diagnosis of rolling bearings based on short-time Fourier transform and convolutional neural network. Shock and Vibration 2020(1), 8857307.

Author biography

Jiangtao Zhang received his B.S. degree from Lanzhou Jiaotong University, Lanzhou, China, in 2020. He is currently working toward the Ph.D. degree in Traffic Information Engineering and Control in the School of Automation and Intelligence at Beijing Jiaotong University, Beijing, China. His recent research focuses on intelligent operation and maintenance of heavy-haul railway lines, weak signal feature extraction. Email: 23111065@bjtu.edu.cn

Yuan Cao received his B.S. degree in Electric Engineering and Automation from Dalian Jiaotong University and Ph.D. degree in Traffic Information Engineering and Control from Beijing Jiaotong University in 2004 and 2011 respectively, where he is now a Professor. Since 2006, he has participated in many engineering practice, especially in the signal and communication system of high-speed railway. He has taken part in several key national research projects in the field of high-speed train control systems. His research interests include health management in high speed railway system. Email: ycao@bjtu.edu.cn

Yuntong An received his B.S. degree in Automation (Railway Signal) in Traffic Information Engineering and Control from Beijing Jiaotong University in 2021, where he is currently working toward the Ph.D. degree at School of Automation and Intelligence. His research interests include fault diagnosis and condition monitoring in train control systems, and life prediction for railway key equipment.Email: 21111059@bjtu.edu.cn

Feng Wang (Member, IEEE) obtained Ph.D. degree in traffic information engineering and control from Beijing Jiaotong University (BJTU), Beijing, China, in 2021.

He is currently an Assistant Professor with the School of Automation and Intelligence, Beijing Jiaotong University. His research interests include statistical modeling and prognostics with high-dimensional data.

Yongkui Sun received his B.S. degree in Automation (Railway Signal) and Ph.D. degree in Traffic Information Engineering and Control from Beijing Jiaotong University in 2016 and 2021, respectively. Now he is an Associate Professor in National Engineering Research Center of Rail Transportation Operation Control System, School of Automation and Intelligence, Beijing Jiaotong University, Beijing, China. His research interests include fault diagnosis and condition monitoring in train control systems, and life prediction for railway key equipments. Email: sunyk@bjtu.edu.cn

Shuai Su received the PH.D. degree from Beijing Jiaotong University, Beijing, China, in 2016. He is currently working as the deputy director in the Frontiers Science Center for Smart High-speed Railway System, Beijing Jiaotong University. His current research interests include energy-efficient operation and control in railway systems, intelligent train control and dispatching, such as timetable optimization, optimal driving strategy and rescheduling.. Email: shuaisu@bjtu.edu.cn