This is a Just Accepted version, keep updated with https://doi.org/10.26599/LLES.2025.9660014

Research Article

Application of an Improved RepVGG Algorithm to DAS-Based Train Operation Detection

Zhongsheng Jiang[1,2], Zhouchang Hu[1,2], Shuang Yang[1,2], Xingrong Jiang[2,3], Yang Zhang[1,2], Yuquan Tang[1,2,✉], Saifullah Jamali[2,4], Zhirong Zhang[1,2,3], Nek Muhammad Shaikh[4], Baddar ul ddin Jamali[5]

[1] University of Science and Technology of China, Hefei 230026, China

[2] Anhui Provincial Key Laboratory of Photonic Devices and Materials, Anhui Institute of Optics and Fine Mechanics, HFIPS, Chinese Academy of Sciences, Hefei 230031, China

[3] School of Advanced Manufacturing Engineering, Hefei University, Hefei 230601, Anhui, China

[4] Institute of Physics, University of Sindh, Jamshoro 71000, Pakistan

[5] Dr. M.A. Kazi Institute of Chemistry, University of Sindh, Jamshoro 71000, Pakistan

✉Address correspondence to Yuquan Tang, laserway@aiofm.ac.cn

Received: 29 September 2025; Revised: 18 November 2025; Accepted: 28 November 2025

Abstract

Train operation monitoring based on distributed acoustic sensing (DAS) has attracted increasing attention for its potential to ensure railway safety and support intelligent transportation systems. However, accurately recognizing train vibration signals and distinguishing them from noise remains a significant challenge. This study proposes a classification-model-based framework to identify both the train direction of travel and interfering noise signals. The approach transforms orthogonally-demodulated vibration power spectrum features into image representations, where conventional image-processing techniques are first applied to extract candidate vibration regions. A deep-learning-based classifier is then employed to perform the final recognition. To improve discriminative feature modeling, we adopt the RepVGG architecture with structural re-parameterization and further integrate a lightweight Normalization-based Attention Module (NAM). This design enhances the model's ability to capture direction-of-travel cues while suppressing redundant information, thereby reducing false positives and false negatives. The dataset is randomly divided into training and testing subsets with an 80/20 ratio. Four models including VGG16, ResNet18, RepVGG, and the proposed RepVGG-NAM are trained and evaluated on this dataset. Experimental results demonstrate that RepVGG-NAM achieves superior classification performance, attaining an accuracy of 97.45% on the test set, which corresponds to absolute improvements of 0.2%, 0.3%, and 0.2% over VGG16, ResNet18, and RepVGG, respectively. These consistent improvements verify the effectiveness of NAM in dynamically calibrating channel weights to highlight salient features. Overall, the proposed RepVGG-NAM framework provides a high-accuracy and computationally efficient solution for DAS-based train vibration signal classification, contributing to the advancement of intelligent railway monitoring systems.

Keywords: DAS, classification, RepVGG, NAM

0 Introduction

With the rapid expansion of Chinese railway network, the need for reliable along-line safety monitoring has become increasingly critical. Accurate localization and real-time monitoring of trains are essential for issuing early warnings to trackside maintenance personnel, thereby ensuring operational safety. DAS has emerged as a promising technology in this domain. By capturing Rayleigh backscattering and analyzing the demodulated vibration power-spectrum signals together with their time-delay information, DAS enables vibration detection and localization along optical fibers. Its advantages, including high sensitivity, long-range coverage, and ease of deployment, make DAS-based systems particularly suitable for large-scale railway corridor monitoring [1-4].

Nevertheless, the high sensitivity of DAS also results in highly diverse and complex vibration signals, which complicates the task of accurately identifying vibration types. Consequently, vibration recognition has become a key research focus in distributed fiber-optic sensing [5-6]. Early efforts employed classical machine-learning methods. For instance, Lu et al. (2017) combined threshold-based screening with support vector machines (SVMs) to achieve an 88% recognition rate for train-related vibrations [7]. Liu et al. (2018) used random forest classifiers to distinguish between watering, tapping, and rolling vibrations, reporting accuracies of 93.79%, 97.36%, and 97.06%, respectively [8]. Deep learning methods have also been introduced. Sun et al. (2020) reconstructed DAS signals into pulse-scanning images and applied convolutional neural networks (CNNs) to detect vehicle, pedestrian, and leak-induced vibrations, achieving a recognition rate of 98% [9]. More recently, Bu et al. (2022) applied an SVM to classify tapping, shaking, walking, and noise signals with an accuracy exceeding 97% [10]. Meng et al. (2023) employed a ResNet-based network to classify wind, knocking, and background noise signals acquired by a Φ-OTDR system, and compared the performance of ResNet on temporal-spatial images and time-frequency images. The recognition accuracies based on temporal-spatial and time-frequency representations reached 99.49% and 98.23% [11]. While these approaches demonstrate high accuracy, traditional machine-learning pipelines rely on hand-crafted feature extraction, and many deep learning studies have focused primarily on architectural complexity [12-15], which can improve accuracy but often at the cost of slower inference and reduced deployment efficiency.

To overcome these limitations, this work develops a deep learning framework based on the RepVGG architecture [16-17] and integrates a lightweight NAM [18-19]. RepVGG leverages structural re-parameterization to decouple training and inference, thereby achieving both high accuracy and efficient inference. The incorporation of NAM enables dynamic calibration of channel weights, suppressing redundant features and enhancing discriminative cues. This design significantly improves the recognition of train position and direction in DAS signals while maintaining computational efficiency. The proposed approach thus provides a practical and scalable solution for high-accuracy, low-overhead railway vibration monitoring.

1 Acquisition of Materials

1.1 Data Collection with DAS

The DAS system employs a high-repetition-rate, narrow-linewidth pulsed laser as the light source. Pulses are launched into the fiber, and Rayleigh backscattered light is collected at the fiber input. By exploiting time-of-flight differences of the received backscatter, vibrations at different positions along the fiber can be detected and localized. In the train-detection setup, the pulserepetition rate is set to 1 kHz[20]. An acquisition card digitizes the backscattered intensity; every 250 spatial samples (range bins) is grouped as one data block and subjected to digital quadrature demodulation to recover the optical amplitude. A moving-average filter is then applied to obtain the vibration power-spectrum data. One minute of power-spectrum data is assembled into an image with 240 rows, which facilitates subsequent image-processing-based segmentation of regions that may contain vibration signals; within these regions, the signals are further identified as true vibrations or noise. The overall architecture is shown in Fig. 1.

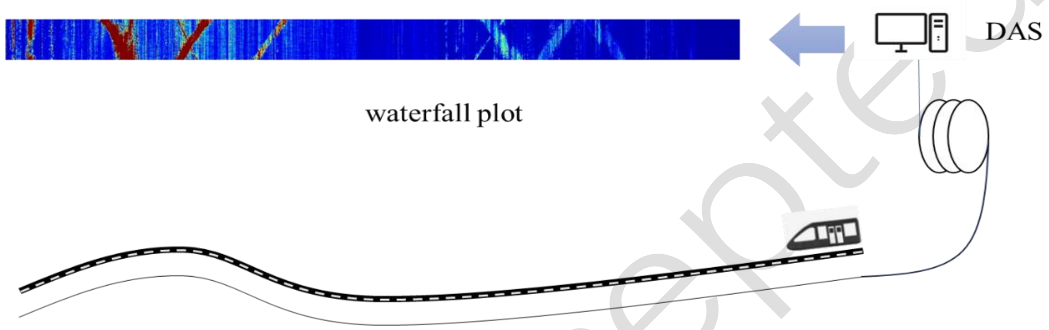

Fig. 1 System architecture diagram

1.2 Data Preprocessing

OpenCV-based vibration-signal extraction: (1) Apply mean filtering with a 5×5 convolution kernel to denoise the image. (2) Perform morphological closing (using morphologyEx) with a 17×7 rectangular structuring element to fill horizontal and vertical voids and reconnect broken segments in the vibration power-spectrum image. (3) Conduct block-wise thresholding: divide the image horizontally into three segments and apply grayscale thresholds of 160, 120, and 50 to the first, second, and third segments, respectively, to compensate for the attenuation of signal intensity along the fiber. (4) Use OpenCV's findContours to extract contours, compute their minimum bounding rectangles, and crop/save the corresponding ROIs. The saved images are categorized into four classes-simultaneous trains in both directions (bi-directional traffic), upbound trains, downbound trains, and noise-as illustrated in Fig. 2.

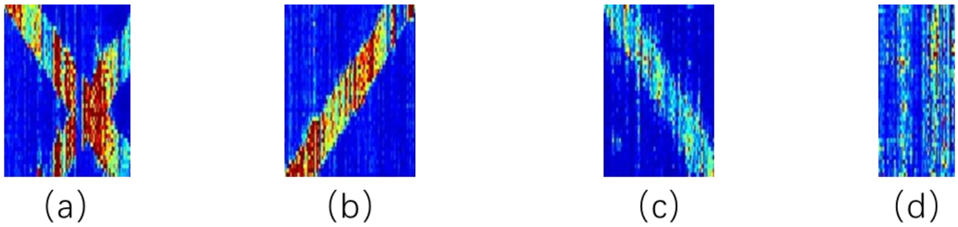

Fig. 2 (a) Trains meeting (b) Upward (c) Downward (d) Noise

The dataset is split at an 80/20 ratio into a training set and a test set. The training set is used to train the classification network to fit the data features, whereas the test set is used to tune the network hyperparameters during training and to evaluate the model performance. The sample distribution of the dataset is summarized in Table 1.

Table 1. Sample distribution

Dataset	Trains meeting	Upward	Downward	Noise	Total
Training Set	2000	2000	2000	2000	8000
Test Set	500	500	500	500	2000
Total	2500	2500	2500	2500	10000

2 Train-Signal Classification Model

To accurately recognize and localize trains, this study employs a RepVGG network augmented with the NAM attention mechanism (RepVGG-NAM) to classify train vibration signals. We further compare against ResNet18 [21], VGG16 [22-23], and the vanilla RepVGG classifier to validate the effectiveness of image-based recognition algorithms for train vibration signals. VGG16 represents the classic deep plain convolutional architecture, whereas ResNet18 represents a modern efficient design with residual connections. The two models provide a complementary contrast in terms of architectural philosophy and computational complexity, and both are widely recognized as standard baselines in the field.

2.1 RepVGG Module

In conventional CNN architecture design, multi-branch models represented by ResNet and Inception [24] alleviate vanishing gradients through residual connections or parallel convolutional paths, but their complex topologies markedly reduce inference efficiency[25]. In contrast, plain single-path architectures such as VGG are hardware-friendly yet difficult to scale to very deep networks due to insufficient gradient diversity. RepVGG introduces an innovative “train-inference decoupling” philosophy: via structural re-parameterization, it combines the training advantages of multi-branch topologies with the inference efficiency of a single-path architecture, thereby redefining the design paradigm for efficient models. Inspired by ResNet, RepVGG adopts a “train then convert” strategy (train-time multi-branch, inference-time single-path). During training, each basic module (RepVGG block) is constructed as a three-branch structure:

$$y = x + g(x) + f(x) \tag{1}$$

In this formulation, x, $g(x)$, and $f(x)$ denote the identity mapping, a 1×1convolution, and a 3×3convolution, respectively. Compared with ResNet’s periodically repeated residual design, RepVGG introduces a multi-branch structure at every layer to maximize gradient diversity (see Fig. 3). During inference, mathematically equivalent transformations merge the multi-branch topology into a single 3×3 convolution, eliminating additional computational overhead.

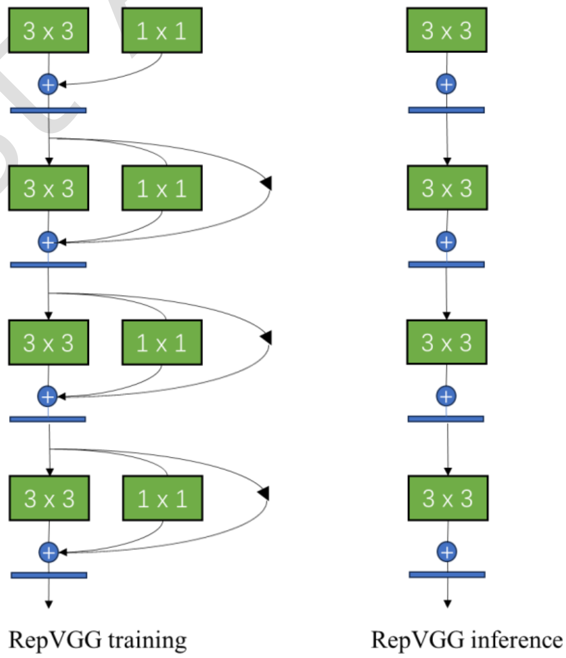

Fig. 3 Sketch of RepVGG architecture

2.2 NAM Module

The NAM is a lightweight and efficient attention mechanism. Its core idea is to exploit the internal statistics of batch normalization (BN) layers to automatically identify salient feature channels and spatial regions, while suppressing redundant information via a sparsity-inducing weight penalty. Inspired by the dual-path design of CBAM, NAM redefines the computations of the channel and spatial attention submodules, thereby substantially reducing parameters and computational overhead.

$$B_{out} = BN(B_{in}) = \gamma \frac{B_{in} - \mu_b}{\sqrt{\sigma_b^2 + \epsilon}} + \beta \qquad (2)$$

In the formulation, γ and β denote the trainable affine transformation parameters representing scaling and shifting, respectively. μ_b and σ_b^2 represent the mean and variance computed over a mini-batch B. The channel attention submodule, as illustrated in Fig. 4 and defined by Equation (3), outputs a modulation tensor denoted as Mc. Here, γ serves as a scaling factor assigned to each channel, with corresponding weights denoted by w_i. This mechanism leverages the scaling coefficients from BN to emphasize the relative importance of different channels by adaptively adjusting their contributions. Furthermore, spatial attention is incorporated by applying BN-derived scaling factors across the spatial dimensions, enabling the model to estimate the significance of individual pixels. The corresponding spatial attention submodule, also depicted in Fig. 4 and described by Equation (4), produces the attention map Ms, where λ represents the spatial scaling coefficient and w_i again denotes the learnable weights associated with each spatial location.

$$w_i = \frac{\gamma_i}{\sum_{j=0} \gamma_j} \qquad (3)$$

$$w_i = \frac{\lambda_i}{\sum_{j=0} \lambda_j} \qquad (4)$$

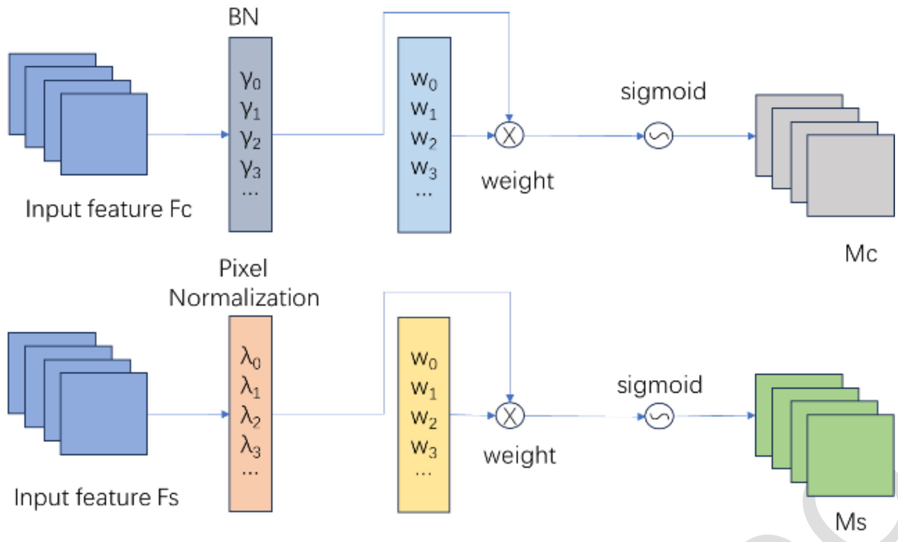

Fig. 4 Sketch of NAM architecture

2.3 RepVGG-NAM Classification Model

RepVGG leverages structural re-parameterization to achieve a strong balance between computational efficiency and predictive performance. During training, the network adopts a multi-branch topology that facilitates feature learning, while at inference the structure is converted into a streamlined single-path architecture, thereby reducing latency and improving efficiency. Despite these advantages, the backbone of RepVGG exhibits limited capacity in modeling inter-channel dependencies, which restricts its ability to emphasize salient features in complex scenarios.

To address this limitation, we integrate a NAM into the outputs of Stage 3 and Stage 4 of RepVGG. NAM introduces channel attention by adaptively reweighting feature channels, thus enhancing the representation of discriminative cues. Specifically, NAM derives attention weights directly from the scaling factors of BN layers. Since these factors inherently capture distributional statistics of the feature maps, they provide a principled means to estimate channel importance: larger absolute scaling values indicate channels with richer and more salient feature content, whereas smaller values suggest redundancy.

By embedding NAM at the output of each residual block in Stages 3 and 4, the network dynamically adjusts channel-weight distributions during deeper feature extraction. This design choice is particularly effective because Stage 3 encodes mid-level semantic patterns (such as object parts), whereas Stage 4 captures higher-level abstractions (such as global object contours). As a result, the integration of NAM strengthens RepVGG's ability to discriminate critical features in DAS-based vibration signals, while preserving the efficiency of single-path inference. The overall architecture of the proposed RepVGG-NAM network is illustrated in Fig. 5.

From an interpretability perspective, the proposed RepVGG-NAM framework can be understood as a channel-wise feature selection mechanism tailored to DAS vibration spectrograms. NAM first normalizes the intermediate feature maps and then computes channel-wise importance scores based on their global statistics. Channels that capture continuous, high-energy spectral traces produced by train-induced vibrations and direction-of-travel patterns are assigned larger weights, whereas channels dominated by low-energy, irregular background disturbances receive smaller weights. In this way, the classifier effectively suppresses redundant noise information while highlighting physically meaningful DAS responses along the fiber. This behavior is consistent with expert knowledge that train passages generate structured, coherent time–frequency signatures, whereas environmental noise tends to be fragmented and less coherent. Therefore, the performance gains of RepVGG-NAM can be attributed not only to a more powerful backbone, but also to the attention-driven emphasis on salient DAS vibration characteristics.

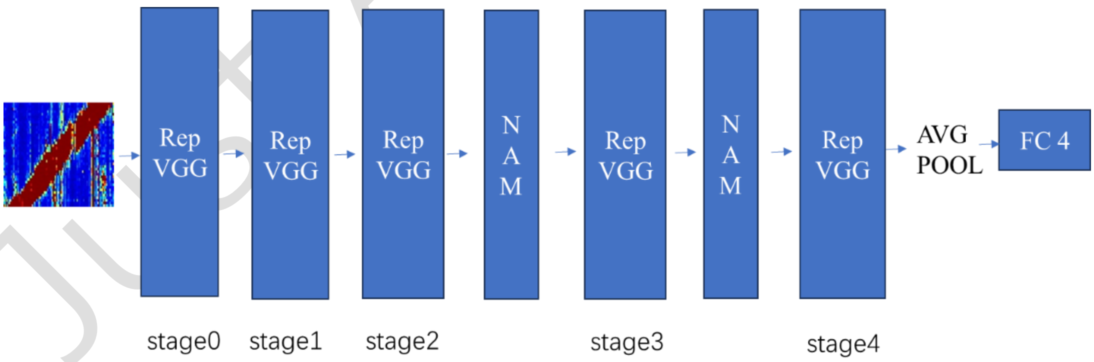

Fig. 5 Structure of RepVGG-NAM network

3 Experimental Analysis

All experiments were conducted on a workstation running Ubuntu 20.04. The hardware configuration included two Intel Xeon E5-2680 v3 CPUs, an NVIDIA RTX 3090 GPU with 24 GB memory, and 32 GB of system RAM. The deep learning models were implemented using the

PyTorch framework.

3.1 Evaluation Metrics

For multi-class recognition of train vibration signals, the confusion matrix characterizes the correspondence between predicted and ground-truth classes on the test set. To more comprehensively describe model performance, we compute statistics derived from the confusion matrix and report accuracy, precision, recall, and F1-score as evaluation metrics for the classification models.

3.2 Parameter Settings

To validate the effectiveness of the model, the dataset is randomly split into 80% for training and 20% for testing. The RepVGG-NAM model is trained on the 80% training portion and evaluated on the 20% test portion, with comparative experiments conducted against VGG16, ResNet18, and RepVGG. The initial learning rate is set to 0.05, and the weight decay to 0.005. Cross-entropy is used as the loss function, and the parameters are updated with stochastic gradient descent (SGD) to minimize the loss. The input image size is set to 64 × 64, the batch size to 128, and the number of epochs to 100.

3.3 Results and Analysis

After 100 training epochs, the curves of accuracy and loss are obtained as shown in Fig. 6. Confusion matrices for RepVGG-NAM, VGG16, ResNet18, and RepVGG are constructed to compare classification performance; the matrices on the validation set are presented in Fig. 7.

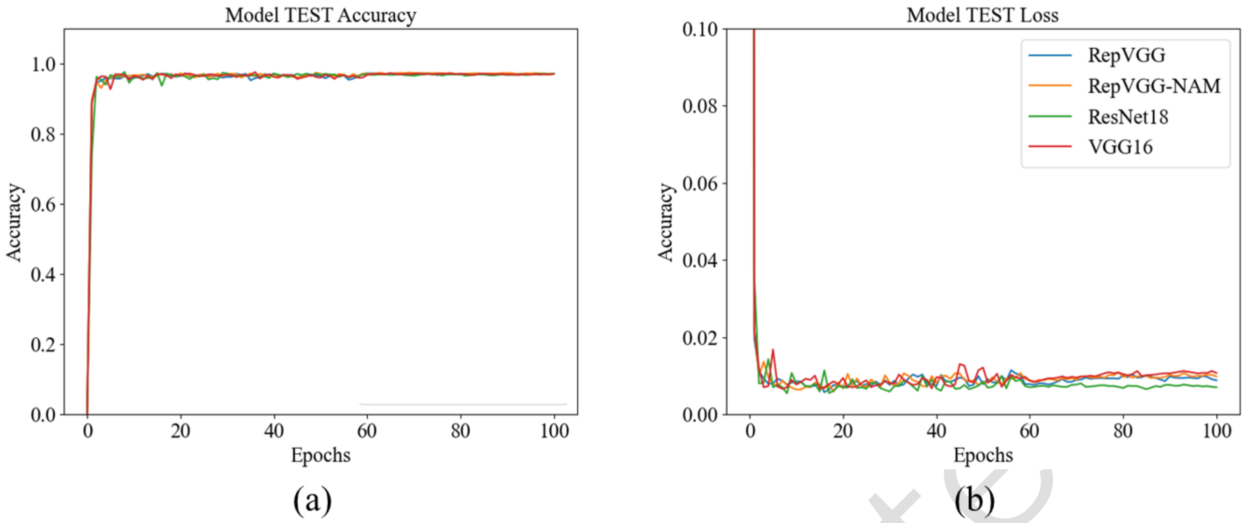

Fig. 6 (a) Classification accuracy curve and (b) loss curve of training.

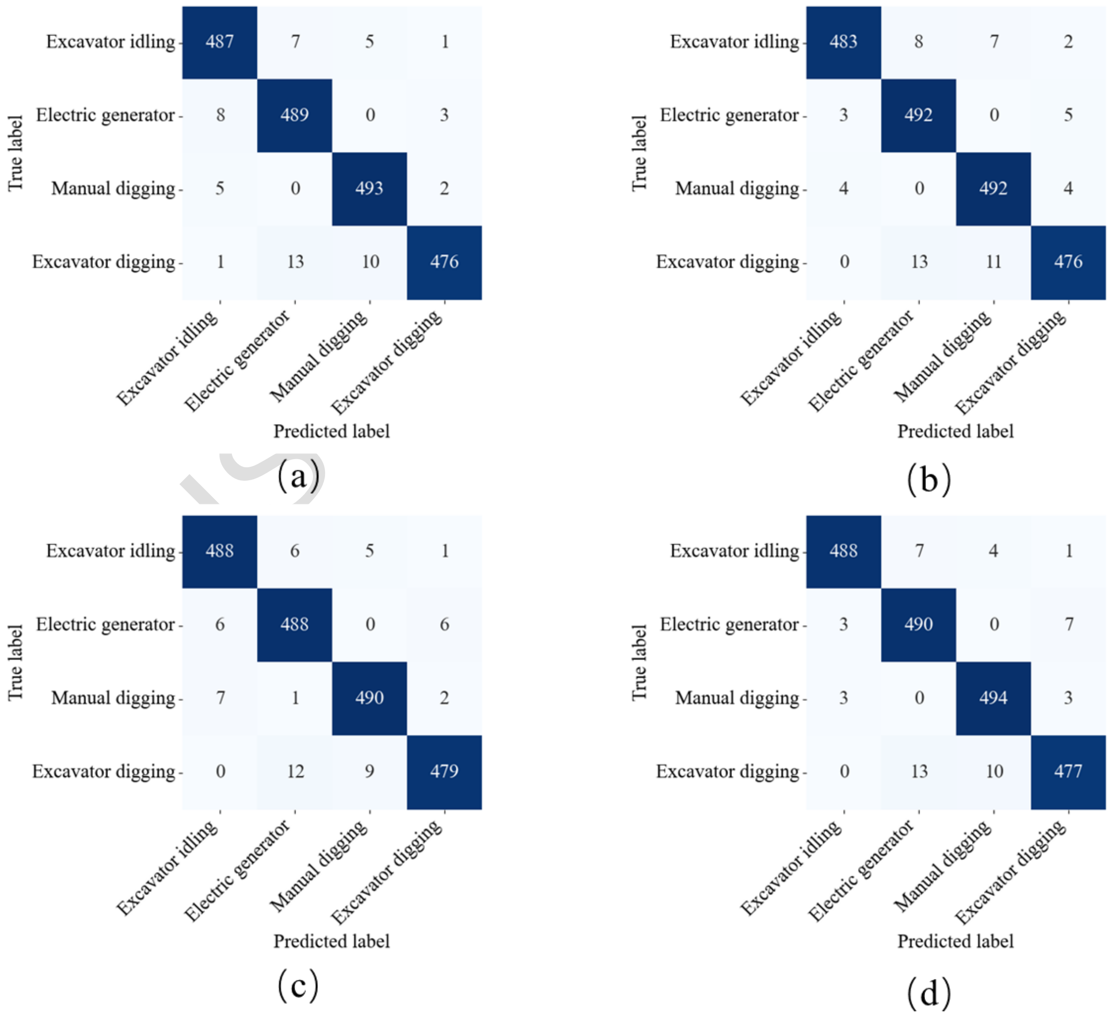

Fig. 7 Confusion matrix (a) VGG16 (b) ResNet18 (c) RepVGG (d) RepVGG-NAM

The performance of the four classification models was quantitatively evaluated based on their respective confusion matrices, from which standard evaluation metrics including Accuracy, Precision, Recall, and F1-Score were derived. The detailed metric values are presented in Table 2.

Table 2. Comparison of the experimental results

Method	Accuracy	Precision	Recall	F1-Score
VGG16	0.9725	0.9727	0.9725	0.9726
ResNet18	0.9715	0.9717	0.9715	0.9716
RepVGG	0.9725	0.9726	0.9725	0.9726
RepVGG-NAM	0.9745	0.9746	0.9745	0.9746

Based on the comparative classification results presented in Table 2, the RepVGG-NAM model, which is RepVGG combined with the lightweight NAM attention mechanism, achieves modest yet consistent improvements across all four evaluation metrics in the classification of train vibration signals acquired by DAS. Specifically, RepVGG-NAM attains an accuracy of 97.45%, representing a 0.20% absolute improvement over both vanilla RepVGG (97.25%) and VGG16 (97.25%), and surpassing ResNet18 (97.15%). Although the absolute gains are relatively small, the concurrent increases in all core indicators, namely precision (97.46%), recall (97.45%), and F1-score (97.46%), demonstrate that NAM, through adaptive channel-weight calibration, effectively enhances the model's focus on discriminative features, thereby improving robustness while maintaining high classification confidence.

From a more fine-grained perspective, the F1-score of RepVGG-NAM (97.46%) is 0.30% higher than that of ResNet18 (97.16%), and its balance between precision and recall is superior to both VGG16 and RepVGG. This reflects NAM's capability to suppress background-noise interference and strengthen salient feature extraction in DAS data. Furthermore, RepVGG-NAM preserves the architectural efficiency of RepVGG, maintaining the same parameter count while still achieving comprehensive improvements over conventional VGG16 and ResNet18. This finding validates the effectiveness of integrating structural re-parameterization with a lightweight attention mechanism in deep learning-based vibration classification. Although the overall performances of the models are closely aligned, with maximum differences within 0.3%, the stable performance of RepVGG-NAM on the test set suggests that dynamic enhancement of feature channels provides a more reliable and consistent solution for high-accuracy DAS-based classification tasks. These results highlight the potential of combining RepVGG's structural optimization with NAM's adaptive calibration to deliver both accuracy and efficiency in practical railway monitoring applications.

3.4 Structural Re-parameterization Analysis

To further validate the effectiveness of structural re-parameterization (Rep), we examined its impact on the RepVGG-NAM classifier in comparison with the baseline RepVGG. As illustrated in Fig. 8, the confusion matrices on the test set after applying structural re-parameterization are presented for both models. A detailed comparison with Fig. 7, which depicts the confusion matrices of the non-re-parameterized models, indicates that classification performance and error distributions remain unchanged. Both models exhibit highly consistent discriminative behavior before and after re-parameterization. This result provides strong evidence that structural re-parameterization does not alter the final decision-making process of the models. Instead, it preserves the classification accuracy and robustness of RepVGG-NAM while maintaining the lightweight inference advantages of RepVGG. In the context of DAS-based vibration monitoring, this property is particularly valuable, as it ensures that the integration of lightweight NAM attention mechanisms and structural optimization can enhance feature representation without compromising recognition reliability.

To further provide quantitative validation, we conducted a systematic comparison of the exact numerical outputs at the network's output layer before and after structural re-parameterization for RepVGG-NAM. As shown in Table 3, for identical input samples, the output tensors generated by the re-parameterized and the original models exhibit element-wise agreement to within four decimal places. This high level of numerical consistency offers conclusive evidence at the computational level, confirming that structural re-parameterization preserves functional equivalence without introducing deviations in the classification results.

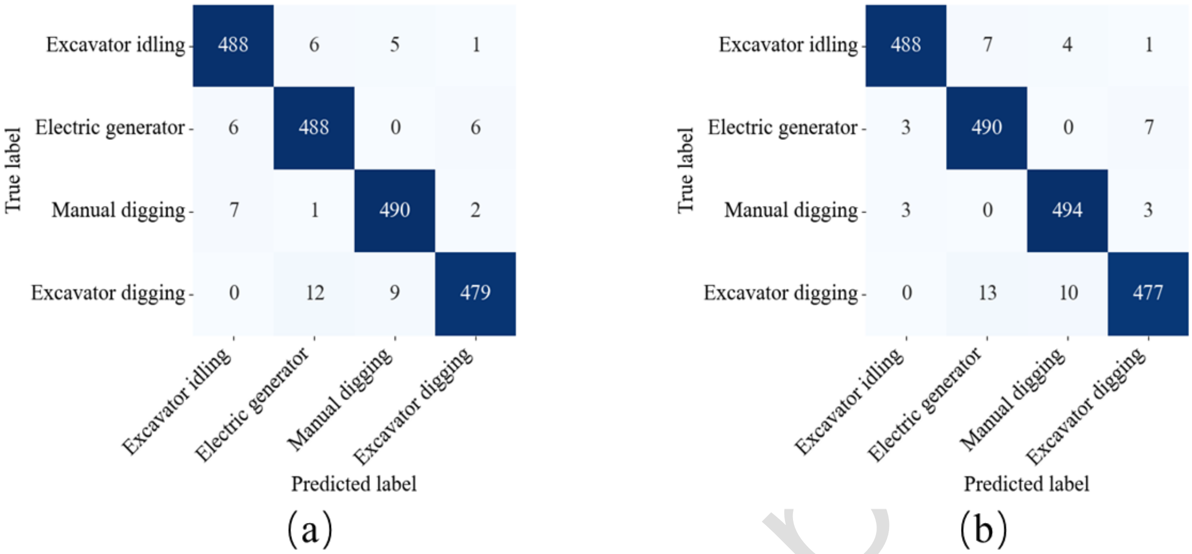

Fig. 8 Confusion matrix (a) RepVGG-RE (b) RepVGG-NAM-RE

Table 3. Comparison of the outputs of different methods.

Input	Method	Output			
Sample.1	RepVGG-NAM	7.9741	-1.059	-3.9332	-2.9121
	RepVGG-NAM-RE	7.9741	-1.059	-3.9332	-2.9121
Sample.2	RepVGG-NAM	-1.9911	8.8968	-3.3247	-3.6428
	RepVGG-NAM-RE	-1.9911	8.8968	-3.3247	-3.6428
Sample.3	RepVGG-NAM	-1.5548	-3.5806	8.4438	-3.4207
	RepVGG-NAM-RE	-1.5548	-3.5806	8.4438	-3.4207
Sample.4	RepVGG-NAM	-9.1626	-0.4659	-2.5147	11.995
	RepVGG-NAM-RE	-9.1626	-0.4659	-2.5147	11.995

Additionally, we evaluated the parameter counts, computational complexity, and inference speed of the different deep learning models, as summarized in Table 4. Benefiting from structural re-parameterization, the RepVGG family demonstrates substantial optimization at inference time. After incorporating the NAM, the re-parameterized model (RepVGG-NAM-RE) exhibits an increase in parameters to 13.3M, corresponding to a 60.0% rise compared with the baseline RepVGG. However, the computational cost increases only modestly by 18.2% (0.13 vs. 0.11 GFLOPs), remaining below that of ResNet18. This result highlights that the lightweight design of NAM effectively avoids the computational inflation typically associated with attention mechanisms. In terms of inference efficiency, RepVGG-NAM-RE maintains a throughput of 55,210 FPS, which, although reflecting a 27.3% reduction relative to vanilla RepVGG, still substantially outperforms VGG16 and ResNet18. These findings confirm the compatibility advantages of structural re-parameterization: the multi-branch NAM modules used during training can be seamlessly fused into standard convolutional operations during inference, ensuring no additional branches or latency. Taken together, the results validate that the integration of NAM into RepVGG not only improves feature discrimination but also preserves the high-throughput characteristics of RepVGG. This balance between accuracy and efficiency underscores the practical value of the proposed RepVGG-NAM framework for DAS-based train vibration classification, where both recognition reliability and real-time responsiveness are critical.

Table 4. Performance of algorithms.

Method	Params(M)	FLOPs(B)	Speed(img/s)
VGG16	138.35	1.38	7930
ResNet18	11.68	0.15	52015
RepVGG-RE	8.31	0.11	75910
RepVGG-NAM-RE	13.3	0.13	55210

4 Conclusion

This paper investigates the classification of train vibration signals acquired through DAS and proposes an efficient deep learning classifier, RepVGG-NAM. By embedding a lightweight Normalization-based Attention Module (NAM) into the RepVGG architecture, the model effectively enhances the representation of salient features through dynamic channel-weight calibration, thereby suppressing redundant information while strengthening discriminative cues.

The overall framework consists of preprocessing and coarse screening of raw vibration signals, followed by precise classification of four train-related signal categories using the improved RepVGG-NAM network. Experimental evaluations demonstrate that RepVGG-NAM achieves consistent performance improvements over VGG16, ResNet18, and baseline RepVGG. Specifically, the model attains an accuracy of 97.45%, precision of 97.46%, recall of 97.45%, and F1-score of 97.46%, with absolute gains of 0.20–0.30% across all four metrics. Although the numerical improvements appear modest, the uniform enhancement across all indicators validates the robustness of NAM in refining channel responses and optimizing feature utilization. In addition, RepVGG-NAM preserves the structural re-parameterization advantage of RepVGG, whereby the multi-branch training topology is converted into a streamlined single-branch inference architecture. This design ensures high recognition accuracy without incurring significant computational overhead, underscoring the practical feasibility of integrating lightweight attention with structurally optimized deep learning models. In summary, the proposed RepVGG-NAM provides a high-accuracy, low-overhead solution for DAS-based train vibration signal classification, offering valuable insights for the broader application of deep learning in intelligent railway monitoring and safety assurance. For future work, we will collect more along-line railway data to further improve the accuracy and generalization capability of the proposed model.

References:

[1] Jousset P, Currenti G, Schwarz B, et al. Fibre optic distributed acoustic sensing of volcanic events[J]. Nature Communications, 2022, 13: 1753.

[2] Han B, Guan H J, Yao J Z, et al. Distributed acoustic sensing with sensitivity-enhanced optical cable[J]. IEEE Sensors Journal, 2021, 21(4): 4644-4651.

[3] Zhang R, Liu Y. Zhang Z, Research on Real-time Safety Monitoring Methods for Railway Communication Optical Cables [J]. Railway signalling&Communication, 2019, 55(3).

[4] Diao D, Fault monitoring of Coal transportation system based on phase-OTDR[D]. ChongqingUniversity,2016.

[5] Zheng Y, Duan F J, Tu Q C. Signal processing of intrusion monitoring based on Φ-OTDR[J]. Acta Photonica Sinica, 2015, 44(1): 0106004.

[6] Li X Y, Wu H J, Peng Z P. A novel time sequence singular spectrum analysis method for Φ-OTDR disturbance detection system[J]. Acta Photonica Sinica, 2014, 43(4): 0428001.
[7] Lu X, An B, Chen Y, Li Y. Pattern recognition for fiber vibration system based on modern power spectral estimation. optical communication technology,2017, 41(5):59-62.
[8] Wang X, Liu Y, Liang S, et al. Event identification based on random forest classifier for phase-sensitive OTDR fiber-optic distri-buted disturbance sensor[J]. Infrared Physics & Technology, 2019, 97: 319-325.
[9] Sun Q, Li Q, Chen L, et al. Pattern recognition based on pulse scanning imaging and convolutional neural network for vibrational events in Φ-OTDR[J]. Optik, 2020, 219: 165205.
[10] Bu Z, Mao B, Si Z. Signal recognition of φ-OTDR system based on wavelet packet decomposition and SVM[J]. Acta Photonica Sin, 2022, 11: 87-95.
[11] Meng F, Zhang W, Liu X, et al. Comparative analysis of temporal-spatial and time-frequency features for pattern recognition of φ-OTDR[J]. Chinese Optics Letters, 2023, 21(4): 040601.
[12] Huang G, Liu Z, Van Der Maaten L, et al. Densely connected convolutional networks[C]//Proceedings of the IEEE conference on computer vision and pattern recognition. 2017: 4700-4708.
[13] Xie S, Girshick R, Dollár P, et al. Aggregated residual transformations for deep neural networks[C]//Proceedings of the IEEE conference on computer vision and pattern recognition. 2017: 1492-1500.
[14] Dosovitskiy A. An image is worth 16x16 words: Transformers for image recognition at scale[J]. arXiv preprint arXiv:2010.11929, 2020.
[15] Li D Q, Yi D, Zhou X H, et al. Multi-source threatening event recognition scheme targeting drone intrusion in the fiber optic DAS system[J]. IEEE Sensors Journal, 2024.
[16] Ding X, Zhang X, Ma N, et al. Repvgg: Making vgg-style convnets great again[C]//Proceedings of the IEEE/CVF conference on computer vision and pattern recognition. 2021: 13733-13742.
[17] Yin A, Lu M, Yang M, et al. Bearing fault diagnosis based on the combined use of RepVGG and CapsNet[J]. JOURNAL OF VIBRATION AND SHOCK ,2024,43(14):301-307.
[18] Liu Y, Shao Z, Teng Y, et al. NAM: Normalization-based attention module[J]. arXiv preprint arXiv:2111.12419, 2021.

[19] Sirisha U, Chandana B S, Harikiran J. Nam-yolov7: An improved yolov7 based on attention model for animal death detection[J]. Traitement du Signal, 2023, 40(2): 783-789.

[20] Hu Z, Tang Y, Yang S, et al. A new phase demodulation method for φ-OTDR based on coherent detection[C]//Optics Frontiers Online 2020: Distributed Optical Fiber Sensing Technology and Applications. SPIE, 2021, 11607: 124-131.

[21] He M, Qing W, Qu J. Defect Recognition in Ballastless Track Structures Based on Distributed Acoustic Sensors[J]. Applied Sciences, 2023, 13(17): 9663.

[22] Simonyan K, Zisserman A. Very deep convolutional networks for large-scale image recognition[J]. arXiv preprint arXiv:1409.1556, 2014.

[23] Li J, Yao R, Zhang J, et al. Pipeline threat event identification based on GAF of distributed fiber optic signals[J]. IEEE Sensors Journal, 2023, 23(21): 26796-26803.

[24] Szegedy C, Vanhoucke V, Ioffe S, et al. Rethinking the inception architecture for computer vision[C]//Proceedings of the IEEE conference on computer vision and pattern recognition. 2016: 2818-2826.

[25] He K, Zhang X, Ren S, et al. Deep residual learning for image recognition[C]//Proceedings of the IEEE conference on computer vision and pattern recognition. 2016: 770-778.

Funding：

Funded by Anhui Science and Technology Innovation Breakthrough Program (No. 202423i08050031)、Anhui Quality Infrastructure Standardization Special Program (No. 2023MKS22)、Anhui Postdoctoral Scientific Research Program Foundation (No. 2024C973)、The Open Project Program of Guangdong Provincial Key Laboratory of Intelligent Disaster Prevention and Emergency Technologies for Urban Lifeline Engineering (No. 2022ZA03)、Hefei Municipal Natural Science Foundation (No. 202331、202335)、Suzhou Major Science and Technology Special Program (No. SZKJXM202418)