Journal of Electronic Research and Application, 2026, Volume 10, Issue 1
http://ojs.bbwpublisher.com/index.php/JERA
ISSN Online: 2208-3510
ISSN Print: 2208-3502

A Small-Sample Bearing Fault Diagnosis Method Based on Multi-Image Fusion and Multi-Scale Dynamic Residual Dual Attention Mechanism

Ao Xu, Mu Li*

School of Information and Electrical Engineering, Hunan University of Science and Technology, Xiangtan 411100, China

**Corresponding author:* Mu Li, mli@hnust.edu.cn*

Abstract: In recent years, fault diagnosis methods based on convolutional neural networks (CNNs) have garnered significant attention in the field of rotating bearing fault diagnosis. Addressing the challenge of extremely limited fault signal samples, this paper proposes a small-sample bearing fault diagnosis method based on multi-image fusion and a dual-attention mechanism incorporating multi-scale dynamic residuals. This method first converts the fault signal into a two-dimensional image through continuous wavelet transform and Gram angle field (GASF/GADF), thereby transforming the fault diagnosis problem into an image feature learning problem. The model extracts basic features through the initial convolutional layer and sequentially learns deep features via multi-scale dynamic residual blocks and dual attention mechanisms. Among these, the multi-scale architecture captures features across different receptive fields through parallel convolutional branches, while the dual attention mechanism performs feature recalibration in both the channel and spatial dimensions. Experimental results demonstrate that the proposed method achieves an accuracy rate of 97.47% in bearing fault diagnosis tasks, representing a significant improvement over traditional CNN models. This validates the model's effectiveness and superiority in complex fault diagnosis scenarios.

Keywords: Fault diagnosis; Convolutional neural networks; Continuous wavelet transform; Gram angle field; Channel attention mechanism; Spatial attention mechanism; Rolling bearing fault; Multiscale residual blocks

Online publication: February 12, 2026

1. Introduction

As the core carrier of modern industrial production, mechanical equipment directly impacts a company's production efficiency and operational costs. Its health status has become crucial for ensuring the efficient operation of industrial systems. Rolling bearings, indispensable mechanical components within such equipment, support rotating shafts and their attached parts. Serving as the core elements of precision mechanical transmission systems, rolling bearings comprise a coordinated working system consisting of four major components: the inner ring, outer ring, rolling elements, and cage [1,2]. The inner ring is rigidly connected to the rotating shaft via an interference fit. Its precision-machined outer raceway forms a dynamic contact surface with the rolling elements, transmitting shaft system loads to the rolling elements. The outer ring provides support and constraint through the bearing housing. Its inner raceway forms a relative motion path with the inner ring, with load distribution optimized through geometric parameter adjustments. However, rolling bearings also face risks of various potential defects such as wear and fracture. Once these issues manifest, they severely impair normal equipment operation, not only causing significant economic losses but also posing serious threats to personnel safety [3,4].

Signals used for rolling bearing fault diagnosis include acoustic emission signals, vibration signals, temperature signals, and current signals, among others [5,6]. Tao *et al.* converted one-dimensional raw vibration signals into two-dimensional time-frequency maps via short-time Fourier transform as input for a classification generative adversarial network (GAN). Leveraging the unsupervised clustering capabilities of this GAN, they achieved rolling bearing fault diagnosis [7]. Han *et al.* proposed a fault diagnosis model based on the fusion of multi-level wavelet packets and a dynamic ensemble convolutional neural network (DECNN). A multi-level wavelet coefficient matrix was constructed via wavelet packet transform to comprehensively represent non-stationary vibration signals [8]. Xia *et al.* combined vibration signals collected from multiple sensors of the same type at different locations into a two-dimensional matrix, which served as input for a 2DCNN model to identify faults in rolling bearings and gearboxes [9]. Jiao *et al.* proposed a complementary data-driven deep coupled dense convolutional network (CDCN) model for planetary gearbox fault identification [10].

2. Dataset selection and preprocessing

This study utilizes the bearing data set from Case Western Reserve University in the United States as experimental data [11]. The experimental setup is illustrated in **Figure 1**. The test bench comprises a 2 HP (2 horsepower) motor, a torque sensor/encoder, a dynamometer, and control electronics (not shown in the figure).

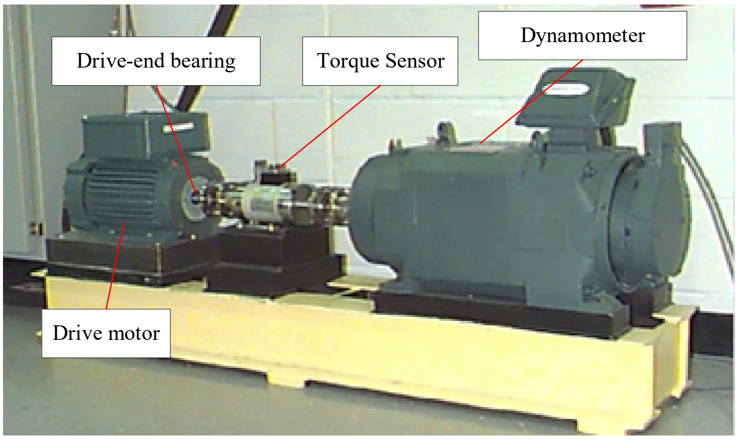

Figure 1. Experimental setup table.

3. Multi-image fusion

As shown in **Table 1** and **Figure 2**, in bearing fault diagnosis, faults cause signals to exhibit strong non-stationary characteristics and cross-scale energy distribution differences. CWT characterizes the temporal location of impact occurrence and its corresponding frequency band response through a combined time-frequency representation, highlighting local transient phenomena and multi-scale textures.

Table 1. System parameters

Label	Fault location	Image type	Label	Fault location	Image type
0	Normal	CWT	6	IR007	CWT
1	Normal	GADF	7	IR007	GADF
2	Normal	GASF	8	IR007	GASF
3	B007	CWT	9	OR007 @6	CWT
4	B007	GADF	10	OR007 @6	GADF
5	B007	GASF	11	OR007 @6	GASF

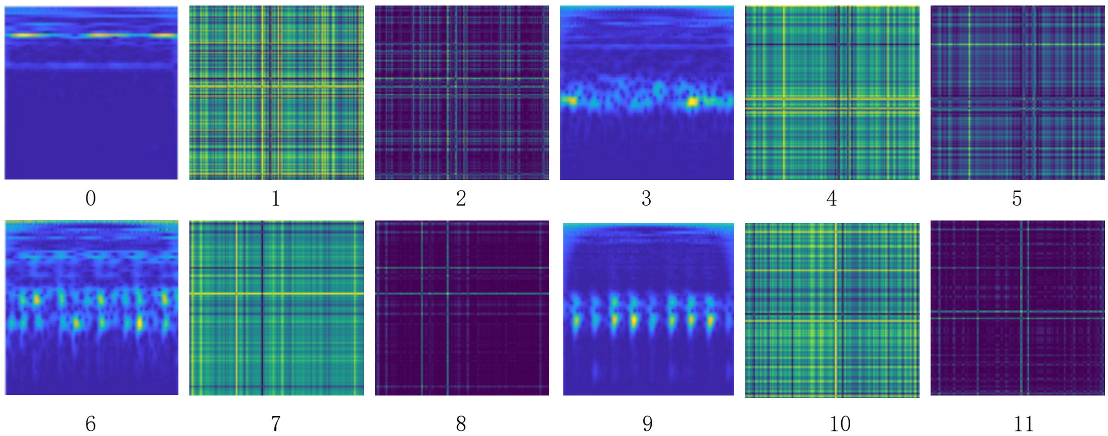

Figure 2. Multi-source images.

The wavelet energy map visually represents the energy distribution across different frequencies, where blue indicates low energy density, green and yellow denote moderate energy levels, and red signifies high energy content. The deeper the color, the greater the energy concentration. GASF depicts the consistency of shape/phase across two time points. If the amplitude variation trends of the sequence are similar at both moments, the corresponding GASF region will exhibit more consistent and continuous texture. GADF expression characterizes the direction and differences in changes at two time points, exhibiting greater sensitivity to mutations, impact edges, and upward/downward directions. Consequently, it tends to produce more pronounced texture contrasts near the impact location. This paper achieves bearing fault localization by transforming time-frequency domain signals into two-dimensional graphs via continuous wavelet transform and Gram angle field, which are then jointly fed as input to a 2DCNN.

4. Model construction

This paper proposes a small-sample bearing fault diagnosis method (MSDR-DAM) based on a dual attention mechanism that integrates multi-image fusion and multi-scale dynamic residuals. This model employs multi-image fusion, multi-scale analysis, dynamic residual learning, and a dual attention mechanism, wherein the dual attention mechanism comprises a channel attention mechanism and a spatial attention mechanism. This fault diagnosis process first transforms the raw fault signal into a time-frequency feature map via continuous wavelet transform and Gram angular field transform as model input. It then extracts basic features through an initial convolutional layer, followed by deep feature learning through multi-scale dynamic residual blocks and a dual self-attention mechanism. where the multi-scale architecture captures features across distinct receptive fields through parallel convolutional branches, while the dual attention mechanism enables feature recalibration at both the channel and spatial dimensions. Next, a feature pyramid is constructed through two rounds of downsampling, progressively expanding the receptive field while compressing spatial dimensions. Finally, global features are aggregated using global average pooling, and a fully connected classifier outputs fault category probability. The entire training process employs dynamic learning rate adjustment and gradient clipping optimization strategies to ensure model convergence stability and optimal diagnostic accuracy. The overall model framework is illustrated in **Figure 3**.

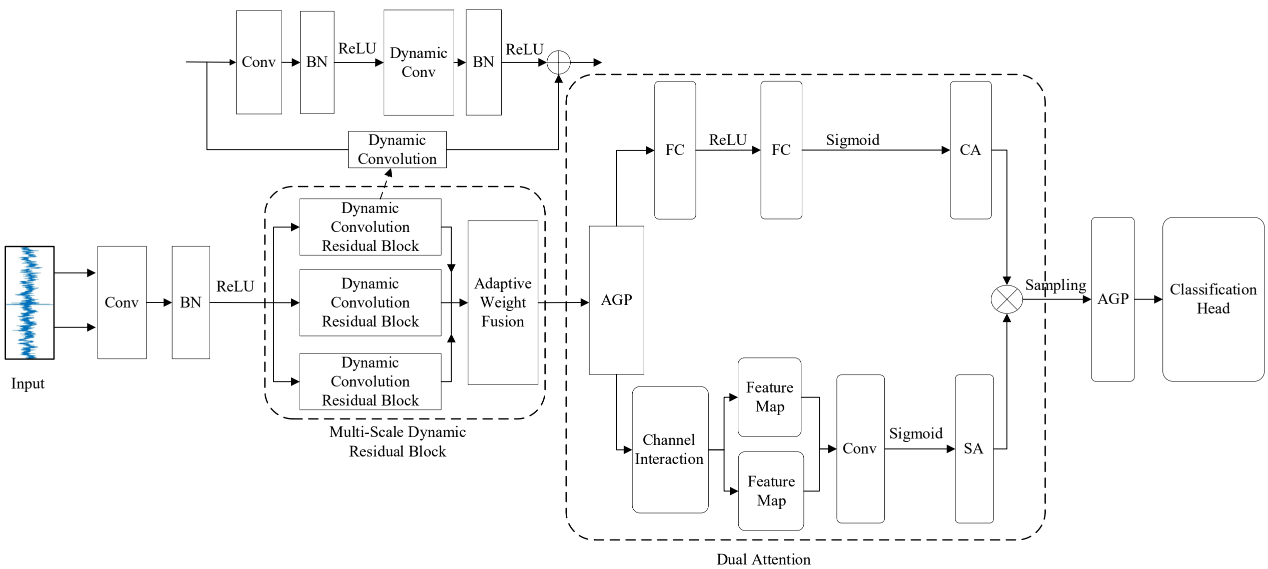

Figure 3. Overall framework of the model

4.1. Multi-scale dynamic residuals

The Multi-scale dynamic residual block is an innovative deep learning module whose core design philosophy lies in organically integrating parallel multi-scale feature extraction with a dynamic weight fusion mechanism. It processes input features through multiple parallel convolutional branches with distinct receptive fields and introduces an adaptive attention mechanism to dynamically adjust the contribution of each branch. Architecturally, the module comprises three parallel convolutional branches employing 1×1, 3×3, and 5×5 kernel sizes, respectively. The 1×1 convolution focuses on inter-channel information exchange and feature reorganization, the 3×3 convolution captures local spatial patterns, while the 5×5 convolution acquires a broader receptive field to capture global contextual information. The core innovation lies in its dynamic weight generation mechanism, mathematically expressed as follows:

$$W=Softmax(W_2 \cdot GAP(X))$$

where $W2$ is the weight matrix for the 1×1 convolution and GAP denotes global average pooling.

The implementation first acquires global spatial information from input features via adaptive average pooling, then compresses the channel dimension to the number of branches using 1×1 convolution to generate preliminary weights. Finally, the Softmax function ensures the sum of weights across branches equals 1, achieving a probabilistic distribution. During forward propagation, the feature fusion strategy can be expressed as follows:

$$Y = \Sigma wi \cdot Fi(X),$$

where wi is the dynamic weight of the i-th branch and Fi(·) denotes the transformation function of the i-th branch.

In the code implementation, the attention weight tensor is first generated. Then, the outputs of each branch are weighted by their respective weights, and finally, the weighted outputs of all branches are summed. The residual connection design adheres to the identity mapping principle, mathematically expressed as follows:

$$Z = \mathrm{ReLU}(Y + G(X)),$$

where G(·) is the shortcut connection function.

When input and output channel dimensions match, it acts as an identity mapping; otherwise, it adjusts dimensions via 1×1 convolution. This design ensures both smooth gradient flow and the integrity of feature information.

4.2. Dual attention mechanism

This model employs a dual attention mechanism comprising two parallel branches: channel attention and spatial attention. The channel attention branch first compresses the spatial dimension through a global average pooling layer to extract channel-level statistical features. Subsequently, it utilizes a bottleneck structure composed of two fully connected layers to facilitate information exchange between channels. It generates channel attention weights via a sigmoid activation function, enabling adaptive recalibration across different feature channels. Simultaneously, the spatial attention branch generates two spatial feature maps by computing the mean and maximum values along the channel dimension. These maps are concatenated along the channel dimension and passed through a 7×7 convolutional layer to capture large-scale spatial dependencies. The sigmoid function then produces the spatial attention weight map. Ultimately, the channel attention output is multiplied position-wise with the spatial weights, enabling simultaneous fine-tuning of feature maps across both channel and spatial dimensions. This allows the model to adaptively focus on the feature channels and critical regions most relevant to fault diagnosis, significantly enhancing the discriminative power of feature representations.

4.3. Fault identification process

Task identification based on MSDR-DAM specifically involves the following three steps:

(1) Step 1: Collect bearing signals as raw diagnostic data, construct a dataset encompassing normal operation and various fault conditions, and employ continuous wavelet transform and Gram angular field to generate two-dimensional maps for image fusion;

(2) Step 2: Input the preprocessed data into the MSDR-DAM model for training. Once the model achieves stable accuracy that meets the criteria, validate its precision using the test set data and save the model's optimal hyperparameters and weights;

(3) Step 3: Output model diagnostic logs and reports.

5. Experimental verification and analysis

5.1. Experimental platform

To validate the diagnostic effectiveness of the proposed model in this paper, the experimental platform employs an Intel Core i5-14600KF processor, an NVIDIA GeForce RTX 5060 Ti graphics card, and 32 GB of memory. The software environment is based on the CUDA 12.8 acceleration library, with the deep learning framework utilizing the PyTorch programming language on Python 3.9.

5.2. The experimental results

To validate the feasibility of the feature extraction method and diagnostic model under small sample conditions, this experiment utilizes the data from Section 2 for bearing fault diagnosis testing. By inputting data into the MSDR-DAM model, this paper employs a confusion matrix to quantitatively analyze diagnostic outcomes. The horizontal axis represents the diagnostic model's predicted results, while the vertical axis denotes the actual diagnostic results. Elements on the main diagonal indicate the number of samples where predictions match actual outcomes, whereas off-diagonal elements represent misclassified samples. The test diagnostic confusion matrix is shown in **Figure 4**. Samples under different fault categories can be correctly classified, with few misclassified samples, high classification accuracy, and strong generalization capability. Throughout the process, no complex coordinate transformations, expert prior knowledge, or manual debugging experience are required to successfully diagnose the location of the fault. This demonstrates that the method proposed in this paper has certain feasibility for small-sample bearing fault diagnosis.

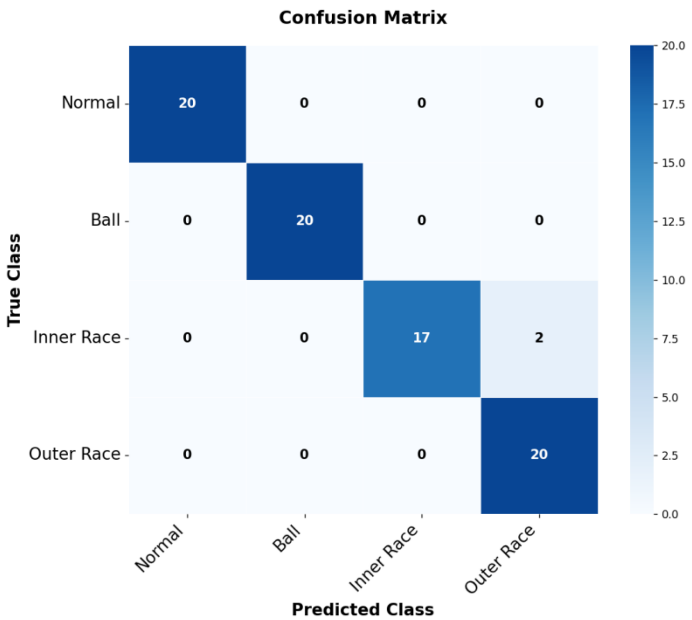

Figure 4. Confusion matrix.

As shown in **Table 2**, the proposed model achieves an accuracy of 98.43% in IGBT localization, demonstrating exceptionally high diagnostic precision. Furthermore, the high values of precision, recall, and F1 score confirm the model's capability in identifying samples.

Table 2. Comparative experiment

Diagnosis method	Accuracy	Precision	Recall	F1 score
1D-CNN	91.03%	92.50%	90.56%	90.69%
2D-CNN	95.00%	95.49%	95.00%	94.89%
Proposed Method	97.47%	97.73%	97.37%	97.42%

Table 2 demonstrates that the proposed model exhibits significant advantages across multiple key performance metrics. Compared to the 1D-CNN model, our approach enhances diagnostic accuracy by converting fault signals into two-dimensional data. Compared to the 2D-CNN model, this paper introduces multi-image fusion and an attention mechanism. This mechanism guides the network to more effectively capture discriminative features related to faults by simultaneously computing channel attention and spatial attention, thereby enhancing the diagnostic performance and robustness of the convolutional network. This model fusion approach enables the MSDR-DAM diagnostic model to excel in accuracy, efficiency, stability, and resource consumption for small-sample fault diagnosis in bearings.

6. Conclusion

This paper proposes a small-sample bearing fault diagnosis method based on multi-image fusion and a dual-attention mechanism for multi-scale dynamic residuals. The multi-scale dynamic residual module adaptively fuses fault features with varying receptive fields to capture multi-scale patterns ranging from transient impacts to steady-state modulations. The dual-attention mechanism for both channels and spatial dimensions simultaneously calibrates feature responses along these dimensions, precisely focusing on fault-sensitive regions. To validate performance, the dataset from the Bearing Data Center at Case Western Reserve University was employed for experiments. Results demonstrate that the proposed method achieves high-accuracy diagnosis even under minimal sample conditions, with bearing fault localization accuracy reaching 97.47%, significantly outperforming traditional models such as 1D-CNN and 2D-CNN.

Funding

The National Natural Science Foundation of China (Project No.: 61404049, 62273141); Natural Science Foundation of Hunan Province (Project No.: 2020JJ6031)

Disclosure statement

The author declares no conflict of interest.

References

[1] Zhang T, Liu S, Zhang S, 2021, Review on Fault Diagnosis on the Rolling Bearing. Journal of Physics: Conference Series. IOP Publishing, 1820(1): 012107.

[2] Zhang S, Ye F, Wang B, et al., 2019, Semi-Supervised Learning of Bearing Anomaly Detection via Deep Variational

Autoencoders. arXiv, https://doi.org/10.48550/arXiv.1912.01096

[3] Zhang C, Guo J, Zhen D, et al., 2020, Rolling Element Bearing Fault Diagnosis based on the Wavelet Packet Transform and Time-Delay Correlation Demodulation Analysis. Advances in Asset Management and Condition Monitoring: COMADEM 2019, 1195–1203.

[4] Xue B, Yi W, Jing F, et al., 2021, Complex ISAR Target Recognition using Deep Adaptive Learning. Engineering Applications of Artificial Intelligence, 2021(97): 1–9.

[5] Wong P, Zhong J, Yang Z, et al., 2016, A New Framework for Intelligent Simultaneous-Fault Diagnosis of Rotating Machinery Using Pairwise-Coupled Sparse Bayesian Extreme Learning Committee Machine. Proceedings of The Institution of Mechanical Engineers, Part C: Journal of Mechanical Engineering Science, 231(6): 1146–1161.

[6] Safizadeh M, Yari B, 2017, Pump Cavitation Detection Through Fusion of Support Vector Machine Classifier Data Associated with Vibration and Motor Current Signature. Insight, 2017(59): 669–673.

[7] Tao H, Wang P, Chen Y, et al., 2020, An Unsupervised Fault Diagnosis Method for Rolling Bearing Using STFT and Generative Neural Networks. Journal of the Franklin Institute, 357(11): 7286–7307.

[8] Han Y, Tang B, Deng L, 2018, Multi-Level Wavelet Packet Fusion in Dynamic Ensemble Convolutional Neural Network for Fault Diagnosis. Measurement, 2018(127): 246–255.

[9] Xia M, Li T, Xu L, et al., 2017, Fault Diagnosis for Rotating Machinery Using Multiple Sensors and Convolutional Neural Networks. IEEE/ASME Transactions on Mechatronics, 2017(99): 1.

[10] Jiao J, Zhao M, Lin J, et al., 2019, Deep Coupled Dense Convolutional Network with Complementary Data for Intelligent Fault Diagnosis. IEEE Transactions on Industrial Electronics, 2019: 1.

[11] Smith W, Randall R, 2015, Rolling Element Bearing Diagnostics using the Case Western Reserve University Data: A Benchmark Study. Mechanical Systems and Signal Processing, 2015(64): 100–131.