IEEE Access
Multidisciplinary ⋮ Rapid Review ⋮ Open Access Journal

Received 4 August 2023, accepted 24 September 2023, date of publication 27 September 2023, date of current version 4 October 2023.

Digital Object Identifier 10.1109/ACCESS.2023.3320065

RESEARCH ARTICLE

A Rotating Machinery Fault Diagnosis Method Based on Multi-Sensor Fusion and ECA-CNN

HONGXING WANG, HUA ZHU, AND HUAFENG LI

State Key Laboratory of Mechanics and Control for Aerospace Structures, Nanjing University of Aeronautics and Astronautics, Nanjing 210016, China

Corresponding authors: Hua Zhu (hzhu103@nuaa.edu.cn) and Huafeng Li (lihuaf@nuaa.edu.cn)

This work was supported by the National Natural Science Foundation of China under Grant 52277055.

ABSTRACT Fault diagnosis is critical to maintaining the performance of rotating machinery and ensuring the safe operation of the equipment. Convolutional neural networks (CNNs) have recently shown great potential with excellent automatic feature learning and nonlinear mapping abilities in the field of rotating machinery fault diagnosis. However, the CNN-based methods still suffer from some defects, such as inadequate data utilization and uneconomical computational efficiency, which limit the further improvement of diagnosis performance. Therefore, this paper proposes a fault diagnosis method based on multi-sensor fusion and Convolutional Neural Network with Efficient Channel Attention (ECA-CNN). First, multi-sensor vibration signals are sampled, converted, and channel fused into multi-channel images with rich and comprehensive features. Then, the efficient channel attention mechanism is introduced into CNN to increase the feature learning ability by adaptively scoring and assigning weights to the channel features. The ECA-CNN is proposed to learn representative fault features from multi-sensor fusion data to achieve fault identification. Finally, two experimental cases on the bearing and gearbox datasets prove that the proposed method has excellent performance, strong generalization capability, and high computational efficiency.

INDEX TERMS Fault diagnosis, multi-sensor fusion, convolutional neural networks, channel attention mechanism.

I. INTRODUCTION

With the vigorous development of modern industrial systems, rotating machinery plays an irreplaceable role in intelligent equipment, and its safety has received wide attention from academia and industry [1]. The key transmission components of rotating machinery (e.g., bearings and gears) are inevitably subject to wear, corrosion, deformation, cracks, and other failures under complex operating environments such as heavy loads and high speeds [2]. Faults of transmission components directly affect the operational reliability of rotating machinery and may cause significant accidents, resulting in substantial economic losses and even casualties [3]. Therefore, it is of great research value to carry out fault diagnosis and predictive maintenance of rotating machinery [4].

In recent years, with strong fault features learning ability (without the hand-feature design) and end-to-end diagnosis characteristics (mapping the original data to machine health states), deep learning (DL) methods have gradually become a research hotspot in the field of rotating machinery fault diagnosis [5], [6]. DL methods have only two steps. The first step is to sample the original data to obtain a set containing samples and labels. The second step is to classify the samples using DL models integrating feature design, extraction, and classification [7], [8].

For the first step of DL, with the rapid development of sensing and communication technologies, lots of low-cost vibration sensors can be deployed at different positions and directions of rotating machinery to monitor a large amount of operating state data of the equipment [9], [10]. Vibration data are usually waveform signals (such as acceleration and acoustic signals), and the features of multi-sensor signals vary considerably [11], [12]. DL models require a large amount of data containing rich features to be involved in training, so multi-sensor signals not only bring new life to DL-based fault diagnosis but also present many application challenges, such as data idleness and data misuse, and so on [13], [14].

The associate editor coordinating the review of this manuscript and approving it for publication was Mehrdad Saif.

© 2023 The Authors. This work is licensed under a Creative Commons Attribution-NonCommercial-NoDerivatives 4.0 License.
For more information, see https://creativecommons.org/licenses/by-nc-nd/4.0/

For the second step of DL, a variety of DL models, such as multi-layer perception (MLP), deep belief network (DBN), auto-encoder (AE), recurrent neural network (RNN), and convolutional neural network (CNN), have been extensively studied and applied to fault diagnosis of rotating machinery [15], [16]. For instance, Liu et al. [17] proposed an improved auto-encoder based on recurrent neural networks to achieve fault diagnosis of bearings. Qin et al. [18] proposed an optimized deep belief network and applied it to fault diagnosis for planetary gearboxes of wind turbines. Zhang et al. [19] proposed a rotating machinery fault diagnosis method based on gated recurrent unit recurrent neural network. In particular, CNNs are more suitable for processing periodic signals and learning features from mechanical vibration signals [20]. They stand out among these algorithms and perform better in various fault diagnosis tasks [21], [22]. For example, Zhang et al. [23] proposed a method named deep convolutional neural networks with wide first-layer kernels (WDCNN) for fault diagnosis, which has good anti-noise and domain adaptation ability on raw vibration signals. Liu et al. [24] proposed a lightweight multi-task CNN for fault diagnosis, which achieved good diagnosis results on the wheelset bearing dataset. Zhao et al. [25] proposed an efficient and lightweight fault diagnosis model based on CNN with mixed information (MIXCNN), which performed well on multiple rotating machinery datasets.

According to the above study, although CNN-based fault diagnosis methods have performed well in many tasks, it still suffers from the following shortcomings.

(1) Most existing methods suffer from inadequate data utilization. They only utilize signals from a single sensor or position for fault diagnosis, ignoring the differences in fault features between multi-sensor signals [26]. This prevents the model from thoroughly learning representative fault features, limiting the further improvement of diagnosis performance [27].
(2) In CNN, each convolutional kernel performs independent convolutional operations on corresponding input channels and produces feature maps (output channels) [28]. This means that CNN only processes different channels independently, ignoring the importance of different channel features in the diagnosis results [29].
(3) To maximize performance, most newly developed networks usually require complex structures and too many parameters to obtain strong feature learning capability, which leads to low computational efficiency and waste of computational resources [30].

To overcome the above drawbacks, this paper proposes a fault diagnosis method based on multi-sensor fusion and ECA-CNN. The method has three primary purposes: (1) Fully exploit multi-sensor data. (2) Increase the feature learning ability of CNN. (3) Improve computational efficiency. The main contributions of this paper are summarized as follows:

(1) The channel fusion of multi-sensor vibration signals is realized, which makes the multi-channel fusion data have rich and comprehensive abnormal features.
(2) The efficient channel Attention mechanism is introduced into CNN to increase the feature learning ability and focus attention on the more sensitive features.
(3) The ECA-CNN with low computational complexity is proposed to learn representative fault features from multi-sensor fusion data to achieve fault identification.
(4) Two experimental cases on the bearing and gearbox datasets verify the effectiveness and generalization of the proposed method.

The remaining content of this paper is organized as follows. Section II describes the proposed method in detail. Section III presents the experimental results and analysis of two datasets. Section IV analyzes the visualization of three aspects of the proposed method. Finally, Section V summarizes this paper and concludes future work.

II. PROPOSED METHOD

In this section, a fault diagnosis method based on multi-sensor fusion and ECA-CNN is developed, which consists of three parts: multi-sensor data preprocessing, the proposed ECA-CNN, and overall framework of the proposed method.

A. MULTI-SENSOR DATA PREPROCESSING

1) SAMPLING OF MULTI-SENSOR SIGNALS

Suppose the components of the rotating machinery have c health states, including one normal state and several fault states. Under certain working conditions, the monitoring data from m sensors of rotating machinery is described as the following formula,

$$\{X, y\} = \{(X_i, y_i)\}, \quad i \in [1, c] \tag{1}$$

$$\begin{cases} X_i = \left(X_i^1, X_i^2, \ldots, X_i^m\right) \\ y_i \in \{label\ 1, label\ 2, \ldots, label\ c\} \end{cases} \tag{2}$$

$$X_i^j = \left[X_i^j(1), X_i^j(2), \ldots, X_i^j(N)\right], \quad j \in [1, m] \tag{3}$$

where $\{X, y\}$ denotes the monitoring signals and labels in c health states. X_i and y_i are the signals and labels under state i. X_i^j is the signal of the j-th sensor under state i, $X_i^j(N)$ is the N-th signal point of X_i^j.

Insufficient training samples in data-driven intelligent fault diagnosis likely cause model overfitting [5]. Therefore, data augmentation techniques are usually used to alleviate the problem, and the sliding window sampling is the most commonly used method [15]. As shown in FIGURE 1, the monitoring multi-sensor signals are sampled according to the sample length L and the sliding stride S, and the signal sample set obtained is defined as follows,

$$\{x, y\} = \{(x_i, y_i)\}, \quad i \in [1, c] \tag{4}$$

$$x_i = \left(x_i^1, x_i^2, \ldots, x_i^m\right) \tag{5}$$

$$x_i^j = \left(x_{i,1}^j, x_{i,2}^j, \ldots, x_{i,k-1}^j, x_{i,k}^j\right), \quad j \in [1, m], \ k \in N^* \tag{6}$$

$$x_{i,k}^j = \left[X_i^j((k-1) \times S), \ldots, X_i^j((k-1) \times S + L)\right] \tag{7}$$

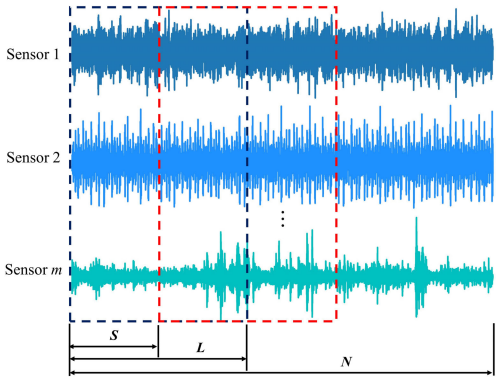


FIGURE 1. Sampling process of multi-sensor signals.

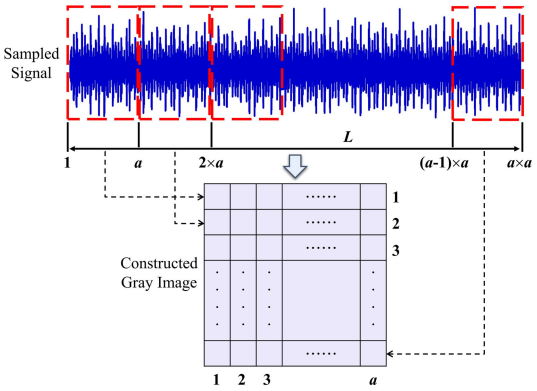


FIGURE 2. Conversion process of signal-to-image.

where $\{x, y\}$ represents the signal samples and labels in c health states. x_i and y_i are the signal samples and labels under state i. x_i^j is the signal samples of the j-th sensor under state i, $x_{i,k}^j$ is the k-th sample of x_i^j.

2) SIGNAL-TO-IMAGE CONSTRUCTION

This paper adopts a simple and effective conversion method to convert the 1D signals into 2D images [31]. The conversion method provides a way to directly explore the 2D features of 1D signals [19], and 2D convolution can simultaneously extract the features of different periods in the time-domain signals, breaking the spatial limitation of 1D convolution [32]. FIGURE 2 shows the conversion process of signal-to-image, where a consecutive signal segment of length L sequentially fills the rows of the constructed image. The pixel values of the constructed image are defined as follows,

$$x_{i,k}^j(p)_{img} = round\left\{\frac{x_{i,k}^j(p) - \min\left(x_{i,k}^j\right)}{\max\left(x_{i,k}^j\right) - \min\left(x_{i,k}^j\right)}\right\} \times 255, \quad p = 1, 2, \ldots, L \tag{8}$$

where $x_{i,k}^j(p)$ is the p-th points of $x_{i,k}^j$. $max(\cdot)$ and $min(\cdot)$ represent the maximum and minimum functions, respectively. $round(\cdot)$ represents the rounding function. The pixel values of the constructed image are normalized to an integer ranging between 0 to 255, corresponding to the pixel intensity of the gray image. In this paper, the sample length L is set to the commonly used 1024, which means the edge length a of the constructed image is 32.

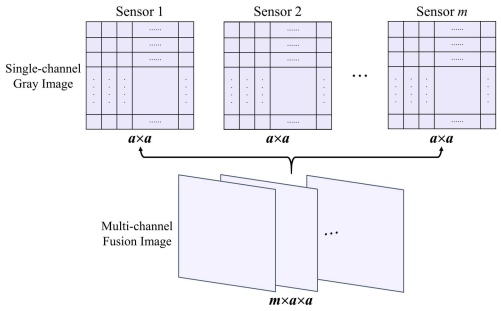


FIGURE 3. Channel fusion process of multi-sensor images.

3) CHANNEL FUSION OF MULTI-SENSOR IMAGES

In image recognition and detection tasks, RGB trichromatic channels are usually used to describe image characteristics comprehensively [33]. Thus, fusing the images constructed from multi-sensor signals into multi-channel images to express features fully is logical. As shown in FIGURE 3, the single-channel images constructed from each sensor are fused into a multi-channel image in the channel dimension. The multi-channel fusion image sample set is described as the following formula,

$$\{x, y\}^{fusion} = \left\{\left(x_i^{fusion}, y_i\right)\right\}, \quad i \in [1, c] \tag{9}$$

$$x_i^{fusion} = \left(x_{i,1}^{fusion}, x_{i,2}^{fusion}, \ldots, x_{i,k}^{fusion}\right), \quad k \in N^* \tag{10}$$

$$x_{i,k}^{fusion} = \left[x_{i,k}^1, x_{i,k}^2, \ldots, x_{i,k}^m\right] \tag{11}$$

where $\{x, y\}^{fusion}$ denotes the multi-channel image samples and labels in c health states. x_i^{fusion} and y_i are the multi-channel image samples and labels under state i. $x_{i,k}^{fusion}$ is the k-th sample of x_i^{fusion}. $x_{i,k}^m$ represents the k-th single-channel image of the m-th sensor under state i.

B. THE PROPOSED ECA-CNN

1) CONVOLUTIONAL NEURAL NETWORK

CNN is a widely used feed-forward DL model. It utilizes the spatial weight shared to solve the overfitting problem that the fully connected (FC) networks are prone to [20]. It mainly comprises convolutional, activation, and pooling layers, which learn useful local features by alternately stacking convolutional and pooling operations [22]. This paper focuses on 2D CNN since the input is multi-channel fusion images.

Convolutional layer: In the convolutional layer, multiple convolutional kernels convolve with the input features to obtain the output features. Each convolution kernel has the same size, and local features are extracted by sliding. The number of convolution kernels corresponds to the number of

output features. The formula is defined as follows,

$$h_j^l = \sum_i x_i^{l-1} * w_{ij}^l + b_j^l \tag{12}$$

where $*$ represents the convolution operation, x_i^{l-1} is the i-th feature map of layer $l-1$, w_{ij}^l is the weight of the j-th convolutional kernel of l-th layer, b_j^l is the j-th bias of l-th layer, h_j^l is the j-th output of l-th layer.

Batch normalization (BN): BN can achieve a stable distribution of activation values during the training process, reduce internal covariance shifts, and avoid gradient explosion and disappearance, all of which facilitate learning representative features in the potential space by each layer [34]. The BN layer usually comes after the convolutional layer and before the activation layer. Suppose $h^l = \{h^{l(1)}, \ldots, h^{l(N)}\}$ is the input to the l-th BN layer with mini-batch size N, and $h^{l(n)} = \{h_1^{l(n)}, \ldots, h_k^{l(n)}\}$. The formula for the BN operation is as follows,

$$\widehat{h}_j^{l(n)} = \frac{h_j^{l(n)} - \mu_j}{\sqrt{\sigma_j^2 + \varepsilon}} \tag{13}$$

$$y_j^{l(n)} = \gamma_j^l \widehat{h}_j^{l(n)} + \beta_j^l \tag{14}$$

$$\mu_j = \frac{1}{N} \sum_n h_j^{l(n)} \tag{15}$$

$$\sigma_j^2 = \frac{1}{N} \sum_n \left(h_j^{l(n)} - \mu_j\right)^2 \tag{16}$$

where $y_j^{l(n)}$ is the output of the BN layer of the convolution result of the j-th kernel. u_j and σ_j^2 are the mean and variance of h_j^l, respectively. ε is a tiny constant to prevent the invalid calculation when the variance is 0. γ_j^l and β_j^l are the scale and shift parameters to be learned, respectively.

Activation layer: The activation function can add nonlinearity to make the features learned in the convolution layer easier to distinguish. The ReLU function can enhance the sparsity of the network to alleviate the overfitting problem and accelerate the convergence of the network. Therefore, it is widely used in the activation layer of various models. The formula is described as follows,

$$\mathrm{ReLU}(x) = \begin{cases} x, & x \geq 0 \\ 0, & x < 0 \end{cases} = \max(0, x) \tag{17}$$

Pooling layer: The pooling layer is a down-sampling operation that extracts crucial local information from the feature map. It can reduce the dimension of the feature space, thus effectively controlling the risk of overfitting. The maximum pooling layer is the most commonly used, which performs local maximum operations on the input features. The maximum pooling operation is defined as follows,

$$z_j^l = max\left(y_j^l; s\right) \tag{18}$$

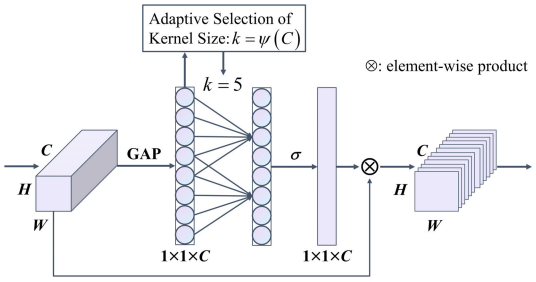


FIGURE 4. Diagram of ECA module.

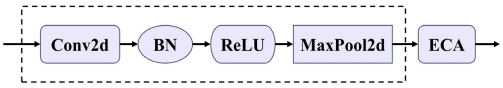


FIGURE 5. The structure of CNN-ECA block.

where $max(\cdot)$ denotes the down-sampling function, y_j^l is the j-th input feature map of l-th layer, s is the pooling size, z_j^l is the output feature map of the maximum pooling layer.

2) EFFICIENT CHANNEL ATTENTION MECHANISM

The convolution kernel can adaptively extract the features of the input during the convolution process and treat them as the output channels. However, convolutional operation ignores the importance of different channel features in the prediction results [28]. Channel attention mechanisms can automatically learn global information to obtain the importance of each channel feature, and it has been shown to have great potential in improving the feature learning ability of CNNs models [35]. In this paper, an efficient channel attention (ECA) module [36] is introduced to the feature learning of multi-sensor fusion data, and FIGURE 4 shows the diagram of the ECA module.

Previous channel attention mechanisms usually have high computational complexity [35]. The ECA module improves computational efficiency by introducing lightweight 1D convolution operation and proposes an adaptive selection of cross-channel size strategy. Suppose $X \in R^{C \times H \times W}$ is the feature map after convolution, and C, H, and W are the number of channels, height, and width of the feature map, respectively. First, X performs the global average pooling (GAP) operation to obtain aggregated feature $g(X)$ for each channel. Then, C is used to adaptively calculate the number of cross-channels k. Finally, the weight w of each channel is calculated using a 1D convolution with kernel size k and a Sigmoid function and assigned to the raw feature map. The specific equations are as follows,

$$g(X) = \frac{1}{W} \sum_{i=1}^{W} X_i \tag{19}$$

$$k = \psi(C) = \left| \frac{\log_2(C)}{\gamma} + \frac{b}{\gamma} \right|_{odd} \tag{20}$$

$$\omega = \sigma\left(C1D_k\left(g(X)\right)\right) \tag{21}$$

where $g(\cdot)$ denotes the GAP function. $\psi(\cdot)$ is the function of adaptively calculating the number of cross-channels k.

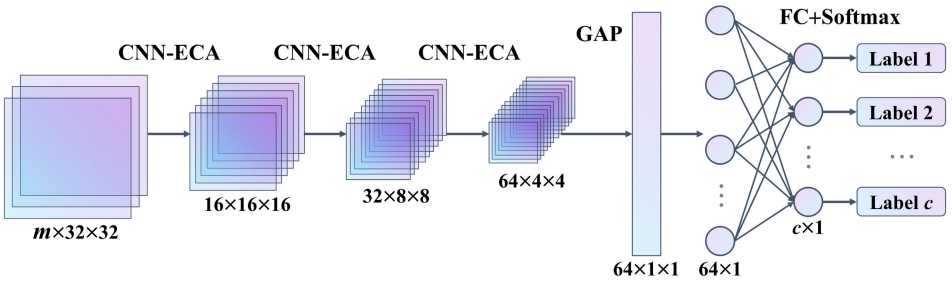


FIGURE 6. The architecture of ECA-CNN.

TABLE 1. The parameter configuration of ECA-CNN.

Layer	Type	Conv2d	Maxpool2d	ECA block	Output
1	Input Layer	-	-	-	$m \times 32 \times 32$
2	CNN-ECA	(16, 3 × 3, 1, 1)	2 × 2	16	16 × 16 × 16
3	CNN-ECA	(32, 3 × 3, 1, 1)	2 × 2	32	32 × 8 × 8
4	CNN-ECA	(64, 3 × 3, 1, 1)	2 × 2	64	64 × 4 × 4
5	GAP	-	-	-	64
6	FC + Softmax	-	-	-	c

$C1D_k(\cdot)$ indicates 1D convolution operation. $\sigma(\cdot)$ represents the Sigmoid function. b and γ are constants, in this paper $b = 1, \gamma = 2$ [36].

3) ARCHITECTURE OF THE PROPOSED ECA-CNN

This paper constructs a CNN-ECA block consisting of a 2D convolutional layer, a BN layer, a ReLU, a 2D maximum pooling layer, and an ECA Module, as is shown in FIGURE 5. The ECA Module can adaptively score and assign weights to the channel features learned by CNN, enhance features associated with the fault information, and ignore irrelevant features, thus facilitating feature learning of multi-sensor fusion data. In addition, the ECA Module has lower computational complexity.

The architecture of the proposed ECA-CNN is shown in FIGURE 6, it mainly consists of three CNN-ECA blocks, and the number of CNN-ECA blocks can be tuned to fit different task needs. Multi-channel fusion images are first input to CNN for feature learning. Then the learned channel features are scored and assigned weights adaptively by ECA Module. In ECA-CNN, the above operations are performed multiple times, meaning that the required feature information is learned layer by layer. The output dimension of each layer varies slightly depending on the parameters of the convolution and pooling layers. In this paper, the output dimension of each layer is half of the input dimension. The kernel size of the convolutional layer is 3 × 3, and the stride and the padding are both 1. The kernel size of the maximum pooling layer is 2 × 2. Finally, a GAP layer is used to establish the relationship between features and results, and a FC layer with a Softmax function is used to classify the output into c categories. TABLE 1 shows the detailed parameter configuration of ECA-CNN.

C. OVERALL FRAMEWORK OF THE PROPOSED METHOD

The overall framework of the proposed method is shown in FIGURE 7, which mainly consists of the following three steps.

Step 1: Data acquisition. Multiple sensors are installed at different positions or directions near the component of rotating machinery to be diagnosed, and multi-sensor vibration signals with various health states are collected using data acquisition devices.

Step 2: Data preprocessing. The original multi-sensor vibration signals from each state are preprocessed to obtain a multi-channel image sample set, then split into training, validation, and test set according to the time sequences [15].

Step 3: Fault diagnosis. Use the training and validation set to train the proposed ECA-CNN. The trained model is used to predict the fault types on the test set to verify the performance of the model.

III. EXPERIMENTAL VERIFICATION

In this section, to verify the effectiveness and generalization of the proposed method, two experimental cases are studied on the Case Western Reserve University (CWRU) bearing dataset [37] and the Southeast University (SEU) gearbox dataset [38], which are widely used in rotating machinery fault diagnosis.

A. EXPERIMENTAL SETTING

In those two experimental cases, the computer is a Core i5-9300H CPU @ 2.40 GHz with 16 GB of Ram and works in

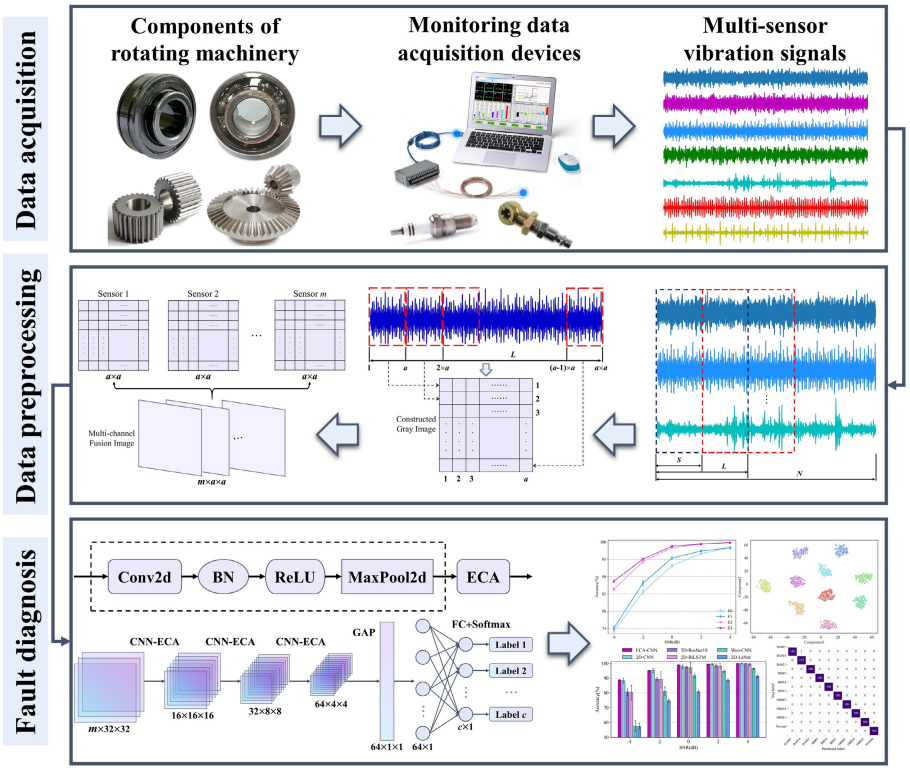


FIGURE 7. Overall framework of the proposed method.

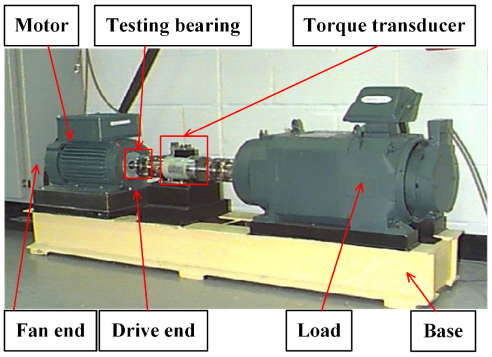


FIGURE 8. CWRU bearing test rig.

the Windows 64-bit operating system and a GPU (GTX1650) with 4 GB of memory. The programming language is Python 3.8.13, and the DL framework is Pytorch 1.10.1. During the training process, the Adam optimization algorithm is adopted, which has high computational efficiency and a small memory requirement to accelerate the convergence of the network [39], and the learning rate is 0.001. The batch size is set to 64, and the total number of training epochs is set to 100. In addition, this paper adopts accuracy to evaluate the performance of the model. Each experiment is conducted five times to reduce the effects of the randomness, and the mean and standard deviation are considered the final experimental results for analysis.

TABLE 2. Description of CWRU bearing dataset.

State	Fault diameter	Label
Ball fault	0.007 inch	BA007
	0.014 inch	BA014
	0.021 inch	BA021
Inner race fault	0.007 inch	IR007
	0.014 inch	IR014
	0.021 inch	IR021
Outer race fault	0.007 inch	OR007
	0.014 inch	OR014
	0.021 inch	OR021
Normal	-	BA007

B. CASE STUDY 1: FAULT DIAGNOSIS OF BEARINGS

1) DATASET DESCRIPTION

CWRU bearing dataset was acquired from the Electrical Engineering Laboratory of Case Western Reserve University [37]. As FIGURE 8, the test rig mainly consists of a driving motor, a torque transducer, and a power dynamometer. Vibration signals were collected from four different loads (0-3 hp), with three fault types occurring in the ball, inner race, and outer race, and there were three levels of each type of fault, as list in TABLE 2. In this paper, vibration signals at the drive end (DE) and fan end (FE) under 0 hp with a 12 kHz

TABLE 3. The results of four experiments under different levels of noise.

Exp	SNR (dB)				
	−4	−2	0	2	4
E3	**88.68 ± 0.44**	**95.04 ± 0.31**	**98.78 ± 0.23**	**99.42 ± 0.14**	**99.82 ± 0.06**
E2	86.37 ± 0.49	94.31 ± 0.43	98.36 ± 0.24	99.40 ± 0.16	99.78 ± 0.09
E1	75.15 ± 0.42	88.10 ± 0.81	95.33 ± 0.36	97.39 ± 0.31	98.33 ± 0.08
E0	74.31 ± 0.55	85.56 ± 0.77	93.15 ± 0.63	96.64 ± 0.57	98.31 ± 0.32

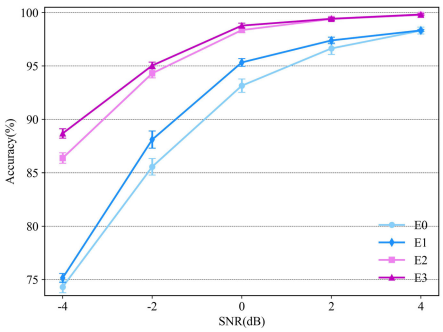


FIGURE 9. The results of four experiments under different levels of noise.

sampling frequency are selected for analysis. Multi-sensor vibration signals from each state are preprocessed to obtain 910 samples, where the first 550 samples are for training, the middle 180 samples for validation, and the rest 180 samples for testing.

2) THE EFFECTIVENESS OF THE PROPOSED METHOD

This subsection set up the following four comparative experiments to fully explore the effectiveness of the proposed method.

Experiment 0 (E0): Single-sensor vibration data (DE) + NoECA-CNN (the parameter configuration is the same as that of ECA-CNN but without the ECA module).

Experiment 1 (E1): Single-sensor vibration data (DE) + ECA-CNN.

Experiment 2 (E2): Multi-sensor vibration data (DE and FE) + NoECA-CNN.

Experiment 3 (E3): Multi-sensor vibration data (DE and FE) + ECA-CNN.

The collected vibration signals already inevitably contain a certain degree of noise. However, signals may contain higher noise levels in actual industrial working conditions. Therefore, it is necessary to study the performance of these methods under different noise environments. The Gaussian white noise with signal-to-noise ratio (SNR) values of −4 dB, −2 dB, 0 dB, 2 dB, and 4 dB are added to the raw vibration signals, respectively. The experimental results are shown in FIGURE 9 and TABLE 3.

The following conclusions can be drawn from FIGURE 9 and TABLE 3: First, E0 only obtains a mean accuracy of 74.31% when SNR = −4 dB, but when SNR = 4 dB, the mean accuracy is 98.31%. This indicates that ordinary CNN can obtain high accuracy when there is no or less noise, but its performance is significantly reduced when the noise is intense. This phenomenon has been reported in many studies. Next, E1 performs better than E0 under all noise levels, which confirms that the ECA Module can help CNN extract valuable information and ignore irrelevant information (such as noise), thus improving the interference resistance of CNN. Then, the performance of E2 is significantly improved compared to E0 and E1. For instance, the mean accuracy is 86.37% when SNR = −4 dB, which is 11.22% higher than E1. When SNR = −2 dB, E2 obtains a mean accuracy of 94.31%, which is 6.21% higher than E1. This shows that using multi-sensor fusion data can effectively improve the diagnosis performance of the model. Finally, E3 has the best diagnosis results under all noise levels, which indicates that the multi-sensor fusion data (more rich and comprehensive abnormal features) and the ECA-CNN (more powerful feature extraction capability) can synergistically improve diagnosis results.

FIGURE 10 shows the training and validation accuracy curves of four experiments when SNR = 4 dB and SNR = −4 dB. As FIGURE 10 (a), when SNR = 4 dB, the training and validation accuracy of four experiments increase rapidly and then gradually stabilize when the epoch is greater than 40. However, the validation accuracy of E0 and E1 is slightly lower than the training accuracy, which is a slight overfitting phenomenon. As FIGURE 10 (a), the convergence speed of four experiments becomes significantly slower when SNR = −4 dB. The overfitting phenomenon of E0 and E1 becomes serious. When the epoch exceeds 60, E0 and E1 can reach the training accuracy of 100%, but the validation accuracy is only about 70%. This demonstrates that learning useful features from vibration signals in strong noise conditions is challenging. In addition, although E2 and E3 also suffer from overfitting, their situation is much better than E0 and E1, and they can still achieve a validation accuracy of more than 85%.

FIGURE 11 shows the confusion matrix of E3 when SNR = 4 dB and SNR = −4 dB. The rows of the matrix represent the true labels, and the columns of the matrix represent the predicted labels. The values on the diagonal line are the number of test samples correctly classified for each category. When SNR = 4 dB, good classification results are acquired for each category. Only one sample of BA014 is misclassified as BA007, and samples from other categories are correctly classified. When SNR = −4 dB, the classification results for some categories become poor, mainly occurring in

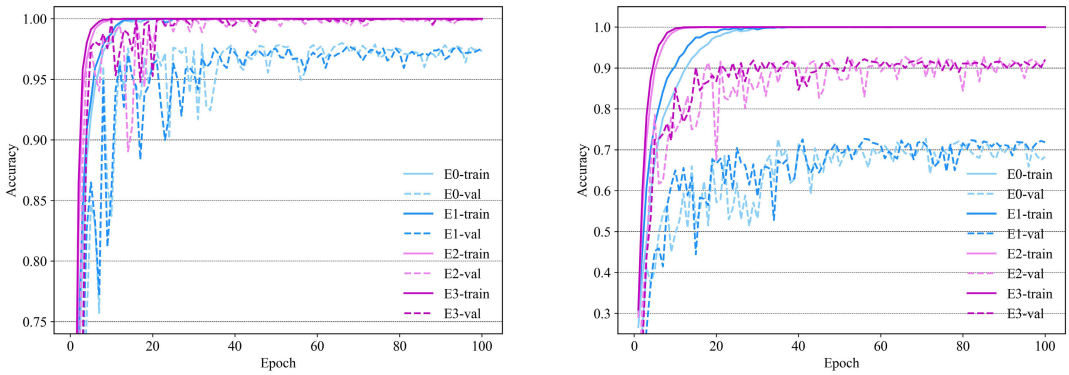


FIGURE 10. The training and validation accuracy curves of four experiments: (a) SNR = 4 dB; (b) SNR = −4 dB.

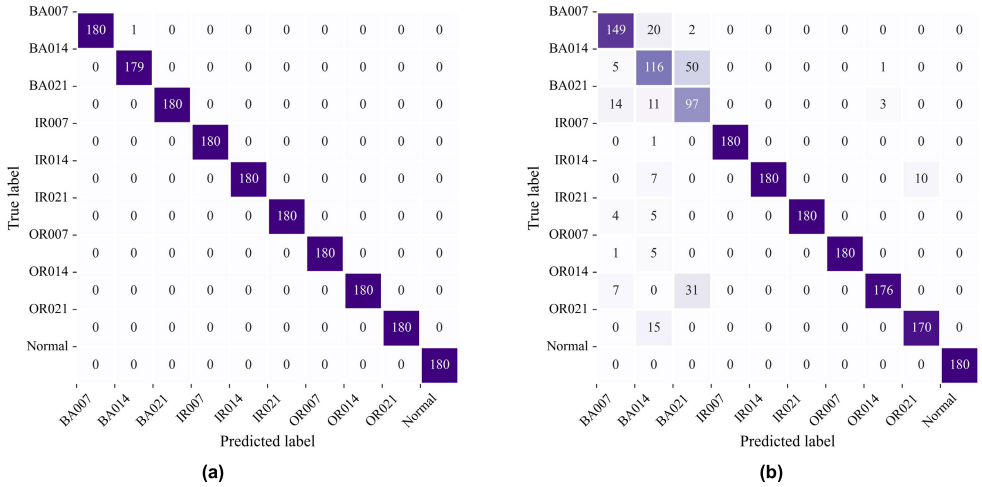


FIGURE 11. The confusion matrix of E3: (a) SNR = 4 dB; (b) SNR = −4 dB.

three levels of ball fault. BA007, BA014, and BA021 have many samples misclassified from each other, which indicates that the ball fault is susceptible to interference from noise. In addition, OR014 and OR021 are also sensitive to noise, and some of their samples are misclassified as inner race fault and ball fault.

FIGURE 12 shows the feature distribution visualization of four experiments. It is implemented by t-SNE (t-Distributed Stochastic Neighbor Embedding) technology [40], where different color shapes represent different categories. When SNR = 4 dB, the distribution of features learned by E2 and E3 have very well distinguishability, and test samples of different categories are distinguished. There is a slight overlap in the distribution of features learned by E0 and E1. When SNR = −4 dB, the distribution of features learned by E0 and E1 is entirely indistinguishable, and samples of different categories are mixed, resulting in poor diagnosis performance. This indicates that it is difficult for E0 and E1 to learn distinguishable features in strong noise conditions. E2 and E3 perform better than E0 and E1, and the distribution of features learned has only a small overlap. This confirms that the multi-sensor fusion data and the ECA-CNN can reduce intra-category distance and increase inter-category distance, effectively improving the recognition ability of differences between different categories.

3) PERFORMANCE COMPARISON

This subsection selects five existing CNN-based models for comparative analysis to verify the superiority of the proposed method. These models are Wen-CNN, 2D-CNN, 2D-LeNet, 2D-ResNet18, and 2D-BiLSTM. Wen-CNN (Wen et al., 2017) [31] is a classical CNN model for fault diagnosis. The improved LeNet-5 is used to classify the images converted from raw signals. The remaining four methods are all benchmark models proposed by Zhao et al. [15] for rotating machinery fault diagnosis. 2D-CNN, 2D-LeNet, and 2D-ResNet18 are CNN models designed for 2D input data. 2D-BiLSTM is a model that combines CNN and BiLSTM. It should be noted that the above models can classify multi-channel images by changing the number of input channels, and all models adopt the multi-sensor fusion data and

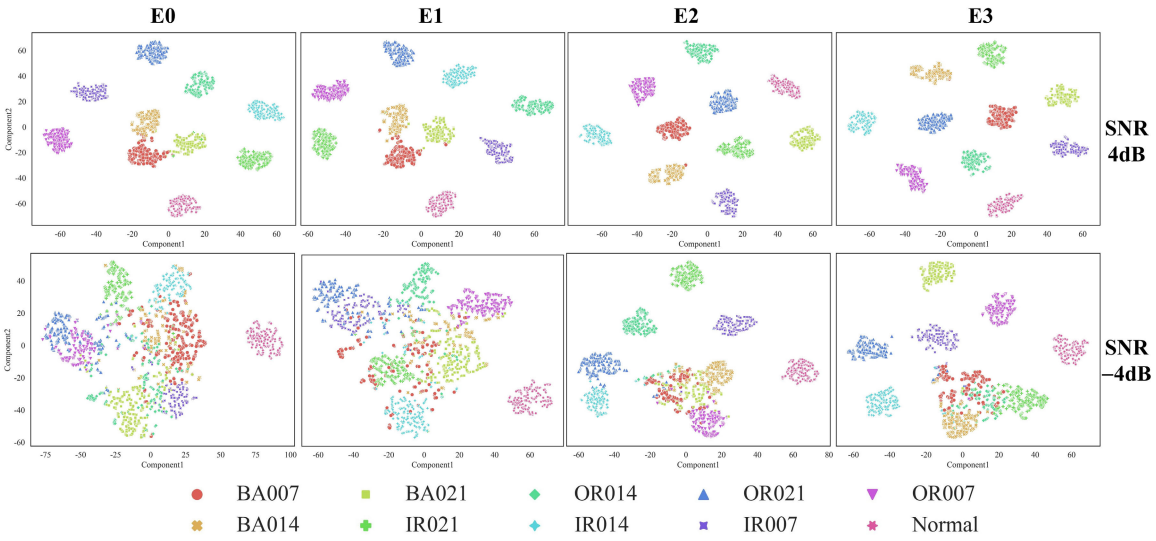


FIGURE 12. The feature distribution visualization of four experiments.

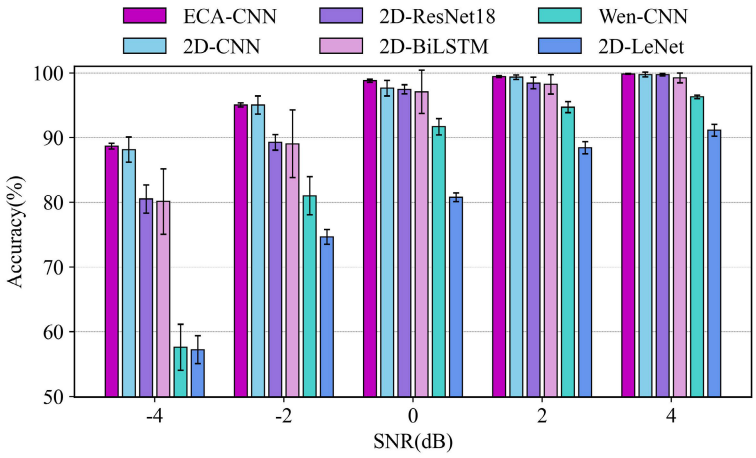


FIGURE 13. The experimental results under different levels of noise.

the same training strategy for a fair comparison. FIGURE 13 and TABLE 4 show the experimental results.

As FIGURE 13 and TABLE 4, ECA-CNN has the best diagnosis results under all noise levels. 2D-CNN performs the best among the five comparison models, its mean accuracy is similar to that of ECA-CNN, but the standard deviation of accuracy is somewhat gap. For instance, when SNR = −4 dB, ECA-CNN obtains an accuracy standard deviation of 0.49%, while that of the 2D-CNN is 1.96%. When SNR = −2 dB, the accuracy standard deviation of ECA-CNN is 0.31%, while that of the 2D-CNN is 1.38%. This indicates that ECA-CNN not only has good diagnosis performance but also has strong stability. In addition, five comparison models are noise-sensitive, causing significant performance degradation. For example, when SNR = 4 dB, the mean accuracy of Wen-CNN and 2D-LeNet can reach more than 90%, but when SNR = −4 dB, their mean accuracies are less than 60%. Significantly, when the SNR changes from −2 dB to −4 dB, the mean accuracy of the five comparison models decreases substantially, and the accuracy standard deviation increases substantially. This indicates that the interference resistance and stability of these comparison models are insufficient in strong noise conditions.

C. CASE STUDY 2: FAULT DIAGNOSIS OF GEARBOX

1) DATASET DESCRIPTION

SEU gearbox dataset was obtained from the drivetrain dynamic simulator (DDS), including a bearing dataset and a gear dataset [38]. The test rig mainly comprises a motor, a planetary gearbox, a parallel gearbox, and a brake, as shown in FIGURE 14. Vibration signals were collected from the planetary and parallel gearbox housing in three directions with a sampling frequency of 2 kHz. There were two working conditions with the rotating speed-load configuration (RS-LC) set to be 20 Hz - 0 V and 30 Hz - 2 V. This paper adopts vibration signals of the planetary gearbox in three directions (x, y, and z) for analysis. The bearings and gears have five different states under different working conditions,

TABLE 4. The experimental results under different levels of noise.

Model	SNR (dB)				
	−4	−2	0	2	4
ECA-CNN	**88.68 ± 0.44**	**95.04 ± 0.31**	**98.78 ± 0.23**	**99.42 ± 0.14**	**99.82 ± 0.06**
Wen-CNN	57.60 ± 3.56	81.00 ± 2.93	91.67 ± 1.23	94.69 ± 0.86	96.28 ± 0.26
2D-CNN	88.14 ± 1.96	95.03 ± 1.38	97.64 ± 1.21	99.33 ± 0.37	99.74 ± 0.35
2D-LeNet	57.23 ± 2.16	74.64 ± 1.16	80.77 ± 0.67	88.44 ± 0.92	91.12 ± 0.90
2D-ResNet18	80.50 ± 2.19	89.26 ± 1.20	97.45 ± 0.69	98.42 ± 0.91	99.71 ± 0.19
2D-BiLSTM	80.12 ± 5.05	89.04 ± 5.23	97.07 ± 3.35	98.24 ± 1.50	99.22 ± 0.75

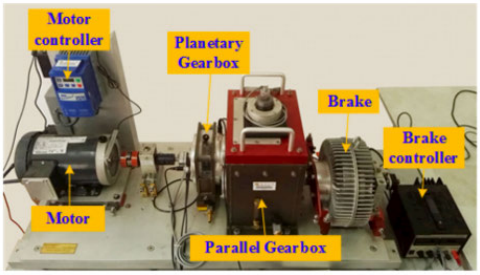


FIGURE 14. SEU gearbox test rig.

as listed in TABLE 5. At each working condition, multi-sensor vibration signals from each state are preprocessed to obtain 1022 samples, of which the first 614 are training samples, the middle 204 are validation samples, and the rest 204 are testing samples.

2) PERFORMANCE COMPARISON

This subsection investigates the generalization of ECA-CNN and five comparison models of Case study 1, under different conditions of the dataset, including the same dataset under different working conditions, the same dataset under the same working condition, the mixed dataset under the same working condition, and the complete dataset. TABLE 6 shows the experimental results.

As TABLE 6, ECA-CNN performs better than the five comparison models on each sub-dataset, with a mean accuracy of more than 99% and a small standard deviation. This indicates that ECA-CNN has good adaptability and stability. Moreover, different models have different performances under different datasets. For example, Wen-CNN obtains good diagnosis results on the bearing dataset but performs poorly on the gear dataset. 2D-LeNet also performs well on the bearing dataset but obtains poor results on the mixed dataset. It can be found that the bearing dataset is relatively simple, and the diagnosis results of all models are generally better than those of the gear dataset. Specifically, ECA-CNN reaches the mean accuracy of 100% on all three conditions of the bearing dataset. Although 2D-CNN, 2D-ResNet18, and 2D-BiLSTM perform less well than ECA-CNN, with the help of multi-sensor fused data, their diagnosis results on each dataset are acceptable.

TABLE 5. Description of SEU gearbox dataset.

Component	Speed-Load	State	Label
Bearing	20 Hz - 0 V & 30 Hz - 2 V	Ball creak	Ball
		Inner and outer ring creak	Combination
		Healthy	Health-B
		Inner ring creak	Inner
		Outer ring creak	Outer
Gear	20 Hz - 0 V & 30 Hz - 2 V	Chipping tip	Chipped
		Healthy	Health-G
		Missing tooth	Miss
		Root crack	Root
		Surface wear	Surface

D. COMPUTATIONAL COMPLEXITY COMPARISON

This subsection analyzes the computational complexity of ECA-CNN and six other models. Due to the differences in computer hardware and model optimization strategies, comparing the running times of different models can lead to significant errors [25]. Therefore, Parameters (Params) and Floating-point Operations (FLOPs) are often used as evaluation metrics for model computational complexity [41]. Params are the number of parameters of the model. FLOPs represent the number of floating-point operations in the forward propagation of the model. The Params and FLOPs of these models on the CWRU bearing dataset are compared in TABLE 7.

The architecture of NoECA-CNN is the same as ECA-CNN but without the ECA modules. Although ECA-CNN incorporates an ECA module in each layer, the computational complexity hardly increases. This proves that the ECA module has lower computational complexity. The Params of ECA-CNN and NoECA-CNN are only 0.024 M, and FLOPs are only 2.77 M, much smaller than the other models except 2D-LeNet. 2D-LeNet has smaller FLOPs than ECA-CNN, but the diagnosis performance is significantly worse than ECA-CNN. The diagnosis performance of 2D-CNN is only slightly worse than ECA-CNN, but the Params and FLOPs are much larger than ECA-CNN. The

TABLE 6. The experimental results under different conditions of the dataset.

Component	Speed-Load	ECA-CNN	Wen-CNN	2D-CNN	2D-LeNet	2D-ResNet18	2D-BiLSTM
Bearing	20 Hz - 0 V	**100.0 ± 0.00**	97.94 ± 0.36	99.98 ± 0.04	89.86 ± 0.41	99.90 ± 0.06	99.02 ± 1.62
	30 Hz - 2 V	**100.0 ± 0.00**	99.37 ± 0.21	99.95 ± 0.11	97.80 ± 0.58	98.84 ± 2.31	98.92 ± 0.92
	all	**100.0 ± 0.00**	98.91 ± 0.41	99.65 ± 0.46	92.09 ± 0.30	99.16 ± 1.66	98.75 ± 2.05
Gear	20 Hz - 0 V	**99.69 ± 0.07**	88.69 ± 1.97	99.08 ± 0.69	83.28 ± 0.38	98.41 ± 0.45	97.80 ± 1.34
	30 Hz - 2 V	**99.78 ± 0.14**	85.39 ± 2.58	99.71 ± 0.41	91.18 ± 1.31	99.41 ± 0.69	98.82 ± 1.67
	all	**99.90 ± 0.03**	81.62 ± 2.37	99.69 ± 0.30	87.61 ± 1.00	99.14 ± 0.39	99.32 ± 0.56
Mix	20 Hz - 0 V	**99.95 ± 0.02**	92.49 ± 1.43	99.79 ± 0.16	35.23 ± 0.77	97.68 ± 2.19	99.51 ± 0.40
	30 Hz - 2 V	**99.93 ± 0.05**	89.16 ± 3.13	99.88 ± 0.10	74.17 ± 5.74	98.47 ± 2.42	98.45 ± 1.67
	all	**99.99 ± 0.01**	94.45 ± 2.29	99.89 ± 0.09	51.49 ± 4.43	96.93 ± 1.05	98.54 ± 1.50

TABLE 7. The Params and FLOPs of seven models.

Model	Params (M)	FLOPs (M)
ECA-CNN	0.024323	2.77824
NoECA-CNN	0.024314	2.77062
Wen-CNN	3.039	18.441
2D-CNN	2.328	16.374
2D-LeNet	0.062	0.534
2D-ResNet18	11.179	36.418
2D-BiLSTM	0.975	6.132

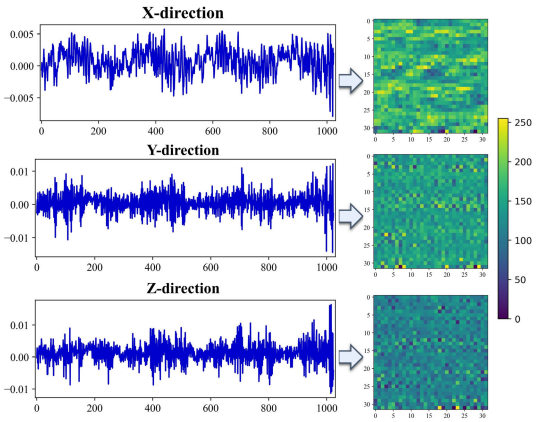


FIGURE 15. Multi-sensor signal-to-image visualization.

computational complexity of Wen-CNN is comparable to that of 2D-CNN, but the diagnosis performance is significantly worse than 2D-CNN. 2D-ResNet18 has the highest computational complexity among all models, but the diagnosis performance is comparable to that of 2D-BiLSTM, which has much smaller Params and FLOPs.

IV. VISUALIZATION ANALYSIS

This section visualizes the three aspects of the multi-sensor signal-to-image, the learned features, and the learned feature distribution to get a more intuitive view of the internal operation of the proposed method.

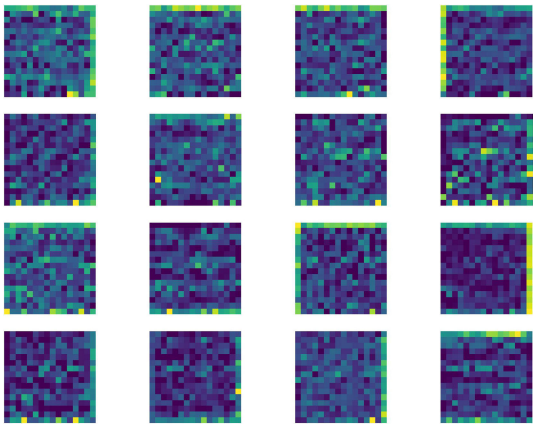


FIGURE 16. Feature map visualization.

A. VISUALIZATION OF MULTI-SENSOR SIGNAL-TO-IMAGE

To more intuitively understand the change of multi-sensor signal-to-image, the detailed conversion results of a ball creak fault sample in Case study 2 are visualized. As shown in FIGURE 15, the differences between the channel images converted from different sensors are more significant than the original signals. The 2D images of each channel show the spatial features of different periods in the 1D time-domain signals, which have more rich and comprehensive features.

B. VISUALIZATION OF LEARNED FEATURES

To further reveal the feature representations, feature maps (the number of channels is 16) extracted from the first CNN-ECA layer are shown in FIGURE 16. It can be observed that the feature map contains many bright strip lines, with brighter colors indicating higher activation degrees, which corresponds to the fault peaks response in the original time-domain signals. Further, feature maps with different activation areas show various interlayer features, which proves that the convolution kernels act as filters with different resolutions, and the channel attention mechanism attaches the importance of each channel, extracting the various impact information from the signal.

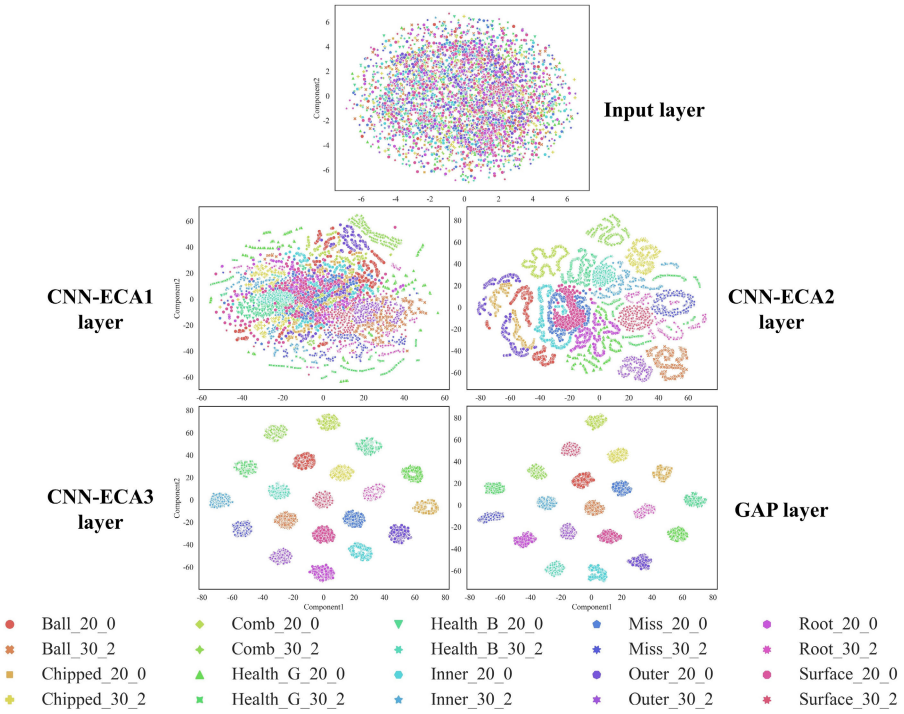


FIGURE 17. Feature distribution visualization of each layer.

C. VISUALIZATION OF LEARNED FEATURE DISTRIBUTION

To better understand the change of feature distribution in each layer of the model, the classification process of the complete dataset of Case study 2 is visualized using t-SNE. As shown in FIGURE 17, the initial distribution is messy, and as the network goes deeper, the learned features become increasingly divisible. After the first CNN-ECA layer, the feature points in the twenty categories tend to separate. After the second CNN-ECA layer, the feature points from different categories start to separate, and there are apparent boundary areas between different categories. After the third CNN-ECA layer, the feature points of different categories are all separate, and the feature points of the same categories are cluster together. Finally, the parameters of the classification layer are reduced by the GAP layer.

V. CONCLUSION

This paper proposes a rotating machinery fault diagnosis method based on multi-sensor fusion and ECA-CNN. It aims to fully exploit multi-sensor data, increase the feature learning ability of CNN, and improve computational efficiency. Multi-sensor vibration signals are sampled, converted, and channel fused into multi-channel images with rich and comprehensive features. The ECA-CNN is proposed to learn representative fault features from multi-sensor fusion data to achieve fault identification. Two experimental cases on the bearing and gearbox datasets prove that the proposed method has excellent performance, strong generalization capability, and high computational efficiency.

Although the proposed method can perform the fault diagnosis task of multi-sensor fusion well, the multi-sensor signals adopted in this paper are collected from the same type of sensors at different positions and directions. In the future, we will explore how to fuse the signals collected from different types of sensors. In addition, the study of fault diagnosis driven by multi-sensor fusion data is more suitable to current industrial practical scenarios. It can lead to more reliable fault diagnosis results with high efficiency, which is worth further exploration.

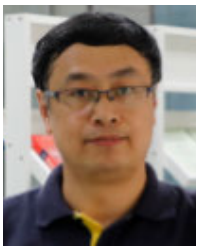
REFERENCES

[1] Y. Sun, J. Wang, and X. Wang, "Fault diagnosis of mechanical equipment in high energy consumption industries in China: A review," *Mech. Syst. Signal Process*, vol. 186, Mar. 2023, Art. no. 109833.
[2] A. Sahu, S. Palei, and A. Mishra, "Data-driven fault diagnosis approaches for industrial equipment: A review," *Exp. Syst.*, May 2023, Art. no. e13360.
[3] S. Bi, C. Wang, B. Wu, S. Hu, W. Huang, W. Ni, Y. Gong, and X. Wang, "A comprehensive survey on applications of AI technologies to failure analysis of industrial systems," *Eng. Failure Anal.*, vol. 148, Mar. 2023, Art. no. 107172.
[4] F. Kibrete and D. Woldmichael, "Applications of artificial intelligence for fault diagnosis of rotating machines: A review," in *Proc. Int. Conf. Adv. Sci. Technol. (ICAST)*, vol. 455, 2023, pp. 41–62.
[5] Y. Lei, B. Yang, X. Jiang, F. Jia, N. Li, and A. Nandi, "Applications of machine learning to machine fault diagnosis: A review and roadmap," *Mech. Syst. Signal Process.*, vol. 138, Apr. 2020, Art. no. 106587.
[6] Z. Zhu, Y. Lei, G. Qi, Y. Chai, N. Mazur, Y. An, and X. Huang, "A review of the application of deep learning in intelligent fault diagnosis of rotating machinery," *Measurement*, vol. 206, Dec. 2023, Art. no. 112346.
[7] J. Cen, Z. Yang, X. Liu, J. Xiong, and H. Chen, "A review of data-driven machinery fault diagnosis using machine learning algorithms," *J. Vib. Eng. Technol.*, vol. 10, pp. 2481–2507, Oct. 2022.
[8] C. Lou, M. Atoui, and X. Li, "Recent deep learning models for diagnosis and health monitoring: A review of research works and future challenges," *Trans. Inst. Meas. Control*, pp. 1–38, Mar. 2023.
[9] M. Huang, Z. Liu, and Y. Tao, "Mechanical fault diagnosis and prediction in IoT based on multi-source sensing data fusion," *Simul. Modell. Pract. Theory*, vol. 102, Jul. 2020, Art. no. 101981.

[10] C. Weng, B. Lu, Q. Gu, and X. Zhao, "A novel multisensor fusion transformer and its application into rotating machinery fault diagnosis," *IEEE Trans. Instrum. Meas.*, vol. 72, 2023, Art. no. 3507512.
[11] J. Yao, C. Liu, K. Song, C. Feng, and D. Jiang, "Fault diagnosis of planetary gearbox based on acoustic signals," *Appl. Acoust.*, vol. 181, Oct. 2021, Art. no. 108151.
[12] Z. He, H. Shao, Z. Ding, H. Jiang, and J. Cheng, "Modified deep autoencoder driven by multisource parameters for fault transfer prognosis of aeroengine," *IEEE Trans. Ind. Electron.*, vol. 69, no. 1, pp. 845–855, Jan. 2022.
[13] Z. Huo, M. Martínez-García, Y. Zhang, and L. Shu, "A multisensor information fusion method for high-reliability fault diagnosis of rotating machinery," *IEEE Trans. Instrum. Meas.*, vol. 71, 2022, Art. no. 3500412.
[14] Z. Zhao, J. Wu, T. Li, C. Sun, R. Yan, and X. Chen, "Challenges and opportunities of AI-enabled monitoring, diagnosis & prognosis: A review," *Chin. J. Mech. Eng.*, vol. 34, pp. 16–44, Dec. 2021.
[15] Z. Zhao, T. Li, J. Wu, C. Sun, S. Wang, R. Yan, and X. Chen, "Deep learning algorithms for rotating machinery intelligent diagnosis: An open source benchmark study," *ISA Trans.*, vol. 107, pp. 224–255, Dec. 2020.
[16] Z. Zhao, Q. Zhang, X. Yu, C. Sun, S. Wang, R. Yan, and X. Chen, "Applications of unsupervised deep transfer learning to intelligent fault diagnosis: A survey and comparative study," *IEEE Trans. Instrum. Meas.*, vol. 70, 2021, Art. no. 3525828.
[17] H. Liu, J. Zhou, Y. Zheng, W. Jiang, and Y. Zhang, "Fault diagnosis of rolling bearings with recurrent neural network-based autoencoders," *ISA Trans.*, vol. 77, pp. 167–178, Jun. 2018.
[18] Y. Qin, X. Wang, and J. Zou, "The optimized deep belief networks with improved logistic sigmoid units and their application in fault diagnosis for planetary gearboxes of wind turbines," *IEEE Trans. Ind. Electron.*, vol. 66, no. 5, pp. 3814–3824, May 2019.
[19] Y. Zhang, T. Zhou, X. Huang, L. Cao, and Q. Zhou, "Fault diagnosis of rotating machinery based on recurrent neural networks," *Measurement*, vol. 171, Feb. 2021, Art. no. 108774.
[20] F. Jia, Y. Lei, N. Lu, and S. Xing, "Deep normalized convolutional neural network for imbalanced fault classification of machinery and its understanding via visualization," *Mech. Syst. Signal Process.*, vol. 110, pp. 349–367, Sep. 2018.
[21] H. Wang, Z. Liu, D. Peng, M. Yang, and Y. Qin, "Feature-level attention-guided multitask CNN for fault diagnosis and working conditions identification of rolling bearing," *IEEE Trans. Neural Netw. Learn. Syst.*, vol. 33, no. 9, pp. 4757–4769, Sep. 2022.
[22] B. Zhao, X. Zhang, H. Li, and Z. Yang, "Intelligent fault diagnosis of rolling bearings based on normalized CNN considering data imbalance and variable working conditions," *Knowl.-Based Syst.*, vol. 199, Jul. 2020, Art. no. 105971.
[23] W. Zhang, G. Peng, C. Li, Y. Chen, and Z. Zhang, "A new deep learning model for fault diagnosis with good anti-noise and domain adaptation ability on raw vibration signals," *Sensors*, vol. 17, no. 2, p. 425, Jan. 2017.
[24] Z. Liu, H. Wang, J. Liu, Y. Qin, and D. Peng, "Multitask learning based on lightweight 1DCNN for fault diagnosis of wheelset bearings," *IEEE Trans. Instrum. Meas.*, vol. 70, 2021, Art. no. 3501711.
[25] Z. Zhao and Y. Jiao, "A fault diagnosis method for rotating machinery based on CNN with mixed information," *IEEE Trans. Ind. Informat.*, vol. 19, pp. 1–11, 2022.
[26] Z. He, H. Shao, X. Zhong, and X. Zhao, "Ensemble transfer CNNs driven by multi-channel signals for fault diagnosis of rotating machinery cross working conditions," *Knowl.-Based Syst.*, vol. 207, Nov. 2020, Art. no. 106396.
[27] J. Zheng, C. Yang, F. Zheng, and B. Jiang, "A rolling bearing fault diagnosis method using multi-sensor data and periodic sampling," in *Proc. IEEE Int. Conf. Multimedia Expo (ICME)*, Jul. 2022, pp. 1–6.
[28] Y. Huang, A. Liao, D. Hu, W. Shi, and S. Zheng, "Multi-scale convolutional network with channel attention mechanism for rolling bearing fault diagnosis," *Measurement*, vol. 203, Nov. 2022, Art. no. 111935.
[29] B. Chen, T. Liu, C. He, Z. Liu, and L. Zhang, "Fault diagnosis for limited annotation signals and strong noise based on interpretable attention mechanism," *IEEE Sensors J.*, vol. 22, no. 12, pp. 11865–11880, Jun. 2022.
[30] T. Jin, C. Yan, C. Chen, Z. Yang, H. Tian, and S. Wang, "Light neural network with fewer parameters based on CNN for fault diagnosis of rotating machinery," *Measurement*, vol. 181, Aug. 2021, Art. no. 109639.
[31] L. Wen, X. Li, L. Gao, and Y. Zhang, "A new convolutional neural network-based data-driven fault diagnosis method," *IEEE Trans. Ind. Electron.*, vol. 65, no. 7, pp. 5990–5998, Jul. 2018.
[32] W. Li, X. Zhong, H. Shao, B. Cai, and X. Yang, "Multi-mode data augmentation and fault diagnosis of rotating machinery using modified ACGAN designed with new framework," *Adv. Eng. Informat.*, vol. 52, Apr. 2022, Art. no. 101552.
[33] S. Speth, A. Goncalves, B. Rigault, S. Suzuki, M. Bouazizi, Y. Matsuo, and H. Predinger, "Deep learning with RGB and thermal images onboard a drone for monitoring operations," *J. Field Robot.*, vol. 39, pp. 840–868, Sep. 2022.
[34] S. Ioffe and C. Szegedy, "Batch normalization: Accelerating deep network training by reducing internal covariate shift," in *Proc. Int. Conf. Mach. Learn.*, vol. 37, Jul. 2015, pp. 448–456.
[35] Z. Niu, G. Zhong, and H. Yu, "A review on the attention mechanism of deep learning," *Neurocomputing*, vol. 452, pp. 48–62, Mar. 2021.
[36] Q. Wang, B. Wu, P. Zhu, P. Li, W. Zuo, and Q. Hu, "ECA-Net: Efficient channel attention for deep convolutional neural networks," in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jun. 2020, pp. 11531–11539.
[37] W. A. Smith and R. B. Randall, "Rolling element bearing diagnostics using the Case Western Reserve University data: A benchmark study," *Mech. Syst. Signal Process.*, vol. 64, pp. 100–131, Dec. 2015.
[38] S. Shao, S. McAleer, R. Yan, and P. Baldi, "Highly accurate machine fault diagnosis using deep transfer learning," *IEEE Trans. Ind. Informat.*, vol. 15, no. 4, pp. 2446–2455, Apr. 2019.
[39] D. Kingma and J. Ba, "Adam: A method for stochastic optimization," in *Proc. Int. Conf. Learn. Represent. (ICLR)*, 2015, p. 13.
[40] N. Pezzotti, B. P. F. Lelieveldt, L. V. D. Maaten, T. Höllt, E. Eisemann, and A. Vilanova, "Approximated and user steerable tSNE for progressive visual analytics," *IEEE Trans. Vis. Comput. Graphics*, vol. 23, no. 7, pp. 1739–1752, Jul. 2017.
[41] K. He and J. Sun, "Convolutional neural networks at constrained time cost," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jun. 2015, pp. 5353–5360.

HONGXING WANG received the B.S. degree from Liaocheng University, in 2019, and the M.S. degree from the Nanjing University of Aeronautics and Astronautics, Nanjing, China, in 2022, where he is currently pursuing the Ph.D. degree in instrument science and technology with the College of Aerospace Engineering. His research interest includes deep learning with applications in rotating machinery fault diagnosis.

HUA ZHU received the B.S., M.S., and Ph.D. degrees from the Nanjing University of Aeronautics and Astronautics, Nanjing, China, in 2000, 2003, and 2007, respectively. He is currently an Associate Researcher with the College of Aerospace Engineering, Nanjing University of Aeronautics and Astronautics. His current research interests include precision drive and control, intelligent manufacturing, and industrial control systems.



HUAFENG LI received the B.S. and Ph.D. degrees from the Huazhong University of Science and Technology, Wuhan, China, in 1997 and 2002, respectively. He is currently a Professor with the College of Aerospace Engineering, Nanjing University of Aeronautics and Astronautics. His current research interests include precision drive and controls, motion control systems, and wireless transmission.