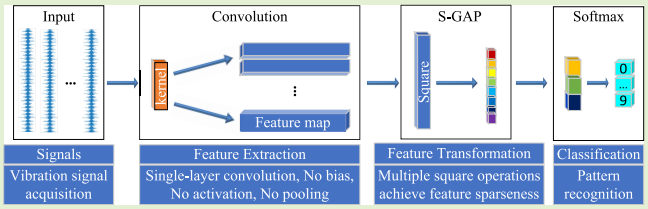
An Explainable and Lightweight Improved 1-D CNN Model for Vibration Signals of Rotating Machinery

Pengfei Pang, Jian Tang, Jiqing Luo, Miao Chen, Hui Yuan, and Lei Jiang

***Abstract*—Previous 1-D convolutional neural network (1-D CNN) models for vibration fault diagnosis have high computational complexity and poor interpretability, which cannot meet the higher requirements of model storage, computational efficiency, and reliability for airborne and portable devices. Considering these challenges, an explainable and lightweight 1-D CNN (ELCNN) model based on square global average pooling (S-GAP) and improved vibration signals is proposed. The feature extraction and classification layers of 1-D CNN are optimized to minimize the model parameters and computational complexity and improve interpretability while ensuring diagnostic accuracy. The model compresses the number of convolutional layers, removes unnecessary bias, activation function, and pooling layer, and replaces the fully connected layer (FCL) with S-GAP. Improved 1-D CNN models of different methods are evaluated and analyzed on public datasets of rolling bearings. Results show that the ELCNN improved for vibration signals is more lightweight, antinoise, and explainable than other models, and the diagnostic accuracy is further improved.**



***Index Terms*—Convolutional neural networks (CNNs), explainable, fault diagnosis, lightweight, vibrational signal.**

I. Introduction

With the development of intelligent equipment, it is of great significance to study how to intelligently perceive the operating state, detect faults as early as possible, and avoid losses [1], [2]. Previous fault diagnosis methods can be divided into two categories: one is based on models (such as stochastic resonance systems [3] and harmonic detection models [4]) and the other is based on vibration signals to extract fault features (such as short-time Fourier transform (STFT) [5], wavelet transform (WT) [6], and empirical mode decomposition (EMD) [7]). First, these methods rely on professional knowledge, such as mathematics, physical models, and fault mechanisms, and require a high level of diagnostic experts. Second, the above methods include signal feature learning, state feature selection and transformation, fault pattern recognition, and other links, and many processes require human participation and do not have incremental and adaptive learning capabilities [8].

Artificial intelligence (AI)-based intelligent fault diagnosis methods get rid of the dependence on experts and experience, and realize intelligent fault diagnosis by learning knowledge from historical data, such as clustering algorithm (CA) [9] and principal component analysis (PCA) [10] based on unsupervised learning, decision tree (DT) [11], and support vector machine (SVM) [12] based on supervised learning. However, these methods are only applicable to small and medium-sized datasets and have limited ability to diagnose large-scale and complex data.

Since 2015, fault diagnosis algorithms based on deep learning (DL) have become a hot spot [13]. Among these algorithms, the 1-D convolutional neural network (CNN) is particularly suitable for the overall optimization of each link of fault diagnosis and the improvement of the intelligent level of diagnosis due to its powerful data abstraction, end-to-end structure, and adaptive learning feature ability based on 1-D vibrational signals [13], [14], [15], [16], [17]. To make the 1-D CNN model better fit the characteristics of intelligent fault diagnosis based on vibration signals, the improvement of the model and algorithm is mainly carried out in two directions. One is to combine traditional diagnostic models (such as multidomain feature fault diagnosis [18] and WaveCluster clustering analysis [19]) and shallow learning

Manuscript received 12 September 2023; revised 23 October 2023; accepted 23 October 2023. Date of publication 24 January 2024; date of current version 29 February 2024. The associate editor coordinating the review of this article and approving it for publication was Dr. Dong Wang. *(Corresponding author: Jian Tang.)*

Pengfei Pang, Jian Tang, Jiqing Luo, and Lei Jiang are with the Engineering Institute of Engineering Corps, the Army Engineering University, Nanjing 210007, China (e-mail: 664522431@qq.com; lgdx_tj@163.com; ljqlgd@163.com; jl4573@126.com).

Miao Chen is with the Mechanical Engineering, the Nanjing University of Technology, Nanjing 211816, China (e-mail: 707403322@qq.com).

Hui Yuan is with the Advanced Equipment Manufacturing Research Institute, the Nanjing University of Technology, Nanjing 211816, China (e-mail: yzx9807@sina.com).

Digital Object Identifier 10.1109/JSEN.2023.3327783

models (such as logistic regression [20] and SVM [21]) with 1-D CNN for optimization. The second is to introduce advanced network architectures (such as deep residual learning (DRL) [22], multibranch multiscale convolutional neural networks (MBMSCNNs) [23], and multiattention mechanism (MAM) [24]) from computer images to improve the feature extraction ability of 1-D CNN. Although the above methods improve the performance of the model to a certain extent, there are still deficiencies, such as high computational complexity and poor interpretability, which cannot meet the higher requirements of airborne and portable devices. Lightweight and interpretability have always been difficult problems in intelligent fault diagnosis, which is also a direction worthy of studies in the future 1-D CNN.

Aiming to achieve the lightweight of the model, many studies have focused on using depthwise separable convolution (DSC) [2] and adaptive convolution (AC) [25] to reduce the network size, but there is no specific analysis of the network structure for vibration signals to make the extracted features physically interpretable. On interpretability issues, many works combine attention mechanisms with multiscale convolution (MSC) and DRL [26], [27], [28] to remove redundant information. These methods are only based on statistical probability to guide 1-D CNN learning, without considering the frequency characteristics of complex faults.

Some scholars have made effective explorations in improving the 1-D CNN structure and realizing a lightweight interpretable model. For example, Li et al. [29] selected $\text{sign}(x)$ as the activation function to solve the problem of gradient disappearance and removed the pooling layer to avoid feature loss. Kim et al. [30] used global power pooling to make the model focus on more effective time–frequency features. These methods only proposed improvement measures for the model and neither analyzed its internal diagnostic mechanism nor did they evaluated its performance from the aspects of lightweight and interpretability.

To meet the above challenges, a new explainable and lightweight 1-D CNN (ELCNN) diagnostic model for vibration signals is proposed. The main work of this article is given as follows.

1) Traditional time–frequency analysis methods, such as STFT [31] and WT [32], are essentially linear convolution processes. There is no nonlinear mapping such as bias and activation, and there is no use of a pooling layer to remove high-frequency features. Therefore, this article is not limited to the traditional neuron model. While removing the bias, activation function, and pooling layer, the network scale is compressed as much as possible, and a single-layer CNN ELCNN is constructed.
2) Considering that low signal-to-noise ratio (SNR) signals in strong noise environments may affect fault diagnosis performance, in order to extract fault features more effectively, a method based on square global average pooling (S-GAP) enhanced signal time-domain envelope is proposed. This method extracts the frequency modulation (FM) and amplitude modulation (AM) characteristics of complex faults through multiple square operations.
3) In order to meet the higher requirements of airborne and portable devices for the storage, computational efficiency, and reliability of diagnostic models, the performance of various models is comprehensively evaluated on the public bearing datasets to verify the effectiveness of the improved strategies and the lightweight interpretability advantages of the ELCNN model.

The structure of this article is given as follows. Section II introduces the improved 1-D CNN network framework for vibration signals. Section III tests the effectiveness and superiority of the various improved strategies proposed in this article through the ablation experiments and comparative experiments of the ELCNN feature extraction layer and the classification layer. Section IV verifies the advanced nature of the proposed model in terms of lightweight and interpretability through comparative experiments of different improved 1-D CNN models. Finally, Section V summarizes the main work and innovation of this article.

II. Improved 1-D CNN for Vibration Signals

As shown in Fig. 1, the 1-D CNN is an end-to-end network, including an input layer, a convolutional layer, a pooling layer, a fully connected layer (FCL), and a SoftMax layer. The convolution layer and the pooling layer correspond to the feature extraction layer, while the FCL and the SoftMax layer are used to complete the pattern classification function (classification layer). Similar to Hilbert–Huang transform (HHT) [32], which uses EMD to adaptively select the basis function and extracts the time–frequency characteristics of the signal through the Hilbert transform [33], 1-D CNN first adaptively learns the weights, biases, and other parameters of the model through gradient descent and backpropagation. Then, the signal features are extracted layer by layer from low to high through the feature extraction layer, and finally, the feature dimension reduction and classification are realized through the classification layer. To ensure the effect of fault diagnosis and realize the lightweight and interpretability of a 1-D CNN, we improve the model feature extraction and classification layers for vibration signals.

A. Improvement of Feature Extraction Layer

Traditional vibration signal feature extraction methods, such as STFT, WT, and HHT, are essentially convolution operations based on a certain convolution kernel. A 1-D CNN similarly relies on the convolutional layer and the pooling layer to extract (learn, select, and transform) the local features of the signals [34]. Hence, it is very important to optimize the convolutional layer and the pooling layer [35].

As shown in Fig. 2, the essence of convolution is a dot product in the filter weights ω and the local space of input data x; then, the bias b is added; and finally, the activation operation δ is performed. Considering the convolution operation of traditional feature extraction methods, there is no nonlinear mapping, such as bias and activation, and there is no feature extraction by pooling. Based on the above characteristics, this section analyzes the hyperparameters of the feature extraction layer for the vibration signal and removes the unnecessary

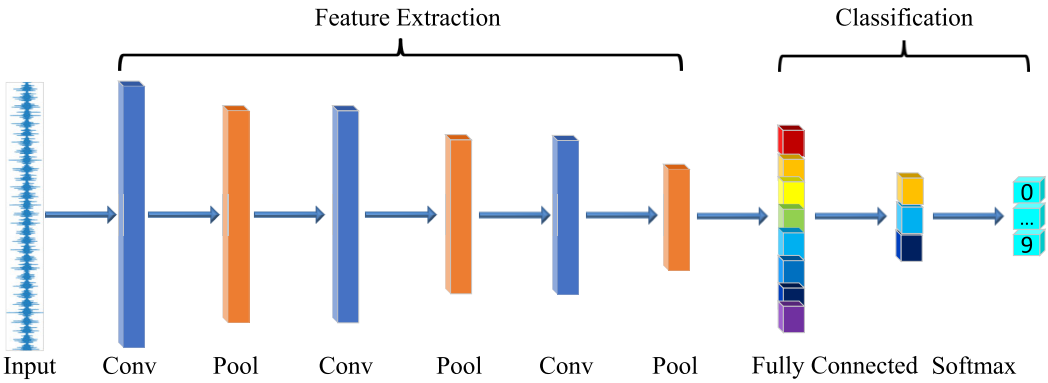


Fig. 1. Classic 1-D CNN.

and nonlinear feature extraction links while retaining the time–frequency characteristics of the vibration signal as much as possible.

1) Removal of Bias: The output of each neuron in the 1-D CNN convolutional layer is $y = \delta(\omega x + b)$ [36], where the bias $-b$ can be regarded as the threshold to control each neuron and δ is the activation function. The activation function of the neuron is defined as sign(x), and the output of each neuron is defined as

$$\text{sign}\,(\omega x + b) = \begin{cases} +1, & \omega x + b \geq 0 \\ -1, & \omega x + b < 0. \end{cases} \tag{1}$$

When $WX < -b$ and output $= -1$, the neuron is inhibited, and when $WX \geq -b$ and output $= 1$, it is activated. Therefore, the essence of convolution is to extract some local information and discard the information outside the threshold during the extraction process. The bias is used to adjust the proportion of discarded local information. Without this parameter, there is a lack of flexibility to adjust the information discarding rate. For a vibration signal of rotating machinery, especially for early faults, the fault characteristics are often very weak, submerged in normal vibration signals and even noise, and cannot be ignored. Therefore, in the design of a 1-D CNN for vibration signals, the bias should be removed to prevent the loss of weak fault features due to neuron suppression.

2) Removal of Activation Function: Activation functions commonly used in image recognition include Rectified Linear Unit (ReLU)(x), Sigmoid(x), Tanh(x), and Softsign(x) [37]. The outputs of ReLU(x) and Sigmoid(x) functions are only positive [see Fig. 3(b)]. The vibration signal fluctuates up and down at the equilibrium position, with a positive or negative amplitude. The positive operation is easy to discard the negative component of the signal, which weakens the time–frequency characteristics and introduces the nonfault characteristics [as shown in the nonfault frequency in the green box near 1000 Hz in Fig. 4(b)]. The Tanh(x) function is easily oversaturated in the range (–1,1), resulting in gradient explosion or overfitting. The Softsign(x) function can better avoid the above problems, but, because the activation function is a nonlinear function [38], the features learned by the convolutional layer will have a certain deviation, and nonfault features are introduced [as shown in the green box in Fig. 4(c), especially the nonfault component near 1000 Hz].

TABLE I
STATISTICAL CHARACTERISTICS OF THE SIGNALS LEARNED FROM THE CONVOLUTIONAL LAYERS USING DIFFERENT ACTIVATION FUNCTIONS. THE KURTOSIS IS USED TO REFLECT THE SHARPNESS OF THE VIBRATION SIGNAL WAVEFORM, AND THE PULSE FACTOR IS USED TO EVALUATE THE IMPACT AND TRANSIENT CHARACTERISTICS OF THE VIBRATION SIGNAL

Characteristics	Raw	ReLu	Softsign	No Activation
Kurtosis	10.01	13.00	7.31	11.79
Pulse Factor	13.32	15.80	9.39	15.69

Compared with Raw (unprocessed original signal), the time-domain amplitude features learned by ReLU(x) and Softsign(x) have errors (Fig. 3(b) introduces weak nonfault components, and Fig. 3(c) expands the signal amplitude gap) and generate nonfault components in the frequency domain [see Fig. 4(b) and (c)]. The impact pulse signal can be detected by kurtosis and pulse factor. As shown in Table I, the positive operation of ReLU(x) introduces a sharp nonfault component [see Fig. 4(b)] to strengthen the impact excitation, while the nonlinear operation of Softsign(x) learns more stationary nonfault components [see Fig. 4(c)], and the signal mean is pulled up, which weakens the impact excitation. Compared with ReLU(x) and Softsign(x), which introduce nonfault components, the convolutional layer without an activation function can better learn the low-frequency feature of Raw and strengthen the impact excitation [as shown in Fig. 4(d)], which is more interpretable. The reason for using the activation function is to prevent gradient explosion and overfitting. We focus on exploring the lightweight and interpretability of 1-D CNN. First, the network structure is simple, and there is no gradient explosion. Second, the overfitting problem can be solved by the S-GAP of the classification layer, so we remove the activation function.

3) Removal of Pooling Layer: The common pooling in 1-D CNN includes maximum pooling, average pooling, and GAP [39], which are mainly used to reduce the feature dimension of the input signal and prevent overfitting [40]. The essence of the pooling layer is to remove the high-frequency

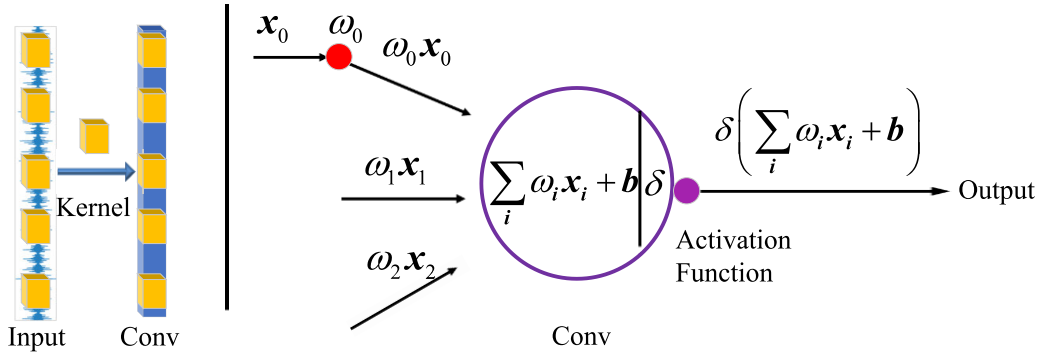


Fig. 2. Operation of the 1-D CNN convolutional layer.

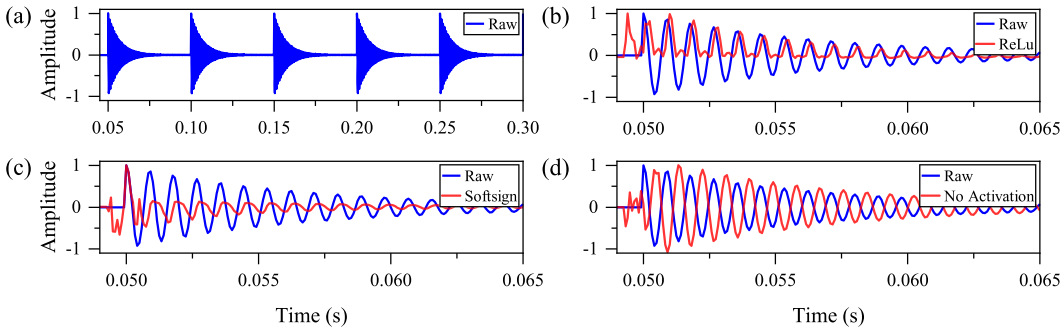


Fig. 3. Influence of activation function on time-domain characteristics of signal: (a) original signal, (b) ReLU, (c) Softsign, and (d) no activation. (a) Time-domain waveform of the unprocessed original signal. (b)–(d) Time-domain waveform of a single cycle.

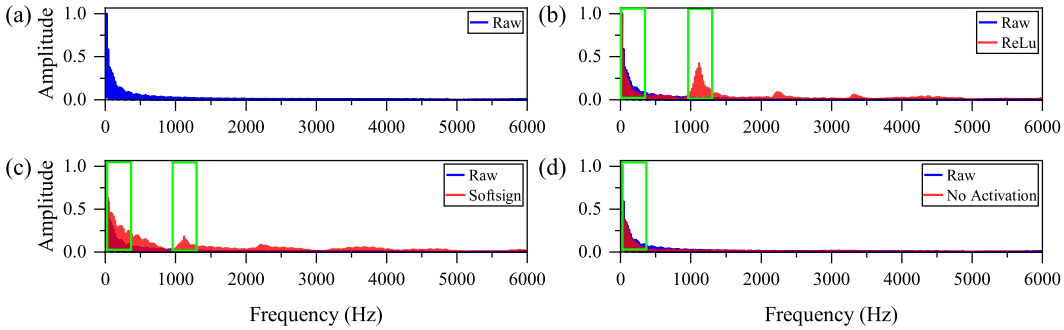


Fig. 4. Influence of activation function on frequency-domain characteristics of signal: (a) original signal, (b) ReLU, (c) Softsign, and (d) no activation. (a) Frequency-domain diagram of the original signal. The green box of (b)–(d) is the frequency band to be focused on.

characteristics of the signal and retain the low-frequency characteristics of the signal, and moderate pooling can also suppress the high-frequency noise. However, the traditional pooling operation often suppresses the broadband low-amplitude fault features (average pooling) or discards them (maximum pooling). As shown in Fig. 5(b) and (c), average pooling takes the average value of the signal and weakens the fault impact excitation, while maximum pooling ignores the low-amplitude information because of the maximum value [30]. Fault characteristics are often hidden in broadband and low-amplitude impact excitation. To improve the interpretability of the extracted features, the pooling layer is no longer used. In Fig. 6(b) and (c), the signals learned by average pooling and maximum pooling lose the fault frequency characteristics (greater than 100 Hz). In particular, the average pooling also introduces pseudofrequency components. It is obvious that 1-D CNN without a pooling layer can perfectly retain the time–frequency characteristics of the signal, so the pooling layer is removed.

B. Improvement of Classification Layer

As shown on the left-hand side of Fig. 7, the FCL can be understood as a convolution of the previous features to obtain the output of a neuron. Through the nonlinear mapping of the activation function, the FCL can simulate any nonlinear transformation to achieve dimensionality reduction and classification. However, it has obvious disadvantages, such as parameter redundancy and the inability to maintain spatial structure.

To overcome shortcomings of FCL such as a large number of parameters and overfitting, the reference [41] and network models, such as ResNet and GoogLeNet, use GAP instead

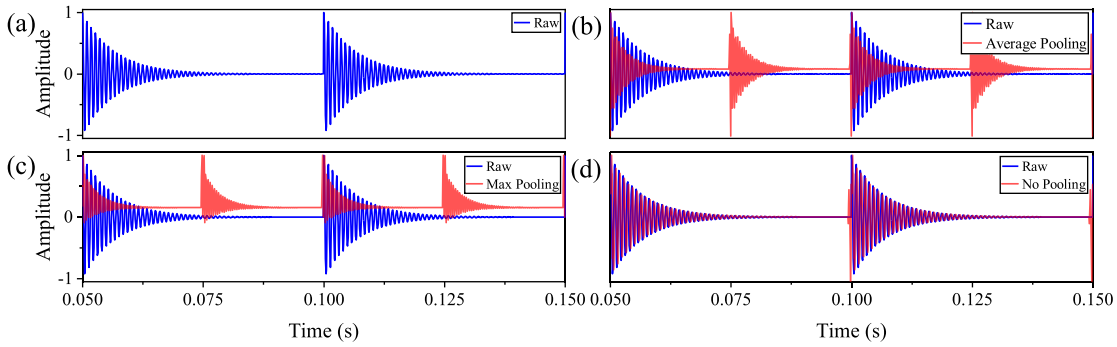


Fig. 5. Influence of different pooling layers on time-domain characteristics of signal: (a) original signal, (b) average pooling, (c) maximum pooling, and (d) no pooling.

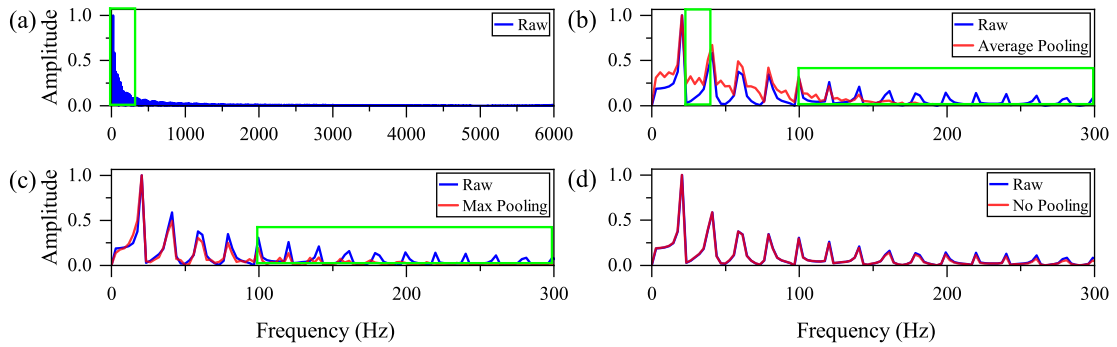


Fig. 6. Influence of different pooling layers on frequency-domain characteristics of signal: (a) original signal, (b) average pooling, (c) maximum pooling, and (d) no pooling. (a) Frequency-domain feature map of the original signal. (b)–(d) Key frequency band in the green box of (a). The green parts of (b) and (c) are the spectral features affected by pooling.

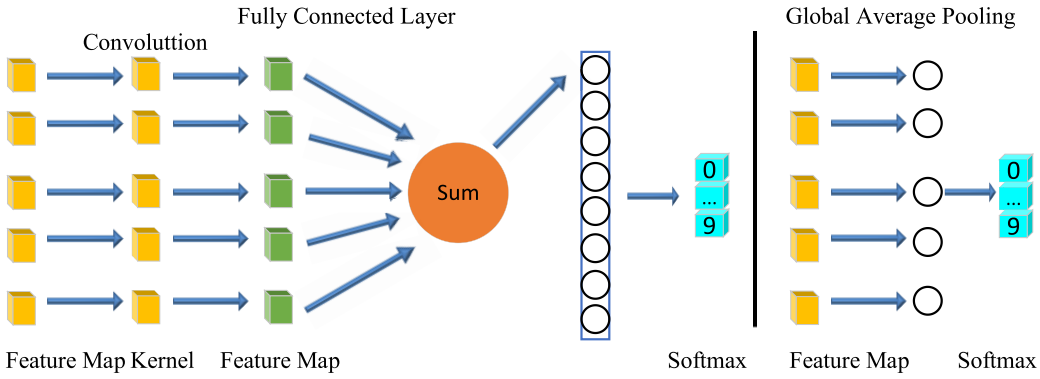


Fig. 7. Classification layer of 1-D CNN. (Left) FCL. (Right) GAP.

of FCL to fuse learned deep features (right-hand side of Fig. 7) and achieve a better prediction performance [30]. However, the above two methods are unsuitable for rotating machinery fault diagnosis because Flatten in FC concatenates multiple feature maps and then classifies them, losing the ability of category feature localization; while GAP has the ability of feature location [42], it offsets positive and negative signals and weakens features to reduce the classification effect. To overcome the above shortcomings and improve the generalization performance of the model, the signal features after convolution are transformed by square and positive operation, respectively, and then, the results are sent to the GAP layer to obtain the output features of the improved model after signal transformation (see Fig. 8).

From the analysis of Fig. 8, it can be seen that the square and positive operations of the signal learned by the convolutional layer overcome the problem of positive and negative offset caused by the GAP layer. However, from the perspective of the learning effect, there are obvious differences between the two. Among them, the ReLU operation removes the negative information and weakens the model's ability to learn fault features; the square operation expands the peak-to-valley ratio, which not only strengthens the time-domain envelope of the signal to highlight the characteristics of local extreme points but also suppresses overfitting, which is more in line with the characteristics of the vibration signal.

C. Improved ELCNN Architecture

Combined with the various improvement strategies proposed above, we construct an improved ELCNN model based on S-GAP. At the same time, to more effectively realize the lightweight and feature interpretability of the model, the

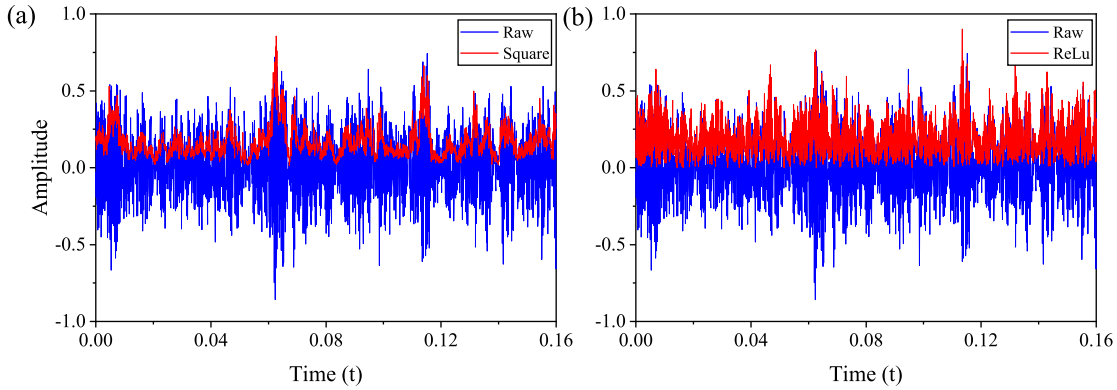


Fig. 8. Time-domain waveform obtained by square or positive operation of the features learned by the convolutional layer: (a) square and (b) ReLU.

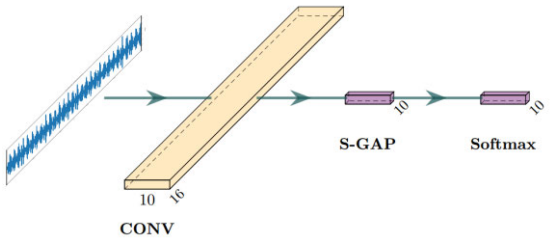


Fig. 9. Improved ELCNN structure.

convolution layers of the model are further compressed, and the number of square operations of the S-GAP layer is increased. The network parameters and structure of ELCNN are shown in Table II and Fig. 9.

III. Effectiveness and Superiority Verification of the Improved Strategies

To verify the effectiveness and superiority of the improvement strategies, this section evaluates the performance of ELCNN based on public datasets.

A. Sources and Classifications of Samples

The signal samples are from the single-condition-bearing dataset of Case Western Reserve University (CWRU) and the multicondition-bearing dataset of Xi'an Jiaotong University (XJTU).

1) CWRU Single-Condition Bearing Dataset: The experimental platform is a motor drive system, including a motor, a torque sensor (decoder), and a power tester [43]. The driving end bearing model is SKF6205, and the bearing assembly consists of inner race (IR), outer race (OR), and balls (B). Bearing damage is single-point damage by EDM. The bearing conditions are the motor load of 0 hp, and the rotating frequency is 29.95 Hz. The experimental data collect the vibration acceleration signals of the fault bearing through the acceleration sensor, and the sampling frequency is 12 kHz. The experimental samples include three types of faults: IR, OR, and B. The damage degrees of IR, OR, and B faults are 007, 014, and 021, respectively, and the damage diameter corresponds to 0.1778, 0.3556, and 0.5334 mm, respectively [44]. To improve the availability of data and avoid contingency, the original experimental samples are divided at equal intervals. The 1500 samples are sampled for each type. The training, validation, and test sets are in a 7:1:2 ratio.

2) XJTU Multicondition Bearing Dataset: The experimental platform is Spectra Quest, and the bearing model is NSK6203. The bearing damage is a single-point damage processed by a grinding pen. The load of the bearing motor is 0 hp, and there are three kinds of rotating frequencies: 19.05, 29.05, and 39.05 Hz. The experimental data are from the piezoelectric acceleration sensor, and the sampling frequency is 25.6 kHz. The signal samples include normal, IR, and OR faults, and the damage of each fault has three levels: mild, moderate, and severe. The preprocessing of signal samples is consistent with the CWRU.

B. Settings of Model Parameters

The model parameters mainly considered include kernel size, filter number, and signal sampling length.

1) Optimization of Kernel Size and Filter Number: van den Hoogen et al. [45] used wide convolution kernels to improve the performance of 1-D CNN. It can be seen that it is necessary to study the hyperparameters of convolution kernels, such as the kernel size and number of filters. To select the appropriate model parameters, we analyze the ELCNN diagnostic performance of different sizes of kernels and different numbers of filters in Fig. 10 and the spectral characteristics of the signals extracted by different sizes of kernels in Fig. 11.

As shown in Fig. 10, when the size of the convolution kernel is greater than 8 and the number of filters is greater than 5, it can basically achieve satisfactory classification results, but it is not the case that the larger the number of filters and the larger the kernel size, the better.

An excessive number of filters will increase the computational burden and is not conducive to classification; the larger kernel cannot effectively extract signal features by using translation invariance. As shown in Fig. 11 and Table III, when the kernel size is equal to 16, the features extracted by the convolution layer have a good correlation with the original signal and can learn the fault features more effectively. On the whole, when the kernel size is 16 and the filter number is 10, the model performance and the feature learning are the best.

2) Optimization of Signal Sampling Length: There are rich fault characteristics in the vibration signal, and a reasonable

TABLE II
NETWORK PARAMETER CONFIGURATION OF THE IMPROVED 1-D CNN

Layer	Type	Kernel	Channel	Bias	Padding	Activation	Pooling	Output
0	Input	-	-	-	-	-	-	4096*1
1	Conv	16	10	None	Same	None	None	4096*10
S-GAP								
Softmax								
C0 C1 C2 C3 C4 C5 C6 C7 C8 C9								

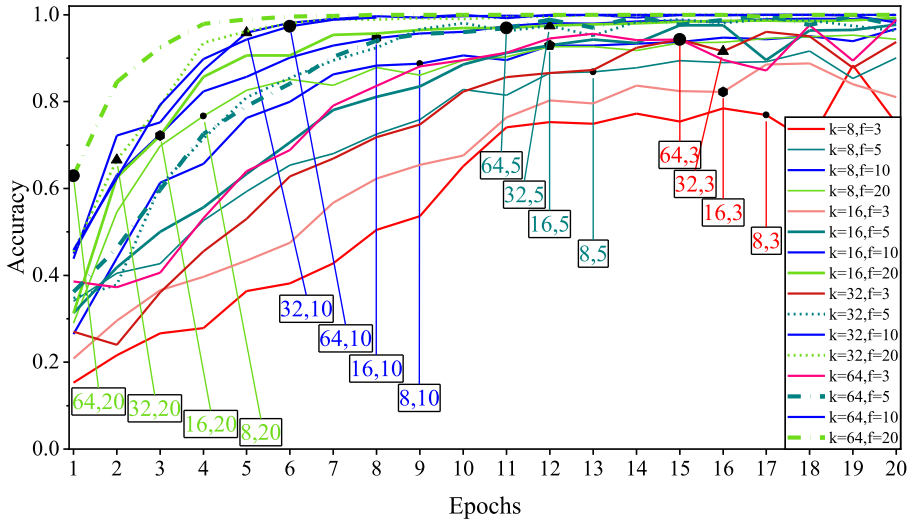


Fig. 10. Classification accuracy of ELCNN under different sizes of convolutional kernels and numbers of filters, where k is greater than 8 and f is greater than 5; the model diagnosis effect is better. (k denotes the size of the convolutional kernel, and f denotes the number of filters).

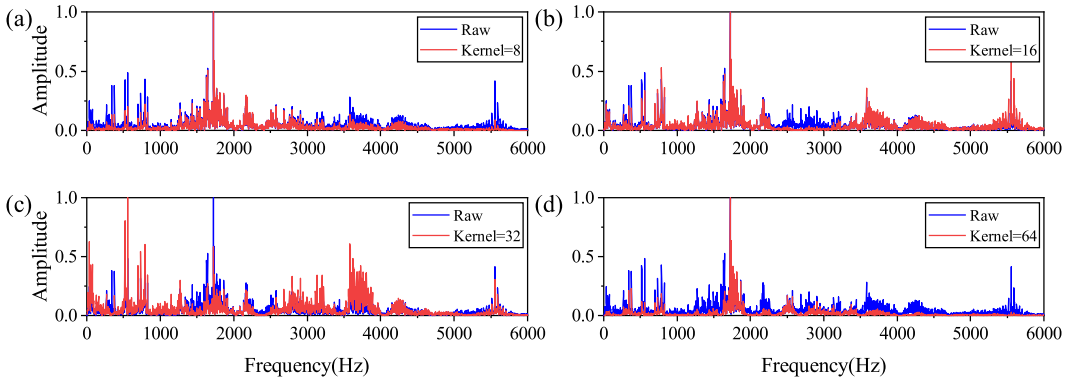


Fig. 11. IR fault spectrums of XJTU extracted by kernels of different sizes (Raw is the original input signal that is not processed): (a) kernel = 8, (b) kernel = 16, (c) kernel = 32, and (d) kernel = 64. When kernel = 8 or kernel = 16, the signal features extracted by the convolution layer have a good correlation with the Raw signal.

sampling length is the basic condition to ensure that the sample truly reflects the original signal. The sampling length refers to the time record length required to be able to analyze the lowest frequency in the signal. The lowest fault frequency in rolling bearings is generally the cage fault frequency (11.93 Hz in CWRU and 17.6 Hz in XJTU). The sample length occupied by CWRU cage fault is 1006 (12 000/11.93 $\approx$ 1006 and 1006*4 = 4024, where 12 kHz is the sampling frequency of CWRU), and XJTU is 1455 (25 600/17.6 $\approx$ 1455 and 1455*2 = 2910, where 25.6 kHz is the sampling frequency of XJTU). To ensure the integrity of the data cycle and avoid the marginal effect, we use 4096 as the experimental data length to ensure that the signal sample has more than two cage fault cycles.

C. Verification of the Improved Strategies' Effectiveness

Different combinations of improvement strategies have different effectiveness. The following will analyze the effectiveness of 1-D CNN improvement strategies from both internal and external perspectives.

TABLE III
CORRELATION COEFFICIENT BETWEEN RAW SIGNAL AND SIGNAL FEATURES EXTRACTED BY DIFFERENT DATASETS AND DIFFERENT SIZES OF KERNELS. COMPARED WITH OTHER SIZE CONVOLUTION LAYERS, THE SIGNAL FEATURES EXTRACTED BY THE KERNEL = 16 CONVOLUTION LAYER MARKED IN RED HAVE A BETTER CORRELATION WITH RAW

No.	Data Set	Sampling Frequency (kHz)	Fault Frequency (Hz)	Fault Category	kernel=8	kernel=16	kernel=32	kernel=64
1	CWRU	12	162.19	IR	0.86	0.95	0.89	0.86
2			107.36	OR	0.93	0.93	0.9	0.92
3			141.17	B	0.54	0.58	0.6	0.54
4	XJTU	25.6	192.56	IR	0.91	0.91	0.73	0.79
5			119.84	OR	0.8	0.82	0.8	0.76

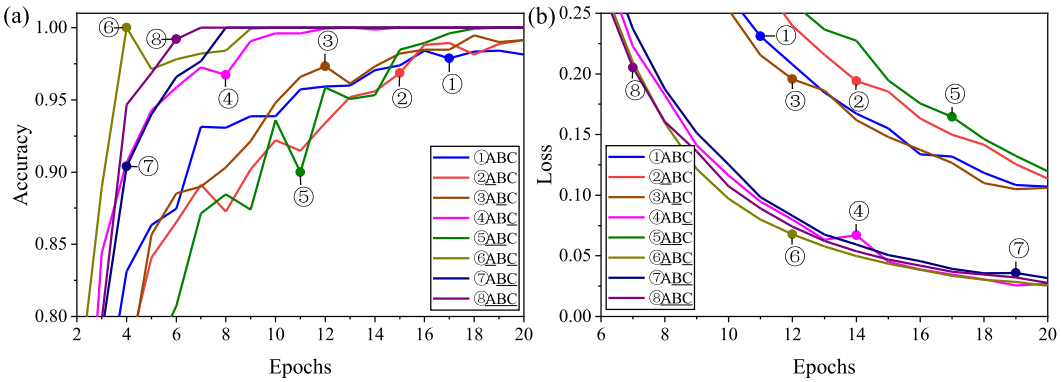


Fig. 12. Classification accuracy and loss of different improvement strategies in the feature extraction layer. (a) Accuracy. (b) Loss. A, B, and C represent different strategies used by the model, where A represents bias, B represents the Softsign activation function, C represents average pooling, A represents removing bias, B represents the removing activation function, and C represents the removing pooling layer.

1) Perspective of Feature Extraction Mechanism: The S-GAP classification layer of the 1-D CNN model is fixed, and then, the effectiveness of different improvement strategies of bias, activation, and pooling in the feature extraction layer is studied.

As shown in Fig. 12 and Table IV (A, B, and C), their different combinations can accelerate the convergence speed of the model and improve the classification accuracy. The combination with A can also reduce the model parameters. The specific analysis is given as follows.

Firstly, for single improvement strategies A, B and C, C is better than B, B is better than A, and A is better than ABC in improving the performance of 1-D CNN. It shows that the removals of pooling, bias, and activation functions make the features extracted by the model more effective for classification, and the removal of pooling is more effective.

Second, for the combination improvement strategy, the proposed model combining the three strategies has the best performance, that is, ABC has the fastest convergence speed and the shortest operation time while ensuring diagnostic accuracy.

The reasons for the above results are given as follows. On the one hand, the downsampling operation of the pooling layer weakens the fault features; especially, the average pooling discards the signal features with weaker amplitude, and often, the weak features are very important for fault pattern recognition. On the other hand, when using bias and activation functions for nonlinear learning, it is possible to discard weak fault features (such as bias) and learn unnecessary nonfault features (such as activation functions).

2) Perspective of Influence of Classification Results: This section mainly studies the influence of different combinations of improvement strategies of the feature extraction layer and classification layer on the performance of 1-D CNN. By comparing and analyzing the combined application effects of the improved strategies of the feature extraction layer and classification layer in Fig. 13 and Table V, the following three conclusions can be obtained.

First, the GAP classification layer is retained, and only the 1-D CNN feature extraction layer is improved; that is, the bias, activation function, and pooling layer are removed. The improved model has fewer calculation parameters and fast operation speed, but the classification accuracy is low. This is because, although the bias, activation function, and pooling layer may lose some fault features, they are obviously easier to extract nonlinear features and avoid overfitting, thereby improving the generalization performance of the model.

Second, the bias, activation function, and pooling layer of the feature extraction layer are retained, and only the 1-D CNN classification layer is improved, that is, the signal features extracted by the convolutional layer are squared. The convergence time of the improved model is prolonged, but the diagnostic effect is significantly improved. This is because the S-GAP that can strengthen the time-domain envelope of

TABLE IV

INFLUENCE OF DIFFERENT IMPROVEMENT STRATEGIES OF FEATURE EXTRACTION LAYER ON THE PERFORMANCE OF 1-D CNN, AMONG WHICH ABC LABELED RED HAS THE BEST PERFORMANCE. A, B, AND C REPRESENT DIFFERENT STRATEGIES USED BY THE MODEL, WHERE A REPRESENTS BIAS, B REPRESENTS SOFTSIGN ACTIVATION FUNCTION, C REPRESENTS AVERAGE POOLING, A REPRESENTS REMOVING BIAS, B REPRESENTS REMOVING ACTIVATION FUNCTION, AND C REPRESENTS REMOVING POOLING LAYER

No.	Strategy Combination	Parameter Numbers	Convergence Epochs	Convergence Times(s)	Accuracy	Epoch=20 Loss	Output Length
1	ABC	320	20	17.97	98.1%	0.1071	2048
2	ABC	310	20	16.98	99.1%	0.1135	2048
3	ABC	320	20	17.52	99.1%	0.1061	2048
4	ABC	320	15	14.58	100%	0.027	4096
5	ABC	310	19	15.90	100%	0.1197	2048
6	ABC	310	10	11.97	100%	0.0253	4096
7	ABC	320	8	8.62	100%	0.0317	4096
8	ABC	310	7	8.60	100%	0.0276	4096

TABLE V

INFLUENCE OF DIFFERENT COMBINATIONS OF FEATURE EXTRACTION LAYER AND CLASSIFICATION LAYER ON THE PERFORMANCE OF 1-D CNN. AMONG THEM, THE KG MARKED IN BLUE HAS THE LOWEST CLASSIFICATION ACCURACY, AND THE KG MARKED IN RED HAS THE BEST PERFORMANCE. K REPRESENTS THE RETENTION OF BIAS, ACTIVATION FUNCTION, AND POOLING LAYER; K REPRESENTS THE REMOVAL OF BIAS, ACTIVATION FUNCTION, AND POOLING LAYER; G REPRESENTS THE MODEL USING GAP CLASSIFICATION LAYER; AND G REPRESENTS THE MODEL USING S-GAP CLASSIFICATION LAYER

No.	Strategy Combination	Parameter Numbers	Convergence Epochs	Convergence Times(s)	Accuracy	Epoch=20 Loss	Output Length
1	KG	320	20	18.56	82.1%	0.4506	2048
2	KG	310	20	16.58	20.9%	2.0354	4096
3	KG	320	20	19.85	97.2%	0.1141	2048
4	KG	310	8	9.38	100%	0.0305	4096

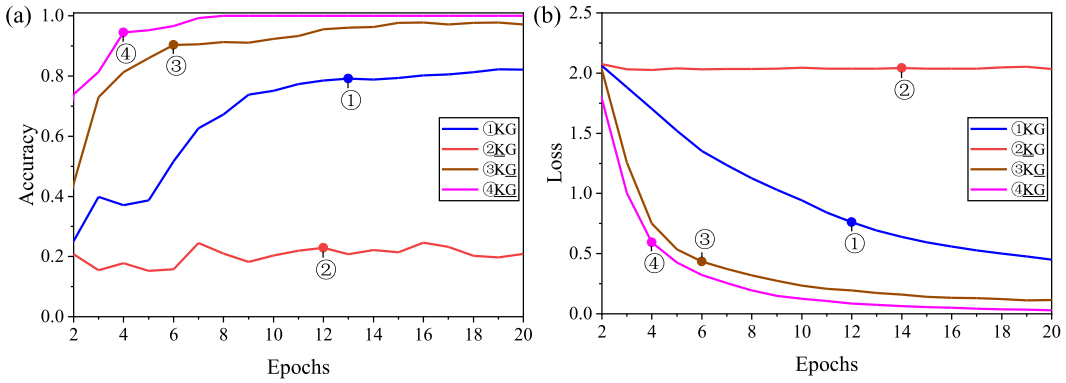


Fig. 13. Classification accuracy and loss of different combinations of feature extraction layer and classification layer. (a) Accuracy. (b) Loss. K represents the retention of bias, activation function, and pooling layer; K represents the removal of bias, activation function, and pooling layer; G represents the model using the GAP classification layer; and G represents the model using the S-GAP classification layer.

the signal is more conducive to extracting the instantaneous frequency characteristics, so the diagnosis effect is better.

Finally, the feature extraction layer and the classification layer of 1-D CNN are improved, which not only reduces the model parameters and improves the operation speed and classification accuracy but also converges faster and more stable. It can be seen that the improvement based on the S-GAP classification layer can make up for the overfitting problem caused by the optimization of the feature extraction layer, while the optimization of the feature extraction layer retains the signal features to the greatest extent and improves the lightweight performance of the ELCNN.

On the whole, improving the 1-D CNN feature extraction layer or the S-GAP classification layer alone can effectively

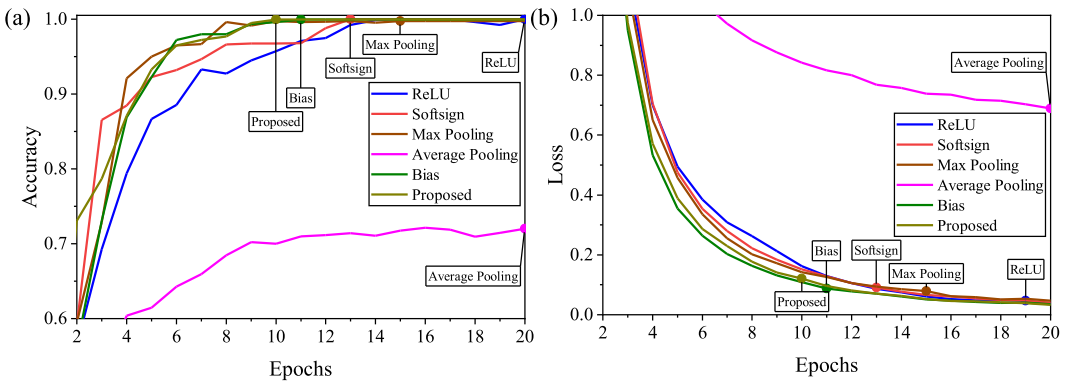


Fig. 14. Classification accuracy and loss of different strategies in the feature extraction layer. (a) Accuracy. (b) Loss.

TABLE VI
INFLUENCE OF DIFFERENT STRATEGIES OF FEATURE EXTRACTION LAYER ON THE PERFORMANCE OF 1-D CNN (MAXIMUM POOLING REFERRED TO AS MP AND AVERAGE POOLING REFERRED TO AS AP). AMONG THEM, THE AP MARKED IN BLUE HAS THE LOWEST CLASSIFICATION ACCURACY, AND THE PROPOSED MODEL MARKED IN RED HAS THE BEST OVERALL PERFORMANCE

No.	Strategy	Parameter Numbers	Convergence Epochs	Convergence Times(s)	Accuracy	Epoch=20 Loss	Output Length
1	ReLU	310	19	17.50	99.2%	0.0363	4096
2	Softsign	310	13	12.08	100%	0.0407	4096
3	MP（size=4）	310	15	12.76	99.7%	0.0466	1024
4	AP（size=4）	310	20	15.26	72.0%	0.6893	1024
5	Bias	320	11	11.69	99.9%	0.0355	4096
6	Proposed	310	10	9.49	99.9%	0.0332	4096

improve the fault diagnosis performance, but the model performance is optimal after improving the feature extraction layer and the classification layer at the same time.

D. Verification of the Improved Strategies' Superiority

This section distinguishes the feature extraction layer and the classification layer, and compares the proposed improvement strategy with other similar strategies to verify the superiority of the proposed improvement strategy.

1) Feature Extraction Layer: Five different feature extraction layer strategies are introduced, and the corresponding parameters in the proposed model are replaced by the control variable method to analyze their influence on the performance of ELCNN. The classification accuracy, loss, and overall performance of the model after applying various strategies are shown in Fig. 14 and Table VI. First, according to the number of parameters, the bias uses more parameters. Second, from the perspective of convergence time, the proposed model consumes the shortest time of 9.49 s. Finally, from the perspective of classification accuracy, Softsign, bias, and proposed strategies have better diagnostic results, while the other strategies are relatively worse, especially average pooling has the lowest classification accuracy. This is because the ReLU of the convolutional layer will lose the negative component of the signal, and the removal of bias and activation can significantly improve the computational efficiency of the model. The maximum pooling operation of the pooling layer ignores the low amplitude, and the average pooling weakens the fault impact excitation, resulting in low classification accuracy of the model.

On the whole, the proposed improved ELCNN neither uses bias and activation functions for nonlinear learning nor does it use pooling to reduce the feature dimension and prevent overfitting. However, it has fewer fault diagnosis steps, fewer parameters, shorter time, and better model performance.

2) Classification Layer: Six different classification layer improvement strategies, including Flatten, Flatten+ReLU, GAP, ReLU+GAP, S-GAP(3) based on three-layer convolution, and S-GAP(1) based on one-layer convolution, are selected to discuss the influence of classification layers on the performance of 1-D CNN.

As shown in Fig. 15 and Table VII, the GAP classification layer can effectively reduce the number of parameters. In particular, S-GAP(1) has the least number of convolution layers, the least parameters, the least convergence time, the least classification loss, and an accuracy rate of 100%. There are three main reasons why S-GAP(1) is superior to other methods.

First, the amplitude of the signal may be positive or negative. The Flatten and GAP operations retain the amplitude symbol, so the offset of the positive and negative values will reduce the classification probability of the SoftMax calculation.

Second, the positive operation in Flatten + ReLU and ReLU+GAP can avoid the problem of positive and negative offset, but it will remove the negative value information of the

TABLE VII
INFLUENCE OF DIFFERENT CLASSIFICATION LAYERS ON THE PERFORMANCE OF 1-D CNN. AMONG THEM, THE CLASSIFICATION ACCURACY OF THE MODEL MARKED IN BLUE USING THE FLATTEN OR GAP STRATEGY IS LOW, AND THE S-GAP(1) LABELED AS RED HAS THE BEST OVERALL PERFORMANCE

No.	Strategy Combination	Layers	Parameter Numbers	Convergence Epochs	Convergence Times(s)	Accuracy	Epoch=20 Loss
1	Flatten	3	413,090	20	71.90	27.5%	3.8057
2	Flatten+ReLu	3	413,090	20	66.67	94.9%	0.5380
3	GAP	3	3,590	20	64.83	17.8%	2.0415
4	ReLu+GAP	3	3,590	20	66.09	99.5%	0.1843
5	S-GAP(3)	3	3,590	7	25.95	100%	0.0330
6	S-GAP(1)	1	310	11	11.39	100%	0.0319

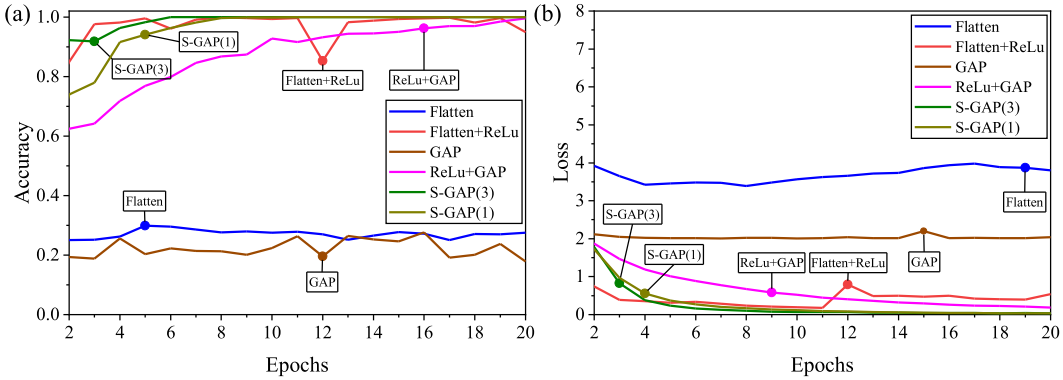


Fig. 15. 1-D CNN classification accuracy and loss of different classification layers. (a) Accuracy. (b) Loss.

signal. However, the time-domain information of the signal is squared by S-GAP, which can perfectly solve the above problems.

Finally, although the diagnostic effect of S-GAP(3) is good, some fault features will be lost, and the operation efficiency will be reduced because of multilayer convolution for nonlinear learning. On the whole, the overall performance of the model using S-GAP(1) is better.

IV. EVALUATION OF DIAGNOSTIC PERFORMANCE OF ELCNN

To comprehensively evaluate the advancement of the ELCNN model, eight improved representative 1-D CNNs were selected, and comparative experiments were conducted based on CWRU and XJTU from three aspects: overall model performance, antinoise performance, and feature interpretability.

A. Different Improved 1-D CNN Models

As shown in Fig. 16, MBMSCNN is a multibranch and multiscale 1-D CNN. The signal features are extracted by convolution kernels of different sizes and further fused and transformed by concatenation and pooling operations. The general CNN (GCNN) uses GAP to replace the FCL of 1-D CNN (see Fig. 17).

It is a good idea to improve the GCNN network structure and optimize the parameters for vibration signals. For example, Li et al. [29] selected a Softsign activation function and removed the pooling layer to improve a 1-D CNN (explainable CNN-1 and ECNN-1). Kim et al. [30] replaced FC with global power pooling (the output signal of the convolutional layer is first calculated by square, then GAP, and S-GAP) to improve the classification layer (explainable CNN-2 and ECNN-2).

As shown in Fig. 18, we propose an improved 1-D CNN (ELCNN) based on S-GAP. The main work is given as follows.

1) We remove unnecessary biases, activation functions, and pooling layers.

2) We construct only one layer of convolution and select the convolution kernel size and the number of filters that make the model perform well, and the learning features are better.

3) We use S-GAP to replace the FCL and perform one to three square operations on the output of the convolutional layer, where ELCNN-3 represents three-layer convolution, ELCNN-1x denotes one-layer convolution, and x denotes the square operation x times.

Table VIII shows the network structure parameters of the eight models. There are three kinds of convolution kernel sizes of MBMSCNN, which are 1×1, 8×1, and 16×1, respectively. The convolution kernel size of the remaining 1-D CNNs is 16×1, and the number of filters is 10. In addition, GCNN uses the dropout technique, 1S-GAP in ELCNN refers to the square operation of the features learned in the last convolutional layer, 2S-GAP refers to two square operations (fourth power), and 3S-GAP refers to three square operations (eighth power).

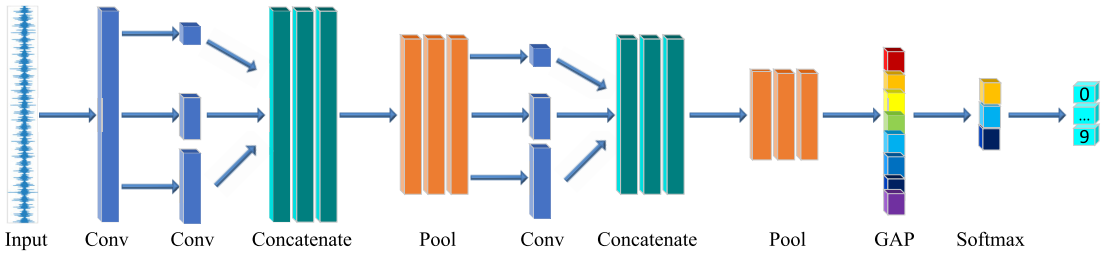


Fig. 16. Architecture of MBMSCNN.

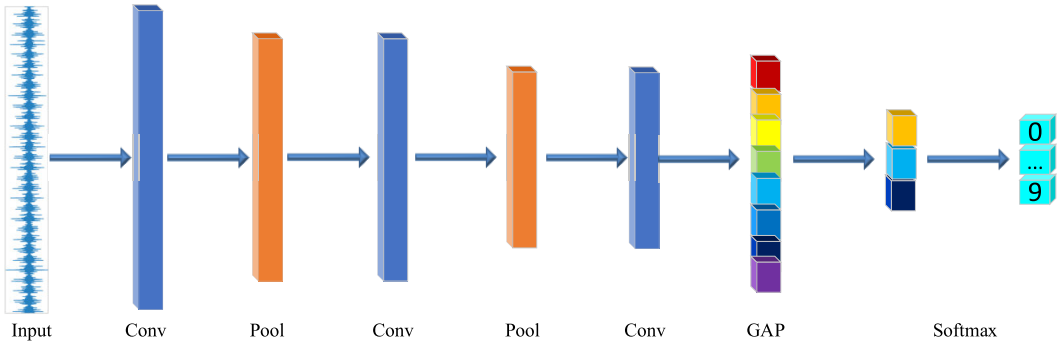


Fig. 17. Architecture of GCNN.

TABLE VIII
NETWORK STRUCTURE PARAMETERS OF EIGHT TYPES OF 1-D CNNS

No.	Network	Layers	Use bias	Activation	Pooling	Classification	Multiple Branch
1	MBMSCNN	3	True	ReLU	Average	GAP	Yes
2	GCNN	3	True	ReLU	Average	GAP	No
3	ECNN-1	3	True	Softsign	None	GAP	No
4	ECNN-2	3	True	Softsign	None	1S-GAP	No
5	ELCNN-3	3	False	None	None	1S-GAP	No
6	ELCNN-11	1	False	None	None	1S-GAP	No
7	ELCNN-12	1	False	None	None	2S-GAP	No
8	ELCNN-13	1	False	None	None	3S-GAP	No

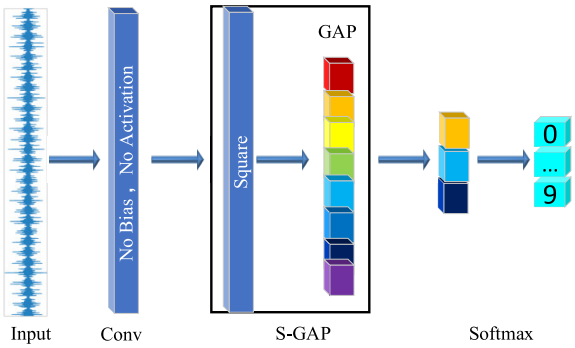


Fig. 18. Architecture of ELCNN.

B. Model Performance Evaluation Based on CWRU

Based on the CWRU, this section compares and analyzes eight types of improved 1-D CNN from three aspects: model overall performance, antinoise performance, and feature interpretability.

1) Overall Performance: As shown in Fig. 19(a) and Table IX, comparing the convergence of the classification accuracy of each 1-D CNN, we have the following findings.

First, the classification accuracy of ECNN-1 without pooling is higher than that of GCNN with pooling.

Second, compared with the ECNN-1 using GAP, the ECNN-2 using S-GAP has a shorter and more stable convergence time. This is because the square operation transforms the original signal into a similar power spectrum, which can better highlight the peak frequencies.

Third, the improved ELCNN-3, ELCNN-11, and ELCNN-12 classification accuracy curves are smoother and less volatile than other models. It shows the advancement of the improved model; especially, ELCNN-12 has fewer calculation parameters, shorter running time, and smaller weight files.

Finally, the classification accuracy of ELCNN-13 is reduced, and the operation time is prolonged, indicating that the

TABLE IX
PERFORMANCE COMPARISON OF EIGHT TYPES OF 1-D CNNS ON THE CWRU BEARING DATASET, WHERE EPOCHS AND TIMES ARE THE NUMBER OF STEPS AND TIME REQUIRED FOR THE MODEL TO CONVERGE, RESPECTIVELY. AMONG THEM, ELCNN-12 MARKED IN RED HAS THE BEST PERFORMANCE

No.	Type	Layers	Parameter Numbers	Epochs	Times(s)	Accuracy	Epoch=20 Loss	Output Length	Weight Files Size (KB)
1	MBMSCNN	3	10,820	18	142.74	100%	0.0388	1024	273.0
2	GCNN	3	3,620	20	40.30	99.6%	0.0324	1024	117.0
3	ECNN-1	3	3,620	18	65.06	100%	0.0265	4096	114.0
4	ECNN-2	3	3,620	10	39.66	100%	0.0287	4096	115.0
5	ELCNN-3	3	3,590	7	25.34	100%	0.0547	4096	107.0
6	ELCNN-11	1	310	14	14.43	100%	0.0548	4096	39.6
7	ELCNN-12	1	310	9	9.20	100%	0.0140	4096	40.4
8	ELCNN-13	1	310	20	18.46	91.7%	1.7453	4096	42.0

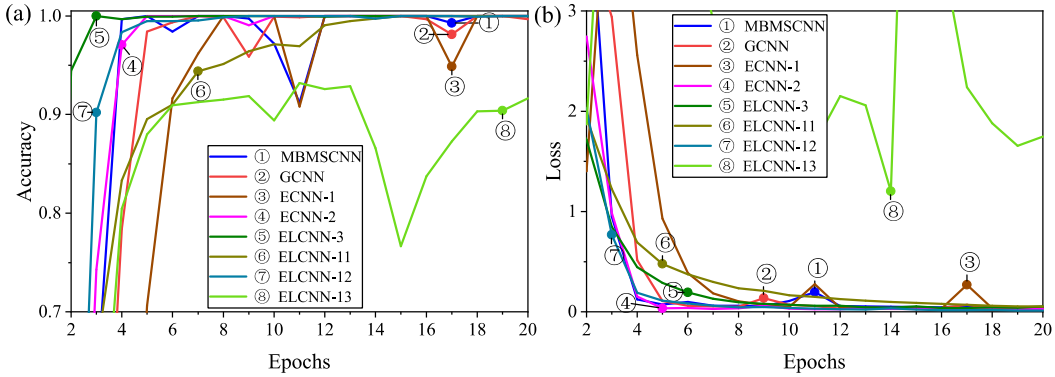


Fig. 19. Classification accuracy and loss of eight kinds of 1-D CNNs on the CWRU dataset. (a) Accuracy. (b) Loss.

square operation is not proportional to the performance of the model.

In summary, compared with the traditional 1-D CNN using MSC, pooling layer, bias, and activation function (Softsign and ReLU), the proposed ELCNN framework, especially ELCNN-12, has a better performance and lighter weight [46].

2) Antinoise Performance: As shown in Table X, the noised signals are simulated with additive Gaussian white noises that are proportional to the clean signals [8], and the fault signals with SNR of 15, 10, 5, 0, and −5 dB are selected to evaluate the noise immunity of the models.

Through the comparative analysis of the classification accuracy, three conclusions are drawn.

First, compared with other models, the antinoise performance of GCNN without MSC and only pooling decreases most obviously, which indicates that pooling loses signal features while suppressing noises.

Second, the antinoise performance of ECNN-1, ECNN-2, and ECNN-3 in the orange part of Table X is special. When the SNR is 15 dB, the antinoise performance of ECNN-1 is less than that of ECNN-2 and less than that of ELCNN-3. When the SNR is less than 15 dB, the antinoise performance of ECNN-1 is greater than that of ECNN-2 and greater than that of ELCNN-3. It shows that, when the signal is interfered with by strong noise, compared with ECNN-1, ECNN-2 and ELCNN-3 with multilayer convolution using S-GAP enhanced features have no obvious advantages; especially, the classification accuracy of ELCNN-3, which removes bias and activation functions, decreases significantly.

Finally, the proposed ELCNN using single-layer convolution for the feature extraction layer and removing bias, activation, and pooling and using S-GAP for the classification layer to enhance the time-domain envelope of the signal is more antinoise. In particular, ELCNN-12 still has a classification accuracy of 78.76% when the SNR is 5 dB, and the model deviation is smaller (±0.1).

On the whole, CWRU is easily affected by noise. The improved multilayer convolution model is not as good as the traditional MSC due to the extraction of nonlinear features and the introduction of many noises. Single-layer convolution can overcome the above problems and improve the antinoise performance of the model.

3) Feature Interpretability: This section mainly analyzes the interpretability of the features extracted by various improved 1-D CNNs to IR, OR, and B faults from two aspects of statistical characteristics and frequency characteristics.

TABLE X
CLASSIFICATION ACCURACY OF EIGHT KINDS OF 1-D CNNS FOR CWRU BEARING FAULTS WITH DIFFERENT SNRS. AMONG THEM, ELCNN-12 LABELED AS THE RED PART HAS THE BEST PERFORMANCE, AND THE ORANGE PARTS ARE 1-D CNNS THAT NEED TO BE SPECIFICALLY ANALYZED

SNR / Method	15dB	10dB	5dB	0dB	-5dB
MBMSCNN	95.11±4.05	77.79±15.40	49.25±11.37	29.67±9.96	11.43±3.13
GCNN	84.33±10.67	63.39±16.09	31.28±9.67	17.2±7.21	10.11±0.14
ECNN-1	97.46±2.29	82.23±17.23	40.13±8.35	16.47±4.94	10.05±0.12
ECNN-2	98.63±0.99	74.43±19.06	29.55±10.00	15.70±4.08	11.73±3.88
ELCNN-3	99.38±0.60	69.59±8.81	24.39±6.12	15.41±5.00	10.52±1.18
ELCNN-11	99.88±0.11	83.78±8.84	50.27±5.29	23.66±4.01	13.18±1.47
ELCNN-12	99.95±0.05	98.18±0.33	78.76±0.10	37.64±0.21	15.78±0.57
ELCNN-13	99.7±0.09	92.65±0.24	70.82±0.50	18.01±0.21	10.10±0.08

TABLE XI
STATISTICAL CHARACTERISTICS OF IR FAULTS IN THE CRWU DATASET LEARNED BY THE LAST LAYER CONVOLUTION OF EIGHT TYPES OF 1-D CNNS. THE KURTOSIS IS USED TO REFLECT THE SHARPNESS OF THE VIBRATION SIGNAL WAVEFORM, AND THE PULSE FACTOR IS USED TO EVALUATE THE IMPACT AND TRANSIENT CHARACTERISTICS OF THE VIBRATION SIGNAL. THE BLUE PART INDICATES THAT THE SHOCK EXCITATION OF THE MODEL IS WEAK, AND THE RED PART INDICATES THAT THE SHOCK EXCITATION OF THE MODEL IS THE STRONGEST

Characteristics	Raw	MBMS CNN	GCNN	ECNN-1	ECNN-2	ELCNN-3	ELCNN-11	ELCNN-12	ELCNN-13
Kurtosis	3.87	-1.18	-0.52	-0.20	14.12	4.56	12.80	63.71	404.06
Pulse Factor	9.44	1.98	1.73	3.08	10.97	7.30	10.88	50.36	226.46

a) IR: Bearing defects often cause impact excitation. To measure the difference of impact excitation learned by different models, kurtosis and pulse factors are introduced to quantitatively analyze the fault of IR.

As shown in Table XI, MBMSCNN, GCNN, and ECNN-1 weaken the impulse excitation, and the remaining models use S-GAP to strengthen the signal characteristics (among which ELCNN-13 has the most obvious effect). Among them, compared with ELCNN-3 without bias and activation function, the impact excitation of ECNN-2 with bias and activation function is more obvious, but it does not mean that its model performance is better. This is because, although the impact excitation reflects the sensitivity of the model to the learned fault to a certain extent, the features learned by the model are not necessarily fault features. Therefore, it is necessary to analyze the interpretability of the extracted features based on the frequency characteristics. Fig. 20 shows the spectrum characteristics of the IR fault learned by the eight types of models, in which the rotating frequency is 29.95 Hz and the IR fault frequency is 162.19 Hz. By analyzing the spectral characteristics of the model in Figs. 20 and 21, the following conclusions can be drawn.

First, MBMSCNN and GCNN with pooling layer pay more attention to low-frequency features (29.29 Hz) and discard many fault features (IR fault frequency: 162.19 Hz and double frequency: 322.26 Hz) to weaken impact excitation. ECNN-1 and ECNN-2 without pooling can learn more fault features.

Second, although ECNN-1 and ECNN-2 without the pooling layer can learn more features and they have a large deviation from the Raw signal (preventing overfitting), the bias and activation function lead to the loss of some features learned by the model (rotating frequency: 29.29 Hz), and not all the features learned are needed for fault diagnosis. For example, ECNN-2 with more obvious impact excitation in Fig. 21 pays more attention to high-frequency nonfault characteristics than ELCNN-3 without bias and activation function, while bearing defects often cause low-frequency impact excitation [47]. It shows that bias and activation functions can prevent model overfitting and improve nonlinear learning ability, but, at the same time, some features will be discarded, and nonfault high-frequency features will be learned.

Third, compared with ECNN-2, the features learned by the three-layer convolutional ELCNN-3 are rough, while the single-layer convolutional ELCNN-11 can better fit the low-frequency fault features of the signal (such as rotating frequency: 29.29 Hz and sideband: 131.8 Hz). It shows that the feature extraction ability of the proposed single-layer convolution model is better than that of the multilayer convolution model, and the feature learning effect of the model is better after removing the bias and activation function.

Finally, ELCNN-12 and ELCNN-13 with multiple square operations learn more harmonic features (131.8 and 193.4 Hz) and cage fault frequency (11.7 Hz). It shows that S-GAP can indeed strengthen the FM and AM characteristics.

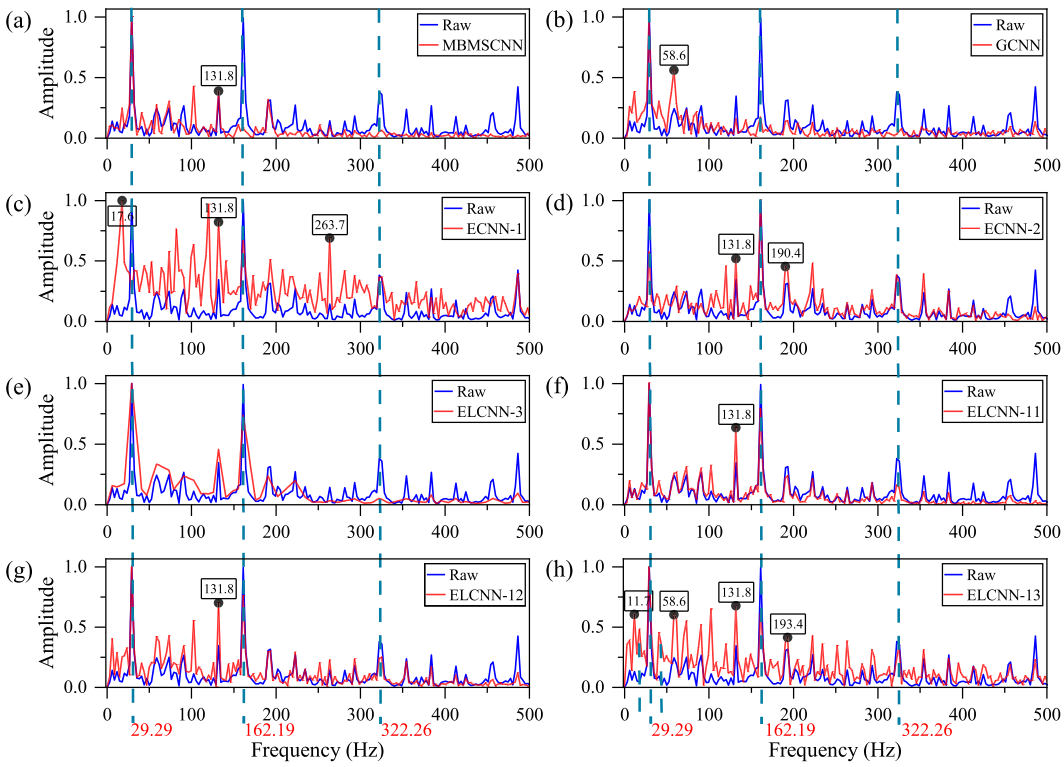


Fig. 20. Frequency-domain features of IR faults in the CRWU dataset learned by the last layer convolution of eight types of 1-D CNNs. (a) MBMSCNN. (b) GCNN. (c) ECNN-1. (d) ECNN-2. (e) ELCNN-3. (f) ELCNN-11. (g) ELCNN-12. (h) ELCNN-13.

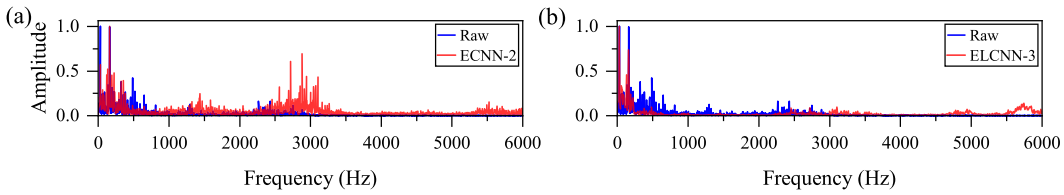


Fig. 21. Frequency-domain features of the IR fault of the CRWU learned by the last layer convolution of (a) ECNN-2 and (b) ELCNN-3, where ECNN-2 uses biases and activation functions, while ELCNN-3 does not use.

In particular, the model using multiple square operations not only effectively avoids overfitting but also has a better impact on incentive learning and more interpretability for extracting features.

On the whole, the IR fault is modulated by the rotating frequency of 29.29 Hz and the cage fault of 11.7 Hz at the same time, which has a very complex phenomenon of FM and AM.

b) OR: Similar to the IR fault, the analysis of the OR fault (108.4 Hz) in Figs. 22 and 23 shows that the single-layer convolutional ELCNN without bias and activation function can better fit the Raw signal characteristics (e.g., ELCNN-11 learned 108.4-Hz fault frequency and 46.9-Hz sideband). In particular, ELCNN-12 and ELCNN-13 with multiple square operations learn more FM features (such as double frequency: 213.87 Hz), AM features (such as 137.7 Hz), and cage fault frequency (11.7 Hz).

The OR fault also has a complex FM and AM phenomenon, which is modulated by the rotating frequency of 29.03 Hz and the cage fault frequency of 11.7 Hz. For example, the rotating frequency of 29.03 Hz is modulated by the cage fault to obtain a frequency of 17.6 Hz; the fault frequency of 108.4 Hz and its double frequency of 213.87 Hz are modulated by the rotating frequency of 29.03 Hz to produce 137.7, 184.6, and so on.

c) B: As shown in Fig. 24, 1-D CNN feature extraction characteristics similar to IR and OR faults can also be obtained from B faults (141.17 Hz). In multilayer convolution, GCNN with a pooling layer loses fault features (fault frequency greater than 100 Hz), while ECNN-1 and ECNN-2 without a pooling layer learn high-frequency nonfault features due to bias and activation function [see Fig. 25(a)]. The proposed single-layer convolutional ELCNN can better fit ball faults (such as ELCNN-11), and the models ELCNN-12 and ELCNN-13 with multiple square operations can learn more FM and AM features.

The fault frequency of B is 141.17 Hz, which is not obvious. It is more affected by the 131.84-Hz frequency modulated by the cage fault frequency of 11.7 Hz. This may be caused by the mutual modulation of the rotating frequency of 29.30 Hz and the cage fault frequency of 11.7 Hz. For example, the sidebands of 131.84 Hz (131.84-11.7 = 120.14 Hz) and 161.1 Hz (161.1-11.7 = 149.4 Hz) modulated by 11.7 Hz

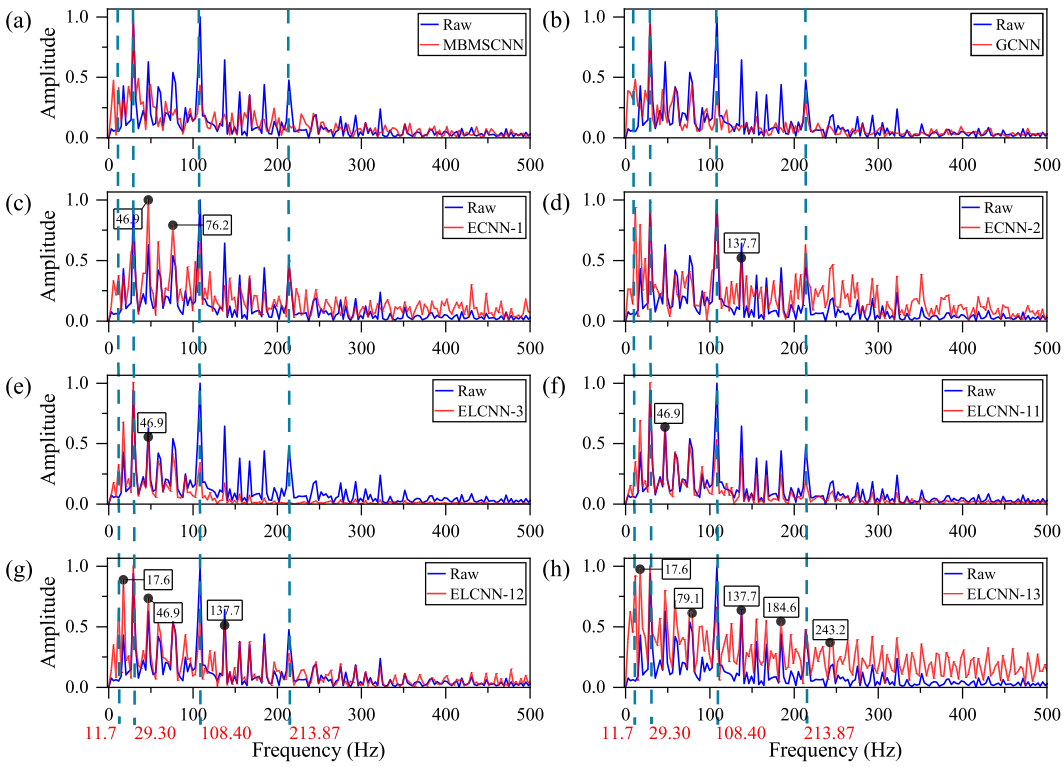


Fig. 22. Frequency-domain features of OR faults in the CRWU dataset learned by the last layer convolution of eight types of 1-D CNNs. (a) MBMSCNN. (b) GCNN. (c) ECNN-1. (d) ECNN-2. (e) ELCNN-3. (f) ELCNN-11. (g) ELCNN-12. (h) ELCNN-13.

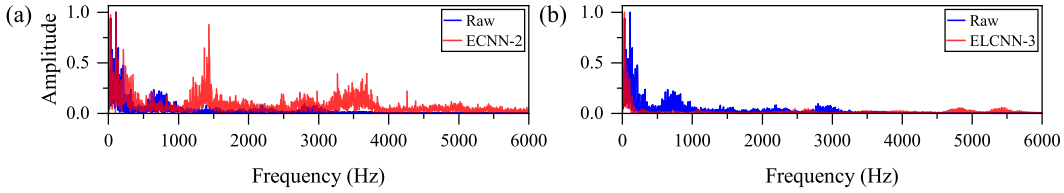


Fig. 23. Frequency-domain features of the OR fault of the CRWU learned by the last layer convolution of (a) ECNN-2 and (b) ELCNN-3, where ECNN-2 uses biases and activation functions, while ELCNN-3 does not use.

learned by the ELCNN-13 model are also 4 times and 5 times of the rotation frequency of 29.30 Hz. Therefore, the ball fault has weaker and more complex FM and AM features than IR and OR faults.

C. Model Performance Evaluation Based on XJTU

To ensure the credibility of the experimental results, the XJTU is selected, and the comparative experiments are carried out again on the eight types of improved 1-D CNNs. In addition, considering the similarity of data, this section only analyzes one of the working conditions, that is, IR and OR faults with a rotating frequency of 39.05 Hz.

1) Overall Performance: As shown in Fig. 26 and Table XII, we can see the classification accuracy and model performance of various improved 1-D CNNs.

First, the performance of GCNN with a pooling layer is better than that of ECNN-1, ECNN-2, and ELCNN-3 without a pooling layer, and the convergence is faster and more stable. Although the convergence effect of ECNN-2 that uses S-GAP to enhance signal features and ELCNN-3 that removes bias and activation functions has been improved, the classification accuracy of these models has not been significantly improved. It shows that the classification characteristics of the XJTU are obvious. The multilayer convolution model uses the pooling layer to achieve classification more easily. After removing the pooling, bias, and activation functions, the model learning is not ideal, and the convergence is not stable.

Second, compared with the multilayer convolutional model, the proposed single-layer convolutional model ELCNN, especially ELCNN-12, can learn signal features more smoothly and avoid overfitting. Moreover, it has better lightweight performance (fewer calculation parameters, shorter running time, and smaller weight file), which further illustrates the advanced nature of the proposed model.

2) Antinoise Performance: As shown in Tables X and XIII, the classification accuracy based on XJTU is significantly higher than that of CWRU under the same models and the same SNR fault conditions. It shows that the signal characteristics of the XJTU are more obvious and less affected by noise. Two conclusions can be drawn from Table XIII.

First, ECNN-1 has the lowest classification accuracy because only the pooling layer is removed, which brings noise

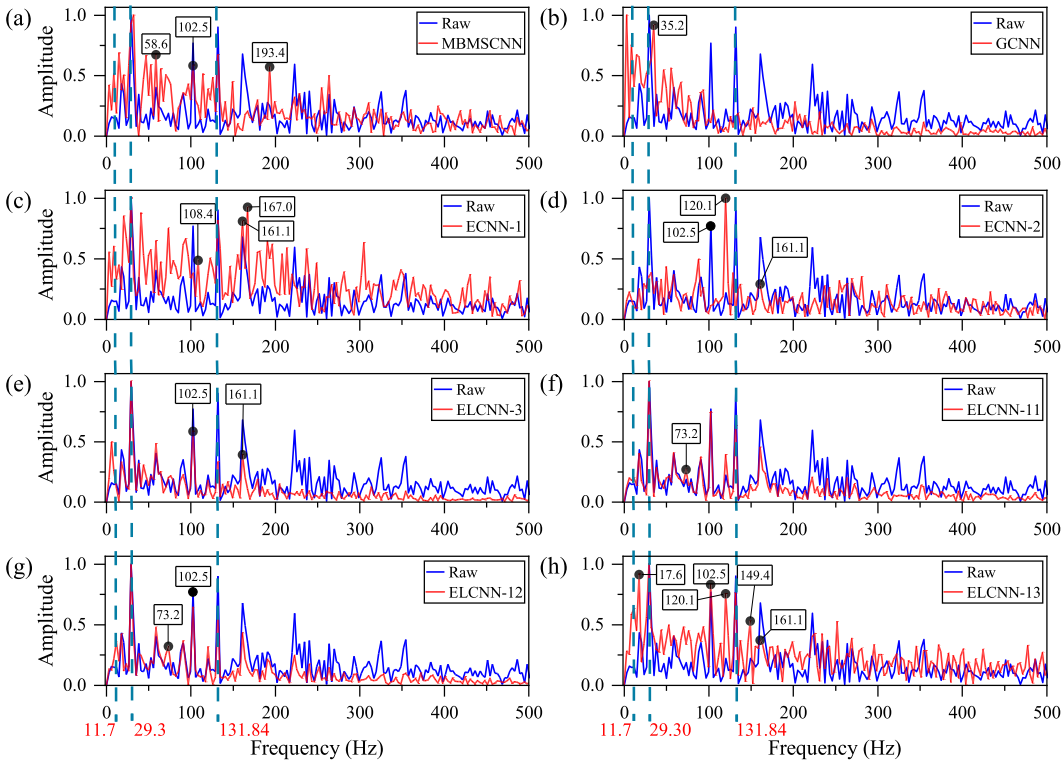


Fig. 24. Frequency-domain features of B faults in the CRWU dataset learned by the last layer convolution of eight types of 1-D CNNs. (a) MBMSCNN. (b) GCNN. (c) ECNN-1. (d) ECNN-2. (e) ELCNN-3. (f) ELCNN-11. (g) ELCNN-12. (h) ELCNN-13.

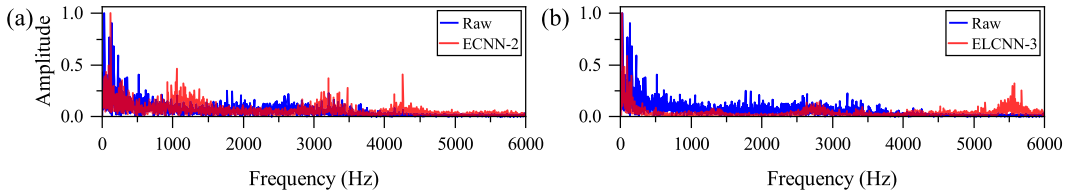


Fig. 25. Frequency-domain features of the B fault of the CRWU learned by the last layer convolution of (a) ECNN-2 and (b) ELCNN-3, where ECNN-2 uses biases and activation functions, while ELCNN-3 does not use.

TABLE XII

PERFORMANCE COMPARISON OF EIGHT KINDS OF 1-D CNNS ON THE XJTU DATASET, WHERE EPOCHS AND TIMES ARE THE NUMBER OF STEPS AND TIME REQUIRED FOR THE MODEL TO CONVERGE, RESPECTIVELY. THE CLASSIFICATION ACCURACY OF ECNN-2, ECNN-3, AND ELCNN-13 MARKED IN BLUE IS LOW, WHILE THE COMPREHENSIVE PERFORMANCE OF ELCNN-12 MARKED IN RED IS THE BEST

No.	Type	Layers	Parameter Numbers	Epochs	Times(s)	Accuracy	Epoch=20 Loss	Output Length	Weight Files Size (KB)
1	MBMSCNN	3	11,099	13	196.90	100%	0.0364	1024	276.0
2	GCNN	3	3,719	18	68.99	100%	0.0241	1024	118.0
3	ECNN-1	3	3,719	20	134.43	100%	0.0331	4096	114.0
4	ECNN-2	3	3,719	20	147.29	99.8%	0.0366	4096	115.0
5	ELCNN-3	3	3,689	18	114.09	99.8%	0.0174	4096	109.0
6	ELCNN-11	1	409	7	11.75	100%	0.0196	4096	41.6
7	ELCNN-12	1	409	5	10.39	100%	0.0087	4096	43.0
8	ELCNN-13	1	409	20	33.29	97.5%	0.1862	4096	43.6

interference while retaining the signal characteristics. S-GAP is needed to enhance the fault features.

Second, when the signal is interfered with by strong noise, the ECNN-2 and ELCNN models using S-GAP enhanced features still have good performance when the SNR is 5 dB for the XJTU with obvious features. It shows that the S-GAP model can effectively suppress high-frequency noise.

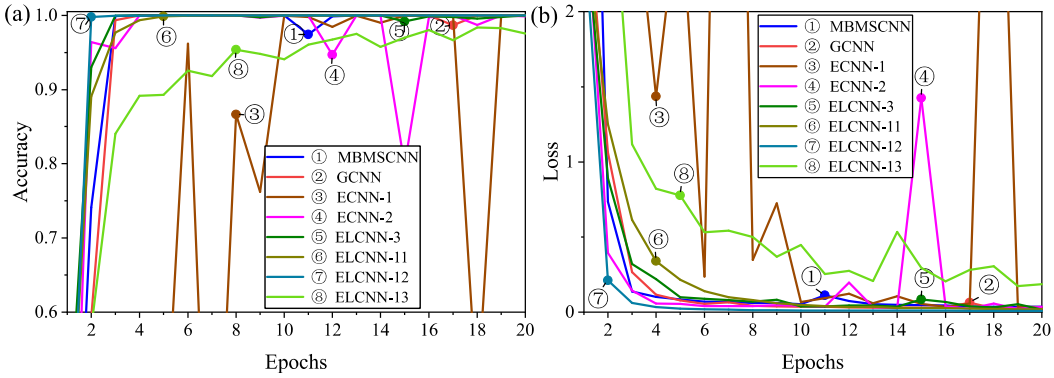


Fig. 26. Classification accuracy and loss of eight kinds of 1-D CNNs on the XJTU dataset. (a) Accuracy. (b) Loss.

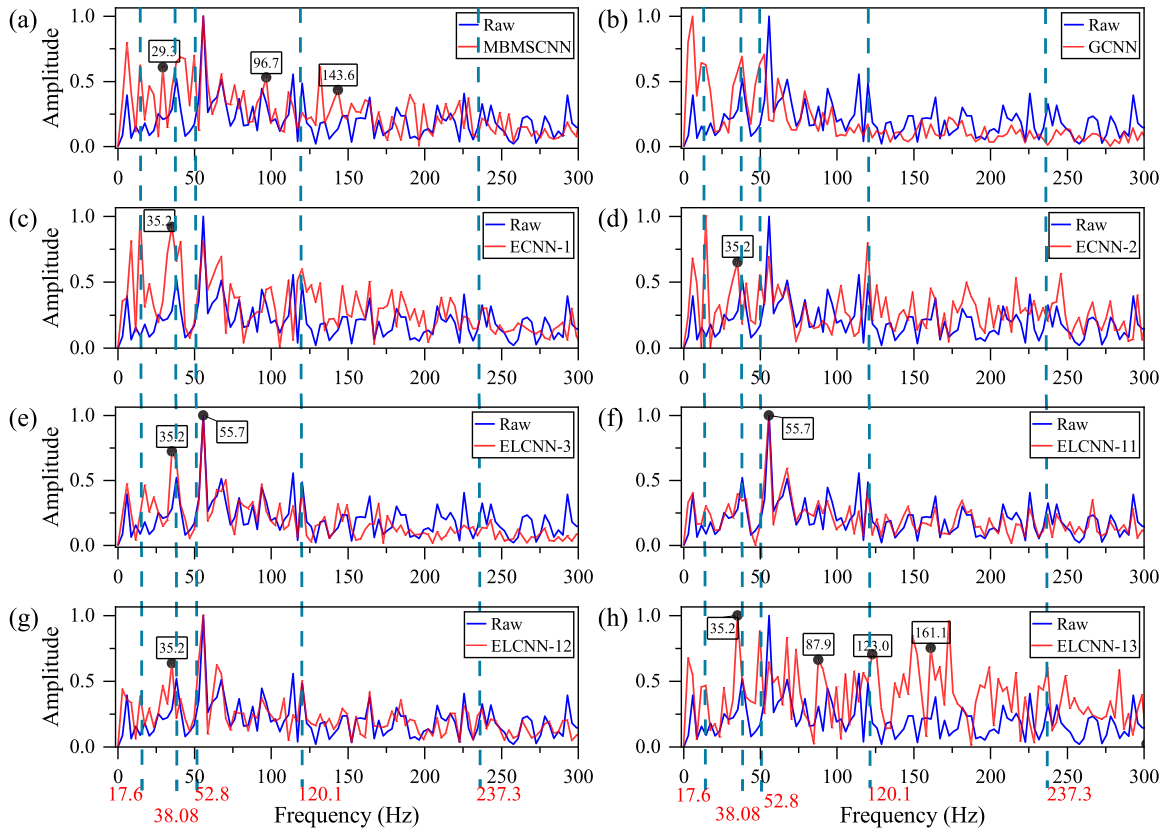


Fig. 27. Frequency-domain features of OR faults in the XJTU dataset learned by the last layer convolution of eight types of 1-D CNNs. (a) MBMSCNN. (b) GCNN. (c) ECNN-1. (d) ECNN-2. (e) ELCNN-3. (f) ELCNN-11. (g) ELCNN-12. (h) ELCNN-13.

The single-layer convolutional model (such as ELCNN-11 and ELCNN-12) does not show better performance than the multilayer convolutional model (such as ECNN-2 and ELCNN-3). However, the square operation of the S-GAP classification layer still has significant feature enhancement advantages. For example, the ELCNN-13 labeled red still has a classification accuracy of 95.95% when the SNR is up to 5 dB.

3) Feature Interpretability: This section mainly analyzes the interpretability of the features extracted by various improved 1-D CNNs based on the IR and OR fault frequency characteristics of the XJTU and verifies the credibility of the conclusions drawn from the CWRU.

a) OR: As shown in the OR fault (119.8 Hz) spectrum learned by different models in Fig. 27, the feature extraction characteristics of the improved 1-D CNN based on the XJTU are similar to those based on the CWTU.

First, the multilayer convolutional GCNN using the pooling layer only focuses on fewer fault frequencies (cage fault: 17.6 Hz; triple frequency: 52.8 Hz; and rotating frequency: 35.2 Hz). Although it can effectively avoid overfitting and improve classification accuracy, it loses fault features (fault frequency greater than 100 Hz).

Second, the multilayer convolution models with removed pooling layers, especially ECNN-1 and ECNN-2, learn more fault features (such as fault frequency: 120.1 Hz). However,

TABLE XIII

CLASSIFICATION ACCURACY OF EIGHT KINDS OF 1-D CNNS FOR XJTU BEARING FAULTS WITH DIFFERENT SNRS. AMONG THEM, THE ECNN-1 MARKED IN BLUE HAS THE WORST ANTINOISE PERFORMANCE, WHILE THE ELCNN-13 MARKED IN RED HAS THE BEST ANTINOISE PERFORMANCE

SNR / Method	15dB	10dB	5dB	0dB	-5dB
MBMSCNN	99.95±0.05	96.63±2.61	43.30±13.19	19.11±8.62	9.16±3.70
GCNN	99.94±0.05	94.00±7.82	56.61±13.05	22.58±8.97	12.20±5.42
ECNN-1	97.99±3.94	85.27±8.21	25.50±7.38	9.11±3.06	6.35±1.93
ECNN-2	99.96±0.03	99.42±0.67	81.76±20.11	30.32±20.91	8.96±3.50
ELCNN-3	99.95±0.07	99.59±0.42	77.30±17.44	29.83±17.53	10.95±5.06
ELCNN-11	99.96±0.05	99.60±0.51	80.28±15.45	26.41±6.93	10.12±4.07
ELCNN-12	99.95±0.06	99.75±0.19	81.74±21.55	33.20±17.56	7.72±3.226
ELCNN-13	100±0	99.91±0.06	95.95±2.22	49.84±0.34	8.96±0.16

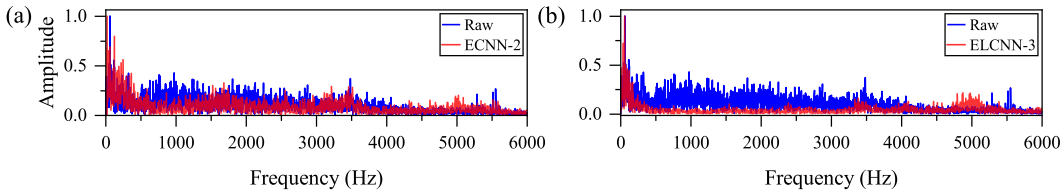


Fig. 28. Frequency-domain features of the OR fault of the XJTU learned by the last layer convolution of (a) ECNN-2 and (b) ELCNN-3, where ECNN-2 uses biases and activation functions, while ELCNN-3 does not use.

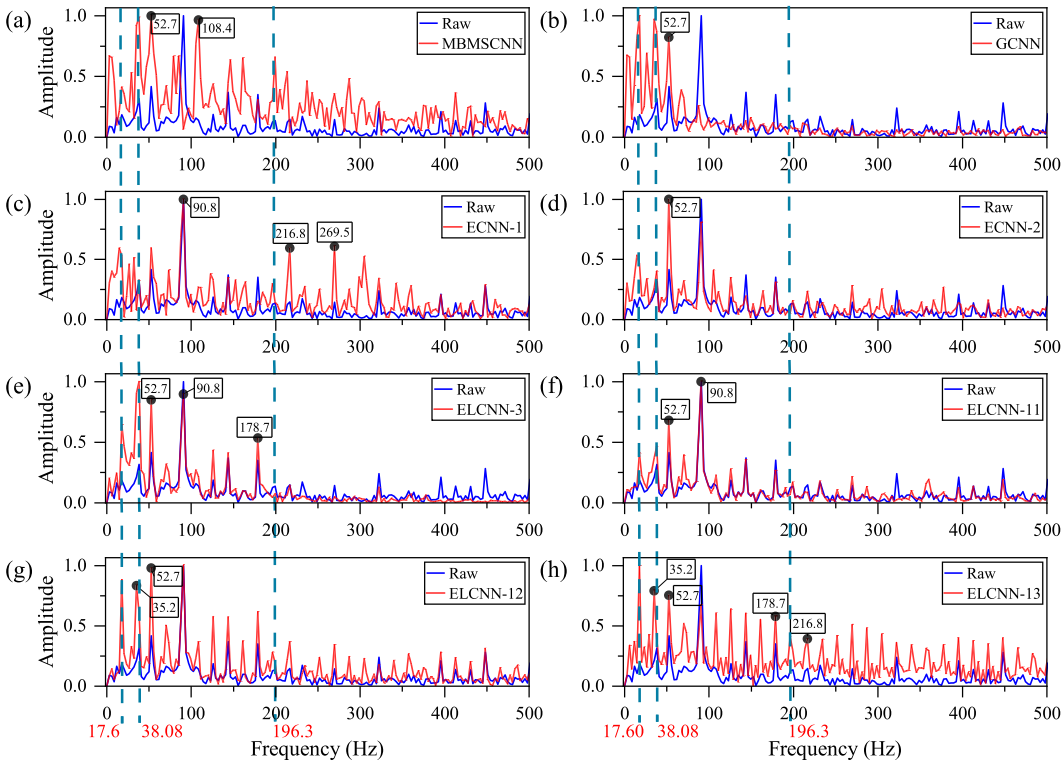


Fig. 29. Frequency-domain features of IR faults in the XJTU dataset learned by the last layer convolution of eight types of 1-D CNNs. (a) MBMSCNN. (b) GCNN. (c) ECNN-1. (d) ECNN-2. (e) ELCNN-3. (f) ELCNN-11. (g) ELCNN-12. (h) ELCNN-13.

ECNN-2 learns more high-frequency nonfault features [see Fig. 28(a)] than ELCNN-3, which removed bias and activation functions.

Finally, the proposed single-layer convolutional ELCNN pays more attention to the low-frequency characteristics; especially, the model using square operation multiple times is

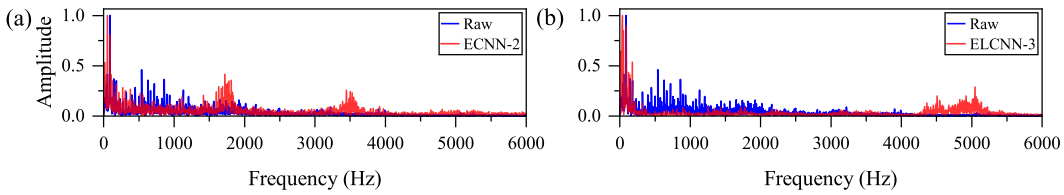


Fig. 30. Frequency-domain features of the IR fault of the XJTU learned by the last layer convolution of (a) ECNN-2 and (b) ELCNN-3, where ECNN-2 uses biases and activation functions, while ELCNN-3 does not use.

easier to extract FM and AM characteristics. For example, ELCNN-13 not only learns the fault frequency of 123 Hz but also learns the sidebands of 87.9 and 161.1 Hz and the double frequency of 237.3 Hz.

In summary, XJTU bearing faults also have complex FM and AM phenomena. The rotating frequency of 35.2 Hz (actual 38.08 Hz) learned by ELCNN-13 is the two times frequency of cage fault of 17.6 Hz, 52.8 Hz is the three times frequency of cage fault, and the fault frequency of 123 Hz is the seven times frequency of cage fault. Therefore, the OR fault of XJTU is obviously affected by the modulation of cage fault.

b) IR: As shown in Figs. 29 and 30, similar to the OR fault, we can obtain the feature extraction characteristics of the 1-D CNN model from the IR fault (196.3 Hz) spectrum, which further confirms that the proposed single-layer convolutional ELCNN model can better fit the fault features. In particular, ELCNN-12 and ELCNN-13, which use square operation many times, can learn more FM and AM features.

From the IR fault learned by ELCNN-13, it can be seen that the double frequency of the cage fault frequency of 17.6 Hz is 35.2 Hz, covering the rotating frequency of 38.08 Hz, and many frequency doublings of 17.6 Hz are generated (such as 11 times the frequency of 17.6 and 193.6 Hz, covering the fault frequency of 196.3 Hz). Therefore, the IR fault of the XJTU is more affected by the cage fault frequency of 17.6 Hz modulation, resulting in regular FM characteristics.

V. Conclusion

We are not limited to the traditional neuron model but return to the essential attributes of fault diagnosis (pattern recognition problem based on feature extraction). A novel interpretable and lightweight ELCNN model is proposed to achieve higher requirements for model storage, computational efficiency, and reliability of airborne and portable devices. First, a single-layer CNN is constructed for vibration signal analysis, and the bias, activation function, and pooling layer of the 1-D CNN feature extraction layer are removed. Then, the S-GAP classification layer with multiple square operations is used to replace the traditional FCL. Finally, the effectiveness and advancement of the proposed method are proved by ablation experiments and comparative experiments.

1) Removing the bias, the activation function and the pooling layer of the feature extraction layer can not only realize the lightweight of the model but also be more conducive to the convolution layer to learn the low-frequency fault features.
2) The multiple square operations of the classification layer S-GAP are beneficial for the model to extract the FM and AM characteristics, avoid the overfitting problem, and improve its antinoise performance and interpretability.

Experiments show that the model optimization idea of removing unnecessary hyperparameters of 1-D CNN by lightweight (improving the operating efficiency of the model and retaining signal features to the greatest extent) and then strengthening and extracting fault features by S-GAP is feasible, which is worthy of reference for other diagnostic models.

References

[1] H. Sun, X. Cao, C. Wang, and S. Gao, "An interpretable anti-noise network for rolling bearing fault diagnosis based on FSWT," *Measurement*, vol. 190, Feb. 2022, Art. no. 110698, doi: 10.1016/j.measurement.2022.110698.

[2] P. Zhang et al., "A novel multiscale lightweight fault diagnosis model based on the idea of adversarial learning," *IEEE Trans. Instrum. Meas.*, vol. 70, pp. 1–15, 2021, doi: 10.1109/TIM.2021.3076841.

[3] P. Xu and Y. Jin, "Stochastic resonance in multi-stable coupled systems driven by two driving signals," *Phys. A, Stat. Mech. Appl.*, vol. 492, pp. 1281–1289, Feb. 2018, doi: 10.1016/j.physa.2017.11.056.

[4] A. Sapena-Baño et al., "Harmonic order tracking analysis: A novel method for fault diagnosis in induction machines," *IEEE Trans. Energy Convers.*, vol. 30, no. 3, pp. 833–841, Sep. 2015, doi: 10.1109/TEC.2015.2416973.

[5] H. Wu, X. Ma, and C. Wen, "Multilevel fine fault diagnosis method for motors based on feature extraction of fractional Fourier transform," *Sensors*, vol. 22, no. 4, p. 1310, Feb. 2022, doi: 10.3390/s22041310.

[6] Y. Cheng, M. Lin, J. Wu, H. Zhu, and X. Shao, "Intelligent fault diagnosis of rotating machinery based on continuous wavelet transform-local binary convolutional neural network," *Knowl.-Based Syst.*, vol. 216, Mar. 2021, Art. no. 106796, doi: 10.1016/j.knosys.2021.106796.

[7] J. Wang, G. Du, Z. Zhu, C. Shen, and Q. He, "Fault diagnosis of rotating machines based on the EMD manifold," *Mech. Syst. Signal Process.*, vol. 135, Jan. 2020, Art. no. 106443, doi: 10.1016/j.ymssp.2019.106443.

[8] X. Liu, Q. Zhou, J. Zhao, H. Shen, and X. Xiong, "Fault diagnosis of rotating machinery under noisy environment conditions based on a 1-D convolutional autoencoder and 1-D convolutional neural network," *Sensors*, vol. 19, no. 4, p. 972, Feb. 2019, doi: 10.3390/s19040972.

[9] J. Shi, Q. He, and Z. Wang, "GMM clustering-based decision trees considering fault rate and cluster validity for analog circuit fault diagnosis," *IEEE Access*, vol. 7, pp. 140637–140650, 2019, doi: 10.1109/ACCESS.2019.2943380.

[10] C. Wang, J. Hu, and C. Wen, "Multi-level PCA and its application in fault diagnosis," in *Proc. 26th Chin. Control Decis. Conf. (CCDC)*, May 2014, Accessed: Jul. 6, 2023, pp. 2810–2814.

[11] M. M. Taheri, H. Seyedi, and B. Mohammadi-Ivatloo, "DT-based relaying scheme for fault classification in transmission lines using MODP," *IET Gener., Transmiss. Distribution*, vol. 11, no. 11, pp. 2796–2804, Aug. 2017, doi: 10.1049/iet-gtd.2016.1821.

[12] Q. Shi and H. Zhang, "Fault diagnosis of an autonomous vehicle with an improved SVM algorithm subject to unbalanced datasets," *IEEE Trans. Ind. Electron.*, vol. 68, no. 7, pp. 6248–6256, Jul. 2021, doi: 10.1109/TIE.2020.2994868.

[13] S. Kiranyaz, O. Avci, O. Abdeljaber, T. Ince, M. Gabbouj, and D. J. Inman, "1D convolutional neural networks and applications: A survey," *Mech. Syst. Signal Process.*, vol. 151, Apr. 2021, Art. no. 107398, doi: 10.1016/j.ymssp.2020.107398.

[14] S. Kiranyaz, T. Ince, R. Hamila, and M. Gabbouj, "Convolutional neural networks for patient-specific ECG classification," in *Proc. 37th Annu. Int. Conf. Eng. Med. Biol. Soc. (EMBC)*, Aug. 2015, pp. 2608–2611, doi: 10.1109/EMBC.2015.7318926.

[15] S. Kiranyaz, A. Gastli, L. Ben-Brahim, N. Al-Emadi, and M. Gabbouj, "Real-time fault detection and identification for MMC using 1-D convolutional neural networks," *IEEE Trans. Ind. Electron.*, vol. 66, no. 11, pp. 8760–8771, Nov. 2019, doi: 10.1109/TIE.2018.2833045.

[16] L. Eren, T. Ince, and S. Kiranyaz, "A generic intelligent bearing fault diagnosis system using compact adaptive 1D CNN classifier," *J. Signal Process. Syst.*, vol. 91, no. 2, pp. 179–189, Feb. 2019, doi: 10.1007/s11265-018-1378-3.

[17] L. Eren, "Bearing fault detection by one-dimensional convolutional neural networks," *Math. Problems Eng.*, vol. 2017, pp. 1–9, 2017, doi: 10.1155/2017/8617315.

[18] X. Shao, L. Wang, C. S. Kim, and I. Ra, "Fault diagnosis of bearing based on convolutional neural network using multi-domain features," *KSII Trans. Internet Inf. Syst.*, vol. 15, no. 5, pp. 1610–1629, 2021, doi: 10.3837/tiis.2021.05.002.

[19] J. Liu, M. Zhang, H. Wang, W. Zhao, and Y. Liu, "Sensor fault detection and diagnosis method for AHU using 1-D CNN and clustering analysis," *Comput. Intell. Neuroscienc*, vol. 2019, Sep. 2019, Art. no. 5367217, doi: 10.1155/2019/5367217.

[20] H. Zhang, P. Chen, and Q. Wang, "Retracted: Fault diagnosis method based on EEMD and multi-class logistic regression," in *Proc. 3rd Int. Conf. Smart City Syst. Eng. (ICSCSE)*, Dec. 2018, pp. 859–863, doi: 10.1109/ICSCSE.2018.00185.

[21] H. Wu et al., "One-dimensional CNN-based intelligent recognition of vibrations in pipeline monitoring with DAS," *J. Lightw. Technol.*, vol. 37, no. 17, pp. 4359–4366, 2019, doi: 10.1109/JLT.2019.2923839.

[22] D. Peng, Z. Liu, H. Wang, Y. Qin, and L. Jia, "A novel deeper one-dimensional CNN with residual learning for fault diagnosis of wheelset bearings in high-speed trains," *IEEE Access*, vol. 7, pp. 10278–10293, 2019, doi: 10.1109/ACCESS.2018.2888842.

[23] J. He, P. Wu, Y. Tong, X. Zhang, M. Lei, and J. Gao, "Bearing fault diagnosis via improved one-dimensional multi-scale dilated CNN," *Sensors*, vol. 21, no. 21, p. 7319, Nov. 2021, doi: 10.3390/s21217319.

[24] H. Wang, Z. Liu, D. Peng, and Y. Qin, "Understanding and learning discriminant features based on multiattention 1DCNN for wheelset bearing fault diagnosis," *IEEE Trans. Ind. Informat.*, vol. 16, no. 9, pp. 5735–5745, Sep. 2020, doi: 10.1109/TII.2019.2955540.

[25] Z. Ye and J. Yu, "AKSNet: A novel convolutional neural network with adaptive kernel width and sparse regularization for machinery fault diagnosis," *J. Manuf. Syst.*, vol. 59, pp. 467–480, Apr. 2021, doi: 10.1016/j.jmsy.2021.03.022.

[26] Y. Chen, D. Zhang, H. Ni, J. Cheng, and H. R. Karimi, "Multi-scale split dual calibration network with periodic information for interpretable fault diagnosis of rotating machinery," *Eng. Appl. Artif. Intell.*, vol. 123, Aug. 2023, Art. no. 106181, doi: 10.1016/j.engappai.2023.106181.

[27] 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, doi: 10.1109/JSEN.2022.3169341.

[28] Y. Shi, A. Deng, M. Deng, J. Zhu, Y. Liu, and Q. Cheng, "Enhanced lightweight multiscale convolutional neural network for rolling bearing fault diagnosis," *IEEE Access*, vol. 8, pp. 217723–217734, 2020, doi: 10.1109/ACCESS.2020.3041735.

[29] G. Li et al., "An explainable one-dimensional convolutional neural networks based fault diagnosis method for building heating, ventilation and air conditioning systems," *Building Environ.*, vol. 203, Oct. 2021, Art. no. 108057, doi: 10.1016/j.buildenv.2021.108057.

[30] M. S. Kim, J. P. Yun, and P. Park, "An explainable convolutional neural network for fault diagnosis in linear motion guide," *IEEE Trans. Ind. Informat.*, vol. 17, no. 6, pp. 4036–4045, Jun. 2021, doi: 10.1109/TII.2020.3012989.

[31] H. Tao, P. Wang, Y. Chen, V. Stojanovic, and H. Yang, "An unsupervised fault diagnosis method for rolling bearing using STFT and generative neural networks," *J. Franklin Inst.*, vol. 357, no. 11, pp. 7286–7307, Jul. 2020, doi: 10.1016/j.jfranklin.2020.04.024.

[32] J. Chen, J. Pan, Z. Li, Y. Zi, and X. Chen, "Generator bearing fault diagnosis for wind turbine via empirical wavelet transform using measured vibration signals," *Renew. Energy*, vol. 89, pp. 80–92, Apr. 2016, doi: 10.1016/j.renene.2015.12.010.

[33] Z. Liu, D. Peng, M. J. Zuo, J. Xia, and Y. Qin, "Improved Hilbert–Huang transform with soft sifting stopping criterion and its application to fault diagnosis of wheelset bearings," *ISA Trans.*, vol. 125, pp. 426–444, Jun. 2022, doi: 10.1016/j.isatra.2021.07.011.

[34] Z. J. Wang et al., "CNN explainer: Learning convolutional neural networks with interactive visualization," *IEEE Trans. Vis. Comput. Graphics*, vol. 27, no. 2, pp. 1396–1406, Feb. 2021, doi: 10.1109/TVCG.2020.3030418.

[35] E. Bochinski, T. Senst, and T. Sikora, "Hyper-parameter optimization for convolutional neural network committees based on evolutionary algorithms," in *Proc. IEEE Int. Conf. Image Process. (ICIP)*, Sep. 2017, pp. 3924–3928, doi: 10.1109/ICIP.2017.8297018.

[36] Q. Sun, X. Yu, and H. Li, "Open-circuit fault diagnosis based on 1D-CNN for three-phase full-bridge inverter," in *Proc. 11th Int. Conf. Prognostics Syst. Health Manage. (PHM-Jinan)*, Oct. 2020, pp. 322–327, doi: 10.1109/PHM-Jinan48558.2020.00064.

[37] W. You, C. Shen, D. Wang, L. Chen, X. Jiang, and Z. Zhu, "An intelligent deep feature learning method with improved activation functions for machine fault diagnosis," *IEEE Access*, vol. 8, pp. 1975–1985, 2020, doi: 10.1109/ACCESS.2019.2962734.

[38] Z. Pan, Y. Wang, K. Wang, G. Ran, H. Chen, and W. Gui, "Layer-wise contribution-filtered propagation for deep learning-based fault isolation," *Int. J. Robust Nonlinear Control*, vol. 32, no. 17, pp. 9120–9138, Nov. 2022, doi: 10.1002/rnc.6328.

[39] W. Zhang, G. Peng, and C. Li, "Bearings fault diagnosis based on convolutional neural networks with 2-D representation of vibration signals as input," in *Proc. MATEC Web Conf.*, vol. 95, 2017, p. 13001, doi: 10.1051/matecconf/20179513001.

[40] 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, doi: 10.1109/TIE.2017.2774777.

[41] M. Lin, Q. Chen, and S. Yan, "Network in network," 2014, *arXiv:1312.4400*.

[42] B. Zhou, A. Khosla, A. Lapedriza, A. Oliva, and A. Torralba, "Learning deep features for discriminative localization," 2015, *arXiv.1512.04150*.

[43] Z. Wang, Q. Liu, H. Chen, and X. Chu, "A deformable CNN-DLSTM based transfer learning method for fault diagnosis of rolling bearing under multiple working conditions," *Int. J. Prod. Res.*, vol. 59, no. 16, pp. 4811–4825, Aug. 2021, doi: 10.1080/00207543.2020.1808261.

[44] M. Wang, W. Wang, X. Zhang, and H. H.-C. Iu, "A new fault diagnosis of rolling bearing based on Markov transition field and CNN," *Entropy*, vol. 24, no. 6, p. 751, May 2022, doi: 10.3390/e24060751.

[45] J. O. D. van den Hoogen, S. D. Bloemheuvel, and M. Atzmueller, "An improved Wide-Kernel CNN for classifying multivariate signals in fault diagnosis," in *Proc. Int. Conf. Data Mining Workshops (ICDMW)*, Nov. 2020, pp. 275–283, doi: 10.1109/ICDMW51313.2020.00046.

[46] D. Yao, H. Liu, J. Yang, and X. Li, "A lightweight neural network with strong robustness for bearing fault diagnosis," *Measurement*, vol. 159, Jul. 2020, Art. no. 107756, doi: 10.1016/j.measurement.2020.107756.

[47] 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.*, vols. 64–65, pp. 100–131, Dec. 2015, doi: 10.1016/j.ymssp.2015.04.021.

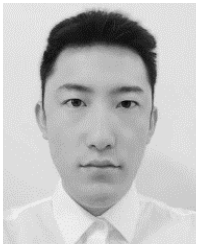
Pengfei Pang was born in Shanxi, China, in 1985. He received the B.S. degree from the Taiyuan University of Technology, Taiyuan, China, in 2007, and the M.S. degree from the Beijing Institute of Technology, Beijing, China, in 2018. He is currently pursuing the Ph.D. degree with Army Engineering University, Nanjing, China.

His research interests include mechanical fault diagnosis algorithms based on artificial intelligence and computer vision technology.

Jian Tang was born in Chongqing, China, in 1977. She received the Ph.D. degree from the PLA University of Science and Technology, Nanjing, China, in 2013.

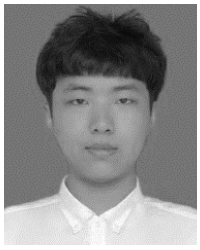
She is currently a Professor with Army Engineering University, Nanjing. She has written or copublished more than 25 papers [Elsevier Engineering Information (EI)/Science Citation Index (SCI)]. She has presided over the National Natural Science Foundation of China and the Natural Science Foundation of Jiangsu Province. She published three books and monographs, and obtained seven patents and five software copyrights. Her current research interests are artificial intelligence algorithms and their applications in mechanical fault diagnosis.

Dr. Tang is a member of China Computer Federation (CCF).



Jiqing Luo was born in Shanxi, China, in 1996. He received the B.S. degree from the Taiyuan University of Technology, Taiyuan, China, in 2019, and the M.S. degree from Army Engineering University, Nanjing, China, in 2021, where he is currently pursuing the Ph.D. degree.

His current research interests include image processing, machine learning, and pattern recognition.



Miao Chen was born in Jiangsu, China, in 2000. He received the B.S. degree from the Nanjing University of Technology, Nanjing, China, in 2022, where he is currently pursuing the M.S. degree in machinery.

His research interests include mechanical fault diagnosis, deep learning, and machine learning.

Hui Yuan was born in Hunan, China, in 1972. He received the Ph.D. degree from the PLA University of Science and Technology, Nanjing, China, in 2013.

He is currently a Professor with the Nanjing University of Technology, Nanjing, where he is also the Vice-President of the Advanced Equipment Manufacturing Research Institute. He has written or coauthored more than 40 publication articles [Science Citation Index (SCI)/Elsevier Engineering Information (EI)], published five books and monographs, and obtained five software copyrights. He presided over 15 scientific research projects, such as the subproject of the National Science and Technology Support Program, the Natural Science Foundation of Jiangsu Province, and scientific research in the military. His research interests are safety monitoring of engineering structures and computer vision technology.

Lei Jiang was born in Anhui, China, in 1994. She received the B.S. degree from the Hefei University of Technology, Hefei, China, in 2017, and the M.S. degree from Army Engineering University, Nanjing, China, in 2013.

Her research interests include deep learning, mechanical fault diagnosis, and mechanical information processing.